Moroccan Journal of Algebra and Geometry with Applications

Volume 2, Issue 1 (2023), pp 1-173

ISSN: 2820-7114

Supported by:

Sidi Mohamed Ben Abdellah University, Fez, Morocco

جامعة سيدي محمد بن عبد الله بفاس

UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

Moroccan Journal of Algebra
and Geometry with Applications
Volume 2 (1) (2023)
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

Moroccan Journal of Algebra
and Geometry with Applications
Volume 2 (1) (2023)
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

The Moroccan Journal of Algebra and Geometry with Applications publishes two issues per year composed of high-quality original research papers with significant contributions in the fields of algebra, geometry and related fields. It publishes also invited survey articles on hot topics, from distinguished mathematicians from across the world. All published research articles are subject to rigorous peer review, based on initial screening by editors, refereeing by independent expert referees, and consequent revision by article authors when required. The published article constitutes the final and definitive version of the work. All manuscripts submitted to the journal must be original contributions, and must not be under consideration for publication with another journal, nor have been previously published, in part or whole.

Moroccan Journal of Algebra
and Geometry with Applications
Volume 2 (1) (2023)
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

Author Guidelines / Submissions

Manuscript Submission :

Submission of a manuscript implies that

- the work described has not been published before;
- it is not under consideration for publication anywhere else;
- its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out.

Articles must be written in English or, exceptionally, in French and prepared in Latex at MJAGA format. The format can be found here (https://ced.fst-usmba.ac.ma/p/mjaga/author-guidelines-submissions). Articles may be submitted directly via email as a PDF-file to any of the Journal Editors.

The manuscript should contain :

- A concise and informative title.
- The name(s) of the author(s), affiliation(s) of the author(s), a clear indication and an active e-mail address of the corresponding author, and the 16-digit ORCID of the author(s), if available.
- An abstract of 150 to 250 words.
- 4 to 6 keywords which can be used for indexing purposes.
- Appropriate numbers of MSC codes.

Peer review :

The submission will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of one independent expert reviewer to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final.

After acceptance :

After the paper is accepted for publication, the corresponding author will be required to submit the LaTeX file. Final formatting matters will follow, and publication will then be immediate.

Publication Charges :

There are no submission fees, publication fees or page charges for this journal.

Moroccan Journal of Algebra
and Geometry with Applications
Volume 2, Issue 1, June 2023
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

Contents :

Remembering Muhammad Zafrullah

By Ayman Badawi
Department of Mathematics & Statistics
The American University of Sharjah
United Arab Emirates
abadawi@aus.edu

I was asked to write a brief biography of the commutative algebraist Muhammad Zafrullah, a Pakistani nationalist who died on July 28, 2022. I shall limit my writing to discussing his impact on commutative algebra. Muhammad Zafrullah’s journey began in 1974 when he graduated from the University of London with a Ph.D. in mathematics under the guidance of Paul Moritz Cohn. To the best of my knowledge, he worked in Libya for a while after graduating, moved to the United States, and remained there until his passing. In the United States of America, he met Daniel D. Anderson, a well-known commutative ring theorist who passed away on April 24, 2022, and the two co-authored several fundamental papers. The multiplicative ideal theory, a subfield of commutative algebra that attracted many algebraists, is the key area in which Muhammad Zafrullah made contributions. To name a few of these contributions: A few examples include the D+M construction, t-invertibility, almost prufer domain, factorization in integral domains, t-ideals, t-locally domains, h-local domains, GCD-domains, and many more. His high citations, for instance, have an h-index of 33 on Google Scholar and total of 3746 citations, which serve as evidence of his effect in commutative algebra. Muhammad co-authored numerous publications with well-known algebraists, including Daniel D. Anderson, David F. Anderson, Alain Bouvier, Gyu Whan Chang, Scott Chapman, Douglas Costa, Jim Coykendall, David E. Dobbs, Tiberiu Dumitrescu, Said Elbaghdadi, Marco Fontana, Stefania Gabelli, Robert Gilmer, Franz Halter-Koch, Evan Houston, Dong Je Kwak, Heakyung Lee, Yves Lequain, Thomas Lucas, Joe Mott, Moshe Roitman. A few months before his natural death, he continued to appreciate and be passionate about research. He is a co-author of four articles that were published in 2022. The reader can check the following links for more information about Muhammad Zafrullah's research.
https://scholar.google.ae/citations?hl=en&user=56pdMcUAAAAJ
https://mathscinet.ams.org/mathscinet/search/author.html?mrauthid=186170
https://www.scopus.com/authid/detail.uri?authorId=6701711532

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco

Volume 2, Issue 1 (2023), pp 1-13

ISSN: 2820-7114

Title :

On the possible inequalities involving the mean, the median and the mode in a divisible ordered abelian group

Author(s):

David E. Dobbs

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 1–13

Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco **ISSN: 2820-7114**

On the possible inequalities involving the mean, the median and the mode in a divisible ordered abelian group

David E. Dobbs
Department of Mathematics, University of Tennessee, Knoxville, Tennessee 37996-1320
e-mail: *ddobbs1@utk.edu*

Communicated by Najib Mahdou
(Received 15 September 2022, Revised 10 October 2022, Accepted 02 November 2022)

Abstract. Let $(G, \leq)$ be a nonzero additive divisible ordered abelian group. By a *list* L (*of length* n) *in* G, we mean a finite multiset $x_1, \ldots, x_n$ with each $x_i \in G$ (and where possibly $x_i = x_j$ for some $i \neq j$) for some $n \geq 2$ such that L has a unique mode. There are exactly 13 (pairwise distinct) possible continued inequalities involving the mean, the median and the mode, along with the relations < and/or = (such as "mean < median = mode") that a given list in G may satisfy. For each of these 13 situations, it is proved that there exists a list in G that satisfies that inequality, the minimal length of such a list is determined and is at most 6, infinitely many lists in G are constructed that each satisfy that minimal length (for the given inequality), at least one of those lists also has the property that its minimal entry is any preassigned element of G, and the minimal length of a satisfying list (for the given inequality) is independent of G. Because of the first, second and fifth of these facts, the search for a suitable list (satisfying a given inequality) can be restricted to the case $G = \mathbb{Q}$ (under addition). For that setting, a proof, programming assignment, or discovery activity could be carried out in courses at various levels.

Key Words: Mean, median, mode, ordered abelian group, torsionfree, divisible abelian group.
2010 MSC: Primary 06F20, 62A01; Secondary 13C11, 20K99.

In memory of Muhammad Zafrullah and his interest in Riesz groups

1 Introduction

Because of the widespread use of statistics, the current curriculum for middle school and high school introduces students to the concepts of the mean, the median and the mode of a nonempty finite set of rational numbers. As each of these concepts is often introduced as a "center-indicating measure" of a finite set (or of a finite mutiset, that is, a finite list), some students are led to ask about the possible inequalities that could involve the three concepts of mean, median and mode. Our interest here is in answering all such questions in an organized way.

It turns out that the methods (and the answers) that work for rational numbers also apply, with no essential changes, when the context is generalized (from $\mathbb{Q}$) to working with the elements of a nonzero additive divisible ordered abelian group G. Since any such group is torsionfree, the "divisible" and "ordered abelian group" conditions allow the mean and median of any finite list in G to be unambiguously defined. In that regard, recall that the (*arithmetic*) *mean* of a list $\mathcal{L}$ (given by $x_1, \ldots, x_n$) is defined to be $(\sum_{1=1}^{n} x_i)/n$. Also, by convention, the *median* of a list $\mathcal{L}$ (given by $x_1, \ldots, x_n$) is obtained by first permuting (the entries in) $\mathcal{L}$ so that the list becomes weakly increasing (also known as non-decreasing), in the usual sense that $x_{i-1} \leq x_i$ whenever $2 \leq i \leq n$; then the median of such a weakly increasing list $\mathcal{L}$ is defined to be

$$\frac{x_k + x_{k+1}}{2} \text{ if } n = 2k \text{ is even, and } x_k \text{ if } n = 2k+1 \text{ is odd.}$$

To avoid trivialities (that is, singleton lists), we will require that **all lists under consideration here have some finite** *length* $n \geq 2$. Also, to simplify matters, each list $\mathcal{L}$ (given by $x_1, \ldots, x_n$) under consideration here will be assumed to have a **unique mode** (that is, a unique value $w \in G$ so that the cardinality of the set $\{i \mid x_i = w \text{ and } 1 \leq i \leq n\}$ is maximal). A last bit of terminology: if $\{\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3\} = \{\text{mean}, \text{median}, \text{mode}\}$, the relations $\aleph_1$ and $\aleph_2$ are each elements of $\{<, =\}$, and the list $\mathcal{L} : x_1, \ldots, x_n$ (of length $n \geq 2$) in G satisfies the continued inequality "$\mathrm{M}_1 \ \aleph_1 \ \mathrm{M}_2 \ \aleph_2 \ \mathrm{M}_3$", we will say that $\mathcal{L}$ is an *optimal list* (for the given continued inequality) if, in addition, the following two conditions hold: $\mathcal{L}$ is weakly increasing; and n is minimal among the lengths of solutions for the given continued inequality.

For G as above and for each of the 13 possible inequalities of the form "$\mathrm{M}_1 \ \aleph_1 \ \mathrm{M}_2 \ \aleph_2 \ \mathrm{M}_3$", our main conclusions may be summarized as follows: we construct infinitely many optimal lists $\mathcal{L}$ (given by $x_1, \ldots, x_n$ in G) for the given continued inequality; the length of any such $\mathcal{L}$ is determined, it is at most 6, and it is independent of G (so, all relevant calculations can be done in $\mathbb{Q}$); for each nonzero element $g \in G$, some optimal list has its minimal element $x_1 = g$ and each of its elements x_i can be expressed as $m_i g$ for some integers m_i. In case G is an additive subgroup of $\mathbb{Q}$ (with the usual ordering) that contains $\mathbb{Z}$, we can also construct infinitely many optimal lists such that each x_i is a positive integer (and at least one of those lists is also such that $x_1 = 1$). Finally, note that Remark 2.5 collects several comments about a variety of theoretical and/or pedagogic matters.

2 Results

We begin with some background that will explain the comment in the Introduction that the mean and the median of a list (in a nonzero additive divisible ordered abelian group G) were "unambiguously defined." The following comments will also serve to justify the calculations involving the "fractional" notation in the proofs of Lemma 2.1, Theorem 2.2 and Proposition 2.3.

Suppose, for the moment, that an additive group H is divisible. Recall that this means, by definition, that if $h \in H$ and n is a nonzero integer, then there exists $f \in H$ such that $nf = h$. One's reflexes stemming from childhood experiences in solving linear equations lead naturally to asking whether one could/should denote such an element f via the notation h/n. A use of such notation may well suggest that f is uniquely determined by h and n. While that kind of uniqueness need not be the case in general, it *does* hold if H is also assumed to be torsionfree (for instance, if H is also assumed to be an ordered abelian group). Thus, if H is an additive divisible torsionfree abelian group, then the "fractional notation" h/n identifies a uniquely determined element of H (such that $n(h/n) = h$), for all $h \in H$ and $0 \neq n \in \mathbb{Z}$. In the hint for [5, Exercise 8, page 198], Hungerford goes on to explain how this fractional notation can be extended so that H (satisfying the above hypotheses) is viewed as a vector space over $\mathbb{Q}$ (in a way that, essentially, defines $(m/n)h$ to be $m(h/n)$ for all $h \in H$ and $m, n \in \mathbb{Z}$ with $n \neq 0$). In fact, the copious hints in [5, Exercises 8-11, pages 198-199] sketch a development of the entire structure theory of divisible abelian groups, including the fact that any additive divisible torsionfree group is isomorphic (as an abelian group) to a direct sum of copies of $\mathbb{Q}$ (the cardinality of whose index set is uniquely determined). This point of view permits and justifies calculations of the following kind. Consider $d, x_1, \ldots, x_n \in H$ and nonzero integers p, k. Then

$$\frac{\sum_{i=1}^n p(d + x_i)}{k} = \frac{(\sum_{i=1}^n pd) + (\sum_{i=1}^n px_i)}{k} =$$

$$\frac{npd + p\sum_{i=1}^n x_i}{k} = (\frac{np}{k})d + (\frac{p}{k})\sum_{i=1}^n x_i.$$

The just-displayed kind of calculations will be fundamental in verifying several steps in the proofs of Lemma 2.1, Theorem 2.2 and Proposition 2.3.

The above-mentioned theory of the structure of divisible abelian groups can be subsumed in a more general structure theory, due to Matlis [7], of the structure of injective modules over a commutative Noetherian ring R. According to that theory (whose uniqueness aspect need not concern us here), each injective R-module (for R as above) is isomorphic to a direct sum of R-modules that are each an injective envelope of an R-module R/P as P varies (and is possibly listed with repetition) over certain prime ideals of R. The application of Matlis' theory to abelian groups (that is, to $\mathbb{Z}$-modules) is due to the following facts: if R is a (commutative unital) integral domain, then each injective R-module is divisible (cf. [9, Theorem 3.23]); and each divisible R-module is an injective R-module if and only if R is a Dedekind domain [3, Proposition 5.1, page 134]; principal ideal domain $\Rightarrow$ Dedekind domain $\Rightarrow$ commutative Noetherian ring; $\mathbb{Z}$ is a principal ideal domain; the injective envelope, as a $\mathbb{Z}$-module, of $\mathbb{Z}$ (resp., of $\mathbb{Z}/p\mathbb{Z}$ for some prime number p) is, up to isomorphism, $\mathbb{Q}$ (resp., $\mathbb{Z}(p^\infty)$). Consequently, by specializing to the case $R = \mathbb{Z}$, we get that every divisible abelian group is isomorphic to a direct sum of abelian groups that are each of the form either $\mathbb{Q}$ or $\mathbb{Z}(p^\infty)$ (possibly listed with repetition). As any $\mathbb{Z}(p^\infty)$ (or, for that matter, any $\mathbb{Z}/p\mathbb{Z}$) is a *torsion* abelian group, we thus recover the conclusion that the divisible torsionfree abelian groups are the same, up to isomorphism, as the direct sums of copies of $\mathbb{Q}$ (with index sets of arbitrary cardinality). By the same reasoning, we get, more generally, that if R is a principal ideal domain with quotient field K, then the divisible torsionfree R-modules M are, up to R-module isomorphism, the direct sums of copies of K (with index sets of arbitrary cardinality). It follows that any such M is a vector space over K (and, as a special case, we recover the fact that any divisible torsionfree abelian group is a vector space over $\mathbb{Q}$). Indeed, since $K \otimes_R K \cong K$ and tensor product commutes with arbitrary direct sums of modules, we get that $M \cong K \otimes_R M$, which is naturally a vector space over K.

Since the additive group of real numbers is a nonzero additive divisible ordered abelian group, it follows from the above comments that $\mathbb{R}$ is isomorphic, as an abelian group under addition, to $\oplus_I \mathbb{Q}$, a direct sum of copies of $\mathbb{Q}$ indexed by a set I. Of course, this conclusion also holds because I is essentially a $\mathbb{Q}$-vector space basis of $\mathbb{R}$ (that is, a *Hamel basis* of $\mathbb{R}$ over $\mathbb{Q}$). As we are assuming ZFC foundations, the usual rules for arithmetic with cardinal numbers can be used to show that $|I|$ (which denotes the cardinal number of I) satisfies

$$c := |\mathbb{R}| = |\oplus_I \mathbb{Q}| = \aleph_0 \cdot |I| = |I|;$$

that is, $|I| = 2^{\aleph_0}$. While $\mathbb{Q}$ can be viewed as an additive subgroup of $\oplus_I \mathbb{Q}$ (and, hence, of $\mathbb{R}$) in uncountably many ways, cardinality considerations show that there is no way to embed a sufficiently "large" additive divisible ordered abelian group G, when viewed up to isomorphism as $\oplus_J \mathbb{Q}$ for some "large" index set J, as a subgroup of $\mathbb{R}$. (Indeed, if $|J| > c$, then $|G| > c$.) Some of the ordered abelian subgroups of $\mathbb{R}$ will be discussed further in parts (f) and (g) of Remark 2.5.

We begin with a lemma whose parts (a), (b), (c) and (d) will allow us to convert an optimal list (for a given continued inequality) into an optimal list (for the same inequality) that has a presassigned element of G as its minimal entry. Moreover, in case G is an additive subgroup of $\mathbb{Q}$ with the usual ordering, Lemma 2.1 (e) will allow us (once Theorem 2.2 has been proved) to produce, for each of the 13 possible continued inequalities, an optimal list consisting of positive integers.

As usual, if H is an additive (semi)group, with $h \in H$ and and p a positive integer, then ph denotes the sum of p copies of h (in H).

Lemma 2.1. *Let $(G, \leq)$ be a nonzero additive divisible ordered abelian group. Let $L_1 : x_1, \dots, x_n$ be a list (of length $n \geq 2$) in G with a unique mode. Let $d \in G$; and let p be a positive integer. Consider the multisets*

$$L_2 : x_1 + d, \dots, x_n + d \quad \text{and} \quad L_3 : px_1, \dots, px_n.$$

Then:

(*a*) *L_2 and L_3 are each a list of length n in G with a unique mode.*

(*b*) *L_2 is weakly increasing $\Leftrightarrow$ L_1 is weakly increasing $\Leftrightarrow$ L_3 is weakly increasing.*

(*c*) $x_1 + d = \min(\{x_j + d \mid 1 \le j \le n\}) \Leftrightarrow x_1 = \min(\{x_j \mid 1 \le j \le n\}) \Leftrightarrow px_1 = \min(\{px_j \mid 1 \le j \le n\})$.

(*d*) *The mean* (*resp., the median; resp., the mode*) *of L_2 is the sum of d and the mean* (*resp., the median; resp., the mode*) *of L_1.*

(*e*) *The mean* (*resp., the median; resp., the mode*) *of L_3 is the product of p and the mean* (*resp., the median; resp., the mode*) *of L_1.*

Proof. The reader can easily verify that the hypotheses on G ensure that if $x, y, x_1, \dots, x_n \in G$, then the following six assertions hold:
(i) $x + d = y + d \Leftrightarrow x = y \Leftrightarrow px = py$;
(ii) $x + d \le y + d \Leftrightarrow x \le y \Leftrightarrow px \le py$;
(iii) $\frac{\sum_{i=1}^{n}(x_i+d)}{n} = d + \frac{\sum_{i=1}^{n} x_i}{n}$;
(iv) $\frac{(x+d)+(y+d)}{2} = d + \frac{x+y}{2}$;
(v) $\frac{\sum_{i=1}^{n} px_i}{n} = p(\frac{\sum_{i=1}^{n} x_i}{n})$;
(vi) $\frac{px+py}{2} = p(\frac{x+y}{2})$.

(a) It is clear from the definitions that L_2 and L_3 each inherit from L_1 the property of being a list in G of length n. Also, it follows from the first (resp., second) equivalence in (i) that L_2 (resp., L_3) inherits from L_1 the property of having a unique mode.

(b), (c): The first (resp., second) equivalence in each of these assertions follows from the first (resp., second) equivalence in (ii).

(d) The assertion concerning the means (resp., medians; resp., modes) follows from (iii) (resp., from (iv); resp, from the first equivalence in (i)).

(e) The assertion concerning the means (resp., medians; resp., modes) follows from (v) (resp., from (vi); resp., from the second equivalence in (i)). The proof is complete. □

In any nonzero ordered ring R (such as $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$), the fact that $(-1)^2 = 1 = 1^2$, together with the axioms of an ordered ring, ensures that $1 > 0$ and, consequently, that any positive integer (when viewed as an element of R) is positive. Thus, in such a ring, there is no ambiguity as to the meaning of a "positive integer." That observation leads to the fact that $\mathbb{Z}$ has a unique structure as a (complete) ordered integral domain, and it follows easily that $\mathbb{Q}$ has a unique structure as an ordered field (with m/n being deemed "positive", when m and n are nonzero integers, if and only if either both m and n are positive or both m and n are negative, equivalently, if and only if mn is positive: cf. [2, Theorem 12, page 12]). Along those lines, it is interesting to note that $\mathbb{R}$ has a unique structure as a (complete) ordered field (cf. [2, Theorem 7, page 99; Exercise 9, page 100]). However, the situation is different for ordered abelian *groups*. For instance, there are two distinct ways to view the additive abelian group $\mathbb{Z}$ as an ordered abelian group, according as to whether the integer 1 is deemed to be positive or negative. Although the resulting two distinct structures of $\mathbb{Z}$ as an ordered abelian group are isomorphic (in the category of ordered abelian groups), we will adopt the following convention in this paper: **in any ordered abelian group containing $\mathbb{Z}$ as an additive subgroup, we assume that** $1 > 0$. One upshot of this convention (which can be stated informally as "The integer 1 is always positive") is the following: if G is an additive divisible ordered abelian group, with $0 \ne h \in G$ and n a positive integer, then h/n is positive (resp., negative) in G if and only if h is positive (resp., negative) in G. This convention also has other advantages: it accords with common terminology and facts from the context of ordered rings; and, up to isomorphism, its adoption loses nothing. Without further comment, this convention will be used in the proofs of Theorem 2.2 (e) and Proposition 2.3.

Recall that, prior to Lemma 2.1, we defined the symbol "ph", in case p is a positive integer and h is an element of some additive (semi)group H, essentially as follows: $ph := 1 + \cdots + 1$, the sum of p copies of h in H. In case the additive group $\mathbb{Z}$ is a sub(semi)group of H and H admits a binary operation of multiplication (which need not lead to H being a ring), there is another reasonable meaning for the symbol "ph" that was not necessarily intended earlier, namely, as the product of p and h in H. If multiplication distributes over addition in H, then (with the above assumption that $\mathbb{Z}$ is a subsystem of H) the two possible interpretations of the symbol "ph" agree (for all p and h as above) if and only if $1h = h$ for all $h \in H$. If H is a ring whose multiplication has neutral element 1, this condition is certainly satisfied. However, this condition would not be satisfied in the following situation: use the classical method, as in [8, page 8], of embedding $\mathbb{Z}$ as a subsystem of a "ring with unity" H. We suggest that this observation seems to vindicate our earlier choice of definition for the symbol "ph" and the reasonableness of the convention in the preceding paragraph.

Many of the details of the necessary calculations involving the "fractional" notation h/n in the proofs of the following results will be left to the reader, as the rationale for those details was explained at the beginning of this section and the reader has now had experience in carrying out such calculations (while verifying the assertions (iii)-(vi) in the proof of Lemma 2.1).

Our main conclusions are collected in the following three results.

Theorem 2.2. Let G be a nonzero additive divisible ordered abelian group. Let $g > 0$ in G. Then:

(a) Each of the 13 possible continued inequalities in G involving the mean, the median, and the mode, together with the relations $<$ and/or $=$, has an optimal list in G whose first entry is g and whose length is at most 6. In detail:

(i) For the continued inequality mean $<$ median $<$ mode: one optimal list (in G) with first entry g and length $n = 4$ is $g, 5g, 7g, 7g$;

(ii) For the continued inequality mean $<$ median $=$ mode: one optimal list (in G) with first entry g and length $n = 3$ is $g, 4g, 4g$;

(iii) For the continued inequality mean $=$ median $<$ mode: one optimal list (in G) with first entry g and length $n = 5$ is $g, 3g, 4g, 6g, 6g$;

(iv) For the continued inequality mean $=$ median $=$ mode: one optimal list (in G) with first entry g and length $n = 2$ is g, g;

(v) For the continued inequality mean $<$ mode $<$ median: one optimal list (in G) with first entry g and length $n = 6$ is $g, 16g, 16g, 18g, 19g, 20g$;

(vi) For the continued inequality mean $=$ mode $<$ median: one optimal list (in G) with first entry g and length $n = 6$ is $g, 10g, 10g, 12g, 13g, 14g$;

(vii) For the continued inequality median $<$ mean $<$ mode: one optimal list (in G) with first entry g and length $n = 5$ is $g, 2g, 3g, 7g, 7g$;

(viii) For the continued inequality median $<$ mean $=$ mode: one optimal list (in G) with first entry g and length $n = 6$ is $g, 2g, 3g, 5g, 5g, 14g$;

(ix) For the continued inequality median $<$ mode $<$ mean: one optimal list (in G) with first entry g and length $n = 6$ is $g, 2g, 3g, 5g, 5g, 20g$;

(x) For the continued inequality median $<$ mode $=$ mean: one optimal list (in G) with first entry g and length $n = 6$ is $g, 2g, 5g, 7g, 7g, 20g$;

(xi) For the continued inequality median $=$ mode $<$ mean: one optimal list (in G) with first entry g and length $n = 3$ is $g, g, 4g$;

(xii) For the continued inequality mode $<$ mean $<$ median: one optimal list (in G) with first entry g and length $n = 5$ is $g, g, 5g, 6g, 7g$;

(xiii) For the continued inequality mode $<$ median $=$ mean: one optimal list (in G) with first entry g and length $n = 5$ is $g, g, 3g, 4g, 6g$.

(b) For each of the 13 continued inequalities in G that were addressed in (i)-(xiii) of (a), there exist an optimal list (in G) with first entry $-g$ and an optimal list (in G) with first entry 0.

(c) For each of the 13 continued inequalities in G that were addressed in (i)-(xiii) of (a), there exist infinitely many optimal lists (in G).

(d) If, in addition, G is a cyclic subgroup of $\mathbb{Q}$ (with the induced ordering) and g generates G, then for each of the 13 continued inequalities in G that were addressed in (i)-(xiii) of (a), there exist infinitely many optimal lists (in G) with the property that their first element is of the form kg (where k is a positive integer that depends on the list).

(e) If, in addition, G is a subgroup of $\mathbb{Q}$ (with the induced ordering), then for each of the 13 continued inequalities in G that were addressed in (i)-(xiii) of (a), there exist infinitely many optimal lists (in G), each of which consists of positive integers.

(f) For each of the 13 continued inequalities in G that were addressed in (i)-(xiii) of (a), there exists an optimal list $\mathcal{L} : m_1 g, \ldots, m_n g$ for this continued inequality in G such that $m_1 = 1$ and each $m_j \in \mathbb{Z}$.

Proof. (a) Suppose that $x, y, d \in G$ and p is a positive integer. Let $\aleph$ be one of the (binary) relations $<$ and $=$ on G. Then

$$x \aleph y \Leftrightarrow x + d \aleph y + d \Leftrightarrow px \aleph py.$$

This easy fact was, of course, involved in verifying some of the assertions in Lemma 2.1 and in the comments preceding the statement of that result. Its uses there suggest (to the author) that it would be helpful to find a way to simplify our task of proving the present theorem. To that end, suppose that the relations $\aleph_1$ and $\aleph_2$ are each elements of $\{<, =\}$, that

$$\{M_1, M_2, M_3\} = \{\text{mean, median, mode}\},$$

that the list $\mathcal{L} : x_1, \ldots, x_n$ (of length $n \geq 2$) in G satisfies

$$M_1 \aleph_1 M_2 \aleph_2 M_3,$$

and that n is minimal for lists that satisfy this continued inequality. As a permutation of (the entries in) a list does not change the mean, the median or the mode of the list, we can and will, without loss of generality, assume that **each list considered in this proof is weakly increasing**. Many lists will be tested to determine whether they satisfy the definition of an "optimal list" (for a given continued inequality). In view of a comment from the Introduction, we also **assume that each list considered here is of finite length** ≥ 2 **and has a unique mode**.

We wish to avoid any ambiguity that may arise from the use of commas in the structure of a sentence in this proof and the use of commas separating an element in a list from the next element in that list. Accordingly, it will be helpful to adopt the following notation/terminology. We can use the generic notation $a < b < c < \ldots$ for a *strictly* increasing sequence in G to uniquely describe any list under consideration here. For instance, the description aabbbc would apply to all (weakly increasing) lists (of length 6), $\mathcal{L} : x_1, x_2, x_3, x_4, x_5, x_6$, such that $x_1 = x_2 < x_3 = x_4 = x_5 < x_6$. In this example, all such lists would be described as being of type (or of kind) aabbbc; and all such lists could be treated together (when working with any specific continued inequality) because their relevant statistics would all be given by mean $= (2a + 3b + c)/6$, median $= (b + b)/2 = b$ and mode $= b$.

Observe that the lists of the type aa ... a are the same as the lists such that mean = median = mode. Thus, the optimal lists for the continued inequality in (iv) are the lists of the type aa, such as the list $L : g, g$. As we have just proved the assertion for (a) (iv), the remainder of the proof of (a) is intended to apply to the contexts (i)-(iii) and (v)-(xiii). The preceding comments explain why we need not further consider any list of the type aa ... a. Moreover, we need not consider any list having one of the types ab, abc, abcd, ... or aabb, aabbc, ..., as each such list fails to have a unique mode. In particular, each list considered below can be assumed to have length at least 3.

One can attack any of the contexts (i)-(iii) and (v)-(xiii) in what the Introduction called "an organized way", as follows. Because it would be desirable to find all the optimal lists for a given continued inequality, we first test all the (types of) lists of length 3. If at least one list having one of those types satisfies the given continued inequality, we choose a suitable realization of that notation and place it into the assertion for the appropriate context of (a). If no example of a type of list of length 3 has satisfied the given continued inequality, test all the (types of) lists of length 4. If at least one list having one of those types satisfies the given continued inequality, we choose one such realization of that notation and place it into the assertion for the appropriate context of (a). If no example of a type of list of length 4 has satisfied the given continued inequality, test all the (types of) lists of length 5. Iterate. While studying all the candidate (types of) lists of a fixed length, we will be sure to list the candidates in increasing number of pairwise distinct elements of $\{a,b,c,\dots\}$ that are needed to describe the list; and for candidates of the same length with the same number of such pairwise distinct elements, we will list those candidates in a lexicographic-like manner that will become clear from the following two examples.

Having already proved (a) (iv), we will give complete proofs for (a) (i) and for (a) (ix). The first of these is not entirely trivial (its optimal lists end up having length 4); on the other hand, the second of the contexts proved below is as difficult as possible, since its optimal lists end up having length 6. In the proof for (i), we will document an inequality or equality that a rejected candidate has failed to satisfy; if a candidate has failed to satisfy both parts of the continued inequality at hand, we will mention the first of those failures, in the interest of conserving some space. In that same interest, the proof given below for (ix) will leave to the reader the task of confirming that a rejected candidate has failed to satisfy an inequality that we have specified.

(i) For aab: Each list of this kind is to be rejected because it fails to satisfy mean < median (since $(2a+b)/3 \nless a$).

For abb: Reject it, as it fails to satisfy median < mode (since $b \nless b$).

For aaab: Reject it, as it fails to satisfy mean < median (since $(3a+b)/4 \nless a = (a+a)/2$).

For abbb: Reject it, as it fails to satisfy median < mode (since $b = (b+b)/2 \nless b$).

For aabc: Reject it, as it fails to satisfy mean < median (since $(2a+b+c)/4 \nless (a+b)/2$).

For abbc: Reject it, as it fails to satisfy median < mode (since $(b+b)/2 = b \nless b$).

For abcc: Each list of this type satisfies mean < median < mode, since

$$\frac{a+b+2c}{4} < \frac{b+c}{2} < c.$$

Consequently, the optimal lists for (i) are the same as those lists of the type abcc (with $a < b < c$, as in the above conventions). Many readers are likely to agree that the most evident example of such an optimal list, with $x_1 = g$, is $g, 2g, 3g, 3g$. This example of an optimal list for (i) has the additional desirable feature that all the numerical coefficients of g (in the descriptions of the members of the optimal list) are positive integers. However, for reasons that will become clear in Corollary 2.4, the statement of (a) (i) mentions, instead, the optimal list $g, 5g, 7g, 7g$ (which also has the just-mentioned “desirable feature"). This completes the proof for context (i) of (a).

Before passing to the proof of context (ix) for (a), we would like to mention a feature of the possible structure of optimal lists that did not arise in the above solution for context (i). The readers who carry out the methodology (that was suggested above) to determine the optimal lists for context (v) will find that those optimal lists are only *some* of the lists of a certain type. Specifically, the optimal lists for (a) (v) are the lists of the type abbcde such that $a+c+d+e < 4b$. In particular, the list $g, 9g, 9g, 10g, 11g, 12g$ is an optimal list for (v); but the list $g, 8g, 8g, 10g, 11g, 12g$ is not an optimal list for (v), even though this list is of the type abbcde.

(ix) For aab: Each list of this kind is to be rejected because it fails to satisfy median < mode.

For abb: Reject it, as it fails to satisfy median < mode.

For aaab: Reject it, as it fails to satisfy median < mode.

For abbb: Reject it, as it fails to satisfy median < mode.

For aabc: Reject it, as it fails to satisfy median < mode.

For abbc: Reject it, as it fails to satisfy median < mode.

For abcc: Reject it, as it fails to satisfy mode < mean.
For aaaab: Reject it, as it fails to satisfy the inequality median < mode.
For aaabb: Reject it, as it fails to satisfy median < mode.
For aabbb: Reject it, as it fails to satisfy median < mode.
For abbbb: Reject it, as it fails to satisfy the inequality median < mode.
For aaabc: Reject it, as it fails to satisfy median < mode.
For abbbc: Reject it, as it fails to satisfy median < mode.
For abccc: Reject it, as it fails to satisfy mode < mean.
For aabcd: Reject it, as it fails to satisfy median < mode.
For abbcd: Reject it, as it fails to satisfy median < mode.
For abccd: Reject it, as it fails to satisfy median < mode.
For abcdd: Reject it, as it fails to satisfy mode < mean.
For aaaaab: Reject it, as it fails to satisfy median < mode.
For aaaabb: Reject it, as it fails to satisfy median < mode.
For aabbbb: Reject it, as it fails to satisfy median < mode.
For abbbbb: Reject it, as it fails to satisfy median < mode.
For aaaabc: Reject it, as it fails to satisfy median < mode.
For aaabbc: Reject it, as it fails to satisfy median < mode.
For aaabcc: Reject it, as it fails to satisfy median < mode.
For aabbbc: Reject it, as it fails to satisfy median < mode.
For aabccc: Reject it, as it fails to satisfy mode < mean.
For aabcde: Reject it, as it fails to satisfy median < mode.
For abbcde: Reject it, as it fails to satisfy median < mode.
For abccde: Reject it, as it fails to satisfy median < mode.
For abcdde: A list of this type satisfies median < mode < mean if and only if

$$\frac{c+d}{2} < d < \frac{a+b+c+2d+e}{6} \Leftrightarrow e > 4d-a-b-c.$$

For abcdee: Reject it, as it fails to satisfy mode < mean.

Consequently, the optimal lists for (ix) are the same as those lists of the type abcdde (with $a < b < c < d < e$, as in the above conventions) such that $e > 4d - a - b - c$. Perhaps the most evident example of such an optimal list, with $x_1 = g$, is $g, 2g, 3g, 4g, 4g, 11g$; but, for reasons that will become clear in Corollary 2.4, the statement of (a) (ix) mentions, instead, the optimal list $g, 2g, 3g, 5g, 5g, 20g$. This completes the proof for context (ix) of (a). The reader is encouraged to use the above methodology to complete the proofs for the remaining ten contexts of (a).

In the proofs of (b)-(e), (*) will denote (a fixed) one of the 13 continued inequalities in G that were addressed in (i)-(xiii) of (a), and $L_1 : x_1, \ldots, x_n$ will denote the optimal list for (*) in G, with first entry $x_1 = g$, that was found in (a).

(b) By (a), there exists a finite multiset $L_1 : x_1, \ldots, x_n$ which is an optimal list for (*) in G whose first entry is $x_1 = g$. Let $L_2 : y_1, \ldots, y_n$ and $L_3 : z_1, \ldots, z_n$ be the multisets given by $y_i := x_i - 2g$ and $z_i := x_i - g$ for all $1 \leq i \leq n$. In view of the conditions satisfied by L_1 and the definition of an optimal list, it follows from parts (a), (b), (c) and (d), respectively, of Lemma 2.1 that L_2 and L_3 are each a list of length n in G with a unique mode, that L_2 and L_3 are each weakly increasing, that the minimal value of an entry of L_2 is $y_1 = x_1 - 2g = g - 2g = -g$ and the minimal value of an entry of L_3 is $z_1 = x_1 - g = g - g = 0$, and that L_2 and L_3 are each a solution of (*). Thus, L_2 and L_3 are optimal lists with the asserted properties.

(c) For each integer $d \geq 2$, let $L_d : x_{d,1}, \ldots, x_{d,n}$ be the multiset given by $x_{d,i} := dg + x_i$. As in the proof of (b), successive appeals to parts (a)-(d) of Lemma 2.1 reveal, for each $d \geq 2$, that L_d is a list of length n in G with a unique mode, that L_d is weakly increasing, that the minimal value of an entry of

L_d is $x_{d,1} = dg + x_1$, and that L_d is a solution of (*). Thus, for each $d \geq 2$, L_d satisfies the definition of being an optimal list for (*) in G. Finally, as d varies, there are infinitely many such L_d, for if $d_1 \neq d_2$, then $L_{d_1} \neq L_{d_2}$ (since L_{d_1} and L_{d_2} have unequal minimal entries: $x_{d_1,1} = d_1g + x_1 \neq d_2g + x_1 = x_{d_2,1}$).

(d) For each integer $d \geq 2$, let $L_d : x_{d,1}, \dots, x_{d,n}$ be the multiset given by $x_{d,i} := dg + x_i$. As in the proof of (c), each L_d is an optimal list for (*) in G, with $L_{d_1} \neq L_{d_2}$ whenever $d_1 \neq d_2$, and so there are infinitely many such L_d. Note that for each $d \geq 2$, the first entry of L_d, namely, $x_{d,1} := dg + x_1 = dg + g = (d+1)g$, is of the desired form, as $d + 1$ is a positive integer.

(e) As $x_1 = g$ is a positive rational number and L_1 is weakly increasing, one can express the entries of L_1 as fractions having a common denominator, as follows: $x_i = m_i/p$, for certain positive integers $m_1, \dots, m_n, p$. Consider the multiset $\mathcal{L} : w_1, \dots, w_n$ given by $w_i := px_i$ for all $1 \leq i \leq n$. As $p > 0$, successive appeals to parts (a), (b), (c) and (e) of Lemma 2.1 reveal that $\mathcal{L}$ is a list of length n in G with a unique mode, that $\mathcal{L}$ is weakly increasing, that the minimal value of an entry of $\mathcal{L}$ is $w_1 = px_1 = pg$, and $\mathcal{L}$ is a solution of (*). Thus, $\mathcal{L}$ is an optimal list for (*) in G. Moreover, each entry of $\mathcal{L}$ is a positive integer, since $w_i = px_i = m_i$. Even though G need not be a cyclic group, we can now mimic part of the proof of (d), as follows. For each integer $d \geq 2$, let $\mathcal{L}_d : w_{d,1}, \dots, w_{d,n}$ be the multiset given by $w_{d,i} := d + w_i$. As in the proof of (b), successive appeals to parts (a)-(d) of Lemma 2.1 lead to the conclusion that for each $d \geq 2$, $\mathcal{L}_d$ is an optimal list for (*) in G, with $\mathcal{L}_{d_1} \neq \mathcal{L}_{d_2}$ whenever $d_1 \neq d_2$ (since $w_{d_1,1} = d_1 + w_1 = d_1 + m_1 \neq d_2 + m_1 = d_2 + w_1 = w_{d_2,1}$) and each entry $w_{d,i}$ $(= d + w_i = d + m_i)$ is a positive integer.

(f) The reader may have noticed that some of the entries in the examples of optimal lists illustrating parts (i)-(xiii) in (a) had unnecessarily large numerical coefficients multiplying g. For instance, part (i) was illustrated by the list "$g, 5g, 7g, 7g$", but it could also have been illustrated by the list "$g, 2g, 3g, 3g$". However, for our present purposes in (f), the latter list would not be satisfactory, since its associated mean (resp., associated median), namely $(9/4)g$ (resp., $(5/2)g$), is not of the form mg for some positive integer m. On the other hand, each of the examples of optimal lists illustrating parts (i)-(xiii) in the statement of (a) was constructed/chosen so that its mean (resp., median; resp., mode) is of the form mg for some positive integer m. We leave to the reader the task of performing the calculations that verify the assertions in the preceding sentence. The proof is complete. □

Parts (e) and (f) of Theorem 2.2 suggest Proposition 2.3 (e) and Corollary 2.4. These results illustrate the following maxim that some mathematicians find helpful in conducting research: it is often fruitful to begin by deeply analyzing a typical relevant example (I would add: provided that it is feasible to access such an example). Some pedagogic upshots of that point of view will be offered in Remark 2.5 (h).

Proposition 2.3. *Let $(G, \leq)$ be a nonzero additive divisible ordered abelian group. Let $g > 0$ in G. Let $L : q_1, \dots, q_n$ be a list (of length $n \geq 2$) in $\mathbb{Q}$ (viewed as an ordered abelian group under addition, with the usual ordering) with a unique mode. Express the q_i as fractions having a common denominator, as follows: $q_i = m_i/k$ (for all $1 \leq i \leq n$) for certain integers $m_1, \dots, m_n, k$, with $k > 0$ in $\mathbb{Z}$. Consider the multisets*

$$\mathcal{L} : q_1g, \dots, q_ng \text{ in } G \quad \text{and} \quad \mathcal{E} : m_1, \dots, m_n \text{ in } \mathbb{Q}.$$

Then:

(a) If q_j is the mode of L (in $\mathbb{Q}$), then q_jg is the unique mode of $\mathcal{L}$ (in G) and m_j is the unique mode of $\mathcal{E}$ (in $\mathbb{Q}$).

(b) $\mathcal{L}$ is weakly increasing in $G \Leftrightarrow L$ is weakly increasing in $\mathbb{Q} \Leftrightarrow \mathcal{E}$ is weakly increasing in $\mathbb{Q}$.

(c) In G, let μ_1, λ_1 and ν_1 denote the mean, median and mode, respectively, of $\mathcal{L}$. In $\mathbb{Q}$, let μ_2, λ_2 and ν_2 denote the mean, median and mode, respectively, of L; and let μ_3, λ_3 and ν_3 denote the mean, median and mode, respectively, of $\mathcal{E}$. Then $\mu_2 = \frac{\mu_3}{k}$, $\lambda_2 = \frac{\lambda_3}{k}$, $\nu_2 = \frac{\nu_3}{k}$, $\mu_1 = \mu_2 g = (\frac{\mu_3}{k})g$, $\lambda_1 = \lambda_2 g = (\frac{\lambda_3}{k})g$ and $\nu_1 = \nu_2 g = (\frac{\nu_3}{k})g$.

(*d*) *Let* M_1 *and* M_2 *be distinct members of* {mean, median, mode} *and let the relation* $\aleph \in \{<,=\}$. *Then* $\mathcal{L}$ *satisfies* (*the continued inequality*) $\mathrm{M}_1 \aleph \mathrm{M}_2$ *in* $G \Leftrightarrow L$ *satisfies* $\mathrm{M}_1 \aleph \mathrm{M}_2$ *in* $\mathbb{Q} \Leftrightarrow \mathcal{E}$ *satisfies* $\mathrm{M}_1 \aleph \mathrm{M}_2$ *in* $\mathbb{Q}$.

(*e*) *Let* (*) *be one of the 13 continued inequalities that were addressed in* (*i*)-(*xiii*) *of Theorem* 2.2 (*a*). *Then the following conditions are equivalent:*

(*1*) $\mathcal{L}$ *is an optimal list* (*of length* n) *for* (*) *in* G;
(*2*) L *is an optimal list* (*of length* n) *for* (*) *in* $\mathbb{Q}$;
(*3*) $\mathcal{E}$ *is an optimal list* (*of length* n) *for* (*) *in* $\mathbb{Q}$.

Proof. Some of the techniques in the proofs of (a) and (b) that are given below are easy and familiar facts about ordered abelian groups, and in fact, some of those techniques were used implicitly in calculations supporting the assertions in some of the cases in Theorem 2.2 (a). For the sake of completeness, and because (a) and (b) will be used in the proof of (e), which is central to this work, we will provide complete proofs for (a) and (b) here.

(a) Since G is torsionfree and $g \neq 0$, $q_{i_1}g = q_{i_2}g$ (if and) only if $q_{i_1} = q_{i_2}$. It follows that $q_j g$ is the unique mode of $\mathcal{L}$. Similarly, since $q_{i_1} = q_{i_2}$ (if and) only if $m_{i_1} = m_{i_2}$, it follows that m_j is the unique mode of $\mathcal{E}$.

(b) Since G is an ordered group and $g > 0$, it is clear that $q_{i_1} < q_{i_2}$ implies $q_{i_1}g < q_{i_2}g$. As G is torsionfree (and $g > 0$), the converse also holds (for if $q_{i_1}g < q_{i_2}g$ and $q_{i_2} < q_{i_1}$, then $q_{i_2}g < q_{i_1}g < q_{i_2}g$, whence $q_{i_2}g < q_{i_2}g$, a contradiction). Thus, $q_{i_1} < q_{i_2}$ if and only if $q_{i_1}g < q_{i_2}g$. When combined with the proof of (a), this gives the conclusion that L is weakly increasing if and only if $\mathcal{L}$ is weakly increasing. By reasoning similarly, a proof that L is weakly increasing if and only if $\mathcal{E}$ is weakly increasing is reduced to showing the following: $q_{i_1} < q_{i_2}$ if and only if $m_{i_1} < m_{i_2}$. That, in turn, is evident, since $m_{i_t} = km_{i_t}$ and $k > 0$ in $\mathbb{Q}$.

(c) By definition, $\nu_2 = q_j$. Hence, by (a), $\nu_1 = q_j g$ ($= \nu_2 g$); and $\nu_3 = m_j$, so that $\nu_2 = m_j/k = \nu_3/k$. It remains only to prove that $\mu_2 = \mu_3/k$, $\lambda_2 = \lambda_3/k$, $\mu_1 = \mu_2 g$ and $\lambda_1 = \lambda_2 g$. We will prove the first and third of these assertions (that is, the assertions concerning means) and leave the similar (and slightly less detailed) proofs of the second and fourth of these assertions (concerning medians) to the reader.

By using the definition of a mean, we get

$$\mu_2 = \frac{\sum_{i=1}^n q_i}{n} = \frac{\sum_{i=1}^n (\frac{m_i}{k})}{n} = (\frac{1}{k})(\frac{\sum_{i=1}^n m_i}{n}) = \frac{\mu_3}{k}.$$

Recalling from the earlier discussion that G can be viewed as a vector space over $\mathbb{Q}$, we get via similar reasoning that

$$\mu_1 = \frac{\sum_{i=1}^n q_i g}{n} = \sum_{i=1}^n (\frac{q_i}{n})g = (\frac{\sum_{i=1}^n q_i}{n})g = \mu_2 g.$$

(d) Because there are six choices for the ordered pair $(\mathrm{M}_1, \mathrm{M}_2)$ and two choices for the relation $\aleph$, we must establish 12 assertions in order to prove (d). The six assertions involving the relation "<" can be stated as follows:

(i) $\mu_1 < \lambda_1$ (in G) $\Leftrightarrow \mu_2 < \lambda_2$ (in $\mathbb{Q}$) $\Leftrightarrow \mu_3 < \lambda_3$ (in $\mathbb{Q}$);
(ii) $\mu_1 < \nu_1 \Leftrightarrow \mu_2 < \nu_2 \Leftrightarrow \mu_3 < \nu_3$;
(iii) $\lambda_1 < \mu_1 \Leftrightarrow \lambda_2 < \mu_2 \Leftrightarrow \lambda_3 < \mu_3$;
(iv) $\lambda_1 < \nu_1 \Leftrightarrow \lambda_2 < \nu_2 \Leftrightarrow \lambda_3 < \nu_3$;
(v) $\nu_1 < \mu_1 \Leftrightarrow \nu_2 < \mu_2 \Leftrightarrow \nu_3 < \mu_3$;
(vi) $\nu_1 < \lambda_1 \Leftrightarrow \nu_2 < \lambda_2 \Leftrightarrow \nu_3 < \lambda_3$.

By using (c), we can reformulate (i)-(vi) as follows:

(i)′ $\mu_2 g < \lambda_2 g \Leftrightarrow \mu_2 < \lambda_2 \Leftrightarrow k\mu_2 < k\lambda_2$;
(ii)′ $\mu_2 g < \nu_2 g \Leftrightarrow \mu_2 < \nu_2 \Leftrightarrow k\mu_2 < k\nu_2$;

(iii)′ $\lambda_2 g < \mu_2 g \Leftrightarrow \lambda_2 < \mu_2 \Leftrightarrow k\lambda_2 < k\mu_2$;
(iv)′ $\lambda_2 g < \nu_2 g \Leftrightarrow \lambda_2 < \nu_2 \Leftrightarrow k\lambda_2 < k\nu_2$;
(v)′ $\nu_2 g < \mu_2 g \Leftrightarrow \nu_2 < \mu_2 \Leftrightarrow k\nu_2 < k\mu_2$;
(vi)′ $\nu_2 g < \lambda_2 g \Leftrightarrow \nu_2 < \lambda_2 \Leftrightarrow k\nu_2 < k\lambda_2$.
Since $g > 0$ in G and $k > 0$ in $\mathbb{Q}$, each of (i)′-(vi)′ can be proved by reasoning as in the proof of (b).

Finally, the six assertions involving the relation "=" can be obtained from (i)-(vi) by replacing each occurrence of "<" in (i)-(vi) with "=". The resulting six assertions can be reformulated by using (c). This has the effect of reducing our task to proving the six assertions that can be obtained from (i)′-(vi)′ by replacing each occurrence of "<" in (i)′-(vi)′ with "=". Since $g \neq 0$ in G and $k \neq 0$ in $\mathbb{Q}$, each of the resulting six assertions can then be proved by reasoning as in the proof of (a).

(e) Observe that $\mathcal{L}$ and $\mathcal{E}$ each have the same length (namely, n) as L. Therefore, in view of the definition of an "optimal list", the assertion follows by combining (b) and (d). The proof is complete. □

Corollary 2.4. *Let* (*) *be one of the 13 possible continued inequalities involving the mean, the median, and the mode* (*and the symbols "<" and/or "="*) *of a finite nonempty list in a nonzero additive divisible ordered abelian group, and let G be such a group. Then there exists an optimal list for* (*) *in G, the length of any such optimal list is at most* 6 *and, although that length depends on* (*), *that length is independent of G.*

Proof. Pick a positive element g of G. By Theorem 2.2 (a), there exists an optimal list, $\mathcal{L} : a_1 g, \dots, a_n g$, for (*) in G, for some integer n such that $2 \le n \le 6$ and some list, $L : a_1, \dots, a_n$, of positive integers. The detailed statement of Theorem 2.2 (a) reveals that the length of L does depend on (*). However, since Proposition 2.3 (e) gives that L is an optimal list for (*) in $\mathbb{Q}$ (with the usual ordering), we have that n is determined by (*). In that sense, n does not depend on the specific G in the hypotheses. □

We close by collecting some theoretical and/or pedagogic comments.

Remark 2.5. (a) It is reasonable to ask, with G as above, what can be said about an analogous meaning of "optimal list" for a continued inequality of the form

$$M_1 \ \mathcal{R}_1 \ M_2 \ \mathcal{R}_2 \ M_3$$

if $\{M_1, M_2, M_3\} = \{\text{mean, median, mode}\}$ and the relations $\mathcal{R}_1$ and $\mathcal{R}_2$ are each elements of $\{<, \le, =\}$. Such an inclusion of the relation $\le$ as a possible participant in the relevant continued inequalities does lead to the need to consider some new continued inequalities, but Theorem 2.2 can be used to address each of those new ones. To illustrate how this can be done, consider the (new) continued inequality

$$\text{mean} \ \le \ \text{median} \ < \ \text{mode}.$$

This is, of course, the disjunction "Either mean = median < mode or mean < median < mode." (For any of the new continued inequalities, we would be considering, at this point, the disjunction of at most four of the continued inequalities that were addressed in Theorem 2.2.) Accordingly, parts (iii) and (ii) of Theorem 2.2 (a), together with Proposition 2.3 (e) (see also Corollary 2.4), draw our attention to the two lists (in $\mathbb{Q}$, with $g := 1$)

$$1, 3, 4, 6, 6 \text{ of length } 5 \quad \text{and} \quad 1, 4, 4 \text{ of length } 3.$$

It follows that an optimal list (in the obvious sense) for "mean ≤ median < mode" is 1,4,4, which is of length 3.

(b) The proof of Theorem 2.2 (a) shows that, instead of assuming that the ambient G is a nonzero additive divisible ordered abelian group, it would have sufficed to assume that G is a nonzero additive ordered abelian group which is also 2-divisible, 3-divisible and 5-divisible (hence, also 4-divisible and 6-divisible). However, making those weaker assumptions at the outset of the statement of Theorem 2.2, before knowing the "$n \leq 6$" conclusion for each of the 13 parts of Theorem 2.2 (a), would likely have seemed artificial and unmotivated.

(c) Speaking of generalizing the context of Theorem 2.2, one could have considered some of the above questions for an ambient group G that is assumed to be a certain kind of nonabelian left-orderable (or right-orderable or bi-orderable) group. Pursuing such matters may be of interest to some readers, although we would not encourage a nonabelian study of center-indicating measures for unsophisticated audiences. On the other hand, a nonabelian and "sufficiently divisible" context may lead to some interesting results. In any such research, one would likely need to decide how to use permutations to get appropriate analogues of "mean" and "median."

(d) For readers who may be interested in generalizing the context of Theorem 2.2 while maintaining the assumption that the ambient G has a commutative addition, we suggest that it may be fruitful to investigate possible results when the ambient G is assumed to be a certain kind of abelian monoid M. For that context, a "cancellative" assumption on M would seem reasonable if one seeks to generalize some of the methods of proof in this paper.

(e) It may also be interesting to see what happens if the above role of the "(arithmetic) mean" is played instead by the geometric mean (provided that the appropriate n^{th} roots exist in G), as the concept of "geometric mean" of a finite set of positive real numbers is often studied in high school (or early in one's college studies). Following the lead of classical analysis, one may also seek to investigate what happens if the "mean" studied in Theorem 2.2 is replaced by $(\sum |x_i|)/n$ (or by $((\sum |x_i|^p)^{1/p})/n$ for some positive integer p if sufficiently many p^{th} roots exist in G). Similarly, analogues of the "median" could also be studied in such research. We suspect that noncommutative analysis may suggest additional variants of Theorem 2.2 that would be harder but interesting to study, but we will not comment further on this because of our relative inexperience with noncommutative analysis.

(f) One of the closing comments in the Introduction mentioned additive subgroups of $\mathbb{Q}$ (with the usual ordering) that contain $\mathbb{Z}$ and, more generally, Theorem 2.2 (e) considered additive subgroups of $\mathbb{Q}$ (with the induced ordering). Still more generally, there has been widespread interest in ordered abelian groups, for several reasons. One of these reasons concerns torsion. In this paper, we have used the easy fact that an ordered abelian group is torsionfree. The converse of that fact is also true. Indeed, in [6], F. Levi proved that a torsionfree abelian group is bi-orderable (that is, is an ordered group).

Another natural reason for interest in ordered groups is related to some familiar characterizations of the field of real numbers. One of those characterizations states that up to isomorphism (as a field), $\mathbb{R}$ is the unique (Dedekind-) complete ordered field. Another characterization describes $\mathbb{R}$ as being, up to isomorphism, the only Cauchy-complete Archimedean ordered field. Both of these characterizations have inspired research in adjacent mathematical areas such as Riesz groups/spaces. A more fundamental, but perhaps less widely known, variant of the second of these characterizations is due to Otto Hölder [4] and states that an Archimedean group (that is, a bi-ordered group satisfying the Archimedean property) is isomorphic to a subgroup of $\mathbb{R}$ under addition (and hence is an abelian group). As is occasionally the case for results that are considered to be of fundamental importance in several areas, the attribution of the just-mentioned result has been varied and, perhaps, has not always met the highest scholarly standard. Concerning this observation/opinion, we next discuss some pertinent history.

(g) In (f), we mentioned the result that any Archimedean group is isomorphic to an additive subgroup of $\mathbb{R}$. The history concerning the proof of that classic result deserves some scrutiny. In (f), we attributed that result to O. Hölder [4]. We were led to investigate the origins of that result in the summer of 2022, when we read in the Wikipedia piece on ordered groups that the result was being attributed to the 1942 paper of F. Levi [6] (which was appropriately cited for a different reason in (f)). Wikipedia gave the reference for that citation as a survey monograph that had been written by two eminent experts and published by the American Mathematical Society early in this century.

As I felt intuitively certain that the result was *much* older than I, I consulted an earlier publication of the American Mathematical Society, specifically [1], the Math Review by R. Baer of the paper [6] that was mentioned above. In [1], Baer mentions a sequence of papers proving the result and/or a generalization of it, beginning with Hölder's paper [4]. In chronological order of the papers, their authors were O. Hölder (1901), R. Baer (1929), H. Cartan (1939), Garrett Birkhoff (1942) and F. W. Levi (1942). In attributing the result to Hölder, Baer referred specifically to [4, pages 13-14].

(h) To enhance accessibility and to make for a relatively short paper, we have not pursued any of the directions for future research that were suggested above in (c)-(e). In closing, we wish to elaborate on the final sentence of the Abstract and on the end of the first paragraph of the Introduction. The latter essentially promised that this paper would be written in an "organized way." The proof of Theorem 2.2 (a) was, in particular, written to fit that description. It would not be difficult to convert that proof into a computer program involving positive integers. By Proposition 2.3 (e) and Corollary 2.4, the results of such computer activity could discover the facts and examples in the statement of Theorem 2.2 (a). The proof of Theorem 2.2 (a) was intentionally written in a style that could accommodate the needs of a variety of classes, students and instructors, as their pursuits of mathematics may proceed by various mixtures of methods or emphases, ranging from calculations with axioms to numerical experiments possibly aided by technology. Accordingly, we believe that this paper could be used for classes/courses at a variety of levels in a variety of ways, such as lectures with formal proofs; individual or group projects or discovery activities, possibly involving programming (either in class, as homework or on examinations); or other ways that fit into the teaching/learning styles of the instructor/audience.

References

[1] R. Baer, Review of [6], Math. Reviews, MR0007779 (4, 192b), Amer. Math. Soc.

[2] G. Birkhoff and S. Mac Lane, A survey of modern algebra, Third edition, Macmillan, New York, 1953.

[3] H. Cartan and S. Eilenberg, An introduction to homological algebra, Princeton University Press, Princeton, 1956.

[4] O. Hölder, Die Axiome der Quantität und die Lehre vom Mass, Ber. Verh. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Kl., 53 (1901), 1–64.

[5] T. W. Hungerford, Algebra, Graduate Texts in Math. **73**, Springer-Verlag, New York-Berlin, 1980.

[6] F. W. Levi, Ordered groups, Proc. Indian Acad. Sci., Sect. A, 16 (1942) (4), 256–263.

[7] E. Matlis, Injective modules over Noetherian rings, Pac. J. Math., 8 (1958), 511–528.

[8] N. H. McCoy, The theory of rings, Macmillan, New York, 1964.

[9] J. J. Rotman, An introduction to homological algebra, Academic Press, New York, 1979.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 14-37
ISSN: 2820-7114

Title :

Solutions of differential equations through Lie symmetry algebras

Author(s):

Abdelaziz Ouhadan & El Hassan El Kinani

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 14–37
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

Solutions of differential equations through Lie symmetry algebras

Abdelaziz Ouhadan[1] and El Hassan El Kinani[2]
[1] Department of Mathematics, Mathematical and Didactical Research Team,
Regional Center of Education and Training Trades, Meknes, Morocco
e-mail: *aouhadan@gmail.com*
[2] Department of Mathematics, Modeling Analysis and Control Systems (MACS)
Faculty of Sciences, Moulay Ismail University, Meknes, Morocco
e-mail: *e.elkinani@umi.ac.ma*

Communicated by Mohammed Tamekkante
(Received 10 October 2022, Revised 22 December 2022, Accepted 28 December 2022)

Abstract. In this paper, on the one hand, the necessary fundamental results of Lie's theory are presented. Precisely, we expose the notion of infinitesimal generators of a given group action, prologation formula and invariance criterion. On the other hand, some applications of symmetry analysis to some differential equations as construction of similarity solutions, linearization and so on are presentand.

Key Words: Lie group transformations, Differential equation, Similarity solutions.
2010 MSC: 26A33, 70G65, 76M60.

In memory of Muhammad Zafrullah

1 Introduction

Lie symmetry analysis is one of the efficient approches to study ordinary and partial differential equations. To address the area of linear differential equations appears easly then to study non linear ones. This is due to the standard technique of Fourier transform. However, to examine a non linear differential equation requirs to deal with a linear equation known as its approximation equation. One of the adventages of Lie symmetry method is to preserve the studied non linear equation without linear approximation. So the original differential equation is maintained unchanged.

Lie groups and Lie algebras play an important role in the understanding for generating integrable equations, construction of exact solutions, conservation laws, recursion operators and so on [17, 12, 2]. Generally, they are an integral part to understand the structure of differential equations. Sophus Lie introduced the concept of the Lie group in the purpose to unify various methods and techniques known in solving differential equations [10, 9]. The author of continuous transformations inspired the invariance idea from Galois's theory precisely from the invariance of an algebraic equation. He was the first to show that a one parameter symmetry group admitted by a differential equation allows us to reduce its order by one [10, 9]. Thereafter, several applications of Lie groups were developed including construction of an integrating factor, linearization, finding lax pairs, construction of similarity solutions and so on [18, 1, 7, 8].

In this paper, we briefly recall basic definitions and examples of Lie groups and Lie algebras necessary for the study of invariance properties of differential equations. Next, we present how a given function is transformed under a group action. After, we expose the prolongation of a group action and how an infinitesimal generator can be naturally prolonged to act on the jet space of independent variables, dependent variables and derivatives of dependent variables up to any fixed order. Finally, we present the invariance criterion and some applications of Lie symmetry analysis illustrated in the case of Thomas equation and Burger's equation.

2 Lie groups and Lie algebras

In this section, we recall in brief the definition of both a Lie group and a Lie algebra associated with numerous classical examples found in the literature precisely matrices groups and algebras.

2.1 Lie groups

Definition 2.1. An r-parameter Lie group G, which also carries the structure of an r-dimensional differentiable manifold in such a way that both the group composition $c : G \times G \to G, c(g_1, g_2) = g_1.g_2$ with $g_1, g_2 \in G$ and the inversion $i : G \to G, i(g) = g^{-1}$ are smooth maps between manifolds.

Example 2.2. The general linear group $GL(n, \mathbb{R})$ is a Lie group. The manifold structure can be identified with the open subset $GL(n, \mathbb{R}) = \{A : det(A) \neq 0\}$ of the linear space of all $n \times n$ nonsingular matrices. This space is isomorphic to $\mathbb{R}^{n^2}$ with matrix entries A_{ij} of A. Thus $GL(n, \mathbb{R})$ is also an n^2-dimensional manifold.
Furthermore, some other useful examples of classical Lie groups found in the literature are

1. The special linear groups $SL(n) := \{A \in GL(n, \mathbb{R}) : det(A) = 1\}$, with the dimension $n^2 - 1$,
2. The orthogonal groups $O(n) := \{A \in GL(n, \mathbb{R}) : A^T A = I\}$, with the dimension $\frac{1}{2}n(n-1)$,
3. The special orthogonal groups $SO(n) := \{A \in O(n) : det(A) = 1\}$, with the dimension $\frac{1}{2}n(n-1)$,
4. The unitry groups $U(n) := \{A \in GL(n, \mathbb{C}) : A^\dagger A = I\}$, with the dimension n^2 where the dragger $\dagger$ denotes the transpose and complex conjugate,
5. The special unitry groups $SU(n) := \{A \in U(n) : det(A) = 1\}$, with the dimension $n^2 - 1$.

2.2 Lie algebras

Definition 2.3. A Lie algebra $\mathcal{G}$ is a vector space over some field $\mathbb{K}$ on which a product $[,]$ called the Lie bracket or commutator, is defined with the properties:

1. Closure: $\forall (X, Y) \in \mathcal{G}^2$ it follows $[X, Y] \in \mathcal{G}$;
2. Bilinearity: $\forall (X, Y, Z) \in \mathcal{G}^3, \forall (\alpha, \beta) \in \mathbb{K}^2$:
$$[X, \alpha Y + \beta Z] = \alpha[X, Y] + \beta[X, Z];$$
3. Antisymmetry: $[X, Y] = -[Y, X]$;
4. Jacobi identity: $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$

Remark 2.4.
- If $\mathbb{K}$ is the field of real numbers $\mathbb{R}$ so $\mathcal{G}$ is called a real Lie algebra and a complex Lie algebra if $\mathbb{K}$ is the field of complex numbers $\mathbb{C}$.
- $\mathcal{G}$ is called an abelian Lie algebra if for any $X, Y \in \mathcal{G}$ we have
$$[X, Y] = 0.$$

Definition 2.5. Let $G \subset Gl(n,\mathbb{R})$ be a Lie group. The Lie algebra of the Lie group G is the subset

$$\mathcal{G} = \left\{X \in M_n(\mathbb{R}) : e^{tX} \in G, \forall t \in \mathbb{R}\right\}.$$

Remark 2.6. We can replace the global condition $e^{tX} \in G, \forall t \in \mathbb{R}$, by a local condition, as follows:

$$\forall \varepsilon > 0 : \quad \mathcal{G} = \left\{X \in M_n(\mathbb{R}) : e^{tX} \in G, \mid t \mid < \varepsilon\right\}.$$

In fact, let t' be in $\mathbb{R}$. then $\exists n \in \mathbb{N}$ (we can chose $n = E(\frac{2t'}{\varepsilon})$ such that $\alpha = t' - \frac{n\varepsilon}{2} \in [0, \frac{\varepsilon}{2}[$. So, if $e^{tX} \in G$ for $\mid t \mid < \varepsilon$ we get

$$e^{t'X} = e^{(\alpha + \frac{n\varepsilon}{2})X} = e^{\alpha X}(e^{\frac{\varepsilon}{2}X})^n \in G.$$

According to the definition of a Lie algebra of a Lie group $G \subset Gl(n,\mathbb{R})$, we obtain that an element $X \in M_n(\mathbb{R})$ is an element of the Lie algebra of G if and only if the one parameter subgroup $\varphi(t) = e^{tX}$ generated by X is a sub group of G. Then, $X = \frac{d}{dt}\varphi(t)_{|t=0}$ is the tangent vector to the curve $\varphi(t)$ at $\varphi(0) = I$. Consequently, the Lie algebra of a Lie group is therefore the set of tangent vectors in the identity element I to one-parameter subgroups of G.

Example 2.7. 1. Let X be an element of $M(n,\mathbb{R})$. For each $t \in \mathbb{R}$ there is an inverse of e^{tX} that is e^{-tX}. Hence, $M(n,\mathbb{R}) \subset \mathcal{G}l(n,\mathbb{R})$ so,

$$\mathcal{G}l(n,\mathbb{R}) = M(n,\mathbb{R}).$$

2. In the same way, we get

$$\begin{aligned} sl &= \left\{X \in M(n,\mathbb{R}) : \quad e^{tX} \in SL(n,\mathbb{R}), \forall t \in \mathbb{R}\right\} \\ &= \left\{X \in M(n,\mathbb{R}) : \quad det(e^{tX}) = 1, \forall t \in \mathbb{R}\right\} \\ &= \left\{X \in M(n,\mathbb{R}) : \quad e^{t.tr(X)} = 1, \forall t \in \mathbb{R}\right\}. \end{aligned}$$

Consequently, the Lie algebra of the Lie group $SL(n,\mathbb{R}$ is given by

$$sl = \{X \in M(n,\mathbb{R}) : \quad tr(X) = 0\}.$$

In the same way, we can obtain the Lie algebras of the orthogonal groups $O(n)$, the special orthogonal groups $SO(n)$, the unitry groups $U(n)$ and so on [12, 18].

2.3 Lie transformation groups

Definition 2.8. Let M be a differentiable manifold and G be a Lie group. The group G is called a Lie transformation group of M if there is a differentiable map $\varphi : G \times M \to M, \quad \varphi(g,x) = gx$ such that :

1. $(g_1.g_2)x = g_1.(g_2x)$ for $x \in M$ and $g_1, g_2 \in G$
2. $ex = x$ for the identity element e of G and $x \in M$

are satisfied.

Remark 2.9. This transformations group is also known as the group action on M.

Example 2.10 (Translation group)**.** Let be $T(n)$ the translation group on the vector space $\mathbb{R}^n : T(n) := \{\tau(a) : a \in \mathbb{R}^n\}$ where $\tau(a)(x) = x + a$ for $x \in \mathbb{R}^n$.

Then the dimension of $T(n)$ is n. Furthermore, $T(n)$ is an abelian group since

$$\begin{aligned}\tau(v').\tau(v)(x) &= (x+v)+v' \\ &= x+(v+v') \\ &= \tau(v+v')(x) \\ &= \tau(v).\tau(v')(x).\end{aligned}$$

The inverse is given by $\tau(a)^{-1} = \tau(-a)$ and the identity element is $e = \tau(0)$.

3 Infinitesimal transformations and infinitesimal generators

This section has two aims. First, to introduce the notion of an infinitesimal matrix for an r-parameter Lie group of transformations and its corresponding infinitesimal generators. Second, to recall the three Lie's fondamental theorems.

3.1 Infinitesimal transformations

Let φ be a one parameter Lie group of transformations of the form

$$y = \varphi(x, \varepsilon), \tag{1}$$

where $y = (y_1, \cdots, y_m)$ and $x = (x_1, \cdots, x_m)$ belong in an open subset $U \subset \mathbb{R}^m$. The group composition denoted by $\phi(\varepsilon, \eta)$ where ε and η are the group parameters in an open subset of $\mathbb{R}$ and ϕ is considered to be analytic in its domain. Thus, if $z = \varphi(y, \eta)$ and $y = \varphi(x, \varepsilon)$ then $z = \varphi(x, \phi(\varepsilon, \eta))$.

Example 3.1 (Scaling group in $\mathbb{R}^2$)**.** Let φ be a one parameter Lie group of transformations defined by

$$\begin{aligned}\varphi :\ ]0,+\infty[\times \mathbb{R}^2 &\to \mathbb{R}^2 \\ (\alpha,(x,y)) &\mapsto (\alpha x, \alpha^2 y).\end{aligned}$$

From $(\alpha,(x,y)) \overset{\alpha}{\mapsto} (\alpha x, \alpha^2 y) \overset{\beta}{\mapsto} (\beta\alpha x, \beta^2\alpha^2 y)$ we get $\phi(\alpha, \beta) = \alpha\beta$, so the identity element is $\alpha = 1$.

Remark 3.2. If we put $\varepsilon = \alpha - 1$, so that the identity element is $\varepsilon = 0$ and φ can be redefined to be of the form

$$\begin{aligned}\varphi :\ ]-1,+\infty[\times \mathbb{R}^2 &\to \mathbb{R}^2 \\ (\varepsilon,(x,y)) &\mapsto ((1+\varepsilon)x, (1+\varepsilon)^2 y).\end{aligned}$$

Hence, as done in the previous example, we obtain $\phi(\varepsilon, \eta) = 1 + \eta + \varepsilon + \varepsilon\eta$ and $\varepsilon^{-1} = \dfrac{-\varepsilon}{1+\varepsilon}$.

Obviously, the above algebraic manipulation yields the identity element $\varepsilon = 0$. Moreover, it doesn't give the standard form of the group composition. This question was solved by the first Lie's theorem that will be illustread.

Definition 3.3. The transformation

$$y = x + \varepsilon\xi(x), \tag{2}$$

where $\xi(x)$ is finding by expanding $\varphi(x, \varepsilon)$ about 0:

$$y = x + \varepsilon \frac{\partial\varphi}{\partial\varepsilon}(x,\varepsilon)\mid_{\varepsilon=0} + \frac{\varepsilon^2}{2}\frac{\partial^2\varphi}{\partial\varepsilon^2}(x,\varepsilon)\mid_{\varepsilon=0} + o(\varepsilon^2),$$

with $\xi(x) := \dfrac{\partial \varphi}{\partial \varepsilon}(x,\varepsilon)\mid_{\varepsilon=0}$ is called the infinitesimal transformation of the one parameter Lie transformation group (1) and the components of $\xi(x)$ are called infinitesimals of (1).

Theorem 3.4 (First fundamental theorem of Lie)**.** *There exists a parametrization $\tau(\varepsilon)$ such that the Lie group of transformation (1) is equivalent to the solution of the initial value problem for the autonomous system of first order differential equations*

$$\frac{dy}{d\tau} = \xi(y), \tag{3}$$

with $y = x$ when $\tau = 0$. In particular

$$\tau(\varepsilon) = \int_0^{\varepsilon} \Gamma(\varepsilon')\, d\varepsilon', \tag{4}$$

where

$$\Gamma(\varepsilon) = \frac{\partial \phi(a,b)}{\partial b}\mid_{(a,b)=(\varepsilon^{-1},\varepsilon)} . \tag{5}$$

Example 3.5 (Scaling group)**.** We consider the one parameter Lie group of transformation given by

$$\begin{array}{llll} \varphi : &]-1,+\infty[\, \times \mathbb{R}^2 & \to \mathbb{R}^2 & \\ & (\varepsilon,(x_1,x_2)) & \mapsto \varepsilon.(x_1,x_2) = (y_1,y_2) = ((1+\varepsilon)x_1,(1+\varepsilon)^2 x_2), \end{array}$$

with

$$\phi(\varepsilon,\eta) = 1+\varepsilon+\eta+\varepsilon\eta \qquad \text{and} \qquad \varepsilon^{-1} = \frac{-\varepsilon}{1+\varepsilon}.$$

We have

$$\begin{aligned} \Gamma(\varepsilon) &= \frac{\partial \phi(a,b)}{\partial b}\mid_{(a,b)=(\varepsilon^{-1},\varepsilon)} \\ &= 1 + a_{(a,b)=(\varepsilon^{-1},\varepsilon)} \\ &= 1 + \frac{-\varepsilon}{1+\varepsilon} \\ &= \frac{1}{1+\varepsilon}, \end{aligned}$$

and

$$\xi(x) = \frac{\partial \varphi}{\partial \varepsilon}(x,\varepsilon)\mid_{\varepsilon=0} = (x_1, 2(1+\varepsilon)x_2)\mid_{\varepsilon=0} = (x_1, 2x_2) = (\xi_1,\xi_2).$$

So

$$\left\{ \begin{array}{lll} \dfrac{dy}{d\varepsilon} &=& \Gamma(\varepsilon)\xi(y) \\ y(0) &=& x \end{array} \right. \qquad \textit{hence} \qquad \left\{ \begin{array}{lll} \dfrac{dy_1}{d\varepsilon} &=& \dfrac{1}{1+\varepsilon} y_1 \\ \dfrac{dy_2}{d\varepsilon} &=& \dfrac{1}{1+\varepsilon} 2y_2 \end{array} \right.$$

Consequently, the parametrization is given by

$$\tau(\varepsilon) = \int_0^{\varepsilon} \Gamma(t)dt = \int_0^{\varepsilon} \frac{1}{1+t}dt = \ln(1+\varepsilon),$$

hence, $\varepsilon = e^{\tau} - 1$ and the parametrized group is obtain to be of the forme

$$\left\{ \begin{array}{lll} y_1 &=& e^{\tau} x_1 \\ y_2 &=& e^{2\tau} y_1, \end{array} \right. \qquad \text{with} \qquad \tau \in\,]-\infty,+\infty[$$

and the low of composition for this parametrized group is $\phi(a,b) = a+b$.

3.2 Infinitesimal generators

Definition 3.6. Let $\varphi(x,\varepsilon)$ be a one parameter group of transformations. The infinitesimal generator of φ is defined by the operator

$$X = X(x) := \Sigma_{j=1}^{n} \xi_j(x) \frac{\partial}{\partial x_j}. \tag{6}$$

The operator X is also known as the vector field of the one parameter Lie group of transformations (1) and if g is a differentiable function so

$$X f(x) = \Sigma_{j=1}^{n} \xi_j(x) \frac{\partial f(x)}{\partial x_j}.$$

Paricularly, for $f(x) = x$ we obtain

$$X f(x) = \xi(x).$$

The relation between the Lie symmetry vector field and its corresponding one parameter Lie group of transformations is given by the following theorem

Theorem 3.7. *The one-parameter Lie group of transformations (1) can be written as*

$$\begin{aligned} y = \varphi(x,\varepsilon) &= e^{\varepsilon X} x \\ &= \sum_{i=0}^{+\infty} \frac{\varepsilon^i}{i!} X^i(x), \end{aligned} \tag{7}$$

where the operator $X(x)$ is defined by (6) and

$$X^i f(x) = X(X^{i-1} f(x)), \quad i \geqslant 1.$$

Remark 3.8. The series (7) is known as a Lie series.

Example 3.9. Let φ be the function which defines the one parameter group of transformations acting on $\mathbb{R}^2$ by

$$\begin{aligned} y = \varphi(x,\varepsilon) &= A(\varepsilon)x \\ &= \begin{pmatrix} \cosh\varepsilon & -\sinh\varepsilon \\ -\sinh\varepsilon & \cosh\varepsilon \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \end{aligned} \tag{8}$$

The infinitesimals are given by

$$\begin{aligned} \xi(x) &= (\xi_1(x_1,x_2), \xi_2(x_1,x_2)) \\ &= \left(\frac{\partial \varphi_1(x,\varepsilon)}{\partial \varepsilon}\Big|_{\varepsilon=0}, \frac{\partial \varphi_2(x,\varepsilon)}{\partial \varepsilon}\Big|_{\varepsilon=0}\right) \\ &= (-x_2, -x_1). \end{aligned}$$

Consequently,

$$X = -x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2}.$$

By using of the last theorem, we have

$$\begin{aligned} \varphi(x,\varepsilon) &= e^{\varepsilon X} x \\ &= (e^{\varepsilon X} x_1, e^{\varepsilon X} x_2), \end{aligned}$$

with

$$\begin{aligned} e^{\varepsilon X}x_1 &= \Sigma_{k=0}^{+\infty}\frac{\varepsilon^k}{k!}X^k x_1 \\ &= x_1 + \varepsilon X x_1 + \frac{\varepsilon^2}{2!}X(Xx_1) + \cdots \end{aligned}$$

as

$$X(x_1) = (-x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2})(x_1) = -x_2,$$

and

$$X(x_2) = (-x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2})(x_2) = -x_1,$$

then,

$$\begin{aligned} e^{\varepsilon X}x_1 &= x_1 - \varepsilon x_2 + \frac{\varepsilon^2}{2!}x_1 - \frac{\varepsilon^3}{3!}x_2 + \cdots \\ &= x_1(1 + \frac{\varepsilon^2}{2!} + \frac{\varepsilon^4}{4!} + \cdots) - x_2(\varepsilon + \frac{\varepsilon^3}{3!} + \frac{\varepsilon^5}{5!} + \cdots) \\ &= x_1 \cosh\varepsilon - x_2 \sinh\varepsilon \\ &= \varphi_1(x, \varepsilon). \end{aligned}$$

In the same way we obtain $e^{\varepsilon X}x_2 = \varphi_2(x, \varepsilon)$.

Definition 3.10 (Infinitesimal matrix of an r-parameter Lie group of transformations)**.** We consider an r-parameter Lie group of transformations given by

$$y = \varphi(x, \varepsilon), \tag{9}$$

assumed to be analytic in its domain of definition where $y = (y_1, y_2, \ldots, y_m)$, $x = (x_1, x_2, \ldots, x_m)$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_r)$ and $\varphi = (\varphi_1, \ldots, \varphi_r)$. The corresponding infinitesimal generators are given by

$$X_i = \Sigma_{i=1}^m \xi_{ij}(x)\frac{\partial}{\partial x_j},$$

with $\xi_{ij}(x) = \dfrac{\partial\varphi_j(x,\varepsilon)}{\partial\varepsilon_i}|_\varepsilon = 0, \qquad i = 1, \ldots, r$ and $j = 1, \ldots, m$.

Theorem 3.11. *The infinitesimal generators of an r-parameter Lie group of transformations form a Lie algebra.*

Definition 3.12. Let $\mathcal{G}$ be a Lie algebra with basis elements $\{X_1, \ldots, X_r\}$. The commutators may be expressed in terms of the basis elements:

$$\left[X_\alpha, X_\beta\right] = \Sigma_{i=1}^r C_{\alpha\beta}^i X_i, \qquad \alpha, \beta = 1, \ldots, r,$$

where $C_{\alpha\beta}^i$ are called the structure constants.

Now we recall the second and third fondamental theorems of Lie.

Theorem 3.13. *The structure constants $C_{\alpha\beta}^i$ are constants.*

Theorem 3.14. *The structure constants $C_{\alpha\beta}^i$ satisfy the relations:*

1. $C^i_{\alpha\beta} = -C^i_{\beta\alpha}$,

2. $\Sigma^r_{i=1}\left(C^i_{\alpha\beta}C^m_{in} + C^i_{\alpha n}C^m_{i\alpha} + C^i_{n\alpha}C^m_{i\beta}\right) = 0$,

where $\alpha, \beta, m, n = 1, \ldots, r$.

Example 3.15. We consider the 3-parameter group (Euclidean group) $E(2)$ acting on $\mathbb{R}^2$as follows:

$$Y = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \varphi(X,\varepsilon) = \begin{pmatrix} \cos\varepsilon_1 & -\sin\varepsilon_1 \\ -\sin\varepsilon_1 & \cos\varepsilon_1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}.$$

The 3-dimentional Lie algebra $e(2)$ for $E(2)$ is spanned by

$$X_1 = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}, \qquad X_2 = \frac{\partial}{\partial x}, \qquad X_3 = \frac{\partial}{\partial y},$$

and its commutaor table is given by

$[X_i, X_j]$	X_1	X_2	X_3
X_1	0	$-X_3$	X_2
X_2	X_3	0	0
X_3	$-X_2$	0	0

Example 3.16 (Projective Lie group transformations in $\mathbb{R}^2$)**.** Projective transformations in $\mathbb{R}^2$ is defined by:

$$Y = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \varphi(X,\varepsilon) = \begin{pmatrix} \dfrac{(1+\varepsilon_3)x + \varepsilon_4 y + \varepsilon_5}{\varepsilon_1 x + \varepsilon_2 y + 1} \\ \dfrac{\varepsilon_6 x + (1+\varepsilon_7)y + \varepsilon_8}{\varepsilon_1 x + \varepsilon_2 y + 1} \end{pmatrix}.$$

The infinitesimal generators of the corresponding 8-dimensional Lie algebra are

$$X_1 = x^2\frac{\partial}{\partial x} + xy\frac{\partial}{\partial y}, \qquad X_2 = xy\frac{\partial}{\partial x} + y^2\frac{\partial}{\partial y}, \qquad X_3 = x\frac{\partial}{\partial x}, \qquad X_4 = y\frac{\partial}{\partial x},$$

$$X_5 = \frac{\partial}{\partial x}, \qquad X_6 = x\frac{\partial}{\partial y}, \qquad X_7 = y\frac{\partial}{\partial y}, \qquad X_8 = \frac{\partial}{\partial y}.$$

The following commutator table is find to be as follows:

	X_1	X_2	X_3	X_4
X_1	0	0	$-X_1$	$-X_2$
X_2	0	0	0	0
X_3	X_1	0	0	$-X_4$
X_4	X_2	0	X_4	0
X_5	$2X_3 + X_7$	X_4	X_5	0
X_6	0	X_1	$-X_6$	$X_3 - X_7$
X_7	0	X_2	0	X_4
X_8	X_6	$X_3 + 2X_7$	0	X_5

	X_5	X_6	X_7	X_8
X_1	$-2X_3 - X_7$	0	0	$-X_6$
X_2	$-X_4$	$-X_1$	$-X_2$	$-X_3 - 2X_7$
X_3	$-X_5$	X_6	0	0
X_4	0	$X_7 - X_3$	$-X_4$	$-X_5$
X_5	0	X_8	0	0
X_6	$-X_8$	0	X_6	0
X_7	0	$-X_6$	0	$-X_8$
X_8	0	0	X_8	0

From the above table, and the expression

$$[X_4, X_3] = \sum_{i=1}^{8} C_{43}^{i},$$

we get

$$C_{43}^{4} = 1, \qquad \text{and} \qquad C_{43}^{i} = 0, \quad i = 1,2,3,5,6,7,8.$$

4 Invariance of differential equations

In this section, the process of invariance of a system of differential equations will be clarified. First, through basic results illustrated with examples especially, the rotation group $SO(2)$. This group will be used frequently to clarify both the prolongation of a group action and the prolongation of a vector field. Second, through the invariance criterion, the latter will be of great interest, it allows to overcome the difficulties imposed by complicated calculations that appear during the prolongation of a given group action.

4.1 Systems of differential equations

Definition 4.1. A system of n-th order differential equations in p independent and q dependent variables is given as a system of equations

$$\Delta_i(x, u^{(n)}) = 0, \qquad i = 1, \ldots, l \tag{10}$$

involving $x = (x_1, \ldots, x_p), u = (u^1, \ldots, u^q)$ and the derivatives of u with respect to x up to order n.

In order to represent a system of differential equations with a vanishing of certain functions, we need to prolong the basic space $X \times U$ of independent and dependent variables to the jet space $X \times U^{(n)}$ that represents the various partial derivatives occuring in the system.

Remark 4.2. The functions $\Delta(x, u^{(n)}) = (\Delta_1(x, u^{(n)}, \ldots, \Delta_l(x, u^{(n)}))$ will be assumed to be smooth in their arguments, hence Δ can be viewed as a smooth map from the jet space $X \times U^{(n)}$ to some $\mathbb{R}^l$.

Example 4.3 ($p = 2, q = 1$)**.** In this case $X = \mathbb{R}^2$, and $U = \mathbb{R}$, then $U^{(n)} = U \times U_1 \times \ldots \times U_n$ with $U_k \simeq \mathbb{R}^{k+1}$ represents the $k+1$ distinct k-th order partial derivatives of u. Particularly, we have $U^{(2)} = U \times U_1 \times U_2 \simeq \mathbb{R}^6$ with coordinates $u^{(2)} = (u, u_x, u_y, u_{xx}, u_{xy}, u_{yy})$ represents all derivatives of u with respect to x and y of order at most 2.

4.2 The n-th prolongation of a function

Definition 4.4. Let be $f = u(x) : X \subset \mathbb{R}^p \to U \subset \mathbb{R}^q$ a smooth function. The induced function $u^{(n)} = Pr^{(n)} f(x)$ called the n-th prolongation of f is given by

$$u_J^\alpha = \partial_J f^\alpha(x) = \frac{\partial^k f(x)}{\partial x^{J_1} \partial x^{J_2} \cdots \partial x^{J_k}}, \qquad \alpha = 1, \ldots, q, \tag{11}$$

where $J = (J_1, \ldots, J_k)$ is a k-th order multi-index with the order denoted by $\neq J = k$, $1 \leqslant j_k \leqslant p$ for each k and u_J^α represents all the different k-th order derivatives of the components of f at a point x.

So, $Pr^{(n)} f : X \to U^{(n)}$ is a function from X to the space $U^{(n)}$ that represents all its derivatives up to order n at the point x.

Example 4.5 ($p = 2, q = 1$)**.** . In this case, for a given function $f(x,y)$ we get

$$Pr^{(1)} f(x,y) = (f(x,y), \frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)),$$

and

$$Pr^{(2)} f(x,y) = (f(x,y), \frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)), \frac{\partial^2 f}{\partial x^2}(x,y), \frac{\partial^2 f}{\partial x \partial y}(x,y), \frac{\partial^2 f}{\partial y}(x,y)),$$

4.3 Lie group transformation of a function

To illustrate how a group G transforms a function f, we identify the function f with its graph Γ_f:

$$\Gamma_f = \left\{(x, f(x)) / x \in D_f\right\} \subset X \times U,$$

where D_f is the domain of definition of f.

Definition 4.6. Let Γ_f be the graph of a function $f : X \to U$ and assuming Γ_f to be contained in the domain of definition of the group transformation g. Then, the transformation of Γ_f by g is given by

$$g.\Gamma_f = \left\{(\tilde{x}, \tilde{u}) = g.(x,u) / (x,u) \in \Gamma_f\right\}.$$

Definition 4.7. The set $g.\Gamma_f$ is not necessarily the graph of another function $\tilde{u} = \tilde{f}(\tilde{x})$. However, we assume that for an element g near the identity, the transformation $g.\Gamma_f = \Gamma_{\tilde{f}}$ is the graph of some function $\tilde{u} = \tilde{f}(\tilde{x})$. We write $\tilde{f} = g.f$ and call the function $\tilde{f}$ the transformation of f by g.

Example 4.8 ($G = SO(2), X = \mathbb{R}, U = \mathbb{R}$)**.** The group G acts on $X \times U$ as follows:

$$\theta.(x,u) = (x\cos\theta - u\sin\theta, x\sin\theta + u\cos\theta).$$

In this case, $\theta.\Gamma_u$ which is the rotated graph is not necessarily the graph of a single valuted function. However, it is for a finite interval of the independent variable x and θ to be not too large.

Particularly, if $f(x) = ax + b, \quad a \neq 0$, then the rotated graph $\theta.\Gamma_f$ will be another straight line which is the graph of another function provided $\theta.\Gamma_f$ is not vertical.

As a point (x,u) in Γ_f is rotated to the point

$$(\tilde{x}, \tilde{u}) = (x\cos\theta - (ax+b)\sin\theta, x\sin\theta + (ax+b)\cos\theta).$$

So provided $\cot\theta \neq 0$ (which is possible. In particular, for θ sufficently near 0) we obtain [12, 2]

$$\tilde{u} = \tilde{f}(\tilde{x}) = \frac{\sin\theta + a\cos\theta}{\cos\theta - a\sin\theta}\tilde{x} + \frac{b}{\cos\theta - a\sin\theta},$$

which is a linear function.

4.4 The n-th prolongation of a group action

Now, let G be a group action on an open subset M in the space of independent variables X and dependent variables U.

Definition 4.9. The n-th prolongation of the action of G on M denoted $Pr^{(n)}G$ is so that it transforms the derivatives of a function f into the corresponding derivatives of the transformed function $\tilde{f}$.

Remark 4.10. The n-th prolongation of the action of G on M is an induced local action of G on the n-jet space $M^{(n)}$. In other words, if $g \in G$ is sufficiently near the identity so let $u = f(x)$ to be a smooth function defined on a neighbourhood of x_0. Then, $g.f$ is defined in a neibourhood of $\tilde{x_0}$ such that $(\tilde{x_0}, \tilde{u_0}) = g.(x_0, u_0)$ and $Pr^{(n)}g$ is determined by evaluating the drivatives of the transformed function $g.f$ at $\tilde{x_0}$:

$$Pr^{(n)}g.(x_0, u_{0^{(n)}}) = (\tilde{x_0}, \tilde{u_0}^{(n)}),$$

where

$$\tilde{u_0}^{(n)} = Pr^{(n)}(g.f)(\tilde{x_0}).$$

Remark 4.11. For the existence of a function f, it is garanted by the possibility to choose the function f to be the n-th order Taylor polynomial at x_0 corresponding to the given values $u_0^{(n)}$.

Example 4.12 ($Pr^{(1)}G$, $G = SO(2)$, $X = \mathbb{R}$, $U = \mathbb{R}$)**.** Let (x^0, u^0, u_x^0) be a point in the jet-space $X \times U^{(1)}$ with the Taylor polynomial

$$f(x) = u^0 + u_x^0(x - x^0).$$

Then, we have

$$f(x^0) = u^0 \qquad \text{and} \qquad f'(x^0) = u_x^0.$$

As done in the previous example, the transformed function $\tilde{f} = \varepsilon.f$ is given by

$$\tilde{f}(\tilde{x}) = \frac{\sin\varepsilon + u_x^0 \cos\varepsilon}{\cos\varepsilon - u_x^0 \sin\varepsilon}\tilde{x} + \frac{u^0 - u_x^0 x^0}{\cos\varepsilon - u_x^0 \sin\varepsilon}.$$

Hence,

$$\tilde{x^0} = x^0\cos\varepsilon - u^0 \sin\varepsilon, \qquad \tilde{u^0} = x^0 \sin\varepsilon + u^0\cos\varepsilon,$$

and

$$\tilde{u_x^0} = \tilde{f}'(\tilde{x^0}) = \frac{\sin\varepsilon + u_x^0\cos\varepsilon}{\cos\varepsilon - u_x^0 \sin\varepsilon}, \qquad \text{provided} \quad u_x^0 \neq \cot\varepsilon.$$

Consequently,

$$Pr^{(1)}\varepsilon.(x, u, u_x) = (x\cos\varepsilon - u\sin\varepsilon, x\sin\varepsilon + u\cos\varepsilon, \frac{\sin\varepsilon + u_x\cos\varepsilon}{\cos\varepsilon - u_x\sin\varepsilon}),$$

provided it exists.

4.5 The n-th prolongation of a vector fields

We will define the prolongation of a vector field as the infinitesimal generator of the prolongation of its corresponding group action.

Definition 4.13. Let V be a vector field on an open subset $M \subset X \times U$, with corresponding (local) one parameter group $\exp(\varepsilon V)$. The n-th prolongation of V denotted $Pr^{(n)}V$ will be a vector field on the n-jet space $M^{(n)}$ and is defined to be the infinitesimal generator of the corresponding prolonged one-parameter group $Pr^{(n)}[\exp(\varepsilon V)]$. i.e:

$$Pr^{(n)}V_{|(x,u^{(n)})} = \frac{d}{d\varepsilon}_{|_{\varepsilon=0}} Pr^{(n)}[\exp(\varepsilon V)](x,u^{(n)}),$$

for any $(x,u^{(n)}) \in M^{(n)}$.

Example 4.14 ($Pr^{(1)}G$, $G = SO(2)$, $X = \mathbb{R}$, $U = \mathbb{R}$)**.** For thsi group, the corresponding infinitesimal generator is

$$V = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u},$$

and

$$\exp(\varepsilon V)(x,u) = (x\cos\varepsilon - u\sin\varepsilon, x\sin\varepsilon + u\cos\varepsilon).$$

As discussed before, the first prolongation of $SO(2)$ is given by:

$$Pr^{(1)}\varepsilon.(x,u,u_x) = \left(x\cos\varepsilon - u\sin\varepsilon, x\sin\varepsilon + u\cos\varepsilon, \frac{\sin\varepsilon + u_x\cos\varepsilon}{\cos\varepsilon - u_x\sin\varepsilon}\right).$$

Consequently,

$$\begin{aligned} Pr^{(1)}V_{(x,u,u_x)} &= \frac{d}{d\varepsilon}_{|_{\varepsilon=0}} Pr^{(1)}[\exp(\varepsilon V)](x,u,u_x) \\ &= \left(-u, x, 1+u_x^2\right), \end{aligned}$$

so

$$Pr^{(1)}V_{(x,u,u_x)} = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u} + (1+u_x^2)\frac{\partial}{\partial u_x}.$$

Finding the n-th prolongation of an arbitrary group action is often complicate. However, we have a prolongation formulae which allows us to get the n-th prolongation of a given vector field given by the following theorem.

Theorem 4.15. *[12][Prolongation formula] Let*

$$V = \Sigma_{i=1}^{p}\xi^i(x,u)\frac{\partial}{\partial x^i} + \Sigma_{\alpha=1}^{q}\varphi_\alpha(x,u)\frac{\partial}{\partial u^\alpha},$$

be a vector field defined on an open subset $M \subset X \times U$. The n-th prolongation of V is the vector field

$$Pr^{(n)}V = V + \Sigma_{\alpha=1}^{q}\Sigma_J\varphi_\alpha^J(x,u^{(n)})\frac{\partial}{\partial u_J^\alpha},$$

defined on the corresponding jet-space $M^{(n)} \subset X \times U^{(n)}$. The second summation being over all (unordered) multi-indices $J = (J_1,\ldots,J_k)$ with $1 \leqslant J_k \leqslant p$, $1 \leqslant k \leqslant n$. The coefficient functions φ_α^J of $Pr^{(n)}V$ are given by the following formula:

$$\varphi_\alpha^J(x,u^{(n)}) = D_J\left(\varphi_\alpha - \Sigma_{i=1}^{p}\xi^i u_i^\alpha\right) + \Sigma_{i=1}^{p}\xi^i u_{J,i}^\alpha,$$

where $u_i^\alpha = \dfrac{\partial u^\alpha}{\partial x^i}$ and $u_{J,i}^\alpha = \dfrac{\partial u_J^\alpha}{\partial x^i}$.

Example 4.16 ($Pr^{(1)}G$, $G = SO(2)$, $X = \mathbb{R}$, $U = \mathbb{R}$)**.** The infinitesimal generator is

$$V = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u}.$$

The first prolongation of V is given by

$$Pr^{(1)}V_{(x,u,u_x)} = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u} + (1+u_x^2)\frac{\partial}{\partial u_x}.$$

The second prolongation of V is given by

$$Pr^{(2)}V_{(x,u,u_x,u_{xx})} = V + (1+u_x^2)\frac{\partial}{\partial u_x} + \varphi^{xx}\frac{\partial}{\partial u_{xx}},$$

with φ^{xx} is given by the prolongation formula:

$$\begin{aligned}
\varphi^{xx} &= D_{xx}(\varphi - \xi u_x) + \xi u_{xxx} \\
&= D_{xx}(x + u u_x) - u u_{xxx} \\
&= D_x\left(1 + u_x^2 + u u_{xx}\right) - u u_{xxx} \\
&= 2u_x u_{xx} + u_x u_{xx} + u u_{xxx} - u u_{xxx} \\
&= 3u_x u_{xx}.
\end{aligned}$$

Then,

$$Pr^{(2)}V_{(x,u,u_x,u_{xx})} = V + (1+u_x^2)\frac{\partial}{\partial u_x} + 3u_x u_{xx}\frac{\partial}{\partial u_{xx}}.$$

Theorem 4.17. *[2] Let $Pr^{(n)}V, Pr^{(n)}X$ be the n-th prolongation of the infinitesimal generators V, X and let $Pr^{(n)}[V,X]$ be the n-th prolongation of the commutator $[V,X]$. Then*

$$Pr^{(n)}[V,X] = [Pr^{(n)}V, Pr^{(n)}X], \qquad n = 1,2,\ldots$$

Hence, if $[V,X] = Y$, so $[Pr^{(n)}V, Pr^{(n)}X] = Pr^{(n)}Y$.

4.6 Invariance of differential equations

Definition 4.18. Let Δ be a system of differential equations. A symmetry group of the system Δ is a local group of transformations G acting on an open subset M of $X \times U$ (the space of independent variables and dependent variables) with the property that it transforms solutions of Δ to other solutions.

More precisely, whenever f is a solution of Δ and whenever $\tilde{f} = g.f$ is defined for $g \in G$, then $\tilde{f} = g.f$ is also a solution of Δ.

Theorem 4.19. *[12] Let Δ be an n-th order system of differential equations defined on an open subset M of $X \times U$. If G is a local group of transformations acting on M such that whenever*

$$(x, u^{(n)}) \in \left\{(x, u^{(n)}) \in M^{(n)} : \Delta(x, u^{(n)}) = 0\right\},$$

we have $Pr^{(n)}g.(x, u^{(n)})$ is a solution for all $g \in G$ (provided $Pr^{(n)}g.(x, u^{(n)})$ is defined). Then G is a symmetry group of the system Δ.

As already mentiened, the complexity of prolonging a group can be overcome. Then, instead of passing through the extension of a group action we have a result called invariance criterion and which directly uses the prolongation formula of a vector field.

Theorem 4.20 (Invariance criterion)**.** *[12, 2] Let*

$$\Delta_\nu(x, u^{(n)}) = 0, \qquad \nu = 1, \ldots, l$$

be an n-th order system of differential equations of maximal rank defined over an open subset $M \subset X \times U$. If G is a local group of transformations acting on M, and

$$Pr^{(n)}V\left[\Delta_\nu(x, u^{(n)})\right] = 0, \qquad \nu = 1, \ldots, l$$

whenever $\Delta(x, u^{(n)}) = 0$, for every infinitesimal generator V of G. Then G is a symmetry group of the system Δ.

5 Lie symmetry algebra of some differential equations

This section is devoted to present some applications of Lie symmetry method including: construction of similarity solutions, linearization and so on.

5.1 Invariance and some exact solutions of Thomas equation

Proposed by Thomas [14] and has the form

$$u_{xy} + \alpha u_x + \beta u_y + \gamma u_x u_y = 0, \tag{12}$$

where α, β and γ are constants. For exchange processus $\alpha > 0, \beta > 0$ and $\gamma \neq 0$.
A general vector field on the space of independent and dependent variables is written in the form

$$X = \xi(x,y,u)\frac{\partial}{\partial x} + \eta(x,y,u)\frac{\partial}{\partial t} + \varphi(x,y,u)\frac{\partial}{\partial u},$$

where ξ, η and φ are the infinitesimals and depend on x, y and u. The invariance criterion of the Thomas equation under an infinitesimal generator is given by

$$Pr^2(\Delta) = 0 \qquad \text{whenever} \qquad \Delta = 0,$$

with $\Delta = u_{xy} + \alpha u_x + \beta u_y + \gamma u_x u_y$. The obtained system of determining equations for the symmetry group from the invariance criterion leads to the explicite expressions of the infinitesimals ξ, η and φ to be the following:

$$\begin{aligned}
\xi &= -k\gamma x + c, \\
\eta &= k\gamma y + b, \\
\varphi &= -\frac{g(x,y)}{\gamma}e^{-\gamma u} + k(\beta x - \alpha y) + a,
\end{aligned}$$

where a, b, c and k are arbitrary constants and $g(x, y)$ an arbitrary solution of the Thomas equation. Consequently, the Lie symmetry algebra of Thomas equation is spanned by the four vectors fields [14]:

$$\begin{aligned}
X_1 &= \frac{\partial}{\partial x}, \\
X_2 &= \frac{\partial}{\partial y}, \\
X_3 &= \frac{\partial}{\partial u}, \\
X_4 &= -\gamma x\frac{\partial}{\partial x} + \gamma y\frac{\partial}{\partial y} + (\beta x - \alpha y)\frac{\partial}{\partial u},
\end{aligned}$$

and the infinite dimensional subalgebra :

$$X_g = -\frac{g(x,y)}{\gamma}\exp(-\gamma u)\frac{\partial}{\partial u},$$

with $g(x,y)$ is an arbitrary solution of the Thomas equation.

For example, the one-parameter groups G_4 and G_g generated respectively by X_4 and X_g are given by:

$$\begin{aligned} G_4 : \exp(\varepsilon X_4).(x,y,u) &= \left(xe^{-\gamma\varepsilon}, ye^{\gamma\varepsilon}, \frac{\beta}{\gamma}x(1-e^{-\gamma\varepsilon}) + \frac{\beta}{\gamma}y(1-e^{\gamma\varepsilon}) + u\right), \\ G_g : \exp(\varepsilon X_g).(x,y,u) &= \left(x, y, \frac{1}{\gamma}\ln\left(\gamma g(x,y)\varepsilon + e^{\gamma u}\right)\right). \end{aligned}$$

Consequently, if $u = f(x,y)$ is a solution of the Thomas equation, so are the functions :

$$u_4 = \frac{\beta}{\gamma}x(e^{\gamma\varepsilon}-1) + \frac{\beta}{\gamma}y(e^{-\gamma\varepsilon}-1) + f(xe^{\gamma\varepsilon}, ye^{-\gamma\varepsilon}),$$

and

$$u_g = \frac{1}{\gamma}\ln\left(\gamma g(x,y)\varepsilon + e^{\gamma f(x,y)}\right),$$

where ε is a parameter groups.
The commutation relations are given in the following table :

$[X_i, X_j]$	X_1	X_2	X_3	X_4	X_g
X_1	0	0	0	$-\gamma X_1 + \beta X_3$	X_{g_x}
X_2	0	0	0	$\gamma X_2 - \alpha X_3$	X_{g_y}
X_3	0	0	0	0	$X_{-\gamma g}$
X_4	$\gamma X_1 - \beta X_3$	$-\gamma X_2 + \alpha X_3$	0	0	X_ψ
X_g	$-X_{g_x}$	$-X_{g_y}$	$X_{\gamma g}$	$-X_\psi$	0

where $\psi = -\gamma x f_x + \gamma y f_y - \gamma(\beta x - \alpha y)f$.

Using the adjoint representation and the Lie series, we listed some invariant solutions from an optimal system constructed through generators X_1, X_2 and X_3 (for more details see [14]).

Every one-dimensional subalgebra of $\mathcal{G}$ is determined by a nonzero vector field X of the form :

$$X = a_1X_1 + a_2X_2 + a_3X_3,$$

where $a_1, \ldots, a_4$ are arbitrary constants. For more details of simplifications and different reductions of the studied equation (12) see [14]).

Case.1 ($a_3 = 1, a_1a_2 \neq 0$ and $(\alpha a_1 - \beta a_1 + \gamma)^2 + 4\gamma a_1\beta \geqslant 0$)

The invariant solution in this case is obtained to be of the form:

$$u(x,y) = \frac{1}{\gamma}\ln\left(A - \frac{\gamma}{C}e^{-C(a_2x - a_1y)}\right) + \theta_0(a_2x - a_1y) + \frac{y}{a_2} + cte,$$

where A is an arbitrary constant, θ_0 satisfies the second order equation

$$(\alpha a_1 - \beta a_1 + \gamma)z - a_1 a_2 \gamma z^2 + \frac{\beta}{a_2} = 0,$$

and C is given by

$$C = \frac{2a_1 a_2 \gamma \theta_0 - \alpha a_2 + \beta a_1 - \gamma}{a_1 a_2}.$$

Case.2 ($a_3 = 1, a_1 a_2 \neq 0$ and $(\alpha a_1 - \beta a_1 + \gamma)^2 + 4\gamma a_1 \beta < 0$)

In this case, the corresponding exact solution is given by the following:

$$u(x,y) = \frac{1}{2A_1} \ln\Big(1 + \tan(A_2 \sqrt{K}(a_2 x - a_1 y) + A_0)\Big) - \frac{A_1}{2A_2}(a_2 x - a_1 y) + \frac{y}{a_2} + cte,$$

where $A_1 = \dfrac{\alpha a_2 - \beta a_1 + \gamma}{a_1 a_2}, A_2 = -\gamma$ and $K = \dfrac{4A_2 A_3 - A_1^2}{4A_2^2}$ with $A_3 = \dfrac{\beta}{a_1 a_2^2}$ and A_0 an arbitrary constant.

Case.3 ($a_3 = 1, a_2 = 0, a_1 \neq 0$ and $a_1 \neq \dfrac{-\gamma}{\beta}$)

The infinitesimal generator is reduced to be of the form:

$$X = a_1 \frac{\partial}{\partial x} + \frac{\partial}{\partial u},$$

and the invaraint solution obtained in this case is find to be of the form:

$$u(x,y) = \frac{x}{a_1} - \frac{\alpha}{\beta a_1 + \gamma} y + cte.$$

Case.4 ($a_3 = 0, a_2 a_1 \neq 0$ and $a_2 = \dfrac{\beta}{\alpha}$)

The obtained solution in this case is as follows:

$$u(x,y) = \frac{1}{\gamma} \ln\left(\gamma(x - \frac{y}{a_2}) + k_0\right) + cte,$$

where k_0 is an arbitrary constant.

Case.5 ($a_3 = 0, a_2 a_1 \neq 0$ and $a_2 \neq \dfrac{\beta}{\alpha}$)

In this case, the exact solution constructed from reduction technique is given by:

$$u(x,y) = -\frac{1}{\gamma} \ln\left(1 + \frac{\gamma}{\nu(\beta - a_2\alpha) e^{(\beta - a_2 \alpha)(x - \frac{y}{a_2})} - \gamma}\right) + cte,$$

where ν is an arbitrary constant.

Remark 5.1. The list of exact solutions obtained here, is not exaustive and other exact solutions can be found in [14].

5.2 Invariance and some exact solutions of Burger's equation

Here, we consider the Burger's equation given by:

$$u_t = u_{xx} + u u_x, \tag{13}$$

which is an important case of the general non linear diffusion equation [4, 3]. This equation gives a description of waves in nonviscous medium. It was firstly introduced by Burgers to describe one dimensional turbulence, and used later to study other wave phenomena [19, 5].

Let $\mathcal{G}$ be a symmetry Lie algebra admitted by the Burger's equation. It is known that it is spanned by the vector fields:

$$\begin{aligned}
X_1 &= tx\frac{\partial}{\partial x} + t^2\frac{\partial}{\partial t} - (tu + x)\frac{\partial}{\partial u},\\
X_2 &= x\frac{\partial}{\partial x} + 2t\frac{\partial}{\partial t} - u\frac{\partial}{\partial u},\\
X_3 &= t\frac{\partial}{\partial x} - \frac{\partial}{\partial u},\\
X_4 &= \frac{\partial}{\partial x},\\
X_5 &= \frac{\partial}{\partial t}.
\end{aligned}$$

A general element X of this algebra is written as

$$X = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5,$$

where $a_1, \ldots, a_5$ are constants. Utilizing the commutator table given by:

$[X_i, X_j]$	X_1	X_2	X_3	X_4	X_5
X_1	0	$-2X_1$	0	X_3	$-X_2$
X_2	$2X_1$	0	X_3	$-X_4$	$-2X_5$
X_3	0	$-X_3$	0	0	$-X_4$
X_4	$-X_3$	X_4	0	0	0
X_5	X_2	$2X_5$	X_4	0	0

and the Adjoint representation obtained by using the Lie series

$$Ad(\exp(\varepsilon X_i))X_j = X_j - \varepsilon[X_i, X_j] + \frac{\varepsilon^2}{2}[X_i, [X_i, X_j]] - \cdots$$

presented in the following table:

Ad	X_1	X_2	X_3
X_1	X_1	$X_2 + 2\varepsilon X_1$	0
X_2	$e^{-2\varepsilon}X_1$	X_2	$e^{-\varepsilon}X_3$
X_3	X_1	$X_2 + \varepsilon X_3$	X_3
X_4	$X_1 + \varepsilon X_3$	$X_2 - \varepsilon X_4$	X_3
X_5	$X_1 - \varepsilon X_2 + \varepsilon^2 X_5$	$e^{-2\varepsilon}X_2$	$X_3 - \varepsilon X_4$

Ad	X_4	X_5
X_1	$X_4 - \varepsilon X_3$	$X_5 + \varepsilon X_2 + \varepsilon^2 X_1$
X_2	$e^{\varepsilon} X_4$	$e^{2\varepsilon} X_5$
X_3	X_4	$X_5 + \varepsilon X_4$
X_4	X_4	X_5
X_5	X_4	X_5

Employing the above table, a list of inequivalent generators that constitutes an optimal system [13] is given by:

$$\begin{aligned} Y_1 &= X_3 + X_5, \qquad Y_2 = -X_3 + X_5, \qquad Y_3 = X_5, \\ Y_4 &= X_1 + X_5, \qquad Y_5 = -X_1 + X_5, \qquad Y_6 = \beta X_1 + X_2 + X_4, \\ Y_7 &= \beta X_1 - X_2 + X_4, \qquad Y_8 = \beta X_1 + X_4, \qquad Y_9 = X_3, \\ Y_{10} &= X_1, \qquad Y_{11} = X_2, \end{aligned}$$

where β is an arbitrary constant.
Now, let us present a family of exact solutions obtained from some generators listed above. We note that, the exposed exact solutions are limited to the expressed explicitly ones. Furthermore, the explicitly solutions listed here are given directly without presenting the constructed similarity variables and corresponding reduced equations. For solutions written in terms of special functions: Airy, Airy Bi, confluent hypergeometric functions, Laguerre and Hermite polynomials see [16].

The corresponding solution of the infinitesimal generator X_5

The solution of the equation is given by

$$u(x,t) = \begin{cases} \dfrac{2}{x+k_1}, & \\ \left(k_2 \exp(\sqrt{b}x) - \dfrac{1}{2\sqrt{b}}\right)^{-1} + \sqrt{b}, & b > 0 \\ \sqrt{-2b}\tan\left(\dfrac{\sqrt{-2b}}{2}x + k_3\right), & b < 0 \end{cases}$$

where k_1, k_2 and k_3 are arbitrary constants (provided the expression is defined).

The corresponding solution of the infinitesimal generators $X_1 + X_5$ and $-X_1 + X_5$

In these cases, the solutions are written in terms of special functions precisely Airy and Airy Bi functions. However, specific values of constants arising in the general solutions yield exact solutions given by :

$$u_1(x,t) = -\frac{t}{1+t^2} + i\frac{x}{1+t^2}, \qquad u_2(x,t) = -\frac{x}{1+t}.$$

The corresponding solution of the infinitesimal generator $\beta X_1 + X_2 + X_4$

As in the previous case, the general solution is written in terms of special functions precisely Hermite polynomial and confluent hypergeometric functions. Hence, for specific values of arbitrary constants appeared in this general solution, an explicite form of an exact solution is given by:

$$u(x,t) = \frac{\beta(1-x))}{\beta t + 2} - \frac{2(x + \beta t + 1)}{\beta t^2 + 2t}.$$

The corresponding solution of the infinitesimal generator $\beta X_1 - X_2 + X_4$

In this case, the general solution is written in terms of Hermite polynomial and confluent hypergeometric functions. Hence, for some specific values of arbitrary constants appeared in this general solution an explicite solution is given by:

$$u(x,t) = \frac{4(x+\beta t-1)}{(x+\beta t-1)^2-(\beta t^2-2t)} - \frac{\beta t(x+1)}{\beta t^2-2t}.$$

The corresponding solution of the infinitesimal generator X_3

The invariant solution of the Burger's equation corresponding to X_3 is given by:

$$u(x,t) = \frac{\mu - x}{t},$$

with μ is an arbitrary constant.

The corresponding solution of the infinitesimal generator X_1

In this case, a new solution is obtained after a reduced differential equation is solved and it is of the following form:

$$u(x,t) = \begin{cases} \dfrac{2}{x+k_1 t} - \dfrac{x}{t}, & \\ \left(k_2 t \exp(\sqrt{b}\dfrac{x}{t}) - \dfrac{t}{2\sqrt{b}}\right)^{-1} + \dfrac{\sqrt{b}-x}{t}, & b>0 \\ \dfrac{\sqrt{-2b}\tan\left(-\dfrac{\sqrt{-2b}}{2t}x+k_3\right)-x}{t}, & b<0 \end{cases}$$

where k_1, k_2, k_3 are arbitrary constants and b a constant of integration appeared in the reduced equation.

The corresponding solution of the infinitesimal generator X_2

In this last case, the general solution is written in terms of special functions precisely hypergeometric function and Laguerre polynomial. However, a specific values of constants arising in the general solution yield an exact solutions given by:

$$u_1(x,t) = \frac{4t}{6t-x^2} + \frac{2t}{x^2} - 1.$$

Remark 5.2. We can generat, other new explicit solutions admitted by the Burger's equation by considering specific values of arbitrary constants appearing in the general forme of solutions.

5.3 Invariance and linearization of $g(u)$-Burger's equation

It is known that the Burger's equation studied in the precedent example can be linearized in the sens that it can be transformed into a linear equation especially the heat equation. The transformation used to linearized it was given by $w = e^u$. In this section, we will study the $g(u)$-Burger's equation given by

$$u_t = u_{xx} + g(u)u_x, \tag{14}$$

where $g(u) = f'(u)$ is a smooth function depending only on the dependent variable u. The results obained here were generalized in [11] and in [6].

The choice of a specific expressions of $g(u)$ leads to a particular cases of our equation:

$g(u) = 0$, the $g(u)$-Burger's equation becomes the standart heat equation:

$$u_t = u_{xx}. \tag{15}$$

$g(u) = 1$, in this case we obtain the standard Buger's equation

$$u_t = u_{xx} + u_x^2. \tag{16}$$

$g(u) = u^{-1}$, this choice of $g(u)$ leads to the modified Burger's equation

$$u_t = u_{xx} + u^{-1}u_x^2. \tag{17}$$

According to the invariance criterion

$$Pr^2X(u_t - u_{xx} - g(u)u_x) = 0,$$

whenever, $u_t = u_{xx} + g(u)u_x$ and an infinitesimal generatot X of a one parameter group is given by

$$X = \xi(x,y,u)\frac{\partial}{\partial x} + \eta(x,y,u)\frac{\partial}{\partial t} + \varphi(x,y,u)\frac{\partial}{\partial u},$$

we obtain the general form of infinitesimals ξ, η and φ to be of the form

$$\begin{aligned}
\xi(x,y,u) &= 4a_1tx + a_2x + 2a_4t + a_5, \\
\eta(x,y,u) &= 4a_1t^2 + 2a_2t + a_3, \\
\varphi(x,y,u) &= \beta(t,x)\exp(-f(u)) + (-a_1x^2 - a_4x - 2a_1t + a_5)h(u),
\end{aligned}$$

where $a_1,\ldots,a_5$ are arbitrary constants and β is an arbitaray solution of the heat equation.

The Lie symmetry algebra admitted by the $g(u)$-Burger's equation is spanned by the vector fields [15]

$$\begin{aligned}
X_1 &= \frac{\partial}{\partial x}, \\
X_2 &= \frac{\partial}{\partial t}, \\
X_3 &= h(u)\frac{\partial}{\partial u}, \\
X_4 &= x\frac{\partial}{\partial x} + 2t\frac{\partial}{\partial t}, \\
X_5 &= 2t\frac{\partial}{\partial x} - xh(u)\frac{\partial}{\partial u}, \\
X_6 &= 4tx\frac{\partial}{\partial x} + 4t^2\frac{\partial}{\partial t} - (2t + x^2)h(u)\frac{\partial}{\partial u},
\end{aligned}$$

and an infinite dimensional subalgebra of the form

$$X_\beta = \beta(t,x)\exp(-f(u))\frac{\partial}{\partial u},$$

where $\beta(x,t)$ is an arbitrary solution of the heat equation.

Remark 5.3. From the closure of Lie bracket, especially

$$\begin{array}{rcl} [X_1, X_\beta] & = & X_{\beta_x}, \quad [X_2, X_\beta] = X_{\beta_t}, \\ \left[X_4, X_\beta\right] & = & X_{\beta_1}, \quad [X_5, X_\beta] = X_{\beta_2}, \quad [X_6, X_\beta] = X_{\beta_3}, \end{array}$$

where

$$\begin{array}{rcl} \beta_1 & = & x\beta_x + 2t\beta_t, \\ \beta_2 & = & 2t\beta_x + x\beta, \\ \beta_3 & = & 4tx\beta_x + 4t^2\beta_t + (2t + x^2)\beta, \end{array}$$

we conclude that if $\beta(t,x)$ is a solution of the heat equation so are the functions β_1, β_2 and β_3. In this way, we can generate an infinite family of exact solutions of the Heat equation (15).

The corresponding one parameter groups of generators

For convenience, we put

$$L(s) = \ln \int \exp(f(s))ds,$$

Hence, the one parameter groups G_i generated by X_i are given as follows:

$$\begin{array}{lcl} G_1 & : & (\tilde{x}, \tilde{t}, \tilde{u}) = (x + \varepsilon, t, u); \\ G_2 & : & (\tilde{x}, \tilde{t}, \tilde{u}) = (x, t + \varepsilon, u); \\ G_3 & : & (\tilde{x}, \tilde{t}, \tilde{u}) = (2t\varepsilon + x, t, \tilde{u}), \quad \text{with} \quad L(\tilde{u}) = \varepsilon + L(u), \\ G_4 & : & (\tilde{x}, \tilde{t}, \tilde{u}) = (xe^{\varepsilon}, te^{2t}, u), \\ G_5 & : & (\tilde{x}, \tilde{t}, \tilde{u}) = (2t\varepsilon + x, t, \tilde{u}), \quad \text{with} \quad L(\tilde{u}) = -t\varepsilon^2 - x\varepsilon - L(u), \\ G_6 & : & (\tilde{x}, \tilde{t}, \tilde{u}) = (\dfrac{x}{1 - 4t\varepsilon}, \dfrac{t}{1 - 4t\varepsilon}, \tilde{u}), \end{array}$$

where

$$\exp(L(\tilde{u})) = \sqrt{1 - 4t\varepsilon}\exp\left(\frac{-\varepsilon x^2}{1 - 4t\varepsilon}\right)\exp(L(u)).$$

For the last one parameter group generated by X_β, we obtain

$$G_\beta : (\tilde{x}, \tilde{t}, \tilde{u}) = (x, t, \tilde{u}), \quad \text{with} \quad L(\tilde{u}) = -\beta\varepsilon + \exp(L(u)).$$

Exact solution from trivial ones

As each one parameter group G_i is a symmetry group of the equation, then if $u = y(x,t)$ is a solution of the $g(u)$-Burger's equation (17) so are the functions

$$\begin{array}{rcl} y_1 & = & y(x - \varepsilon, t); \\ y_2 & = & (x, t - \varepsilon); \\ y_3 & = & e^{\varepsilon} y(x,t), \\ y_4 & = & y(e^{-\varepsilon}x, e^{-2\varepsilon}t), \\ y_5 & = & y(x - 4t\varepsilon, t)\exp(-\varepsilon x + 2t\varepsilon^2), \\ y_6 & = & (1 + 8\varepsilon t)^{-\frac{1}{4}} y(\dfrac{x}{1 + 8t\varepsilon}, \dfrac{t}{1 + 8t\varepsilon})\exp(\dfrac{-\varepsilon x^2}{1 + 8t\varepsilon}), \end{array}$$

where ε is a group parameter. Now, using of G_6 and the trivial solution $y(t,x) = 1$, we get that the function

$$y_6(x,t) = (1+8\varepsilon t)^{-\frac{1}{4}} \exp(\frac{-\varepsilon x^2}{1+8t\varepsilon}),$$

is also a solution. According to the invariance under G_2, we obtain an exact solution of the $g(u)$-Burger's equation (17) given by:

$$t^{-\frac{1}{4}} \exp\left(-\frac{x^2}{8t}\right).$$

Linearization of the $g(u)$-Burger's equation

Here, we are looking for a map which transforms the $g(u)$- Burger's equation into a linear equation. The idea was introduced firstly by Bluman and Kumei [2] and requires the equation to be invariant under an infinite parameter-Lie group of transformations.

Let us recall the basic result that connects linearization and symmetries [2]

Theorem 5.4. *Let Δ be a non-linear system of differential equations that admits an infinite-parameter Lie group of transformations with an infinitesimal generator*

$$X = \xi_i(x,u)\frac{\partial}{\partial x_i} + \eta_j(x,u)\frac{\partial}{\partial u^j},$$

$$\xi_i(x,u) = \Sigma_{\sigma=1}^{q} \alpha_i^{\sigma}(x,u) F^{\sigma}(x,u), \qquad i = 1,\ldots,p \tag{18}$$

and

$$\eta_j(x,u) = \Sigma_{\sigma=1}^{q} \beta_j^{\sigma}(x,u) F^{\sigma}(x,u), \qquad j = 1,\ldots,q \tag{19}$$

with $F = (F^1,...,F^q)$ is a solution of some linear system Δ', then there exists a mapping

$$z_i = \phi_i(x,u), \quad i = 1,...,p$$

$$w_j = \theta_j(x,u), \quad j = 1,...,q$$

that transforms the given non-linear system Δ to the linear system Δ'. In such a case, the components of the mapping ϕ_i and θ_j satisfy the following system:

$$\Sigma_{i=1}^{p} \alpha_i^{\sigma}(x,u)\frac{\partial \phi_k}{\partial x_i} + \Sigma_{j=1}^{p} \beta_j^{\sigma}(x,u)\frac{\partial \phi_k}{\partial u^j} = 0, \qquad \sigma, k = 1,\ldots,q \tag{20}$$

$$\Sigma_{i=1}^{p} \alpha_i^{\sigma}(x,u)\frac{\partial \theta_l}{\partial x_i} + \Sigma_{j=1}^{p} \beta_j^{\sigma}(x,u)\frac{\partial \theta_l}{\partial u^j} = \delta^{\sigma l}, \qquad \sigma, l = 1,\ldots,q \tag{21}$$

where $\delta^{\sigma l}$ is a kronecker symbol, $\sigma, l = 1,\ldots,q$.

Example 5.5. In this example, we show that the $g(u)$-Burger's equation can be linearized by using of its infinite parameter group of transformations generated by the vector field:

$$X_\beta = \beta(t,x)\exp(-f(u))\frac{\partial}{\partial u},$$

where $\beta(x,t)$ is an arbitrary solution of the heat equation.

According to the above theorem, and equations(18) and (19), the studied $g(u)$-Burger's equation can be transformed to the heat equation. In fact, we put

$$\alpha_1^1(x,u) = \alpha_2^1(x,u) = 0, \qquad \beta_1^1(x,u) = \exp(-f(u)).$$

Hence, equation (20) becomes

$$\phi_u = 0.$$

Consequently, two independent solutions of equation (20) can be chosen as a new independent variables

$$z_1 = x, \qquad \text{and} \qquad z_2 = t.$$

According to the second equation (21) of θ

$$\exp(-f(u))\frac{\partial \theta}{\partial u} = 1,$$

we get the solution

$$\theta = \int \exp(f(u))\,du.$$

Finally, a map obtained to be of the form

$$z_1 = x, \qquad z_2 = t, \qquad \text{and} \qquad \omega = \theta = \int \exp(f(u))\,du,$$

takes the $g(u)$-Burger's equation to the linear heat equation

$$\theta_{z_2} = \theta_{z_1 z_1}.$$

Remark 5.6. For $g(u) = 1$, we retrieve the classical Hopf-Cole transformation turning the nonlinear Burgers equation into the linear heat equation.

6 Conclusion

The Lie symmetry theory is a powerful method to study differential equations. We tried in this paper to illustrate the great importance of Lie symmetry analysis through a short review on fundamental and basic results relieted to Lie groups, Lie algebras, transformations groups, infinitesimal generators and the invariance criterion of a system of differential equations. The usefulness of symmetry method was showed using some applications namely, construction of exact solutions, linearization, construction of a non trivial solutions from trivial ones and construction of explicit solutions from the closure of the symmetry algebra under Lie Bracket. In this paper, the exposed list of solutions is not exaustive. Hence, other exact solutions and applications can be found in the literature.

References

[1] G. W. Bluman and J. D. Cole, Similarity Methods for Differential Equations, in Applied Mathematical Sciences, No.13 Equations, Springer, New York, 1974.

[2] G. W. Bluman and S. Kumei, Symmetries and Differentials Equations. Springer, New York. 1989.

[3] J. M. Burgers, A Mathematical Model Illustrating the Theory of Turbulence, Advances in Applied Mechanics, Academic Press, New York. 1948.

[4] J. M. Burgers, The Nonlinear Diffusion Equation: Asymptotic Solutions and Statistical Problems (Dordrecht:Reidel), 1974.

[5] M. V. Foursov, On integrable coupled Burgers-type equations. Phys. Lett. A 272 (2000), 57–64.

[6] I. L. Freire, Note on Lie point symmetries of Burgers equations, Trends Appl. Comput. Math. 11(2) (2010), 15–157.

[7] N. H. Ibragimov, (ed.), CRC Handbook of Lie Group Anaysis of Differential Equations, Vol.I : Symmetries, Exact Solutions, and Conservations Laws, CRC Press, Boca Raton, 1994.

[8] N. H. Ibragimov, Elementary Lie Group Analysis and Ordinary Differential Equations, Wiley, New York, 1999.

[9] S. Lie, Theorie der transformationsgruppen Teubner: Leipzig, Germany, 1988.

[10] S. Lie, Uber die intergation durch bestimmte intergrals von einer Klasse Linearer partieller differentialgleichungen. Archiv der Mathematik. 6 (1881), 328–368.

[11] M. Nadjafikhah, Lie symmetries of inviscid Burgers' equation, Adv. Appl. Clifford Alg., 19 (2009), 101–112.

[12] P. J. Olver, Application of Lie groups to Differential Equation. Springer, New York. 1986.

[13] A. Ouhadan and E. H. El Kinani, An Optimal System and Symmetry Reduction of the Burgers Equation. Algebras Groups Geom., 23 (2006), 153–166.

[14] A. Ouhadan and E. H. El Kinani, Lie Symmetries and Preliminary Classification of Group-Invariant Solutions of Thomas equation, Algebras Groups Geom., 22(4) (2005), 432–446.

[15] A. Ouhadan, E. H. El Kinani, Lie symmetries of the equation $u_t(x,t) + g(u)u_x(x,t) = 0$, Adv. Appl. Clifford Alg., 17(1) (2007), 95–106.

[16] A. Ouhadan, T. Mekkaoui and E. H. Elkinani, Invariant and Some New Exact Solutions of Burgers Equation. J Nonlinear Sci, 11(1) (2011), 51–58.

[17] L. V. Ovsiannikov, Group Analysis of differential Equations. Academic Press. New York, USA. 1982.

[18] H. Stephani, Differential Equations: Their Solution Using Symmetries, Cambridge Press. 1989.

[19] V. E. Zakharov, What is Integrability? Berlin, Springer-Verlag. 1991.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 38-69
ISSN: 2820-7114

Title :

When Two Definitions of an Additive Functor of Commutative Algebras Agree

Author(s):

David E. Dobbs

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 38–69
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

When Two Definitions of an Additive Functor of Commutative Algebras Agree

David E. Dobbs
Department of Mathematics, University of Tennessee, Knoxville, Tennessee 37996-1320
e-mail: *ddobbs1@utk.edu*

Communicated by Najib Mahdou
(Received 20 October 2022, Revised 26 December 2022, Accepted 29 December 2022)

Abstract. Let R be a commutative ring and $\underline{C}_R$ the category of commutative unital R-algebras. We show that $\underline{C}_R$ is a pre-additive category if and only if R is a zero ring. When these conditions hold, a functor F from $\underline{C}_R$ to a pre-additive category $\underline{D}$ with finite products is an additive functor (in the classical sense) if and only if F is additive in the sense due to Chase-Harrison-Rosenberg (the latter sense of "additive functor" meaning that F commutes with finite products), if and only if $F(R)$ is a terminal object of $\underline{D}$. More generally, if $\underline{C}$ and $\underline{D}$ are additive categories (that is, pre-additive categories with finite products) and $F : \underline{C} \to \underline{D}$ is a functor, then F is additive if and only if F commutes with finite products. For such categories $\underline{C}$ and $\underline{D}$, we also give four other new characterizations of the additive functors $F : \underline{C} \to \underline{D}$.

Key Words: Commutative ring, unital algebra, pre-additive category, additive functor, CHR-additive functor, zero ring, sheaf.
2010 MSC: Primary 13A99; Secondary 18E05.

In memory of Jon Beck and his stimulating teaching

1 Introduction

Let $\underline{C}$ be a category, with $|\underline{C}|$ its class of objects; and for all $C_1, C_2 \in |\underline{C}|$, let $\underline{C}(C_1, C_2)$ denote the "hom-set" of all morphisms in $\underline{C}$ having domain C_1 and codomain C_2. It will be convenient to let Ab denote the category of abelian groups and (abelian) group (homo)morphisms. In the early days of homological algebra, one often said [19, page 32] (cf. also [4, page 19]) that a category $\underline{C}$ is an *additive category* if, for all $C_1, C_2 \in |\underline{C}|$, there is an "addition function" $+ = +_{C_1,C_2} : \underline{C}(C_1, C_2) \times \underline{C}(C_1, C_2) \to \underline{C}(C_1, C_2)$, with the accompanying notation $(\varphi, \psi) \mapsto \varphi + \psi := +(\varphi, \psi)$, such that $\underline{C}(C_1, C_2)$ is thereby an additive abelian group (that is, an object of Ab). Subsequently, the definition of an "additive category" evolved and now also includes, at least, the requirement that the following two properties hold for all (not necessarily pairwise distinct) objects C_1, C_2 and C_3 of $\underline{C}$:

$$f(g+h) = fg + fh \text{ for all } g, h \in \underline{C}(C_1, C_2) \text{ and all } f \in \underline{C}(C_2, C_3); \text{and}$$

$$(f+g)h = fh + gh \text{ for all } h \in \underline{C}(C_1, C_2) \text{ and all } f, g \in \underline{C}(C_2, C_3).$$

One should not regard the just-displayed properties as indicating a logical gap in either [19] or[4]. Indeed, those early texts had a special interest, for any unital ring R, in the category ${}_R$Mod consisting of unital left R-modules and R-module homomorphisms (and in its special case ${}_\mathbb{Z}$Mod = Ab); of course, the naturally occurring "addition functions" in ${}_R$Mod are given by pointwise addition (more precisely, if φ, ψ are each left R-module homomorphisms $A \to B$, then $(\varphi + \psi)(a) = \varphi(a) + \psi(a)$ for all $a \in A$) and evidently satisfy the two just-displayed properties. So, ${}_R$Mod satisfies all the above axioms/properties. However, as category theory (and, with it, homological algebra) has continued to evolve, it would be more appropriate to use the terminology "$\underline{C}$ is a *pre-additive category*" [20, page 6] or "$\underline{C}$ is an Ab-*category*" [16, page 28] to describe a category $\underline{C}$ that satisfies all the above properties. Pre-additive categories will suffice as a general context for the ring-theoretic work that is the main object of this note. (For the sake of completeness, let us record that it is nowadays commonly agreed to define an *additive category* as a pre-additive category $\underline{C}$ that has a null object (that is, an object which is both an initial object of $\underline{C}$ and a terminal object of $\underline{C}$) and is such that, for all objects C_1 and C_2 of $\underline{C}$, there exists a biproduct (in the sense of [16, Definition, page 194]) of C_1 and C_2 in $\underline{C}$; cf. also [13, page 126] and [9, page 60].) Fortunately, all relevant references agree (cf. [16, page 29], [20, page 7]) that if $\underline{D}$ and $\underline{E}$ are pre-additive categories and $F : \underline{D} \to \underline{E}$ is a (covariant) functor, then F is called an *additive functor* if, for all objects D_1 and D_2 of $\underline{D}$, the assignment $f \mapsto Ff$ $(:= F(f))$ determines an abelian group homomorphism $\underline{D}(D_1, D_2) \to \underline{E}(F(D_1), F(D_2))$ (that is, if, for all objects D_1 and D_2 of $\underline{D}$ and for all $\varphi, \psi \in \underline{D}(D_1, D_2)$, one has $F(\varphi + \psi) = F(\varphi) + F(\psi)$).

As mentioned above, our main interest here is in studying rings. Indeed, **from this point on, all rings are assumed commutative and unital; all algebras are assumed to be commutative and unital; and all algebra homomorphisms and all modules are also assumed to be unital**. For any (commutative unital) ring R, we let $\underline{C}_R$ denote the category of (commutative unital) R-algebras and (unital) R-algebra homomorphisms. In particular, $\underline{C}_{\mathbb{Z}}$ is the category of (commutative unital) rings and (unital) ring homomorphisms. For a ring R, we have already mentioned one connection with the above material, namely, the fact that ${}_R\text{Mod}$ is a pre-additive category. It is well known that ${}_R\text{Mod}(R, B)$ $(= \text{Hom}_R(R, B)) \cong B$ for all R-modules B. Consequently, ${}_R\text{Mod}(R, B)$ is a singleton set (namely, $\{0\}$) if and only if $B = 0$. Thus ${}_R\text{Mod}(R, 0) = 0$ for any ring R. However, ${}_R\text{Mod}(D, E) = 0$ for all (unital) R-modules D and $E \Leftrightarrow R = 0$ (for if the identity map ι on R is identically 0 and $r \in R$, then $r = \iota(r) = 0$; and if $R = 0$, any (unital) R-module is (isomorphic to) 0). In other words, every hom-set of ${}_R\text{Mod}$ is a singleton set if and only if R is a singleton set (if and only if $1 = 0$ in R; if and only if R is a *zero ring*). It seems natural to ask whether $\underline{C}_R$ exhibits the same kind of categorical behavior as ${}_R\text{Mod}$. It is true that a (commutative unital) ring R is a zero ring if and only if every hom-set of $\underline{C}_R$ is a singleton set. (To find a proof of this, consider what it would mean to be a polynomial ring over a zero ring – such musings will lead to the proof in Remark 2.3.) However, it is also true that, unlike ${}_R\text{Mod}$, $\underline{C}_R$ is rarely a pre-additive category. Indeed, as recorded in Corollary 2.4, $\underline{C}_R$ is a pre-additive category if and only if R is a zero ring. Also, as recorded in Corollary 2.7 (b) and Remark 2.11 (b), if R *is* a zero ring, then some general conclusions can be drawn concerning the additive functors from $\underline{C}_R$ to some pre-additive categories (such as $\underline{C}_R$ or Ab).

While developing Galois Theory for rings, Chase, Harrison and Rosenberg introduced a rather different meaning for the terminology "additive functor" in [5, page 30]. To avoid confusion, we will call that notion a "CHR-additive functor." If $\underline{C}$ and $\underline{D}$ are categories with finite products, then a functor $F : \underline{C} \to \underline{D}$ is called a *CHR-additive functor* if F preserves finite products; that is, if, for all finite lists $C_1, \ldots, C_n$ (possibly with repetition) of objects of $\underline{C}$, the canonical morphism

$$F(\prod_{i=1}^{n} C_i) \to \prod_{i=1}^{n} F(C_i),$$

induced by applying F to the projection maps $\prod_{i=1}^{n} C_i \to C_j$ (for $1 \leq j \leq n$), is an isomorphism in $\underline{D}$. It was noted without proof in [5] that for any ring R, the unit functor and the Picard group functor are each (in the above terminolgy) CHR-additive functors from $\underline{C}_R$ to Ab. The assertion concerning the unit functor is clear; for the sake of completeness, a proof of the assertion concerning the Picard group functor was given in [6, Theorem 1.27].

Frankly, [6] paid little, if any, attention to zero rings and empty products (and perhaps that can also be said of [5]). It is easy to see that the unit functor and the Picard group functor each send any zero ring to 0 (that is, to a/the abelian group with only one element). The rest of this paragraph and the next paragraph will go beyond those contexts and develop some facts that will be used in

Section 2. Observe that for any category $\underline{C}$, an empty product in $\underline{C}$ is the same as a terminal object of $\underline{C}$. Thus, the case $n = 0$ in the definition in the preceding paragraph implies that if $\underline{C}$ and $\underline{D}$ are categories with finite products, then any CHR-additive functor $F : \underline{C} \to \underline{D}$ sends some, hence each, terminal object of $\underline{C}$ to a terminal object of $\underline{D}$. With an eye to applications in situations where ($\underline{C}$ and $\underline{D}$ are each assumed to have finite products and) $\underline{C}$ is a full subcategory of $\underline{C}_{\mathbb{Z}}$ (note that in any such situation, any zero ring in $|\underline{C}|$ is an empty product and a terminal object of $\underline{C}$), we can conclude that any CHR-additive functor $F : \underline{C} \to \underline{D}$ sends each zero ring in $|\underline{C}|$ to a terminal object of $\underline{D}$. In particular, in any such situation where $\underline{D} = \mathrm{Ab}$, we can conclude that any CHR-additive functor sends any zero ring in $|\underline{C}|$ to 0.

It is natural to ask whether additive functors exhibit behavior that is somewhat like the behavior that was just noted for certain CHR-additive functors. A familiar argument (cf. the proof of Lemma 2.1 (a) below) shows that if $\underline{K}$ and $\underline{L}$ are pre-additive categories and $F : \underline{K} \to \underline{L}$ is an additive functor, then F sends any zero morphism f to a zero morphism (that is, if $K_1, K_2 \in |\underline{K}|$ and f is the neutral element in the abelian group $\underline{K}(K_1, K_2)$, then Ff is the zero morphism in $\underline{L}(F(K_1), F(K_2))$). Under these assumptions, it follows that if K_3 is a terminal object of $\underline{K}$, then $F(K_3)$ is a null object of $\underline{L}$. (Indeed, the preceding sentence implies that the identity map on K_3 (which is the only endomorphism of K_3 and, hence, is the neutral element in $\underline{K}(K_3, K_3)$) is sent to the neutral element in $\underline{L}(F(K_3), F(K_3))$ and, because F is a functor, it is also sent to the identity map on $F(K_3)$. Then [16, Proposition 1, page 194] can be applied to conclude that $F(K_3)$ is a null object of $\underline{L}$.) We have just shown that any additive functor sends any terminal object to a null object. When this fact is compared with what was shown in the preceding paragraph (specifically, that for categories with finite products, any CHR-additive functor sends any terminal object to a terminal object), it is natural to ask whether the concepts of “additive functor" and “CHR-additive functor" are equivalent in categorical settings where the context for the definition of each of these concepts is satisfied. For domain categories $\underline{C}$ of the form $\underline{C}_R$, that question will be answered in the affirmative in Corollary 2.7 (b) (i). The next paragraph will recall an apparently different categorical context for which [5] gave an affirmative answer. The final paragraph of the Introduction briefly summarizes the connections between the above material and the main results in this paper.

In [5], Chase, Harrison and Rosenberg noted one context where the notions of an additive functor and (what we have called) a CHR-additive functor agree, namely, for a functor $F : \underline{C} \to \underline{D}$ where both $\underline{C}$ and $\underline{D}$ are abelian categories. (Cf. also [9, Theorem 3.11].) Perhaps the most familiar examples of abelian categories are ${}_R\mathrm{Mod}$ (for any unital ring R) and the category of Ab-valued sheaves on X (for any topological space X). The term “abelian category" is due to Grothendieck [13, page 127], who defined an abelian category as an additive category in which every morphism has a kernel and a cokernel and which satisfies the so-called AB2 axiom (which can be paraphrased as requiring that the First Isomorphism Theorem holds in the category). An equivalent definition of abelian categories was introduced slightly earlier by Buchsbaum in the appendix of [4], where (what are now called) abelian categories were called “exact categories": see, especially, [4, pages 379-381]. We recommend [9] as an excellent introduction to abelian categories. For instance, [9] contains a lucid proof that the dual of any abelian category is an abelian category; this conclusion can also be found elsewhere, for example in [16, page 201].

One problem suggested by the title is to study the rings R and the functors $F : \underline{C}_R \to \mathrm{Ab}$ such that F is both additive (in the classical sense reviewed in the first paragraph) and also CHR-additive (in the sense defined four paragraphs ago). Unfortunately, the observation of Chase, Harrison and Rosenberg that was mentioned at the beginning of the preceding paragraph would seem to be of little help in such studies, since $\underline{C}_R$ is not an abelian category if R is a nonzero ring. (Indeed, since any abelian category is an additive category, it must have a null object. However, if R is a nonzero ring, then $\underline{C}_R$ does not have a null object, since the initial object R of $\underline{C}_R$ is not isomorphic to the terminal object 0 of $\underline{C}_R$.) Recall that the classical definition of an additive functor F requires the domain of F

to be a pre-additive category. We show in Corollary 2.4 that $\underline{C}_R$ can be given the structure of a pre-additive category (relative to some binary operation of "addition" in the hom-sets of $\underline{C}_R$) if and only if R is a zero ring. Our first main result, Corollary 2.7, shows, among other things, that if R is a zero ring, then a functor $\underline{C}_R \to \text{Ab}$ is additive if and only if it is CHR-additive. Despite the comparatively simple nature of $\underline{C}_R$ when R is a zero ring, Example 2.10 constructs, for any such ring R, two non-additive functors $\underline{C}_R \to \text{Ab}$. Our second main result, Theorem 2.18, examines the functors $F : \underline{C} \to \underline{D}$ in the most general relevant context, namely, where both $\underline{C}$ and $\underline{D}$ are pre-additive categories with finite products (and, hence, also with finite coproducts), that is, where both $\underline{C}$ and $\underline{D}$ are additive categories. The main contribution of Theorem 2.18 is the explication of five new characterizations of the additive functors $F : \underline{C} \to \underline{D}$ for any given additive categories $\underline{C}$ and $\underline{D}$.

2 Results

For the sake of completeness, we begin with a lemma that establishes some basic facts about pre-additive categories. Some other useful facts about pre-additive categories will be given in Proposition 2.5 and Lemma 2.6.

Following [16, page 33], we will use the following notation concerning dual categories. Let $\underline{C}$ be a category. Then the *dual category* of $\underline{C}$ is the category $\underline{C}^{\text{op}}$ which is defined as follows: $|\underline{C}^{\text{op}}| = |\underline{C}|$; for any $C_1, C_2 \in |\underline{C}^{\text{op}}|$, there is a bijection of hom-sets $\underline{C}(C_2, C_1) \to \underline{C}^{\text{op}}(C_1, C_2)$, denoted by $h \mapsto h^{\text{op}}$; and for all $C_1, C_2, C_3 \in |\underline{C}^{\text{op}}|$ and all $g^{\text{op}} \in \underline{C}^{\text{op}}(C_1, C_2)$ and $f^{\text{op}} \in \underline{C}^{\text{op}}(C_2, C_3)$, one defines $f^{\text{op}} g^{\text{op}} := (gf)^{\text{op}}$. It is harmless (and customary) to identify $(\underline{C}^{\text{op}})^{\text{op}}$ with $\underline{C}$ (so that a morphism $(h^{\text{op}})^{\text{op}}$ is identified with h).

Lemma 2.1. *Let $\underline{C}$ be a pre-additive category. Then:*

(a) Let $C_1, C_2, C_3 \in |\underline{C}|$, $f \in \underline{C}(C_1, C_2)$, and $g \in \underline{C}(C_2, C_3)$. Also let n_1, n_2 and n_3, respectively, denote the neutral elements in the abelian groups $\underline{C}(C_1, C_2)$, $\underline{C}(C_2, C_3)$ and $\underline{C}(C_1, C_3)$. Then

$$gn_1 = n_3 = n_2 f.$$

(b) $\underline{C}^{\text{op}}$ is a pre-additive category.

Proof. (a) Since $n_1 + n_1 = n_1$ and $\underline{C}$ is a pre-additive category, we have

$$gn_1 = g(n_1 + n_1) = gn_1 + gn_1.$$

By adding $-(gn_1)$ to the extreme members of the last displayed equations and then using group axioms to simplify the resulting expressions, we get

$$n_3 = -(gn_1) + gn_1 = -(gn_1) + (gn_1 + gn_1) = (-(gn_1) + gn_1) + gn_1 =$$

$n_3 + gn_1 = gn_1$. Similarly, $n_2 f = (n_2 + n_2)f = n_2 f + n_2 f$ leads to

$$n_3 = -(n_2 f) + n_2 f = -(n_2 f) + (n_2 f + n_2 f) = (-(n_2 f) + n_2 f) + n_2 f =$$

$n_3 + n_2 f = n_2 f$.

(b) If $C, D \in |\underline{C}^{\text{op}}|$, the canonical bijection $\underline{C}(D, C) \to \underline{C}^{\text{op}}(C, D)$ allows the abelian group structure on $\underline{C}(D, C)$ (which exists because $\underline{C}$ is assumed to be a pre-additive category) to be transferred to an abelian group structure on $\underline{C}^{\text{op}}(C, D)$. In detail: if $\lambda^{\text{op}}, \mu^{\text{op}} \in \underline{C}^{\text{op}}(C, D)$, then

$$\lambda^{\text{op}} + \mu^{\text{op}} := (\lambda + \mu)^{\text{op}}.$$

It only remains to show that for all objects C_1, C_2 and C_3 of $\underline{C}$, the following two "distributivity laws" hold:

$$f^{\rm op}(g^{\rm op}+h^{\rm op}) = f^{\rm op}g^{\rm op}+f^{\rm op}h^{\rm op}$$

$$\text{for all } g^{\rm op}, h^{\rm op} \in \underline{C}^{\rm op}(C_1, C_2) \text{ and all } f^{\rm op} \in \underline{C}^{\rm op}(C_2, C_3); \text{and}$$

$$(f^{\rm op}+g^{\rm op})h^{\rm op} = f^{\rm op}h^{\rm op}+g^{\rm op}h^{\rm op}$$

$$\text{for all } h^{\rm op} \in \underline{C}^{\rm op}(C_1, C_2) \text{ and all } f^{\rm op}, g^{\rm op} \in \underline{C}^{\rm op}(C_2, C_3).$$

We will prove the first of these "laws" and leave to the reader the (similar) proof of the second "law". An interesting feature of the proof will be that the first (resp., second) of the distributivity laws in $\underline{C}^{\rm op}$ will follow from the second (resp., first) of the distributivity laws in $\underline{C}$.

Given $g^{\rm op}, h^{\rm op} \in \underline{C}^{\rm op}(C_1, C_2)$ and $f^{\rm op} \in \underline{C}^{\rm op}(C_2, C_3)$, we have

$$f^{\rm op}(g^{\rm op}+h^{\rm op}) = f^{\rm op}(g+h)^{\rm op} = ((g+h)f)^{\rm op}.$$

As the morphisms in $\underline{C}$ satisfy both distributivity laws (because $\underline{C}$ is a pre-additive category), the right-most member in the last display can be expressed as

$$(gf+hf)^{\rm op} = (gf)^{\rm op}+(hf)^{\rm op} = f^{\rm op}g^{\rm op}+f^{\rm op}h^{\rm op}.$$

The proof is complete. □

In any pre-additive category $\underline{C}$, it is customary to let 0 denote the neutral element in any hom-set $\underline{C}(C_1, C_2)$. One can do so because each such hom-set is an abelian group under some "addition" operation +. Using the "0" notation in this way allows us to restate the conclusion of Lemma 2.1 (a) as $g0 = 0 = 0f$ (for $\underline{C}$, f and g as supposed above). However, this sort of use of the notation "0" can be ambiguous if it is not clear which objects of $\underline{C}$ are intended to be the domain and codomain, respectively, of "0." For that reason, the above statement of Lemma 2.1 (a) used the symbols n_1, n_2 and n_3 (instead of the generic symbol 0), out of an abundance of caution. In most situations, no harm is likely in using the notation "0" in a pre-additive category. The proof of Proposition 2.2 will use the fact that $0f = 0$, but it will not need to explicitly use the fact that $g0 = 0$.

The proof of Lemma 2.1 (a) was necessarily quite fussy. Half of that fussiness could have been avoided if we had proved part (b) of Lemma 2.1 before proving part (a). (We did not choose that reorganization because our usual pedagogic/expository preference is to begin with the easier proofs.) Let us give the details of how one could use Lemma 2.1 (b) and the "$g0 = 0$" conclusion from Lemma 2.1 (a) to prove the "$0f = 0$" conclusion in Lemma 2.1 (a). In detail,

$$0f = (0^{\rm op})^{\rm op}(f^{\rm op})^{\rm op} = (f^{\rm op}0^{\rm op})^{\rm op} = (0^{\rm op})^{\rm op} = 0.$$

The preceding details give a nice example of using dual categories to avoid excessive fussiness. In particular, observe that the third equality in the preceding display used the fact that $0^{\rm op}$ is the neutral element in the appropriate hom-set of $\underline{C}^{\rm op}$ (and applied the first equation in Lemma 2.1 (a) to the pre-additive category $\underline{C}^{\rm op}$ and its morphism $g := f^{\rm op}$).

The proof of Proposition 2.2 will also use [16, Proposition 1, page 194], which is the first result in the section on pre-additive categories in [16]. The proof of that result in [16] is short, slick, very clever and, in our opinion, somewhat incomplete in two ways, the first of which is very minor and the second of which is more noteworthy. First, the proof of [16, Proposition 1, page 194] uses the fact that $0f = 0$. We agree that this fact is available at that point in [16], because our statement of Lemma 2.1 (a) can be gleaned from the first half of the (in our opinion, very terse) final sentence preceding the statement of [16, Proposition 1, page 194]. Indeed, that half of that sentence can be interpreted, using the above terminology, as stating the following:"Again, a composite with the neutral element of a hom-set in a pre-additive category is necessarily the neutral element of the appropriate hom-set in that category". One can find what was apparently intended to serve as a proof of the assertion in the first half of the final sentence preceding the statement of [16, Proposition 1, page 194] by reading the following second half of the final sentence preceding the statement of [16, Proposition 1, page 194]: ", since composition is distributive over addition." We would agree that the preceding quotation does give the second most important step in the proof that $0f = 0$, but it has omitted the most important step (which would begin the argument by observing that $0f = (0+0)f$) and it has also omitted the last step of the proof (which is to use the abelian group structure of the hom-sets by adding $-(0f)$ to both sides of the equation $0f = 0f + 0f$ and then simplifying by using the group axioms satisfied by that hom-set). I suggest that a clearer wording for the second half of the final sentence preceding the statement of [16, Proposition 1, page 194] would have been the following: ": mimic the usual proof that $r0 = 0 = 0r$ for any element r of a ring R". Thus, my first complaint about the proof of [16, Proposition 1, page 194] is only a critique of the terse manner in which it justified the step asserting that $0f = 0$. The second "somewhat incomplete" aspect of that proof is more serious. The statement of [16, Proposition 1, page 194] is that four conditions, (i)-(iv), in a pre-additive category A are equivalent and, hence, that "In particular, any initial (or terminal) object in A is a null object." This "In particular" assertion justifies the main step in the proof given below of Proposition 2.2. Also, one can see at once that this "In particular" assertion is clear *if* one has truly proven that (i)-(iv) are equivalent (for the ambient pre-additive category). However, an objective report on the complete published three-sentence proof of [16, Proposition 1, page 194] is the following: its first sentence implicitly uses $0f = 0$ to show that (i) $\Rightarrow$ (iii) $\Rightarrow$ (iv), its second sentence explicitly uses $0f = 0$ to prove that (iii) $\Rightarrow$ (ii), and its third sentence states that "The rest follows by duality." I agree that what Mac Lane has called "the rest" *would* follow by duality *if* one knew that the dual of the ambient pre-additive category is itself a pre-additive category. Unfortunately, I cannot find anything in [16] prior to its page 194 that would suggest that one should (or does) know that the class of pre-additive categories is stable under the formation of dual categories. That deficiency in the exposition in [16, page 194] is why Lemma 2.1 (b) was given above. With both parts of Lemma 2.1 in hand, one can now use the preceding comments to give what I would consider to be a complete proof of [16, Proposition 1, page 194]. With that in hand, the proof of Proposition 2.2 that is given below will be seen as also being complete. However, I am frankly concerned that such a category-laden approach to proving Proposition 2.2 may deter some inexperienced readers. So, let me mention here that another proof of Proposition 2.2 will be given in Remark 2.3 and this alternate proof will use only the fact that $0f = 0$ from Lemma 2.1 (a) and the universal mapping property of a polynomial ring over a nonzero commutative ring. I would like to end this long paragraph with a three-part *apologia* of sorts for its existence. This paragraph has allowed me to rectify what I have long considered to be one of the exceedingly rare blemishes in the writings of Saunders Mac Lane (I know of only one other serious blemish in his writings) – Mac Lane was an outstanding creative mathematician and expositor; this paragraph has allowed me a forum to publish something that I discovered in April 1967 during my first week of doctoral research (my doctoral advisor, who was one of the authors of [5], advised me that the contents of what are here called Remark 2.3 and Corollary 2.4 should not appear in my eventual doctoral thesis [6], and I was frankly too intimidated to request an explanation from him or to otherwise pursue the matter further at that time); and this paragraph has given me the opportunity to alert any commutative ring-theorists who would prefer to read as little category theory as possible that one can proceed to the proofs in Remark 2.3 and Corollary 2.4 at once after reading the proof of Lemma 2.1 (a).

Recall that a *zero ring* is a ring with a unique element (equivalently, a singleton set R, with the unique function $R \times R \to R$ necessarily taken as both the "addition" operation on R and the "multiplication" operation on R, and with the unique element of R necessarily playing both the additive role of 0 in R and the multiplicative role of 1 in R). It is easy to see (and we will need to use this triviality later) that if R and S are rings and R is a zero ring, then S is a zero ring if and only if $R \cong S$ (as rings).

We can now give a necessary condition for $\underline{C}_R$ to be a pre-additive category.

Proposition 2.2. *Let R be a ring such that $\underline{C}_R$ is a pre-additive category. Then R is a zero ring.*

Proof. There is a unique way to view R as a (commutative unital) R-algebra (namely, via $s \cdot r := sr$ for all $s, r \in R$). Observe that R is then an initial object of $\underline{C}_R$. Choose T to be any zero ring, and let $t \in T$ denote the unique element of T. There is exactly one way to view T as a (commutative unital) R-algebra (namely, via $r \cdot t := t$ for each $r \in R$). Observe that T is then a terminal object of $\underline{C}_R$. Since $\underline{C}_R$ is assumed to be a pre-additive category, it follows from [16, Proposition 1, page 194] that every initial object of $\underline{C}_R$ is a terminal object of $\underline{C}_R$. Hence, both R and T are terminal objects of $\underline{C}_R$. But any two terminal objects of a category are isomorphic in that category. Consequently, R and T are isomorphic in $\underline{C}_R$. Thus, by equating cardinalities, we get $|R| = |T| = 1$. Therefore, R is a zero ring, as asserted. □

The assertion in the preceding proof that every initial object of $\underline{C}_R$ is a terminal object of $\underline{C}_R$ (assuming that $\underline{C}_R$ is a pre-additive category) follows from the implication (i) ⇒ (ii) in [16, Proposition 1, page 194]. That implication was proved in the first and second sentences of the proof of [16, Proposition 1, page 194]. Thus, all that the above proof of Proposition 2.2 needed from [16, Proposition 1, page 194] was the first and second sentences of the latter's proof. If the proof of Proposition 2.2 had, instead, used the fact that any two *initial* objects of a category are isomorphic, we would have needed to know that any terminal object of a pre-additive category is an initial object of that category; that, in turn, would have required the proof of Proposition 2.2 to use/explicate the third sentence of the proof of [16, Proposition 1, page 194] (namely, the above-mentioned sentence, "The rest follows by duality."); that, in turn, would have required us to develop Lemma 2.1 (b). By the way, the above proof of Proposition 2.2 also needed (because of its role in proving that (i) ⇒ (ii)) the part of the second sentence of the proof of [16, Proposition 1, page 194] which asserted that $0f = 0$, which is half of the content of Lemma 2.1. In summary, a justifiable appeal to [16, Proposition 1, page 194] in the proof of Proposition 2.2 required us to develop half of Lemma 2.1 (a), while contemplation of the proof that was published for [16, Proposition 1, page 194] led us to develop Lemma 2.1 (b) and the "other" half of Lemma 2.1 (a). We hope that much of the rest of this paper will convince the reader that there is merit in our emphasis here on pre-additive categories, as that emphasis will lead to new technical information about pre-additive categories (in Proposition 2.5 and Lemma 2.6) and, ultimately, to the solution of the paper's motivating question in Corollary 2.7 (b) (i), along with several categorical characterizations of the category of zero rings in Corollary 2.7 (a), as well as more substantial categorical generalizations in Theorem 2.18.

We next give an alternate, less categorical proof of Proposition 2.2. The results and arguments given in Remark 2.3 and Corollary 2.4 were found by the author in April 1967.

Remark 2.3. While readers who are comfortable with the basics of category theory may find the above proof of Proposition 2.2 to be terse (and perhaps self-contained and elegant), we believe that many readers will find the following alternate proof of Proposition 2.2 to be more direct and accessible than the proof which was given above. We will give an indirect argument (that is, a "proof by contradiction"). So, we assume that $\underline{C}_R$ is a pre-additive category and that the ring R is not a zero ring, and our task is to find a contradiction. Let X be an indeterminate over the ring R. By the universal mapping property of a polynomial ring over a nonzero commutative ring (cf. [14, Theorem 5.5, page 152]), the assignment $\varphi \mapsto \varphi(X)$ determines a bijection $\underline{C}_R(R[X], R) \to R$. The inverse of this bijection sends any $r \in R$ to the R-algebra homomorphism $\psi_r : R[X] \to R$ such that $\psi_r(X) = r$. Since $\underline{C}_R$ is a pre-additive category, the hom-set $\underline{C}_R(R[X], R)$ is an abelian group (under some "addition" operation which we need not specify). The neutral element in this abelian group (which we hesitate to denote by the overworked symbol "0") is, by the preceding observation, of the form ψ_n for some uniquely detemined element $n \in R$. (This choice of notation is motivated by the condition that ψ_n is a *n*eutral element.) By the second assertion in Lemma 2.1 (a), we get that if A is any (commutative unital) R-algebra and $\rho : A \to R[X]$ is any (unital) R-algebra homomorphism, then $\psi_n\rho$ is the neutral element of the abelian group $\underline{C}_R(A,R)$. In particular, if $A = R[X]$ and ρ is taken to be the R-algebra endomorphism of $R[X]$ determined by $X \mapsto X+1$, then $\psi_n\rho$ is the neutral element of the abelian group $\underline{C}_R(R[X],R)$. In other words, $\psi_n\rho = \psi_n$. Applying these equal functions to $X \in R[X]$, we get that $\psi_n(X)+1 = \psi_n(X)+\psi_n(1) =$

$$\psi_n(X+1) = \psi_n(\rho(X)) = (\psi_n\rho)(X) = \psi_n(X) = \psi_n(X)+0,$$

whence $1 = 0$ in R, whence each $r \in R$ satisfies $r = r \cdot 1 = r \cdot 0 = 0$, whence R is a zero ring, the desired contradiction. This completes the proof. This completes the remark.

It is natural to ask if the converse of Proposition 2.2 is valid. The next result answers this question.

Corollary 2.4. *Let R be a (commutative unital) ring. Then the following conditions are equivalent:*
(1) *$\underline{C}_R$ is a pre-additive category;*
(2) *R is a zero ring.*

Proof. (1) $\Rightarrow$ (2): Apply Proposition 2.2 (or Remark 2.3).

(2) $\Rightarrow$ (1): Assume that R is a zero ring. It follows easily that each (unital) R-algebra (that is, each object of $\underline{C}_R$) is also a zero ring. Hence, it also follows easily that for any objects S and T of $\underline{C}_R$, the unique function $f_{S,T} : S \to T$ is an R-algebra homomorphism, and so the hom-set $\underline{C}_R(S,T)$ is the singleton set $\{f_{S,T}\}$. Of course, this singleton set can be given the structure of an additive abelian group in a unique way (by defining $f_{S,T} + f_{S,T}$ to be $f_{S,T}$). It remains only to prove that, with addition having been explicated (in fact, forced) in all the hom-sets of $\underline{C}_R$, both distributivity laws hold in $\underline{C}_R$. We will prove the first of those laws, leaving to the reader the (similar) proof of the second distributivity law.

Suppose, then, that S, T and U are objects of $\underline{C}_R$, with $g,h \in \underline{C}_R(S,T)$ and $f \in \underline{C}_R(T,U)$. It remains only to prove that $f(g+h) = fg+fh$. This, in turn, is evident, since $f(g+h)$ and $fg+fh$ are each elements of the singleton set $\underline{C}_R(S,U)$. The proof is complete. □

We would caution the reader not to rework the above proof of Corollary 2.4 by using copious occurrences of the symbol "0". A more careful approach (for example, using notation such as the above "$f_{S,T}$") will yield clear benefits in Corollary 2.7 where, among other things, we will answer this paper's motivating question. We would also caution the reader, in case R is a zero ring, not to view $\underline{C}_R$ as having a unique object. While it is true (cf. [2, Chapter II, 1.2]) that a strong version of the Axiom of Choice (specifically, that the universe can be well-ordered) implies that every category is equivalent to a "skeletal" category (that is, to a category in which any two isomorphic objects are equal) and it has been known for more than 80 years (cf. [12]) that such a strong Axiom of Choice is consistent with ZFC, we recommend that one should not decide to make such an additional foundational assumption simply because of a desire to simplify some notation.

We next give two categorical results. Proposition 2.5 shows one way in which CHR-additive functors and additive functors behave similarly in relevant contexts, where each of these properties of functors is shown to be stable under natural equivalence. While the proof of Proposition 2.5 will seem routine for readers who are comfortable with category theory, we will provide full details for that proof, in order to enhance accessibility. That result can be seen as motivation for some of Example 2.10. The path to our first main result, Corollary 2.7, will be eased by Lemma 2.6, which collects/states some facts that were proved in the Introduction.

Proposition 2.5. (*a*) *Let* $\underline{C}$ *and* $\underline{D}$ *be pre-additive categories, let* $F:\underline{C}\to\underline{D}$ *be an additive functor, and let* $G:\underline{C}\to\underline{D}$ *be a functor such that F and G are naturally equivalent. Then G is an additive functor.*

(*b*) *Let* $\underline{C}$ *and* $\underline{D}$ *each be categories with finite products, let* $F:\underline{C}\to\underline{D}$ *be a CHR-additive functor, and let* $G:\underline{C}\to\underline{D}$ *be a functor such that F and G are naturally equivalent. Then G is a CHR-additive functor.*

Proof. (a) By hypothesis, we can pick a natural equivalence $\eta: F\to G$. Thus for each object C of $\underline{C}$, one has a "natural" isomorphism $\eta_C: F(C)\to G(C)$ in $\underline{D}$. Our task is to prove that if $f,g\in\underline{C}(C_1,C_2)$ (for some objects C_1 and C_2 of $\underline{C}$), then $G(f+g)=G(f)+G(g)$. Of course, $F(f+g)=F(f)+F(g)$, since F is assumed to be an additive functor. Moreover, when the "naturality" of the above-mentioned isomorphisms of the form η_C is applied to the morphisms f, g and $f+g$, we get

$$G(f)\eta_{C_1}=\eta_{C_2}F(f),\, G(g)\eta_{C_1}=\eta_{C_2}F(g) \text{ and } G(f+g)\eta_{C_1}=\eta_{C_2}F(f+g).$$

Therefore, since composition distributes over addition of morphisms in (the pre-additive category) $\underline{D}$, we get

$$G(f)+G(g)=\eta_{C_2}F(f)(\eta_{C_1})^{-1}+\eta_{C_2}F(g)(\eta_{C_1})^{-1}=$$

$\eta_{C_2}(F(f)+F(g))(\eta_{C_1})^{-1}$. This simplifies to $\eta_{C_2}F(f+g)(\eta_{C_1})^{-1}=G(f+g)$, as desired.

(b) Let us first deal with the case of empty products. In that regard, it will suffice to show that if T is a terminal object of $\underline{C}$ such that $F(T)$ is a terminal object of $\underline{D}$, then $G(T)$ is also a terminal object of $\underline{D}$. This can be shown by using basic category theory (without the hypothesis that F is CHR-additive and without the hypothesis that the object T is terminal in $\underline{C}$), as follows. Our task is to show that if $D\in|\underline{D}|$, then $\underline{D}(D,G(T))$ is a singleton set. By hypothesis, $\underline{D}(D,F(T))$ is a singleton set. Let φ denote its unique element. Pick a natural equivalence $\eta: F\to G$. Then $\psi:=\eta_T\varphi\in\underline{D}(D,G(T))$, where as usual, η_T denotes the "natural" isomorphism $F(T)\to G(T)$ given by η. It remains only to prove that if $\psi^*\in\underline{D}(D,G(T))$, then $\psi^*=\psi$ (that is, $\psi^*=\eta_T\varphi$). This, in turn, holds since $(\eta_T)^{-1}\psi^*=\varphi$, the point being that $(\eta_T)^{-1}\psi^*\in\underline{D}(D,F(T))=\{\varphi\}$.

It remains to consider nonempty finite products. Let $C_1,C_2,\ldots,C_n$ be a finite list (possibly with repetition) of elements in $|\underline{C}|$, for some integer $n\geq 2$. Fix a (direct) product $P=\prod_{i=1}^n C_i$ in $\underline{C}$; also fix products $\prod_i F(C_i)$ and $\prod_i G(C_i)$ in $|\underline{D}|$. The structures of these products include projection maps $p_j: P\to C_j$, $\pi_j:\prod_i F(C_i)\to F(C_j)$ and $\rho_j:\prod_i G(C_i)\to G(C_j)$, for $j=1,\ldots,n$. The universal mapping property of products gives uniquely determined morphisms

$$\alpha: F(P)\to\prod_{i=1}^n F(C_i) \text{ and } \beta: G(P)\to\prod_{i=1}^n G(C_i)$$

such that $\pi_j\alpha=F(p_j)$ and $\rho_j\beta=G(p_j)$ for $j=1,\ldots,n$. By hypothesis, α is an isomorphism. Our task is to show that β is an isomorphism.

Pick a natural equivalence $\eta: F\to G$. Recall that η_C is an isomorphism for each $C\in|\underline{C}|$. The universal mapping property of products gives uniquely determined morphisms

$$\gamma:\prod_{i=1}^n F(C_i)\to\prod_{i=1}^n G(C_i) \text{ and } \delta:\prod_{i=1}^n G(C_i)\to\prod_{i=1}^n F(C_i)$$

such that $\rho_j\gamma=\eta_{C_j}\pi_j$ and $\pi_j\delta=(\eta_{C_j})^{-1}\rho_j$ for $j=1,\ldots,n$.

We next make the following two claims: the composite morphisms $\gamma\delta$ and $\delta\gamma$ are each identity maps. As the proofs of these claims are similar, we will prove the claim about $\gamma\delta$ and leave the claim about $\delta\gamma$ to the reader. By the "uniqueness" aspect of the universal mapping property of products, the claim will follow if we show that $\rho_j(\gamma\delta)=\rho_j$ for $j=1,\ldots,n$. For each j, we have

$$\rho_j(\gamma\delta)=(\rho_j\gamma)\delta=(\eta_{C_j}\pi_j)\delta=\eta_{C_j}(\pi_j\delta)=\eta_{C_j}((\eta_{C_j})^{-1}\rho_j)=\rho_j,$$

thus proving the above claim(s).

It follows (now that we have proved the above claims) that γ is an isomorphism, with $\gamma^{-1} = \delta$. Hence, being a composite of isomorphisms, $\gamma\alpha(\eta_P)^{-1}$ is an isomorphism. Therefore, to complete the proof, it will suffice to show that $\beta = \gamma\alpha(\eta_P)^{-1}$. By the universal mapping property of $\prod_i G(C_i)$, an equivalent task is to show that

$$\rho_j\beta = \rho_j(\gamma\alpha(\eta_P)^{-1}) \text{ if } j = 1, \ldots, n.$$

Fix j. Note via the naturality of η that $G(p_j)\eta_P = \eta_{C_j}F(p_j)$, and so $G(p_j) = (\eta_{C_j}F(p_j))(\eta_P)^{-1}$. Hence,

$$\rho_j\beta = G(p_j) = \eta_{C_j}F(p_j)(\eta_P)^{-1} = \eta_{C_j}(\pi_j\alpha)(\eta_P)^{-1} = (\eta_{C_j}\pi_j)\alpha(\eta_P)^{-1} =$$

$$(\rho_j\gamma)\alpha(\eta_P)^{-1} = \rho_j(\gamma\alpha(\eta_P)^{-1}),$$

as desired. The proof is complete. □

Lemma 2.6. *(a) Let $\underline{C}$ and $\underline{D}$ each be categories with finite products, let $F : \underline{C} \to \underline{D}$ be a CHR-additive functor, and let C be a terminal object of $\underline{C}$. Then $F(C)$ is a terminal object of $\underline{D}$.*

(b) Let $\underline{C}$ and $\underline{D}$ be pre-additive categories and let $F : \underline{C} \to \underline{D}$ be an additive functor. Then F sends any zero morphism to a zero morphism (that is, if $C_1, C_2 \in |\underline{C}|$ and f is the neutral element of the abelian group $\underline{C}(C_1, C_2)$, then Ff is the zero morphism in $\underline{D}(F(C_1), F(C_2))$). Moreover, it follows that, under these assumptions, F sends any terminal object of $\underline{C}$ to a null object of $\underline{D}$.

Proof. (a) This assertion was proved in the fourth paragraph of the Introduction.

(b) This assertion was proved in the fifth paragraph of the Introduction. □

It will be convenient to let $\underline{\mathcal{Z}}$ denote the category of zero rings, that is, the full subcategory of $\underline{C}_{\mathbb{Z}}$ whose class of objects is the collection of zero rings.

We next present our first main result.

Corollary 2.7. *Let R be a (commutative unital) ring. Let $\underline{\mathcal{Z}}$ denote the category of zero rings. Then:*

(a) The following conditions are equivalent:

(1) R is a zero ring;

(2) $\underline{C}_R = \underline{\mathcal{Z}}$;

(3) $\underline{C}_R$ is a pre-additive category;

(4) $\underline{C}_R$ is an additive category;

(5) $\underline{C}_R$ is an abelian category;

(6) $\underline{C}_R = {}_R$Mod.

(b) Assume, moreover, that R is a zero ring. Then the following assertions, (i)-(iii), are valid:

(i) Let $\underline{D}$ be a pre-additive category with finite products. (For instance, $\underline{D}$ could be $\underline{C}_R$ or Ab.) Let $F : \underline{C}_R \to \underline{D}$ be a functor. Then F is an additive functor if and only if F is a CHR-additive functor.

(ii) Every hom-set $\underline{C}_R(S,T)$ in $\underline{C}_R$ is a singleton set, and every morphism in $\underline{C}_R$ is an isomorphism.

(iii) Let $\underline{D}$ be a pre-additive category (resp., let $\underline{D}$ be a category with finite products). Let $F : \underline{C}_R \to \underline{D}$ be an additive functor (resp., a CHR-additive functor). Let $S, T \in |\underline{C}_R|$ and let $f_{S,T}$ denote the unique morphism $S \to T$ in $\underline{C}_R$. Then $F(S)$ and $F(T)$ are each null objects (resp., terminal objects) of $\underline{D}$. Moreover $F(S) \cong F(T)$ in $\underline{D}$, and $F(f_{S,T}) : F(S) \to F(T)$ is an isomorphism in $\underline{D}$, with inverse $F(f_{T,S})$.

Proof. (a) (1) ⇔ (3): Apply Corollary 2.4.

(1) ⇒ (2): Suppose that R is a zero ring. If $S \in |\underline{C}_R|$ and $s \in S$, then $s = 1 \cdot s = 0 \cdot s = 0$, so S is a zero ring, that is, $S \in |\underline{\mathcal{Z}}|$. It follows easily that if S and T are objects of $|\underline{C}_R|$, then the unique function $S \to T$ is an R-algebra homomorphism (and hence a ring homomorphism). Thus, every object (resp., morphism) of $\underline{C}_R$ is an object (resp., morphism) of $\underline{\mathcal{Z}}$. Now, consider any object V of $\underline{\mathcal{Z}}$. Let 0_V

(resp., 0_R) denote the (unique) element of V (resp., of R). There is exactly one way to endow the zero ring V with the structure of an R-algebra, namely, via $0_R \cdot 0_V := 0_V$. Moreover, if zero rings V and $W = \{0_W\}$ are thus endowed with R-algebra structures, it is easy to see that the unique function $h : V \to W$ (that is, the unique ring homomorphism $V \to W$) is an R-algebra homomorphism. (In detail: $0_R \cdot h(0_V) = 0_R \cdot 0_W = 0_W = h(0_V) = h(0_R \cdot 0_V)$.) Thus, every object (resp., morphism) of $\underline{\mathcal{Z}}$ is an object (resp., morphism) of $\underline{C}_R$. This concludes the proof of (2).

(2) ⇒ (1): Suppose that $\underline{C}_R = \underline{\mathcal{Z}}$. Since R is a (unital) R-algebra (and hence R is an object of $\underline{C}_R$), it follows that R is an object of $\underline{\mathcal{Z}}$ (and hence R is a zero ring).

(6) ⇒ (5): It suffices to use the fact that for all (unital but not necessarily commutative) rings Λ, the category ${}_{\Lambda}$Mod of (left) R-modules (and R-module homomorphisms) is an abelian category.

(5) ⇒ (4) ⇒ (3): These implications hold for arbitrary categories.

(1) ⇒ (6): Of course, any object (resp., morphism) in $\underline{C}_R$ is an object (resp., morphism) in ${}_R$Mod. Assume (1), with 0_R denoting the unique element of R. Let $M \in |{}_R\text{Mod}|$. Let 0_M denote the additive identity element of M. Since M is an (unital) R-module and 0_R is the multiplicative identity element of R, each $m \in M$ satisfies $m = 0_R \cdot m = 0_M \in M$, and so $M = \{0_M\}$. Moreover, by defining "multiplication on M" to be the unique binary operation on M, one checks easily that M is a zero ring. Thus, since (2) ⇒ (1), $M \in |\underline{C}_R|$. It remains only to show that if $M, N \in |_R \text{ Mod}|$ and $f_{M,N} : M = \{0_M\} \to N = \{0_N\}$ is an R-module homomorphism (that is, if $f_{M,N}$ is the unique function $M \to N$), then $f_{M,N}$ is an R-algebra homomorphism. This, in turn, follows since

$$f_{M,N}(0_M \cdot 0_M) = f_{M,N}(0_M) = 0_N = 0_N \cdot 0_N = f_{M,N}(0_M) \cdot f_{M,N}(0_M).$$

Although a proof of (a) is complete at this point, we next provide an alternate direct proof that (1) ⇒ (5), in order to have a self-contained proof of the equivalence of conditions (1)-(5) in (a) that would avoid any mention of condition (6).

(1) ⇒ (5): Assume that R is a zero ring. As (1) ⇔ (3), we already know that $\underline{C}_R$ is a pre-additive category. We will prove that $\underline{C}_R$ is an abelian category. This can be done by verifying that $\underline{C}_R$ satisfies the conditions in the characterization of abelian categories in [13] that was mentioned in the Introduction. We will leave the details of that kind of verification to the reader. Instead, we will next sketch how to verify that $\underline{C}_R$ satisfies the conditions in the (possibly more accessible) five-part characterization of abelian categories that can be found in [16, Definition, page 198].

• $\underline{C}_R$ is a pre-additive category: This was observed above (since R is a zero ring).
• $\underline{C}_R$ has a null object: This holds since R being a zero ring ensures that R is a null object of $\underline{C}_R$.
• $\underline{C}_R$ has binary biproducts: This follows from [16, Theorem 2, page 194], since $\underline{C}_R$ is a pre-additive category and $S \times S \cong S$ for each object S of $\underline{C}_R$ (the latter fact being an easy consequence of the fact that $\underline{C}_R(S,T)$ is a singleton set for all objects S and T of $\underline{C}_R$).
• Each morphism in $\underline{C}_R$ has a kernel and a cokernel: This holds by the following reasoning. Since $\underline{C}_R(S,T)$ is a singleton set for all objects S and T of $\underline{C}_R$, it follows from the discussion of equalizers (resp., coequalizers) on page 70 (resp., page 64) of [16] that for each/the morphism $f : S \to T$ in $\underline{C}_R$, the identity map on S (resp., the identity map on T) is an equalizer (resp., a coequalizer) of the pair consisting of f and f, and thus that identity map is a kernel (resp., a cokernel) of f.
• Each monomorphism in $\underline{C}_R$ is a kernel and each epimorphism in $\underline{C}_R$ is a cokernel: This can be shown to hold by using the facts (including the references) in the proof of the preceding bulleted item. Indeed, one can thus show that each morphism $f : S \to T$ in $\underline{C}_R$ is both a kernel and a cokernel of the unique morphism $T \to S$. This completes the proof of (a).

(b) (i) It suffices to combine (a) with both parts of Lemma 2.6. For the sake of completeness, we next provide the details.

Suppose first that F is a CHR-additive functor. Since each object of $\underline{C}_R$ is a terminal object, Lemma 2.6 (a) ensures that F sends each object of $\underline{C}_R$ to a terminal object of $\underline{D}$. Let $S, T \in |\underline{C}_R|$. Since $F(T)$ is a terminal object, $\underline{D}(F(S), F(T))$ is a singleton set. On the other hand, $\underline{C}_R(S,T)$ is also a singleton set, with unique element, say, f. To show that F is an additive functor, it suffices to prove that $F(f+f) = F(f)+F(f)$. Necessarily, $f+f = f$. Thus, $F(f+f) = F(f)$. Moreover, since $\underline{D}(F(S),F(T))$ is a singleton set, $F(f)+F(f) = F(f)$. Hence, $F(f+f) = F(f)+F(f)$, as desired.

Conversely, suppose that F is an additive functor. Since each object of $\underline{C}_R$ is a terminal object, Lemma 2.6 (b) ensures that F sends each object of $\underline{C}_R$ to a terminal object of $\underline{D}$. Moreover, if $C_1,\ldots,C_n$ is a nonempty finite list (possibly with repetition) of objects of $\underline{C}_R$, then $\mathcal{T} := \prod_{i=1}^{n} F(C_i)$ is a product of finitely many terminal objects of $\underline{D}$, and so $\mathcal{T}$ is a terminal object of $\underline{D}$. As $F(\prod_{i=1}^{n} C_i)$ is also a terminal object of $\underline{D}$, there exists an isomorphism $h : F(\prod_{i=1}^{n} C_i) \to \mathcal{T}$ (in $\underline{D}$). Any such h must be the unique element of $\underline{D}(F(\prod_{i=1}^{n} C_i),\mathcal{T})$. Thus, the canonical morphism $F(\prod_{i=1}^{n} C_i) \to \mathcal{T}$ must be h and so is an isomorphism, whence F is a CHR-additive functor.

To accommodate readers who may have preferred the second approach to the proof of (b) (i) (which avoided using condition (6) in (a)), we will next give proofs of parts (ii) and (iii) of (b) that will avoid explicit mention of that condition (6). Readers who preferred the first approach to the proof of (b) (i) (which used condition (6) in (a)) are advised that some of the details in the following self-contained proofs of (ii) and (iii) necessarily repeat some observations from the above proof that (1) $\Rightarrow$ (6) in (a).

(ii) Let $S,T \in |\underline{C}_R|$. By the implication (1) $\Rightarrow$ (2) in (a), $\underline{C}_R = \underline{\mathcal{Z}}$, and so both S and T are zero rings. It will be convenient to let 0_S (resp., 0_T) denote the unique element of S (resp., T). It is easy to check that the unique function $f_{S,T} : S \to T$ (sending 0_S to 0_T) is an R-algebra homomorphism, and so $\underline{C}_R(S,T) = \{f_{S,T}\}$, which is a singleton set. Necessarily, the composite functions $f_{S,T}f_{T,S}$ and $f_{T,S}f_{S,T}$ are identity maps (on T and S, respectively), and so $f_{S,T}$ (the typical morphism in $\underline{C}_R$) is an isomorphism in $\underline{C}_R$ (with $f_{T,S}$ serving as its inverse).

(iii) The nature of $f_{S,T}$ was exposed in the proof of (ii). To prove the assertion that $F(S)$ and $F(T)$ are each null objects (resp., terminal objects) of $\underline{D}$, use the first assertion in (ii) to conclude that S and T are each terminal objects of $\underline{C}_R$ and then apply part (b) (resp, part (a)) of Lemma 2.6. The final assertions can be proven by combining the following facts: the proof of (ii) showed that $f_{S,T}$ is an isomorphism with inverse $f_{T,S}$, and functors preserve isomorphisms and their inverses. Note that an alternate proof that $F(S) \cong F(T)$ is available, since any two null (resp., terminal) objects of a category are isomorphic. □

We can now give a companion for Proposition 2.5.

Corollary 2.8. *Let R be a zero ring and let $\underline{D}$ a pre-additive category with finite products. Let F and G be additive functors (equivalently, CHR-additive functors) $\underline{C}_R \to \underline{D}$. Then F and G are naturally equivalent.*

Proof. The parenthetical equivalence follows from Corollary 2.7 (b) (i). By using either part of Lemma 2.6, we see that for all objects T of $\underline{C}_R$, the functors F and G each send T to (a possibly different) terminal object of $\underline{D}$, whence there is a (unique) isomorphism $\eta_T : F(T) \to G(T)$ in $\underline{D}$. Now, consider any objects T_1, T_2 of $\underline{C}_R$ and the (unique) morphism $f \in \underline{C}_R(T_1,T_2)$. Since $G(T_2)$ is a terminal object of $\underline{D}$, we have $\eta_{T_2}F(f) = G(f)\eta_{T_1}$. Hence, η is a natural transformation from F to G. As η_T is an isomorphism for each object T, it follows that η is a natural equivalence. The proof is complete. □

Note, as a consequence of Corollary 2.7 (a), that the category $\underline{\mathcal{Z}}$ of zero rings is an abelian category.

In view of the equivalence (1) $\Leftrightarrow$ (5) in Corollary 2.7 (a), one sees that, for the special case where $\underline{D}$ is an abelian category, the conclusion in Corollary 2.7 (b) (i) follows from an observation of Chase, Harrison and Rosenberg [5] that we mentioned in the penultimate paragraph of the Introduction. The conclusion obtained above in Corollary 2.7 (b) (i) (where $\underline{D}$ a pre-additive category with finite products) is a stronger result. Indeed, in the Foreword to [10] (a reprint of [9]), Freyd [10, page 21 of Foreword] has given an example of a pre-additive category with finite products which is not an abelian category. For the sake of completeness, the next result states the specifics of Freyd's example.

Example 2.9. (Freyd [10, page 21 of Foreword]) There exists a pre-additive category with finite products which is not an abelian category. One way to construct such a category $\underline{D}$ is the following. Let K be a field; let $X_1, X_2, \ldots$ be denumerably many (commuting) algebraically independent indeterminates over K; let A be the polynomial ring $K[\{X_i \mid i \geq 1\}]$; let I be the ideal of A generated by $\{X_i X_j \mid 1 \leq i \leq j\}$; let $R := A/I$; and let $\underline{D}$ be the full subcategory of ${}_R\text{Mod}$ whose class of objects is the collection of finitely presented R-modules. Indeed, the (endo)morphism in $\underline{D}(R, R)$ which is given by multiplication by $X_1 + I$ does not have a kernel in $\underline{D}$.

Let R be a zero ring and let $\underline{D}$ be a pre-additive category with finite products. Perhaps because of the nature of what has been emphasized in the existing literature, much of the above material has focused on functors $\underline{C}_R \to \underline{D}$ that are additive (equivalently, CHR-additive). We next give two examples showing that for suitable such $\underline{D}$ (for instance, take $\underline{D}$ to be Ab), a functor $\underline{C}_R \to \underline{D}$ need not be additive (equivalently, need not be CHR-additive). Some readers may find the following fact to be interesting. The functor constructed in Example 2.10 (a) sends every object to an object G such that $G \times G$ is not isomorphic to G, but the functor constructed in Example 2.10 (b) sends every object to an object G such $G \times G \cong G$.

Example 2.10. (a) Let R be a zero ring. Let $\underline{D}$ be a pre-additive category with finite products such that there exists an object G of $\underline{D}$ with the property that $G \times G$ and G are not isomorphic in $\underline{D}$. (For instance, take $\underline{D}$ = Ab and take G to be any nontrivial finite abelian group.) One can obtain a functor $F_1 : \underline{C}_R \to \underline{D}$ via the following construction. For each $S \in |\underline{C}_R|$, put $F_1(S) := G$; and for each morphism f in $\underline{C}_R$, define $F_1(f)$ to be the identity map on G. Then F_1 is not an additive functor and F_1 is not a CHR-additive functor.

(b) Let R be a zero ring. Let G be an infinite abelian group such that $G \times G \cong G$ in Ab. (For instance, take G to be the direct product of $\aleph_0$ many copies of $\mathbb{Z}/2\mathbb{Z}$.) One can obtain a functor $F_2 : \underline{C}_R \to$ Ab via the following construction. For each $S \in |\underline{C}_R|$, put $F_2(S) := G$; and for each morphism f in $\underline{C}_R$, define $F_2(f)$ to be the identity map on G. Then F_2 is not an additive functor and F_2 is not a CHR-additive functor.

Proof. (a) We can use considerations of cardinality to verify the parenthetical assertion that if G is a nontrivial finite abelian group, then $G \times G$ is not isomorphic to G in Ab. Indeed, since $n := |G|$ satisfies $2 \leq n < \infty$ by hypothesis, we have $|G| = n < n^2 = |G|^2 = |G \times G|$.

Let us now return to the main assertion. It is straightforward to check that F_1 is a functor. By Corollary 2.7 (b) (i), F_1 is not an additive functor if and only if F_1 is not a CHR-additive functor. We will show directly that the functor F_1 is neither additive nor CHR-additive. Recall that in any category $\underline{K}$, if T is a terminal object of $\underline{K}$, then the product $T \times T$ exists in $\underline{K}$ and $T \times T \cong T$ in $\underline{K}$. Thus, since $G \times G$ is not isomorphic to G in $\underline{D}$ by hypothesis, we get that G is not a terminal object of $\underline{D}$. Therefore, it follows from the second assertion in part (b) (resp., from part (a)) of Lemma 2.6 that F_1 is not an additive functor (resp., that F_1 is not a CHR-additive functor). For an alternative proof in case $\underline{D}$ = Ab and G is a nontrivial finite abelian group, combine the fact that the identity map $G \to G$ is not a zero morphism in Ab with the first assertion in Lemma 2.6 (b) to conclude that F_1 is not an additive functor (and then use Corollary 2.7 (b) (i) to conclude that F_1 is not CHR-additive).

(b) Let us first address the parenthetical assertion. This can be done via considerations of cardinality. Indeed, if G is the direct product in Ab of $\aleph_0$ many copies of $\mathbb{Z}/2\mathbb{Z}$, then in Ab, $G \times G$ is the direct product of $(\aleph_0)^2$ many copies of $\mathbb{Z}/2\mathbb{Z}$. However, a standard fact about arithmetic with infinite cardinal numbers (assuming, as we do, the ZFC foundations) gives $(\aleph_0)^2 = \aleph_0$, whence $G \times G \cong G$ in Ab, as desired.

Let us now return to the main assertion. Observe that the terminal objects in Ab are the trivial (necessarily abelian) groups. Of course, G is not a trivial group (since G is infinite), and so G is not a terminal object of Ab. To complete the proof, it is now straightforward to tweak the final two sentences of the above proof of (a). □

Much of what we have said here (and much of what we will say below) has to do with the fact that if R is a zero ring, then the category $\underline{C}_R$ has the property that each of its morphisms is an isomorphism. This property can be restated, using standard terminology in category theory, as saying that under the stated conditions, $\underline{C}_R$ is a *groupoid*. (Some users restrict the "groupoid" terminology to small categories in which each morphism is an isomorphism, but I will ignore that "small" foundational issue in this comment.) Many mathematicians are somewhat familiar with such a concept, having studied the fundamental group(oid) of a (possibly path-connected) topological space as part of an introduction to homotopy in a course on algebraic topology or winding numbers in a course on complex analysis. Interested readers are invited to examine how far one go in extending our methods here so as to generalize our results on $\underline{C}_R$ when R is a zero ring (that is, our work here on the category $\underline{\mathcal{Z}}$ of all zero rings) to categorical facts about groupoids in which each hom-set is a singleton set. See Remark 2.11 for an indication of what may/should be possible along these lines.

Remark 2.11. (a) Following Corollary 2.4, we mentioned that certain strong foundational assumptions that are consistent with ZFC allow one to prove that any category is equivalent to a skeletal category. In the spirit of the above comments about groupoids, part (a) of this remark will examine how the use of the above-mentioned foundational assumptions would simplify and strengthen the work in Example 2.10.

Consider the category $\underline{C}_R$ for some zero ring R. (By Corollary 2.7 (a), this $\underline{C}_R$ is equal to the category of all zero rings.) By the above-mentioned strong foundational assumptions, $\underline{C}_R$ is equivalent to a skeletal category, say $\underline{K}$. Since each hom-set in $\underline{C}_R$ is a singleton set, the "skeletal" property (in conjunction with the fact that a categorical equivalence is a certain kind of fully faithful functor) shows that $\underline{K}$ is the simplest kind of nonempty category, namely, a category with a unique object and a unique morphism. It is straightforward to verify directly that $\underline{K}$ is a pre-additive category with finite products. In fact, $\underline{K}$ is an abelian category. Let us consider what happens when the role of $\underline{C}_R$ in Example 2.10 is played instead by the equivalent category $\underline{K}$. The earlier roles of a pre-additive category $\underline{D}$ with finite products and of Ab will not change. The reader may wish to pause reading at this point in order to consider whether the possible additive or CHR-additive nature of a functor is (un)affected by this change of functorial domains. That issue will, in effect, be handled by the discussion in (b).

Because of the simple nature of $\underline{K}$, it is clear that the functors F from $\underline{K}$ to $\underline{D}$ (resp., to Ab) are in one-to-one correspondence with the objects G of $\underline{D}$ (resp., of Ab), via the assignment sending each F to its value at the unique object of $\underline{K}$. For each object G of $\underline{D}$, let F_G denote the functor F associated to G; that is, G is the value of F_G at the unique object of $\underline{K}$. It is a straightforward (and not overly long) exercise in basic category theory to show, by using both parts of Lemma 2.6 and [16, Proposition 1, page 194] (cf. also the second paragraph of the proof of Corollary 2.7 (b) (i)), that the following holds for each functor F as above (that is, $F = F_G$ for some object G): F is an additive functor $\Leftrightarrow$ F is an CHR-additive functor $\Leftrightarrow$ G is a terminal (equivalently, a null) object (of $\underline{D}$ or of Ab, depending on the context). This result suggests a underlying reason which explains why each of F_1 and F_2 in Example 2.10 was neither additive nor CHR-additive, namely, neither of the abelian groups G in parts (a) and (b) of Example 2.10 was a null object of Ab (that is, neither of those groups G was a trivial group). The present analysis via an equivalent skeletal category suggests/reveals that the key tool that has been introduced here for such questions is Lemma 2.6, whereas the fact that exactly one of the two abelian groups G in Example 2.10 satisfies $G \times G \cong G$ is, however interesting it may have seemed, only of peripheral importance.

(b) It seems reasonable to expect that some readers would not wish to make the strong foundational assumption that led us in (a) to replace $\underline{C}_R$ (for a zero ring R) with an equivalent skeletal category. Now that (a) has indicated what may be expected, we will proceed (with the help of some of the above material, especially Corollary 2.7) to show that those expectations are realized even if we use only the usual ZFC foundations. Once again, let R be a zero ring and let $\underline{D}$ be a pre-additive category with finite products. We will characterize when a functor from $\underline{C}_R \to \underline{D}$ is additive (resp., CHR-additive). As above, if $S, T \in |\underline{C}_R|$ (that is, if S and T are zero rings), we let $f_{S,T}$ denote the unique function (equivalently, the unique R-algebra homomorphism; equivalently, the isomorphism in $\underline{C}_R$) $S \to T$.

"Having" a functor $F : \underline{C}_R \to \underline{D}$ (or "letting" F be such a functor) is equivalent to having the following four items:

(i) a nonempty class $\mathcal{D}$ of pairwise isomorphic objects of $\underline{D}$;

(ii) for each (ordered) pair of objects $D_1, D_2 \in \mathcal{D}$, a singleton set $\{h_{D_1,D_2}\} \subseteq \underline{D}(D_1, D_2)$, such that

(ii)$_1$: for all $D_3, D_4, D_5, D_6 \in \mathcal{D}$, we have $h_{D_4,D_5} h_{D_3,D_4} = h_{D_3,D_5}$ and h_{D_6,D_6} is the identity map on D_6;

(iii) an assignment $S \mapsto F(S)$ sending each object S of $\underline{C}_R$ to some element $F(S) \in \mathcal{D}$, such that

(iii)$_1$: $\mathcal{D}$ is the collection of all objects of $\underline{D}$ that are of the form $F(S)$ for some object S of $\underline{C}_R$; and

(iv) for each (ordered) pair of objects S, T of $\underline{C}_R$, a function $\underline{C}_R(S, T) \to \underline{D}(F(S), F(T))$ sending the morphism $f_{S,T}$ to $h_{F(S),F(T)}$.

Indeed, the "object assignment" aspect of any functor F of the kind being considered must satisfy (i) and (iii), since a functor must preserve isomorphisms; and (ii)$_1$ and (iv) are required so that F behaves functorially on morphisms (that is, so that F behaves "homomorphically" on composites of morphisms and sends identity maps to the appropriate identity maps). Note that while the particular functors F_1 and F_2 that were constructed in Example 2.10 necessarily satisfied (i)-(iv), their construction was as simple as possible, in the sense that the corresponding sets playing the role of $\mathcal{D}$ were chosen to be singleton sets. The above (more complicated) characterization of relevant functors F in terms of (i)-(iv) will be needed in the next paragraph where we will characterize when an arbitrary such functor is additive (resp., CHR-additive). As a pedagogic aside, the explicitness of (i)-(iv) also serves as a reminder that the construction of a functor requires more than simply stipulating the associated object assignment. (Along those lines, one may examine the route that I took in proving the first significant result in my doctoral research, [7, Chapter I, Theorem 3.10]; we will have reason to mention [7] again in Remark 2.20.)

We can now state the desired result. Let R be a zero ring, let $\underline{D}$ be a pre-additive category with finite products, and let $F : \underline{C}_R \to \underline{D}$ be a functor. Let $\mathcal{D}$ be the collection of all objects of $\underline{D}$ that are of the form $F(S)$ for some object S of $\underline{C}_R$. Then the following five conditions are equivalent:

(1) F is an additive functor;

(2) F is a CHR-additive additive functor;

(3) Some (equivalently, every) element of $\mathcal{D}$ is a terminal object of $\underline{D}$;

(4) Some (equivalently, every) element of $\mathcal{D}$ is an initial object of $\underline{D}$;

(5) Some (equivalently, every) element of $\mathcal{D}$ is a null object of $\underline{D}$.

Let us very briefly sketch a proof of the result. Since $\underline{D}$ is a pre-additive category, it follows from [16, Proposition 1, page 194] that an object D of $\underline{D}$ is a terminal object of $\underline{D}$ if and only if D is an initial object of $\underline{D}$, if and only if D is a null object of $\underline{D}$. It follows that (3), (4) and (5) are equivalent. In view of the above methods and accumulated information (especially Corollary 2.7), we can give almost the same hint for a proof of the equivalence of (1), (2), (3) that was given for the proof of the analogous equivalences in (a): use both parts of Lemma 2.6, [16, Proposition 1, page 194], and the second and third paragraphs of the proof of Corollary 2.7 (b) (i). This completes the remark.

The constructions in Example 2.10, which gave functors F_1 and F_2 that were neither additive nor CHR-additive, had the property that F_1 and F_2 each sent all objects to an object that was not a terminal object. Nevertheless, one could use Corollary 2.7 (b) (i) to recover the fact that if R is a zero ring, then the restrictions of the unit functor and the Picard group functor each give a CHR-additive functor $\underline{C}_R \to$ Ab, since one can see directly and easily that each of these restrictions is an additive functor. For such a direct analysis, the underlying fact is that the unit group of a zero ring and the Picard group of a zero ring are each trivial groups. We will say more about "zero-ish" matters in Remark 2.19. In closing, Remark 2.20 will recall some facts about some variants/applications of

CHR-additive functors from some of our early work and point the way to possible future work at the interface of commutative algebra and algebraic geometry.

Before proceeding to the two above-mentioned remarks, we will give our second main result, Theorem 2.18, which addresses the more general categorical question that is suggested by the results in Corollary 2.7 (b) (i) and Remark 2.11 (b). Theorem 2.18 not only generalizes those earlier results, but it also serves to validate the insight of Chase, Harrison and Rosenberg in [5] that [9, Theorem 3.11] is relevant to finding contexts for which the concepts of “additive functor" and (what we have called) “CHR-additive functor" are equivalent.

The next sentence pertains to some terminology used in Lemma 2.12 (and later). We will use 1_E to denote the identity map $E \to E$ on an object E of a given category; and a “zero morphism," usually denoted by 0, in a pre-additive category $\underline{E}$ will refer to the neutral element (with respect to addition) in some hom-set of $\underline{E}$. Also, in a pre-additive category $\underline{E}$ with a null object N, the zero morphism in a hom-set $\underline{E}(E_1, E_2)$ is the same as a morphism $E_1 \to E_2$ that factors through N. (A proof of this fact can be found easily by using Lemma 2.1 (a).)

Lemma 2.12. *Let $F : \underline{C} \to \underline{D}$ be a functor, where $\underline{C}$ and $\underline{D}$ are each a pre-additive category with a null object. Then the following conditions are equivalent:*

(1) F sends each zero morphism of $\underline{C}$ to a zero morphism of $\underline{D}$;
(2) F sends some null object of $\underline{C}$ to a null object of $\underline{D}$;
(3) F sends each null object of $\underline{C}$ to a null object of $\underline{D}$.

Proof. (3) ⇒ (2): Trivial.

(2) ⇒ (3): This equivalence can be proven by applying the following three facts in order. Any two null objects of a given category are isomorphic; functors preserve isomorphisms; and null objects are preserved by isomorphisms.

(1) ⇒ (3): Assume (1). Let N be a null object of $\underline{C}$. Then 1_N is a zero morphism, by [16, Proposition 1, page 194]. Thus, by (1), $1_{F(N)} = F(1_N)$ is a zero morphism. Hence, by [16, Proposition 1, page 194], $F(N)$ is a null object.

(3) ⇒ (1): Assume (3). Let $f = 0 \in \underline{C}(C_1, C_2)$. Our task is to show that $g := F(f) \in \underline{D}(F(C_1), F(C_2))$ satisfies $g = 0$. By the above comments, $f = f_2 f_1$ where $f_1 = 0 \in \underline{C}(C_1, N)$ and $f_2 = 0 \in \underline{C}(N, C_2)$ for some null object N of $\underline{C}$. As (3) ensures that $F(N)$ is a null object of $\underline{D}$, it follows that $1_{F(N)} = 0$ by [16, Proposition 1, page 194], and so $F(f_2) = F(f_2)1_{F(N)} = F(f_2)0 = 0$ by Lemma 2.1 (a). Therefore, $g = F(f_2)F(f_1) = 0F(f_1) = 0$ by Lemma 2.1 (a). The proof is complete. □

Much in the results 2.15-2.18 will concern when the “canonical morphisms" $\alpha : F(A \times B) \to F(A) \times F(B)$ and $\beta : F(A) \oplus F(B) \to F(A \oplus B)$ are isomorphisms. Since (direct) product and (direct) sum are defined only up to isomorphism (even when they exist), one may well ask the apparently more basic question of whether α and β are actually well-defined morphisms. A specialist in category theory would perhaps reply, for good reason, "Yes, up to isomorphism." But such a reply may worry some readers who wish to avoid the conceptual problems associated with the 19$^{\text{th}}$ century's studies (before the foundations of the theory of Riemann surfaces were rigorously developed) of the supposed domain and range of a so-called “many-valued function." Propositions 2.13 and 2.14 will carefully identify and then resolve some of the underlying issues. Readers who are familiar with the definitions of Amitsur cohomology in a functor or of Cech cohomology in a presheaf (or perhaps only in a sheaf) have already dealt with similar issues. After all, the definitions of those cohomology groups use the definitions of some underlying cochain complexes, and the latter definitions assume that it is meaningful to apply a functor to certain tensor products or to apply a presheaf to certain fiber (co)products, even though those tensor products or fiber products are only defined up to isomorphism. Such readers who have already made their peace with that aspect of the literature (perhaps by emulating the spirit of [16, pages 195-196]) will likely not be surprised by the proofs of Propositions 2.13 and 2.14.

Proposition 2.13 will prove in detail that "α is an isomorphism" is a well-defined property (even though α is only "defined up to isomorphism"). The proof of Proposition 2.14, while being largely left to the reader, does give hints for ways to prove that "β is an isomorphism" is a well-defined property (even though β is only "defined up to isomorphism"). Our approach to proving Propositions 2.13 and 2.14 will not take the somewhat draconian step of replacing each relevant category with an equivalent skeletal category; nor will we pretend that it would suffice to merely point out that two objects of a certain comma category are isomorphic. Instead, in Proposition 2.13, we will assume only that we are given a functor $F : \underline{C} \to \underline{D}$, along with (possibly equal) objects A and B of $\underline{C}$, such that a product of A and B exists in $\underline{C}$ and a product of $F(A)$ and $F(B)$ exists in $\underline{D}$; we will show that, under these assumptions, various versions of α are certainly well defined and that, if one of those versions of α is an isomorphism, then all the other versions of α are also isomorphisms. In short, Proposition 2.13 gives a precise sense in which "α is an isomorphism" is a well defined property and shows that this property holds under what are arguably the most general conditions for which such a study should be pursued. One can say that Proposition 2.14 will essentially do for β (resp., sums) what Proposition 2.13 will have done for α (resp., products). Indeed, Proposition 2.14 gives the corresponding conclusions about the various analogous versions of β (assuming the existence of the appropriate sums, $A \oplus B$ in $\underline{C}$ and $F(A) \oplus F(B)$ in $\underline{D}$).

Proposition 2.13. *Let $\underline{C}$ be a category, and let A and B be (possibly isomorphic) objects of $\underline{C}$. Let P_1 be a product of A and B in $\underline{C}$, with projection morphisms $p_1 : P_1 \to A$ and $p_2 : P_1 \to B$. Let P_2 be a (possibly different) product of A and B in $\underline{C}$, with projection morphisms $p_1^* : P_2 \to A$ and $p_2^* : P_2 \to B$. Let $\theta : P_1 \to P_2$ be the unique morphism (actually, an isomorphism, given by the universal mapping property of the product P_2) such that $p_1^*\theta = p_1$ and $p_2^*\theta = p_2$. Let $\underline{D}$ be a category and let $F : \underline{C} \to \underline{D}$ be a functor. Let Q_1 be a product of $F(A)$ and $F(B)$ in $\underline{D}$, with projection morphisms $\pi_1 : Q_1 \to F(A)$ and $\pi_2 : Q_1 \to F(B)$. Let Q_2 be a (possibly different) product of $F(A)$ and $F(B)$ in $\underline{D}$, with projection morphisms $\pi_1^* : Q_2 \to F(A)$ and $\pi_2^* : Q_2 \to F(B)$. Let $\psi : Q_1 \to Q_2$ be the unique morphism (actually, an isomorphism, given by the universal mapping property of the product Q_2) such that $\pi_1^*\psi = \pi_1$ and $\pi_2^*\psi = \pi_2$. Let $\alpha_1 : F(P_1) \to Q_1$ be the (uniquely determined, since Q_1 is a product) canonical morphism such that $\pi_1\alpha_1 = F(p_1)$ and $\pi_2\alpha_1 = F(p_2)$. Let $\alpha_2 : F(P_2) \to Q_2$ be the (uniquely determined, since Q_2 is a product) canonical morphism such that $\pi_1^*\alpha_2 = F(p_1^*)$ and $\pi_2^*\alpha_2 = F(p_2^*)$. Then:*

(a) $\alpha_2 F(\theta) = \psi\alpha_1$.

(b) α_1 is an isomorphism (in $\underline{D}$) if and only if α_2 is an isomorphism (in $\underline{D}$).

Proof. Of course, θ and ψ are isomorphisms because in any category, any two objects with the same universal mapping property are isomorphic (cf. [14, proof of Theorem 7.3, page 54; also page 57]). As functors preserve isomorphisms, $F(\theta)$ is also an isomorphism. This fact will be used in the proof of (b). Also, a piece of the above information can be rewritten as $\pi_k^* = \pi_k\psi^{-1}$ for $k \in \{1,2\}$; this fact will be used in the proof of (a).

(a) Since ψ is an isomorphism, our task can be rephrased as the requirement to prove that $\psi^{-1}\alpha_2 F(\theta) = \alpha_1$. Thus, in view of the above characterization of α_1, our task can be rephrased as the requirement to prove that

$$\pi_1\psi^{-1}\alpha_2 F(\theta) = F(p_1) \text{ and } \pi_2\psi^{-1}\alpha_2 F(\theta) = F(p_2).$$

To accomplish this task, note that for $k \in \{1,2\}$, we have

$$F(p_k) = F(p_k^*\theta) = F(p_k^*)F(\theta) = \pi_k^*\alpha_2 F(\theta) = \pi_k\psi^{-1}\alpha_2 F(\theta).$$

(b) It will be enough to assume that α_1 is an isomorphism and then prove that α_2 is an isomorphism. Recall that $F(\theta)$ and ψ are isomorphisms. Thus, by (a), $\alpha_2 = \psi\alpha_1(F(\theta))^{-1}$, which is a composition of isomorphisms, and so α_2 is an isomorphism. □

Proposition 2.14. *Let $\underline{C}$ be a category, and let A and B be (possibly isomorphic) objects of $\underline{C}$. Let S_1 be a (direct) sum of A and B in $\underline{C}$, with injection morphisms $i_1 : A \to S_1$ and $i_2 : B \to S_1$. Let S_2 be a (possibly different) sum of A and B in $\underline{C}$, with injection morphisms $i_1^* : A \to S_2$ and $i_2^* : B \to S_2$. Let $\varphi : S_1 \to S_2$ be the unique morphism (actually, an isomorphism, given by the universal mapping property of the sum S_1) such that $\varphi i_1 = i_1^*$ and $\varphi i_2 = i_2^*$. Let $\underline{D}$ be a category and let $F : \underline{C} \to \underline{D}$ be a functor. Let T_1 be a sum of $F(A)$ and $F(B)$ in $\underline{D}$, with injection morphisms $j_1 : F(A) \to T_1$ and $j_2 : F(B) \to T_1$. Let T_2 be a (possibly different) sum of $F(A)$ and $F(B)$ in $\underline{D}$, with injection morphisms $j_1^* : F(A) \to T_2$ and $j_2^* : F(B) \to T_2$. Let $\Psi : T_1 \to T_2$ be the unique morphism (actually, an isomorphism, given by the universal mapping property of the sum T_1) such that $\Psi j_1 = j_1^*$ and $\Psi j_2 = j_2^*$. Let $\beta_1 : T_1 \to F(S_1)$ be the (uniquely determined, since T_1 is a sum) canonical morphism such that $\beta_1 j_1 = F(i_1)$ and $\beta_1 j_2 = F(i_2)$. Let $\beta_2 : T_2 \to F(S_2)$ be the (uniquely determined, since T_2 is a sum) canonical morphism such that $\beta_2 j_1^* = F(i_1^*)$ and $\beta_2 j_2^* = F(i_2^*)$. Then:*

(*a*) $\beta_2\Psi = F(\varphi)\beta_1$.

(*b*) *β_1 is an isomorphism (in $\underline{D}$) if and only if β_2 is an isomorphism (in $\underline{D}$).*

Proof. We leave to the reader the details involved in intuitively "dualizing" the proof of Proposition 2.13. Readers seeking a more rigorous approach to such "dualizing" arguments are encouraged to skip ahead to Remark 2.15 (a), to familiarize themselves with the one-to-one correspondence of functors $F \leftrightarrow \mathcal{F}$ (with $F : \underline{C} \to \underline{D}$ and $\mathcal{F} : \underline{C}^{\text{op}} \to \underline{D}^{\text{op}}$) which is established there, and then to use that correspondence to fashion an alternate proof of Proposition 2.14. The details of that alternate proof are also left to the reader. □

Some of the considerations in Theorem 2.18 will require us to go a step further than what was done in the preceding two results. In view of the statements of Propositions 2.13 and 2.14, one can ask whether, when given objects $A_1, \dots, A_n$ (possibly listed with repetition) of a pre-additive category $\underline{C}$ with finite products and a functor $F : \underline{C} \to \underline{D}$, where $\underline{D}$ is also a pre-additive category with finite products, one can say, for the "canonical morphisms" $\alpha : F(\prod_{i=1}^n A_i) \to \prod_{i=1}^n F(A_i)$ and $\beta : \coprod_{i=1}^n F(A_i) \to F(\coprod_{i=1}^n A_i)$, that "$\alpha$ is an isomorphism" and "β is an isomorphism" are well defined properties. (Notice that this question about α is basically asking whether the concept of a CHR-additive functor is well defined.) The answer(s) is/are in the affirmative, and we expect that the interested reader will be able to obtain this/these answer(s) by building on the proofs of Propositions 2.13 and 2.14. If necessary, some readers may wish to review the general associative laws (up to natural isomorphisms) for finite (direct) products and finite coproducts in such categories: in this regard, see Remark 2.15 (a) and the third paragraph of Remark 2.16.

In the spirit of the characterization of additive functors F of abelian categories given in [9, Theorem 3.11], one of the characterizations given in Theorem 2.18 will be that F carries biproduct diagrams to biproduct diagrams. Remark 2.15 (a) will recall the definition of a biproduct diagram (for the more general context of Theorem 2.18) and some of its useful consequences. Parts (b) and (c) of Remark 2.15 will develop a useful categorical technique that involves dual categories and will then examine some of its applications, especially to pre-additive categories with finite products.

Remark 2.15. (a) Let $\underline{C}$ and $\underline{D}$ be (possibly isomorphic) pre-additive categories with finite products. Let A and B be (possibly isomorphic) objects of $\underline{C}$. Fix a product P of A and B (in $\underline{C}$). By [16, Theorem 2, page 194] (and its proof), the existence of the product P leads to a biproduct diagram, say $\mathcal{D}$, for A and B in $\underline{C}$. According to the definition of a biproduct diagram (see [16, page 194]), $\mathcal{D}$ consists of injection morphisms, $i_1 : A \to P$ and $i_2 : B \to P$, and projection morphisms, $p_1 : P \to A$ and $p_2 : P \to B$, such that $p_1 i_1 = 1_A$, $p_2 i_2 = 1_B$ and $i_1 p_1 + i_2 p_2 = 1_P$ (and, necessarily, $p_1 i_2 = 0$ and $p_2 i_1 = 0$, as in [16, page 195, lines 3-4]). It follows that the maps i_1 and i_2 determine P as a sum (that is, as a coproduct) of A and B (in $\underline{C}$) while the maps p_1 and p_2 determine P as a product of A and B (in $\underline{C}$).

It will be interesting and useful to ask whether applying F to the data in $\mathcal{D}$ produces a biproduct diagram in $\underline{D}$. By definition, an affirmative answer would mean that the object $\mathcal{P} := F(P)$ of $\underline{D}$ has injection morphisms, $j_1 := F(i_1) : F(A) \to \mathcal{P}$ and $j_2 := F(i_2) : F(B) \to \mathcal{P}$, and projection morphisms, $\pi_1 := F(p_1) : \mathcal{P} \to F(A)$ and $\pi_2 := F(p_2) : \mathcal{P} \to F(B)$, such that $\pi_1 j_1 = 1_{F(A)}$, $\pi_2 j_2 = 1_{F(B)}$ and $j_1\pi_1 + j_2\pi_2 = 1_{\mathcal{P}}$ (and, necessarily, $\pi_1 j_2 = 0$ and $\pi_2 j_1 = 0$). Then $\mathcal{P}$ would be a sum of $F(A)$ and $F(B)$ (in $\underline{D}$) determined by the injection morphisms j_1 and j_2, while $\mathcal{P}$ would be a product of $F(A)$ and $F(B)$ (in $\underline{D}$) determined by the projection morphisms π_1 and π_2.

A nontrivial consequence of the reasoning two paragraphs ago is that if $\underline{C}$, A and B are as above (not necessarily such that F carries $\mathcal{D}$ to a biproduct diagram), then any product of A and B in $\underline{C}$ is a sum of A and B in $\underline{C}$. Combining this observation with Lemma 2.1 (b), we get (for $\underline{C}$, A and B as above) that any sum of A and B in $\underline{C}$ is a product of A and B in $\underline{C}$. It is not difficult to conclude therefrom that a pre-additive category with finite products is the same as a pre-additive category with finite coproducts. These basic facts will be used in Theorem 2.18. While the just-mentioned facts in this paragraph and the facts in the preceding two paragraphs are true, we believe that the above exposition of them here has been incomplete, owing to what we consider to be some incomplete or vague passages in [16]. (For the same reason, one should also comment further on the use of the proof of [16, Theorem 2, page 195] which will be implicitly appealed to later in part (c) of the present remark in order to have certain biproduct diagrams.) Remark 2.16 will, in our opinion, provide enough details to rectify matters.

(b) This paragraph will develop the following result: for *any* categories $\underline{K}_1$ and $\underline{K}_2$, there is a natural one-to-one correspondence between the (covariant) functors $F : \underline{K}_1 \to \underline{K}_2$ and the (covariant) functors $\mathcal{F} : \underline{K}_1^{\mathrm{op}} \to \underline{K}_2^{\mathrm{op}}$. (Of course, the notation $\underline{K}_i^{\mathrm{op}}$ means $(\underline{K}_i)^{\mathrm{op}}$.) To see this, let us begin by observing that any F as above induces a (covariant) functor $\mathcal{F} : \underline{K}_1^{\mathrm{op}} \to \underline{K}_2^{\mathrm{op}}$, where the object assignment of $\mathcal{F}$ is the same as the object assignment of F, while $\mathcal{F}$ is defined on morphisms by $\mathcal{F}(f^{\mathrm{op}}) := (F(f))^{\mathrm{op}}$. Noting that any category $\underline{K}$ satisfies $(\underline{K}^{\mathrm{op}})^{\mathrm{op}} = \underline{K}$, one checks easily that if a functor F induces $\mathcal{F} : \underline{K}_1^{\mathrm{op}} \to \underline{K}_2^{\mathrm{op}}$ as above, then the functor that $\mathcal{F}$ induces from $\underline{K}_1$ to $\underline{K}_2$ is F itself. Similarly, if one uses a functor $\mathcal{G} : \underline{K}_1^{\mathrm{op}} \to \underline{K}_2^{\mathrm{op}}$ to induce a functor $G : \underline{K}_1 \to \underline{K}_2$, then one checks easily that the functor which G induces from $\underline{K}_1^{\mathrm{op}}$ to $\underline{K}_2^{\mathrm{op}}$ is $\mathcal{G}$ itself. This completes (a sketch of) the proof of the above-asserted one-to-one correspondence between the functors $\underline{K}_1 \to \underline{K}_2$ and the functors $\underline{K}_1^{\mathrm{op}} \to \underline{K}_2^{\mathrm{op}}$. (We have not used notation such as "$\mathcal{F} = F^{\mathrm{op}}$", nor will we do so, for the following reason. A number of speakers and authors have seen fit to convert a (possibly naturally occurring) "contravariant functor" $H : \underline{K}_1 \to \underline{K}_2$ to a (covariant) functor, which they have denoted by H^{op}, either from $\underline{K}_1^{\mathrm{op}}$ to $\underline{K}_2$ or from $\underline{K}_1$ to $\underline{K}_2^{\mathrm{op}}$. That sort of construction should not be confused with the above assignment $F \mapsto \mathcal{F}$ that induced the above one-to-one correspondence, as our construction of $\mathcal{F}$ in terms of F involved dualizing *both* the domain of F *and* the codomain of F. Apart from this parenthetical aside, all the functors considered in this paper are assumed to be covariant.)

This paragraph collects some material that can be useful in applying the preceding paragraph to pre-additive categories. First, recall from the proof of Lemma 2.1 (b) that the definition of addition of morphisms in the dual of a pre-additive category, which was essentially given by

$$\lambda^{\mathrm{op}} + \mu^{\mathrm{op}} := (\lambda + \mu)^{\mathrm{op}},$$

led to the fact that a category $\underline{K}^{\mathrm{op}}$ is a pre-additive category if (and only if) $\underline{K}$ is a pre-additive category. The next two observations will be useful for the context of Theorem 2.18, that is, whenever when $\underline{C}$ and $\underline{D}$ are each pre-additive categories with finite products and $F : \underline{C} \to \underline{D}$ is a functor. Let $\mathcal{F} : \underline{C}^{\mathrm{op}} \to \underline{D}^{\mathrm{op}}$ be the functor induced by F via the construction in the preceding paragraph. Loosely stated, here is the next observation: if F satisfies condition (2) in the statement of Theorem 2.18, then so does $\mathcal{F}$. More precisely put: if F carries each biproduct diagram in $\underline{C}$ to a biproduct diagram in $\underline{D}$, then $\mathcal{F}$ carries each biproduct diagram in $\underline{C}^{\mathrm{op}}$ to a biproduct diagram in $\underline{D}^{\mathrm{op}}$. (For a proof, combine the following three items: the definition of a biproduct diagram in [16, Definition, page 194], as recalled in (a) above; the equivalence of (1) and (2) in Theorem 2.18 (taken directly from [16, Proposition 4, page 197]); and the fact (which is a consequence of the second sentence in this paragraph) that additive functors preserve each of the three identities appearing in the definition of a biproduct diagram.) Here is another useful observation: it is easy to see that F sends each zero morphism in $\underline{C}$ to a zero morphism if and only if $\mathcal{F}$ sends each zero morphism in $\underline{C}^{\text{op}}$ to a zero morphism. (In view of the definition of $\mathcal{F}$ on morphisms, the following elementary categorical observation provides the appropriate detail to prove the preceding "easy" comment. It follows from Lemma 2.1 (b) and the definition of addition of morphisms in the dual of a pre-additive category $\underline{E}$ that if μ is the neutral element in a hom-set (abelian group) $\underline{E}(E_1, E_2)$ (that is, if μ is the zero morphism from E_1 to E_2 in $\underline{E}$), then μ^{op} is the zero morphism from E_2 to E_1 in $\underline{E}^{\text{op}}$.)

(c) Let $\underline{C}$ and $\underline{D}$ be (possibly isomorphic) pre-additive categories with finite products, let $H : \underline{C} \to \underline{D}$ be a functor, and fix (possibly isomorphic) objects A and B of $\underline{C}$. It will often be convenient to replace the previously used notations α and β with α_H and β_H, respectively. Thus, for fixed A and B as above, α_H denotes the canonical morphism $H(A \times B) \to H(A) \times H(B)$ in $\underline{D}$ and β_H denotes the canonical morphism $H(A) \oplus H(B) \to H(A \oplus B)$ in $\underline{D}$. If $F \leftrightarrow \mathcal{F}$ is the one-to-one correspondence constructed in (b) and if $A, B \in |\underline{C}|$, then $\alpha_F = (\beta_{\mathcal{F}})^{\text{op}}$ and (so, by replacing F with $\mathcal{F}$, which is permissible in view of Lemma 2.1 (b), we also get that) $\beta_F = (\alpha_{\mathcal{F}})^{\text{op}}$.

For a proof, we begin by fixing a product P of A and B in $\underline{C}$. This leads to a biproduct diagram, say $\mathcal{D}$, for A and B in $\underline{C}$. By definition, $\mathcal{D}$ consists of injection morphisms, $i_1 : A \to P$ and $i_2 : B \to P$, and projection morphisms, $p_1 : P \to A$ and $p_2 : P \to B$, such that $p_1 i_1 = 1_A$, $p_2 i_2 = 1_B$ and $i_1 p_1 + i_2 p_2 = 1_P$ (and, necessarily, $p_1 i_2 = 0$ and $p_2 i_1 = 0$). It follows that i_1 and i_2 determine P as a coproduct of A and B (in $\underline{C}$) while p_1 and p_2 determine P as a product of A and B (in $\underline{C}$). Next, fix a product $\mathcal{P}$ of $F(A)$ and $F(B)$ in $\underline{D}$. As above, an ensuing biproduct diagram (this time, in $\underline{D}$) features injection morphisms, $j_1 : F(A) \to \mathcal{P}$ and $j_2 : F(B) \to \mathcal{P}$, and projection morphisms, $\pi_1 : \mathcal{P} \to F(A)$ and $\pi_2 : \mathcal{P} \to F(B)$, such that $\pi_1 j_1 = 1_{F(A)}$, $\pi_2 j_2 = 1_{F(B)}$ and $j_1\pi_1 + j_2\pi_2 = 1_{\mathcal{P}}$ (and, necessarily, $\pi_1 j_2 = 0$ and $\pi_2 j_1 = 0$). Then $\mathcal{P}$ is a coproduct of $F(A)$ and $F(B)$ (in $\underline{D}$) determined by the injection morphisms j_1 and j_2, while $\mathcal{P}$ is a product of $F(A)$ and $F(B)$ (in $\underline{D}$) determined by the projection morphisms π_1 and π_2.

Let us repeat the above reasoning, this time focusing on the functor $\mathcal{F}$ (rather than on F) and on A and B as objects of $\underline{C}^{\text{op}}$ (rather than $\underline{C}$). The upshot is that the injection morphisms, $(i_1)^* := (p_1)^{\text{op}} : A \to P$ and $(i_2)^* := (p_2)^{\text{op}} : B \to P$ in $\underline{C}^{\text{op}}$, determine P as a *coproduct* of A and B in $\underline{C}^{\text{op}}$, while the injection morphisms, $(j_1)^* := (\pi_1)^{\text{op}} : \mathcal{F}(A) = F(A) \to \mathcal{P}$ and $(j_2)^* := (\pi_2)^{\text{op}} : \mathcal{F}(B) = F(B) \to \mathcal{P}$ in $\underline{D}^{\text{op}}$, determine $\mathcal{P}$ as a *coproduct* of $\mathcal{F}(A)$ and $\mathcal{F}(B)$ in $\underline{D}^{\text{op}}$.

It remains to prove that $\alpha_F = (\beta_{\mathcal{F}})^{\text{op}}$. It follows from the universal mapping properties of product and coproduct that α_F is (uniquely) determined by the conditions $\pi_1\alpha_F = F(p_1)$ and $\pi_2\alpha_F = F(p_2)$; and that $\beta_{\mathcal{F}}$ is determined by the conditions $\beta_{\mathcal{F}}(j_1)^* = \mathcal{F}((i_1)^*)$ and $\beta_{\mathcal{F}}(j_2)^* = \mathcal{F}((i_2)^*)$. It suffices to prove that $(\beta_{\mathcal{F}})^{\text{op}}$ has the just-mentioned properties which determine α_F. In other words, it suffices to prove that

$$\pi_1(\beta_{\mathcal{F}})^{\text{op}} = F(p_1) \text{ and } \pi_2(\beta_{\mathcal{F}})^{\text{op}} = F(p_2).$$

As the proofs of the two just-displayed equations are similar, we will give the first of those proofs next, while leaving the proof of the second equation to the reader. By applying the $^{\text{op}}$ operator, we see that our task is equivalent to proving that

$$(\pi_1(\beta_{\mathcal{F}})^{\text{op}})^{\text{op}} = (F(p_1))^{\text{op}}.$$

Accordingly, the proof concludes via the following calculation:

$$(\pi_1(\beta_{\mathcal{F}})^{\text{op}})^{\text{op}} = \beta_F(\pi_1)^{\text{op}} = \beta_F(j_1)^* = \mathcal{F}((i_1)^*) =$$

$\mathcal{F}((p_1)^{\text{op}}) := (F(p_1))^{\text{op}}$.

We are optimistic that additional uses of the one-to-one correspondence $F \leftrightarrow \mathcal{F}$ will be noticed and become popular in presentations of a variety of topics in category theory. This completes the remark.

The next remark fulfills the expository purposes that were mentioned in the third paragraph of Remark 2.15 (a).

Remark 2.16. Let A and B be objects of a pre-additive category $\underline{E}$. At first glance, the proof of [16, Theorem 2, pages 194-195] would seem to prove the following two things: the first sentence of the statement of that result, namely, that a (direct) product $A \prod B$ (also denoted by $A \times B$) exists in $\underline{E}$ if and only if A and B have a biproduct diagram (in the sense defined in [16, Theorem 2, page 194] and recalled in Remark 2.15 (a)) in $\underline{E}$; and the first part of the second sentence in the statement of [16, Theorem 2, page 194], which essentially explains how to extract a product of A and B from a biproduct diagram associated to A and B. The statement of [16, Theorem 2, page 194] asserts more (and the proof of Theorem 2.18 and some argumentation leading to that proof will need it), namely, for A, B and $\underline{E}$ as above, the following two things: the second part of the second sentence in the statement of [16, Theorem 2, page 194], which essentially explains how to extract a coproduct of A and B from a biproduct diagram associated to A and B; and the assertion that $A \prod B$ exists in $\underline{E}$ if and only if a (direct sum, that is, a) coproduct $A \coprod B$ (also denoted by $A \oplus B$) exists in $\underline{E}$. (Note that the "only if" part of the preceding equivalence was given in [9, Exercise A1, page 60], although Freyd's usage ruled out the easy case where $\underline{E}$ has at most one object. In the just-mentioned exercise, Freyd went on conclude that $A \coprod B$ is isomorphic to $A \prod B$ if the latter exists (cf. also [9, Theorem 2.35], which is a result on abelian categories). That same conclusion can be drawn from the statement of [16, Theorem 2, page 194].) To be fair, these extra assertions in the statement of [16, Theorem 2, page 194] can be viewed as proven, as the statement of that result includes the word "dually" and we can see, thanks to Lemma 2.1 (b), that such usage is appropriate. (For instance, a product of A and B in the pre-additive category $\underline{E}^{\mathrm{op}}$ would be the same as a coproduct of A and B in $\underline{E}$; and an instructive calculation using the operator ${}^{\mathrm{op}}$ reveals how a biproduct diagram associated to A and B in $\underline{E}^{\mathrm{op}}$ leads to a biproduct diagram associated to A and B in $\underline{E}$.)

We next follow up on the above comment about "Freyd's usage". As category theory was a quickly developing field in the early 1960s, it is perhaps not surprising that with the passage of time, some of the terminology that had been used in [9] has changed its meaning. We will next give two instances of such changes. (These are pertinent to the statement of some results in [9] that are related to [9, Theorem 3.11]). First, what was called an "additive category" in [9, page 60] is what would nowadays be called a "pre-additive category with finite products and a null object". (Equivalently, by [16, Proposition 1, page 194], a result whose validity we explicated above with the aid of Lemma 2.1 (b), this kind of category would nowadays be called a "pre-additive category with finite products".) Second, what was called a "pre-additive category" in [9, page 60] would now be called a "pre-additive category with a null object and greater than one object."

Lastly, we address something that was mentioned in the Introduction and is implicit in the statement of conditions (5) and (7) of Theorem 2.18, namely, the fact that a pre-additive category $\underline{K}$ with finite products necessarily also has finite coproducts. The issue of the existence of an empty coproduct (that is, an initial object) of $\underline{K}$ is settled by [16, Proposition 1, page 194], which guarantees that any empty product in $\underline{K}$ (that is, any terminal object of $\underline{K}$) is an initial object of $\underline{K}$. The issue of the existence of binary coproducts $A \oplus B$ in $\underline{K}$ was handled two paragraphs ago. Finally, for integers $n \geq 3$, the existence of coproducts $\coprod_{i=1}^{n} A_i$ in $\underline{K}$ can be discerned from the proof that (6) $\Rightarrow$ (7) in Theorem 2.18 below. This completes the remark.

The next result contains the final technical information that will be needed in the proof of Theorem 2.18. Readers seeking a shorter path to Theorem 2.18 may be interested to know that the proof of Proposition 2.17 (a) was the last proof that I completed while doing this research, as it enabled me to complete the proof that (4) $\Rightarrow$ (3) in Theorem 2.18.

To avoid possible confusion, Proposition 2.17 and Theorem 2.18 will occasionally use the following enhanced notation for zero morphisms. If E_1 and E_2 are (possibly equal) objects of a pre-additive category $\underline{E}$, then the neutral element of the abelian group $\underline{E}(E_1, E_2)$ will be denoted by $0_{E_1,E_2}$.

Proposition 2.17. *Let $\underline{C}$ and $\underline{D}$ be (possibly isomorphic) pre-additive categories with finite products, and let $F : \underline{C} \to \underline{D}$ be a functor. Let $\mathcal{F} : \underline{C}^{\text{op}} \to \underline{D}^{\text{op}}$ be the functor that is induced by F by using the construction in Remark 2.15 (b). Let A and B be (possibly equal) objects of $\underline{C}$. Let P be a product of A and B in $\underline{C}$ (and hence also a sum of A and B in $\underline{C}$), with projection morphisms $p_1 : P \to A$ and $p_2 : P \to B$, and also with injection morphisms $i_1 : A \to P$ and $i_2 : B \to P$, such that the set of data $\{A, B, P, p_1, p_2, i_1, i_2\}$ gives a (uniquely determined) biproduct diagram $\mathcal{D}$ in $\underline{C}$ (that is, such that $p_1 i_1 = 1_A$, $p_2 i_2 = 1_B$ and $i_1 p_1 + i_2 p_2 = 1_P$, and, necessarily, $p_1 i_2 = 0_{B,A}$ and $p_2 i_1 = 0_{A,B}$). Put $\mathcal{P} := F(P)$. Let Q be a product of $F(A)$ and $F(B)$ in $\underline{D}$ (and hence also a sum of $F(A)$ and $F(B)$ in $\underline{D}$), with projection morphisms $\pi_1 : Q \to F(A)$ and $\pi_2 : Q \to F(B)$, and also with injection morphisms $j_1 : F(A) \to Q$ and $j_2 : F(B) \to Q$, such that the set of data $\{F(A), F(B), Q, \pi_1, \pi_2, j_1, j_2\}$ gives a (uniquely determined) biproduct diagram $\mathcal{E}$ in $\underline{D}$ (that is, such that $\pi_1 j_1 = 1_{F(A)}$, $\pi_2 j_2 = 1_{F(B)}$ and $j_1\pi_1 + j_2\pi_2 = 1_Q$, and, necessarily, $\pi_1 j_2 = 0_{F(B),F(A)}$ and $\pi_2 j_1 = 0_{F(A),F(B)}$). Let α be the canonical morphism $F(A \times B) \to F(A) \times F(B)$, viewed as the morphism $\alpha : \mathcal{P} \to Q$ in $\underline{D}$ that is uniquely determined by $\pi_1\alpha = F(p_1)$ and $\pi_2\alpha = F(p_2)$. Also, consider the canonical morphism $\beta : F(A) \oplus F(B) \to F(A \oplus B)$, viewed as the morphism $\beta : Q \to \mathcal{P}$ in $\underline{D}$ that is uniquely determined by $\beta j_1 = F(i_1)$ and $\beta j_2 = F(i_2)$. Then:*

(a) Suppose that B is a null object of $\underline{C}$ and that α is an isomorphism. Then $F(0_{A,B}) = 0_{F(A),F(B)}$.

(b) Suppose that A is a null object of $\underline{C}$ and that β is an isomorphism. Then $F(0_{A,B}) = 0_{F(A),F(B)}$.

(c) Suppose that α is an isomorphism, $F(0_{A,B}) = 0_{F(A),F(B)}$ and $F(0_{B,A}) = 0_{F(B),F(A)}$. Then β is an isomorphism.

(d) Suppose that β is an isomorphism, $F(0_{A,B}) = 0_{F(A),F(B)}$ and $F(0_{B,A}) = 0_{F(B),F(A)}$. Then α is an isomorphism.

(e) Suppose that F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$. Then both α and β are isomorphisms.

(f) Suppose that α is an isomorphism, $F(0_{A,B}) = 0_{F(A),F(B)}$ and $F(0_{B,A}) = 0_{F(B),F(A)}$. Then F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$.

(g) Suppose that β is an isomorphism, $F(0_{A,B}) = 0_{F(A),F(B)}$ and $F(0_{B,A}) = 0_{F(B),F(A)}$. Then F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$.

Proof. Note that the existence of the product P (resp., Q) implies the existence of the biproduct diagram $\mathcal{D}$ (resp., $\mathcal{E}$) with the stated properties by virtue of the proof of [16, Theorem 2, pages 194-195] (as supplemented by Remark 2.16). Note also that the various assumptions or conclusions that α (resp., β) is an isomophism are unambiguous, by Proposition 2.13 (b) (resp., Proposition 2.14 (b)).

(a) Since B is a terminal object of $|\underline{C}|$, there is a unique morphism, say u, from A to B in $|\underline{C}|$. Since $\underline{C}$ is a pre-additive category, u is a zero morphism; that is, $u = 0_{A,B}$. Let $C \in |\underline{C}|$. Let v denote the unique morphism from C to B in $|\underline{C}|$. By the uniqueness of v, it is clear that $u\theta = v$ for each $\theta \in \underline{C}(C, A)$. It follows that for each $\lambda \in \underline{C}(C, A)$, there exists a unique $\psi \in \underline{C}(C, A)$ such that $1_A\psi = \lambda$ and $u\psi = v$, namely, $\psi = \lambda$. Therefore, A is a product of A and B in $\underline{C}$ when considered together with the projection maps $p_1^* = 1_A : A \to A$ and $p_2^* = u : A \to B$. This view of A as a product with respect to these projection maps leads to an associated biproduct diagram in $\underline{C}$, by the proof of [16, Theorem 2, pages 194-195] (cf. also Remark 2.15 (a) and the first paragraph of Remark 2.16). In view of Proposition 2.13 (b), it is clear that, in regard to the task of proving (a), there is no harm in taking $A = P$, $p_1^* = p_1$ and $p_2^* = p_2$, with the just-mentioned biproduct diagram being $\mathcal{D}$, together with the items i_1 and i_2 as in the statement of (a). Then we also have $\mathcal{P} := F(P) = F(A)$.

Let $\tilde{u}$ denote the morphism $F(u) : \mathcal{P} \to F(B)$. We claim that $\mathcal{P}$ is a product of $F(A)$ and $F(B)$ in $\underline{D}$ when considered together with the projection maps $1_{F(A)} : \mathcal{P} \to F(A)$ and $\tilde{u}$. Recall that $\alpha : \mathcal{P} = F(A \times B) \to F(A) \times F(B) = Q$ is assumed to be an isomorphism; and that Q is a product of $F(A)$ and $F(B)$ with associated projection morphisms $\pi_1 : Q \to F(A)$ and $\pi_2 : Q \to F(B)$. Consequently, $\mathcal{P}$ is a product of $F(A)$ and $F(B)$ in $\underline{D}$ when considered together with the projection maps $\pi_1\alpha : \mathcal{P} \to F(A) = \mathcal{P}$ and $\pi_2\alpha : \mathcal{P} \to F(B)$. Recall from Proposition 2.13 that the definition of α entails that $\pi_1\alpha = F(p_1)$ and

$\pi_2\alpha = F(p_2)$. Observe that $F(p_1) = F(p_1^*) = F(1_A) = 1_{F(A)}$; and $F(p_2) = F(p_2^*) = F(u) = \tilde{u}$. This completes the proof of the above claim.

In view of the above established claim, it is clear that, in regard to the task of proving (a), there is no harm in viewing $F(A)\times F(B)$ as $Q = \mathcal{P}$, together with the projection maps $\pi_1 = 1_{F(A)}$ and $\pi_2 = \tilde{u}$. Now, consider any object D of $\underline{D}$, along with any morphisms $\rho_1 \in \mathcal{D}(D, F(A))$ and $\rho_2 \in \mathcal{D}(D, F(B))$. Since $\mathcal{P}$ has the universal mapping property of a product of $F(A)$ and $F(B)$ in $\underline{D}$, there exists a unique morphism $\varphi : D \to \mathcal{P}$ in $\underline{D}$ such that $1_{F(A)}\varphi = \rho_1$ and $\tilde{u}\varphi = \rho_2$. Necessarily, $\varphi = 1_{F(A)}\varphi = \rho_1$, and so $\tilde{u}\rho_1 = \rho_2$.

Now, consider the special case of the result in the preceding paragraph when we take $D := F(A)$ and $\rho_1 := 1_{\mathcal{P}}$. With ρ now playing the role of ρ_2 from the preceding paragraph, we get the following conclusion: for each morphism $\rho : F(A) \to F(B)$ in $\mathcal{D}$, $\varphi = 1_{\mathcal{P}}$ is the unique morphism $\mathcal{P} \to \mathcal{P}$ in $\underline{D}$ such that $\tilde{u}\varphi = \rho$, and so $\rho = \tilde{u}\varphi = \tilde{u}1_{\mathcal{P}} = \tilde{u}$. Hence, $\tilde{u}$ is the only element of the set $\underline{D}(F(A), F(B))$. Since $\underline{D}$ is a pre-additive category, $\tilde{u}$ is a zero morphism; that is, $\tilde{u} = 0_{F(A),F(B)}$. Thus,

$$F(0_{A,B}) = F(u) = \tilde{u} = 0_{F(A),F(B)},$$

as desired. The proof of (a) is complete.

(b) We have an object P that is both a product of A and B in $\underline{C}$ and a sum of A and B in $\underline{C}$; and we also have an object Q that is both a product of $F(A)$ and $F(B)$ in $\underline{D}$ and a sum of $F(A)$ and $F(B)$ in $\underline{D}$. Note that P is both a sum of A and B in $\underline{C}^{\text{op}}$ and a product of A and B in $\underline{C}^{\text{op}}$; and Q is both a sum of $F(A)$ and $F(B)$ in $\underline{D}^{\text{op}}$ and a product of $F(A)$ and $F(B)$ in $\underline{D}^{\text{op}}$. Also, recall that $\mathcal{P} := F(P)$.

For reasons that will become clear, let α_F (rather than simply α) denote the canonical morphism $F(A\times B) \to F(A)\times F(B)$, viewed as $\alpha_F : \mathcal{P} \to Q$, in $\underline{D}$; and similarly, let β_F (rather than simply β) denote the canonical morphism $F(A)\oplus F(B) \to F(A\oplus B)$, viewed as $\beta_F : Q \to \mathcal{P}$, in $\underline{D}$. Let $\mathcal{F} : \underline{C}^{\text{op}} \to \underline{D}^{\text{op}}$ be the functor induced by F via the construction in Remark 2.15 (b). In the spirit of two sentences ago, we introduce the following notation: let $\alpha_{\mathcal{F}}$ denote the canonical morphism $\mathcal{F}(A\times B) \to \mathcal{F}(A)\times\mathcal{F}(B)$, viewed as $\alpha_{\mathcal{F}} : \mathcal{P} \to Q$, in $\underline{D}^{\text{op}}$; and let $\beta_{\mathcal{F}}$ denote the canonical morphism $\mathcal{F}(A)\oplus\mathcal{F}(B) \to \mathcal{F}(A\oplus B)$, viewed as $\beta_{\mathcal{F}} : Q \to \mathcal{P}$, in $\underline{D}^{\text{op}}$. Recall from Remark 2.15 (c) that $\alpha_F = (\beta_{\mathcal{F}})^{\text{op}}$ and $\beta_F = (\alpha_{\mathcal{F}})^{\text{op}}$.

There are two hypotheses in (b). The first of these is the assumption that A is a null object of $\underline{C}$, and this assumption is equivalent to A being a null object of $\underline{C}^{\text{op}}$. The second hypothesis in (b) is that β_F is an isomorphism in $\underline{D}$; equivalently, that $(\alpha_{\mathcal{F}})^{\text{op}}$ is an isomorphism in $\underline{D}$. We claim that this second hypothesis implies that $\alpha_{\mathcal{F}}$ is an isomorphism in $\underline{D}^{\text{op}}$. (Actually, one can strengthen that "implies" to "is equivalent to" here, but we will not need that stronger fact.) To prove the above claim, it suffices to show that if $h : E_1 \to E_2$ is an isomorphism in a category $\underline{E}$, then $h^{\text{op}} : E_2 \to E_1$ is an isomorphism in $\underline{E}^{\text{op}}$. To accomplish this task will require two elementary categorical observations. For those, see the next paragraph.

First, it follows easily from the definition of composition of morphisms in a dual category that if E is an object of a category $\underline{E}$ and $f := 1_E : E \to E$ is the identity morphism on E in $\underline{E}$, then $f^{\text{op}} : E \to E$ is the identity morphism on E in $\underline{E}^{\text{op}}$. Second, if $h : E_1 \to E_2$ is an isomorphism in a category $\underline{E}$ with inverse $h^{-1} : E_2 \to E_1$ (in $\underline{E}$), then it follows easily from the preceding sentence that $h^{\text{op}} : E_2 \to E_1$ is an isomorphism in $\underline{E}^{\text{op}}$ with inverse $(h^{-1})^{\text{op}} : E_1 \to E_2$ (in $\underline{E}^{\text{op}}$). It is clear that the above claim follows from the two observations in this paragraph.

We can use the above (established) claim that $\alpha_{\mathcal{F}}$ is an isomorphism in $\underline{D}^{\text{op}}$ by applying (a) to $\mathcal{F}$. (Note that (a) is applicable to $\mathcal{F}$ by Lemma 2.1 (b) since $\underline{C}^{\text{op}}$ and $\underline{D}^{\text{op}}$ inherit from $\underline{C}$ and $\underline{D}$, respectively, the property of being a pre-additive category with finite products. That application shows that $\mathcal{F}$ sends the zero morphism $u : B \to A$ in $\underline{C}^{\text{op}}$ to the zero morphism $v : \mathcal{F}(B) = F(B) \to \mathcal{F}(A) = F(A)$ in $\underline{D}^{\text{op}}$.

By the final comment in Remark 2.15 (b), it follows that if $\underline{E}$ is a pre-additive category and if μ is the neutral element in a hom-set (abelian group) $\underline{E}(E_1, E_2)$ (that is, μ is the zero morphism from E_1 to E_2 in $\underline{E}$), then μ^{op} is the zero morphism from E_2 to E_1 in $\underline{E}^{\text{op}}$. In particular, if z denotes the zero morphism from A to B in $\underline{C}$, then z^{op} is the zero morphism from B to A in $\underline{C}^{\mathrm{op}}$ (that is, $z^{\mathrm{op}} = u$); and if ζ denotes the zero morphism from $F(A)$ to $F(B)$ in $\underline{D}$, then ζ^{op} is the zero morphism from $F(B)$ to $F(A)$ in $\underline{D}^{\mathrm{op}}$ (that is, $\zeta^{\mathrm{op}} = v$, the zero morphism from $\mathcal{F}(B)$ to $\mathcal{F}(A)$ in $\underline{D}^{\mathrm{op}}$). Therefore, by applying (a) to $\mathcal{F}$ and also using the definition of $\mathcal{F}$ on morphisms from Remark 2.15 (b) (together with several applications of the fact that any morphism g satisfies $(g^{\mathrm{op}})^{\mathrm{op}} = g$), we get

$$F(0_{A,B}) = F(z) = F(u^{\mathrm{op}}) = (\mathcal{F}(u))^{\mathrm{op}} = v^{\mathrm{op}} = \zeta = 0_{F(A),F(B)},$$

as desired. The proof of (b) is complete.

(c) Since α is an isomorphism, $\mathcal{P}$ is a product (and hence also a sum) of $F(A)$ and $F(B)$ in $\underline{D}$. Therefore, it follows from Proposition 2.13 (b) and Proposition 2.14 (b) that, in proving (c), we can assume, without loss of generality, that $Q = \mathcal{P}$. Hence, α is the uniquely determined endomorphism of $\mathcal{P}$ in $\underline{D}$ such that $\pi_1\alpha = F(p_1)$ and $\pi_2\alpha = F(p_2)$; and β is the uniquely determined endomorphism of $\mathcal{P}$ in $\underline{D}$ such that $\beta j_1 = F(i_1)$ and $\beta j_2 = F(i_2)$.

We claim that it suffices to prove (under the assumptions that α is an isomorphism and F sends both $0_{A,B}$ and $0_{B,A}$ to zero morphisms) that $\alpha\beta\alpha = \alpha$. Indeed, if $\alpha\beta\alpha = \alpha$, then (since α is an isomorphism),

$$\beta = \alpha^{-1}(\alpha\beta) = \alpha^{-1}(\alpha\alpha^{-1}) = \alpha^{-1}1_{\mathcal{P}} = \alpha^{-1},$$

which is an isomorphism. This proves the above claim.

We will next prove that $\alpha\beta\alpha = \alpha$. This is equivalent to showing that $\pi_1\alpha\beta\alpha = F(p_1)$ and $\pi_2\alpha\beta\alpha = F(p_2)$. It is interesting to observe that the proof of the first (resp., second) of these equations will use the fact that F sends $0_{B,A}$ (resp., $0_{A,B}$) to a zero morphism. We will provide a detailed proof of the first of these equations, leaving the similar proof of the second equation to the reader.

Note that $F(p_1)F(i_1) = F(p_1i_1) = F(1_A) = 1_{F(A)}$ and, similarly, $F(p_2)F(i_2) = 1_{F(B)}$; and $F(p_1)F(i_2) = F(p_1i_2) = F(0_{B,A}) = 0_{F(B),F(A)}$ (by hypothesis) and, similarly, $F(p_2)F(i_1) = F(0_{A,B}) = 0_{F(A),F(B)}$ (by hypothesis). Therefore, using at a crucial point that composition of morphisms distributes over addition in a pre-additive category, we get that

$$\pi_1\alpha\beta\alpha = F(p_1)\beta(1_{\mathcal{P}})\alpha = F(p_1)\beta(j_1\pi_1 + j_2\pi_2)\alpha =$$

$$F(p_1)\beta j_1\pi_1\alpha + F(p_1)\beta j_2\pi_2\alpha.$$

Since $F(p_1)\beta j_1\pi_1\alpha = F(p_1)F(i_1)\pi_1\alpha = F(p_1i_1)\pi_1\alpha = F(1_A)\pi_1\alpha = 1_{F(A)}\pi_1\alpha = \pi_1\alpha = F(p_1)$, we need only show that $F(p_1)\beta j_2\pi_2\alpha = 0_{\mathcal{P},F(A)}$, the neutral element in the additive abelian group $\underline{D}(\mathcal{P}, F(A))$. That, in turn, holds (thanks, in part, to Lemma 2.1 (a)), since

$$F(p_1)\beta j_2\pi_2\alpha = F(p_1)F(i_2)\pi_2\alpha = F(p_1i_2)\pi_2\alpha = F(0_{B,A})\pi_2\alpha =$$

$0_{F(B),F(A)}\pi_2\alpha = 0_{\mathcal{P},F(A)}$. The proof of (c) is complete.

(d) We will explain how (d) follows from (c) in the same spirit of the above proof which explained how (b) follows from (a). As above, let α_F denote the canonical morphism $F(A \times B) \to F(A) \times F(B)$, viewed as $\alpha_F : \mathcal{P} \to Q$, in $\underline{D}$; and similarly, let β_F denote the canonical morphism $F(A) \oplus F(B) \to F(A \oplus B)$, viewed as $\beta_F : Q \to \mathcal{P}$, in $\underline{D}$. Let $\mathcal{F} : \underline{C}^{\mathrm{op}} \to \underline{D}^{\mathrm{op}}$ be the functor induced by F via the construction in Remark 2.15 (b). Let $\alpha_{\mathcal{F}}$ denote the canonical morphism $\mathcal{F}(A \times B) \to \mathcal{F}(A) \times \mathcal{F}(B)$, viewed as $\alpha_{\mathcal{F}} : \mathcal{P} \to Q$, in $\underline{D}^{\mathrm{op}}$; and let $\beta_{\mathcal{F}}$ denote the canonical morphism $\mathcal{F}(A) \oplus \mathcal{F}(B) \to \mathcal{F}(A \oplus B)$, viewed as $\beta_{\mathcal{F}} : Q \to \mathcal{P}$, in $\underline{D}^{\mathrm{op}}$. Recall from Remark 2.15 (c) that $\alpha_F = (\beta_{\mathcal{F}})^{\mathrm{op}}$ and $\beta_F = (\alpha_{\mathcal{F}})^{\mathrm{op}}$.

As we have assumed in (d) that $F(0_{A,B}) = 0_{F(A),F(B)}$ and $F(0_{B,A}) = 0_{F(B),F(A)}$, it follows from the reasoning in the final paragraph of Remark 2.15 (b) that $\mathcal{F}$ sends both the zero morphism $B \to A$ in $\underline{C}^{\mathrm{op}}$ and the zero morphism $A \to B$ in $\underline{C}^{\mathrm{op}}$ to zero morphisms (in $\underline{D}^{\mathrm{op}}$). Moreover, by hypothesis, β_F is an isomorphism; that is, $(\alpha_{\mathcal{F}})^{\mathrm{op}}$ is an isomorphism. Therefore, by reasoning as in the third and fourth paragraphs of the proof of (b), we get that $\alpha_{\mathcal{F}}$ is an isomorphism. Recall also that both $\underline{C}^{\mathrm{op}}$

and $\underline{D}^{\text{op}}$ are pre-additive categories with finite products. Consequently, by applying (c) to $\mathcal{F}$, we get that $\beta_{\mathcal{F}}$ is an isomorphism; that is, $(\alpha_F)^{\text{op}}$ is an isomorphism. Hence, by another appeal to the just-mentioned part of the proof of (b), α_F is an isomorphism. The proof of (d) is complete.

(e) Recall that the biproduct diagram $\mathcal{D}$ in $\underline{C}$ is given by the data set $\{A, B, P, p_1, p_2, i_1, i_2\}$, such that $p_1 i_1 = 1_A$, $p_2 i_2 = 1_B$ and $i_1 p_1 + i_2 p_2 = 1_P$, and, necessarily, $p_1 i_2 = 0_{B,A}$ and $p_2 i_1 = 0_{A,B}$. Assume that F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$; that is, that $F(p_1)F(i_1) = 1_{F(A)}$, $F(p_2)F(i_2) = 1_{F(B)}$ and $F(i_1)F(p_1)+F(i_2)F(p_2) = 1_{F(P)}$ $(=1_{\mathcal{P}})$, and, necessarily, $F(p_1)F(i_2) = 0_{F(B),F(A)}$ and $F(p_2)F(i_1) = 0_{F(A),F(B)})$. Our task is to show that both α and β are isomorphisms.

Recall from the proof of [16, Theorem 2, pages 194-195] that the fact that $\mathcal{D}$ is a biproduct diagram implies that P is a product of A and B (in $\underline{C}$) with projection morphisms $p_1 : P \to A$ and $p_2 : P \to B$. For the sake of completeness, we next show how to modify the just-cited argument in [16] to prove that the fact that $\mathcal{D}$ is a biproduct diagram implies that P is also a sum of A and B (in $\underline{C}$) with injection morphisms $i_1 : A \to P$ and $i_2 : B \to P$.

Consider any morphisms $f_1 : A \to C$ and $f_2 : B \to C$ for some object C of $\underline{C}$. To prove the above “sum" assertion, it suffices to show that there exists a unique morphism $h : P \to C$ such that $hi_1 = f_1$ and $hi_2 = f_2$. As for existence, it suffices to show that $h := f_1 p_1 + f_2 p_2$ satisfies $hi_1 = f_1$ and $hi_2 = f_2$. We will prove the first of these equations, leaving the similar proof of the second equation to the reader. We have

$$hi_1 = (f_1 p_1 + f_2 p_2)i_1 = f_1(p_1 i_1) + f_2(p_2 i_1) = f_1 1_A + f_2 0_{A,B} =$$

$f_1 + 0_{A,C} = f_1$. As for uniqueness, suppose that a morphism $h^* : P \to C$ satisfies $h^* i_1 = f_1$ and $h^* i_2 = f_2$. Then

$$h^* = h^* 1_P = h^*(i_1 p_1 + i_2 p_2) = (h^* i_1)p_1 + (h^* i_2)p_2 = f_1 p_1 + f_2 p_2 = h.$$

This completes the proof of the above “sum" assertion.

Recall that F is assumed to carry $\mathcal{D}$ to a biproduct diagram, say Δ, in $\underline{D}$. Thus, by the reasoning in the preceding two paragraphs, $F(P)$ $(= \mathcal{P})$ is a product of $F(A)$ and $F(B)$ in regard to the projection morphisms $F(p_1) : \mathcal{P} \to F(A)$ and $F(p_2) : \mathcal{P} \to F(B)$, and $\mathcal{P}$ is also a sum of $F(A)$ and $F(B)$ in regard to the injection morphisms $F(i_1) : F(A) \to \mathcal{P}$ and $F(i_2) : F(B) \to \mathcal{P}$. We claim that, in view of Proposition 2.13 (b) and Proposition 2.14 (b), we can assume, without loss of generality, that $Q = \mathcal{P}$. To prove this claim, one must show that if one takes $Q = \mathcal{P}$, along with $\pi_1 = F(p_1)$, $\pi_2 = F(p_2)$, $j_1 = F(i_1)$ and $j_2 = F(i_2)$, then these changes of variables still give a biproduct diagram in $\underline{D}$. In other words, one must show that

$$F(p_1)F(i_1) = 1_{F(A)}, F(p_2)F(i_2) = 1_{F(B)}, F(i_1)F(p_1) + F(i_2)F(p_2) = 1_{\mathcal{P}},$$

(and, necessarily, $F(p_1)F(i_2) = 0_{F(B),F(A)}$ and $F(p_2)F(i_1) = 0_{F(A),F(B)})$. The three just-displayed desired equations are *precisely* what it means to say that Δ is a biproduct diagram in $\underline{D}$. This proves the above claim (that we can take $Q = \mathcal{P}$ along with the above identifications of the associated structural morphisms).

It remains to prove that $\alpha : \mathcal{P} \to \mathcal{P}$ and $\beta : \mathcal{P} \to \mathcal{P}$ are isomorphisms. Recall that α is uniquely determined by the conditions $\pi_1 \alpha = F(p_1)$ and $\pi_2 \alpha = F(p_2)$; and that β is uniquely determined by the conditions $\beta j_1 = F(i_1)$ and $\beta j_2 = F(i_2)$. Hence, in view of the above changes of variables, α is uniquely determined by $F(p_1)\alpha = F(p_1)$ and $F(p_2)\alpha = F(p_2)$; and β is uniquely determined by $\beta F(i_1) = F(i_1)$ and $\beta F(i_2) = F(i_2)$. The uniqueness of those determinations ensures that $\alpha = 1_{\mathcal{P}}$ and $\beta = 1_{\mathcal{P}}$. In particular, both α and β are isomorphisms. The proof of (e) is complete.

(f), (g): Suppose that α is an isomorphism (resp., β is an isomorphism), $F(0_{A,B}) = 0_{F(A),F(B)}$ and $F(0_{B,A}) = 0_{F(B),F(A)}$. Then by (c) (resp, by (d)), β is an isomorphism (resp., α is an isomorphism). Recall that the set of data $\{A, B, P, p_1, p_2, i_1, i_2\}$ gives a uniquely determined biproduct diagram $\mathcal{D}$ in $\underline{C}$ (that is, such that $p_1 i_1 = 1_A$, $p_2 i_2 = 1_B$ and $i_1 p_1 + i_2 p_2 = 1_P$, and, necessarily, $p_1 i_2 = 0_{B,A}$ and $p_2 i_1 = 0_{A,B}$). Our task is to show that F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$; that is, that $F(p_1)F(i_1) = 1_{F(A)}$,

$F(p_2)F(i_2) = 1_{F(B)}$ and $F(i_1)F(p_1) + F(i_2)F(p_2) = 1_{F(P)}$ $(= 1_{\mathcal{P}})$. The first and second of these equations follow easily from the corresponding equations given above since F is a functor. Thus, it remains only to prove that

$$F(i_1)F(p_1) + F(i_2)F(p_2) = 1_{\mathcal{P}}.$$

To that end, let us use the determining conditions $\pi_1\alpha = F(p_1)$, $\pi_2\alpha = F(p_2)$, $\beta j_1 = F(i_1)$ and $\beta j_2 = F(i_2)$. These lead to

$$F(i_1)F(p_1) + F(i_2)F(p_2) = \beta j_1\pi_1\alpha + \beta j_2\pi_2\alpha = \beta(j_1\pi_1 + j_2\pi_2)\alpha =$$

$\beta 1_{\mathcal{P}}\alpha = \beta\alpha$. Hence, we need only prove that $\beta\alpha = 1_{\mathcal{P}}$. Since α is an isomorphism, we need only prove that $\alpha\beta\alpha = \alpha$. That, in turn, can be shown by repeating the final two paragraphs of the proof of (c). The proof is complete. □

Recall (cf. [16, page 196]) that an *additive category* is a pre-additive category with a null object and binary products. (Equivalently, one could define an additive category as a pre-additive category with finite products.) One could summarize Theorem 2.18, which is our second main result, as giving, for additive categories $\underline{C}$ and $\underline{D}$, five new characterizations of the additive functors $\underline{C} \to \underline{D}$.

Theorem 2.18. Let $\underline{C}$ and $\underline{D}$ be (possibly isomorphic) pre-additive categories with finite products, and let $F : \underline{C} \to \underline{D}$ be a functor. Then the following seven conditions are equivalent:

(1) F is an additive functor;

(2) F carries each biproduct diagram (in the sense defined in [16, Definition, page 194]) in $\underline{C}$ to a biproduct diagram in $\underline{D}$;

(3) F is a CHR-additive functor;

(4) If $A, B \in |\underline{C}|$, the canonical morphism $F(A \times B) \to F(A) \times F(B)$ is an isomorphism in $\underline{D}$;

(5) If $A_1, \dots, A_n \in |\underline{C}|$ for some integer $n \geq 2$, then the canonical morphism $F(\prod_{i=1}^n A_i) \to \prod_{i=1}^n F(A_i)$ is an isomorphism in $\underline{D}$;

(6) If $A, B \in |\underline{C}|$, then the canonical morphism $F(A) \oplus F(B) \to F(A \oplus B)$ is an isomorphism in $\underline{D}$;

(7) If $A_1, \dots, A_n \in |\underline{C}|$ for some integer $n \geq 2$, then the canonical morphism $\coprod_{i=1}^n F(A_i) \to F(\coprod_{i=1}^n A_i)$ is an isomorphism in $\underline{D}$.

Proof. We begin by proving the following useful facts: if (4) holds (resp., if (6) holds) and if A and B are (possibly equal) objects of $\underline{C}$, then $F(0_{A,B}) = 0_{F(A),F(B)}$. To see this, note first that $u := 0_{A,B}$ factors through some null object N of $\underline{C}$; that is, there exists a null object N of $\underline{C}$ such that $u = vw$ for some morphisms $w \in \underline{C}(A,N)$ and $v \in \underline{C}(N,B)$. Since $\underline{C}$ is a pre-additive category and N is a null object, we have $w = 0_{A,N}$ and $v = 0_{N,B}$. Moreover, since (4) holds (resp., since (6) holds) and N is a null object, it follows from part (a) (resp., part (b)) of Proposition 2.17 that $F(w) = 0_{F(A),F(N)}$ (resp., that $F(v) = 0_{F(N),F(B)}$). Therefore, by Lemma 2.1 (a), $F(0_{A,B}) = F(u) = F(vw) = F(v)F(w)$ equals $F(v)0_{F(A),F(N)} = 0_{F(A),F(B)}$ (resp., equals $0_{F(N),F(B)}F(w) = 0_{F(A),F(B)}$), as asserted.

(4) ⇔ (6) : By the preceding paragraph, part (c) (resp., part (d)) of Proposition 2.17 gives (4) ⇒ (6) (resp., gives (6) ⇒ (4)).

(4) ⇔ (5) : It is trivial that (5) ⇒ (4). Conversely, the implication (4) ⇒ (5) follows from the associativity, up to natural isomorphism, of nonempty (direct) products. (*That*, in turn, follows from the proof, not the statement, of [16, Proposition 1, page 73].)

(6) ⇔ (7) : It is trivial that (7) ⇒ (6). Conversely, the implication (6) ⇒ (7) can be proved by adapting the above proof that (4) ⇒ (5). That adaptation, which is being left to the reader, proceeds via a straightforward dualization that focuses on coproducts rather than products.

(1) ⇔ (2) : This equivalence was proved in [16, Proposition 4, page 197]. While the statement of [16, Proposition 4, page 197] includes fewer explicit assumptions than the statement of the present Theorem 2.18, an examination of the proof of [16, Proposition 4, page 197] reveals that it uses all the assumptions of our Theorem 2.18.

(2) $\Rightarrow$ (3): Assume (2). Since (2) $\Rightarrow$ (1), it follows from Lemma 2.6 (b) that F sends any terminal object N of (that is, any empty product in) $\underline{C}$ to a null (hence, terminal) object of (hence, an empty product in) $\underline{D}$. Hence, the canonical morphism in $\underline{D}$ from $F(N)$ to an empty product, being the unique morphism between two terminal (actually, null) objects in a pre-additive category, is necessarily an isomorphism. Thus, by the definition of a CHR-additive functor, our task of proving (3) has been reduced to proving (5). As we proved above that (4) $\Rightarrow$ (5), the task of proving (3) can be reduced to proving (4). That, in turn, follows since Proposition 2.17 (e) ensures that (2) $\Rightarrow$ (4) (and, incidentally, also that (2) $\Rightarrow$ (6)).

(3) $\Rightarrow$ (4): This implication follows at once from the definition of a CHR-additive functor.

(4) $\Rightarrow$ (2): It suffices to combine the first paragraph of this proof with Proposition 2.17 (f). The proof is complete. □

The equivalence of conditions (1) and (3) in Theorem 2.18 makes precise a statement of the result that was promised in the penultimate sentence of the Abstract. I trust that it would not be considered immodest or inaccurate for me to add that, because of Example 2.9, one can conclude that the equivalence (1) $\Leftrightarrow$ (3) in Theorem 2.18 gives a strict generalization of the above-mentioned observation of Chase, Harrison and Rosenberg [5] concerning abelian categories.

In regard to the results that were promised in the final sentence of the Abstract: the equivalence of (1), (4), (5), (6) and (7) in Theorem 2.18 provides four additional new characterizations of the additive functors $F : \underline{C} \to \underline{D}$ whenever $\underline{C}$ and $\underline{D}$ are pre-additive categories with finite products. Thus, for such categories, Theorem 2.18 has provided five new characterizations of the associated additive functors. Of course, the equivalence of conditions (1) and (2) in Theorem 2.18 also serves to characterize those additive functors, but as noted in the proof of Theorem 2.18, its equivalence (1) $\Leftrightarrow$ (2) can be found in [16].

One consequence of the equivalence (1) $\Leftrightarrow$ (4) in Theorem 2.18 is that, for the data in Example 2.10 (b), the canonical morphism $\alpha : F_2(A \times B) \to F_2(A) \times F_2(B)$ is not an isomorphism for some zero rings A and B. (This follows since it was shown in Example 2.10 (b) that the functor F_2 is not additive.) This consequence may seem surprising, as those data satisfy $G \times G \cong G$ in Ab and F_2 sends every object of $\underline{C}_R$ to G. So, it may be of interest to have the following short proof that if $A = B = N$ is an arbitrary (necessarily null) object of $\underline{C}_R$ (for the ambient zero ring R), then α is *not* an isomorphism. Let $p_1 : N \times N \to N$ and $p_2 : N \times N \to N$ be the projection morphisms that are pertinent to the product $N \times N$ in $\underline{C}_R$. Then $p_1 = p_2$ since N is a terminal object of $\underline{C}_R$. Then $F_2(p_1) = F_2(p_2)$. Recall that $F_2(N) = G$ in |Ab|. As G is a nontrivial (in fact, infinite) group, one can pick distinct elements $a, b \in G$. Suppose, contrary to the above assertion, that α is an isomorphism (in Ab). As α is then surjective, there exists a (in fact, unique) element $\xi \in F_2(N \times N)$ (= G) such that $\alpha(\xi) = (a, b)$. Let π_1 and π_2 be the projection morphisms $F_2(N) \times F_2(N) \to F_2(N)$ (that is, $G \times G \to G$) that are pertinent to the product $F_2(N) \times F_2(N)$ in Ab. Hence, $\pi_1(a, b) = a$ and $\pi_2(a, b) = b$. By the universal mapping property of this (direct) product, α is determined by the two conditions $\pi_1 \alpha = F_2(p_1)$ and $\pi_2 \alpha = F_2(p_2)$. Therefore,

$$a = \pi_1(a, b) = \pi_1(\alpha(\xi)) = F_2(p_1)(\xi) = F_2(p_2)(\xi) = \pi_2(\alpha(\xi)) =$$

$\pi_2(a, b) = b$, the desired contradiction, thus completing the promised "short proof."

The result in the previous paragraph gives a sense in which we cannot weaken condition (4) in the statement of Theorem 2.18. In particular, the preceding paragraph shows that if $\underline{C}$ and $\underline{D}$ are each pre-additive categories with finite products and $F : \underline{C} \to \underline{D}$ is a functor such that $F(A \times B) \cong F(A) \times F(B)$ for all objects A and B of $\underline{C}$, then it need not be the case that F is an additive functor. Thus, the functor F_2 from Example 2.10 (b) illustrates the importance of requiring that the isomorphisms stipulated in condition (4) of Theorem 2.18 be "canonical" or "natural". The reader is invited to use the $F \leftrightarrow \mathcal{F}$ correspondence to construct an example that makes the analogous point about the isomorphisms stipulated in condition (6) of Theorem 2.18.

This paragraph and the next two paragraphs will discuss some theoretical and/or pedagogical matters that we believe to be of some interest. The multitude of conditions in the statements of Proposition 2.17 and Theorem 2.18 leads naturally to several different reasonable ways to organize the proofs of those results. Exploring those ways can lead to some material for use in a graduate course on category theory. First, in regard to the proof of Proposition 2.17, one could ask whether a proof of its part (g) would be possible (as was the case for the above proofs of its parts (b) and (d)) by using the $F \leftrightarrow \mathcal{F}$ correspondence from Remark 2.15 (b). The answer is in the affirmative, but for reasons of space, we will only sketch the relevant details in the next paragraph.

Let $F : \underline{C} \to \underline{D}$ be as in the setting for Proposition 2.17. Let $\mathcal{F} : \underline{C}^{\text{op}} \to \underline{D}^{\text{op}}$ be the functor induced by F using the construction in Remark 2.15 (b). Let $\mathcal{D}$ be a biproduct diagram in $\underline{C}$, featuring projection morphisms p_1 and p_2 and injection morphisms i_1 and i_2. We have seen in Proposition 2.17 and Theorem 2.18 that it can be fruitful to study whether F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$. One can ask if it would be fruitful to ask the analogous question about $\mathcal{F}$. The answer is in the affirmative, but before stating it, we need to make precise what is meant by the opposite of a biproduct diagram. In short, by definition, $\mathcal{D}^{\text{op}}$ features projection morphisms i_1^{op} and i_2^{op} and injection morphisms p_1^{op} and p_2^{op}. One can show that F carries $\mathcal{D}$ to a biproduct diagram in $\underline{D}$ if and only if $\mathcal{F}$ carries $\mathcal{D}^{\text{op}}$ to a biproduct diagram in $\underline{D}^{\text{op}}$. (Since the correspondence $F \leftrightarrow \mathcal{F}$ is a one-to-one correspondence, it suffices to prove the “only if" part of the preceding statement.) This result can be used to show that part (f) of Proposition 2.17 (when applied to $\mathcal{F}$ instead of F) implies part (g) of Proposition 2.17 (and *vice versa*), thus answering a question that was raised in the preceding paragraph. The calculations involving in proving this result, as one may expect from having worked out the details in the above proofs of parts (b) and (d) of Proposition 2.17, give some worthwhile and instructive experience in dealing with the definitions of composition and addition of morphisms in a dual category, as well as the definition of the action of $\mathcal{F}$ on morphisms. For instance, a proof of the result that we just stated includes verifying that

$$\mathcal{F}(p_1^{\text{op}})\mathcal{F}(i_1^{\text{op}}) + \mathcal{F}(p_2^{\text{op}})\mathcal{F}(i_2^{\text{op}}) = (1_{F(P)})^{\text{op}}$$

and a complete verification of the just-displayed equation involves (at least) five steps.

Next, any search for alternate ways to present a proof of Theorem 2.18 comes down to asking for different ways of stating and organizing the various parts of Proposition 2.17. In that regard, our preliminary work did find direct proofs (for given objects A and B) that led to the conclusion that condition (2) in Theorem 2.18 implies each of conditions (3)-(7) in Theorem 2.18. For a lecture or homework, an instructor could ask a class to find some of those direct proofs and to see how such arguments could be used to create different presentations of Proposition 2.17 and Theorem 2.18. A somewhat harder assignment would be to ask for direct proofs that each of (4)-(7) implies (2). The latter task would be somewhat easier if the assignment allowed students to also assume that F sends both $0_{A,B}$ and $0_{B,A}$ to zero morphisms.

Remark 2.19. Although we have had reason to discuss several “zero-ish” concepts here, I hesitated to title this paper, “Much ado about zero”, for two reasons. First, our work here would likely suffer in comparison with a similarly titled play by Shakespeare. Second (and more seriously), one must admit that there are some mathematical situations where consideration of a zero element or a zero ring would be inconvenient and, ultimately, irrelevant for the study at hand. (We will discuss a family of such situations in the next paragraph.) However, one must also admit that in some other mathematical situations, one cannot ignore zero rings, as they can provide answers to some natural questions (as in this paper’s Corollaries 2.4 and 2.7.) Moreover, in the final paragraph of this remark, we will discuss an anecdote illustrating how objects such as zero rings can be part of some mathematicians’ fundamental views about the basis and nature of mathematics. That anecdote will also serve to explain this paper’s dedication.

Since the turn of the century, there has been considerable interest in, and research on, a variety of graphs that are defined in terms of the structure of a given nonzero (commutative unital) ring R. To a large extent, the history of such research began with a remarkable paper [3] on “colorings" by István Beck that was published in 1986. To some readers, the graphs that resulted from the methodology in [3] were unnecessarily complicated while studying the zero-divisors of R because that methodology required the element $0 \in R$ to be treated in the same way as each nonzero (possibly zero-divisor) element of R. A much more attractive approach to such questions was begun by D. F. Anderson and P. S. Livingston in a paper [1] that was published in 1999. The methodology that was introduced in [1] to study the set of zero-divisors of R produced more tractable graphs than those which would have resulted from [3] because Anderson and Livingston took the set of vertices of the appropriate graph to be the set of *nonzero* zero-divisors of R. By thus not allowing the element $0 \in R$ to be a vertex of the graph, [1] produced a more intelligible graph (with, for a nonzero ring, fewer vertices and typically many fewer edges). Once it had been decided to disallow consideration of the *element* $0 \in R$, it was clearly pointless (pun intended) to consider any zero ring as a possible R. (Indeed, if one deletes that element from a zero ring, one gets the empty set, and no one would seriously suggest that graph theory could/should be used to deeper our understanding of $\emptyset$.) Moreover, I believe that most ring theorists would find no merit in considering the set of zero-divisors of a *zero ring*. In that regard, perhaps Kaplansky said it best [15, Note, page 34]: “It is perhaps treacherous to try to talk about zero-divisors on the zero module".

During my final year in graduate school (1968-69), I was a student in a course on category theory that was taught by Jon Beck (not to be confused with István Beck). The highlight of the course was Professor Beck’s presentation of his famous “tripleability theorem" (cf. [16, pages 151-159]). Much earlier in the course, one of the students interrupted a lecture by asking the following question: “Why do we need to consider zero to be a ring?". (Of course, everyone understood that, by “zero", the student meant “$\{0\}$".) Perhaps Professor Beck knew that the student who had asked the question was doing doctoral research in algebraic geometry (and so was I, with more of an emphasis on “algebraic" and less on “geometry" than the other student). Professor Beck immediately replied, “Because [the category of] Schemes needs an initial object." My initial reaction to that reply was that it must have been intended as a joke, as it could be translated, at least for affine schemes, via duality, as saying that the category of commutative unital rings needs a terminal object. Within moments, I understood more deeply that Professor Beck’s reply had not been meant as a joke. After all, everyone agrees that $\emptyset$ is an initial object in the category of sets, and it is only a small step from there to agree that the empty scheme is an initial object in the category of (not necessarily affine) schemes. Of course, the empty scheme can be realized as Spec(R) where R is any zero ring. Elucidating the (admittedly easy/trivial) sheaf-theoretic details of the structure of that (empty) scheme as a local ringed space took me only a few more moments, and then I was able to resume listening to the lecture. The way in which Professor Beck handed the question has, over the years, given me much food for thought as I considered how to teach advanced graduate courses, because of the following aspects, each of which slowly dawned upon me as time passed. In saying just six words, Professor Beck had managed to do all of the following: he welcomed the question; he answered it in a way that was consistent with the course’s point of view (and, as I learned later, with his personal point of view of mathematics); he treated the audience with respect, seeing them as young professionals by giving an answer that would be clear to some of the students but would possibly require other students to think and study before being able to understand his reply; and he exuded authenticity by using his view that category theory is central to mathematics in order to inform his teaching practices. Many readers will be familiar with the following saying of a 19^{th} century historian and journalist, Henry Adams: “A teacher affects eternity; he can never tell where his influence stops." Whenever I hear or read that saying, I think of three of my teachers. In chronological order, the first of these, who was female, was my History teacher in high school; the second of these directed my masters thesis; and the third of these memorable teachers was Jon Beck. This completes the remark.

We close with some background and a recommendation for one possible direction of future research.

Remark 2.20. For any field k, let $\underline{D}_k$ be the full subcategory of $\underline{C}_k$ whose objects are (isomorphic to) finite products of finite-dimensional separable field extensions of k, and let Ad be the category whose objects are the CHR-additive functors from $\underline{D}_k$ to Ab. See [7, Chapter I, especially Theorem 3.13, pages 29-30, also pages 56-57] for a result showing how $\underline{C}_k$ and $\underline{D}_k$ can be used (along with Cech cohomology in the étale topology for (Spec)(k)) to determine the cohomological dimension of k (that is, the cohomological dimension of the Galois group of the separable closure of k, in the sense of Serre and Tate). For base rings R (that need not be fields), [7, Chapter II, especially pages 86-87] developed tools for use on "R-based topologies" T (which are certain affine-inspired variants of Grothendieck topologies) and the associated notions of a T-additive functor (which is a certain kind of CHR-additive functor), a T-sheaf, and Cech cohomology in T. In Chapter III (resp., Chapter IV) of [7], a specific R-based topology T was introduced that reduced to the classical étale setting from Chapter I if R is a field but also, in case R is a certain kind of one-dimensional valuation domain, produced (by use of the above-mentioned tools, especially Cech cohomology in T-additive functors) a T-cohomological dimension of R that coincides with the classical cohomological dimension of the quotient field (resp., with the classical cohomological dimension of the residue field) of R.

Having thus demonstrated the applicability of the notion of a CHR-additive functor in various settings, we returned, in [8], to the context of a base field k. Let Ad denote the category of CHR-additive functors from $\underline{D}_k$ to Ab. This category is amenable to the classical methods of homological algebra, as it was shown in [8, Corollary 2.3] that Ad is a Grothendieck category with a generator. Moreover, precise categorical descriptions of Ad were obtained via an exact left adjoint functor in [8, Theorem 2.2] and a categorical equivalence in [8, Remark 2.5]. In [8, Section 3], these descriptions of Ad led to examples of behavior of Amitsur cohomology of certain finite-dimensional non-Galois field extensions (for certain associated CHR-additive functors) which were qualitatively different from the behavior of group cohomology, Grothendieck sheaf cohomology, or Cech cohomology of CHR-additive functors in the above-mentioned étale setting. The interpretation in terms of the étale topology of Spec(k) is that, while the direct limit of certain cohomology groups, when indexed by a geometrically interesting set of covers, may exhibit classical behavior, very different behavior can be exhibited by those cohomology groups when one focuses on only one cover which is a singleton set. An affine analogue of that conclusion is the following fact, which has surely been observed by many commutative algebraists. While an integral domain D may be severely restricted when one requires *every* overring of D (inside the quotient field of D) to have a certain property $\mathcal{P}$, there may exist more general integral domains D with some overrings that satisfy $\mathcal{P}$ and other overrings (of D) that do not satisfy $\mathcal{P}$. For instance, if every proper overring of a Noetherian domain D is Noetherian, then D and each of its overrings have Krull dimension at most 1 (cf. [15, Exercise 20, page 64], [11, page 363]), but any Noetherian domain of finite Krull dimension $n \geq 3$ has a proper Noetherian overring of Krull dimension 2: cf. [15, Theorem 85].

Recent decades have witnessed a variety of transfusions connecting commutative algebra with algebraic geometry and cohomology theories. In our opinion, much remains to be learned along these lines, that is, by suitably translating various results from multiplicative ideal theory into the modern language of algebraic geometry, and *vice versa*, and that some of that prospective research should involve CHR-additive functors. This opinion may receive some support from algebraic geometers, since many interesting sheaves are CHR-additive functors, as one can see from the method of proof of a result [6, Proposition 5.2, page 51] in the classical étale setting (cf. also [18, page 707]). In particular, given the intuitive geometric meaning of a "cover", it would seem reasonable in many situations to expect an Ab-valued T-sheaf of some commutative algebras (resp., an Ab-valued sheaf of some affine schemes) to send the 0 algebra (resp., the empty scheme) to the abelian group 0. In that regard and in view of the attention that was paid to null objects and zero morphisms leading up to and during the proof of Theorem 2.18, one should note our long-held interest in ensuring that certain cohomologically relevant functors of commutative algebras send the 0 algebra to the abelian group 0: cf. "$M^*(0) = 0$" in [7, Definition 3.8, page 24]. Note, however, as a closing counterpoint, that a sheaf in an R-based topology need not be a T-additive functor: see [7, page 176, line 4] (where the reference there to "page 33" was intended to be to page 33 of Chapter II, that is, to page 101 of that volume).

References

[1] D. F. Anderson and P. S. Livingston, The zero-divisor graph of a commutative ring, J. Algebra 217 (2) (1999), 434–447.

[2] H. Bass, Lectures on topics in algebraic K-theory, Tata Inst. Fund. Research, Bombay, 1967.

[3] I. Beck, Colorings of commutative rings, J. Algebra 116 (1) (1988), 208-226.

[4] H. Cartan and S. Eilenberg, An introduction to homological algebra, Princeton University Press, Princeton, 1956.

[5] S. U. Chase, D. K. Harrison and A. Rosenberg, Galois theory and Galois cohomology of commutative rings, Mem. Amer. Math. Soc., Volume 52 (1965), 15–33.

[6] D. E. Dobbs, Cech cohomology and a dimension theory for commutative rings, Ph. D. thesis, Cornell University, Ithaca (NY), 1969.

[7] D. E. Dobbs, Cech cohomological dimensions for commutative rings, Lecture Notes in Math., Volume 147, Springer-Verlag, New York, 1970.

[8] D. E. Dobbs, Amitsur cohomology in additive functors, Canad. Math. Bull. 16 (3) (1973), 417–426.

[9] P. Freyd, Abelian categories: an introduction to the theory of functors, Harper & Row, New York, 1964.

[10] P. J. Freyd, Abelian categories: an introduction to the theory of functors, Repr. Theory Appl. Categ. 3, 2003.

[11] R. Gilmer, Multiplicative ideal theory, Dekker, New York, 1972.

[12] K. Gödel, The consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory, Princeton Univ. Press, Princeton, 1940.

[13] A. Grothendieck, Sur quelques points d'algèbre homologique, Tohoku Math. J. 9 (2) (1957), 119–221.

[14] T. W. Hungerford, Algebra, Graduate Texts in Math. **73**, Springer-Verlag, New York-Berlin, 1980.

[15] I. Kaplansky, Commutative rings, revised edition, Univ. of Chicago, Chicago-London, 1974.

[16] S. Mac Lane, Categories for the working mathematician, second edition, Graduate Texts in Math., Volume 5, Springer-Verlag, New York, 1998.

[17] R. A. Morris, The inflation-restriction theorem for Amitsur cohomology, Pacific J. Math. 41 (1972), 791–797.

[18] R. A. Morris, Reciprocity isomorphisms for separable field extensions, Rocky Mountain J. Math. 9 (4) (1979), 707–719.

[19] D. G. Northcott, An introduction to homological algebra, Cambridge University Press, London, 1966.

[20] J. J. Rotman, An introduction to homological algebra, Academic Press, New York, 1979.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco

Volume 2, Issue 1 (2023), pp 70-82

ISSN: 2820-7114

Title :

What can we say about the Pólya Group of a Bicyclic Biquadratic Number Field?

Author(s):

Jean-Luc Chabert

Moroccan Journal of Algebra
and Geometry with Applications
Vol. 2(1) (2023), 70–82
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

What can we say about the Pólya Group of a Bicyclic Biquadratic Number Field?

Jean-Luc Chabert
LAMFA CNRS-UMR 7352, Université de Picardie, 80039 Amiens, France.
e-mail: *jean-luc.chabert@u-picardie.fr*

Communicated by Abdelmalek Azizi
(Received 21 October 2022, Revised 12 January 2023, Accepted 13 January 2023)

Abstract. The Pólya group $\mathcal{P}o(K)$ of a finite Galois extension K of $\mathbb{Q}$ is the subgroup of the class group of K formed by the strong ambiguous classes of K. In this paper, we state a general formula which gives the order of $\mathcal{P}o(K)$ when K is a bicyclic biquadratic number field by means of classical indices, namely, the unit index of K, the number of ramified primes, and the number of fundamental units with norm 1 of the quadratic subfields of K. Then we study separately the imaginary case and the real case.

Key Words: Pólya group, Pólya field, Biquadratic number field, Ambiguous ideal.
2010 MSC: Primary 11R20, 11R29; Secondary 13F20.

1 Introduction

Recall first some definitions, notations, and results.

Definition 1.1. [2, Definition II.3.8] and [25, § 1]

1. The *Pólya-Ostrowski group* or simply the *Pólya group* of an algebraic number field K is the subgroup $\mathcal{P}o(K)$ of the class group $\mathcal{C}l(K)$ of K generated by the classes of the products of all the maximal ideals of the ring of integers $\mathcal{O}_K$ with the same norm.

2. A *Pólya field* is a number field whose Pólya group is trivial.

Notation. We classicaly denote by $\Pi_q(K)$ the product of all the maximal ideals of $\mathcal{O}_K$ with norm q. If q is the norm of no maximal ideal, we set $\Pi_q(K) = \mathcal{O}_K$. Obviously, if $\Pi_q(K) \neq \mathcal{O}_K$, then q is a prime power: $q = p^f$.

The Galois Case. When $K/\mathbb{Q}$ is a Galois extension, that is in the case that interests us here, denoting by e_p and f_p the ramification index and the inertial degree of p in the extension $K/\mathbb{Q}$, we have

$$\Pi_{p^{f_p}}(K) = \prod_{\mathfrak{m}\in\mathrm{Max}(\mathcal{O}_K),\,\mathfrak{m}|p} \mathfrak{m} \quad \text{and} \quad p\mathcal{O}_K = \Pi_{p^{f_p}}(K)^{e_p}. \tag{1}$$

Consequently, following Ostrowski [20], if p in not ramified in the Galois extension $K/\mathbb{Q}$, $\Pi_{p^{f_p}}(K)$ is principal generated by p. Thus, the Pólya group $\mathcal{P}o(K)$ is generated by the classes of the $\Pi_{p^{f_p}}(K)$ where p is ramified in $K/\mathbb{Q}$, so that, $\mathcal{P}o(K)$ is nothing else than the group of strongly ambiguous classes of K.

There are already contributions of several authors who studied the Pólya group of a bicyclic biquadratic number field, for instance: Chattopadhyay and Saikia [6], Heidaryan and Rajaei ([10] and

[11]), Leriche ([16] and [17]), Maarefparvar ([18] and [19]), Taous and Zekhnini ([22], [23] and [26]), and Tougma [24].

We begin in the next section with some preliminaries, in fact some well known facts that will be useful for our study of the bicyclic biquadratic number fields K. Then, in the following section, we establish a formula giving the order of the Pólya group of K by means of several indices and we deduce upper bounds for the number s_K of ramified primes for K to be a Pólya field. In the fourth section, we consider the imaginary case and end, in the last section, with the real case.

2 Preliminaries

General Notation

For every algebraic number field K, we denote by

$\mathcal{O}_K$ the ring of integers of K,
$\mathcal{O}_K^\times$ the units of $\mathcal{O}_K$,
$\mathcal{I}_K$ the group of nonzero fractional ideals of $\mathcal{O}_K$,
$\mathcal{P}_K$ the subgroup of $\mathcal{I}_K$ formed by the principal ideals,
$\mathcal{C}l(K) = \mathcal{I}_K/\mathcal{P}_K$ the class group of K.

Moreover, when $K/\mathbb{Q}$ is a Galois extension, we denote by

$G = \mathrm{Gal}(K/\mathbb{Q})$ the Galois group of $K/\mathbb{Q}$,
$\mathcal{I}_K^G$ the subgroup of $\mathcal{I}_K$ formed by the ambiguous ideals,
$\mathcal{P}_K^G = \mathcal{P}_K \cap \mathcal{I}_K^G$ the subgroup of $\mathcal{P}_K$ formed by the principal ambiguous ideals,
$\mathcal{P}o(K) = \mathcal{I}_K^G/\mathcal{P}_K^G$ the Pólya group of K,
$\mathcal{I}_K^G/\mathcal{P}_\mathbb{Q}$ and $\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q}$ the quotient of $\mathcal{I}_K^G$ and $\mathcal{P}_K^G$ by the extension of $\mathcal{P}_\mathbb{Q}$ in $\mathcal{I}_K^G$,
e_p and f_p the ramification index and the inertial degree of p in $K/\mathbb{Q}$.

Some Exact Sequences

Proposition 2.1. *If $K/\mathbb{Q}$ is a Galois extension with Galois group G, the following sequence of abelian groups is exact*:

$$0 \to \mathcal{P}_K^G/\mathcal{P}_\mathbb{Q} \to \oplus_{p\in\mathbb{P}} \mathbb{Z}/e_p\mathbb{Z} \to \mathcal{P}o(K) \to 0. \tag{2}$$

Proof. The containments $\mathcal{P}_\mathbb{Q} \subseteq \mathcal{P}_K^G \subseteq \mathcal{I}_K^G$ lead to the obvious exact sequence:

$$0 \to \mathcal{P}_K^G/\mathcal{P}_\mathbb{Q} \to \mathcal{I}_K^G/\mathcal{P}_\mathbb{Q} \to \mathcal{I}_K^G/\mathcal{P}_K^G \to 0.$$

We already said that by definition $\mathcal{P}o(K) = \mathcal{I}_K^G/\mathcal{P}_K^G$. It remains to show that $\mathcal{I}_K^G/\mathcal{P}_\mathbb{Q} \simeq \oplus_{p\in\mathbb{P}} \mathbb{Z}/e_p\mathbb{Z}$. The group $\mathcal{I}_K^G$ of ambiguous ideals is the free group generated by the ideals $\Pi_{p^{f_p}}(K)$. Thus we can consider the natural isomorphism

$$\mathfrak{I} = \prod_{p\in\mathbb{P}} (\Pi_{p^{f_p}}(K))^{k_p(\mathfrak{I})} \in \mathcal{I}_K^G \mapsto (k_p(\mathfrak{I}))_{p\in\mathbb{P}} \in \oplus_{p\in\mathbb{P}} \mathbb{Z},$$

which induces a surjective morphism

$$\mathfrak{I} \in \mathcal{I}_K^G \mapsto (\overline{k_p(\mathfrak{I})})_{p\in\mathbb{P}} \in \oplus_{p\in\mathbb{P}} \mathbb{Z}/e_p\mathbb{Z}.$$

The kernel of this last morphism is clearly formed by the ideals $\mathfrak{I}$ such that, for each $p \in \mathbb{P}$, $k_p(\mathfrak{I}) = e_p m_p$ with $m_p \in \mathbb{Z}$, that is, the ideals $\mathfrak{I} = (\prod_p p^{m_p})\mathcal{O}_K$, in other words, the kernel is the image by extension of $\mathcal{P}_\mathbb{Q}$ in $\mathcal{I}_K^G$. □

The exact sequence (2) should be compared with the well known following cohomological exact sequence.

Proposition 2.2. [25, p. 163] *If $K/\mathbb{Q}$ is a Galois extension with Galois group G, the following sequence of abelian groups is exact*:

$$0 \to H^1(G, \mathcal{O}_K^\times) \to \oplus_{p\in\mathbb{P}} \mathbb{Z}/e_p\mathbb{Z} \to \mathcal{P}o(K) \to 0\,. \qquad (3)$$

Sequences (2) and (3) together show that

$$\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q} \simeq H^1(G, \mathcal{O}_K^\times) \qquad \text{(Iwasawa [13])}. \qquad (4)$$

It is not a surprise, indeed, from the short exact sequence

$$1 \to \mathcal{O}_K^\times \to K^* \to \mathcal{P}_K \to 1,$$

the left exactness of the functor $U \mapsto U^G$ on the abelian category of G-modules leads, with Hilbert 90, to:

Proposition 2.3. [1, Lemma 2.1] *If $K/\mathbb{Q}$ is a Galois extension with Galois group G, the following sequence of abelian groups is exact*:

$$1 \to \mathbb{Q}^*/\{\pm 1\} \to \mathcal{P}_K^G \to H^1(G, \mathcal{O}_K^\times) \to 1\,. \qquad (5)$$

Quadratic Fields

The first important result about Pólya groups is due to Hilbert who was interested in the ambiguous ideals.

Proposition 2.4. (Hilbert [12, Theorem 105–106]) *Let $k = \mathbb{Q}(\sqrt{d})$ be a quadratic number field where d is a square-free integer. If s_k denotes the number of ramified prime numbers in the extension $k/\mathbb{Q}$, then*

$$|\mathcal{P}o(k)| = \begin{cases} 2^{s_k-2} & \text{if } k \text{ is real and } N_{k/\mathbb{Q}}(\mathcal{O}_K^\times) = \{1\} \\ 2^{s_k-1} & \text{else.} \end{cases} \qquad (6)$$

Let us recall that the reason of this formula comes from the fact that the group of classes of ambiguous ideals is generated by the classes of the ramified prime ideals of k. But, when $d \neq -1$, the principal ambiguous ideal $\sqrt{d}\mathcal{O}_K$ leads to a relation between the previous generators. Moreover, it is known that there is another relation induced by another principal ambiguous ideal if and only if k is real and the norm of the fundamental unit ε is $+1$. In this case, there are $\alpha \in \mathcal{O}_k$ and $a \in \mathbb{Z}$ such that $\alpha^2 = \varepsilon a$ and $a|2d$.

3 The order of the Pólya group of a biquadratic number field

Let us introduce some more notations with some indices.

Notation. From now on, K denotes a bicyclic biquadratic number field, that is, $K = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2})$ where d_1 and d_2 are two distinct square-free integers. The three quadratic subfields of K are

$$k_1 = \mathbb{Q}(\sqrt{d_1}),\, k_2 = \mathbb{Q}(\sqrt{d_2}) \text{ and } k_3 = \mathbb{Q}(\sqrt{d_3}) \text{ where } d_3 = \frac{d_1 d_2}{(\gcd(d_1, d_2))^2}.$$

- We denote by s_K the number of ramified primes in the extension $K/\mathbb{Q}$ and analogously by s_i ($i = 1, 2, 3$) the number of ramified primes in the extension $k_i/\mathbb{Q}$.
- We set $i_2 = 1$ or 0 depending on whether 2 is, or is not, totally ramified in $K/\mathbb{Q}$.
- If k_i is real, $\varepsilon_i > 1$ denotes the fundamental unit, $\mathcal{O}_{k_i}^\times = \{\pm\varepsilon_i^t \mid t \in \mathbb{Z}\} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$.

• If k_i is imaginary, $\mathcal{O}_{k_i}^\times = \mu_{k_i}$ where μ_{k_i} denotes the group of roots of unity.
• For $1 \le i \le 3$, we let $\nu_i = 1$ or 0 according to the fact that k_i is real and the norm $N_{k_i/\mathbb{Q}}(\varepsilon_i)$ is equal to 1 or not.
• Now we introduce two 'global' indices:

$$\nu_K = \nu_1 + \nu_2 + \nu_3$$

and the unit index

$$q_K = (\mathcal{O}_K^\times : \mathcal{O}_{k_1}^\times \mathcal{O}_{k_2}^\times \mathcal{O}_{k_3}^\times).$$

An odd prime number is ramified in k_i if and only if it divides d_i, thus an odd prime number which is ramified in one of the subfields is ramified in exactly two subfields. The prime 2 is ramified in k_i if and only if $d_i \equiv 2$ or $3 \pmod 4$, so that, if 2 is ramified in one of the subfields, then it is ramified in at least two subfields. Moreover, 2 is totally ramified in K if and only if it is ramified in each subfields k_i. Consequently, we have the relation

$$s_1 + s_2 + s_3 = 2s_K + i_2\,. \tag{7}$$

The prime 2 is not totally ramified in $K/\mathbb{Q}$ if and only if at least one of the three integers d_i is congruent to 1 modulo 4. Equivalently, 2 is totally ramified in $K/\mathbb{Q}$ if and only if two of the integers d_i are even and the third one is congruent to 3 modulo 4.

From the natural morphisms

$$j_{k_i}^K : \mathfrak{a}_i \in \mathcal{I}_{k_i} \mapsto \mathfrak{a}_i \mathcal{O}_K \in \mathcal{I}_K \quad (1 \le i \le 3),$$

we deduce a morphism:

$$(\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3) \in \mathcal{I}_{k_1} \times \mathcal{I}_{k_2} \times \mathcal{I}_{k_3} \mapsto \mathfrak{a}_1 \mathfrak{a}_2 \mathfrak{a}_3 \mathcal{O}_K \in \mathcal{I}_K,$$

which itself induces a morphism:

$$\varphi_K : \mathcal{C}l(k_1) \times \mathcal{C}l(k_2) \times \mathcal{C}l(k_3) \to \mathcal{C}l(K).$$

It is easy to see that the Pólya groups behave well by extensions when all the considered fields are Galois extensions of $\mathbb{Q}$ (see [3, Proposition 3.4]). Thus, φ_K itself induces by restriction a natural morphism:

$$\psi_K : \mathcal{P}o(k_1) \times \mathcal{P}o(k_2) \times \mathcal{P}o(k_3) \to \mathcal{P}o(K).$$

Proposition 3.1. *The natural morphism $\psi_K : \mathcal{P}o(k_1) \times \mathcal{P}o(k_2) \times \mathcal{P}o(k_3) \to \mathcal{P}o(K)$ is surjective either if 2 is not totally ramified or if $\Pi_2(K)$ is principal. Else, the quotient $\mathcal{P}o(K)/\mathrm{Im}(\psi_K)$ has order 2.*

Proof. Let p be a ramified prime number which is odd or equal to 2 if 2 is not totally ramified in K. There is a quadratic subfield k_i of K in which p is ramified, and then, $\Pi_p(k_i)$ is not ramified in the extension K/k_i, else p would be totally ramified in $K/\mathbb{Q}$. Thus, we have $\Pi_p(k_i)\mathcal{O}_K = \Pi_p(K)$ or $\Pi_{p^2}(K)$ according to the fact that $\Pi_p(k_i)$ is decomposed or inert in the extension K/k_i.

Assume now that 2 is totally ramified. Then, whatever the quadratic subfield k_i, we have $\Pi_2(k_i)\mathcal{O}_K = (\Pi_2(K))^2$. Thus, denoting by $\mathrm{Im}(\psi_K)$ the image of ψ_K, we have the isomorphism

$$\mathcal{P}o(K)/\mathrm{Im}(\psi_K) \simeq \langle \overline{\Pi_2(K)} / \overline{\Pi_2(K)^2} \rangle$$

which allows us to conclude. □

Finally, following Kubota, we have:

Proposition 3.2. [14, Satz 4] *With the previous notation*

$$|\mathrm{Ker}(\psi_K)| = \begin{cases} \frac{1}{q_K}\prod_p e_p(K/\mathbb{Q}) & \text{if } K \text{ is real and } \nu_K = 0 \\ \frac{1}{2q_K}\prod_p e_p(K/\mathbb{Q}) & \text{else.} \end{cases} \tag{8}$$

We are now able to state a formula for the order of the Pólya group of K.

Theorem 3.3. The order of the Pólya group $\mathcal{P}o(K)$ of K is equal to

$$|\mathcal{P}o(K)| = \begin{cases} q_K \times 2^{s_K+j_2-2-\max(1,\nu_K)} & \text{if } K \text{ is real} \\ q_K \times 2^{s_K+j_2-2-\nu_K} & \text{if } K \text{ is imaginary} \end{cases} \tag{9}$$

where $j_2 = 1$ if 2 is totally ramified and $\Pi_2(K)$ is not principal and $j_2 = 0$ else.

Proof. By Proposition 2.4,

$$|\mathcal{P}o(k_1)\times\mathcal{P}o(k_2)\times\mathcal{P}o(k_3)| = 2^{\sum_{i=1}^{3}(s_i-1-\nu_i)} = 2^{s_1+s_2+s_3-3-\nu_K}.$$

By Kubota's result,

$$|\mathrm{Im}(\psi_K)| = \frac{|\mathcal{P}o(k_1)\times\mathcal{P}o(k_2)\times\mathcal{P}o(k_3)|}{|\mathrm{Ker}(\psi_K)|} = \frac{2^{s_1+s_2+s_3-3-\nu_K}}{2^{s_K+i_2}}\times q_K\times(1 \text{ or } 2).$$

Proposition 3.1 says that $\frac{|\mathcal{P}o(K)|}{|\mathrm{Im}(\psi_K)|} = 2^{j_2}$. We may conclude with Formula (7). □

Remark 3.4. It is known (Hasse [8]) that

$$q_K = (\mathcal{O}_K^\times : \mathcal{O}_{k_1}^\times\mathcal{O}_{k_2}^\times\mathcal{O}_{k_3}^\times) = \begin{cases} 1,2,4 & \text{if } K \text{ is real} \\ 1,2 & \text{if } K \text{ is imaginary.} \end{cases} \tag{10}$$

We will prove the two cases in the next sections by means of elementary remarks. Let us start here with a rough result: q_K divides 8 in the real case and divides 4 in the imaginary case.

Letting $\mathrm{Gal}(k_i/\mathbb{Q}) = \langle\sigma_i\rangle$, then

$$\forall x\in\mathcal{O}_K^\times \quad x^2N_{K/\mathbb{Q}}(x) = x^3\sigma_1(x)\sigma_2(x)\sigma_3(x)\in\mathcal{O}_{k_1}^\times\mathcal{O}_{k_2}^\times\mathcal{O}_{k_3}^\times.$$

Thus,

$$(\mathcal{O}_K^\times)^{(2)} \subseteq \mathcal{O}_{k_1}^\times\mathcal{O}_{k_2}^\times\mathcal{O}_{k_3}^\times \subseteq \mathcal{O}_K^\times.$$

In the real case, $\mathcal{O}_K^\times \simeq \{\pm1\}\times\mathbb{Z}^3$, and hence, $\mathcal{O}_K^\times/(\mathcal{O}_K^\times)^{(2)} \simeq (\mathbb{Z}/2\mathbb{Z})^4$, as moreover $-1\notin(\mathcal{O}_K^\times)^{(2)}$, $q_K|8$. In the imaginary case, $\mathcal{O}_K^\times \simeq \mu_K\times\mathbb{Z}$, and hence, $\mathcal{O}_K^\times/(\mathcal{O}_K^\times)^{(2)} \simeq (\mathbb{Z}/2\mathbb{Z})^2$, consequently, $q_K|4$.

From Formula (9) we easily deduce an upper bound for the number of ramified primes in a bicyclic biquadratic Pólya field.

Corollary 3.5. *If K is a bicyclic biquadratic Pólya field, then*

$$s_K \leq \begin{cases} 2+\nu_K & \text{if } K \text{ is imaginary} \\ 3+\max\{0,\nu_K-1\} & \text{if } K \text{ is real} \end{cases} \tag{11}$$

In particular, we always have $s_K \leq 5$ and, if K is imaginary, $s_K \leq 3$.

Remark 3.6. Zantema [25, §4] proved that this bound $s_K \leq 5$ is sharp since the field $\mathbb{Q}(\sqrt{5.7},\sqrt{3.127})$ where $2,3,5,7,127$ are ramified is a Pólya field. Note that, by Corollary 4.7 below, the bound $s_K \leq 3$ is sharp in the imaginary case.

Corollary 3.5 gives necessary conditions on s_K for K to be a Pólya field. On the other hand, s_K can provide sufficient conditions for K to be Pólya.

Proposition 3.7. *Let $K = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2})$ and assume that 2 is not totally ramified.*
(i) If $s_K \le 2$, then K is a Pólya field.
(ii) If $d_2 \nmid d_1$ and $\nu_1 = 1$, then $s_K \le 3$ implies that K is a Pólya field.
(iii) If $(d_1, d_2) = 1$ and $\nu_1 = \nu_2 = 1$, then $s_K \le 4$ implies that K is a Pólya field.

Proof. (i) Assume that $s_K \le 2$. Note first that, as $s_i \ge 1$ for each $i = 1,2,3$, Formula (7) leads to $3 \le 2s_K + i_2 \le 4 + i_2$. Under the hypothesis $i_2 = 0$, $s_K = 1$ is impossible. Thus, $s_K = 2$ and, for instance, $s_1 = s_2 = 1$ and $s_3 = 2$. Since, $s_1 = s_2 = 1$ implies that k_1 and k_2 are Pólya fields, we may conclude with Proposition 3.8 below.

(ii) As $\nu_1 = 1$, there are $a_1 \in \mathbb{Z}$ and $\alpha_1 \in k_1$ such that $\alpha_1^2 = \varepsilon_1 a_1$ and $a_1 | 2d_1$. Assume that $\alpha_1 \mathcal{O}_K = \sqrt{d_2}\mathcal{O}_K$, and hence, that $d_2\mathcal{O}_K = a_1\mathcal{O}_K$.

Assume also that a_1 does not divide d_1. Then, a_1 is even, d_1 is odd, and 2 is ramified in $\mathbb{Q}(\sqrt{d_1})$. As $d_2\mathcal{O}_K = a_1\mathcal{O}_K$, 2 divides d_2, and hence, d_3. The prime 2 would be totally ramified. Thus, a_1 divides (strictly) d_1 and the hypothesis $d_2 \nmid d_1$ implies that $a_1 \ne \pm d_2$, and hence, $\alpha_1\mathcal{O}_K \ne \sqrt{d_2}\mathcal{O}_K$.

The group $\mathcal{P}_K^G$ contains obviously the ideals of $\mathcal{O}_K$ generated by $\sqrt{d_1}$, $\sqrt{d_2}$, and α_1. These principal ideals are not congruent modulo $\mathcal{P}_\mathbb{Q}$ because their norms in the extension $K/\mathbb{Q}$, namely d_1^2, d_2^2, and a_1^2, are not congruent modulo z^4 for any $z \in \mathbb{Q}^*$ thanks to the fact that $d_2' = \frac{d_2}{(d_1,d_2)} \ne \pm 1$ is coprime to d_1 and a_1. Thus, 2^3 divides the order of the group $\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q}$. It follows then from the exact sequence (2) that the order of $\mathcal{P}o(K)$ divides 2^{s_K-3}.

(iii) Since $\nu_1 = \nu_2 = 1$, for $i = 1,2$, there are $a_i \in \mathbb{Z}$ and $\alpha_i \in k_i$ such that $\alpha_i^2 = \varepsilon_i a_i$ where $a_i | 2d_i$. It follows from the proof of assertion (ii) that a_i is a strict divisor of d_i. Consequently, the group $\mathcal{P}_K^G$ contains in particular the ideals of $\mathcal{O}_K$ generated by $\sqrt{d_1}$, $\sqrt{d_2}$, α_1, and α_2. These ideals are not congruent modulo $\mathcal{P}_\mathbb{Q}$ because their norms in the extension $K/\mathbb{Q}$, namely d_1^2, d_2^2, a_1^2, and a_2^2, cannot be congruent modulo z^4 for any $z \in \mathbb{Q}^*$. Thus, 2^4 divides the order of the group $\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q}$ and, it follows from the exact sequence (2) that the order of $\mathcal{P}o(K)$ divides 2^{s_K-4}. □

Another way to obtain sufficient conditions for K to be Pólya comes from the idea that if two number fields are Pólya, then the field they generate is likely to be too. This is not completely wrong, nor completely true according to the following proposition.

Proposition 3.8. [16, Proposition 4.3] *The compositum K of two quadratic Pólya fields is a Pólya field if and only the ideal $\Pi_2(K)$ is principal. This is the case in particular if 2 is not totally ramified.*

Proof. Let k_i $(1 \le i \le 3)$ be the three quadratic subfields of K. If p be a prime such that $e_p(K/\mathbb{Q}) = 2$, then p is ramified in exactly two of the three subfields. By hypothesis, at least two subfields k_i are Pólya fields. Clearly, there is a Pólya subfield k_i such that p is ramified in $k_i/\mathbb{Q}$. Thus, on the one hand, p is not ramified in K/k_i and, on the other hand, $\Pi_p(k_i)$ is principal. Consequently, $\Pi_{p^{f_p(K/\mathbb{Q})}}(K) = \Pi_p(k_i)\mathcal{O}_K$ is principal. □

4 Imaginary Bicyclic Biquadratic Number Fields

The biquadratic field K is imaginary, that is, is not real, if one, and only one, of the three integers d_i is > 0. We introduce some specific notation for the imaginary case.

Notation. In this section, $K = \mathbb{Q}(\sqrt{n}, \sqrt{-m})$ where n and m are two square-free positive integers. The real quadratic subfield $\mathbb{Q}(\sqrt{n})$ is denoted by K^+ and its fundamental unit by ε_{K^+}. We let $\nu_K = \nu_{K^+} = 1$ or 0 according to the fact that $N_{K^+/\mathbb{Q}}(\varepsilon_K{}^+) = 1$ or -1.

While the integer n is uniquely determined, we can replace m by $\frac{m\times n}{\gcd(m,n)^2}$. Most often we choose a priori the smaller of the two, but not always.

The Unit Index. If $K \neq \mathbb{Q}(\sqrt{2}, \sqrt{-1}) = \mathbb{Q}(\zeta_8)$, we always have the equality $\mu_K \mathcal{O}_{K^+}^\times = \mathcal{O}_{k_1}^\times \mathcal{O}_{k_2}^\times \mathcal{O}_{k_3}^\times$, and therefore, q_K is equal to Hasse's unit index

$$q_K = (\mathcal{O}_K^\times : \mu_K \mathcal{O}_{K^+}^\times) \qquad (K \neq \mathbb{Q}(\zeta_8)).$$

We already said that $q_K = 1$ or 2 [8, Satz 14]. Indeed, recall the containments

$$(\mathcal{O}_K^\times)^{(2)} \subseteq \mathcal{O}_{k_1}^\times \mathcal{O}_{k_2}^\times \mathcal{O}_{k_3}^\times \subseteq \mathcal{O}_K^\times.$$

If the first inclusion was an equality, $\sqrt{-1} = i$ would be in some k_j, then $\zeta_8 = \sqrt{i} = \frac{1}{2}(1+i)\sqrt{2}$ would be in K and $\sqrt{2}$ would be in some k_l, and finally, $K = \mathbb{Q}(\sqrt{-1}, \sqrt{2}) = \mathbb{Q}(\zeta_8)$. But the fundamental unit $1+\sqrt{2}$ of $\mathbb{Q}(\sqrt{2})$ does not belong to $(\mathcal{O}_{\mathbb{Q}(\zeta_8)}^\times)^{(2)}$. Thus, q_K is a strict divisor of 4.

Following Zantema [25], being a cyclotomic field, $\mathbb{Q}(i, \sqrt{2}) = \mathbb{Q}(\zeta_8)$ is a Pólya field, where 2 is totally ramified. This is a very particular case as shown by the following lemma.

Lemma 4.1. *Let K be an imaginary bicyclic biquadratic number field distinct from $\mathbb{Q}(\zeta_8)$. If 2 is totally ramified in the extension $K/\mathbb{Q}$, the ideal $\mathfrak{P}$ lying over 2 is not principal and K is not a Pólya field.*

Proof. Let $K = \mathbb{Q}(\sqrt{n}, \sqrt{-m})$ and assume that 2 is totally ramified in $K/\mathbb{Q}$. Denote by $\mathfrak{P}$ and $\mathfrak{p}$ the prime ideals lying over 2 of respectively K and $\mathbb{Q}(\sqrt{-m})$. As $K \neq \mathbb{Q}(\sqrt{-1}, \sqrt{2})$, either $m > 2$ or $\frac{mn}{(m,n)^2} > 2$. Thus, we may assume that $m \geq 3$, and then there is no element of $\mathbb{Q}(\sqrt{-m})$ with norm ± 2. Consequently, $\mathfrak{p}$ is not principal and Lemma 4.2 below shows that $\mathfrak{P}$ cannot be principal. As $\Pi_2(K) = \mathfrak{P}$, K is not a Pólya field. □

Lemma 4.2. *Let L be a number field and p be a prime number which is totally ramified in $L/\mathbb{Q}$. Assume that the ideal $\Pi_p(L)$ is principal, and hence, generated by an element y of $\mathcal{O}_L$ such that $N_{L/\mathbb{Q}}(y) = p$ or $-p$. Then, for every subfield K of L, $\Pi_p(K)$ is principal generated by an element $x \in \mathcal{O}_K$ such that $N_{K/\mathbb{Q}}(x) = N_{L/\mathbb{Q}}(y)$ $(= p$ or $-p)$.*

Proof. As p is totally ramified in $L/\mathbb{Q}$, and hence, in $K/\mathbb{Q}$, $\Pi_p(L)$ is the prime of L and $\Pi_p(K)$ the prime of K lying over p. Then we have $N_L^K(\Pi_p(L)) = \Pi_p(K)$. Consequently, if $\Pi_q(L) = y\mathcal{O}_L$ for some $y \in \mathcal{O}_L$, then $\Pi_p(K) = N_{L/K}(y)\mathcal{O}_K$ and $x = N_{L/K}(y)$ satisfies $N_{K/\mathbb{Q}}(x) = N_{L/\mathbb{Q}}(y)$. □

Proposition 4.3. *An imaginary bicyclic biquadratic field K which is the compositum of two quadratic Pólya fields is a Pólya field if and only if, either $K = \mathbb{Q}(\zeta_8)$, or 2 is not totally ramified in $K/\mathbb{Q}$.*

Proof. The necessary condition is an obvious consequence of Lemma 4.1 while Proposition 3.8 implies the sufficient condition. □

Example 4.4. According to Leriche's assertion [16, Proposition 4.3] (our Proposition 3.8), the following fields are Pólya fields.

(a) $K = \mathbb{Q}(\sqrt{2}, \sqrt{-q})$ where $q \equiv 3 \pmod 4) : s_K = 2, i_2 = 0, \nu_{K^+} = 0, q_K = 1$.

(b) $K = \mathbb{Q}(\sqrt{p}, \sqrt{-2})$ where $p \equiv 1 \pmod 4) : s_K = 2, i_2 = 0, \nu_{K^+} = 0, q_K = 1$.

(c) $K = \mathbb{Q}(\sqrt{2q}, \sqrt{-q'})$ where $q \equiv q' \equiv 3 \pmod 4) : s_K = 3, i_2 = 0, \nu_{K^+} = 1, q_K = 1$.

Note that these three examples do not agree with [23, Theorem 3] which describes the Pólya groups of all imaginary bicyclic biquadratic number fields and says that in the three examples $\mathcal{P}o(K) \simeq \mathbb{Z}/2\mathbb{Z}$.

Remark 4.5. There are examples of imaginary biquadratic Pólya fields that are not the compositum of two quadratic Pólya fields: for instance, let p and q be two primes such that $p \equiv 1$ and $q \equiv 3 \pmod 4$, then the field $\mathbb{Q}(\sqrt{q}, \sqrt{-p})$ is a Pólya field (by Proposition 3.7(ii)) which contains only one Pólya quadratic subfield (namely $\mathbb{Q}(\sqrt{q})$) by Formula (6). If moreover $\left(\frac{q}{p}\right) = -1$, then $\mathbb{Q}(\sqrt{-p}, \sqrt{-q})$ is another example [10, Theorem 3.3].

Proposition 4.6. *If $K \neq \mathbb{Q}(\zeta_8)$ is an imaginary bicyclic biquadratic number field, then*

$$|\mathcal{P}o(K)| = q_K 2^{s_K + i_2 - 2 - \nu_{K^+}}. \tag{12}$$

More precisely,

$$\mathcal{P}o(K) \simeq \begin{cases} (\mathbb{Z}/2\mathbb{Z})^{s_K + i_2 - 2 - \nu_{K^+} + \log_2 q_K} & if\ \Pi_2(K)^2\ is\ principal, \\ (\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})^{s_K - 3 - \nu_{K^+} + \log_2 q_K} & else. \end{cases} \tag{13}$$

Note in particular that either when 2 is not totally ramified in $K/\mathbb{Q}$ or when $\Pi_2(k_i)$ is principal for some subfield k_i, then $\Pi_2(K)^2$ is principal.

Proof. When $K \neq \mathbb{Q}(\zeta_8)$, Formula (12) is just Formula (9) where i_2 replaced j_2 thanks to Lemma 4.1. Formula (13) is a consequence of Formula (12) since $\Pi_2(K)$ is the only ambiguous ideal whose class could be of order > 2. □

As a consequence, we obtain Zantema's characterization of the imaginary bicyclic biquadratic number fields that are Pólya fields (which confirms that Examples 4.4 are Pólya fields).

Corollary 4.7. [25, Theorem 4.1] *An imaginary bicyclic biquadratic number field K is a Pólya field if and only if one of the following conditions holds:*

1. *2 is the only ramified prime: $K = \mathbb{Q}(\sqrt{2}, \sqrt{-1}) = \mathbb{Q}(\zeta_8)$,*
2. *there are exactly two primes that are ramified in $K/\mathbb{Q}$ and 2 is not totally ramified,*
3. *there are exactly three primes that are ramified in $K/\mathbb{Q}$, 2 is not totally ramified, $\nu_{K^+} = 1$ and $\mathcal{O}_K^\times = \mu_K \mathcal{O}_{K^+}^\times$.*

Proof. If K is a Pólya field distinct from $\mathbb{Q}(\zeta_8)$, then necessarily $s_K \leq 3$ by Corollary 3.5 and 2 is not totally ramified by Lemma 4.1. If $s_K = 2$, $i_2 = 0$ is sufficient for K to be Pólya by Lemma 3.7 (i). It remains the case where $s_K = 3$ and it follows from Formula (12) that K is Pólya if and only if $q_K = \nu_K = 1$. □

Remark 4.8. Zantema obtained Corollary 4.7 by another way. Indeed, the 4 parameters in Formula (12), namely, s_K, i_2, ν_{K^+}, and q_K, which are needed to compute the order of the Pólya group of K, are also those that are needed to describe the first terms of the exact sequence (3): the first two parameters, s_K and i_2, characterize easily the middle term $\oplus_{p \in \mathbb{P}} \mathbb{Z}/e_p\mathbb{Z}$, while the last two parameters, ν_{k^+} and q_K, characterize the first term $H^1(G, \mathcal{O}_K^\times)$ as shown by the following lemma due to Zantema.

Lemma 4.9. *[25, Lemma 4.3] Let $K \neq \mathbb{Q}(\zeta_8)$ be an imaginary bicyclic biquadratic number field. Then,*

$$H^1(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \mathcal{O}_K^\times) \simeq \begin{cases} (\mathbb{Z}/2\mathbb{Z})^3 & if\ \ \nu_{K^+} \times q_K = 1 \\ (\mathbb{Z}/2\mathbb{Z})^2 & else. \end{cases}$$

Since $\mathcal{P}o(K)$ is trivial if and only if $H^1(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \mathcal{O}_K^\times) \simeq \oplus_{p \in \mathbb{P}} \mathbb{Z}/e_p\mathbb{Z}$, we see once again that 2 cannot be totally ramified if K is a Pólya field (except if $K = \mathbb{Q}(\zeta_8)$). Formulas (12) and (13) describe $\mathcal{P}o(K)$ up to an isomorphism, but when 2 is totally ramified we have to know the order of the class of the prime ideal $\mathfrak{P}$ of K lying over 2. By Lemma 4.1, this order is 2 or 4. Here are some particular results about this order.

Lemma 4.10. *Let $K = \mathbb{Q}(\sqrt{n}, \sqrt{-m})$ be an imaginary bicyclic biquadratic number field such that 2 is totally ramified but distinct from $\mathbb{Q}(\zeta_8)$. Let $\mathfrak{P}$ (resp. $\mathfrak{p}^+$) be the prime ideal of K (resp. of K^+) lying over 2. Then*

1. *If, for some $i \in \{1,2,3\}$, the prime ideal $\mathfrak{P} \cap k_i$ of k_i is principal, then the ideal $\mathfrak{P}^2$ is principal.*
2. *If $m \nmid 2n$, the ideal $\mathfrak{P}^2$ is principal if and only if the ideal $\mathfrak{p}^+$ is principal.*

Recall that $\mathfrak{p}_+$ is principal if and only if, either $2\left(\mathrm{Tr}_{K^+/\mathbb{Q}}(\varepsilon_{K^+}) + 2\right) \in \mathbb{Z}^{(2)}$, or $2\left(\mathrm{Tr}_{K^+/\mathbb{Q}}(\varepsilon_{K^+}) - 2\right) \in \mathbb{Z}^{(2)}$. Note that, in particular, the ideal $\mathfrak{P}^2$ is principal if one of the quadratic subfields k_i is a Pólya field. This is the case for $m = 1$, or 2, or any prime $p \equiv 3 \pmod 4$, or for $n = 2, 3, \ldots$

Proof. (1) If, in one of the three quadratic subfields k_i, the ideal $\mathfrak{p}_i = \mathfrak{P} \cap k_i$ is principal, then $\mathfrak{P}^2$ is principal since $\mathfrak{p}_i \mathcal{O}_K = \mathfrak{P}^2$.
(2) Assume that $m \nmid 2n$ and that $\mathfrak{P}^2$ is principal and let us prove that $\mathfrak{p}^+$ is principal. If $n = 2$ or 3, then $\mathcal{O}_{K^+}$ is a principal ideal domain, in particular $\mathfrak{p}^+$ is principal. Thus, we also assume that $n > 3$. If we can prove that the morphism $\varepsilon_{K^+}^K : \mathcal{C}l(K^+) \to \mathcal{C}l(K)$ is injective, then the fact that $\mathfrak{p}^+\mathcal{O}_K = \mathfrak{P}^2$ is principal will imply that $\mathfrak{p}^+$ itself is principal.

The fact that $m \nmid 2n$ implies obviously that $m \neq 1$, which implies that $|\mu_K| \equiv 2 \pmod 4$, indeed it is easy to check that $|\mu_K| \equiv 0 \pmod 4$ if and only if $m = 1$. Moreover, as m is assumed to be square-free, the fact that $m \nmid 2n$ implies the existence of some odd prime number p dividing m and not n. Such a p is not ramified in $K^+/\mathbb{Q}$ but is ramified in the extension $K^+(\sqrt{-m})/K^+$, which is then 'essentially ramified' in Hasse's sense [8, Chapter III] or in Lemmermeyer' sense [15, § 1]. Finally, the injectivity of $\varepsilon_{K^+}^K$ follows from Lemma 4.11 below. ☐

Lemma 4.11. (Hasse [8, Satz 17] or Lemmermeyer [15, Theorem 1 (i)]) *If $|\mu_K| \equiv 2 \pmod 4$ and the extension K/K^+ is essentially ramified, then $q_K = 1$ and $\varepsilon_{K^+}^K : \mathcal{C}l(K^+) \to \mathcal{C}l(K)$ is injective.*

Noticing that the hypotheses of Lemma 4.11 are satisfied when $m \nmid 2n$, we are able to describe the group $\mathcal{P}o(K)$ in this case.

Proposition 4.12. [5, Proposition V.30] *Let $K = \mathbb{Q}(\sqrt{n}, \sqrt{-m})$ be an imaginary bicyclic biquadratic number field where 2 is totally ramified. Assume that $m \nmid 2n$. Then $\mathcal{P}o(K)$ is isomorphic to*

1. *$(\mathbb{Z}/2\mathbb{Z})^{s_K - 1}$ if $n = 2$,*
2. *$(\mathbb{Z}/2\mathbb{Z})^{s_K - 3} \times (\mathbb{Z}/4\mathbb{Z})$ if $n \neq 2$ and $\nu_{K^+} = 0$,*
3. *$(\mathbb{Z}/2\mathbb{Z})^{s_K - 4} \times (\mathbb{Z}/4\mathbb{Z})$ if $\nu_{K^+} = 1$ and $\mathcal{O}_{K^+}$ has no element with norm ± 2,*
4. *$(\mathbb{Z}/2\mathbb{Z})^{s_K - 2}$ if $n \neq 2$ and $\mathcal{O}_{K^+}$ has an element with norm ± 2.*

Proof. By Lemma 4.11, $q_K = 1$ and, by hypothesis, $i_2 = 1$. Thus, $|\mathcal{P}o(K)| = 2^{s_K - 1 - \nu_{K^+}}$. Let $\mathfrak{P}$ (resp. $\mathfrak{p}^+$) be the prime of K (resp. of K^+) lying over 2. By Lemma 4.10, the class of $\mathfrak{P}$ is of order 2 or 4 depending on whether $\mathfrak{p}^+$ is principal or not. In order to use Lemma 4.9 we just have to know whether $\nu_{K^+} = 1$ or 0. Note also that, if $\mathfrak{p}^+$ is principal, then $\nu_{K^+} = 1$ or $n = 2$. ☐

Propositions 4.6 and 4.12 characterize $\mathcal{P}o(\mathbb{Q}(\sqrt{n}, \sqrt{-m}))$ up to an isomorphism. One finds descriptions of $\mathcal{P}o(K)$ for imaginary bicyclic biquadratic number fields based on Corollary 4.7 due to Zantema [25] when the field is of the form $\mathbb{Q}(\sqrt{-1}, \sqrt{n})$ in [22], or of the form $\mathbb{Q}(\sqrt{-2}, \sqrt{n})$ in [26] (note that Formula (13) gives also such descriptions since the class number of the fields $\mathbb{Q}(\sqrt{-1})$ and $\mathbb{Q}(\sqrt{-2})$ is one). We already spoke about [23] and its Theorem 3.

5 Real Bicyclic Biquadratic Number Fields

In the real case, the three square-free integers are positive. Let us recall Formula (9) in the real case:

$$|\mathcal{P}o(K)| = q_K \times 2^{s_K + j_2 - 2 - \max(1,\nu_K)}. \tag{14}$$

Luckily, our formula agrees with the different formulas provided by [19, Theorem 3.3]. About the unit index q_K, we proved in Remark 3.4 that, in the real case, $q_K | 8$, but said that in fact $q_K | 4$ (Formula (10)). Let us see that $q_K \neq 8$ by means of the simple arguments that Jacques Boulanger suggested to me.

Proposition 5.1. *If $\sqrt{\varepsilon_1} \in K$, then either $d_2 d_3 = d_1$ or $d_2 d_3 = 4d_1$.*

Proof. Since $\sqrt{\varepsilon_1} \in K$, for every $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$, $\sigma(\varepsilon_1) = (\sigma(\sqrt{\varepsilon_1}))^2 > 0$, in other words, ε_1 is totally positive, in particular $N_{k_1/\mathbb{Q}}(\varepsilon_1) = 1$. Let $\mathrm{Gal}(k_1/\mathbb{Q}) = \langle \sigma_1 \rangle$ and to simplify write tr_1 instead of $Tr_{k_1/\mathbb{Q}}$. Then for $\delta \in \{\pm 1\}$, one see easily that

$$(\varepsilon_1 + \delta)(\sigma_1(\varepsilon_1) + \delta) = tr_1(\varepsilon_1) + 2\delta \ \ and \ \ \varepsilon_1(\varepsilon_1 + \delta)(\sigma_1(\varepsilon_1) + \delta) = (\varepsilon_1 + \delta)^2.$$

Consequently,

$$\sqrt{\varepsilon_1} = \frac{\varepsilon_1 + \delta}{\sqrt{tr_1(\varepsilon_1) + 2\delta}}$$

If $\sqrt{tr_1(\varepsilon_1) + 2\delta}$ was a square in $\mathbb{N}$, then $\sqrt{\varepsilon_1}$ would be in $\mathcal{O}_{k_1}$, but this is impossible since ε_1 is a fundamental unit in k_1, thus $\sqrt{\varepsilon_1}$ is of degree 4. Consequently, the integers $tr_1(\varepsilon_1) + 2\delta$ are not square in $\mathbb{N}$ and the fields $\mathbb{Q}(\sqrt{tr(\varepsilon_1) + 2\delta})$ are quadratic subfields of K. Both subfields are distinct from k_1 because, if one of them was equal to k_1, it would contain $\sqrt{\varepsilon_1}$. They can not be equal to each other since else they would be a quadratic field containing $\sqrt{tr_1(\varepsilon_1)^2 - 4}$, and then equal to k_1. Indeed, if $\varepsilon_1 = a + b\sqrt{d_1}$, then $\sqrt{tr_1(\varepsilon_1)^2 - 4} = b\sqrt{d_1}$ since $1 = N_{k_1/\mathbb{Q}}(\varepsilon_1) = a^2 - d_1 b^2$.

Thus, for instance, $\mathbb{Q}(\sqrt{tr(\varepsilon_1) + 2}) = \mathbb{Q}(\sqrt{d_2})$ and $\mathbb{Q}(\sqrt{tr(\varepsilon_1) - 2}) = \mathbb{Q}(\sqrt{d_3})$. Consequently, $\gcd(d_2, d_3) \mid \gcd(tr_1(\varepsilon_1) + 2, tr_1(\varepsilon_1) - 2) \mid 4$. As $\gcd(d_2, d_3)$ is square-free, it is equal to 1 or 2 which means that $d_2 d_3 = d_1$ or $d_2 d_3 = 4d_1$. □

Corollary 5.2. *If $\sqrt{\varepsilon_1}$ and $\sqrt{\varepsilon_2} \in K$, then $d_3 = 2$.*

Proof. By Proposition 5.1, $d_2 d_3 = 4^u d_1$ and $d_1 d_3 = 4^v d_2$ where $u, v \in \{0, 1\}$. Consequently, $d_3 = 2^{u+v}$, and hence, $d_3 = 2$. □

Corollary 5.3. *The unit index q_K divides 4.*

Proof. The containments

$$(\mathcal{O}_K^\times)^{(2)} \subsetneq \{\pm 1\} \times (\mathcal{O}_K^\times)^{(2)} \subseteq \mathcal{O}_{k_1}^\times \mathcal{O}_{k_2}^\times \mathcal{O}_{k_3}^\times \subseteq \mathcal{O}_K^\times$$

together with $|\mathcal{O}_K^\times / (\mathcal{O}_K^\times)^{(2)}| = 16$ show that, the assumption $q_K = 8$ implies that $\{\pm 1\} \times (\mathcal{O}_K^\times)^{(2)} = \mathcal{O}_{k_1}^\times \mathcal{O}_{k_2}^\times \mathcal{O}_{k_3}^\times$, and hence, that $\sqrt{\varepsilon_i} \in K$ for $i = 1, 2, 3$. By Corollary 5.2, this would imply $d_1 = d_2 = d_3 = \sqrt{2}$, this is a contradiction. □

In view of a partial converse of Lemma 4.2, we recall Setzer's following result.

Proposition 5.4. [21, Theorem 4] *Let K be a real bicyclic biquadratic number field. Let $H = H^1(G, \mathcal{O}_K^\times)$ and denote by $H[2]$ the subgroup of H formed by the elements of order ≤ 2. Then, $(H : H[2]) = 1$ or 2. It is 2 if and only if 2 is totally ramified in $K/\mathbb{Q}$ and there exist integers $x_i \in k_i$ such that*

$$N_{k_1/\mathbb{Q}}(x_1) = N_{k_2/\mathbb{Q}}(x_2) = N_{k_3/\mathbb{Q}}(x_3) = +2 \ or \ -2.$$

Corollary 5.5. *Let K be a real bicyclic biquadratic number field. If 2 is totally ramified in K, the following assertions are equivalent:*

1. *the ideal $\Pi_2(K)$ is principal,*
2. *there exist integers x_i in each quadratic subfields k_i such that*

$$N_{k_1/\mathbb{Q}}(x_1) = N_{k_2/\mathbb{Q}}(x_2) = N_{k_3/\mathbb{Q}}(x_3) = +2 \text{ or } -2. \tag{15}$$

Proof. We already know that (1) implies (2). Assume that (2) holds. By Proposition 5.4, $H^1(G, \mathcal{O}_K^\times) \neq H^1(G, \mathcal{O}_K^\times)[2]$. Now consider $\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q}$ and its subgroup $(\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q})[2]$ formed by the elements of order ≤ 2. By Iwasawa's isomorphism (Formula (7)), $\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q} \simeq H^1(G, \mathcal{O}_K^\times)$, and hence, $(\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q})[2] \simeq H^1(G, \mathcal{O}_K^\times)[2]$. Thus, $(\mathcal{P}_K^G/\mathcal{P}_\mathbb{Q})[2] \neq \mathcal{P}_K^G/\mathcal{P}_\mathbb{Q}$ which means that 2 is totally ramified and that the ideal $\Pi_2(K)$ is principal. □

As a consequence of Corollary 5.5 and Proposition 3.8, we have:

Corollary 5.6. *Let K be the compositum of two real quadratic Pólya fields. The following assertions are equivalent:*

1. *K is a Pólya field,*
2. *the ideal $\Pi_2(K)$ is principal,*
3. *either 2 is not totally ramified, or there exist integers x_i in each quadratic subfields k_i such that $N_{k_1/\mathbb{Q}}(x_1) = N_{k_2/\mathbb{Q}}(x_2) = N_{k_3/\mathbb{Q}}(x_3) = +2$ or -2.*

The study of the composita of two real quadratic Pólya fields is undertaken by Leriche in [16] and [17] and more precise results are then given by Tougma [24].

Remark 5.7. There are Pólya bicyclic biquadratic number fields which are not obtained as such a compositum: the number of ramified primes in the compositum of two Pólya real quadratic number fields is bounded by 4, while, following Zantema [25], there exist bicyclic biquadratic Pólya fields where 5 primes are ramified, namely the number field $\mathbb{Q}(\sqrt{5.7}, \sqrt{3.127})$ which contains only one quadratic Pólya subfield. Maarefparvar [18, Theorem 4.4] generalizes Zantema's example by providing a family of biquadratic Pólya fields with five ramified primes and only one quadratic Pólya subfield. There are other counterexamples with less ramified primes: $\mathbb{Q}(\sqrt{3}, \sqrt{35})$ is a Pólya field with 4 ramified primes and only one quadradic Pólya subfield [11, Theorem C]; $\mathbb{Q}(\sqrt{7}, \sqrt{10})$ is a Pólya field with 3 ramified primes and only one quadradic Pólya subfield. But, following Maarefparvar [18, Theorem 4.1], there are examples of biquadratic Pólya fields with no Pólya quadratic subfield.

References

[1] A. Brumer and M. Rosen, Class number and ramification in number fields, *Nagoya Math. J.* **23** (1963), 97–101.

[2] P.-J. Cahen and J.-L. Chabert, Integer-Valued Polynomials, *Amer. Math. Soc. Surveys and Monographs*, **48**, Povidence, R.I., 1997.

[3] J.-L. Chabert, Factorial groups and Polya groups in Galoisian extensions of $\mathbb{Q}$, in *Commutative Ring Theory and Applications*, Lecture Notes in Pure and Applied Mathematics, vol. **231**, pp. 77-86, Marcel Dekker, New York, 2003.

[4] J.-L. Chabert, From Pólya fields to Pólya groups (I), Galois number fields, *J. Number Theory* **203** (2019), 360–375.

[5] J.-L. Chabert, Pólya Groups of Number Fields of Small Degrees in *Integer-Valued Polynomials* (II), to appear.

[6] J. Chattopadhyay and A. Saikia, Totally real bi-quadratic fields with large Pólya groups, *Res. Number Theory* **8** (2022), n^o 2, Paper 30.

[7] P.E. Conner and J. Hurrelbrink, *Class Number Parity*, vol. 8, World Scientific Series in Pure Math., 1988.

[8] H. Hasse, *Über die Klassenzahl abelscher Zahlkörper*, Berlin, 1952.

[9] Y. Hayashi, Reelle biquadratische Zahlkörper mit ungerader Klassenzahl, *Manuscripta math.* 62 (1988), 65–82.

[10] B. Heidaryan, A. Rajaei, Biquadratic Pólya fields with only one quadratic Pólya subfield, *J. Number Theory* **143** (2014), 279–285.

[11] B. Heidaryan, A. Rajaei, Some non-Pólya biquadratic fields with low ramification, *Rev. Mat. Iberoam* **33** (2017), 1037–1044.

[12] D. Hilbert, *Die Theorie der algebaischen Zahlkörper* (1897), in Jahresbericht der Deutschen Mathematiker-Vereinigung **4** (1894–95), 175–546.

[13] K. Iwasawa, A note on the group of units of an algebraic number field, *J. Math. Pures Appl.* **35** (1956), 189–192.

[14] T. Kubota, Über den bizyklischen biquadratischen Zahlkörper, *Nagoya Math. J.* **10** (1956), 65–85.

[15] Fr. Lemmermeyer, Ideal class groups of cyclotomic fields I, *Acta Arithmetica* **72** (1995), 347–359.

[16] A. Leriche, Pólya fields, Pólya groups and Pólya extensions: A question of capitulation, *J. Theor. Nombres Bordeaux* **23** (2011), 235–249.

[17] A. Leriche, Cubic, quartic and sextic Pólya fields, *J. Number Theory* **133** (2013), 59–71.

[18] A. Maarefparvar, Pólya group in some real biquadratic fields, *J. Number Theory* **228** (2021), 1-7.

[19] A. Maarefparvar, Pólya-Ostrowski group and unit index in real biquadratic fields, ArXiv:2108.01904v2.

[20] A. Ostrowski, Über ganzwertige Polynome in algebraischen Zahlkörpern, *J. reine angew. Math.* **149** (1919), 117–124.

[21] C.B. Setzer, Units over totally real $C_2 \times C_2$ fields, *J. Number Theory* **12** (1980), 160–175.

[22] M. Taous, On the Pólya group of some imaginary biquadratic fields, in *Non-associative and non-commutative algebra and operator theory*, 175–182, Springer, 2016.

[23] M. Taous and A. Zekhnini, Pólya groups of the imaginary bicyclic biquadratic number fields, *J. Number Theory* **177** (2017), 307–327.

[24] C. Tougma, Some questions on biquadratic Pólya fields, *J. Number Theory*, **229** (2021), 386–398.

[25] H. Zantema, Integer valued polynomials over a number field, *Manuscripta Math.* **40** (1982), 155–203.

[26] A. Zekhnini, Imaginary biquadratic Pólya fields of the form $\mathbb{Q}(\sqrt{d}, \sqrt{-2})$, *Gulf. J. Math.* **4** (2016), 182–188.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 83-92
ISSN: 2820-7114

Title :

Note on bi-Amalgamated modules along ideals

Author(s):

Adam Anebri

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 83–92
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco **ISSN: 2820-7114**

Note on bi-Amalgamated modules along ideals

Adam Anebri
Laboratory of Modelling and Mathematical Structures Department of Mathematics,
Faculty of Science and Technology of Fez, Box 2202, University S.M. Ben Abdellah Fez, Morocco.
e-mail: *adam.anebri@usmba.ac.ma*

Communicated by El Hassan El Kinani
(Received 20 November 2022, Revised 10 January 2023, Accepted 20 January 2023)

Abstract. Let $f : A \to B$ and $g : A \to C$ be two commutative ring homomorphisms and let J and J' be two ideals of B and C, respectively, such that $f^{-1}(J) = g^{-1}(J')$. The bi-amalgamation of A with (B,C) along (J,J') with respect to (f,g), denoted by $A \bowtie^{f,g} (J,J')$, is the subring of $B \times C$ given by $A \bowtie^{f,g} (J,J') := \left\{(f(a)+j, g(a)+j') \mid a \in A, (j,j') \in J \times J'\right\}$. In this paper, we study some basic properties of a special kind of $A \bowtie^{f,g} (J,J')$-modules, called the bi-amalgamation of M with (N,P) along (J,J') with respect (φ,ψ), and defined by $M \bowtie^{\varphi,\psi} (JN,J'P) := \{(\varphi(m)+n, \psi(m)+p) \mid m \in M \text{ and } (n,p) \in JN \times J'P\}$. The new results generalize some known results on the bi-amalgamation of rings and the amalgamation of modules along an ideal.
Key Words: Bi-amalgamation, amalgamation, Noetherian module, prime module, reduced module, coherent module.
2010 MSC: 13E05, 13D05, 13D02, 13C60.

Dedicated to the memory of Professor Muhammad Zafrullah

1 Introduction

Throughout this paper, all rings are assumed to be commutative with nonzero identity and all modules are nonzero unital. Let $f : A \to B$ and $g : A \to C$ be two ring homomorphisms and let J and J' be two ideals of B and C, respectively, such that $f^{-1}(J) = g^{-1}(J')$. Kabbaj, Louartiti and Tamekkante in [16] introduced and studied the subring

$$A \bowtie^{f,g} (J,J') := \left\{(f(a)+j, g(a)+j') \mid a \in A, (j,j') \in J \times J'\right\}$$

of $B \times C$ called the bi-amalgamation of A with (B,C) along (J,J') with respect to (f,g). The amalgamated algebra (introduced and studied in [5]) can be regarded as a bi-amalgamation (see [16, Example 2.1]). Let A and B be two rings, J be an ideal of B and $f : A \to B$ be a ring homomorphism. The amalgamation of A and B along J with respect to f was defined by $A \bowtie^f J := \{a, f(a)+j) \mid a \in A \text{ and } j \in J\}$. The basic properties of amalgamated algebra are summarized in [6, 7, 10, 15].
Also, the fact that the bi-amalgamation may be investigated in the context of pullback constructions is one of the important tools for understanding $A \bowtie^{f,g} (J,J')$. This viewpoint enables the authors in [16] to provide a thorough description of many properties of $A \bowtie^{f,g} (J,J')$ in connection with the properties of A, J, J', f and g. More precisely, in [16], the authors studied the basic properties of this construction (e.g., characterizations for $A \bowtie^{f,g} (J,J')$ to be a Noetherian ring, an integral domain, a reduced ring and a local ring) and they characterized those distinguished pullbacks that can be expressed as a bi-amalgamation. This construction have been studied by several authors, see for instance [2, 3, 11, 12, 14, 16, 19, 20].
Let $f : A \to B$ be a ring homomorphism, J be an ideal of B, M be an A-module, N be a B-module (which is an A-module induced naturally by f) and $\varphi : M \to N$ be an A-module homomorphism. The authors of [9] recently introduced the amalgamation of M and N along J with respect to φ

denoted by $M \bowtie^{\varphi} JN$ and gave the following definition

$$M \bowtie^{\varphi} JN := \{(m, \varphi(m)+n) \mid m \in M \text{ and } n \in JN\}$$

which is an $A \bowtie^{f} J$-module with the multiplication given by

$$(a, f(a)+j)(m, \varphi(m)+n) := (am, \varphi(am)+f(a)n+j\varphi(m)+jn),$$

where $a \in A, j \in J$ and $(m, \varphi(m)+n) \in M \bowtie^{\varphi} JN$. In their paper, they studied some basic properties of the amalgamation of A-modules along an ideal. More precisely, they studied when $M \bowtie^{f} J$ is a Noetherian or a coherent $A \bowtie^{f} J$-module.
In this paper, we introduce the bi-amalgamation of modules along ideals. Let $f : A \to B$ and $g : A \to C$ be two ring homomorphisms and let J and J' be two ideals of B and C, respectively, such that $f^{-1}(J) = g^{-1}(J')$. Let M be an A-module, N be a B-module (which is an A-module induced naturally by f), P be a C-module (which is an A-module induced naturally by g), and $\varphi : M \to N$ and $\psi : M \to P$ be two A-module homomrphisms such that $\varphi^{-1}(JN) = \psi^{-1}(J'P)$. We define the bi-amalgamation of M with (N, P) along (J, J') with respect (φ, ψ) by

$$M \bowtie^{\varphi,\psi} (JN, J'P) := \{(\varphi(m)+n, \psi(m)+p) \mid m \in M \text{ and } (n, p) \in JN \times J'P\}.$$

One can see that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is an $A \bowtie^{f,g} (J, J')$-module by the following multiplication

$$(f(a)+j, g(a)+j')(\varphi(m)+n, \psi(m)+p) := (\varphi(am)+f(a)n+j\varphi(m)+jn, \psi(am)+g(a)p+j'\psi(m)+j'p).$$

Note that $\varphi(am) = f(a)\varphi(m)$ and $\psi(am) = g(a)\psi(m)$ since φ and ψ are A-module homomorphisms. One can see that if $M = A$, $N = B$, $P = C$, $\varphi = f$ and $\psi = g$ then the bi-amalgamation of the A-module M with (N, P) along (J, J') with respect (φ, ψ) coincides with the bi-amalgamation of the ring A with (B, C) along (J, J') with respect to (f, g). Also, the amalgamation of M and N along J with respect to φ can be viewed as the bi-amalgamation of M with (M, N) along $(f^{-1}(J), J)$ with respect (id_M, φ) if $\varphi^{-1}(JN) \subseteq f^{-1}(J)M$, that is, $M \bowtie^{\varphi} JN = M \bowtie^{id_M,\varphi} (f^{-1}(J)M, JN)$. In particular, the duplication of the A-module M along an ideal I is a bi-amalgamation of modules.
In this work, we study some basic properties of the bi-amalgamation of modules. Namely, we study when $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a Noetherian, a prime, a reduced or a coherent $A \bowtie^{f,g} (J, J')$-module. Our results generalize some known results on the bi-amalgamation of rings and the amalgamation of modules along an ideal.

2 Some basic properties of $M \bowtie^{\varphi,\psi} (JN, J'P)$

Proposition 2.1. *Consider the A-module homomorphisms $\alpha : \varphi(M) + JN \to M/F_0$, $\varphi(m) + n \mapsto \bar{m}$ and $\beta : \psi(M) + J'P \to M/F_0$, $\psi(m) + p \mapsto \bar{m}$. Then, the bi-amalgamation $M \bowtie^{\varphi,\psi} (JN, J'P)$ is determined by the following pullback*

$$\begin{array}{ccc} M \bowtie^{\varphi,\psi} (JN, J'P) & \twoheadrightarrow & \varphi(M) + JN \\ \downarrow & & \downarrow \alpha \\ \psi(M) + J'P & \xrightarrow{\beta} & M/F_0 \end{array}$$

that is

$$M \bowtie^{\varphi,\psi} (JN, J'P) = \alpha \times_{\frac{M}{F_0}} \beta.$$

Proof. It is clear. □

Proposition 2.2. *Let F be an A-submodule of M. Then the following assertions hold:*

(1) $\dfrac{M \bowtie^{\varphi,\psi} (JN,J'P)}{F \bowtie^{\varphi,\psi} (JN,J'P)} \cong \dfrac{M}{F+F_0}$.

(2) $\dfrac{M \bowtie^{\varphi,\psi} (JN,J'P)}{JN \times J'P} \cong \dfrac{M}{F_0}$.

(3) $\dfrac{M \bowtie^{\varphi,\psi} (JN,J'P)}{0 \times J'P} \cong \varphi(M)+JN$ *and* $\dfrac{M \bowtie^{\varphi,\psi} (JN,J'P)}{JN \times 0} \cong \psi(M)+J'P$.

Proof. (1) Consider the mapping

$$\begin{array}{rcccl} \phi & : & M & \rightarrow & \dfrac{M \bowtie^{\varphi,\psi} (JN,J'P)}{F \bowtie^{\varphi,\psi} (JN,J'P)} \\ & & m & \mapsto & \overline{(\varphi(m),\psi(m))}. \end{array}$$

It can be seen that ϕ is an epimorphism of A-modules and $\ker(\phi) = F + F_0$.
(2) is a particular case of (1) for $F = 0$.
(3) Consider the canonical epimorphism of $(f(A)+J)$)-modules $\pi_N : M \bowtie^{\varphi,\psi} (JN,J'P) \twoheadrightarrow \varphi(M)+JN$. If $\varphi(m)+n = 0$ for some $m \in M$ and $n \in JN$, then $\psi(m)+p \in J'P$ for each $p \in J'P$. It follows that the kernel of π_N coincides with $0\times J'P$. Hence, we have the desired isomorphism. The second one follows similarly. □

We will now see when the bi-amalgamation of modules along ideals is a Noetherian module.

Theorem 2.3. The $A \bowtie^{f,g} (J,J')$-module $M \bowtie^{\varphi,\psi} (JN,J'P)$ is Noetherian if and only if $\varphi(M)+JN$ is a Noetherian $f(A)+J$-module and $\psi(M)+J'P$ is a Noetherian $g(A)+J'$-module.

Proof. If $M \bowtie^{\varphi,\psi} (JN,J'P)$ is a Noetherian $A \bowtie^{f,g} (J,J')$-module, then $M \bowtie^{\varphi,\psi} (JN,J'P)$ is a Noetherian $f(A)+J$-module by [4, Exercise 3.1.5], and hence $\varphi(M)+JN$ is a Noetherian $f(A)+J$-module. Likewise for $\psi(M)+J'P$. Conversely, since $\psi(M)+J'P$ is a Noetherian $g(A)+J'$-module, then $J'P$ is a Noetherian $g(A)+J'$-module, which implies that $M \bowtie^{\varphi,\psi} (JN,J'P)$ is a Noetherian $A \bowtie^{f,g} (J,J')$-module, as required. □

Corollary 2.4. *[16, Proposition 4.2] Under the above notations, the ring $A \bowtie^{f,g} (J,J')$ is Noetherian if and only if $f(A)+J$ and $g(A)+J'$ are Noetherian rings.*

A proper A-submodule N of an A-module M is said to be a prime submodule if for each $a \in A$ the trivial multiplication by a, $M/N \to M/N$ is either injective or zero. This implies that $ann_A(M/N) = P$ is a prime ideal of A, and N is said to be P-prime submodule. We say M is a prime module if the zero submodule of M is a prime submodule of M. Clearly, this is equivalent to the following condition: for all $a \in A$ and $m \in M$ we have $(am = 0) \Rightarrow (m = 0$ or $aM = 0)$. In particular, the ring A is a prime A-module if and only if A is an integral domain. Moreover, N is a prime submodule of M if and only if M/N is a prime module (for more details see [16]).

Theorem 2.5. Under the above notation, the following assertions hold:

(1) If φ is surjective, then $M \bowtie^{\varphi,\psi} (JN,J'P)$ is a prime $A \bowtie^{f,g} (J,J')$-module if and only if "$JN = 0$ and $\psi(M)+J'P$ is a prime $g(A)+J'$-module" or " $J'P = 0$ and $\varphi(M)+JN$ is a prime $f(A)+J$-module".

(2) Suppose that N and P are faithful modules, then $M \bowtie^{\varphi,\psi} (JN,J'P)$ is a prime $A \bowtie^{f,g} (J,J')$-module if and only if "$J = 0$ and $\psi(M)+J'P$ is a prime $g(A)+J'$-module" or " $J' = 0$ and $\varphi(M)+JN$ is a prime $f(A)+J$-module".

Proof. (1) Assume that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a prime module over $A \bowtie^{f,g} (J, J')$. If $JN \neq 0$ and $J'P \neq 0$, then there are $(j, j') \in J \times J'$ and $(n, p) \in N \times P$ such that $jn \neq 0$ and $j'p \neq 0$. We have $(j, 0)(0, j'p) = (0, 0)$. Since φ is surjective, $(j, 0)(M \bowtie^{\varphi,\psi} (JN, J'P)) \neq 0$ since $(j, 0)(\varphi(m), \psi(m)) \neq (0, 0)$ for some $m \in M$, a contradiction. Therefore, one of JN and JP must be null. We suppose that $JN = 0$. Let $g(a) + j' \in g(A) + J'$ and $\psi(m) + p \in \psi(M) + J'P$ such that $(g(a) + j')(\psi(m) + p) = 0$. So, $\psi(am) \in J'P$ and hence $(f(a), g(a)+j')(\varphi(m), \psi(m)+p) = (0, 0)$. By hypothesis, we have either $(f(a), g(a)+j')(M \bowtie^{\varphi,\psi} (JN, J'P)) = 0$ or $(\varphi(m), \psi(m) + p) = (0, 0)$, which gives that either $(g(a) + j')(\psi(M) + J'P) = 0$ or $\psi(m) + p = 0$. Thus, $\psi(M) + J'P$ is a prime $g(A) + J'$-module. Conversely, suppose that $JN = 0$ and $\psi(M) + J'P$ is a prime $g(A) + J'$-module. If $(f(a) + j, g(a) + j')(\varphi(m), \psi(m) + p) = (0, 0)$ for some $(f(a) + j, g(a) + j') \in A \bowtie^{f,g} (J, J')$ and $(\varphi(m), \psi(m) + p) \in M \bowtie^{\varphi,\psi} (JN, J'P)$, then $(g(a) + j')(\psi(M) + J'P) = 0$ or $\psi(m) + p = 0$. If $(g(a) + j')(\psi(M) + J'P) = 0$, then $\psi(am') \in J'P$ for each $m' \in M$, and hence $(f(a) + j, g(a) + j')(M \bowtie^{\varphi,\psi} (JN, J'P)) = 0$. If $\psi(m) + p = 0$, then $\psi(m) \in J'P$ and so $(\varphi(m), \psi(m) + p) = (0, 0)$. By similar argument, we prove that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is prime if and only $\varphi(M) + JN$ is prime in the case where $J'P = 0$.
(2) Suppose that N and P are faithful and $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a prime module. If $J \neq 0$ and $J' \neq 0$ then for nonzero elements $j \in J$, $j' \in J'$ and $p \in P$ we have $(j, 0)(0, j'p) = (0, 0)$ and $(j, 0)(M \bowtie^{\varphi,\psi} (JN, J'P)) \neq 0$, a contradiction. By similar arguments of (1), we obtain the desired result. □

Corollary 2.6. *Let $f : A \to B$ be a ring homomorphism, J be an ideal of B and $\varphi : M \to N$ be an A-module homomorphism such that $\varphi(JN) \subseteq f^{-1}(J)M$. Then the amalgamation $M \bowtie^{\varphi} JN$ is a prime $A \bowtie^{f} J$-module if and only if "$\varphi^{-1}(JN) = 0$ and $\varphi(M) + JN$ is a prime $(f(A) + J)$-module" or "$JN = 0$ and M is a prime A-module".*

Corollary 2.7. *The ring $A \bowtie^{f,g} (J, J')$ is an integral domain if and only if "$J = 0$ and $g(A)+J'$ is an integral domain" or $J' = 0$ and $f(A) + J$ is an integral domain".*

According to [17], an A-module M is called a reduced module if for any $a \in A$ and $m \in M$, $am = 0$ implies $Am \cap aM = 0$. It can be easily seen that M is a reduced module if and only if for any $m \in M$ and $a \in A$, $a^2m = 0$ implies $am = 0$. Also, recall from [1] that $Nil_A(M) = \{m \in M \mid (Am : M)^k Am = 0 \text{ for some } k \in \mathbb{N}\}$. On the other hand, A submodule F of M is said to be semiprime if $F \neq M$ and whenever $a \in A$ and $m \in M$ are such that $a^2m \subseteq F$, then $am \in F$ (see [18]).

Theorem 2.8. Under the above notation, consider the following conditions:

(a) $\varphi(M) + JN$ is a reduced $(f(A) + J)$-module and $J'P \cap Nil(P) = 0$.

(b) $\psi(M) + J'P$ is a reduced $(g(A) + J')$-module and $JN \cap Nil(N) = 0$.

(c) $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a reduced $(A \bowtie^{f,g} (J, J'))$-module.

(d) $JN \cap Nil(N) = 0.$ and $J'P \cap Nil(P) = 0$.

Then:

(1) (a) or (b) $\Rightarrow$ (c).

(2) If N (resp., P) is a finitely generated multiplication B-module (resp., C-module), and $\varphi(M) + JN$ (resp., $\psi(M) + J'P$) is a faithful $(f(A) + J)$-module (resp., $(g(A) + J')$-module), then (c) $\Rightarrow$ (d).

(3) If F_0 is radical, then (d) $\Rightarrow$ (a) and (b).

(4) Suppose that f and φ are surjective, $Ker(f) \subseteq Ker(g)$ P is a finitely generated multiplication C-module and $\psi(M) + J'P$) is a faithful $(g(A) + J')$-module, then

$$M \bowtie^{\varphi,\psi} (JN, J'P) \text{ is reduced} \Leftrightarrow N \text{ is reduced and } J'P \cap Nil(P) = 0.$$

Proof. (1) Let $(f(a)+j, g(a)+j') \in A \bowtie^{f,g} (J,J')$ and $(\varphi(m)+n, \psi(m)+p) \in M \bowtie^{\varphi,\psi} (JN, J'P)$ such that $(f(a)+j, g(a)+j')^2(\varphi(m)+n, \psi(m)+p) = (0,0)$. Since $\varphi(M)+JN$ is reduced, we have $(f(a)+j)(\varphi(m)+n) = 0$. An easy calculation reveals that $(g(a)+j')(\psi(m)+p) \in J'P \cap Nil(P)$ and so $(g(a)+j')(\psi(m)+p) = 0$. Thus $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a reduced $(A \bowtie^{f,g} (J,J'))$-module. Likewise for $(b) \Rightarrow (c)$.
(2) Suppose that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a reduced $(A \bowtie^{f,g} (J,J'))$-module. Let $n \in JN \cap Nil(N)$. So, there exists a positive integer k such that $(Bn : N)^{2k}Bn = 0$. By [8, Theorem 3.1], we get $(Bn : N) \subseteq J$. Now, we will prove that $(Bn : N)Bn = 0$. Let $b \in (Bn : N)$, so $(b^k, 0)^2(n, 0) = (0,0)$. By hypothesis, we have $b^k = 0$. We will apply the same reasoning and we obtain that $bn = 0$. This implies that $(Bn : N)^2(\varphi(M)+JN) = 0$. The fact that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is reduced ensures that $(Bn : N)(\varphi(M)+JN) = 0$, hence $(Bn : N) = 0$. It follows that $Bn = (Bn : N)N = 0$ and thus $JN \cap Nil(N) = 0$. Likewise for $J'P \cap Nil(P) = 0$.
(3) We will show that $\varphi(M)+JN$ is a reduced $(f(A)+J)$-module. Suppose that $(f(a)+j)^2(\varphi(m)+n) = 0$ for some $f(a)+j \in f(A)+J$ and $\varphi(m)+n \in \varphi(M)+JN$. So, $a^2m \in F_0$. By assumption, we have $(f(a)+j)(\varphi(m)+n) \in JN \cap Nil(N)$. So, $(f(a)+j)(\varphi(m)+n) = 0$. The same arguments lead to (b).
(4) By (1) and (2), it suffices to prove that if $M \bowtie^{\varphi,\psi} (JN, J'P)$ is reduced then N is reduced. If $f(a)^2\varphi(m) = 0$ for some $a \in A$ and $m \in M$, then $(f(a), g(a))^2(\varphi(m), \psi(m)) = (0,0)$, which implies that $(f(a), g(a))(\varphi(m), \psi(m)) = (0,0)$. This completes the proof. □

Corollary 2.9. *Let $f : A \to B$ be a ring homomorphism, J be an ideal of B and $\varphi : M \to N$ be an A-module homomorphism such that $\varphi(JN) \subseteq f^{-1}(J)M$. Suppose that N is a multiplication B-module and $\varphi(M)+JN$ is a faithful $(f(A)+J)$-module. Then the amalgamation $M \bowtie^{\varphi} JN$ is a reduced $A \bowtie^{f} J$-module if and only if M is a reduced A-module and $JN \cap Nil(N) = 0$.*

Corollary 2.10. *[16, Proposition 4.7] Under the above notation, consider the following conditions:*

(a) *$f(A)+J$ is reduced and $J' \cap Nil(C) = 0$,*

(b) *$g(A)+J'$ is reduced and $J \cap Nil(B) = 0$,*

(c) *$A \bowtie^{f,g} (J,J')$ is reduced,*

(d) *$J \cap Nil(B) = 0$ and $J' \cap Nil(C) = 0$.*

Then:

(1) *(a) or (b) $\Rightarrow$ (c) $\Rightarrow$ (d).*

(2) *If I_0 is radical, then the four conditions are equivalent.*

(3) *If f is surjective and $Ker(f) \subseteq Ker(g)$, then:*

$$A \bowtie^{f,g} (J,J') \text{ is reduced} \Leftrightarrow B \text{ is reduced and } J' \cap Nil(C) = 0.$$

Let $f : A \to B$ and $g : A \to C$ be two ring homomorphisms and let J (resp., J') be an ideal of B (resp., C) such that $f^{-1}(J) = g^{-1}(J')$ and let n be a positive integer. Consider the functions $f^n : A^n \to B^n$ defined by $f^n((\alpha_i)_{i=1}^n) = (f(\alpha_i))_{i=1}^n$ and $g^n : A^n \to C^n$ defined by $g^n((\alpha_i)_{i=1}^n) = (g(\alpha_i)^n)_{i=1}$. Obviously, f^n and g^n are ring homomorphisms and J^n, J'^n are ideals of B^n and C^n, respectively. This allows us to defined $A^n \bowtie^{f^n,g^n} (J^n, J'^n)$. On the other hand, let M be an A-module. We say an element $m \in M$ a regular element if $(0 :_A m) = 0$. Also, a submdoule N of M is said to be a regular submodule if it contains a regular element.

Theorem 2.11. Under the above notation. If J and J' are finitely generated ideals of $(f(A)+J)$ and $(g(A)+J')$ respectively, then the following assertions hold:

(1) If JN and $J'P$ are finitely generated modules over $(f(A)+J)$ and $(g(A)+J')$ respectively, $J \subseteq f(A)$, then $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module if and only if $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module and $\psi(M)+J'P$ is a coherent $(g(A)+J')$-module.

(2) Suppose that If JN and $J'P$ are finitely generated modules over $(f(A)+J)$ and $(g(A)+J')$ respectively, and $J^2N = 0$. Then $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module if and only if $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module and $\psi(M)+J'P$ is a coherent $(g(A)+J')$-module.

(3) Assume that JN and $J'P$ are regular modules over $f(A)+J$ and $(g(A)+J')$ respectively, and $J \subseteq f(A)$. Then $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module if and only if $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module and $\psi(M)+J'P$ is a coherent $(g(A)+J')$-module and JN and $J'P$ are finitely generated modules over $(f(A)+J)$ and $(g(A)+J')$ respectively.

(4) Assume that JN is a regular finitely generated $(f(A)+J)$-module, and $J' \subseteq g(A)$. Then $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module if and only if $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module and $\psi(M)+J'P$ is a coherent $(g(A)+J')$-module and $J'P$ is a finitely generated $(g(A)+J')$-module.

(5) Suppose that JN is a regular finitely generated $(f(A)+J)$-module, and $J'^2P = 0$. Then $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module if and only if $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module and $\psi(M)+J'P$ is a coherent $(g(A)+J')$-module and $J'P$ is a finitely generated $(g(A)+J')$-module.

The proof of this theorem draws on the following results.

Lemma 2.12. (1) *$JN \times \{0\}$ (resp., $\{0\} \times J'P$) is a finitely generated $A \bowtie^{f,g} (J,J')$-module if and only if JN (resp., $J'P$) is a finitely generated $f(A)+J$-module (resp., $g(A)+J'$-module).*

(2) *If $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module, JN and $J'P$ are finitely generated modules over $f(A)+J$ and $g(A)+J'$, respectively, then $\varphi(M)+JN$ is a coherent $f(A)+J$-module and $\psi(M)+J'P$ a coherent $g(A)+J'$-module.*

Proof. (1) Consider the first component projection from $A \bowtie^{f,g} (J,J')$ to $f(A)+J$. So, $(A \bowtie^{f,g} (J,J'))X = (f(A)+J)X$ for any subset X of the $A \bowtie^{f,g} (J,J')$-module $JN \times \{0\}$. Likewise for the second part.
(2) Assume that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J,J')$-module, JN and $J'P$ are finitely generated modules over $f(A)+J$ and $g(A)+J'$, respectively. Then $\{0\} \times J'P$ and $JN \times \{0\}$ are finitely generated $A \bowtie^{f,g} (J,J')$-modules. It follows that $\frac{M \bowtie^{\varphi,\psi} (JN,J'P)}{0 \times J'P}$ and $\frac{M \bowtie^{\varphi,\psi} (JN,J'P)}{JN \times 0}$ are coherent $A \bowtie^{f,g} (J,J')$-modules. By Proposition 2.2 and [13, Theorem 2.2.7], we conclude that $\varphi(M)+JN$ is a coherent $f(A)+J$-module and $\psi(M)+J'P$ a coherent $g(A)+J'$-module. □

Lemma 2.13. *Let J and J' be ideals of $f(A)+J$ and $g(A)+J'$, respectively, with $J \subseteq f(A)$. If JN is a finitely generated $f(A)+J$-module, J' is a finitely generated ideal of $g(A)+J'$ and $\varphi(M)+JN$ is a coherent $f(A)+J$ module, then $JN \times \{0\}$ is a coherent $A \bowtie^{f,g} (J,J')$-module.*

Proof. Since $JN \times \{0\}$ is a finitely generated $A \bowtie^{f,g} (J,J')$-module, it suffices to prove that every finitely generated $A \bowtie^{f,g} (J,J')$-submodule of $JN \times \{0\}$ is finitely presented. Let K be a finitely generated submodule of $JN \times \{0\}$. One can see that $K = F \times \{0\}$, where $F = \sum_{i=1}^{n}(f(A)+J)f_i$ for some positive integer n and $f_i \in F$. Consider the exact sequence of $(f(A)+J)$-modules

$$0 \longrightarrow \ker u \longrightarrow (f(A)+J)^n \longrightarrow F \longrightarrow 0$$

where $u(f(\alpha_i)+j_i)_{i=1}^n) = \sum_{i=1}^n (f(\alpha_i)+j_i)f_i$. So,

$$\begin{aligned}\ker u &= \{(f(\alpha_i)+j_i)_{i=1}^n \in (f(A)+J)^n \mid \sum_{i=1}^n (f(\alpha_i)+j_i)f_i = 0\}\\ &= \{(f(a_i))_{i=1}^n \in (f(A))^n \mid \sum_{i=1}^n (f(a_i))f_i = 0\}\end{aligned}$$

where $a_i = \alpha_i + k_i$ and $f(k_i) = j_i$ for some $k_i \in A$. The fact that $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module implies that $\ker u$ is finitely generated. Let $\{(f^n(a_i^1)_{i=1}^n), (f^n(a_i^2)_{i=1}^n), \ldots, (f^n(a_i^m)_{i=1}^n)\}$ be a generating set of $\ker u$. On the other hand, consider the exact sequence of $A \bowtie^{f,g} (J,J')$-modules

$$0 \longrightarrow \ker u \longrightarrow (A \bowtie^{f,g} (J,J'))^n \longrightarrow K \longrightarrow 0$$

where $v(f(\alpha_i)+j_i, g(\alpha_i)+j_i')_{i=1}^n) = \sum_{i=1}^n (f(\alpha_i)+j_i, g(\alpha_i)+j_i')(f_i, 0)$. Then

$$\begin{aligned}\ker v &= \{(f(\alpha_i)+j_i, g(\alpha_i)+j_i')_{i=1}^n \in (A \bowtie^{f,g} (J,J'))^n \mid \sum_{i=1}^n (f(\alpha_i)+j_i)f_i = 0\}\\ &= \{(f(b_i), g(b_i)+k_i)_{i=1}^n \in (A \bowtie^{f,g} (J,J'))^n \mid \sum_{i=1}^n (f(b_i))f_i = 0\}\end{aligned}$$

where $b_i = \alpha_i + t_i$ and $f(t_i) = j_i$ for some $t_i \in A$.
Let U be the submodule of A^n generated by $\{(a_i^1)_{i=1}^n, (a_i^2)_{i=1}^n, \ldots, (a_i^m)_{i=1}^n\}$. We show that $\ker v = U \bowtie^{f^n,g^n} (0, J'^n)$. Let $x = (f^n((b_i)_{i=1}^n), g^n((b_i)_{i=1}^n + (k_i)_{i=1}^n) \in \ker v$. So, we have $\sum_{i=1}^n (f(b_i))f_i = 0$. Hence $f^n((b_i)_{i=1}^n) \in \ker u$ and thus

$$\begin{aligned}f^n((b_i)_{i=1}^n) &= \sum_{l=1}^m f(\alpha_l) f^n((a_i^l)_{i=1}^n)\\ &= \sum_{l=1}^m f^n((\alpha_l a_i^l)_{i=1}^n)\\ &= f^n(\sum_{l=1}^m \alpha_l a_i^l)_{i=1}^n).\end{aligned}$$

It follows that $x = (f^n(\sum_{l=1}^m \alpha_l a_i^l)_{i=1}^n), g^n(\sum_{l=1}^m \alpha_l a_i^l)_{i=1}^n) + (k_i')_{i=1}^n)$, with $(\sum_{l=1}^m \alpha_l a_i^l)_{i=1}^n) \in U$. This gives that $x \in U \bowtie^{f^n,g^n} (0, J'^n)$. Conversely, let $x = (f^n((b_i))_{i=1}^n, g^n((b_i))_{i=1}^n + (j_i))_{i=1}^n) \in U \bowtie^{f^n,g^n} (0, J'^n)$. Then $(b_i)_{i=1}^n = \sum_{l=1}^m \alpha_l((a_i^l)_{i=1}^n) = \sum_{l=1}^m \alpha_l((a_i^l))_{i=1}^n$. We have

$$\sum_{i=1}^n f(b_i)f_i = \sum_{i=1}^n f(\sum_{l=1}^m \alpha_l a_i^l))f_i = \sum_{i=1}^n (\sum_{l=1}^m f(\alpha_l a_i^l)))f_i = \sum_{l=1}^m f(\alpha_l)(\sum_{i=1}^n f(a_i^l)f_i) = 0.$$

Consequently, $x \in \ker v$. By [12, Lemma 2] , $\ker(v)$ is finitely generated and so K is finitely presented and thus $JN \times \{0\}$ is a coherent $A \bowtie^{f,g} (J,J')$-module. □

Lemma 2.14. *Assume that J (resp., J') is a finitely generated ideal of $f(A)+J$ (resp., $g(A)+J'$), and $J^2N = 0$. If $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module, then $JN \times 0$ is a coherent $A \bowtie^{f,g} (J,J')$-module.*

Proof. Since $JN \times 0$ is a finitely generated $A \bowtie^{f,g} (J,J')$-module (because J and N are finitely generated $(f(A)+J)$-modules), it remains to prove that every finitely generated $A \bowtie^{f,g} (J,J')$-submodule of $JN \times 0$

is finitely presented. Let $L = K\times 0$ be a finitely generated submodule of $JN\times 0$. So, $K = \sum_{i=1}^{n}(f(A)+J)n_i$, for some positive integer n and $n_i \in JN$. Consider the following exact seqeunce of $(f(A)+J)$-modules

$$0 \longrightarrow \ker v \longrightarrow (f(A)+J))^n \longrightarrow K \longrightarrow 0$$

where $v((f(a_i)+j_i)_{i=1}^{n}) = \sum_{i=1}^{n}(f(a_i)+j_i)n_i$. We obtain that

$$\begin{aligned} \ker v &= \{(f^n(\alpha_i)_{i=1}^n + (j_i)_{i=1}^n) \in (f(A)+J)^n \mid \sum_{i=1}^{n}(f(\alpha_i)+j_i)n_i = 0\} \\ &= \{(f^n(\alpha_i)_{i=1}^n + (j_i)_{i=1}^n) \in (f(A)+J)^n \mid \sum_{i=1}^{n} f(\alpha_i)n_i = 0\}. \end{aligned}$$

Since $\varphi(M)+JN$ is a coherent $(f(A)+J)$-module, then $\ker v$ is a finitely generated $(f(A)+J)$-module. Let $\{f^n(\alpha_i^1)_{i=1}^n + (j_i^1)_{i=1}^n), \ldots, (f^n(\alpha_i^m)_{i=1}^n + (j_i^m)_{i=1}^n)\}$ be a generating set of $\ker v$. Now, we consider the following exact sequence of $A \bowtie^{f,g} (J,J')$-modules

$$0 \longrightarrow \ker u \longrightarrow (A \bowtie^{f,g} (J,J'))^n \longrightarrow L \longrightarrow 0$$

where $u((f(a_i)+j_i, g(a_i)+j_i')_{i=1}^n) = \sum_{i=1}^{n}(f(a_i)+j_i, g(a_i)+j_i')(n_i,0) = \sum_{i=1}^{n}(f(a_i)+j_i)n_i, 0)$. Then

$$\begin{aligned} \ker u &= \{(f^n(\beta_i)_{i=1}^n + (j_i)_{i=1}^n, g^n(\beta_i)_{i=1}^n + (j_i')_{i=1}^n) \in (A \bowtie^{f,g} (J,J'))^n \mid \sum_{i=1}^{n}(f(\beta_i)+j_i)n_i = 0\} \\ &= \{(f^n(\beta_i)_{i=1}^n + (j_i)_{i=1}^n, (g^n(\beta_i)_{i=1}^n + (j_i')_{i=1}^n) \in (A \bowtie^{f,g} (J,J'))^n \mid \sum_{i=1}^{n} f(\beta_i)n_i = 0\}. \end{aligned}$$

Take U the A-submodule of A^n generated by $\{(\alpha_i^1)_{i=1}^n, \ldots, (\alpha_i^m)_{i=1}^n\}$. Our task is to prove that $\ker u = U \bowtie^{f^n,g^n} (J^n, J'^n)$. Let $x \in \ker u$. So, $x = (f^n(d_i)_{i=1}^n + (k_i)_{i=1}^n, (g^n(\beta_i)_{i=1}^n + (k_i')_{i=1}^n)$ with $\sum_{i=1}^{n} f(d_i)n_i = 0$, hence $f^n(d_i)_{i=1}^n + (k_i)_{i=1}^n \in \ker v$ and so

$$\begin{aligned} f^n(d_i)_{i=1}^n + (k_i)_{i=1}^n &= \sum_{s=1}^{m}(f(a_s)+t_s)(f^n(\alpha_i^s)_{i=1}^n + (j_i^s)_{i=1}^n) \\ &= \sum_{s=1}^{m} f^n(a_s\alpha_i^s)_{i=1}^n + (l_i)_{i=1}^n \\ &= f^n((\sum_{s=1}^{m} a_s\alpha_i^s)_{i=1}^n) + (k_i)_{i=1}^n. \end{aligned}$$

It follows that $x = (f^n((\sum_{s=1}^{m} a_s\alpha_i^s)_{i=1}^n) + (l_i)_{i=1}^n, g^n((\sum_{s=1}^{m} a_s\alpha_i^s)_{i=1}^n) + (l_i')_{i=1}^n)$, where $(\sum_{s=1}^{m} a_s\alpha_i^s)_{i=1}^n \in U$. So, $x \in U \bowtie^{f^n,g^n} (J^n, J'^n)$. Conversely, let $x = (f^n(d_i)_{i=1}^n + (k_i)_{i=1}^n, (g^n(\beta_i)_{i=1}^n + (k_i')_{i=1}^n) \in U \bowtie^{f^n,g^n} (J^n, J'^n)$, so $(d_i)_{i=1}^n = \sum_{s=1}^{m} a_s(\alpha_i^s)_{i=1}^n = (\sum_{s=1}^{m} a_s\alpha_i^s)_{i=1}^n$. We have

$$\sum_{i=1}^{n} f(d_i)n_i = \sum_{i=1}^{n} f(\sum_{s=1}^{m} a_s\alpha_i^s)n_i = \sum_{s=1}^{m} f(a_s)(\sum_{i=1}^{n} f(\alpha_i^s)n_i) = 0.$$

Consequently, $x \in \ker u$. Since U is a finitely generated A-module, J (resp., J') is a finitely generated ideal of $f(A)+J$ (resp., $g(A)+J')$), then $\ker u$ is finitely generated. Therefore, L is finitely presented which implies that $JN \times 0$ is a coherent $A \bowtie^{f,g} (J,J')$-module. □

Lemma 2.15. (1) *If $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J, J')$-module and JN is a regular submodule of $\varphi(M) + JN$, then $J'P$ is a finitely generated $g(A) + J'$-module.*

(2) *Suppose that JN and $J'P$ are regular submodules of $\varphi(M) + JN$ and $\psi(M) + J'P$ respectively. If $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J, J')$-module then $\varphi(M) + JN$ is a coherent $f(A) + J$-module and $\psi(M) + J'P$ a coherent $g(A) + J'$-module.*

Proof. (1) Suppose that $M \bowtie^{\varphi,\psi} (JN, J'P)$ is a coherent $A \bowtie^{f,g} (J, J')$-module and JN contains a regular element n. So, $(0 : (n, 0)) = 0 \times J'P$ is a finitely generated $A \bowtie^{f,g} (J, J')$-module, and thus $J'P$ is a finitely generated $g(A) + J'$-module.
(2) By applying (1) and Lemma 2.12. □

Proof of theorem.
(1) Follows immediately from Lemma 2.12, Lemma 2.13 and [13, Theorem 2.1.8].
(2) By Lemma 2.12, Lemma 2.14 and [13, Theorem 2.1.8].
(3) By (1) and Lemma 2.15.
(4) Follows immediately from (1) and Lemma 2.15.
(5) Follows immediately from (2) and Lemma 2.15. □

Corollary 2.16. *[12, Theorem 1] Under the above notations, the followings hold:*

(1) *Assume that J (resp., J') is a finitely generated ideal of $f(A) + J$ (resp., $g(A) + J'$), and $J \subseteq f(A)$. Then $A \bowtie^{f,g} (J, J')$ is a coherent ring if and only if $f(A) + J$ and $g(A) + J'$ are coherent rings.*

(2) *Assume that J (resp., J') is a finitely generated ideal of $f(A) + J$ (resp., $g(A) + J'$), and $J^2 = 0$. Then $A \bowtie^{f,g} (J, J')$ is a coherent ring if and only if $f(A) + J$ and $g(A) + J'$ are coherent rings.*

(3) *Assume that J (resp., J') is a regular ideal of $f(A)+J$ (resp., $g(A)+J'$), and $J \subseteq f(A)$. Then $A \bowtie^{f,g} (J, J')$ is a coherent ring if and only if $f(A) + J$ and $g(A) + J'$ are coherent rings J (resp., J') is a finitely generated ideal of $f(A) + J$ (resp., $g(A) + J'$).*

(4) *Assume that J is a regular finitely generated ideal of $f(A) + J$ and $J' \subseteq g(A)$. Then $A \bowtie^{f,g} (J, J')$ is a coherent ring if and only if $f(A)+J$ and $g(A)+J'$ are coherent rings and J' is a finitely generated ideal of $g(A) + J'$.*

(5) *Assume that J is a regular finitely generated ideal of $f(A) + J$ and $J'^2 = 0$. Then $A \bowtie^{f,g} (J, J')$ is a coherent ring if and only if $f(A)+J$ and $g(A)+J'$ are coherent rings and J' is a finitely generated ideal of $g(A) + J'$.*

References

[1] M. M. Ali, Idempotent and nilpotent submodules of multiplication modules, Comm. Algebra 36 (2008), 4620–4642.

[2] A. Anebri, N. Mahdou and Ü. Tekir, Commutative rings and modules that are r-Noetherian, Bull. Korean Math. Soc. 58(5) (2021), 1221-1233.

[3] A. Anebri, N. Mahdou and Ü. Tekir, On modules satisfying the descending chain condition on r-submodules, Comm. Algebra 50(1) (2022), 383–391.

[4] A. J. Berrick and M. E. Keating, An introduction to rings and modules with K-theory in view, Cambridge University Press (2000).

[5] M. D'Anna, C. A. Finocchiaro and M. Fontana, Amalgamated algebras along an ideal, in: Commutative Algebra and Applications, Proceedings of the Fifth International Fez Conference on Commutative Algebra and Applications, Fez, Morocco, 2008, W. de Gruyter Publisher, Berlin, 2009, pp. 155–172.

[6] M. D'Anna and M. Fontana, Amalgamated duplication of a ring along a multiplicative canonical ideal, Ark. Mat. 45(2) (2007), 241–252.

[7] M. D'Anna and M. Fontana, An amalgamated duplication of a ring along an ideal: the basic properties, J. Algebra Appl. 6(3) (2007), 443–459.

[8] Z. A. El-Bast and P. P. Smith, Multiplication modules, Comm. Algebra 16(4) (1988), 755–779.

[9] R. El Khalfaoui, N. Mahdou, P. Sahandi and N. Shirmohammadi, Amalgamated modules along an ideal, Commun. Korean Math. Soc. 36(1) (2021), 1–10.

[10] A. El Khalfi, H. Kim and N. Mahdou, Amalgamation extension in commutative ring theory: a survey, Moroccan Journal of Algebra and Geometry with Applications 1(1) (2022), 139–182.

[11] A. El Khalfi, N. Mahdou and A. Rahmouni Hassani, Bi-amalgamation of rings defined by certain annihilator conditions, Rocky Mountain J. Math. 50(3) (2020), 947–956.

[12] M. El Ouarrachi and N. Mahdou, Coherence in bi-amalgamated algebras along ideals, Springer Proceedings in Mathematics & Statistics, 127–138.

[13] S. Glaz, Commutative coherent rings, Lecture Notes in Mathematics, 1371, Springer- Verlag, Berlin, 1989.

[14] A. Hamed, S-strong Mori domain of $A + XB[X]$, Moroccan Journal of Algebra and Geometry with Applications, 1(2) (2022), 183–188.

[15] S. Kabbaj, Matlis' semi-regularity and semi-coherence in trivial ring extensions: a survey, Moroccan Journal of Algebra and Geometry with Applications 1(1) (2022), 1–17.

[16] S. Kabbaj, K. Louartiti and M. Tamekkante, Bi-Amalgamated algebras along ideals, J. Commut. Algebra 9(1) (2017), 65–87.

[17] T. K. Lee and Y. Zhou, Reduced Modules. Rings, Modules, Algebras and Abelian Groups. Lecture Notes in Pure and Applied Mathematics, Vol. 236. New York: Dekker (2004), pp. 365–377.

[18] B. Saraç, On Semiprime Submodules, Comm. Algebra 37 (2009), 2485–2495.

[19] M. Tamekkante and E. M. Bouba, Bi-amalgamation of small weak global dimension, Int. Electron. J. Algebra 21 (2017), 127–136.

[20] M. Tamekkante and E. M. Bouba, On pm^+ and finite character bi-amalgamation, Bull. Iranian Math. Soc. 43(5) (2017), 1237–1244.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 93-107

ISSN: 2820-7114

Title :

Cohomology of units and Z_2-torsion of the cyclotomic Z_2-extension of some CM fields

Author(s):

Ali Mouhib

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 93–107
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

Cohomology of units and $\mathbb{Z}_2$-torsion of the cyclotomic $\mathbb{Z}_2$-extension of some CM fields

Ali Mouhib
Univ. Mohammed Ben Abdellah, Fes LSI,
Sciences and Engineering Laboratory Polydisciplinary Faculty of Taza B.P 1223 Taza-Gare Morocco.
e-mail: *mouhibali@yahoo.fr*

Communicated by El Hassan El Kinani
(Received 26 November 2022, Revised 10 February 2023, Accepted 14 February 2023)

Abstract. It is well known from the results of Ferrero and Kida that the $\mathbb{Z}_2$-torsion part of the unramified abelian Iwasawa module $X_\infty(K)$ of any imaginary quadratic number field K is trivial or cyclic of order 2. In this article, we will determine an infinite family of imaginary multiquadratic number fields, in which the $\mathbb{Z}_2$-torsion of the Iwasawa module X_∞ is of arbitrary large rank, giving also the exact value of the rank of X_∞. Also, we will compute the first and the second cohomology groups of units of the cyclotomic $\mathbb{Z}_2$-extension of some CM fields K. Hence, as application, using the Iwasawa Riemann-Hurwitz formula, we obtain the Iwasawa $\lambda_2(K)$ invariant of the cyclotomic $\mathbb{Z}_2$-extension of K in terms of the Iwasawa λ_2 invariant of the cyclotomic $\mathbb{Z}_2$-extension of some totally real subfield of K, thus giving an alternative proof of the previous results obtained by Kida.

Key Words: Class group, Unit group, Capitulation problem, $\mathbb{Z}_2$-extension.
2010 MSC: 11R29, 11R32, 11R37, 11R23.

In memory of Muhammad Zafrullah.

1 Introduction

Let p be a prime number, $\mathbb{Z}_p$ be the ring of p-adic integers, K be a number field, K_∞ be the cyclotomic $\mathbb{Z}_p$-extension of K and for each non-negative integer n, K_n be the n-th layer of K_∞, L_n be the Hilbert p-class field of K_n, $A(K)$ be the p-class group of K and for any integer $n \geq 1$, $A_n(K)$ be the p-class group of K_n. By class field theory, we have $A_n(K) \simeq Gal(L_n/K_n)$. The maximal abelian unramified Iwasawa module $X_\infty(K)$ of K is defined as

$$X_\infty(K) := \varprojlim A_n(K) \simeq \varprojlim Gal(L_n/K_n),$$

where the first projective limit is defined with respect to the norm maps and the second one is defined with respect to the restriction maps. It is well known, by Iwasawa's results that $X_\infty(K)$ is a finitely generated torsion $\Lambda := \mathbb{Z}_p[[T]]$-module and for large n, we have:

$$|A_n(K)| = p^{\lambda_p(K)n+\mu_p(K)p^n+\nu_p(K)},$$

where $\lambda_p(K)$, $\mu_p(K)$ and $\nu_p(K)$ are so called the Iwasawa invariants of K_∞/K. Clearly X_∞ is finite if and only if $\lambda_p(K) = \mu_p(K) = 0$. It is well known [F-W], that $\mu_p(K) = 0$, if K is supposed abelian over $\mathbb{Q}$. In that case, the $\mathbb{Z}_p$-torsion $X^0_\infty(K)$ and the maximal finite Λ-submodule of $X_\infty(K)$ are equal and we have the following structure of the Iwasawa module $X_\infty(K)$:

$$X_\infty(K) \simeq \mathbb{Z}_p^{\lambda(K)} \oplus X^0_\infty(K).$$

Let I_{K_∞} be the group of ideals of K_∞, namely the multiplicative group of invertible o_{K_∞}-modules of K_∞, P_{K_∞} be the subgroup of principal ideals in I_{K_∞} and $C_K = I_{K_\infty}/P_{K_\infty}$ be the ideal class group of K_∞. It is well known [I-2], that C_K is a torsion abelian group. Let $A_\infty(K)$ be the p-primary component of C_{K_∞}, which is equal to $\varinjlim A_n(K)$, where the inductive limit is defined with respect to the identity maps. We have [I-2] :

$$A_\infty(K) \simeq (\mathbb{Q}_p/\mathbb{Z}_p)^{\lambda_p(K)} \oplus A',$$

where A' is such that $p^a A' = 0$ for some nonnegative integer a. Precisely $A' = 0$, when X_∞ is a finitely generated $\mathbb{Z}_p$-module, which is satisfied, when $\mu_p(K) = 0$.
Several approaches have been used to determine the Iwasawa's λ-invariant. One of these approaches is the use of Riemann-Hurwits formula applied for a cyclic extension of number fields of degree p [I-2], which turns out to the computation of some cohomology groups of units. In [F-K-O-T], the authors gave a criterion for the vanishing of λ_p-invariant of cyclic extensions of totally real number fields with degree p. An extension of this result has been given in [S], where the author gave an extension of the Riemann Hurwits formula to cyclic extension of degree p^n for any positive integer n.
Next, for each group G which is a finitely generated $\mathbb{Z}_p$-module, we denote by $rk_p(G)$, the dimension of the $\mathbb{F}_p$-vectorial space G/G^p.

For a special case $p = 2$, B. Ferrero [F] and Y. Kida [K], determined independently and explicitly the Iwasawa λ_2 invariant of the cyclotomic $\mathbb{Z}_2$-extension of the imaginary quadratic number fields $\mathbb{Q}(\sqrt{-d})$, where d is a positive square-free integer:

$$\lambda_2(\mathbb{Q}(\sqrt{-d})) = \sum_{p|d,\, p\neq 2} 2^{-3+v_2(p^2-1)} - 1.$$

Precisely, the $\mathbb{Z}_2$-torsion of $X_\infty(\mathbb{Q}(\sqrt{-d}))$ is trivial or isomorphic to $\mathbb{Z}/2\mathbb{Z}$ and we have for $d \neq 1$:

$$X_\infty(\mathbb{Q}(\sqrt{-d})) \simeq \mathbb{Z}_2^{\lambda_2(\mathbb{Q}(\sqrt{-\mathrm{d}}))} \oplus \mathbb{Z}/2\mathbb{Z} \iff \mathrm{d} \equiv 1 \pmod 4.$$

Also in [A], M. Atsuta studied the minus quotient X_∞^- of the Iwasawa module X_∞ for CM number fields, that is

$$X_\infty^- = X_\infty/(1+J)X_\infty,$$

where J be the complex conjugation. He determined the maximal finite submodule of X_∞^- under some mild assumptions. Precisely for a CM number field F such that its totally real maximal subfield F^+ is unramified at 2 and contains a unique 2-adic place, then $X_\infty^-(F)$ has no non-trivial finite Λ-submodule [A, Example 2.8]. So from the exact sequence :

$$0 \longrightarrow X_\infty(F^+) \longrightarrow X_\infty(F) \longrightarrow X_\infty^-(F) \longrightarrow 0,$$

we have the maximal finite Λ-submodule of F coincides with the maximal finite submodule of F^+:

$$X_\infty^0(F) = X_\infty^0(F^+).$$

Hence, in that case, to determine CM number fields with large $\mathbb{Z}_2$-torsion, this turns out to check Greenberg's conjecture for totally real quadratic number fields in which the 2-rank of X_∞ is large (see Section 3).

Also, we will compute the first and the second cohomology groups of units of the cyclotomic $\mathbb{Z}_2$-extension of some CM fields K and hence as an application, using the Iwasawa Riemann-Hurwitz formula [I-2], we obtain the Iwasawa $\lambda_2(K)$ invariant of K in terms of the Iwasawa λ_2 invariant of some totally real subfield of K, giving then an alternative proof of Kida's result [K].

2 Preliminaries and Results

Let p be a prime number and K/k be a cyclic extension of number fields of degree a prime number p and Galois group G. Denote by k_∞ and K_∞ be respectively the cyclotomic $\mathbb{Z}_p$-extensions of k and K. Suppose $K \not\subset k_\infty$, then $Gal(K_\infty/k_\infty) \simeq G \simeq \mathbb{Z}/p\mathbb{Z}$. For each finite place v of k_∞, and any place w of K_∞ lying over v, let $e(w/v)$ be the order of the inertia group of w in the extension K_∞/k_∞. Let E_{K_∞} be the unit group of K_∞ and let $H^i(G, E_{K_\infty})$, $i = 1, 2$ be the first and the second cohomology groups of the group E_{K_∞} as a G-module. Since G is cyclic of order p, then the cohomology groups $H^1(G, E_{K_\infty})$ and $H^2(G, E_{K_\infty})$ are abelian of exponent p, and can be viewed as vectorial spaces over $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

In [I-2], we find the following formula called Riemann-Hurwits formula giving the relation between $\lambda_p(k)$ and $\lambda_p(K)$ in terms of the dimensions of $H^1(G, E_{K_\infty})$ and $H^2(G, E_{K_\infty})$ and the number of ramified primes in the extension K_∞/k_∞

Theorem 2.1. Supppose that K_∞/k_∞ be a cyclic extension of degree p unramified at every infinite prime. For $i = 1, 2$, let h_i be the dimension of the $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$-vectorial space $H^i(G, E_{K_\infty})$. Then we have the following formula for $\lambda_p(K)$ and $\lambda_p(k)$.

$$\lambda_p(K) = p\lambda_p(k) + \sum_w (e(w/v) - 1) + (p-1)(h_2 - h_1),$$

where w ranges over all non p-places of K.

Let $s_\infty(K/k)$ be the number of ramified primes in the extension K_∞/k_∞ outside p, then we have the following formula:

$$\lambda_p(K) = p\lambda_p(k) + (p-1)[s_\infty(K/k) + h_2 - h_1].$$

Let P_{k_∞} and P_{K_∞} be respectively the group of principal ideals of k_∞ and K_∞. From the following exact sequence of G-modules:

$$1 \longrightarrow E_{K_\infty} \longrightarrow K_\infty^* \longrightarrow P_{K_\infty} \longrightarrow 1,$$

we obtain the following long exact sequence

$$1 \longrightarrow E_{k_\infty} \longrightarrow k_\infty \longrightarrow P_{K_\infty}^G \longrightarrow H^1(G, E_{K_\infty}) \longrightarrow H^1(G, K_\infty^*) \longrightarrow \cdots$$

and, in the last exact sequence, the image of the third application is exactly P_{k_∞}. Hence, we obtain $H^1(G, E_{K_\infty}) \simeq P_{K_\infty}^G/P_{k_\infty}$. Clearly, the last quotient group is generated by principal ramified and inert primes in K_∞/k_∞. Also, it is well known that for each p-adic place w of K_∞, the group $I_w = \varinjlim < \mathcal{P}_{w_i} >$, where $\mathcal{P}_{w_i}$ is the associated maximal ideal in the layer K_i of K_∞, is G-cohomologicaly trivial [I-2]. In particular $H^0(G, I_w) = 0$, hence the class of each p-adic place in $P_{K_\infty}^G/P_{k_\infty}$ is trivial.
In the case where $\lambda_p(k) = \mu_p(k) = 0$ which is equivalent to $A_\infty(k) = 0$, the p-primary part of $I_{k_\infty}/P_{k_\infty}$ is trivial. Hence, the group $P_{K_\infty}^G/P_{k_\infty}$ is generated by the classes of principal ramified primes in the extension K_∞/k_∞ outside p. We obtain the inequality $h_1 \le s_\infty(K/k)$ (see also [F-K-O-T]).

2.1 Computation of the Iwasawa λ_2 invariant of the cyclotomic $\mathbb{Z}_2$-extension of an imaginary quadratic number field

In [F] and [K], we have the exact value of $\lambda_2(\mathbb{Q}(\sqrt{-d}))$. In this section we will be interested to give an other proof for computing $\lambda_2(\mathbb{Q}(\sqrt{-d}))$, where d is a postive square-free integer.
We will use the following preliminary results. For each number field F, denote respectively, I_F, E_F, o_F, C_F, the group of fractional ideals, the unit group, the ring of integers and the class group of F. For a cyclic extension K/k of Galois group G, let the maps

$$\begin{array}{rcl} d_{K/k} : I_k & \longrightarrow & I_K \\ I & \longmapsto & I_K^G \\ j_{K/k} : C_k & \longrightarrow & C_K^G \\ \bar{I} & \longmapsto & \bar{I_K^G} \end{array}$$

From the well known long exact sequence of ambiguous class, we will use the the following exact sequence :

$$coker\ d_{K/k} \xrightarrow{\alpha} coker\ j_{K/k} \xrightarrow{\varphi} H^2(G, E_K) \xrightarrow{\psi} E_k/E_k \cap N_{K/k}(K^*) \longrightarrow 1.$$

In the following proposition, we give a necessary and sufficient condition, whenever each ambiguous class relatively to K/k is represented by an ideal fixed by G. We have the following result.

Proposition 2.2. *Each element of C_K^G is represented by an ideal fixed by G if and only if ψ is an isomorphism. In particular, if $|C_k|$ is not divisible by p, then C_K^G is generated by the classes of ramified primes in K/k.*

Proof. Let $H = \{C \in C_K^G \mid \exists I \in I_K,\ C = \bar{I}\}$ be the subgroup of C_K^G generated by ideals fixed by G. We remark that

$$Im(\alpha) = H/j_{K/k}(C_k).$$

It is easy to see that ψ is an isomorphism, if and only if φ is trivial, which is equivalent to ψ is onto. Hence, w is an isomorphism is equivalent to

$$H/j_{K/k}(C_k) = coker(j_{K/k}).$$

On the other hand, $coker j_{K/k} = C_K^G/j_{K/k}(C_k)$, then the first part of the proposition follows. Also, it is clear that H is generated by
(i) the classes of prime ideals in I_K which are either ramified or inert in K/k and
(ii) the classes of the product of the primes in I_K above the same (non principal) prime ideal in k which splits in K.
Consequently, in the case where $|C_k|$ is not divisible by p, C_K^G is clearly generated by the classes of ramified primes in K/k. □

Remark 2.3. Suppose that $[E_k : E_k \cap N_{K/k}(K^*)]$ is maximal equal to $[k : \mathbb{Q}]$. Hence, we obtain

$$H^2(G, E_K) = E_k/N_{K/k}(E_K).$$

This yields to an isomorphism:

$$H^2(G, E_K) \simeq E_k/E_k \cap N_{K/k}(K^*).$$

Therefore, by the proposition, C_K^G is generated by the classes of ramified primes in K/k.

Now, let $K = \mathbb{Q}(\sqrt{-d})$ and G be its Galois group over $\mathbb{Q}$. Since $\mu_2(K) = 0$, then $rk(A_n(K))$ is bounded when n goes to infinity [W]. Also, since the class number of the n-th layer $\mathbb{Q}_n$ of the cyclotomic $\mathbb{Z}_2$-extension of $\mathbb{Q}$ is odd, then respectively the groups of invariants and co-invariants $A_n(K)^G$ and $A_n(K)_G$ are isomorphic elementary 2-groups and we have :

$$rk(A_n(K)) = rk(A_n(K)_G) = rk(A_n^G(K)).$$

The following theorem gives an alternative proof of Ferrero's results [F].

Theorem 2.4. Let d be a positive square-free integer and $K = \mathbb{Q}(\sqrt{-d})$, then we have

$$\lambda_2(K) = \sum_{p|d,\, p\neq 2} 2^{-3+v_2(p^2-1)} - 1.$$

Proof. We have from ambiguous class number formulae:

$$rk(A_n(K)) = r_n + 2^n - 1 - rk(E_{\mathbb{Q}_n}/E_{\mathbb{Q}_n} \cap N_{K_n/\mathbb{Q}_n}(K_n^*)),$$

where r_n is the number of finite primes ramified in $K_n/\mathbb{Q}_n$, 2^n is the number of infinite primes of $\mathbb{Q}_n$ and the last term of the formula is the rank of the elementary 2-group $E_{\mathbb{Q}_n} \cap N_{K_n/\mathbb{Q}_n}(K_n^*)$).

Next, we will compute the unit index of the formula. We consider the quadratic number field $k = \mathbb{Q}(i)$. We have 2 is the unique ramified prime in k_∞, and since the class number of k is odd, then $A_n(k)$ is trivial for each integer $n \geq 0$ [W, Proposition 13.22]. On the other hand, the ramified primes in the extension $k_n/\mathbb{Q}_n$ are the 2-adic place and the infinite primes of $\mathbb{Q}_n$. Hence the ambiguous class number formula applied to the quadratic extension $k_n/\mathbb{Q}_n$ gives

$$rk(A_n(k)) = 2^n - rk(E_{\mathbb{Q}_n}/E_{\mathbb{Q}_n} \cap N_{k_n/\mathbb{Q}_n}(k_n^*)).$$

Consequently, we obtain $rk(E_{\mathbb{Q}_n}/E_{\mathbb{Q}_n} \cap N_{k_n/\mathbb{Q}_n}(k_n^*)) = 2^n$. Hence, from Hasse's local-global principle, a unit of $\mathbb{Q}_n$ which is not a norm in the extension $K_n/\mathbb{Q}_n$, is not locally a norm in the same extension at some infinite prime. Moreover each unit of $\mathbb{Q}_n$ is locally a norm at all infinite primes relatively to the extension $\mathbb{Q}_n(\sqrt{d})/\mathbb{Q}_n$. Hence, we conclude that a unit of $\mathbb{Q}_n$ is not a norm in the extension $K_n/\mathbb{Q}_n$ if and only if it is not a norm in the extension $K_n/\mathbb{Q}_n$. In conclusion, we obtain

$$rk(E_{\mathbb{Q}_n}/E_{\mathbb{Q}_n} \cap N_{K_n/\mathbb{Q}_n}(K_n^*)) = 2^n, \text{ hence } rk(A_n(K)) = r_n - 1.$$

Now, let m be the smallest integer such that none of the primes dividing d decompose in the extension $\mathbb{Q}_\infty/\mathbb{Q}_m$ and let s_∞ be the number of ramified primes in the extension $K_\infty/\mathbb{Q}_\infty$. Let n be an integer such that $n \geq m$, then using the ambiguous class number formula and the following two facts:
-the norm map $(A_n(K))_G \longrightarrow (A_m(K))_G$ is surjective, since K_n/K_m is a ramified extension,
- the unit index $[E_{\mathbb{Q}_n}/E_{\mathbb{Q}_n} \cap N_{K_n/\mathbb{Q}_n}(K_n^*)] = 2^m$ is maximal,
we obtain

$$rk(A_n(K)^G) = rk(A_m(K)^G) = s_\infty - 1, \ \forall n \geq m.$$

We have the product $\prod_{i=1}^{s_\infty} \mathcal{L}_{n,i}$ of all primes of K_n ramified in $K_n/\mathbb{Q}_n$ is principal, because it is equal to :

$$j_{0,n}((\sqrt{d})) = \prod_{i=1}^{s_\infty} \mathcal{L}_{n,i},$$

where $(\sqrt{d})$ is the principal ideal of K generated by $\sqrt{d}$. On the other hand, since $rk(E_{\mathbb{Q}_n}/E_{\mathbb{Q}_n} \cap N_{K_n/\mathbb{Q}_n}(K_n^*)) = 2^n$ is maximal and the class number of $\mathbb{Q}_n$ is odd, then by Remark 2.3, for each nonnegative integer

$n \geq r$, $A_n(K)^G$ is generated by the classes of $s_\infty - 1$ ramified primes in the extension $K_n/\mathbb{Q}_n$, hence for each integer $m \geq n+1$, we distinguish between two cases :
(i) if 2 is unramified in K, then $ker(A_n^G(K) \longrightarrow A_m^G(K))$ is trivial.
(ii) if 2 is ramified in K, then $ker(A_n^G(K) \longrightarrow A_m^G(K)) \simeq \mathbb{Z}/2\mathbb{Z}$, because the 2-adic place $\mathcal{L}_n$ of K_n, lifted to K_m is principal : $\mathcal{L}_n o_{K_m} = \mathcal{P}_m^{2^{m-n}}$, where $\mathcal{P}_m$ is the principal 2-adic place of $\mathbb{Q}_n$. Consequently, we have in the case where 2 is unramified, $ker(A_n(K) \longrightarrow A_\infty(K))$ is trivial and in the case where 2 is ramified, $ker(A_n(K) \longrightarrow A_\infty(K))$ is cyclic non trivial. Hence,

$$A_\infty(K) := \cup_{n\geq 0}(j_{n,\infty}(A_n(K))) \simeq \mathbb{Q}_2/\mathbb{Z}_2{}^{\lambda_2(K)},$$

where

$$\lambda_2(K) = s_\infty - 1, \text{ when 2 is unramified } ; \lambda_2(K) = s_\infty - 2 \text{ otherwise.}$$

Consequently, after a simple computation of s_∞, which is the number of ramified primes in the extension $K_m/\mathbb{Q}_m$, we obtain the formula of the theorem. □

In the rest of this article, let F be an abelian totally real number field and let K be the composite of F with an imaginary multiquadratic number field of the form $\mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_r}, \sqrt{-d})$, where r, d and d_i, $i \in \{1, 2, ..., r\}$ are positive square free integers and the set $\{d_1, d_2, ..., d_r, d\}$ is pairewise coprime:

$$K = F(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_r}, \sqrt{-d}).$$

Since K and $K(\sqrt{2})$ have the same cyclotomic $\mathbb{Z}_2$-extension, then we can suppose that $\sqrt{2} \notin K$. To fix ideas, we suppose that the conductor f of F is relatively prime with $\prod_{i=1}^{r} d_i d$.

Denote, for each nonnegative integer $m \leq r$, $K^{(m)} := F(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_m})$ and for $m = 0$, $K^{(0)} = F$, so clearly the totally real subfield of K is $K^+ = K^{(r)} = F(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_r})$. It is well known [W, Theorem 4.12], that

$$[E_K : W_0 E_{K^+}] = 1 \text{ or } 2,$$

where W_0 is the group of roots of unity contained in K. Denote by W_n the group of roots of unity contained in K_n the n-th layer of ths cyclotomic $\mathbb{Z}_2$-extension of K. Put

$$W_n = W^n(2) \oplus W_n',$$

where $W^n(2)$ is the group of roots of unity belonging in K_n, of order a power of 2 and W_n' is the group of roots of unity belonging in K_n, of odd order. Also, taking the inductive limits with respect to the identity maps, denote by

$$W^\infty(2) := \varinjlim W^n(2) = \bigcup_{n\geq 0} W^n(2) \text{ and } W_\infty' = \varinjlim W_n' = \bigcup_{n\geq 0} W_n'.$$

We denote by

$$W_\infty = \varinjlim W_n = W^\infty(2) \oplus W_\infty'.$$

It is assumed that $\sqrt{2} \notin K$, then clearly, in the case where $d = 2$, $W_0 = \{\pm 1\}$. Furthermore in the case where $d = 1$, $\zeta_8 \notin K$ and we have $W_0 = \{\pm 1, \pm i\}$. We have the following easy Lemma on units in the layers of the cyclotomic $\mathbb{Z}_2$-extension of K:

Lemma 2.5. *If $d \notin \{1, 2\}$, then $W^n(2) = \{\pm 1\}$ for each integer $n \geq 0$. If $d \in \{1, 2\}$, then for each positive integer n, $W^n(2) =< \zeta_{2^{n+2}} >$ the sub-group generated by the root of unity $\zeta_{2^{n+2}}$.*

We have the following Lemma giving the unit index of units $[E_{K_n} : W_n E_{K_n^+}]$ in the layers of K_∞.

Lemma 2.6. *In the case where $d \notin \{1,2\}$, we have for each positive integer n, $[E_{K_n} : W_n E_{K_n^+}] = 1$.*

Proof. We have $[E_{K_n} : W_n E_{K_n^+}] = 1$ or 2, where $W_n = \{\pm 1\}$ by Lemma 2.5. Let ε be a unit of K_n such that $\varepsilon \notin W_n E_{K_n^+}$ and $\varepsilon^2 \in W_n E_{K_n^+}$, then write $\varepsilon^2 = \zeta v$ where ζ is a root of unity of odd order m belonging in W_n, v is a unit of K_n^+. Write $\varepsilon^{2m} = v^m$, then $\varepsilon^m = v'\sqrt{v}$, where v' is a unit belonging in K_n^+. In the case where v is positive, then $\sqrt{v}$ belongs in K_n, so in K_n^+, hence ε belongs in $W_n E_{K_n^+}$, which is impossible. In the case where v is negative $v < 0$, then we have $K_n = K_n^+(\sqrt{v})$. On the other hand, it is well known that the quadratic extension $K_n^+(\sqrt{-v})/K_n^+$ is unramified outside 2 [N, Lemma 5.3] and then $K_n^+(i, v)/K_n^+$ is unramified outside 2 and infinite primes. Thus K_n/K_n^+ is unramified outside 2 and infinite primes, which is impossible, because under the hypothesis $(d, d_i) = 1$ for each $i \in \{1, 2, ..., n\}$, we have K_n/K_n^+ is ramified at the primes lying over d. This completes the proof. □

2.2 $H^1(G, E_K)$ and $H^2(G, E_K)$, where $G \simeq \mathbb{Z}/2\mathbb{Z}$

In the following Lemma, we remark that the group W'_∞ is finite:

Lemma 2.7. *The group W'_∞ is finite.*

Proof. Let $\zeta \in W'_\infty$ be a root of unity of odd order m. We will prove that ζ belongs in K. Otherwise, if $\zeta \notin K$, then $K(\zeta)/K$ is a non trivial extension which is ramified at some prime dividing m. On the other hand, the extension K_∞/K is a $\mathbb{Z}_2$-extension, so unramified outside 2-adic primes. This is impossible, because $K(\zeta)$ is a proper subfield of the extension K_∞/K. Hence, each root of unity belonging in W'_∞ must be contained in K. This finishes the Lemma. □

Denote for each positive integer $m \leq r$, the positive integer $t_m := d_m d_{m+1} \cdots d_r d$. Let G_r be the Galois group of the quadratic extension $K/K^{(r-1)}(\sqrt{-t_r})$:

$$G_r = Gal(K/K^{(r-1)}(\sqrt{-t_r})) \simeq Gal(K^{(r)}/K^{(r-1)})$$

and for each nonnegative integer $m < r$, let G_m be the Galois group of the quadratic extension $K^{(m)}(\sqrt{-t_{m+1}})/K^{(m-1)}\sqrt{-t_m}$):

$$G_m = Gal(K^{(m)}(\sqrt{-t_{m+1}})/K^{(m-1)}\sqrt{-t_m})) \simeq Gal(K^{(m)}/K^{(m-1)}).$$

Since the set $\{d_1, d_2, ..., d_r, d\}$ is pairewise coprime, then, all these extensions are unramified outside 2. For each nonnegative integer n, we can regard the unit groups E_{K_n} and $E_{K_n^+}$ of the n-th layers of the cyclotomic $\mathbb{Z}_2$-extensions respectively of K and K^+, as G_r-modules. Also, for each positive integer $m < r$, we can regard the unit groups of the n-th layers of the cyclotomic $\mathbb{Z}_2$-extensions respectively of $K^{(m-1)}(\sqrt{-t_m})$ and $K^{(m-1)}$, as G_{m-1}-modules .

We have the following exact sequence as G_r-modules, of the n-th layers of the cyclotomic $\mathbb{Z}_2$-extensions of K^+ and K.

$$1 \longrightarrow E_{K_n^+} \longrightarrow E_{K_n} \longrightarrow E_{K_n}/E_{K_n^+} \longrightarrow 1,$$

and for each positive integer $m < r$, let the following exact sequence as G_{m-1}-modules, of the n-th layers of the cyclotomic $\mathbb{Z}_2$-extensions of $K^{(m)}$ and $K^{(m)}(\sqrt{-t_{m+1}})$.

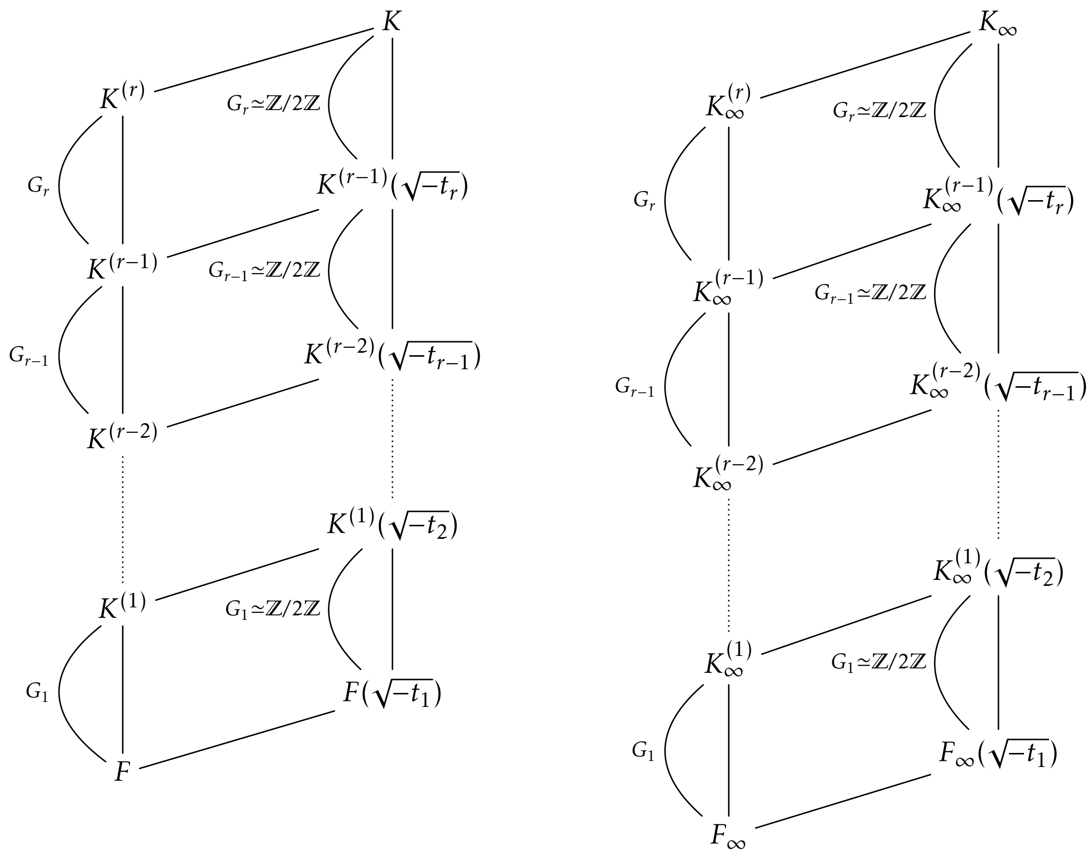
$$1 \longrightarrow E_{K_n^{(m)}} \longrightarrow E_{K_n^{(m)}(\sqrt{-t_{m+1}})} \longrightarrow E_{K_n^{(m)}(\sqrt{-t_{m+1}})}/E_{K_n^{(m)}} \longrightarrow 1,$$

Taking the direct limit with respect to the identity maps, of these exact sequences, we obtain the following exact sequences:

$$1 \longrightarrow E_{K_\infty^+} \longrightarrow E_{K_\infty} \longrightarrow E_{K_\infty}/E_{K_\infty^+} \longrightarrow 1. \qquad (*)$$

$$1 \longrightarrow E_{K_\infty^{(m)}} \longrightarrow E_{K_\infty^{(m)}(\sqrt{-t_{m+1}})} \longrightarrow E_{K_\infty^{(m)}(\sqrt{-t_{m+1}})}/E_{K_\infty^{(m)}} \longrightarrow 1.$$

We have the following diagrams



In the following, we will compute the cohomology groups of units H^1 and H^2 :

$$H^i(G_r, E_{K_\infty}) \text{ and } H^i(G_m, E_{K_\infty^{(m)}(\sqrt{-t_{m+1}})}), \text{ for } i = 1, 2 \text{ and } m < r.$$

We will distinguish between cases where K contains W_∞ or not.

2.2.1 The case $d \notin \{1, 2\}$

Proposition 2.8. *Suppose $d \notin \{1, 2\}$, then we have*

$$H^i(G_r, E_{K_\infty^+}) = H^i(G_r, E_{K_\infty}),$$

for each $i = 1, 2$.

Proof. Since $d \notin \{1, 2\}$, then by Lemma 2.5, $W^\infty(2) = \{\pm 1\}$ and we obtain

$$E_{K_\infty}/E_{K_\infty^+} \simeq W'_\infty,$$

which is a finite 2-torsion free group. On the other hand, we have $G_r \simeq \mathbb{Z}/2\mathbb{Z}$, so the groups $H^i(G_r, E_{K_\infty}/E_{K_\infty^+})$, $i = 1,2$ are 2-torsion groups, hence trivial groups. This completes the proof. □

Corollary 2.9. *For each positive integer $m < r$, we have*

$$H^i(G_m, E_{K_\infty^{(m)}}) = H^i(G_m, E_{K_\infty^{(m)}(\sqrt{-t_{m+1}})}),$$

for each $i = 1,2$.

Proof. This follows from the fact that each positive integer t_m is different from 1 and 2. □

2.2.2 The case $d \in \{1,2\}$

From Lemma 2.5 , K_∞ contains all roots of unit of order a power of 2. For each nonnegative integer n, it is well known that the unit index $[E_{K_n} : W_n E_{K_n}^+] = 1$ or 2. In the following Lemma, we will determine the exact value of the unit index $[E_{K_\infty} : W_\infty E_{K_\infty}^+]$.

Lemma 2.10. *We have $[E_{K_\infty} : W_\infty E_{K_\infty}^+] = 1$.*

Proof. Suppose that $[E_{K_\infty} : W_\infty E_{K_\infty}^+] = 2$, then there exists a unit u belonging in K_∞ such that

$$u \notin W_\infty E_{K_\infty}^+ \text{ and } u^2 \in W_\infty E_{K_\infty^+}.$$

Let $\zeta \in W_\infty$ and v be a unit of K_∞^+ such that $u^2 = \zeta v$. Write $\zeta = \zeta_{2^r}\zeta_m$, where $\zeta_{2^r} \in W^\infty(2)$ and $\zeta_m \in W'_\infty$. The integer m is odd, so write $m = 2k+1$. Then we obtain

$$u^2 = (\zeta_{2^{r+1}})^2(\zeta_m^{k+1})^2 v,$$

hence, $\sqrt{v}$ belongs in E_{K_∞}, so in $E_{K_\infty^+}$, thus we conclude that $u \in W_\infty E_{K_\infty}^+$, which is impossible. This finishes the proof of the lemma. □

Next, for each $i \in \{1,2,\ldots,r\}$, we will denote respectively $H^1(*)$ and $H^2(*)$ instead of $H^1(G_i,*)$ and $H^2(G_i,*)$.

We take the exact sequence :

$$1 \longrightarrow E_{K_\infty^+} \longrightarrow E_{K_\infty} \longrightarrow E_{K_\infty}/E_{K_\infty^+} \longrightarrow 1. \qquad (*)$$

Then, we obtain the cohomology exact sequence of G_r-modules

$$\begin{array}{ccccc} H^1(E_{K_\infty^+}) & \longrightarrow & H^1(E_{K_\infty}) & \longrightarrow & H^1(E_{K_\infty}/E_{K_\infty^+}) \\ \uparrow & & & & \downarrow \\ H^2(E_{K_\infty}/E_{K_\infty^+}) & \longleftarrow & H^2(E_{K_\infty}) & \longleftarrow & H^2(E_{K_\infty^+}) \end{array}$$

Theorem 2.11. We have $H^2(E_{K_\infty}) \simeq H^2(E_{K_\infty^+})$ and the map $H^1(E_{K_\infty^+}) \longrightarrow H^1(E_{K_\infty})$ is onto of kernel isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

Proof. From Lemma 2.10, we have

$$E_{K_\infty}/E_{K_\infty^+} = \overline{W^\infty(2)} \oplus \overline{W'_\infty}$$

where $\overline{W^\infty(2)}$ and $\overline{W'_\infty}$ are the sub-groups of $E_{K_\infty}/E_{K_\infty^+}$, containing respectively the classes of elements of $W^\infty(2)$ and W'_∞. Clearly, such groups are acting on G_r as G_r-modules. We have

$$H^i(E_{K_\infty}/E_{K_\infty^+}) = H^i(\overline{W^\infty(2)}) \oplus H^i(\overline{W'_\infty}), \text{ for } i = 1,2.$$

Since $\overline{W'_\infty}$ is a 2-torsion free group and $G \simeq \mathbb{Z}/2\mathbb{Z}$, then, we obtain :

$$H^i(\overline{W'_\infty}) = 1, \text{ for } i = 1,2.$$

It remains to compute $H^i(\overline{W^\infty(2)})$ for $i = 1,2$. We will calculate $ker(N)$, $Im(\sigma - 1)$, $ker(\sigma - 1)$ and $Im(N)$ with respect to the group $\overline{W^\infty(2)}$, where σ is a generator of G_r and $N = 1 + \sigma$.

We have

$$ker(N) = \{\overline{\zeta} \in \overline{W^\infty(2)} \mid \overline{N(\zeta)} = \overline{1}\} = \{\overline{\zeta} \in \overline{W^\infty(2)} \mid N(\zeta) \in E_{K_\infty^+}\}.$$

For each root of unity $\zeta \in W^\infty(2)$, we have $N\zeta$ is a root of unity belonging in $K_\infty^{(r-1)}(\sqrt{-t_r})$. Since $t_r \notin \{1,2\}$, then by Lemma 2.5, $i \notin K_\infty^{(r-1)}(\sqrt{-t_r})$, hence $N(\zeta) = \pm 1$. Therefore, $Ker(N)$ is the whole group $\overline{W^\infty(2)}$:

$$ker(N) = \overline{W^\infty(2)}.$$

$$Im(\sigma - 1) = \{\overline{\zeta}^{-1}\overline{\sigma(\zeta)} \mid \zeta \in W^\infty(2)\} = \{\frac{\overline{N(\zeta)}}{\overline{\zeta^2}} \mid \zeta \in W^\infty(2)\} = \{\frac{1}{\overline{\zeta^2}} \mid \zeta \in W^\infty(2)\}.$$

Let $\zeta \in W^\infty(2)$ of order 2^n, then it is a square of a root of unity of order 2^{n+1}, furtheremore we see that $Im(\sigma - 1)$ is the whole group $\overline{W^\infty(2)}$:

$$Im(\sigma - 1) = \overline{W^\infty(2)}.$$

Consequently, we conclude that

$$H^1(E_{K_\infty}/E_{K_\infty^+}) = \frac{ker(N)}{Im(\sigma - 1)} \text{ is trivial .}$$

Now, we compute $ker(\sigma - 1)$:

$$ker(\sigma - 1) = \{\overline{\zeta} \in \overline{W^\infty(2)} \mid \sigma(\overline{\zeta}) = \overline{\zeta}\} = \{\overline{\zeta} \mid \sigma(\zeta)\zeta^{-1} \in E_{K_\infty^+}\}$$

Also, since

$$\sigma(\zeta)\zeta^{-1} = \frac{N(\zeta)}{\zeta^2}$$

and $N\zeta$ belongs in $K_\infty^{(r-1)}(\sqrt{-t_r})$, so as before $N\zeta \in \{\pm 1\}$, then we obtain

$$ker(\sigma - 1) = \{\overline{\zeta} \in \overline{W^\infty(2)} \mid \zeta^2 \in E_{K_\infty^+}\} = <\overline{i}>,$$

a sub-group generated by $\overline{i}$, hence

$$ker(\sigma - 1) \simeq \mathbb{Z}/2\mathbb{Z}.$$

We compute $Im(N)$. We have $Im(N) = \{\overline{N(\zeta)} \in \overline{W^\infty(2)} \mid \zeta \in W^\infty(2)\}$.
For each root of unity $\zeta \in W^\infty(2)$, we have $N(\zeta) \in \{\pm 1\}$, then

$$Im(N) = \{\bar{1}\}.$$

Thus, we find

$$H^2(\overline{W^\infty(2)}) = ker(\sigma - 1)/Im(N) \simeq \mathbb{Z}/2\mathbb{Z}.$$

Hence we conclude, that

$$H^2(E_{K_\infty}/E_{K_\infty^+}) \simeq \mathbb{Z}/2\mathbb{Z}.$$

Now, we will see the triviality of the map

$$H^2(E_{K_\infty}) \longrightarrow H^2(E_{K_\infty}/E_{K_\infty^+}).$$

Let $\bar{u} \in H^2(E_{K_\infty})$, so u is fixed by σ, then u belongs in $K_\infty^{(r-1)}(\sqrt{-t_r})$. Since $t_r \notin \{1,2\}$, then by Lemma 2.6, we can write $u = \zeta v$, where $\zeta \in W_\infty$ and v is a unit of $K_\infty^{(r-1)}$. Clearly, we have $v \in E^+_{K_\infty}$, then its image by the map is trivial. On the other hand, since $i \notin K_\infty^{(r-1)}(\sqrt{-t_r})$, then $\zeta = \pm 1$ or ζ belongs in W'_∞. Suppose that ζ belongs W'_∞ of order an odd integer m. So, if we put $m = 2k+1$, then we find $N(\zeta^{k+1}) = \zeta^{2k+2} = \zeta$. Hence ζ belongs in $Im(N)$, and its image in the last map is trivial. We conclude that the map is trivial. Consequently, we obtain

$$H^2(E_{K_\infty}) \simeq H^2(E^+_{K_\infty}).$$

Also the map $H^1(E_{K_\infty^+}) \longrightarrow H^2(E_{K_\infty})$ is onto with kernel isomorphic to $\mathbb{Z}/2\mathbb{Z}$. As a conclusion, we have the following diagram

$$\begin{array}{ccccc}
H^1(E_{K_\infty^+}) & \xrightarrow{\ \beta\ } & H^1(E_{K_\infty}) & \xrightarrow{\ f\ } & H^1(E_{K_\infty}/E_{K_\infty^+}) = 1 \\
\uparrow{\scriptstyle \alpha} & & & & \downarrow{\scriptstyle \gamma} \\
H^2(E_{K_\infty}/E_{K_\infty^+}) \simeq \mathbb{Z}/2\mathbb{Z} & \longleftarrow & H^2(E_{K_\infty}) & \xleftarrow{\ \simeq\ } & H^2(E_{K_\infty^+})
\end{array}$$

□

2.3 The Iwasawa λ_2 invariant of K

Denote for each $i \in \{1,2\}$ and $j \in \{1,2,...,r\}$, $h_i^{(j)} = dim_{\mathbb{F}_2}(H^i(G_j, E_{K_\infty^{(j)}}))$. Now, we state the theorem concerning the Iwasawa λ_2 invariant of the cyclotomic $\mathbb{Z}_2$-extension of K.

Theorem 2.12. For each integer $r \geq 1$, we have

$$(i) \quad \lambda_2(K) = 2^r \lambda_2(F(\sqrt{-t_1})) + \lambda_2(F(\sqrt{d_1},...,\sqrt{d_r})) - 2^r \lambda_2(F) \sum_{i=1}^{r} 2^{r-i} f_i \text{ if } d \notin \{1,2\}$$

$$(ii) \quad \lambda_2(K) = 2^r \lambda_2(F(\sqrt{-t_1})) + \lambda_2(F(\sqrt{d_1},...,\sqrt{d_r})) - 2^r \lambda_2(F) - \sum_{i=1}^{r} 2^{r-i} f_i + 1 \text{ if not },$$

where f_i is the number of ramified primes in the totally real extension $K_\infty^{(i)}/K_\infty^{(i-1)}$.

Proof. The proof of (i) and (ii) are similar. We will give the proof of (i). Denote for each positive integer m, $h_1^{(m)}$ and $h_2^{(m)}$ respectively the dimension of the cohomology groups $H^1(G_m, E_{K_\infty^{(m)}})$ and $H^2(G_m, E_{K_\infty^{(m)}})$.
We give the proof of the theorem, by induction on r. For $r = 1$, we have $K = F(\sqrt{d_1}, \sqrt{-d})$, then applying Theorem 2.1 to the quadratic extension $K_\infty/F_\infty(\sqrt{-t_1})$, where $t_1 = d_1 d$, we obtain:

$$\lambda_2(K) = 2\lambda_2(F_\infty(\sqrt{-t_1})) + s_1 + h_2^{(1)} - h_1^{(1)},$$

here $s_1 = 0$, because the extension $K_\infty/F_\infty(\sqrt{-t_1})$ is unramified outside 2.
By Corollary 2.9, we have $H^i(G_m, E_{K_\infty^{(m)}}) = H^i(G_m, E_{K_\infty^{(m)}(\sqrt{-t_{m+1}})})$, for $i = 1, 2$. Also, applying Theorem 2.1 to the extension K_∞^+/F_∞, we obtain

$$h_2^{(1)} - h_1^{(1)} = \lambda_2(K^+) - 2\lambda(F) - f_1.$$

Therefore, we find

$$\lambda_2(K) = 2\lambda_2(F(\sqrt{-t_1})) + \lambda_2(K^+) - 2\lambda(F) + s_1 - f_1.$$

Now suppose that the formula of the theorem is satisfied for each positive integer $n \leq r-1$ and we will prove the formula for $n = r$.
In the quadratic extension $K_\infty/K_\infty^{(r-1)}(\sqrt{-t_r})$, the Riemann Hurwitz formula yields:

$$\lambda_2(K) = 2\lambda_2(K^{(r-1)}(\sqrt{-t_r})) + s_r + h_2^r - h_1^r, \tag{1}$$

also $s_r = 0$, because the extension $K_\infty/K_\infty^{(r-1)}(\sqrt{-t_r})$ is unramified outside 2 and by Corollary 2.11, we obtain

$$h_2^r - h_1^r = \lambda_2(K^+) - 2\lambda(K_\infty^{(r-1)}) - t_r, \tag{2}$$

hence using the Induction Hypothesis, we find

$$\lambda_2(K^{(r-1)}(\sqrt{-t_r})) = 2^{r-1}\lambda_2(F(\sqrt{-t_1})) + \lambda_2(F(\sqrt{d_1}, ..., \sqrt{d_{r-1}})) - 2^{r-1}\lambda_2(F) - \sum_{i=1}^{r-1} 2^{r-1-i} f_i. \tag{3}$$

Hence, combining the formulas (1), (2) and (3), we conclude the formula of the theorem:

$$\lambda_2(K) = 2^r\lambda_2(F(\sqrt{-t_1})) + \lambda_2(F(\sqrt{d_1}, ..., \sqrt{d_r})) - 2^r\lambda_2(F) - \sum_{i=1}^{r} 2^{r-i} f_i.$$

To find (ii) of the theorem, we will use Theorem 2.11. □

3 Application to the $\mathbb{Z}_2$-torsion of some imaginary biquadratic number fields

In this section, as mentioned in the introduction, we will determine a family of imaginary biquadratic number fields K, such that the $\mathbb{Z}_2$-torsion of the abelian unramified Iwasawa module of the cyclotomic $\mathbb{Z}_2$-extension of K is of arbitrary large rank.

Let N be a positive integer, ℓ and ℓ' be distinct prime numbers such that $\ell \equiv 1 \pmod{2^{N+2}}$, $\ell' \equiv 5 \pmod 8$ and $(\frac{\ell}{\ell'}) = -1$. Put $k = \mathbb{Q}(\sqrt{\ell\ell'})$ and $K = k(i) = \mathbb{Q}(i, \sqrt{\ell\ell'})$.

Clearly ℓ splits completely in $\mathbb{Q}_N$ and ℓ' is inert in $\mathbb{Q}_1$, so in $\mathbb{Q}_\infty$. Also, since $(\frac{\ell}{\ell'}) = -1$, then $A_0(K) \simeq \mathbb{Z}/2\mathbb{Z}$ and generated by the ℓ-adic place of k.
In [O, Lemma 3], M. Ozaki proves that there exist an infinite family of such prime numbers ℓ such that the rank of the elementary 2-group $E_{\mathbb{Q}_N}/E_{\mathbb{Q}_N} \cap N_{k_N/\mathbb{Q}_N} k_N^*$ is :

$$rank(E_{\mathbb{Q}_N}/E_{\mathbb{Q}_N} \cap N_{k_N/\mathbb{Q}_N} k_N^*) = 2^N - 2^{N-1} + 1.$$

Next, the positive integer N can be supposed such that $\mathbb{Q}_N$ is the decomposition field of ℓ in $\mathbb{Q}_\infty$. In terms of valuation, suppose that $v_2(\ell - 1) = 2^{N+2}$. Hence, the number of ℓ-adic primes in $\mathbb{Q}_\infty$ is equal to 2^N. Also in [O, proof of Lemma 3], we have

$$rank_2(X_\infty(k)) \geq rank(A_N(k) \geq 2^{N-1} - 1.$$

Since $rank(A_N(k)) < 2^N$, then $A(k)$ capitulates in k_∞ [O-Y, Lemma 7] or [O, Lemma 1]. Hence, since k contains a unique 2-adic place, then by [G], we have $\lambda_2(k) = 0$, then $X_\infty(k)$ is finite. In the following proposition, we will give the exact value of the 2-rank of $X_\infty(k)$, precisely we find that the last two large inequalities become equalities.

Proposition 3.1. *We have*

$$rank_2(X_\infty(k)) = rank_2(A_N(k)) = 2^{N-1} - 1 \text{ and } X_\infty(k) \simeq A_N(k).$$

Proof. We have the class number of $\mathbb{Q}_N$ is odd, and the general ambiguous class number formula giving the rank of $A_n(K)$ in terms of the rank of $E_{\mathbb{Q}_N}/E_{\mathbb{Q}_N} \cap N_{k_N/\mathbb{Q}_N} k_N^*$ and the number t_N of ramified primes in the quadratic extension $k_N/\mathbb{Q}_N$ gives:

$$rank(A_N(k)) = t_N - 1 - rank(E_{\mathbb{Q}_N}/E_{\mathbb{Q}_N} \cap N_{k_N/\mathbb{Q}_N} k_N^*).$$

It is easy to see that $t_N = 2^N + 1$ and as mentionned before, $rank(E_{\mathbb{Q}_N}/E_{\mathbb{Q}_N} \cap N_{k_N/\mathbb{Q}_N} k_N^*) = 2^N - 2^{N-1} + 1$, then $rank(A_N(K)) = 2^{N-1} - 1$. Let γ be a topological generator of $Gal(K_\infty/K)$ and consider the commutative diagram [I-1, Theorems 6 and 7]:

$$\begin{array}{ccc} A(k) & \xrightarrow{\sim} & X_\infty(k)/w_0 X_\infty(k) \\ \downarrow & & \downarrow \nu_n \\ A_N(k) & \xrightarrow{\sim} & X_\infty(k)/w_n X_\infty(k) \end{array}$$

where $w_0 = T = \gamma - 1$, $w_n = \gamma^{p^n} - 1 = (1+T)^{p^n} - 1$, $\nu_n = \frac{w_n}{w_0}$ and $\Lambda = \mathbb{Z}_p[[T]]$ the local ring of maximal ideal (p, T).

Since $A(k)$ capitulates in k_N, then the left vertical map is trivial, thus $\nu_n X_\infty(k) \subset w_n X_\infty(k)$. Hence, we obtain

$$w_n X_\infty(k) = \nu_n X_\infty(k) = w_0(w_n X_\infty(k)).$$

On the other hand, since $w_n X_\infty(k)$ is a finitely generated Λ-module and w_0 is contained in (p, T), then by Nakayama's Lemma we obtain $w_n X_\infty(k) = 0$, hence $X_\infty(k) \simeq A_N(k)$. □

Now, we are ready to prove the following theorem, giving the exact values of the rank of $X_\infty(K)$ and $\mathrm{X}^0_\infty(K)$. Also, proving that the last group is of arbitrary large rank.

Theorem 3.2. We have the structure of the Iwasawa module

$$X_\infty(K) \simeq \mathbb{Z}_2^{\lambda(K)} \oplus \mathrm{X}_\infty^0(K),$$

where $\lambda_2(K) = 2^N$ and $\mathrm{X}_\infty^0(K)/2 \simeq (\mathbb{Z}/2\mathbb{Z})^{2^{N-1}-1}$.

Proof. From Theorem 2.12, we have

$$\lambda_2(K) = 2\lambda_2((\mathbb{Q}\sqrt{-\ell\ell'})) + \lambda_2(k) - t + 1,$$

where $t = 2^N + 1$ is the number of the primes outside 2 ramified in the extension $k_\infty/\mathbb{Q}_\infty$. From [F], $\lambda_2((\mathbb{Q}\sqrt{-\ell\ell'})) = 2^N$ and by [O], $\lambda_2(k) = 0$, hence, we obtain

$$\lambda_2(K) = 2^{N+1} - (2^N + 1) + 1 = 2^N.$$

On the other hand, since 2 is unramified in $K^+ = k$ and K contains a unique 2-adic place, then by [A, Exampla 2.8], the $\mathbb{Z}_2$-torsion of K is equal to the $\mathbb{Z}_2$-torsion of k. Since $\lambda_2(k) = 0$, then $\mathrm{X}_\infty^0(K) \simeq X_\infty(k)$. hence, combinng theses results with Proposition 3.1, we have the theorem. □

References

[A] M. Atsuta, *Finite Λ-submodules of Iwasawa modules for a CM-field for $p = 2$.* Journal de Théorie de nombres de Bordeaux, Tome 30 (2018), pp. 1017-1035.

[F] B. Ferrero, *The cyclotomic $\mathbb{Z}_2$-extension of imaginary quadratic fields.* Amer. J. Math. 102 (1980), no. 3, 447-459.

[F-K-O-T] T. Fukuda, K. Komatsu, M. Ozaki and H. Taya, *On the Iwasawa λ_p-invariants of relative cyclic extensions of degree p.* Tokyo. J. Math. 20 (1997), no. 2, 475-480.

[F-W] B. Ferrero and L. C. Washington, *The Iwasawa invariant μ_p vanishes for abelian number fields.* Ann. of Math. (2) 109 (1979), no. 2, 377-395.

[G] R. Greenberg, *On the Iwasawa invariants of totally real number fields.* Amer. J. Math. 98 (1976), no. 1, 263-284.

[I-1] K. Iwasawa, *On $\mathbb{Z}_l$-extensions of algebraic number fields,* Ann. of Math. (2) 98 (1973), 246-326.

[I-2] K. Iwasawa, *Riemann-Hurwitz formula and p-adic Galois representations for number fields,* Tohoku Math. J (33), no. 2, (1981), 263-288.

[K] Y. Kida, *Cyclotomic $\mathbb{Z}_2$-extensions of J-fields,* J. Number Theory. 14 (1982), 340-352.

[L] S. Lang,: *Algebraic Number Theory.* Addison-Wesley, Reading (1970)

[N] J. Neukirch, *Class field theory,* Grundlehren der mathematischen Wissenschaften, 280, Springer (1986), 1-142.

[O] M. Ozaki, *Construction of real abelian fields of degree p with $\lambda_p = \mu_p = 0$,* Int. J. Open Problems Compt. Math. 2 (2009), no. 3, 342-351.

[O-Y] M. Ozaki, G. Yamamoto, *Iwasawa λ_3-invariants of certain cubic fields.* Acta Arithmetica. 4 (2001), 387-398.

[S] J. Schettler, *An alternative approach to Kida and Ferrero's computations of Iwasawa λ-invariants.* J. Number Theory. 138 (2014), 84-96.

[W] L. C. Washington, *Introduction to Cyclotomic Fields,* Second edition. Graduate Texts in Mathematics, 83. Springer-Verlag, New York, 1997. xiv+487 pp.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 108-123
ISSN: 2820-7114

Title :

π-dual Baer Modules and π-dual Baer Rings

Author(s):

Derya Keskin Tütüncü & Rachid Tribak

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 108–123
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco **ISSN: 2820-7114**

π-dual Baer Modules and π-dual Baer Rings

Derya Keskin Tütüncü[1] and Rachid Tribak[2]
[1] Department of Mathematics, Hacettepe University, 06800 Beytepe, Ankara, Turkey
e-mail: keskin@hacettepe.edu.tr
[2] *Centre Régional des Métiers de l'Education et de la Formation (CRMEF-TTH)-Tanger, Avenue My Abdelaziz, Souani, B.P. 3117, Tangier, Morocco*
e-mail: tribak12@yahoo.com

Communicated by Mohammed Tamekkante
(Received 08 November 2022, Revised 12 February 2023, Accepted 20 February 2023)

Abstract. Let R be a ring and let M be an R-module with $\mathbf{S} = \mathrm{End}_R(M)$. A submodule N of M is said to be *projection invariant* in M (denoted $N \trianglelefteq_p M$) if $eN \subseteq N$ for all $e = e^2 \in \mathbf{S}$. We call M *π-dual Baer*, if for each $N \trianglelefteq_p M$ there exists $e^2 = e \in \mathbf{S}$ such that $\{f \in \mathbf{S} \mid f(M) \subseteq N\} = e\mathbf{S}$. A characterization of π-dual Baer modules is provided. We show that the class of π-dual Baer modules lies strictly between the classes of dual Baer modules and quasi-dual Baer modules. It is also shown that in general, the class of π-dual Baer modules is neither closed under direct sums nor closed under direct summands. The structure of π-dual Baer modules over Dedekind domains is completely determined. We conclude the paper by studying right π-dual Baer rings. We call a ring R *right π-dual Baer* if the right R-module R_R is right π-dual Baer. A characterization of this class of rings is provided. We also investigate the transfer between a base ring R and many of its extensions (for example, full matrix rings over R or $R[x]$ or $R[[x]]$). In addition, we characterize the 2-by-2 generalized triangular right π-dual Baer matrix rings.

Key Words: dual Baer module; quasi-dual Baer module; π-dual Baer module; endomorphism rings; projection invariant submodule.
2010 MSC: Primary 16D10, 16S50; Secondary 16D80.

Dedicated to the memory of Professor Muhammad Zafrullah

1 Introduction

Throughout this paper R will always be an associative ring with unity and any module will be a unital right R-module unless stated otherwise. Let M be an R-module. By $\mathbf{S} = \mathrm{End}_R(M)$ and $\mathbf{I}_M$, we denote the endomorphism ring of the module M and the subring of $\mathbf{S}$ generated by the idempotents of $\mathbf{S}$, respectively. For a ring R, we use $\mathbf{I}(R)$ to denote the subring of R generated by idempotents. The notations $N \subseteq M$, $N \leq M$ and $N \leq_d M$ mean that N is a subset of M, N is a submodule of M and N is a direct summand of M, respectively. Let $N \leq M$. Then N is called a *fully invariant* submodule of M (denoted $N \trianglelefteq M$) if $f(N) \subseteq N$ for all $f \in \mathbf{S}$, and N is called a *projection invariant* submodule of M (denoted $N \trianglelefteq_p M$) if $e(N) \subseteq N$ for all $e^2 = e \in \mathbf{S}$. Note that every fully invariant submodule is projection invariant and the projection invariant submodules of a module M form a complete sublattice of the lattice of submodules of M. One may observe that if N is fully (projection) invariant in M, then there exists a ring homomorphism $\alpha : \mathbf{S} \to \mathrm{End}_R(N)$ ($\beta : \mathbf{I}_M \to \mathbf{I}_N$) defined by $\alpha(h) = h|_N$ ($\beta(h) = h|_N$) for all $h \in \mathbf{S}$ ($h \in \mathbf{I}_M$) (see [7]). Note that a right ideal I of a ring R is called *projection invariant* in R_R (denoted $I_R \trianglelefteq_p R_R$) if $eI \subseteq I$ for all $e^2 = e \in R$. Moreover, fully invariant right ideals of R coincide with two-sided ideals of R.

The notions of Baer modules and quasi-Baer modules were introduced in 2004 (see [16]). In 2010 (see [13]), Keskin Tütüncü and Tribak dualized the notion of Baer modules. A module M is said to be *dual Baer* if for every submodule N of M, there exists an idempotent $e \in \mathbf{S}$ such that $\{f \in \mathbf{S} \mid f(M) \subseteq N\} = e\mathbf{S}$. The right ideal $\{f \in \mathbf{S} \mid f(M) \subseteq N\}$ of $\mathbf{S}$ will be denoted by $D_{\mathbf{S}}(N)$. For a subset X in $\mathbf{S}$ and a submodule N of M, let $X(N)$ denote the submodule $\sum_{f \in X} f(N)$ of M. Note that a module M is dual Baer if and only if for every subset A of $\mathbf{S}$, $A(M)$ is a direct summand of M if and only if for every right ideal A of $\mathbf{S}$, $A(M)$ is a direct summand of M (see [13, Theorem 2.1]). In 2013 (see [3]), Amouzegar and Talebi introduced the notion of quasi-dual Baer modules by dualizing the notion of quasi-Baer modules. A module M is said to be *quasi-dual Baer* if for every fully invariant submodule N of M, there exists an idempotent $e \in \mathbf{S}$ such that $D_{\mathbf{S}}(N) = e\mathbf{S}$. In [18], the authors continued the study of quasi-dual Baer modules. They showed that a module M is quasi-dual Baer if and only if for every left ideal I of $\mathbf{S}$, $I(M)$ is a direct summand of M (see [18, Proposition 2.4]).

In 2020 (see [7]), Birkenmeier, Kara and Tercan introduced the notion of π-endo Baer (π-e.Baer for short) modules. According to [7, Definition 3.3], a module M is called *π-e.Baer*, if for each $\emptyset \neq X \subseteq M$ such that $j(X) \subseteq X$ for all $j^2 = j \in \mathbf{S}$ there exists $e^2 = e \in \mathbf{S}$ such that $l_{\mathbf{S}}(X) = \{s \in \mathbf{S} \mid s(X) = 0\} = \mathbf{S}e$. By [7, Lemma 3.4], a module M is π-e.Baer if and only if for each $N \trianglelefteq_p M$, there exists $f^2 = f \in \mathbf{S}$ such that $l_{\mathbf{S}}(N) = \mathbf{S}f$ if and only if for each ${}_{\mathbf{S}}Y \trianglelefteq_p {}_{\mathbf{S}}\mathbf{S}$, there exists $e^2 = e \in \mathbf{S}$ such that $\bigcap_{g \in Y} \operatorname{Ker} g = eM$. Later in 2021, this notion was dualized by Kara (see [12]) by introducing the following definition.

Definition 1.1. A module M is called *dual π-endo Baer*, if for each $N \trianglelefteq_p M$, there exists $e^2 = e \in \mathbf{S}$ such that $D_{\mathbf{S}}(N) = e\mathbf{S}$.

Note that in [4] and [7], the authors used the terminology *endomorphism Baer* module, denoted briefly by e-Baer, for the Baer modules defined by Rizvi and Roman in [16]. The rings R for which the right R-module R_R is π-e.Baer were studied in 2018 (see [6]). It was shown in [6, Proposition 2.1] that the π-e.Baer property is left-right symmetric for any ring R. Then (right) π-e.Baer rings were called *π-Baer* rings in [6, Definition 2.2].

Motivated by all these research works ([3], [7], [12] and [13]), we continue to study dual π-endo Baer modules, but under the name *π-dual Baer modules* in this paper. We also study π-dual Baer rings. A ring R is said to be *right* (left) *π-dual Baer* if the right (left) R-module R_R (${}_RR$) is π-dual Baer. The aim of this paper is to show that some results of π-e.Baer modules and π-Baer rings have corresponding duals for π-dual Baer modules and right π-dual Baer rings. In addition, we will obtain the π-dual Baer analogues of certain results appearing in [6] or in [18].

Section 2 is devoted to the study of some basic properties of π-dual Baer modules. We provide some equivalent formulations of being a π-dual Baer module (Theorem 2.4). We show that for an indecomposable $\mathbb{Z}$-module M, M is dual Baer if and only if M is π-dual Baer if and only if M is quasi-dual Baer if and only if $M \cong \mathbb{Q}$ or $M \cong \mathbb{Z}(p^\infty)$ or $M \cong \mathbb{Z}/p\mathbb{Z}$, where p is a prime number (Proposition 2.12). We construct some examples showing that the π-dual Baer condition is strictly between the dual Baer and quasi-dual Baer conditions (Example 2.14).

In Section 3, we investigate direct sums and direct summands of π-dual Baer modules. We first provide examples showing that, in general, the π-dual Baer condition is neither preserved under direct sums nor preserved under direct summands (Examples 3.1 and 3.5). Then we prove that any projection invariant direct summand of a π-dual Baer module inherits the property (Theorem 3.6). It is also shown that if a module $M = \oplus_{i \in I} M_i$ such that $M_i \trianglelefteq_p M$ for all $i \in I$, then M is π-dual Baer if and only if M_i is π-dual Baer for all $i \in I$ (Theorem 3.8). We conclude this section by describing the structure of π-dual Baer modules over Dedekind domains (Theorem 3.15).

In Section 4, we deal with right π-dual Baer rings. We show that the class of right π-dual Baer rings lies strictly between the classes of dual Baer rings and quasi-dual Baer rings (Remark 4.14). We provide a characterization of right π-dual Baer rings (Theorem 4.15). In addition, we study the transfer of the right π-dual Baer property between a base ring R and several extensions. For example, full matrix rings over R or $R[x]$ or $R[[x]]$ (see Propositions 4.19 and 4.21, Examples 4.20 and 4.22).

We conclude the paper by characterizing the 2-by-2 generalized triangular right π-dual Baer matrix rings (Theorem 4.24).

Throughout this paper, by $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{Z}(p^\infty)$ we denote the ring of integer numbers, ring of rational numbers and the Prüfer p-group, respectively where p is a prime number.

2 Some results on π-dual Baer modules

Definition 2.1. A module M is called *π-dual Baer*, if for each $N \trianglelefteq_p M$, there exists $e^2 = e \in \mathbf{S}$ such that $D_{\mathbf{S}}(N) = e\mathbf{S}$.

Example 2.2. (i) Clearly, every semisimple module is π-dual Baer.

(ii) Let M be an indecomposable module. Then 0 and 1 are the only idempotents of $\mathbf{S}$. This implies that all submodules of M are projection invariant. Therefore M is dual Baer if and only if M is π-dual Baer.

(iii) Let R be a commutative ring. Using [13, Corollary 2.9], we see that the R-module R is dual Baer if and only if it is π-dual Baer if and only if it is quasi-dual Baer if and only if R is semisimple.

Recall that an idempotent $e \in R$ is called *left semicentral* if $xe = exe$ for all $x \in R$. The set of left semicentral idempotents of R is denoted by $S_l(R)$. We begin with the following lemma which is taken from [12, Lemmas 2.1 and 2.2] and [7, Lemma 3.1(iii)]. This lemma will be used throughout the paper.

Lemma 2.3. *Let M be a module with $\mathbf{S} = \mathrm{End}_R(M)$.*

(i) *If $N \trianglelefteq_p M$, then $D_{\mathbf{S}}(N) \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$.*

(ii) *If $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$, then $I(M) \trianglelefteq_p M$.*

(iii) *If I is a right ideal of $\mathbf{S}$, then $D_{\mathbf{S}}(I(M))(M) = I(M)$.*

(iv) *If $N \leq M$, then $D_{\mathbf{S}}(D_{\mathbf{S}}(N)(M)) = D_{\mathbf{S}}(N)$.*

(v) *Let $e = e^2 \in \mathbf{S}$. Then $(eM)_R \trianglelefteq_p M_R$ if and only if $(eM)_R \trianglelefteq M_R$ if and only if $e \in S_l(\mathbf{S})$.*

The following characterization of π-dual Baer modules will be used later to obtain other results in this study.

Theorem 2.4. Let M be a module. Then the following are equivalent:

(i) M is π-dual Baer;

(ii) For each $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$, $I(M)$ is a (projection invariant) direct summand of M;

(iii) For each $N \trianglelefteq_p M$, there exists a decomposition $M = M_1 \oplus M_2$ with $M_1 \leq N$, $M_1 \trianglelefteq_p M$ and $\mathrm{Hom}_R(M, N \cap M_2) = 0$;

(iv) For each $N \trianglelefteq_p M$, there exists a decomposition $M = M_1 \oplus M_2$ with $M_1 \leq N$, $M_1 \trianglelefteq M$ and $\mathrm{Hom}_R(M, N \cap M_2) = 0$;

(v) For each $N \trianglelefteq_p M$, there exists a decomposition $M = M_1 \oplus M_2$ with $M_1 \leq N$ and $\mathrm{Hom}_R(M, N \cap M_2) = 0$.

Proof. (i) $\Leftrightarrow$ (ii) This follows from [12, Proposition 2.4] and Lemma 2.3(ii).

(i) $\Rightarrow$ (iii) This implication follows by adapted the proof of [18, Proposition 2.1((i) $\Rightarrow$ (ii))] and using Lemma 2.3.

(iii) $\Rightarrow$ (iv) This follows from Lemma 2.3(v) (see also [1, Proposition 3.1(4)]).

(iv) $\Rightarrow$ (v) This is evident.

(v) $\Rightarrow$ (i) The proof of this implication is similar to that of [18, Proposition 2.1((ii) $\Rightarrow$ (i))]. $\square$

Example 2.5. Let M be a module such that $\mathrm{Hom}_R(M,N) = 0$ for every projection invariant proper submodule N of M. Then M is π-dual Baer by Theorem 2.4. For example, the Prüfer p-group $\mathbb{Z}(p^\infty)$ and the group of rational numbers $\mathbb{Q}$ are π-dual Baer $\mathbb{Z}$-modules, where p is any prime number.

As applications of Theorem 2.4, we obtain the following corollaries.

Corollary 2.6. *Let M be a π-dual Baer module and $N \trianglelefteq_p M$. Then the following are equivalent:*

(i) $N \leq_d M$;

(ii) $D_{\mathbf{S}}(N)(M) = N$.

Proof. (i) $\Rightarrow$ (ii) Let $\pi : M \to N$ be the projection map and $i : N \to M$ be the inclusion map. Then $i\pi \in D_{\mathbf{S}}(N)$ and $i\pi(M) = N$. Hence $D_{\mathbf{S}}(N)(M) = N$.

(ii) $\Rightarrow$ (i) Since $N \trianglelefteq_p M$, $D_{\mathbf{S}}(N) \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$ by Lemma 2.3(i). Applying Theorem 2.4, we get $D_{\mathbf{S}}(N)(M) \leq_d M$. Therefore $N \leq_d M$ by (ii). $\square$

Corollary 2.7. *Let M be a module such that every projection invariant submodule of M is a direct summand of M. Then M is π-dual Baer.*

Proof. Let $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$. Then by Lemma 2.3(ii), $I(M) \trianglelefteq_p M$. So, by hypothesis, $I(M) \leq_d M$. From Theorem 2.4, it follows that M is a π-dual Baer module. $\square$

Corollary 2.8. *Let M be an indecomposable module. Then the following are equivalent:*

(i) *M is a π-dual Baer module;*

(ii) *For every proper submodule N of M,* $\mathrm{Hom}_R(M,N) = 0$.

Proof. Since M is indecomposable, the set of all idempotents of $\mathbf{S}$ is $\{0,1\}$. Therefore all submodules of M are projection invariant.

(i) $\Rightarrow$ (ii) Let N be a proper submodule of M. By Theorem 2.4, $\mathrm{Hom}_R(M,N) = 0$.

(ii) $\Rightarrow$ (i) Let $N \trianglelefteq_p M$ with $N \neq M$. Since $\mathrm{Hom}_R(M,N) = 0$, $D_{\mathbf{S}}(N) = 0$ is a direct summand of $\mathbf{S}_{\mathbf{S}}$. If $N = M$, then $D_{\mathbf{S}}(N) = \mathbf{S}$ is again a direct summand of $\mathbf{S}_{\mathbf{S}}$. This completes the proof. $\square$

Next, we compare the notions of dual Baer, π-dual Baer and quasi-dual Baer modules. From the definitions of these three notions, we infer the following remark.

Remark 2.9. (see also [12, Theorem 2.6]) It is easily seen that the following implications hold for a module M:

M is a dual Baer module $\Rightarrow$ M is a π-dual Baer module $\Rightarrow$ M is a quasi-dual Baer module.

Next, we provide some sufficient conditions under which these three notions coincide. Recall that a ring R is called a *right duo ring* if every right ideal of R is a two-sided ideal.

Example 2.10. Let M be a module such that $\mathbf{S} = \mathrm{End}_R(M)$ is a right duo ring. By [18, Remark 2.8], M is quasi-dual Baer if and only if M is dual Baer. Therefore from Remark 2.9, it follows that M is dual Baer if and only if M is π-dual Baer if and only if M is quasi-dual Baer.

Proposition 2.11. *Let R be a local ring with maximal right ideal m and $M = R/m$. Assume that* $\operatorname{Rad}(E(M)) \neq E(M)$*. Then the following are equivalent:*

(i) *$E(M)$ is a dual Baer R-module;*

(ii) *$E(M)$ is a π-dual Baer R-module;*

(iii) *$E(M)$ is a quasi-dual Baer R-module;*

(iv) *R is a division ring.*

Proof. This follows directly from Remark 2.9 and [18, Corollary 2.14]. □

Proposition 2.12. *Let M be an indecomposable $\mathbb{Z}$-module. Then the following are equivalent:*

(i) *M is dual Baer;*

(ii) *M is π-dual Baer;*

(iii) *M is quasi-dual Baer;*

(iv) *$M \cong \mathbb{Q}$ or $M \cong \mathbb{Z}(p^\infty)$ or $M \cong \mathbb{Z}/p\mathbb{Z}$, where p is a prime number.*

Proof. This is clear by Remark 2.9 and [18, Corollary 3.7]. □

Combining Remark 2.9 and [18, Corollary 3.9], we obtain the following proposition.

Proposition 2.13. *Let M be a nonzero module over a commutative perfect ring R. Then the following conditions are equivalent:*

(i) *M is dual Baer;*

(ii) *M is π-dual Baer;*

(iii) *M is quasi-dual Baer;*

(iv) *M is a semisimple module.*

Next, we present some examples to show that the class of π-dual Baer modules lies properly between the class of dual Baer modules and that of quasi-dual Baer modules (see Remark 2.9).

Example 2.14. (i) Let S be a simple ring and let ${}_SN_S$ be an S-S-bimodule. Consider the generalized matrix ring $R = \begin{bmatrix} S & N \\ N & S \end{bmatrix}$ and the right R-module $M = N \oplus S$. Assume that S is a domain that is not a division ring. We know from [15, p. 1278] that $\operatorname{End}_R(M) \cong S$ (as rings). Then $\operatorname{End}_R(M)$ is a domain and hence M is indecomposable. Therefore all submodules of M are projection invariant. By [18, Example 2.9(ii)], M is a quasi-dual Baer module which is not dual Baer. This implies that M is a quasi-dual Baer module which is not π-dual Baer by [12, Proposition 2.8(ii)].

(ii) Let R be a ring which is a finite product of simple rings such that R is not semisimple. Then R_R is a quasi-dual Baer module by [18, Proposition 2.10]. Let F be a free R-module with a finite rank $n > 1$. Using [3, Theorem 2.7], we conclude that F is a quasi-dual Baer module. Thus F is π-dual Baer by the proof of [12, Corollary 2.9]. On the other hand, the module F is not dual Baer, since otherwise R will be semisimple by [13, Corollaries 2.5 and 2.9].

In the following result, we characterize the class of rings R for which every finitely cogenerated right R-module is π-dual Baer.

Proposition 2.15. *The following conditions are equivalent for a ring R:*

(i) *Every finitely cogenerated right R-module is π-dual Baer;*

(ii) *Every finitely cogenerated right R-module is quasi-dual Baer;*

(iii) *R is a right V-ring.*

Proof. (i) $\Rightarrow$ (ii) This is clear.

(ii) $\Rightarrow$ (iii) Assume that R has a simple right R-module S which is not injective. Then $E(S) \neq S$. Let $M = A \oplus B$ be a right R-module such that $A \cong S$ and $B \cong E(S)$. Let $S_1 = \mathrm{Soc}(B)$. Clearly, $S_1 \cong S$. Note that $N = \mathrm{Soc}(M) = A \oplus S_1$ is an essential submodule of M that is fully invariant in M. By [18, Proposition 2.1], there exists a decomposition $M = M_1 \oplus M_2$ with $M_1 \subseteq N$ and $\mathrm{Hom}_R(M, N \cap M_2) = 0$. Since $N \neq M$, we have $M_2 \neq 0$ and hence $N \cap M_2 \neq 0$. Therefore $N \cap M_2$ contains a simple submodule S_2 with $S_2 \cong S \cong A$. It follows that $\mathrm{Hom}_R(M, N \cap M_2) \neq 0$, a contradiction. This proves that R is a right V-ring.

(iii) $\Rightarrow$ (i) This follows from the fact every finitely cogenerated right module over a right V-ring is semisimple. □

3 Direct sums and direct summands of π-dual Baer modules

A direct sum of π-dual Baer modules may not be π-dual Baer as we see in the following example. Another example is provided in [12, Example 2.13].

Example 3.1. Let L be a simple R-module such that the injective hull of L has no maximal submodules. It is shown in [18, Example 2.17] that the module $M = E(L) \oplus L$ is not quasi-dual Baer. Thus M is not π-dual Baer (see Remark 2.9). Now let R be a discrete valuation ring with maximal ideal $\mathfrak{m}$ and quotient field K. It is well known that $K/R \cong E(R/\mathfrak{m})$. Therefore the R-module $(K/R) \oplus (R/\mathfrak{m})$ is not π-dual Baer. On the other hand, note that both K/R and $R/\mathfrak{m}$ are π-dual Baer by [13, Theorem 3.4].

Next, we deal with a special case of direct sums of π-dual Baer modules. First, we include the following lemma which will be useful to our work in this paper.

Lemma 3.2. [7, Lemma 3.1]

(i) *Let $X_R \leq N_R \leq M$. Then $X \trianglelefteq_p N \trianglelefteq_p M$ implies that $X \trianglelefteq_p M$.*

(ii) *Let $M = \oplus_{i \in I} M_i$ and $X_R \trianglelefteq_p M_R$. Then $X = \oplus_{i \in I}(X \cap M_i)$ and $X \cap M_i \trianglelefteq_p M_i$ for all $i \in I$.*

Theorem 3.3. Let M be a π-dual Baer module. Then every direct sum of copies of M is a π-dual Baer module.

Proof. Let $N = \oplus_{i \in I} M_i$ such that $M_i \cong M$ for all $i \in I$. Let $X \trianglelefteq_p N$. By Lemma 3.2(ii), we have $X = \oplus_{i \in I}(X \cap M_i)$ and $X \cap M_i \trianglelefteq_p M_i$ for all $i \in I$. Fix $i \in I$. Since M_i is π-dual Baer, there exists a decomposition $M_i = K_i \oplus L_i$ with $K_i \leq X \cap M_i$ and $\mathrm{Hom}_R(M_i, X \cap L_i) = 0$ by Theorem 2.4. Put $K = \oplus_{i \in I} K_i$ and $L = \oplus_{i \in I} L_i$. Clearly, $M = K \oplus L$ and $K \subseteq X$. Moreover, we have $X \cap L = \oplus_{i \in I}(X \cap L_i)$. Now assume that $\mathrm{Hom}_R(M, X \cap L) \neq 0$. Then there exist $i, j \in I$ such that $\mathrm{Hom}_R(M_i, X \cap L_j) \neq 0$. But $M_j \cong M_i$. So $\mathrm{Hom}_R(M_j, X \cap L_j) \neq 0$, a contradiction. Hence $\mathrm{Hom}_R(M, X \cap L) = 0$. Applying again Theorem 2.4, it follows that N is a π-dual Baer module. □

The following corollary is an immediate consequence of Theorem 3.3.

Corollary 3.4. *Let R be a ring such that R_R is a right π-dual Baer R-module. Then all free right R-modules are π-dual Baer.*

Note that both the class of dual Baer modules and the class of quasi-dual Baer modules are closed under direct summands (see [13, Corollary 2.5] and [18, Corollary 2.5]). However, the following example illustrates that being π-dual Baer is not preserved by taking direct summands.

Example 3.5. Let R be a simple ring which is a domain but not a division ring. From [18, Proposition 2.10], we infer that R_R is a quasi-dual Baer R-module. On the other hand, R_R is not a π-dual Baer module by [12, Proposition 2.8(ii)] and [13, Corollary 2.9]. Now consider a free right R-module $F_R = \oplus_{i=1}^{n} R_i$ for some integer $n > 1$, where $R_i \cong R$ for all $1 \leq i \leq n$. Note that F is quasi-dual Baer by [3, Theorem 2.7]. Then F is π-dual Baer by [12, Corollary 2.9].

As an application of Theorem 2.4, we can improve and generalize Proposition 2.11 of [12] as follows. The proof and the techniques used are different from those of [12, Proposition 2.11].

Theorem 3.6. Let $M = M_1 \oplus M_2$ be a π-dual Baer module for some submodules M_1 and M_2 of M. If $M_1 \trianglelefteq_p M$, then M_1 and M_2 are π-dual Baer.

Proof. Let us first prove that M_1 is π-dual Baer. Take $N_1 \trianglelefteq_p M_1$. Then $N_1 \trianglelefteq_p M$ by Lemma 3.2(i). Since M is π-dual Baer, there exists a decomposition $M = K_1 \oplus K_2$ with $K_1 \leq N_1$ and $\mathrm{Hom}_R(M, N_1 \cap K_2) = 0$ (see Theorem 2.4). By modularity, we have $M_1 = K_1 \oplus (K_2 \cap M_1)$. Moreover, $N_1 \cap (K_2 \cap M_1) = N_1 \cap K_2$. It is clear that $\mathrm{Hom}_R(M_1, N_1 \cap K_2) = 0$. Using Theorem 2.4, we deduce that M_1 is π-dual Baer. To show that M_2 is π-dual Baer, take $N_2 \trianglelefteq_p M_2$. Then $N = M_1 \oplus N_2 \trianglelefteq_p M$ by [5, Lemma 4.13]. So there exist submodules K and L of M such that $M = K \oplus L$, $K \subseteq N$, $K \trianglelefteq_p M$ and $\mathrm{Hom}_R(M, N \cap L) = 0$ (see Theorem 2.4). Note that $K = (K \cap M_1) \oplus (K \cap M_2)$ by Lemma 3.2(ii). Hence $M = (K \cap M_1) \oplus (K \cap M_2) \oplus L$ and so $M_2 = (K \cap M_2) \oplus [((K \cap M_1) \oplus L) \cap M_2]$. In addition, it is clear that $K \cap M_2 = K \cap N_2 \subseteq N_2$ as $K \subseteq N$. Thus $N_2 = (K \cap N_2) \oplus [((K \cap M_1) \oplus L) \cap N_2]$. Moreover, since $M = (K \cap M_1) \oplus (K \cap N_2) \oplus L$, it follows that $N = (K \cap M_1) \oplus (K \cap N_2) \oplus (N \cap L)$ by modularity. Therefore $N_2 = (K \cap N_2) \oplus [((K \cap M_1) \oplus (N \cap L)) \cap N_2]$. Note that $((K \cap M_1) \oplus (N \cap L)) \cap N_2 \subseteq ((K \cap M_1) \oplus L) \cap N_2$. Then $((K \cap M_1) \oplus (N \cap L)) \cap N_2 = ((K \cap M_1) \oplus L) \cap N_2$. Now assume that $\mathrm{Hom}_R(M_2, N_2 \cap [((K \cap M_1) \oplus L) \cap M_2)]) \neq 0$ and let $f : M_2 \to ((K \cap M_1) \oplus (N \cap L)) \cap N_2$ be a nonzero homomorphism. Let $\pi : (K \cap M_1) \oplus (N \cap L) \to N \cap L$ be the projection map. It is easy to check that $0 \neq \pi f \in \mathrm{Hom}_R(M_2, N \cap L)$. This contradicts the fact that $\mathrm{Hom}_R(M, N \cap L) = 0$. From Theorem 2.4, we infer that M_2 is a π-dual Baer module. □

Proposition 3.7. *Let $M = M_1 \oplus M_2$ for some submodules M_1 and M_2 of M. If M is a π-dual Baer module with $\mathbf{I}_{M_1} = \mathrm{End}_R(M_1)$, then M_1 is π-dual Baer.*

Proof. By Remark 2.9, M is quasi-dual Baer. So M_1 is quasi-dual Baer by [18, Corollary 2.5]. Therefore M_1 is π-dual Baer by [12, Proposition 2.8(iv)]. □

Combining [12, Theorem 2.14] and Lemma 2.3(v), we obtain the following theorem. By using Theorem 2.4, we next provide another proof of this result.

Theorem 3.8. Let $M = \oplus_{i \in I} M_i$, where $M_i \trianglelefteq_p M$ for all $i \in I$. Then M is π-dual Baer if and only if M_i is π-dual Baer for all $i \in I$.

Proof. Assume that M is π-dual Baer. By Theorem 3.6, each M_i $(i \in I)$ is π-dual Baer. Conversely, assume that each M_i is π-dual Baer. By Lemma 2.3(v), $M_i \trianglelefteq M$ for all $i \in I$. So, $\mathrm{Hom}_R(M_i, M_j) = 0$ for all $i \neq j \in I$. Let $N \trianglelefteq_p M$. Thus $N = \oplus_{i \in I}(N \cap M_i)$ and $N \cap M_i \trianglelefteq_p M_i$ for all $i \in I$ by Lemma 3.2(ii). Fix $i \in I$. By Theorem 2.4, there exists a decomposition $M_i = K_i \oplus L_i$ with $K_i \subseteq N \cap M_i$ and $\mathrm{Hom}_R(M_i, N \cap L_i) = 0$. Set $K = \oplus_{i \in I} K_i$ and $L = \oplus_{i \in I} L_i$. Clearly, $M = K \oplus L$ and $K \subseteq N$. Moreover, it is easy to see that $N \cap L = \oplus_{i \in I}(N \cap L_i)$. Combining the facts that $\mathrm{Hom}_R(M_i, M_j) = 0$ for all $i \neq j \in I$ and $\mathrm{Hom}_R(M_i, N \cap L_i) = 0$ for all $i \in I$, we conclude that $\mathrm{Hom}_R(M, N \cap L) = 0$. Using Theorem 2.4, it follows that M is π-dual Baer. □

Let M be a module. The radical of M will be denoted by Rad(M). Note that Rad(M) is a fully invariant submodule of M by [2, Proposition 9.14]. Clearly, if M is semisimple, then $\mathrm{Rad}(M) = 0$.

Corollary 3.9. *Let an R-module $M = M_1 \oplus M_2$ be a direct sum of submodules M_1 and M_2 such that $\mathrm{Rad}(M_1) = M_1$, M_2 is semisimple. If M is π-dual Baer, then M_1 is π-dual Baer. The converse holds when $\mathrm{Hom}_R(M_2, M_1) = 0$.*

Proof. Note that $\mathrm{Rad}(M) = \mathrm{Rad}(M_1) \oplus \mathrm{Rad}(M_2) = M_1 \trianglelefteq M$.

($\Rightarrow$) This follows by Theorem 3.6.

($\Leftarrow$) Since $\mathrm{Hom}_R(M_2, M_1) = 0$, $M_2 \trianglelefteq M$. Now the result follows from Theorem 3.8. □

For the proof of the implication (i) $\Rightarrow$ (ii) in the following proposition, we mainly follow the proof of [18, Proposition 2.15((i) $\Rightarrow$ (ii))].

Proposition 3.10. *Let an R-module $M = M_1 \oplus M_2$ be a direct sum of submodules M_1 and M_2 such that $\mathrm{Rad}(M_1) = M_1$ and M_2 is semisimple. Then the following are equivalent:*

(i) *M is π-dual Baer;*

(ii) *M_1 is π-dual Baer and $I(M_2) \cap M_1 \subseteq I(M_1)$ for all $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$.*

Proof. (i) $\Rightarrow$ (ii) By Corollary 3.9, M_1 is π-dual Baer. Now we will prove that $I(M_2) \cap M_1 \subseteq I(M_1)$ for all $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$. Let $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$. By Lemma 2.3(ii), $I(M_1) + I(M_2) = I(M) \trianglelefteq_p M$. Hence $I(M) = (I(M) \cap M_1) \oplus (I(M) \cap M_2)$ by Lemma 3.2(ii). As $M_1 \trianglelefteq M$, we have $I(M_1) \subseteq M_1$. By modularity, $M_1 \cap I(M) = M_1 \cap (I(M_1) + I(M_2)) = I(M_1) + (M_1 \cap I(M_2))$. Since $M_1 \cap I(M_2)$ is semisimple, there exists a semisimple submodule N of $M_1 \cap I(M_2)$ such that $I(M_1) + (M_1 \cap I(M_2)) = I(M_1) \oplus N$. Therefore $I(M) = (I(M) \cap M_1) \oplus (I(M) \cap M_2) = I(M_1) \oplus N \oplus (I(M) \cap M_2)$. Now by Theorem 2.4, $I(M) = I(M_1) \oplus N \oplus (I(M) \cap M_2) \leq_d M$. Thus $N \leq_d M_1$ and so $\mathrm{Rad}(N) = N \cap \mathrm{Rad}(M_1) = N \cap M_1 = N$. On the other hand, we have $\mathrm{Rad}(N) = 0$ since N is semisimple. Therefore $N = 0$. This implies that $I(M_1) + (M_1 \cap I(M_2)) = I(M_1)$. Consequently, $I(M_2) \cap M_1 \subseteq I(M_1)$.

(ii) $\Rightarrow$ (i) Let $N \trianglelefteq_p M$. Then $N = (N \cap M_1) \oplus (N \cap M_2)$ and $N \cap M_1 \trianglelefteq_p M_1$ (see Lemma 3.2(ii)). Since M_1 is π-dual Baer, there exist submodules K_1 and L_1 of M_1 such that $M_1 = K_1 \oplus L_1$, $K_1 \subseteq N \cap M_1$ and $\mathrm{Hom}_R(M_1, N \cap L_1) = 0$ (see Theorem 2.4). Since M_2 is semisimple, there exists a submodule $L_2 \leq M_2$ such that $M_2 = (N \cap M_2) \oplus L_2$. Put $K = K_1 \oplus (N \cap M_2)$ and $L = L_1 \oplus L_2$. Then $M = K \oplus L$ with $K \subseteq N$. It is easily seen that $N \cap L = (N \cap L_1) \oplus (N \cap L_2)$. But $N \cap L_2 = 0$, so $N \cap L = N \cap L_1$. Applying Theorem 2.4, it remains to prove that $\mathrm{Hom}_R(M, N \cap L_1) = 0$. Let $f \in \mathrm{Hom}_R(M, N \cap L_1)$ and consider the ideal $I = \mathbf{S} f \mathbf{S}$ of $\mathbf{S}$. By (ii), $I(M_2) \cap M_1 \subseteq I(M_1)$. Note that $f(M_1) = 0$ as $\mathrm{Hom}_R(M_1, N \cap L_1) = 0$. Since $M_1 \trianglelefteq M$, we have $I(M_1) = 0$. Therefore $I(M_2) \cap M_1 = 0$ and hence $f(M_2) \cap M_1 = f(M_2) = 0$. It follows that $f = 0$, as desired. □

Next, we provide a characterization of π-dual Baer modules over a commutative semilocal ring. But first we need a lemma.

Lemma 3.11. *Let M be a π-dual Baer module over a commutative ring R. Then $M\mathfrak{a}$ is a direct summand of M for any ideal $\mathfrak{a}$ of R.*

Proof. This follows from Remark 2.9 and [18, Proposition 3.3]. □

Proposition 3.12. *Let M be a nonzero module over a commutative semilocal ring R. Then the following are equivalent:*

(i) *M is π-dual Baer;*

(ii) *$M = M_1 \oplus M_2$ is a direct sum of submodules M_1 and M_2 such that* $\mathrm{Rad}(M_1) = M_1$ *is π-dual Baer and M_2 is semisimple, and $I(M_2) \cap M_1 \subseteq I(M_1)$ for every $I_{\mathbf{S}} \trianglelefteq_p \mathbf{S}_{\mathbf{S}}$.*

Proof. (i) $\Rightarrow$ (ii) By Lemma 3.11 and the proof of [18, Theorem 3.8], the module M has a decomposition $M = M_1 \oplus M_2$ such that $\mathrm{Rad}(M_1) = M_1$ and M_2 is semisimple. The result now follows from Proposition 3.10.

(ii) $\Rightarrow$ (i) This is clear by Proposition 3.10. □

In the remainder of this section we assume that R is a Dedekind domain with quotient field Q such that $Q \neq R$. Let M be an R-module. The set $T(M) = \{x \in M \mid xr = 0$ for some nonzero $r \in R\}$ is a submodule of M which is called the *torsion submodule* of M. The module M is said to be *torsion* (resp., *torsion-free*) if $T(M) = M$ (resp., $T(M) = 0$). Let $\mathbb{P}$ denote the set of all nonzero prime ideals of R. For any $0 \neq \mathfrak{p} \in \mathbb{P}$, let $T_{\mathfrak{p}}(M)$ denote the set $\{x \in M \mid \mathfrak{p}^n x = 0$ for some integer $n \geq 0\}$ which is called the $\mathfrak{p}$-*primary component* of M. The module M is called $\mathfrak{p}$-*primary* if $T_{\mathfrak{p}}(M) = M$. It is well known that if M is a torsion R-module, then M is a direct sum of its $\mathfrak{p}$-primary components. The $\mathfrak{p}$-primary component of the torsion R-module Q/R will be denoted by $R(\mathfrak{p}^\infty)$.

Next, we aim to describe the structure of quasi-dual Baer modules and π-dual Baer modules over Dedekind domains. First, we prove the following needed lemmas.

Lemma 3.13. *Let M be a nonzero torsion-free R-module. If M is quasi-dual Baer, then M is an injective module.*

Proof. Assume that M is quasi-dual Baer and let $0 \neq s \in R$. By [18, Proposition 3.3], there exists a submodule K of M such that $M = sM \oplus K$. Hence $sK = 0$. Therefore $K = 0$ since M is torsion-free. Thus $M = sM$. Hence M is a divisible R-module. By [17, Proposition 2.7], it follows that M is injective. □

Lemma 3.14. *Let M be a torsion R-module. Assume that M is quasi-dual Baer. Then $M = E \oplus F$ is a direct sum of an injective submodule E and a semisimple submodule F.*

Proof. By [18, Corollary 2.5], every primary component $T_{\mathfrak{p}}(M)$ is quasi-dual Baer. Note that every direct sum of injective R-modules is injective since R is a noetherian ring. So without loss of generality we can assume that $M = T_{\mathfrak{p}}(M)$ for some nonzero prime ideal $\mathfrak{p}$ of R. Since $\mathfrak{p}M \trianglelefteq M$, there exists a decomposition $M = M_1 \oplus M_2$ with $M_1 \subseteq \mathfrak{p}M$ and $\mathrm{Hom}_R(M, \mathfrak{p}M \cap M_2) = 0$ (see [18, Proposition 2.1]). Then $\mathfrak{p}M = M_1 \oplus (\mathfrak{p}M \cap M_2)$ by modularity. Moreover, we have $\mathfrak{p}M = \mathfrak{p}M_1 \oplus \mathfrak{p}M_2$. Therefore $\mathfrak{p}M_1 = M_1$ and $\mathfrak{p}M \cap M_2 = \mathfrak{p}M_2$. Thus $\mathrm{Hom}_R(M_2, \mathfrak{p}M_2) = 0$. This implies that $rM_2 = 0$ for all $r \in \mathfrak{p}$, that is, $\mathfrak{p}M_2 = 0$. Hence M_2 is a semisimple module. Moreover, we have $M_1 = \mathfrak{p}M = \mathrm{Rad}(M)$ and $M = \mathfrak{p}M \oplus M_2$. It follows that $\mathfrak{p}M = \mathfrak{p}(\mathfrak{p}M)$. This yields $\mathrm{Rad}(M) = \mathrm{Rad}(\mathrm{Rad}(M))$. Since R is a Dedekind domain, we see that $\mathrm{Rad}(M) = M_1$ is injective. This completes the proof. □

For an R-module M, we will denote the sum of all divisible (injective) submodules of M by $d(M)$. It is well known that $d(M)$ is an injective fully invariant submodule of M. It is shown in [11, Theorem 7] that every injective R-module is a direct sum of copies of Q and $R(\mathfrak{p}^\infty)$ for various nonzero prime ideals $\mathfrak{p}$. An R-module M is said to be *reduced* if M has no divisible submodules (that is $d(M) = 0$).

Theorem 3.15. Let R be a Dedekind domain with quotient field Q such that $Q \neq R$. Then the following assertions are equivalent for an R-module M:

(i) M is dual Baer;

(ii) M is π-dual Baer;

(iii) M is quasi-dual Baer;

(iv) M is a direct sum of copies of Q, $(R(\mathfrak{p}_i^\infty))_{i\in I}$ and $(R/\mathfrak{q})_{j\in J}$, where $(\mathfrak{p}_i)_{i\in I}$ and $(\mathfrak{q})_{j\in J}$ are nonzero prime ideals of R with $\mathfrak{p}_i \neq \mathfrak{q}_j$ for every couple $(i,j) \in I \times J$.

Proof. (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) See Remark 2.9.

(iii) $\Rightarrow$ (iv) Since $d(M)$ is injective, it follows that $M = d(M) \oplus L$ for some reduced submodule L of M. Note that $d(M)$ and L are quasi-dual Baer by [18, Corollary 2.5]. Since $T(L) \trianglelefteq L$, there exists a decomposition $L = N \oplus K$ with $N \subseteq T(L)$ and $\mathrm{Hom}_R(L, T(L) \cap K) = 0$ (see [18, Proposition 2.1]). But $T(L) \cap K = T(K)$. Then $\mathrm{Hom}_R(L, T(K)) = 0$. Now assume that $T(K) \neq 0$. Then K has a direct summand K_0 which is isomorphic to $R/\mathfrak{p}^n$ for some nonzero prime ideal $\mathfrak{p}$ of R and some positive integer n (see [11, Theorem 9]). Since $K_0 \subseteq T(K)$, we have $\mathrm{Hom}_R(K, T(K)) \neq 0$. Hence $\mathrm{Hom}_R(L, T(K)) \neq 0$, a contradiction. Therefore $T(K) = 0$ and so $T(L) = N$. Using again [18, Corollary 2.5], we infer that N and K are quasi-dual Baer. Now taking into account Lemmas 3.13 and 3.14, we conclude that $K = 0$ and $N = L$ is semisimple. Note that $d(M)$ is a direct sum of copies of Q and $R(\mathfrak{p}^\infty)$ for various nonzero prime ideals $\mathfrak{p}$. Moreover, for each nonzero prime ideal $\mathfrak{p}$ of R, the R-module $R(\mathfrak{p}^\infty) \oplus R/\mathfrak{p}$ is not quasi-dual Baer by [18, Example 2.17]. Now (iv) follows from the fact that the class of quasi-dual Baer modules is closed under direct summands (see [18, Corollary 2.5]).

(iv) $\Rightarrow$ (i) This follows from [13, Theorem 3.4]. □

4 π-dual Baer Rings

We will call a ring R a *right π-dual Baer* (resp., *right dual Baer*) ring if the right R-module R_R is π-dual Baer (resp., *dual Baer*). Following [18], a ring R is called a *right quasi-dual Baer* ring if the right R-module R_R is a quasi-dual Baer module. Left π-dual Baer rings, left dual Baer rings and left quasi-dual Baer rings are defined similarly. It was shown in [13, Corollary 2.9] and [18, Corollary 2.11] that dual Baer and quasi-dual Baer properties are left-right symmetric for any ring R. Moreover, the dual Baer rings are exactly the semisimple rings and the class of quasi-dual Baer rings is precisely the class of finite product of simple rings. This implies that a commutative ring R is (right) π-dual Baer if and only if R is semisimple. We begin by characterizing right π-dual Baer rings in some special cases.

Recall that a ring R is called *Abelian* if every idempotent of R is central.

Remark 4.1. (i) Let R be an Abelian ring. By [12, Proposition 2.8(iii)], we infer that R is a right π-dual Baer ring if and only if R is a left π-dual Baer ring if and only if R is a semisimple ring.

(ii) Let R be a ring with $\mathbf{I}(R) = R$. Combining [12, Proposition 2.8(iv)] with [18, Proposition 2.10], we conclude that R is a right π-dual Baer ring if and only if R is a left π-dual Baer ring if and only if R is a quasi-dual Baer ring if and only if R is a finite product of simple rings.

Recall that a ring R is called *projection invariant Baer* (or π-Baer) if for each ${}_RY \trianglelefteq_p {}_RR$, there exists $c^2 = c \in R$ such that $r_R(Y) = \{r \in R \mid Yr = 0\} = cR$ (see [6, Definition 2.2]). It is proven in [6] that π-Baer condition for a ring is left-right symmetric. Therefore R is π-Baer if and only if for each $Y_R \trianglelefteq_p R_R$, there exists $c^2 = c \in R$ such that $l_R(Y) = \{r \in R \mid rY = 0\} = Rc$.

Next, we compare the class of right π-dual Baer rings and that of π-Baer rings.

Remark 4.2. From [12, Proposition 3.1], it follows that every right or left π-dual Baer ring R is a π-Baer ring.

Remark 4.3. It was shown in [6, Corollary 2.2(ii)] that if R is a π-Baer ring and S is a subring of R with $\mathbf{I}(R) \subseteq S$, then S is π-Baer. The analogue of this fact is not true, in general, for right π-dual Baer rings. To see this, consider the ring $\mathbb{Q}$ which is (right) π-dual Baer. However, since the subring $\mathbb{Z}$ of $\mathbb{Q}$ is not semisimple, the ring $\mathbb{Z}$ is not (right) π-dual Baer even if $\mathbf{I}(\mathbb{Q}) = \mathbb{Z}$ (see Remark 4.1(i)).

Note that a ring R is a domain if and only if it is π-Baer and 0 and 1 are its only idempotents. In the following example, we present some rings which are π-Baer, but not right π-dual Baer.

Example 4.4. Let R be a π-Baer ring such that R is not semisimple and the right R-module R_R is indecomposable. Then R cannot be right π-dual Baer by Remark 4.1(i). Explicit examples are:

(i) Let R be the free ring $\mathbb{Z} < x, y >$. Since R is a domain, R is a π-Baer ring (see [6, Example 2.1]). On the other hand, the ring R is not semisimple.

(ii) Let A be a prime ring such that $Z(A_A) \neq 0$, $Z(A_A) \neq A$ and A_A is a uniform module (see specific examples in [8, Example 4.3]). Thus A is not a domain and $\{0, 1\}$ is the set of all idempotent elements of A. Therefore A is not a π-Baer ring. Now let $R = \mathbf{Mat}_n(A)$ be the n-by-n full matrix ring over A for some integer $n > 1$. It is well known that $\mathbf{I}(R) = R$. Moreover, by [6, Example 2.2], R is a π-Baer ring. On the other hand, suppose that the ring R is right π-dual Baer. Then R is quasi-dual Baer (see Remark 4.1(ii)). Hence A is also quasi-dual Baer (see Proposition 4.23 below). Using [18, Proposition 2.10], we deduce that A is a simple ring since A_A is indecomposable. This contradicts the fact that $Z(A_A) \neq 0$ and $Z(A_A) \neq A$. This proves that R is not a right π-dual Baer ring.

Lemma 4.5. *Let e be a central idempotent in a ring R. Then eR is π-dual Baer as a right R-module if and only if eR is π-dual Baer as a right eR-module.*

Proof. This follows directly from Theorem 2.4. □

Proposition 4.6. *Assume that R is a right π-dual Baer ring and let $e^2 = e \in R$. If $eR \trianglelefteq_p R_R$, then e and $1-e$ are central idempotents. Moreover, $eR = eRe$ and $(1-e)R = (1-e)R(1-e)$ are right π-dual Baer rings.*

Proof. Note that R is quasi-dual Baer. Thus R is a semiprime ring by the proof of [18, Proposition 2.10((iii) $\Rightarrow$ (iv))]. Since $eR \trianglelefteq_p R_R$, eR is a two-sided ideal of R by Lemma 2.3(v). Now using [10, Lemma 3.1], it follows that e is central. So $1-e$ is also central. The last assertion follows directly by applying Theorem 3.6 and Lemma 4.5. □

Proposition 4.7. *For a ring R, the following are equivalent:*

(i) *R is a right π-dual Baer ring;*

(ii) *Every projection invariant right ideal of R is a direct summand of R_R;*

(iii) *Every projection invariant right ideal of R is a two-sided ideal of R and R is a quasi-dual Baer ring.*

Proof. Given $a \in R$, let $\varphi_a : R \to R$ be the R-endomorphism of R_R defined by $\varphi_a(x) = ax$ for all $x \in R$.

(i) $\Rightarrow$ (ii) Let $I_R \trianglelefteq_p R_R$. Define the set $\mathcal{I} = \{\varphi_a : a \in I\}$. It is not hard to see that $\mathcal{I}$ is a right ideal of $\mathbf{S} = \mathrm{End}_R(R_R)$. Moreover, $\mathcal{I}_\mathbf{S} \trianglelefteq_p \mathbf{S}_\mathbf{S}$. To see this, let $e^2 = e \in \mathbf{S}$. Then $e = \varphi_{e(1)}$ and $e(1)$ is an idempotent in R. Hence $e(1)I \subseteq I$. Now let $\varphi_b \in \mathcal{I}$, where $b \in I$. Then $\varphi_{e(1)}\varphi_b = \varphi_{e(1)b} \in \mathcal{I}$. Therefore $e\mathcal{I} \subseteq \mathcal{I}$. It follows that $\mathcal{I}_\mathbf{S} \trianglelefteq_p \mathbf{S}_\mathbf{S}$. Now by Theorem 2.4, $\mathcal{I}(R_R) = \sum_{a\in I} \varphi_a(R) = \sum_{a\in I} aR = I \leq_d R_R$.

(ii) $\Rightarrow$ (iii) Note that every two-sided ideal of R is a direct summand of R_R. Thus R is a quasi-dual Baer ring by [18, Proposition 2.10]. Let $I_R \trianglelefteq_p R_R$. By (ii), $I \leq_d R_R$. Hence there exists an idempotent $e \in R$ such that $I = eR$. By Lemma 2.3(v), I is fully invariant in R_R and hence I is a two-sided ideal of R.

(iii) $\Rightarrow$ (i) Let $I_R \trianglelefteq_p R_R$. By (iii), I is a two-sided ideal of R. Therefore $I \leq_d R_R$ by [18, Proposition 2.10]. Hence R is a right π-dual Baer ring by Corollary 2.7. □

Proposition 4.8. *Let $\{R_i : i \in I\}$ be a family of rings. Then the direct product $R = \prod_{i\in I} R_i$ is a right π-dual Baer ring if and only if the indexing set I is finite and each R_i is right π-dual Baer.*

Proof. Using Theorem 3.8 and Lemma 4.5, we are reduced to proving that if R is right π-dual Baer, then I is a finite set. Suppose that R is right π-dual Baer. Assume that I is not finite. Note that $A = \oplus_{i\in I} R_i$ is a two-sided ideal of the ring R. Hence the right ideal A is a direct summand of R_R by Proposition 4.7. Therefore $R_R = A \oplus X$ for some proper right ideal X of R. This is impossible. It follows that I is a finite set. □

To obtain another characterization of right π-dual Baer rings, we introduce the following type of rings which is a stronger form of simple rings.

Definition 4.9. A ring R is said to be a *right* (*left*) *π-simple ring* if 0 and R are the only projection invariant right (left) ideals in R.

It is clear that any right π-simple ring is a simple ring which is right π-dual Baer.

Lemma 4.10. *Let R be a simple ring. Then the following conditions are equivalent:*

(i) *R is a right π-dual Baer ring;*

(ii) *R is a right π-simple ring.*

Proof. (i) ⇒ (ii) Let $I_R \trianglelefteq_p R_R$. By Proposition 4.7, I is a two-sided ideal of R. Since R is a simple ring, it follows that $I = 0$ or $I = R$.

(ii) ⇒ (i) This is immediate. □

In the next example, we exhibit some right π-simple rings.

Example 4.11. Let R be a simple ring such that $\mathbf{I}(R) = R$. Then R is a right and left π-dual Baer ring by Remark 4.1(ii). Therefore R is a right and left π-simple ring by Lemma 4.10. For example, if R' is a simple ring and $n > 1$ is a positive integer, then $\mathbf{Mat}_n(R')$ is a simple ring by [14, Theorem 3.1]. Moreover, we have $\mathbf{I}(\mathbf{Mat}_n(R')) = \mathbf{Mat}_n(R')$. It follows that $\mathbf{Mat}_n(R')$ is a right and left π-simple ring.

Proposition 4.12. *Let R be a right π-simple ring. Then either R is a division ring or R has a non-trivial idempotent element.*

Proof. Assume that R has no idempotent element except 0 and 1. Then clearly every right ideal of R is projection invariant. Since R is right π-simple, it follows that R is a division ring. □

Next, we present some simple rings which are not right π-simple.

Example 4.13. Let R be a simple ring that is not a division ring which has no idempotent element except 0 and 1. Then R is not a right π-simple ring by Proposition 4.12. As explicit examples, we can take:

(a) Weyl algebras, $A_n(F)$, over a field F of characteristic zero (see [14, Corollary 3.17]), or

(b) the Zalesskii-Neroslavskii example (see, for example [9, Example 14.17]).

Remark 4.14. By Remark 2.9, the following implications hold for any ring R:

R is a (right) dual Baer ring ⇒ R is a right π-dual Baer ring ⇒ R is a (right) quasi-dual Baer ring.

The following examples show that these implications are not reversible, in general:

(i) Let R be a simple ring which is not semisimple (see [14]) and let $n > 1$ be a positive integer. Then $\mathbf{Mat}_n(R)$ is a right π-dual Baer ring by Lemma 4.10 and Example 4.11. Let e be the matrix unit E_{11} in $\mathbf{Mat}_n(R)$. Then the rings $e\mathbf{Mat}_n(R)e$ and R are isomorphic (see [14, Example 21.14]). Now using [14, Corollary 21.13], we see that the ring $\mathbf{Mat}_n(R)$ is not semisimple. Hence $\mathbf{Mat}_n(R)$ is not a (right) dual Baer ring by [13, Corollary 2.9].

(ii) Using [18, Proposition 2.10] and Lemma 4.10, it follows easily that the rings given in Example 4.13(a)-(b) are quasi-dual Baer, but not right π-dual Baer.

Theorem 4.15. For a ring R, the following are equivalent:

(i) R is a right π-dual Baer ring;

(ii) R is a finite product of right π-simple rings.

Proof. (i) $\Rightarrow$ (ii) Assume that R is a right π-dual Baer ring. Then R is a (right) quasi-dual Baer ring by Remark 4.14. By [18, Proposition 2.10], there exist nonzero two-sided ideals $R_1, \ldots, R_n$ of R for some positive integer n such that $R = R_1 \oplus \cdots \oplus R_n$ and each R_i $(1 \leq i \leq n)$ is a simple ring. By [2, Proposition 7.6], there exist pairwise orthogonal central idempotents $e_1, \ldots, e_n \in R$ with $1 = e_1 + \cdots + e_n$, and $R_i = e_i R$ for every $i = 1, \ldots, n$. From Proposition 4.6, it follows that each R_i $(1 \leq i \leq n)$ is a right π-dual Baer ring. Now using Lemma 4.10, we infer that each R_i $(1 \leq i \leq n)$ is a right π-simple ring.

(ii) $\Rightarrow$ (i) This follows from Proposition 4.8 and Lemma 4.10. □

Remark 4.16. It would be desirable to investigate if the property of being a π-dual Baer ring is left-right symmetric but we have not been able to do this. Note that from Theorem 4.15, it follows that the π-dual Baer ring property is left-right symmetric if and only if so is the π-simple ring property.

Let R be a ring. For each $A \subseteq R$, the right annihilator of A in R is

$$r_R(A) = \{r \in R \mid ar = 0 \text{ for all } a \in A\}.$$

In the next proposition, we provide a necessary condition for a ring to be right π-simple.

Proposition 4.17. *Let R be a right π-simple ring. Then for every nonzero projection invariant left ideal I of R, we have $r_R(I) = 0$.*

Proof. Note that R is a right π-dual Baer ring by Theorem 4.15. Then R is a π-Baer ring by Remark 4.2. Let $0 \neq {}_R I \trianglelefteq_p {}_R R$. Then $r_R(I) \trianglelefteq_p R_R$ by [6, Lemma 2.1]. Since R is right π-simple, we have $r_R(I) = 0$ or $r_R(I) = R$. But $I \neq 0$. So $r_R(I) = 0$. □

Proposition 4.18. *Let R be a ring with* $\mathrm{Soc}(R_R)$ *essential in R_R. Then the following are equivalent:*

(i) *R is a dual Baer ring;*

(ii) *R is a right π-dual Baer ring;*

(iii) *R is a quasi-dual Baer ring;*

(iv) *R is a semisimple ring.*

Proof. (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are clear by Remark 4.14.

(iii) $\Rightarrow$ (iv) Note that $\mathrm{Soc}(R_R)$ is a two-sided ideal of R. Then $\mathrm{Soc}(R_R)$ is a direct summand right ideal of R by [18, Proposition 2.10]. Hence $R = \mathrm{Soc}(R_R)$ since $\mathrm{Soc}(R_R)$ is essential in R_R.

(iv) $\Rightarrow$ (i) is clear. □

Next, we investigate the transfer of the right π-dual Baer condition between a base ring R and several extensions. We begin with $R[x]$ and $R[[x]]$.

Proposition 4.19. *Let R be a ring satisfying one of the following conditions:*

(i) *$R[x]$ is a right π-dual Baer ring;*

(ii) *$R[[x]]$ is a right π-dual Baer ring.*

Then R is a right π-dual Baer ring.

Proof. (i) Suppose that $R[x]$ is a right π-dual Baer ring and let I be a projection invariant right ideal of R. By [6, Lemma 4.1(iv)], $I[x]$ is a projection invariant right ideal of $R[x]$. This implies that $I[x] = e(x)R[x]$ for some idempotent $e(x) = e_0 + e_1x + \cdots + e_nx^n \in R[x]$ (see Proposition 4.7). Let us show that $I = e_0R$. Since $e(x) \in I[x]$, we have $e_0 \in I$ and so $e_0R \subseteq I$. Now let $a \in I$. Therefore $a \in I[x] = e(x)R[x]$. Hence $a = e(x)f(x)$ for some $f(x) = f_0 + f_1x + \cdots + f_mx^m \in R[x]$. It follows that $a = e_0f_0 \in e_0R$. This proves that $I = e_0R$. Therefore R is a right π-dual Baer ring by Proposition 4.7.

(ii) This follows by the same method as in (i). □

The next example shows that polynomial extensions of right π-dual Baer rings need not be right π-dual Baer.

Example 4.20. Let F be a field. Clearly, F is a right π-dual Baer ring. On the other hand, it is well known that both $F[x]$ and $F[[x]]$ are integral domains, but they are not semisimple. From Remark 4.1(i), it follows that neither $R[x]$ nor $R[[x]]$ is right π-dual Baer.

We conclude this paper by investigating when full or generalized triangular matrix rings are right π-dual Baer.

Proposition 4.21. *Let R be a quasi-dual Baer ring (in particular if R is a right π-dual Baer ring). Then $\mathbf{Mat}_n(R)$ is a right and left π-dual Baer ring for every positive integer $n > 1$.*

Proof. By [18, Proposition 2.10], there exists a positive integer t such that $R = \prod_{i=1}^{t} R_i$ is a finite product of simple rings R_i ($1 \le i \le t$). Let $n > 1$ be a positive integer. Note that $A = \mathbf{Mat}_n(R) \cong \prod_{i=1}^{t} \mathbf{Mat}_n(R_i)$ (as rings). By [14, Theorem 3.1], each $\mathbf{Mat}_n(R_i)$ ($1 \le i \le t$) is a simple ring. Since $\mathbf{I}(A) = A$, it follows from Remark 4.1(ii) that A a right and left π-dual Baer ring. □

The next example illustrates the fact that the right π-dual Baer property is not Morita invariant.

Example 4.22. It is well known that for any ring R and any positive integer m, the rings R and $\mathbf{Mat}_m(R)$ are Morita equivalent (see [2, Corollary 22.6]). Let R be a simple ring which is not right π-simple (see Example 4.13). Then R is not right π-dual Baer by Lemma 4.10. On the other hand, for every positive integer $n > 1$, $\mathbf{Mat}_n(R)$ is a right π-dual Baer ring by Proposition 4.21.

Proposition 4.21 and Example 4.22 should be compared with the following proposition.

Proposition 4.23. *Let R be a ring. Then the following statements are equivalent:*

(i) *R is a quasi-dual Baer ring;*

(ii) *$\mathbf{Mat}_n(R)$ is a quasi-dual Baer ring for every positive integer n;*

(iii) *$\mathbf{Mat}_n(R)$ is a quasi-dual Baer ring for some positive integer $n > 1$.*

Proof. (i) ⇒ (ii) This follows from Remark 4.14 and Proposition 4.21.

(ii) ⇒ (iii) This is immediate.

(iii) ⇒ (i) Let $n > 1$ be a positive integer such that $A = \mathbf{Mat}_n(R)$ is a quasi-dual Baer ring. Then A is a semiprime ring (see the proof of [18, Proposition 2.10]). Let e be the matrix unit E_{11} in A. Clearly, e is an idempotent in A. Moreover, $eAe = \{aE_{11} \mid a \in R\}$ and R are isomorphic rings (see [14, Example 21.14]). Let us show that eAe is a quasi-dual Baer ring. Take a two-sided ideal U of eAe. Then AUA is a two-sided ideal of A. Thus AUA is a direct summand of A_A by [18, Proposition 2.10]. This implies that $AUA = fA$ for some $f^2 = f \in A$. Since A is a semiprime ring, it follows from [10, Lemma 3.1] that f is a central idempotent in A. Now [14, Theorem 21.11(2)] gives that $U = e(AUA)e$. Therefore $U = e(fA)e$. Hence $U = e^2(fAe) = efe(eAe)$ as f is central. Moreover, it is clear that efe is an idempotent in the ring eAe. It follows that U is a direct summand of eAe_{eAe}. Consequently, eAe is a quasi-dual Baer ring by [18, Proposition 2.10]. □

Next, we characterize right π-dual Baer 2-by-2 generalized triangular matrix rings.

Theorem 4.24. Let $T = \begin{bmatrix} R & M \\ 0 & S \end{bmatrix}$ denote a 2-by-2 generalized upper triangular matrix ring where R and S are rings and M is an (R,S)-bimodule. Then the following statements are equivalent:

(i) T is a right π-dual Baer ring;

(ii) R and S right π-dual Baer rings and $M = 0$.

Proof. (i) $\Rightarrow$ (ii) It is well known that $\mathrm{Rad}(T) = \begin{bmatrix} \mathrm{Rad}(R) & M \\ 0 & \mathrm{Rad}(S) \end{bmatrix}$ is a two-sided ideal of T and hence it is a direct summand of T_T by Proposition 4.7. But $\mathrm{Rad}(T)$ is small in T_T. Then $\begin{bmatrix} \mathrm{Rad}(R) & M \\ 0 & \mathrm{Rad}(S) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. This yields $M = 0$. It follows that $T = \begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix} \cong R \times S$ (as rings). Now from Proposition 4.8, we infer that R and S are right π-dual Baer rings.

(ii) $\Rightarrow$ (i) This follows by using again Proposition 4.8. □

Remark 4.25. From the previous theorem, it follows that for any nonzero ring R, the 2-by-2 upper triangular matrix ring over R is never a right π-dual Baer ring.

References

[1] C. Abdioğlu, M. T. Koşan and S. Şahinkaya, On modules for which all submodules are projection invariant and the lifting condition, Southeast Asian Bulletin of Mathematics 34 (2010), 807–818.

[2] F. W. Anderson and K. R. Fuller, Rings and Categories of Modules, Springer, New York, 1974.

[3] T. Amouzegar and Y. Talebi, On quasi-dual Baer modules, TWMS J. Pure Appl. Math. 4 (2013), 78–86.

[4] G. F. Birkenmeier and R. L. LeBlanc, s.Baer and s.Rickart modules, J. Algebra Appl. 14 (2015), (08):1550131.

[5] G. F. Birkenmeier, A. Tercan and C. C. Yücel, The extending condition relative to sets of submodules, Comm. Algebra 42(2) (2014), 764–778.

[6] G. F. Birkenmeier, Y. Kara and A. Tercan, π-Baer rings, J. Algebra Appl. 16(11) (2018), 1850029 (19 pages).

[7] G. F. Birkenmeier, Y. Kara and A. Tercan, π-endo Baer modules, Comm. Algebra 48(3) (2020), 1132–1149.

[8] K. A. Brown, The singular ideals of group rings, Q. J. Math. 28(1) (1977), 41–60.

[9] A. W. Chatters and C. R. Hajarnavis, Rings with Chain Conditions, Research notes in Mathematics, vol. **44**, Pitman Publishing Inc., London, 1980.

[10] K. R. Goodearl, Von Neumann Regular Rings, Pitman, London, 1979.

[11] I. Kaplansky, Modules over Dedekind rings and valuation rings, Trans. Amer. Math. Soc. 72 (1952), 327–340.

[12] Y. Kara, On dual π-endo Baer modules, Malaya J. Mat. 9(2) (2021), 39–45.

[13] D. Keskin Tütüncü and R. Tribak, On dual Baer modules, Glasg. Math. J. 52(2) (2010), 261–269.

[14] T. Y. Lam, A First Course in Noncommutative Rings, 2nd edn, Springer, New York, 2001.

[15] C. Lomp, A counterexample for problem on quasi-Baer modules, Taiwanese Journal of Mathematics 21(6) (2017), 1277–1281.

[16] S. T. Rizvi and C. S. Roman, Baer and quasi-Baer modules, Comm. Algebra 32(1) (2004), 103–123.

[17] D. W. Sharpe, P. Vámos, Injective Modules, Cambridge University Press, Cambridge, 1972.

[18] R. Tribak, Y. Talebi and M. Hosseinpour, Quasi-dual Baer modules, Arab. J. Math. 10 (2021), 497–504.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 124-142
ISSN: 2820-7114

Title :

The implications among the three classical trigonometric laws of hyperbolic geometry

Author(s):

David E. Dobbs

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 124–142

Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco **ISSN: 2820-7114**

The implications among the three classical trigonometric laws of hyperbolic geometry

David E. Dobbs
Department of Mathematics, University of Tennessee, Knoxville, Tennessee 37996-1320
e-mail: *ddobbs1@utk.edu*

Communicated by Najib Mahdou
(Received 04 January 2023, Revised 20 February 2023, Accepted 22 February 2023)

Abstract. If a, b, c are the hyperbolic lengths of the sides of a hyperbolic triangle Δ and α, β, γ are the radian measures of the corresponding interior angles of Δ (necessarily such that $\alpha + \beta + \gamma < \pi$), the usual statement of the Hyperbolic Cosine Law (resp., of the Hyperbolic Sine Law; resp., of the trigonometric law in hyperbolic geometry with no Euclidean counterpart, expressing the hyperbolic cosines of a, b and c in terms of α, β and γ) gives several equations involving a, b, c, α, β and γ. Now, let the entries in a 6-tuple $\mathcal{L} = (a, b, c, \alpha, \beta, \gamma)$ be positive real numbers such that $\alpha + \beta + \gamma < \pi$ (but do not necessarily assume that these entries arise as above from some hyperbolic triangle). We say that $\mathcal{L}$ *satisfies* HCL (resp., $\mathcal{L}$ *satisfies* HSL; resp., $\mathcal{L}$ *satisfies* HOL) if the entries of $\mathcal{L}$ satisfy all the equations in the usual statement of the Hyperbolic Cosine Law (resp., of the Hyperbolic Sine Law; resp., of the just-mentioned trigonometric law in hyperbolic geometry with no Euclidean counterpart). For any such L (not necessarily induced by some hyperbolic triangle), we prove the following: $\mathcal{L}$ satisfies HOL $\Leftrightarrow$ $\mathcal{L}$ satisfies HCL $\Rightarrow$ $\mathcal{L}$ satisfies HSL; and we give an example showing that the last implication cannot be reversed. This paper could to used to enrich a course on the classical geometries that discusses hyperbolic plane geometry.
Key Words: Hyperbolic plane geometry, hyperbolic cosine law, hyperbolic sine law, upper half-plane model, Euclidean geometry, neutral geometry.
2010 MSC: Primary 51-02; Secondary 33B10.

1 Introduction

The literature seems to agree that the three most basic/important trigonometric laws of hyperbolic (plane) geometry are the Hyperbolic Cosine Law (as in [14, Theorem 8.3.2 (i)], [10, formula (13), page 337]; dubbed HCL here), the Hyperbolic Sine Law (as in [14, Theorem 8.3.2 (iii)], [10, formula (14), page 337]; dubbed HSL here), and a law with no Euclidean counterpart (as in [14, Theorem 8.3.2 (ii)], [10, formula (15), page 337]; dubbed HOL here, short for "Hyperbolic Other Law"). It seems natural to ask whether one of these three laws is most basic/important. In other words, does one of these laws imply the other two laws (in a sense to be made precise)? The **purpose of this paper** is to answer this question.

Let us first recall the statements of the three laws that were mentioned above. Let $\Delta = \Delta ABC$ be a hyperbolic triangle; let α, β and γ be the (radian) measures of the (interior) angles of Δ at the vertices A, B and C, respectively; and let a, b and c be the hyperbolic lengths of the sides of Δ that are opposite the vertices A, B and C, respectively. Then, as applied to Δ, the HCL states that

$$\cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}, \cos(\beta) = \frac{\cosh(c)\cosh(a) - \cosh(b)}{\sinh(c)\sinh(a)}$$

$$\text{and} \qquad \cos(\gamma) = \frac{\cosh(a)\cosh(b) - \cosh(c)}{\sinh(a)\sinh(b)};$$

the HSL states that

$$\frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\gamma)}{\sinh(c)};$$

and the HOL states that

$$\cosh(a) = \frac{\cos(\beta)\cos(\gamma)+\cos(\alpha)}{\sin(\beta)\sin(\gamma)}, \cosh(b) = \frac{\cos(\gamma)\cos(\alpha)+\cos(\beta)}{\sin(\gamma)\sin(\alpha)}$$

$$\text{and } \cosh(c) = \frac{\cos(\alpha)\cos(\beta)+\cos(\gamma)}{\sin(\alpha)\sin(\beta)}.$$

Next, to motivate a way to make precise the question that was raised above, let us recall how part of a paper [3] may be viewed as having used the classical Sine Law (SL) and the classical Cosine Law (CL) to obtain equational/algebraic characterizations of Euclidean triangles. That essentially involved the behavior, with respect to the equations in the classical statements of the SL and the CL, of what may be called "Euclidean lists", that is, 6-tuples $\mathcal{E} := (a,b,c,\alpha,\beta,\gamma)$ whose entries are positive real numbers such that $\alpha+\beta+\gamma=\pi$. Explicitly, for such a Euclidean list $\mathcal{E}$, the following three conditions were shown in [3] to be equivalent: (1) there exists a Euclidean triangle (which is necessarily unique up to congruence) with multi-set of side-lengths (a,b,c) and multi-set of radian measures of corresponding interior angles (α,β,γ); (2) $\mathcal{E}$ satisfies SL, in the sense that $\sin(\alpha)/a = \sin(\beta)/b = \sin(\gamma)/c$; (3) $\mathcal{E}$ satisfies CL, in the sense that $a^2 = b^2+c^2-2bc\cos(\alpha)$, $b^2 = c^2+a^2-2ca\cos(\beta)$ and $c^2 = a^2+b^2-2ab\cos(\gamma)$.

Note that the above definition of a "Euclidean list" was motivated, in part, by the fact that the sum of the radian measures of the interior angles of any triangle in Euclidean geometry is π. The relevant corresponding fact for our concerns here is that the sum of the radian measures of the interior angles of any triangle in hyperbolic geometry is a positive number which is less than π (cf. [13, Theorem 6.6], [14, Existence axiom, page 89; and Theorem 7.2.1], [10, Theorem 6.1]). (In regard to the two above-mentioned roles of π, see also our comments after the proof of Theorem 2.9, concerning the measurement of angles.) Motivated by this perspective and the preceding paragraph, we now make the following definitions. By a *hyperbolic list*, we mean any 6-tuple $\mathcal{L} = (a,b,c,\alpha,\beta,\gamma)$, all of whose entries are positive real numbers, such that $\alpha+\beta+\gamma<\pi$ (but do not necessarily assume that these entries arise as above from some hyperbolic triangle). We say that $\mathcal{L}$ *satisfies* HCL (resp., $\mathcal{L}$ *satisfies* HSL; resp., $\mathcal{L}$ *satisfies* HOL) if the entries of $\mathcal{L}$ satisfy all the equations in the usual statement of the Hyperbolic Cosine Law (resp., of the Hyperbolic Sine Law; resp., of the "other" law, HOL). We can now make precise the central question which is asked and answered in this paper. For a hyperbolic list $\mathcal{L}$, what are the implications (if any exist) among the following three conditions: "$\mathcal{L}$ satisfies HCL", "$\mathcal{L}$ satisfies HSL", and "$\mathcal{L}$ satisfies HOL"?

In regard to the above question, we next consider three sources of inspiration for possible conjectures. The first of these sources would be the proofs of the HCL, the HSL and the HOL in some standard textbooks. In [10], a text that may be considered mostly axiomatic or "synthetic", these proofs are left as exercises to the reader, with the advice to use some identities involving hyperbolic functions and the case of a hyperbolic right triangle. (That case had been treated a few pages earlier in [10], by appealing to results involving certain classical geometric transformations, such as inversion in circles, that had been used much earlier in the book to prove that Poincaré's disk model satisfies the SAS axiom.) This paper will not consider further this kind of approach (although it *is* valid) because of our desire to create proofs here that are reasonably short and rather algebraic, while not being overly strewn with infrequently used hyperbolic identities. In [14], a text that is mostly "analytic" (inasmuch as it focuses on Poincaré's upper half-plane model), the proofs of the HCL, the HSL and the HOL in [14, Theorem 8.3.2] also depend on first establishing the hyperbolic version of the Pythagorean Theorem (that is, the case of the HCL for hyperbolic right triangles), together with some effective but technical results concerning angles in the upper half-plane model ([14, Propositions 4.1.1 and 6.1.1]) which (according to my colleagues and my own observations) are considered by most students to be difficult to memorize. That textbook also uses some less familiar hyperbolic identities and relegates some cases to exercises. We will not consider further the approach in [14] to this matter, because we desire proofs here that feature comparatively uncomplicated organizations, without recourse to rarely used identities or special (right triangular) cases.

Next, consider, as a possible second source to motivate conjectures, a known result connecting the SL and the CL in classical (Euclidean, also known as circular) trigonometry. Recall from the above discussion that Euclidean lists were used to give some equational characterizations of Euclidean triangles in [3]. Those characterizations were a consequence of our main result in [3], which is the very title of [3], namely, that the SL and the CL are logically equivalent. I can say, with some modesty and embarrassment, that I consider this to be a beautiful result, since it was discovered and published first by Burton [2] some 35 years before I innocently rediscovered and published it. (Burton did not note anything like the notion of a "Euclidean list" or its above-mentioned uses; a couple of elementary results in [3] that had not been noted in [2] will be mentioned in Remark 2.11 (a).) Given the existence of what I just called a "beautiful result" and the attention paid to certain textbooks in the preceding paragraph, it may seem appropriate to review the typical treatments of the SL and the CL in precalculus texts (in hopes of finding relevant connections between what may be termed the "sinh-like" concept of the sine function and the "cosh-like" concept of the cosine function). Several generations of textbooks on Precalculus (and before that, textbooks for the precursor courses that were commonly called "Algebra, trigonometry and analytic geometry") have given a short proof of the CL that uses the roles of cos and sin in describing Cartesian coordinates, the Euclidean distance formula, and the Pythagorean identity $\sin^2(\theta)+\cos^2(\theta)=1$. (That proof can be found at the beginning of [2] and also in [9, page 397].) As that standard textbook proof of the CL does not immediately suggest (to the author) a way forward in our quest, let us turn to standard textbook (non-vectorial) proofs of the SL. These are all essentially the same as the proof in [9, pages 388-389], which features a case analysis, uses of the subcase for right triangles, and the standard expansion formula for the sine of a difference, but with no other mention of the cosine function. It would seem that the standard textbook proofs of the SL have also failed to suggest a way forward. What, one may ask, can be learned from the separate treatments of the SL and the CL in textbooks? Our limited search in this regard (cf. [9, Sections 7.2 and 7.3]) indicate that the SL and the CL are typically covered in separate sections of the same chapter, with the worked examples and the exercises in each of those sections devised/contrived so that the ambient law would be rather easily applicable while the other law's applicability was neither apparent nor mentioned. In short, perusal of standard textbooks has apparently failed to provide the desired motivation. So, one can now ask if there is any other possible benefit to be gained from further consideration of the Euclidean/circular context. As a last hope in this regard, one may suggest, using the fact that the SL and CL are known to be equivalent, that we conjecture that the same is true for the hyperbolic analogues, the HSL and the HCL. Moreover, since there is no Euclidean analogue of the HOL, one may be tempted to conjecture that the HOL is conspicuously "stronger" or conspicuously "weaker" than the HCL. Unfortunately, as will be shown in this paper, *none* of the three conjectures that were made in the preceding two sentences is valid! It is time to consider the third possible source of inspiration.

Lastly, consider, as a third possible source for motivation, some possibly relevant results in spherical geometry. In the final comment in [2], Burton recalls (without proof) that in spherical geometry, the relevant law of cosines implies the relevant law of sines and also states that the converse is false. These facts may not seem too surprising, even in light of [2] (and [3]), as spherical geometry is not (isomorphic to) a neutral geometry. Indeed, up to isomorphism, plane Euclidean geometry and plane hyperbolic geometry are the only neutral geometries (cf. [1]). Nevertheless, in some ways, spherical geometry does seem more useful than hyperbolic geometry for certain activities and measurements on Earth. Moreover, as Greenberg [10, page 329] points out, all the formulas of hyperbolic trigonometry can be obtained, via the substitution $r = ik$, from formulas known for spherical trigonometry. One may thus be tempted to conjecture, and then to try to use the above-mentioned facts from spherical trigonometry to prove, that the HCL implies the HSL (more precisely, that "$\mathcal{L}$ satisfies HCL" implies "$\mathcal{L}$ satisfies HSL") and that the converse is false. It will be shown in this paper that these two conjectures are valid and that even more (involving the HOL) is true. However, as the next paragraph explains, our proofs will not make use of spherical geometry.

The reason for our decision not to use spherical trigonometry in our proofs here comes from a footnote on page 329 of [10]: to understand why the substitution $r = ik$ has the above-mentioned marvelous consequences, one would need to undertake a study of differential geometry and symmetric spaces. As a graduate student, I became familiar with the reference that Greenberg suggested for such a study and I esteem that book by Helgason highly. But such a study would surely take us too far afield or introduce unnecessarily steep prerequisites for reading this paper. The methods of proof in this paper are much more accessible, in our opinion. Moreover, our methods will naturally also address the HOL alongside our considerations of the HCL and the HSL. For readers who are familiar with spherical geometry and are still wondering about the substitution $r = ik$: of course, i is the usual preferred complex number such that $i^2 + 1 = 0$, and we will say a little more about the constant k prior to Corollary 2.10.

All but one of the main results of the present paper can be summarized as follows. If $\mathcal{L}$ is a hyperbolic list, then: $\mathcal{L}$ satisfies HOL $\Leftrightarrow$ $\mathcal{L}$ satisfies HCL $\Rightarrow$ $\mathcal{L}$ satisfies HSL, and an example shows that the last implication cannot be reversed. Any reader who wishes to take the shortest path to proofs of the just-summarized information need only read the following items herein: either Theorem 2.1 and Corollary 2.2 or Theorem 2.6; Example 2.3; Lemma 2.7; Theorem 2.8; and Theorem 2.9. All readers are advised that the proof of Theorem 2.9 is different in nature from the equation-manipulation methods which are in the spirit of [2], [3] and much of the early material in this paper. Indeed, the proof of Theorem 2.9 uses some basic facts from hyperbolic geometry. A careful reading of the references in that proof will require the reader to read some proofs in [8] that are couched within the upper half-plane model of hyperbolic geometry, and some of those references will require the reader to read some of the early results in [7]. Much of the relevant material in those just-cited references concerns the definition and measurement of angles in the upper half-plane model. That topic has been treated in (in chronological order) [11], [12] and [7]. We will comment further on this topic following the proof of Theorem 2.9.

The above-mentioned "shortest path to proofs" to most of our main results did not list every result in this paper. While there is some redundancy, in a strictly logical sense, in the sequencing of results in this paper, that redundancy is intentional, for the following two reasons. First, I believe that the redundancy serves a good pedagogic purpose, especially for instructors who may choose to use this paper to enrich their courses on the classical geometries. Second, the sequencing of results has been chosen to reflect a likely sequence of questions/answers that would occur to a researcher or student with the following qualities: he/she has expertise in Euclidean geometry and the upper half-plane model of hyperbolic geometry, but does not wish to apply sophisticated geometric or analytic results from the literature for this work; he/she is aware of the existence of [2] (and possibly [3]) and wonders whether the HOL (the law of hyperbolic trigonometry that has no Euclidean counterpart) is "stronger" or "weaker", in some intuitive sense, than its partner laws in hyperbolic trigonometry (the HCL and the HSL) which *do* have Euclidean counterparts. To help readers to navigate that intentional redundancy, the next three paragraphs summarize this paper's results in the order in which they appear. We hope that a reader will agree that our sequencing has addressed what we termed "a likely sequence of questions/answers".

This paragraph contains a summary of the sequence of results that produces an answer to what we termed "the central question which is asked and answered in this paper". Since the above considerations that involved only "sinh-like" or "cosh-like" concepts were not immediately fruitful, Section 2 begins by considering the impact of assuming the HOL. Corollary 2.2 establishes the fact that "$\mathcal{L}$ satisfies HOL" implies "$\mathcal{L}$ satisfies HSL", as a consequence of a sharper result in Theorem 2.1. Turning to the question of a possible converse, we show in Example 2.3 that "$\mathcal{L}$ satisfies HSL" does not, in fact, imply "$\mathcal{L}$ satisfies HOL". Having successfully determined the logical connections between the HOL and the HSL, we return in Example 2.4 to the question of possible connections between the HSL and the HCL. By a further examination of the data constructed in Example 2.3, we show in Example 2.4 that "$\mathcal{L}$ satisfies HSL" does not imply "$\mathcal{L}$ satisfies HCL". With (our) intuition in short supply, we pause in Remark 2.5 to examine some apparently random data and, after some calculation, develop some evidence which indicates "$\mathcal{L}$ satisfies HCL" may possibly/probably imply both "$\mathcal{L}$ satisfies HSL" and "$\mathcal{L}$ satisfies HOL". (Of course, the first of these implications would follow from the second implication, in view of Corollary 2.2, and so an instructor may wish to cover at least part of Remark 2.5 at the very outset.) Theorem 2.6 addresses the easier of the two conjectures which had emanated from Remark 2.5. Indeed, by revisiting Stahl's proof of [14, (iii), page 106], we show in Theorem 2.6 that "$\mathcal{L}$ satisfies HCL" implies "$\mathcal{L}$ satisfies HSL". As one may expect from the cumbersome nature of some of the details in Remark 2.5, the theoretical considerations and numerical calculations needed for the remaining results are more substantial. Following (and using) a lemma that is of some independent interest, we show in Theorem 2.8 that "$\mathcal{L}$ satisfies HOL" implies "$\mathcal{L}$ satisfies HCL". Readers are advised that the calculations in the proof of Theorem 2.8 can be carried out by hand, but some readers may prefer to execute them using technology. Finally, as explained earlier, Theorem 2.9 has a theoretically demanding proof, leading to the conclusion that "$\mathcal{L}$ satisfies HCL" implies "$\mathcal{L}$ satisfies HOL".

The partial summary of "main results" that was given above did not mention one such result. It is Corollary 2.10. This result is in the spirit of the above-mentioned use in [3] of Euclidean lists that explained how the Sine Law and Cosine Law can each be used to give an algebraic/equational characterization of triangles in Euclidean geometry. In a similar sense, Corollary 2.10 makes precise the way in which hyperbolic lists can be used to explain how the HOL and the HCL (but not the HSL) can each be used to give an algebraic/equational characterization of triangles in hyperbolic geometry.

Finally, we close with several comments in Remark 2.11. Its part (d) builds on [4, Theorem 7], a result that produced another way to use suitable 6-tuples of real numbers to give algebraic/equational characterizations of triangles in Euclidean geometry. That result gave a sharp converse to a result of Isaac Newton. Remark 2.11 (d) summarizes the work in [4], for two reasons: for the sake of completeness; and in the hope of stimulating further work on finding algebraic/equational characterizations of various kinds of geometric entities within various hyperbolic universes. In particular, we suggest a possible way to attack the question of whether the hyperbolic tangent function (tanh) can be used to obtain characterizations of triangles in hyperbolic geometry that are analogous to the result for Euclidean geometry in [4, Theorem 7].

2 Results

The Introduction gave an intuitive reason to think that what we have called the "other" trigonometric law may have some strong implicative powers. We begin to confirm that suspicion in Corollary 2.2, by inferring the equations of the Hyperbolic Sine Law from the equations of that "other" law. First, Theorem 2.1 gives an even sharper result.

The proof of Theorem 2.1, along with several of the later proofs, will make frequent use of the standard identities $\sin^2(\theta)+\cos^2(\theta)=1$ and $\cosh^2(d)-\sinh^2(d)=1$. We will also often use the well known facts (cf. [8, page 364]) that cosh and sinh are each differentiable (hence continuous) functions defined on $\mathbb{R}$; the restriction of cosh to the domain $[0,\infty)$ is a strictly increasing (and continuous) function whose range is $[1,\infty)$; and the restriction of sinh to the domain $[0,\infty)$ is a strictly increasing (and continuous) function whose range is $[0,\infty)$. One useful consequence (see, especially, Examples 2.3 and 2.4) is that a positive real number d can be (uniquely) determined by specifying the positive real number $\sinh(d)$.

Theorem 2.1. Let $\mathcal{L} = (a,b,c,\alpha,\beta,\gamma)$ be a hyperbolic list. If $\mathcal{L}$ satisfies (at least) two of the three associated equations from the HOL, then $\mathcal{L}$ satisfies the corresponding equation from the HSL. More precisely: if, for instance,

$$\cosh(a) = \frac{\cos(\beta)\cos(\gamma)+\cos(\alpha)}{\sin(\beta)\sin(\gamma)} \text{ and } \cosh(b) = \frac{\cos(\gamma)\cos(\alpha)+\cos(\beta)}{\sin(\gamma)\sin(\alpha)},$$

then

$$\frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)}.$$

Proof. As $0 < \alpha, \beta < \pi$, both $\sin(\alpha)$ and $\sin(\beta)$ are positive real numbers. By the above comments, the same conclusion holds for $\sinh(a)$ and $\sinh(b)$ (since $0 < a, b$). Hence, we can reformulate our task (after cross-multiplying and squaring both sides of the desired equation) as follows: to prove that

$$\sin^2(\alpha)\sinh^2(b) = \sin^2(\beta)\sinh^2(a); \text{ or, equivalently, that}$$

$\sin^2(\alpha)(\cosh^2(b)-1) = \sin^2(\beta)(\cosh^2(a)-1)$. Using the assumed expressions for $\cosh(a)$ and $\cosh(b)$, we can reformulate our task as finding a proof that

$$\sin^2(\alpha)\left(\frac{(\cos(\gamma)\cos(\alpha)+\cos(\beta))^2}{\sin^2(\gamma)\sin^2(\alpha)} - 1\right) =$$

$$\sin^2(\beta)\left(\frac{(\cos(\beta)\cos(\gamma)+\cos(\alpha))^2}{\sin^2(\beta)\sin^2(\gamma)} - 1\right); \text{ or, equivalently, that}$$

$$\frac{\sin^2(\alpha)(\cos^2(\gamma)\cos^2(\alpha)+2\cos(\gamma)\cos(\alpha)\cos(\beta)+\cos^2(\beta)-\sin^2(\gamma)\sin^2(\alpha))}{\sin^2(\gamma)\sin^2(\alpha)} =$$

$$\frac{\sin^2(\beta)(\cos^2(\beta)\cos^2(\gamma)+2\cos(\beta)\cos(\gamma)\cos(\alpha)+\cos^2(\alpha)-\sin^2(\beta)\sin^2(\gamma))}{\sin^2(\beta)\sin^2(\gamma)}.$$

A final reformulation of the task is found by rewriting the most recent version of the equation by doing the following four steps: cancel the fractions $(\sin^2(\alpha))/(\sin^2(\alpha))$ and $(\sin^2(\beta))/(\sin^2(\beta))$; multiply both sides of the equation by $\sin^2(\gamma)$; then use the identity $\cos^2(\theta) = 1-\sin^2(\theta)$ to rewrite each of $\cos^2(\gamma)$, $\cos^2(\alpha)$ and $\cos^2(\beta)$; and then additively cancel the common term $2\cos(\gamma)\cos(\alpha)\cos(\beta)$. Both sides of the thus rewritten (but still equivalent) equation easily simplify to

$$2-\sin^2(\alpha)-\sin^2(\beta)-\sin^2(\gamma),$$

and so the proof is complete. □

Corollary 2.2. HOL ⇒ HSL. *More precisely, if $\mathcal{L}$ is a hyperbolic list that satisfies* HOL, *then $\mathcal{L}$ satisfies* HSL.

Proof. The assertion follows easily by applying Theorem 2.1 twice. □

It seems natural to ask whether the converse of Corollary 2.2 is valid. This question is answered in the negative by the next example.

Example 2.3. HSL $\not\Rightarrow$ HOL. More precisely, if $\mathcal{L} = (a,b,c,\alpha,\beta,\gamma)$ where $\cosh(a) = \sqrt{2} = \cosh(b)$, $\cosh(c) = \sqrt{5+2\sqrt{3}}$, $\alpha = \pi/12 = \beta$ and $\gamma = 3\pi/4$, then $\mathcal{L}$ is a hyperbolic list that satisfies HSL but does not satisfy HOL.

Proof. As $\sqrt{2}$ and $\sqrt{5+2\sqrt{3}}$ are positive real numbers, the comment prior to Theorem 2.1 ensures that a, b and c are well defined positive real numbers. Also, $\alpha+\beta+\gamma = 11\pi/12 < \pi$. Hence, $\mathcal{L}$ is a hyperbolic list. It is straightforward to verify that $\mathcal{L}$ satisfies HSL. Indeed,

$$\frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\pi/12)}{\sqrt{\cosh^2(a)-1}} = \frac{\frac{\sqrt{2-\sqrt{3}}}{2}}{\sqrt{\sqrt{2}^2-1}} = \frac{\sqrt{2-\sqrt{3}}}{2} \text{ and}$$

$$\frac{\sin(\gamma)}{\sinh(c)} = \frac{\sin(3\pi/4)}{\sqrt{\cosh^2(c)-1}} = \frac{\frac{1}{\sqrt{2}}}{\sqrt{(5+2\sqrt{3})-1}} = \frac{1}{2\sqrt{2+\sqrt{3}}} = \frac{\sqrt{2-\sqrt{3}}}{2}.$$

However, $\mathcal{L}$ does not satisfy HSL. Here is one way to see this:

$$\frac{\cos(\alpha)\cos(\beta)+\cos(\gamma)}{\sin(\alpha)\sin(\beta)} = \frac{\cos^2(\frac{\pi}{12})+\cos(\frac{3\pi}{4})}{\sin^2(\frac{\pi}{12})} =$$

$$\frac{\frac{2+\sqrt{3}}{4}+\frac{-1}{\sqrt{2}}}{\frac{2-\sqrt{3}}{4}} = \frac{2+\sqrt{3}-2\sqrt{2}}{2-\sqrt{3}} \neq \sqrt{5+2\sqrt{3}} = \cosh(c).$$

The proof is complete. □

The preceding proof mentioned "one way to see ...". This raises the following question. Does the hyperbolic list that was studied in Example 2.3 satisfy *any* of the associated equations from the HOL? It is straightforward to check that the answer is "No". That answer, in turn, raises the following questions. Does there exist a hyperbolic list that satisfies HSL and also satisfies exactly one (resp., exactly two) of the associated equations from the HOL? One can ask this type of question in regard to several of the results in this paper. While we have not considered such questions carefully, we hope that some readers will choose to do so, as there are important analogues/precedents in Euclidean geometry that can be traced back, at least in spirit, to some work of Isaac Newton: see Remark 2.11 (d) below.

In light of the Example 2.3, it now seems natural to ask whether the HSL is strong enough to imply the HCL. This question is answered in the negative by the next example.

Example 2.4. HSL $\not\Rightarrow$ HCL. More precisely, if $\mathcal{L}$ is the hyperbolic list studied in Example 2.3, then $\mathcal{L}$ satisfies HSL but $\mathcal{L}$ does not satisfy HCL.

Proof. Recall from the statement and proof of Example 2.3 that $\mathcal{L} = (a,b,c,\alpha,\beta,\gamma)$, where $\cosh(a) = \sqrt{2} = \cosh(b)$, $\cosh(c) = \sqrt{5+2\sqrt{3}}$, $\alpha = \pi/12 = \beta$, $\gamma = 3\pi/4$, $\sin(\alpha) = (\sqrt{2-\sqrt{3}})/2 = \sin(\beta)$, $\sin(\gamma) = 1/\sqrt{2}$, $\cos(\alpha) = (\sqrt{2+\sqrt{3}})/2 = \cos(\beta)$, $\cos(\gamma) = -1/\sqrt{2}$, $\sinh(a) = \sqrt{\sqrt{2}^2-1} = 1 = \sinh(b)$, $\sinh(c) = 2\sqrt{2+\sqrt{3}}$, and $\mathcal{L}$ satisfies HSL. It remains to show that $\mathcal{L}$ does not satisfy HCL. It will be enough to verify that

$$\cos(\alpha) \neq \frac{\cosh(b)\cosh(c)-\cosh(a)}{\sinh(b)\sinh(c)}.$$

That verification will be carried out in the next paragraph. Because the ambient data satisfies $a = b$ and $\alpha = \beta$, it will then follow that

$$\cos(\beta) \neq \frac{\cosh(c)\cosh(a) - \cosh(b)}{\sinh(c)\sinh(a)}.$$

Pursuing a theme that was mentioned following Example 2.3, the interested reader is invited to then also verify that

$$\cos(\gamma) \neq \frac{\cosh(a)\cosh(b) - \cosh(c)}{\sinh(a)\sinh(b)},$$

as that would then complete a proof that the ambient hyperbolic list $\mathcal{L}$ does not satisfy *any* of the associated equations from the HCL.

Observe that $\cos(\alpha) = (\sqrt{2+\sqrt{3}})/2 \approx 0.9659$. On the other hand,

$$\frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)} = \frac{\sqrt{2}\sqrt{5+2\sqrt{3}} - \sqrt{2}}{1 \cdot 2\sqrt{2+\sqrt{3}}} \approx 0.6989,$$

which is certainly not "approximately" 0.9659. (Some readers may prefer a less "numerical" verification, that is, a more traditional "algebraic" verification, that the equation in question fails. They may produce one by using the fact that for positive real numbers r and s, one has that $r = s$ if and only if $r^2 = s^2$. Using that fact twice, together with some related elementary simplifications, one shows easily that if the equation in question were to hold, then it would follow that $25 = 16(5+2\sqrt{3})$, an equation which is absurd for *several* reasons! A similar approach would show that if the above possible equational description of $\cos(\gamma)$ were to hold, then it would follow that $10 + 4\sqrt{3} = 9 + 4\sqrt{2}$, another equation that is absurd for several reasons.) The proof is complete. □

The last two results have shown that the Hyperbolic Sine Law does not imply either of its usual companions. The next result collects some evidence suggesting that the Hyperbolic Cosine Law may have stronger implicative powers.

Remark 2.5. (a) This remark presents evidence indicating that the HCL may imply both the HSL and the HOL. This evidence involves the behavior of a certain hyperbolic list which is arguably "more random" that the list which was studied in Examples 2.3 and 2.4 (inasmuch as that earlier list was "isosceles" in the obvious sense, whereas a, b and c are pairwise distinct in the list studied below).

Consider the list $\mathcal{L} := (a,b,c,\alpha,\beta,\gamma)$, where the ordered triples $(a,b,c) \in (0,\infty)^3$ and $(\alpha,\beta,\gamma) \in (0,\pi)^3$ are (uniquely) determined by $\cosh(a) = 2$, $\cosh(b) = 3$, $\cosh(c) = 4$, $\alpha := \cos^{-1}(5/\sqrt{30})$, $\beta := \cos^{-1}(\sqrt{5}/3)$ and $\gamma := \cos^{-1}(1/\sqrt{6})$. Note that $\mathcal{L}$ is a hyperbolic list, since

$$0 < \alpha + \beta + \gamma = \cos^{-1}(5/\sqrt{30}) + \cos^{-1}(\sqrt{5}/3) + \cos^{-1}(1/\sqrt{6}) <$$

$$0.43 + 0.73 + 1.16 = 2.32 < \pi.$$

In this part (a), we will show that $\mathcal{L}$ satisfies HCL and HSL, while we will show in (b) that $\mathcal{L}$ satisfies HOL.

First, let us collect the numerical information that will be needed for the calculations in this remark. We have $\sinh(a) = \sqrt{\cosh(a)^2 - 1} = \sqrt{2^2 - 1} = \sqrt{3}$; and, similarly, $\sinh(b) = 2\sqrt{2}$ and $\sinh(c) = \sqrt{15}$. Also, $\cos(\alpha) = 5/\sqrt{30}$, $\cos(\beta) = \sqrt{5}/3$, $\cos(\gamma) = 1/\sqrt{6}$ and $\sin(\alpha) = \sqrt{1 - \cos^2(\alpha)} = \sqrt{1 - 5/6} = 1/\sqrt{6}$; similarly, $\sin(\beta) = 2/3$ and $\sin(\gamma) = \sqrt{30}/6$.

The task of showing that $\mathcal{L}$ satisfies HCL consists of verifying the following three equations:

$$\cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}, \cos(\beta) = \frac{\cosh(c)\cosh(a) - \cosh(b)}{\sinh(c)\sinh(a)}$$

$$\text{and} \qquad \cos(\gamma) = \frac{\cosh(a)\cosh(b) - \cosh(c)}{\sinh(a)\sinh(b)}.$$

The above numerical information allows us to reformulate the task as verifying the following three equations:

$$\frac{5}{\sqrt{30}} = \frac{3 \cdot 4 - 2}{2\sqrt{2} \cdot \sqrt{15}}, \ \frac{\sqrt{5}}{3} = \frac{4 \cdot 2 - 3}{\sqrt{15}\sqrt{3}}, \text{ and } \frac{1}{\sqrt{6}} = \frac{2 \cdot 3 - 4}{\sqrt{3} \cdot 2\sqrt{2}}.$$

Each of the three just-displayed equations can be easily checked by using the fact that $\sqrt{rs} = \sqrt{r}\sqrt{s}$ for all positive real numbers r and s. Hence, $\mathcal{L}$ satisfies HCL.

Next, the task of showing that $\mathcal{L}$ satisfies HSL consists of verifying that $\sin(\alpha)/\sinh(a) = \sin(\beta)/\sinh(b) = \sin(\gamma)/\sinh(c)$, that is, that

$$\frac{\frac{1}{\sqrt{6}}}{\sqrt{3}} = \frac{\frac{2}{3}}{2\sqrt{2}} = \frac{\frac{\sqrt{30}}{6}}{\sqrt{15}}.$$

The just-displayed equations can be easily checked (once again, by using $\sqrt{rs} = \sqrt{r}\sqrt{s}$ for all $r, s > 0$). Hence, $\mathcal{L}$ satisfies HSL.

(b) It remains to verify that $\mathcal{L}$ satisfies HOL, namely, that the following three equations hold:

$$\cosh(a) = \frac{\cos(\beta)\cos(\gamma) + \cos(\alpha)}{\sin(\beta)\sin(\gamma)}, \cosh(b) = \frac{\cos(\gamma)\cos(\alpha) + \cos(\beta)}{\sin(\gamma)\sin(\alpha)}$$

$$\text{and } \cosh(c) = \frac{\cos(\alpha)\cos(\beta) + \cos(\gamma)}{\sin(\alpha)\sin(\beta)}.$$

The numerical information in (a) allows us to reformulate this task as verifying that

$$2 = \frac{(\frac{\sqrt{5}}{3})(\frac{1}{\sqrt{6}}) + \frac{5}{\sqrt{30}}}{(\frac{2}{3})(\frac{\sqrt{30}}{6})}, 3 = \frac{(\frac{1}{\sqrt{6}})(\frac{5}{\sqrt{30}}) + \frac{\sqrt{5}}{3}}{(\frac{\sqrt{30}}{6})(\frac{1}{\sqrt{6}})} \text{ and } 4 = \frac{(\frac{5}{\sqrt{30}})(\frac{\sqrt{5}}{3}) + \frac{1}{\sqrt{6}}}{(\frac{1}{\sqrt{6}})(\frac{2}{3})}.$$

We leave it to the reader to check the three just-displayed equations (once again, by using $\sqrt{rs} = \sqrt{r}\sqrt{s}$ for all $r, s > 0$). Hence, $\mathcal{L}$ satisfies HOL. This completes the remark.

We next establish the result that was suggested by the example presented in Remark 2.5 (a). In fact, Theorem 2.6 is essentially known. For the sake of completeness, we will prove it by repurposing Stahl's proof of [14, (iii), page 106].

Theorem 2.6. (Cf. [14, proof of (iii), page 106]) HCL $\Rightarrow$ HSL. More precisely, if $\mathcal{L} = (a, b, c, \alpha, \beta, \gamma)$ is a hyperbolic list that satisfies HCL, then $\mathcal{L}$ satisfies HSL.

Proof. Our task is to show that if

$$\cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}, \cos(\beta) = \frac{\cosh(c)\cosh(a) - \cosh(b)}{\sinh(c)\sinh(a)}$$

$$\text{and} \qquad \cos(\gamma) = \frac{\cosh(a)\cosh(b) - \cosh(c)}{\sinh(a)\sinh(b)},$$

$$\text{then} \qquad \frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\gamma)}{\sinh(c)}.$$

Recall that positive real numbers r and s are equal if and only if $r^2 = s^2$. Since the facts that $a, b, c > 0$ and $0 < \alpha, \beta, \gamma < \pi$ ensure that each of $\sinh(a)$, $\sinh(b)$, $\sinh(c)$, $\sin(\alpha)$, $\sin(\beta)$ and $\sin(\gamma)$ is a positive real number, the same can be said of each of $\sin(\alpha)/\sinh(a)$, $\sin(\beta)/\sinh(b)$ and $\sin(\gamma)/\sinh(c)$. Therefore, it will suffice to show that

$$\frac{\sin^2(\alpha)}{\sinh^2(a)} = \frac{\sin^2(\beta)}{\sinh^2(b)} = \frac{\sin^2(\gamma)}{\sinh^2(c)}.$$

Thus, because of the various symmetries exhibited by the data, it will obviously suffice to prove that there exists an expression, say (*), which is invariant under any permutation of $\{a, b, c\}$ such that if the first of the above three equations arising from the HCL holds, then $\sin^2(\alpha)/\sinh^2(a) = (*)$. Therefore, it will suffice to prove that

$$\text{if } \cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}, \text{ then } \frac{\sin^2(\alpha)}{\sinh^2(a)} =$$

$$\frac{1 - \cosh^2(a) - \cosh^2(b) - \cosh^2(c) + 2\cosh(a)\cosh(b)\cosh(c)}{\sinh^2(a)\sinh^2(b)\sinh^2(c)}.$$

Using the above-mentioned description of $\cos(\alpha)$, along with the identities $\sin^2(\theta) = 1 - \cos^2(\theta)$ and $\sinh^2(d) = \cosh^2(d) - 1$, we get

$$\frac{\sin^2(\alpha)}{\sinh^2(a)} = \frac{1 - \cos^2(\alpha)}{\sinh^2(a)} = \frac{1 - (\frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)})^2}{\sinh^2(a)} =$$

$$\frac{\sinh^2(b)\sinh^2(c) - (\cosh(b)\cosh(c) - \cosh(a))^2}{\sinh^2(a)\sinh^2(b)\sinh^2(c)} =$$

$$\frac{(\cosh^2(b) - 1)(\cosh^2(c) - 1) - (\cosh(b)\cosh(c) - \cosh(a))^2}{\sinh^2(a)\sinh^2(b)\sinh^2(c)} =$$

$$\frac{1 - \cosh^2(a) - \cosh^2(b) - \cosh^2(c) + 2\cosh(a)\cosh(b)\cosh(c)}{\sinh^2(a)\sinh^2(b)\sinh^2(c)},$$

as desired. The proof is complete. □

Theorem 2.8 will establish the result that was suggested by the example presented in Remark 2.5 (b). As the proof of Theorem 2.8 is somewhat deeper than the earlier proofs, we will lighten (or, at least, redistribute) the reader's burden by first isolating part of that proof as Lemma 2.7 (c). We anticipate that Lemma 2.7 (b) may be of some independent interest.

Lemma 2.7. *Let x, y and z be real numbers, possibly listed with repetition. Then:*

(*a*) *Suppose that $z > 0$, $x + y > 0$ and $x + y + z < \pi$. Then*

$$\cos(x)\cos(y) + \cos(z) > \sin(x)\sin(y).$$

(*b*) *Suppose that $z > 0$, $x + y > 0$, $x + y + z < \pi$, neither x nor y is an integral multiple of π, and $\sin(x)$ and $\sin(y)$ have the same (algebraic) sign. Then*

$$2\cos(x)\cos(y)\cos(z) + \cos^2(x) + \cos^2(y) + \cos^2(z) > 1.$$

(*c*) *Suppose that all of x, y and z are positive (real numbers), such that $x + y + z < \pi$. Then*

$$2\cos(x)\cos(y)\cos(z) + \cos^2(x) + \cos^2(y) + \cos^2(z) > 1.$$

Proof. (a) Put $u := \pi/2 - x$ and $v := \pi/2 - y$. Then

$$0 < z < \pi - (x+y) = u + v.$$

Since $\cos|_{(0,\pi)}$ is a strictly decreasing function, the expansion formula for the cosine of a sum (cf. [9, page 332]) now gives

$$\cos(z) > \cos(u+v) = \cos(u)\cos(v) - \sin(u)\sin(v) =$$
$$\cos([\frac{\pi}{2} - x])\cos([\frac{\pi}{2} - y]) - \sin([\frac{\pi}{2} - x])\sin([\frac{\pi}{2} - y]).$$

Using the cofunction identities involving the sine and cosine functions (cf. [9, page 263]), we can rewrite the extreme parts of last display as

$$\cos(z) > \sin(x)\sin(y) - \cos(x)\cos(y).$$

The asserted inequality follows at once.

(b) As $\sin(x)$ and $\sin(y)$ are nonzero real numbers having the same algebraic sign, $\sin(x)\sin(y) > 0$. Also, by (a),

$$\cos(x)\cos(y) + \cos(z) > \sin(x)\sin(y).$$

It follows that the just-displayed inequality persists if we square both of its sides; that is,

$$(\cos(x)\cos(y) + \cos(z))^2 > (\sin(x)\sin(y))^2.$$

Thus,

$$\cos^2(x)\cos^2(y) + 2\cos(x)\cos(y)\cos(z) + \cos^2(z) > \sin^2(x)\sin^2(y) =$$
$$(1 - \cos^2(x))(1 - \cos^2(y)) = 1 - \cos^2(y) - \cos^2(x) + \cos^2(x)\cos^2(y).$$

The asserted inequality follows at once.

(c) The hypotheses for (c) imply the hypotheses for (b). Therefore, an application of (b) completes the proof. □

We pause for some theoretical observations, followed by a pedagogical comment. First, observe that Lemma 2.7(a) generalizes [8, Proposition 2.1]. Furthermore, the former result is strictly stronger than the latter result (as the latter result assumed that each of x, y and z is positive and that $x+y+z < \pi$). Indeed, an example satisfying the hypotheses of Lemma 2.7(a) in which $x > 0$ and $y < 0$ is given by $x := 5\pi/4$, $y := -\pi$ and $z := \pi/2$. A more exotic example satisfying the hypotheses of Lemma 2.7(b) (and of Lemma 2.7(a), but not of [8, Proposition 2.1]) is given by $x := 11\pi/8$, $y := -7\pi/8$ and $z := \pi/4$. Next, not only is Lemma 2.7 (a) a sharper result than [8, Proposition 2.1], the proof given above of the former result is more elementary than the published proof of [8, Proposition 2.1]. Indeed, the latter proof used (differential) calculus, while the proof of the former result used only precalculus.

Theorem 2.8. HOL $\Rightarrow$ HCL. More precisely, if $\mathcal{L} = (a, b, c, \alpha, \beta, \gamma)$ is a hyperbolic list that satisfies HOL, then $\mathcal{L}$ satisfies HCL.

Proof. Our task is to show that if

$$\cosh(a) = \frac{\cos(\beta)\cos(\gamma) + \cos(\alpha)}{\sin(\beta)\sin(\gamma)}, \cosh(b) = \frac{\cos(\gamma)\cos(\alpha) + \cos(\beta)}{\sin(\gamma)\sin(\alpha)}$$

$$\text{and } \cosh(c) = \frac{\cos(\alpha)\cos(\beta) + \cos(\gamma)}{\sin(\alpha)\sin(\beta)}, \text{ then}$$

$$\cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}, \cos(\beta) = \frac{\cosh(c)\cosh(a) - \cosh(b)}{\sinh(c)\sinh(a)}$$

$$\text{and} \qquad \cos(\gamma) = \frac{\cosh(a)\cosh(b) - \cosh(c)}{\sinh(a)\sinh(b)}.$$

We will prove the first of the desired equations, leaving the similar details of the proofs for the desired equational descriptions of $\cos(\beta)$ and $\cos(\gamma)$ to the reader. Our task now is to show that

$$\frac{(\frac{\cos(\gamma)\cos(\alpha)+\cos(\beta)}{\sin(\gamma)\sin(\alpha)})(\frac{\cos(\alpha)\cos(\beta)+\cos(\gamma)}{\sin(\alpha)\sin(\beta)}) - \frac{\cos(\beta)\cos(\gamma)+\cos(\alpha)}{\sin(\beta)\sin(\gamma)}}{\sinh(b)\sinh(c)} = \cos(\alpha).$$

The upcoming calculations will make the introduction of some simplifying notation expedient (and, dare I say, essential if one is to do those calculations by hand). Let

$$d := \sin(\alpha), f := \cos(\alpha), g := \sin(\beta), h := \cos(\beta), j := \sin(\gamma), k := \cos(\gamma).$$

Next, marshal the following: the above equational descriptions of $\cosh(b)$ and $\cosh(c)$; the fact that d, g and j are all positive since $0 < \alpha, \beta, \gamma < \pi$; and the fact that $\sinh(\lambda) = \sqrt{\cosh^2(\lambda) - 1}$ for all $\lambda > 0$. Then, by cross-multiplying, our task is thus reformulated as needing to prove that

$$\frac{(fk+h)(fh+k) - d^2(hk+f)}{d^2gj} =$$

$$f(\frac{\sqrt{(fk+h)^2 - d^2j^2}}{dj})(\frac{\sqrt{(fh+k)^2 - d^2g^2}}{dg}).$$

Since g and j (and d^2) are positive, an equivalent task is to prove that

$$(fk+h)(fh+k) - d^2(hk+f) =$$

$$f(\sqrt{[(fk+h)^2 - d^2j^2][(fh+k)^2 - d^2g^2]}); \text{ or,}$$

by replacing d^2 with $1 - f^2$, equivalently, that

$$2f^2hk + fk^2 + fh^2 - f + f^3 = f\sqrt{T},$$

where T is a certain polynomial expression in six "variables" (d, f, g, h, j and k) such that T is a sum of 16 monomials. The reader should carefully identify T; check, in particular, that each constituent monomial of T is a product of a nonzero integer and integral powers of at least some of the six variables with exponents on the non-trivially participating variables ranging from 1 to 4. Next, to obtain another restatement of our task, do the following: rewrite the last-displayed equation (including the explicit expression for T which the reader has obtained) by replacing d^2, g^2 and j^2 with $1 - f^2$, $1 - h^2$ and $1 - k^2$, respectively; then divide through by f; and then carry out various elementary algebraic simplifications involving the distributive property. The upshot is a reformulation of our task, as needing to prove that

$$2fhk + k^2 + h^2 - 1 + f^2 = \sqrt{S},$$

where S is a certain polynomial expression in three "variables" (f, h and k) such that S is a sum of 45 monomials. The reader should carefully identify S; check, in particular, that each constituent monomial of S is a product of a nonzero integer and integral powers of at least some of the three variables with exponents on the non-trivially participating variables ranging from 1 to 4.

Next, for the most tedious part of this proof, the reader is requested to verify that the result of squaring both sides of the last-displayed equation is a valid equation. (First, expand the left-hand side; then combine, and then compare, “like" terms.) In other words, check that

$$(2fhk + k^2 + h^2 - 1 + f^2)^2 = S.$$

The reader may choose to use some trusted software to verify the just-displayed equation. The author attests that he has checked this equation twice by hand (though, admittedly, his first effort along those lines failed due to human error.)

The main consequence of combining the results from the preceding two paragraphs is that the assertion holds if

$$2fhk + k^2 + h^2 - 1 + f^2 \geq 0; \text{ that is, if}$$
$$2\cos(\alpha)\cos(\beta)\cos(\gamma) + \cos^2(\gamma) + \cos^2(\beta) - 1 + \cos^2(\alpha) \geq 0.$$

Fortunately, Lemma 2.7 established an inequality that is stronger than the just-displayed inequality, and so the proof is complete. □

When compared to the preceding proofs (and to the proofs in [2] and [3]), the proof of the next result is somewhat different in flavor. Indeed, in addition to manipulating certain equations, the proof of Theorem 2.9 will also use actual results of hyperbolic geometry. Note that Theorem 2.9 establishes the result that was suggested by the example presented in Remark 2.5 (b).

Theorem 2.9. HCL ⇒ HOL. More precisely, if $\mathcal{L} = (a, b, c, \alpha, \beta, \gamma)$ is a hyperbolic list that satisfies HCL, then $\mathcal{L}$ satisfies HOL.

Proof. By hypothesis,

$$\cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}, \cos(\beta) = \frac{\cosh(c)\cosh(a) - \cosh(b)}{\sinh(c)\sinh(a)}$$
$$\text{and} \quad \cos(\gamma) = \frac{\cosh(a)\cosh(b) - \cosh(c)}{\sinh(a)\sinh(b)}.$$

Our task is to prove that

$$\cosh(a) = \frac{\cos(\beta)\cos(\gamma) + \cos(\alpha)}{\sin(\beta)\sin(\gamma)}, \cosh(b) = \frac{\cos(\gamma)\cos(\alpha) + \cos(\beta)}{\sin(\gamma)\sin(\alpha)}$$
$$\text{and } \cosh(c) = \frac{\cos(\alpha)\cos(\beta) + \cos(\gamma)}{\sin(\alpha)\sin(\beta)}.$$

Without loss of generality $\alpha \leq \beta \leq \gamma$. (These inequalities were part of the riding hypotheses for our second proof of the realization theorem for hyperbolic geometry in [8]. We will proceed to apply part (a) of that result and its proof.) By [8, Theorem 3.6 (a)], there exist uniquely determined positive real numbers b^* and c^* such that

$$\cosh(b^*) = \frac{\cos(\gamma)\cos(\alpha) + \cos(\beta)}{\sin(\gamma)\sin(\alpha)} \text{ and } \cosh(c^*) = \frac{\cos(\alpha)\cos(\beta) + \cos(\gamma)}{\sin(\alpha)\sin(\beta)}.$$

Also by [8, Theorem 3.6 (a)], there exists a hyperbolic triangle $\tilde{\Delta} = \Delta ABC$ such that the radian measure of the interior angle of $\tilde{\Delta}$ at vertex A (resp., at vertex B; resp., at vertex C) is α (resp., β; resp., γ) and the hyperbolic length of the side of $\tilde{\Delta}$ that is opposite the vertex B is b^*. Therefore, it follows by applying the “other" hyperbolic trigonometric law to $\tilde{\Delta}$ that

$$\cosh(b^*) = \frac{\cos(\gamma)\cos(\alpha) + \cos(\beta)}{\sin(\gamma)\sin(\alpha)}.$$

Hence, $\cosh(b^*) = \cosh(b)$. Thus, since $\cosh|_{[0,\infty)}$ is a bijection $[0,\infty) \to [1,\infty)$, we get $b^* = b$. Also, by the first paragraph of the proof of [8, Theorem 3.6 (a)], the hyperbolic length of the side of $\tilde{\Delta}$ that is opposite the vertex C is c^*. Therefore, we can repeat the above reasoning to get

$$\cosh(c^*) = \frac{\cos(\alpha)\cos(\beta) + \cos(\gamma)}{\sin(\alpha)\sin(\beta)} \ (= \cosh(c)),$$

whence $c^* = c$.

Let a^* be the hyperbolic length of the side of $\tilde{\Delta}$ that is opposite the vertex A. We claim that $a^* = a$. Recall that $b^* = b$ and $c^* = c$. Therefore, by applying the Hyperbolic Cosine Law to $\tilde{\Delta}$, we get

$$\cos(\alpha) = \frac{\cosh(b)\cosh(c) - \cosh(a^*)}{\sinh(b)\sinh(c)}, \text{ and so}$$

$$\frac{\cosh(b)\cosh(c) - \cosh(a^*)}{\sinh(b)\sinh(c)} = \frac{\cosh(b)\cosh(c) - \cosh(a)}{\sinh(b)\sinh(c)}.$$

It follows easily that $\cosh(a^*) = \cosh(a)$. Once again invoking the fact that $\cosh|_{[0,\infty)}$ is an injection, we get that $a^* = a$. This proves the above claim.

We have now proved that the angular measure "parts" of $\tilde{\Delta}$ are α, β and γ and that the corresponding hyperbolic lengths of sides "parts" of $\tilde{\Delta}$ are a, b and c. Therefore, by applying the Hyperbolic Law of Cosines to $\tilde{\Delta}$, we obtain the desired equations. The proof is complete. □

The Introduction made some comments about measuring angles. We devote this paragraph to a few additional comments in that regard. While most pupils first learn to measure (Euclidean) angles in degrees, we typically learn in high school to convert such degree measures to radian measures, if only to simplify the statements of various formulas. Thus, as one would expect, we are here treating models of Euclidean plane geometry (such as the view of $\mathbb{R}^2$ that is used for the purposes of studying calculus for real-valued functions of a single real variable) in which one complete revolution encompasses 2π radians (the point being that $360° = 2\pi$ radians). In publications at this level, *ça va sans dire*. In like manner, we are here considering models of plane hyperbolic geometry (such as the upper half-plane model) in which angles are measured in radians. This means simply that for these models, the constant k in [10, Theorem 10.1 and 10.2] can be taken to be $k = 1$. (For an explanation of the choice $k = 1$ that is given in terms of the "horocycle" concept from hyperbolic geometry, see the beginning of the paragraph on pages 332-333 of [10].) Thus, for us here, the sum of the radian measures of the interior angles of any hyperbolic triangle is less than π, and any positive real number less than π can be realized as such a sum for some triangle in each of the models under consideration.

We next infer a couple of equational/algebraic characterizations of hyperbolic triangles (in models where $k = 1$). Corollary 2.10, when considered in conjunction with the comment which follows it, illustrates the most important way in which the Hyperbolic Law of Cosines is more powerful than the Hyperbolic Law of Sines. Corollary 2.10 also records the possibly surprising fact that the Hyperbolic Law of Cosines and the "other" hyperbolic trigonometric law (the HOL) are equally powerful; that is, they are equivalent.

Corollary 2.10. *Let $\mathcal{L} = (a, b, c, \alpha, \beta, \gamma)$ be a hyperbolic list. Then the following conditions are equivalent:*

(1) *There exists a hyperbolic triangle whose interior angles have respective radian measures α, β and γ and whose corresponding sides have hyperbolic lengths a, b and c;*

(2) *$\mathcal{L}$ satisfies* HCL;

(3) *$\mathcal{L}$ satisfies* HOL.

Moreover, if the above three (equivalent) conditions hold, then the hyperbolic triangle in (1) is uniquely determined up to congruence.

Proof. (1) $\Rightarrow$ (3): Apply the "other" hyperbolic trigonometric law to the hyperbolic triangle posited in (1).

(3) $\Rightarrow$ (2): Apply Theorem 2.8.

(2) $\Rightarrow$ (1): This implication was established and noted in the proof of Theorem 2.9.

We have now proved that (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3). Finally, in view of the final paragraph of the proof of Theorem 2.9, the "Moreover" assertion follows from any of the classical congruence criteria for triangles of hyperbolic geometry (such as SAS or AAA). The proof is complete. □

The data presented in Examples 2.3 and 2.4 show that one *cannot* add "$\mathcal{L}$ satisfies HSL" as another equivalent condition in Corollary 2.10.

Remark 2.11. (a) Recall that, as I was unaware of the existence of [2], I published a(nother) proof that the Cosine Law and the Sine Law (of Euclidean geometry) are equivalent in [3]. The reader may be interested to compare the proofs in [2] with the proofs of the corresponding results in [3]. One feature in [3] that had not been considered in [2] was the identification of the triangles (in Euclidean geometry) that satisfy what may be called the "sine version of the Cosine Law", that is, the identification of the Euclidean triangles whose "parts" (a, b, c, α, β and γ) satisfy the equations obtained from the usual statement of the Cosine Law by replacing each occurrence of "cos" with "sin"; and (in a similar sense) the identification of the Euclidean triangles that satisfy what may be called the "cosine version of the Sine Law". These two curiosities can be handled easily. A decade after I wrote [3], John C. Peterson and I saw fit to include them in our subsequent Precalculus textbook: see [9, Exercise 22, page 401] and [9, Exercises 31 and 32, page 396]. Based on reports that reached Peterson and me, those results/exercises were well received by students in Precalculus classrooms that adopted [9], and so I commend them to the attention of any reader who is also a Precalculus instructor. In that spirit, I wish to raise the question(s) of identifying the hyperbolic triangles whose "parts" satisfy the variants of the usual statements of the HCL (resp., of the HSL; resp., of the HOL) in which the occurrences of "cosh" and "sinh" are interchanged (and one may also wish to consider the effect of interchanging the occurrences of "cos" and "sin" in those classical trigonometry laws of hyperbolic geometry).

(b) In the seventh and eighth paragraphs of the Introduction, we noted that spherical geometry may provide another way to prove the main results of this paper. Spherical geometry has been studied for many centuries, it has numerous connections to other classical geometries, and its principal results have been put into what seems to be final form at least 150 years ago (although newer, vectorial proofs of at least some of them did become available afterwards). Unfortunately, apart from its most elementary aspects, spherical geometry has (to the best of my knowledge) not been part of the standard mathematical curriculum (in North America) since at least 1960. Accordingly, it did not seem appropriate to use the main results of spherical geometry in presenting any of the equational or algebraic argumentation in this paper. Nevertheless, it has long been recognized that there are strong connections (some geometric, some algebraic) among the various trigonometric laws of hyperbolic geometry. For instance, Greenberg [10, page 335] states "All the geometry of a [hyperbolic] right triangle is incorporated in formulas (10) [in Theorem 10.3 on page 334], for the other formulas follow from (10) by pure algebra and identities." Of course, Greenberg's purpose at that point in his book was to prove the various laws of hyperbolic trigonometry, whereas our purpose in this paper was different, namely, to determine all the implications that exist among those various laws. It makes sense to ask if we could eliminate at least some of the geometry that we have used in the course of our algebraic proofs here. (Our only proof that used geometry was that of Theorem 2.9.) We will address that question two and three paragraphs hence, after deeper scrutiny of the above quotation from [10] and related matters.

As indicated by the proofs in this paper, I would prefer to discuss the three main results of hyperbolic trigonometry by considering only the behavior of the four functions which explicitly appear in the statements of those laws, namely, sin, cos, sinh and cosh. Note that Greenberg's explanation of how "the other formulas [for a hyperbolic right triangle] follow from (10) by pure algebra and identities" also involves the behavior of the function tanh. To be fair, the function tanh was already involved in formula (10). Indeed, these "formulas (10)" express the sine and the cosine of a non-right angle of a hyperbolic right triangle Δ in terms of the hyperbolic sine and the hyperbolic tangent of the (hyperbolic) lengths of the sides of Δ. Moreover, the hyperbolic tangent function is usually featured in textbook proofs of the main laws of hyperbolic trigonometry (see, for instance, [14, pages 99-100 and 104-108]; those pages also feature several uses of some geometrically motivated formulas for measuring angles in the upper half-plane model [14, Proposition 6.1.1] that were discussed in the Introduction). Note also that Greenberg's hint in [10, page 337] as to how to prove the main laws of hyperbolic trigonometry for an arbitrary hyperbolic triangle does explicitly suggest some use of identities, specifically, the expansion formulas for the hyperbolic cosine of a sum or difference. That latter hint also involves more geometry, namely, analyzing the two right triangles formed by "dropping an altitude". (Presumably, this analysis is to be done via [10, Theorem 10.3] and what Greenberg's hint calls "some algebra and identities".)

Let us recognize that a considerable amount of geometric background was used in Greenberg's proof of (10). (This observation is not meant as a negative criticism; after all, the assertions in (10) *are* results in hyperbolic *geometry*!) See, specifically, the appeals and/or references to the geometric concepts of inversion in a circle, inverse points and Poincaré lines in [10, pages 335-336]. Our proofs have not needed such explicit appeals to geometric background. However, our proof of Theorem 2.9 *did* use details from our proof(s) in [8] of the realization theorem of hyperbolic geometry. Thus, it would be fair to say that our proof of Theorem 2.9 did expect the reader to know the fact that AAA is a congruence criterion in hyperbolic geometry. That is an important fact, as it is one of the ways that the two kinds of neutral geometries can be distinguished from one another (since, in Euclidean geometry, AAA is a criterion for similarity but not a criterion for congruence). We conclude that the sole geometric aspect of our (mostly) "equational/algebraic" proofs here has occurred in the proof of Theorem 2.9, via ideas centered about the roles of the AAA criterion in hyperbolic geometry (and, perhaps one should add, via use of the upper half-plane model in references cited there). We have not found a purely equational/algebraic proof of Theorem 2.9 that would be motivated and whose details could be checked by hand without undue tedium. So, we raise the question of finding such a proof of Theorem 2.9 in which absolutely no geometric background is explicitly used or assumed.

The fact that the upper half-plane model is a model of hyperbolic geometry was surely used in some of the references that were cited in our proof of Theorem 2.9. Therefore, if one seeks to eliminate any esoteric geometric background in giving another proof of Theorem 2.9, it is important to survey the various verifications that the half-plane model *is* a model of hyperbolic geometry. I say "esoteric" being mindful of the above-mentioned appeals to certain rigid motions, inversions, etc.; in carrying out such a verification, one could not hope to ignore Cartesian equations of lines or circles, because of the nature of the ("straight" or "bowed") geodesics in that model. Nearly all of the verification that the upper half-plane gives a suitable model involves only such elementary considerations from Euclidean geometry: see [5] and [6], in which I modified portions of [14] (or of its first edition) to show how to use elementary Euclidean geometry and technology to teach a more modern course on hyperbolic geometry based on the upper-half plane model. I say "nearly all" because my two above-cited papers did not try to prove that the SAS axiom is satisfied by the upper-half-plane model. (All the other axioms were verified there, along with formulas for hyperbolic length that are easy to implement using the numerical definite integrating function of a graphing calculator.) However, the verification of the SAS axiom for the upper half-plane model in [12] involved the above-mentioned appeals to certain rigid motions, inversions, etc. I have concluded that [14]'s treatment of such matters is somewhat more cursory, but in the same spirit as [12]. As we noted above, Greenberg's proof of his formula (10) also depended on the properties of such geometric transformations. So, it would seem that new ideas will be needed in order to answer the question of finding a motivated proof of Theorem 2.9 that assumes/uses only elementary Euclidean geometry and is reasonable in its computational demands of a reader. In other words, I do not consider this to be an easy question.

(c) Next, we move seemingly far afield to consider the proof in 1976 by K. I. Appel and W. Haken of the Four Color Conjecture and the proof in 2006 by T. C. Hales and S. P. Ferguson of the Kepler Conjecture. Notably, those celebrated proofs of long-standing geometric conjectures were each computer-assisted. In fact, the Hales-Ferguson result was announced in 1998, its proof was published in a series of papers in 2006 but without complete verification by a team of referees, various subsequent publications addressed minor errata in the published proof, and a 29-page paper, "Formal proof of the Kepler Conjecture," was published by a team of 22 authors (Hales *et al*) in 2017 as the "the official published account of the now completed Flyspeck project." It is no exaggeration to say that the reliance on computers and proof assistants in this "formal proof" product of the Flyspeck project seems, at first glance, much deeper than, and different in spirit from, the equation-manipulating approaches in [2], [3] and much of the present paper.

The three-dimensional success (by Hales *et al*) on "cannonball packing" brings to mind several successes during the past few decades in which circle packing and hyperbolic geometry were used to discretize some results of classical complex analysis. (A similar combination of techniques has served to analyze some of the art of M. C. Escher.) It is natural to wonder whether a use of some higher-dimensional generalizations of circle packing and hyperbolic geometry could be developed and then used to reformulate at least some of the "formal proof" aspects of the work of the Flyspeck project as being mere manipulation of equations whose verification would be less reliant on technology. From that perspective, we would raise the following (possibly naïve) question. Can one generalize Corollary 2.10 to obtain equational/algebraic characterizations of various hyperbolic n-gons for $n \geq 4$?

As recalled in the Introduction, part of [3] may be viewed as having used the classical Sine Law and the classical Cosine Law to obtain equational/algebraic characterizations of Euclidean triangles. That point of view motivated us to aim for Corollary 2.10 as one of the goals of the present paper. Toward the end of [3], we raised the question of whether one could obtain generalizations of the above-mentioned equational/algebraic characterizations of Euclidean triangles.(presumably to Euclidean n-gons for $n \geq 4$ or in higher-dimensional Euclidean spaces). To date, no such equational/algebraic characterizations of those Euclidean matters have been brought to our attention.

(d) Besides the motivation for Corollary 2.10 that was attributed to [3] in (c), we close with a summary of most of a paper, [4], to which kindred motivation can also be attributed. Indeed, [4] was our second contribution (the first such being [3]) to the theme of finding equational characterizations of triangles in Euclidean geometry. As one might well expect, the analogue of "Euclidean list" and "hyperbolic list" which is discussed in the next paragraph will be called a "Mollweide list". All triangles considered in the next paragraph will be triangles in Euclidean geometry.

If a triangle Δ (in Euclidean geometry) has sides of lengths a, b, and c, as well as corresponding interior angles with radian measures α, β and γ, then it is known that the following three equations hold:

$$\frac{\cos(\frac{\alpha-\beta}{2})}{\sin(\frac{\gamma}{2})} = \frac{a+b}{c}, \quad \frac{\cos(\frac{\alpha-\gamma}{2})}{\sin(\frac{\beta}{2})} = \frac{a+c}{b}, \text{ and } \frac{\cos(\frac{\beta-\gamma}{2})}{\sin(\frac{\alpha}{2})} = \frac{b+c}{a}.$$

Isaac Newton published a variant of the third of these equations in his *Arithmetica Universalis*. As it is clear, from the symmetry of the notation, that any one of these equations implies the other two equations, one might well expect that these equations would be called the "Newton equations of Δ". However, possibly because some other (probably more important) things are known as "Newton equations" (or "Newton identities"), the just-displayed equations are called the *associated Mollweide* *equations of* Δ. (Mollweide *did* publish this trio of equations in 1808, but others had done so earlier.) It will be convenient to refer to the 6-tuple $\mathcal{M} := (a, b, c, \alpha, \beta, \gamma)$ featuring the above data as the *Mollweide list induced by* Δ, and also to say that Δ *realizes* $\mathcal{M}$. More generally, by analogy with the above terms "Euclidean list" and "hyperbolic list", we will say that a list $\mathcal{M} := (a, b, c, \alpha, \beta, \gamma)$ whose entries are all positive real numbers is a *Mollweide list* if $\alpha + \beta + \gamma = \pi$. In deference to Newton's choice of mentioning (a variant of) only one of the Mollweide equations induced by a given triangle, we began [4] by asking whether a Mollweide list whose entries satisfy at least one of the above-displayed equations must, in fact, be the Mollweide list induced by some triangle. As one would expect, the answer was in the negative: see [4, Example 1]. On a positive note, we took the first step toward a "realization" result by establishing the following fragment of a possible trio of triangle inequalities in [4, Proposition 2]: if a Mollweide list satisfies the first of the above displayed equations, then $a + b > c$. (As one might expect, the proof of this result used a standard trigonometric identity, in this case, the cofunction identity involving sine and cosine.) It would seem clear from [4, Proposition 2] that a more reasonable/natural question is the following: if $\mathcal{M}$ is a Mollweide list whose entries satisfy all three of the above-displayed equations, must $\mathcal{M}$ be the Mollweide list induced by some triangle? Fortunately, we answered this question in the affirmative. Moreover, by extensive use of some trigonometric identities (involving sine and cosine) and some elementary algebra (at the level of solving a "small" linear system of equations for "unknowns" in $\mathbb{R}$), we gave the following stronger result in [4, Theorem 7]: if $\mathcal{M}$ is a Mollweide list whose entries satisfy at least two of the above-displayed equations, then those entries satisfy all three of the above-displayed equations and there exists a triangle Δ (which is necessarily unique up to congruence) such that Δ realizes $\mathcal{M}$.

We view [4, Theorem 7] as a converse of the above-mentioned observation of Isaac Newton, the upshot being an (another) algebraic/equational characterization of triangles in Euclidean geometry. It should now be clear how [4] motivated our interest in obtaining Corollary 2.10 and thus why, along with [3], [4] deserves much of the credit/blame for motivating this entire paper. In view of the role of the hyperbolic tangent function (tanh) in the development of hyperbolic trigonometry in [10], we wish to raise the following question, which essentially asks if there exists an analogue of [4, Theorem 7], possibly involving tanh, in hyperbolic geometry. Does there exist an essentially new equation (*) that can be built using only hyperbolic trigonometric functions and possibly trigonometric functions and has the following two properties: (*) nontrivially involves six real variables; and, for any hyperbolic list $\mathcal{L} = (a, b, c, \alpha, \beta, \gamma)$, the entries of $\mathcal{L}$ satisfy (*) if and only if there exists a hyperbolic triangle with side lengths a, b and c and corresponding interior angles with radian measures α, β and γ? Given that the proofs in [4] involved only elementary algebra, calculus and some classic identities of (Euclidean, also known as circular) trigonometry, answering this question may become a mere exercise in using the table in [10, page 332] (which matched up a long list of identities in circular trigonometry with a list of identities in hyperbolic trigonometry) to "translate" the proofs in [4] (from, as it were, the circular language to the hyperbolic language). Perhaps, a consequence of such work would develop what could be called the "hyperbolic Mollweide equations" of any given triangle in hyperbolic geometry. In any case, although we have not thought deeply about this question and its possible offshoots, we would hope that some reader(s) will have the time and interest to do so.

References

[1] K. Borsuk and W. Szmielew, Foundations of geometry, North Holland, Amsterdam, 1960.

[2] L. J. Burton, The laws of sines and cosines, Amer. Math. Monthly, 56 (8) (1949), 550–551.

[3] D. E. Dobbs, The sine law and the cosine law are equivalent, Computer and Math. Educ. J., 18 (1984), 42–45.

[4] D. E. Dobbs, Two Mollweide equations detect triangles, The AMATYC Review, 15 (1) (1993), 13–19.

[5] D. E. Dobbs, Applications of a distance formula in hyperbolic plane geometry, Inter. J. Math. Educ. Sci. Technol., 32 (1) (2001), 148–156.

[6] D. E. Dobbs, Verification of the hyperbolic parallel postulate in a half-plane model, Math. and Computer Educ. J., 35 (2) (2001), 140–146.

[7] D. E. Dobbs, Constructing an angle with prescribed measure and vertex in the upper half-plane model of hyperbolic plane geometry, Moroccan J. Algebra, Geom. Appl., 1 (1) (2022), 18–48.

[8] D. E. Dobbs, On a realization theorem in plane hyperbolic geometry and a related identity in neutral geometry, Moroccan J. Algebra Geom. Appl., 1 (2) (2022), 361–379.

[9] D. E. Dobbs and J. C. Peterson, Precalculus, Wm. C. Brown, Dubuque, 1993.

[10] M. Greenberg, Euclidean and non-Euclidean geometries: development and history, second edition, W. H. Freeman, San Francisco, 1974.

[11] R. S. Millman, The upper half plane model for hyperbolic geometry, Amer. Math. Monthly, 87 (1) (1980), 48–53.

[12] R. S. Millman and G. D. Parker, Geometry: a metric approach with models, second edition, Undergraduate Texts in Math., Springer-Verlag, New York, 1991.

[13] S. Stahl, The Poincaré half-plane: a gateway to modern geometry, Jones and Bartlett, Boston, 1993.

[14] S. Stahl, A gateway to modern geometry: the Poincaré half-plane, second edition, Jones and Bartlett, Boston, 2008.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco

Volume 2, Issue 1 (2023), pp 143-153

ISSN: 2820-7114

Title :

On Φ-(n,J)-ideals and n-J-ideals of commutative rings

Author(s):

Abdelhaq El Khalfi

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 143–153
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
ISSN: 2820-7114

On ϕ-(n, J)-ideals and n-J-ideals of commutative rings

Abdelhaq El Khalfi
Fundamental and Applied Mathematics Laboratory
Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Morocco.
e-mail: *abdelhaqelkhalfi@gmail.com*

Communicated by Moutu Abdou Salam Moutui
(Received 10 December 2022, Revised 20 March 2023, Accepted 04 April 2023)

Abstract. Let R be a commutative ring with nonzero identity. In this paper, we introduce and investigate a generalization of $(2, J)$-ideals. Let $\phi : \mathcal{I}(R) \to \mathcal{I}(R) \cup \{\emptyset\}$ be a function where $\mathcal{I}(R)$ is the set of ideals of R. A proper ideal I of R is said to be a ϕ-(n, J)-ideal if whenever $x_1 \cdots x_{n+1} \in I \setminus \phi(I)$, for $x_1, \dots, x_{n+1} \in R$, then $x_1 \cdots x_n \in I$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in Jac(R)$, for some $k \in \{1, \dots, n\}$. Also, I is called an n-J-ideal if whenever $x_1 \cdots x_{n+1} \in I$, for $x_1, \dots, x_{n+1} \in R$, then $x_1 \cdots x_n \in Jac(R)$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in I$, for some $k \in \{1, \dots, n\}$. Moreover, we give some basic properties of those classes of ideals and we study the ϕ-(n, J)-ideals and the n-J-ideals of the localization of rings, the direct product of rings, the trivial ring extensions and the amalgamation of rings.

Key Words: ϕ-(n, J)-ideal, ϕ-n-absorbing primary ideal, ϕ-n-absorbing ideal, ϕ-prime ideal, n-J-ideal.
2010 MSC: Primary 13A15; Secondary 13C05.

Dedicated to the memory of Professor Muhammad Zafrullah

1 Introduction

Throughout this paper, all rings are assumed to be commutative with nonzero identity and all modules are nonzero unital. If R is a ring, then $\sqrt{I}$ denotes the radical of an ideal I of R, in the sense of [22, page 17]. We denote the set of all ideals (resp. proper ideals) of a ring R by $\mathcal{I}(R)$ (resp. $\mathcal{I}^*(R)$) and $Jac(R)$ denotes the Jacobson radical of R.
Anderson and Smith [3], defined a weakly prime ideal as a proper ideal P of R with the property that for $a, b \in R$, $0 \neq ab \in P$ implies $a \in P$ or $b \in P$. Later, the authors of [8] defined the notion of almost prime ideal, i.e., an ideal $P \in \mathcal{I}^*(R)$ with the property that if $a, b \in R$, $ab \in P \setminus P^2$, then either $a \in P$ or $b \in P$. Thus a weakly prime ideal is almost prime and any proper idempotent ideal is also almost prime. Moreover, an ideal P of R is almost prime if and only if P/P^2 is a weakly prime ideal of R/P^2. Anderson and Bataineh in [2], extended these concepts to ϕ-prime ideals. Let $\phi : \mathcal{I}(R) \to \mathcal{I}(R) \cup \{\emptyset\}$ be a function. A proper ideal P of R is called ϕ-prime if for $x, y \in R$, $xy \in P \setminus \phi(P)$ implies $x \in P$ or $y \in P$. In fact, P is a ϕ-prime ideal of R if and only if $P/\phi(P)$ is a weakly prime ideal of $R/\phi(P)$. In 2017, J. Bagheri Harehdashti and H. Fazaeli Moghimi defined the ϕ-radical of an ideal I as the intersection of all ϕ-prime ideals of R containing I and investigated when the set of all ϕ-prime ideals of R has a Zariski topology analogous to that of the prime spectrum. Since $P \setminus \phi(P) = P \setminus (P \cap \phi(P))$, there is no loss of generality in assuming that $\phi(P) \subseteq P$. In [6], Badawi introduced the notion of 2-absorbing ideal. A nonzero proper ideal I of R is called a 2-absorbing ideal of R if $abc \in I$ for some elements $a, b, c \in R$, then $ab \in I$ or $ac \in I$ or $bc \in I$. Anderson and Badawi generalized in [1] the concept of 2-absorbing ideals to n-absorbing ideals. They defined a proper ideal I of R to be an n-absorbing ideal if whenever $x_1 \cdots x_{n+1} \in I$ for some $x_1, \dots, x_{n+1} \in R$, then there are n of the x_i's whose product is in I. In [16], Ebrahimpour and Nekooe studied the notion of ϕ-n-absorbing ideals as $(n, n+1)$-ϕ-ideals. On the other hand, Badawi, Tekir and Yetkin [7] introduced the class of 2-absorbing primary ideals.

A proper ideal I of R is said to be a 2-absorbing primary ideal if whenever $a,b,c \in R$ and $abc \in I$, then $ab \in I$ or $ac \in \sqrt{I}$ or $bc \in \sqrt{I}$. In [24], Mostafanasab and Darani generalized the concept of 2-absorbing primary ideals to ϕ-n-absorbing primary ideals. In [5], Anebri et al. introduced the notion of ϕ-(n,N)-ideal. A proper ideal of R is said to be a ϕ-(n,N)-ideal if whenever $x_1 \cdots x_{n+1} \in I \setminus \phi(I)$, for $x_1,\ldots,x_{n+1} \in R$, then $x_1 \cdots x_n \in I$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in \sqrt{0}$, for some $k \in \{1,\ldots,n\}$, where $\sqrt{0}$ denotes the nilradical of R.
In [23], H. A. Khashan and B. B. Bani-Ata introduced the concept of J-ideals. A proper ideal I of R is said to be a J-ideal if for each $a,b \in R$, $ab \in I$ implies that either $a \in Jac(A)$ or $b \in I$. Recently, in [25], Yildiz, Tekir and Koç generalized the class of J-ideals to $(2,J)$-ideals. A proper ideal I of R is called a $(2,J)$-ideal if whenever $a,b,c \in R$ and $abc \in I$, then $ab \in I$ or $ac \in Jac(R)$ or $bc \in Jac(R)$. The purpose of this paper is to introduce and investigate a generalization of $(2,J)$-ideals. Let $\phi : \mathcal{I}(R) \to \mathcal{I}(R) \cup \{\emptyset\}$ be a function where $\mathcal{I}(R)$ is the set of ideals of a ring R. A proper ideal I of R is said to be a ϕ-(n,J)-ideal if $x_1 \cdots x_{n+1} \in I \setminus \phi(I)$ for some $x_1,\ldots,x_{n+1} \in R$, then either $x_1 \cdots x_n \in I$ or the product of x_{n+1} with $(n-1)$ of $x_1,\ldots,x_n$ is in $Jac(R)$. Since $I \setminus \phi(I) = I \setminus (I \cap \phi(I))$, there is no loss of generality in assuming that $\phi(I) \subseteq I$. We henceforth make the assumption that given two functions $\psi_1,\psi_2 : \mathrm{I}(R) \to \mathrm{I}(R) \cup \{\emptyset\}$, then $\psi_1 \leq \psi_2$ if $\psi_1(I) \subseteq \psi_2(I)$, for each $I \in \mathrm{I}(R)$. We can define the map $\phi_\alpha : \mathrm{I}(R) \to \mathrm{I}(R) \cup \{\emptyset\}$ as follows: Let I be a ϕ-(n,J)-ideal of R. Then

(1) $\phi_\emptyset(I) = \emptyset \Rightarrow I$ is an (n,J)-ideal.

(2) $\phi_0(I) = 0 \Rightarrow I$ is a weakly (n,J)-ideal.

(3) $\phi_2(I) = I^2 \Rightarrow I$ is an almost (n,J)-ideal.

(4) $\phi_m(I) = I^m$ $(m \geq 2) \Rightarrow I$ is an m-almost (n,J)-ideal.

(5) $\phi_\omega(I) = \cap_{m=1}^{\infty} I^m \Rightarrow I$ is an ω-(n,J)-ideal.

(6) $\phi_1(I) = I \Rightarrow I$ is any ideal.

Observe that $\phi_\emptyset \leq \phi_0 \leq \phi_\omega \leq \cdots \leq \phi_{m+1} \leq \phi_m \leq \cdots \leq \phi_2 \leq \phi_1$.
Also, a proper ideal I is called an n-J-ideal if whenever $x_1 \cdots x_{n+1} \in I$, for $x_1,\ldots,x_{n+1} \in R$, then $x_1 \cdots x_n \in Jac(R)$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in I$, for some $k \in \{1,\ldots,n\}$. So, we give some basic properties of those classes of ideals and we study the ϕ-(n,J)-ideals and the n-J-ideals of the localization of rings, the direct product of rings, the trivial ring extensions and amalgamation of rings.

2 ϕ-(n,J)-ideals of commutative rings

Let n be a positive integer. Consider the elements $x_1,\ldots,x_n$ and ideals $I_1,\ldots,I_n$ of a ring R. Throughout this paper, we use the following notations:

- $x_1 \cdots \widehat{x_k} \cdots x_n$: k-th term is excluded from $x_1 \cdots x_n$.
- $I_1 \cdots \widehat{I_k} \cdots I_n$: k-th term is excluded from $I_1 \cdots I_n$.

Definition 2.1. Let R be a ring, $\phi : \mathrm{I}(R) \to \mathrm{I}(R) \cup \{\emptyset\}$ be a function. A proper ideal of R is said to be a ϕ-(n,J)-ideal if whenever $x_1 \cdots x_{n+1} \in I \setminus \phi(I)$, for $x_1,\ldots,x_{n+1} \in R$, then $x_1 \cdots x_n \in I$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in Jac(R)$, for some $k \in \{1,\ldots,n\}$.

The following remark follows from the defnition of ϕ-(n,J)-ideals.

Remark 2.2. Let R be a ring, I be a proper ideal of R and $\phi : \mathrm{I}(R) \to \mathrm{I}(R) \cup \{\emptyset\}$ be a function.

(1) I is a ϕ-J-ideal if and only if I is a ϕ-$(1,J)$-ideal.

(2) If $\psi_1,\ \psi_2 : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ two functions with $\psi_1 \leq \psi_2$. Then every ψ_1-(n,J)-ideal of R is a ψ_2-(n,J)-ideal.

(3) I is an (n,J)-ideal $\Rightarrow$ I is a weakly (n,J)-ideal $\Rightarrow$ I is an $\omega-(n,J)$-ideal $\Rightarrow$ I is an m-almost (n,J)-ideal $\Rightarrow$ I is an almost (n,J)-ideal.

(4) If $n \leq n'$ are two positive integers and I is a ϕ-(n,J)-ideal, then I is a ϕ-(n',J)-ideal.

Proposition 2.3. *Let R be a ring, I be a proper ideal of R and $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function.*

(1) *If I is a J-ideal, then is I is a ϕ-(n,J)-ideal.*

(2) *Assume that I is a ϕ-(n,J)-ideal, then $I \setminus \phi(I) \subseteq Jac(R)$.*

(3) *A ϕ-n-absorbing primary ideal I of R is a ϕ-(n,J)-ideal if and only if $I \setminus \phi(I) \subseteq Jac(R)$.*

Proof. (1) Assume that I is a J-ideal and let $x_1 \cdots x_{n+1} \in I \setminus \phi(I)$, for $x_1,\ldots,x_{n+1} \in R$. Suppose that $x_2 \cdots x_{n+1} \notin Jac(R)$. As I is a J-ideal, we get that $x_1 \in I$, and so $x_1 \cdots x_n \in I$. This implies that I is a ϕ-(n,J)-ideal.
(2) Let $x \in I \setminus \phi(I)$. The fact that $1 \cdots 1 \cdot x \in I \setminus \phi(I)$ and $1 \notin I$ imply that $x \in Jac(R)$ and so $I \setminus \phi(I) \subseteq Jac(R)$.
(3) By (2) it suffices to show the converse. Assume that $I \setminus \phi(I) \subseteq Jac(R)$ and $x_1 \cdots x_{n+1} \in I \setminus \phi(I)$ for some $x_1,\ldots,x_{n+1} \in R$. If $x_1 \cdots x_n \notin I$, we then have $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in \sqrt{I} \setminus \phi(I)$ for some $k \in \{1,\ldots,n\}$, which implies that $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in Jac(R)$. Thus I is a ϕ-(n,J)-ideal. □

Proposition 2.4. *Let R be a ring, $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function and $(I_\alpha)_{\alpha \in \Lambda}$ be a family of ϕ-(n,J)-ideals.*

(1) *Assume that ϕ reverses the inclusion, then $\cap_{\alpha \in \Lambda} I_\alpha$ is a ϕ-(n,J)-ideal.*

(2) *Assume that ϕ preserves the inclusion and $(I_\alpha)_{\alpha \in \Lambda}$ is a collection of ascending chain of ϕ-(n,J)-ideals, then $\cup_{\alpha \in \Lambda} I_\alpha$ is a ϕ-(n,J)-ideal.*

Proof. (1) Let $x_1,\ldots,x_{n+1} \in R$ such that $x_1 \cdots x_{n+1} \in \cap_{\alpha \in \Lambda} I_\alpha \setminus \phi(\cap_{\alpha \in \Lambda} I_\alpha)$. Assume that $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \notin Jac(R)$, for each $k \in \{1,\ldots,n\}$. By hypothesis, we must have $x_1 \cdots x_{n+1} \in I_\alpha \setminus \phi(I_\alpha)$, for each $\alpha \in \Lambda$. Moreover, the fact that I_α is a ϕ-(n,J)-ideal implies that $x_1 \cdots x_n \in I_\alpha$ and thus $x_1 \cdots x_n \in \cap_{\alpha \in \Lambda} I_\alpha$.
(2) Assume that $(I_\alpha)_{\alpha \in \Lambda}$ is a collection of ascending chain of ϕ-(n,J)-ideals. Let $x_1,\ldots,x_{n+1} \in R$ satisfying $x_1 \cdots x_{n+1} \in \cup_{\alpha \in \Lambda} I_\alpha \setminus \phi(\cup_{\alpha \in \Lambda} I_\alpha)$ and $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \notin Jac(R)$ for each $k \in \{1,\ldots,n\}$. By hypothesis, there exists $\alpha \in \Lambda$ such that $x_1 \cdots x_{n+1} \in I_\alpha \setminus \phi(I_\alpha)$. Hence $x_1 \cdots x_n \in I_\alpha$ and thus $x_1 \cdots x_n \in \cup_{\alpha \in \Lambda} I_\alpha$, as desired. □

Recall that a ϕ-prime ideal P of R is divided if $P \subseteq (x)$ for every $x \in R \setminus P$.

Theorem 2.5. Let R be a ring, P be a divided ϕ-prime ideal, I be a ϕ-(n,J)-ideal of R with $Jac(R) = P$, $I \subseteq P$ and $\phi(P) \subseteq \phi(I)$. Then I is a ϕ-J-ideal.

Proof. Take $xy \in I \setminus \phi(I)$ for $x,y \in R$ such that $y \notin P$. Since $xy \in P \setminus \phi(P)$ and P is a ϕ-prime ideal, we get $x \in P$. As $y^{n-1} \notin P$, we get that $P \subseteq (y^{n-1})$. Hence there exists $z \in R$ such that $x = y^{n-1}z$ and so $y^n z = y^{n-1}zy = xy \in I \setminus \phi(I)$. Since $y \notin Jac(R)$ and I is a ϕ-(n,J)-ideal, we get $x = y^{n-1}z \in I$. Therefore, I is a ϕ-J-ideal. □

Let I be an ideal of a ring R and $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function. Assume that I is a ϕ-(n,J)-ideal of R and $a_1,\ldots,a_{n+1} \in R$. We say that $(a_1,\ldots,a_{n+1})$ is a ϕ-$(n+1,J)$-tuple of I if $a_1 \cdots a_{n+1} \in \phi(I)$, $a_1 a_2 \cdots a_n \notin I$ and for each $1 \leq k \leq n$, $a_1 \cdots \widehat{a_k} \cdots a_{n+1} \notin Jac(R)$.

Theorem 2.6. Let I be a ϕ-(n,J)-ideal of a ring R and suppose that $(a_1,\ldots,a_{n+1})$ is a ϕ-$(n+1,J)$-tuple of I for some $a_1,\ldots,a_{n+1}\in R$. Then for each element $\alpha\in\{1,2,\ldots,n+1\}$, we have $a_1\cdots\widehat{a_\alpha}\cdots a_{n+1}I\subseteq\phi(I)$.

Proof. Suppose that $a_1\cdots\widehat{a_\alpha}\cdots a_{n+1}I\not\subseteq\phi(I)$. So, there exists $x\in I$ such that $a_1\cdots\widehat{a_\alpha}\cdots a_{n+1}x\notin\phi(I)$. Hence $a_1\cdots\widehat{a_\alpha}\cdots a_{n+1}(a_\alpha+x)\in I\setminus\phi(I)$. Notice that $x\in I\setminus\phi(I)$ and so $x\in Jac(R)$ by Proposition 2.3. If $\alpha<n+1$, we have that either $a_1\cdots\widehat{a_\alpha}\cdots a_n(a_\alpha+x)\in I$ or $a_1\cdots\widehat{a_\alpha}\cdots\widehat{(a_\alpha+x)}\cdots a_{n+1}\in Jac(R)$ or $a_1\cdots\widehat{a_\alpha}\cdots(a_\alpha+x)\cdots\widehat{a_\beta}\cdots a_{n+1}\in Jac(R)$ for some $1\le\beta\le n$ with $\beta\neq\alpha$, which any of these cases has a contradiction. If $\alpha=n+1$, then $a_1\cdots a_n(a_{n+1}+x)\in I\setminus\phi(I)$. Since $a_1\cdots a_n\notin I$, we must have $a_1\cdots\widehat{a_\beta}\cdots a_n(a_{n+1}+x)\in Jac(R)$ for some $1\le\beta\le n$ and so $a_1\cdots\widehat{a_\beta}\cdots a_{n+1}\in Jac(R)$, a desired contradiction. This gives that $a_1\cdots\widehat{a_\alpha}\cdots a_{n+1}I\subseteq\phi(I)$. □

Theorem 2.7. Let R be a ring and $\phi:\mathfrak{I}(R)\to\mathfrak{I}(R)\cup\{\emptyset\}$ be a function. Assume that I is a ϕ-(n,J)-ideal that is not an (n,J)-ideal. Then:

(1) $I^{n+1}\subseteq\phi(I)$.

(2) $\sqrt{\phi(I)}=\sqrt{I}$.

Proof. (1) Assume that $x_1\cdots x_{n+1}\notin\phi(I)$ for some $x_1,\ldots,x_{n+1}\in I$. The fact that I is a ϕ-(n,J)-ideal of R which is not an (n,J)-ideal implies that there exist $a_1,\ldots,a_{n+1}\in R$ such that $(a_1,\ldots,a_{n+1})$ is a ϕ-$(n+1,J)$-tuple of I. By Theorem 2.6, there exists $\xi\in\phi(I)$ such that $(x_1+a_1)\cdots(x_{n+1}+a_{n+1})=\xi+x_1\cdots x_{n+1}$ and hence $(x_1+a_1)\cdots(x_{n+1}+a_{n+1})\in I\setminus\phi(I)$. So, either $(x_1+a_1)\cdots(x_n+a_n)\in I$ or $(x_1+a_1)\cdots\widehat{(x_k+a_k)}\cdots(x_{n+1}+a_{n+1})\in Jac(R)$ for some $k\in\{1,\ldots,n\}$. Since $x_i\in I\setminus\phi(I)\subseteq Jac(R)$ for each i, we must have that either $a_1\cdots a_n\in I$ or $a_1\cdots\widehat{a_k}\cdots a_{n+1}\in Jac(R)$, which is a contradiction. Thus $I^{n+1}\subseteq\phi(I)$.
(2) It remains to prove that $\sqrt{I}\subseteq\sqrt{\phi(I)}$. This, in turn, follows easily from (1). This completes the proof. □

Theorem 2.8. Let R be a chained ring and I be an ideal of R such that I^{n+1} is not principal for some $n\in\mathbb{N}$ and $I\subseteq Jac(R)$. Then I is a ϕ_{n+1}-(n,J)-ideal if and only if I is an (n,J)-ideal.

Proof. Assume that I is a ϕ_{n+1}-(n,J)-ideal that is not an (n,J)-ideal. Then there exist $x_1,x_2,\ldots,x_{n+1}\in R$ such that $x_1x_2\cdots x_{n+1}\in I$ but $x_1x_2\cdots x_n\notin I$ and $x_1\cdots\widehat{x_k}\cdots x_{n+1}\notin Jac(R)$. Thus $(x_k)\not\subseteq I$ for every $1\le k\le n$. Since R is a chained ring, $I^{n+1}\subseteq(x_1x_2\cdots x_{n+1})$. But I^{n+1} is not principal and so, $x_1x_2\cdots x_{n+1}\notin I^{n+1}$. This means that $x_1x_2\cdots x_{n+1}\in I\setminus\phi_{n+1}(I)$. Since I is a ϕ_{n+1}-(n,J)-ideal, $x_1x_2\cdots x_n\in I$ or $x_1\cdots\widehat{x_k}\cdots x_{n+1}\in Jac(R)$ for some $1\le k\le n$, a contradiction. Hence I is an (n,J)-ideal. On the other hand, assume that I is an (n,J)-ideal. Then it is clear that I is a ϕ_{n+1}-(n,J)-ideal. □

Let R and R' be commutative rings with identity and $f:R\to R'$ a ring homomorphism. For an ideal I of R, the extension I^e of I is the ideal of R' generated by $f(I)$; and for an ideal H of R', the contraction of H is the ideal $H^c=\{r\in R\mid f(r)\in H\}$ of R. Let $\phi_{R'}:\mathfrak{I}(R')\to\mathfrak{I}(R')\cup\{\emptyset\}$ be a function. Define $\phi_R:\mathfrak{I}(R)\to\mathfrak{I}(R)\cup\{\emptyset\}$ by $\phi_R(I)=\phi_{R'}(I^e)^c$ and $\phi_R(I)=\emptyset$ if $\phi_{R'}(I^e)=\emptyset$.

Theorem 2.9. Let $f:R\to R'$ be a ring homomorphism such that $f^{-1}(Jac(R'))=Jac(R)$. If H is a $\phi_{R'}$-(n,J)-ideal of R' such that $\phi_{R'}(H)\subseteq\phi_{R'}(H^{ce})$, then H^c is a ϕ_R-(n,J)-ideal of R.

Proof. Let $x_1x_2\ldots x_{n+1}\in H^c\setminus\phi_R(H^c)$. Then $f(x_1x_2\ldots x_{n+1})=f(x_1)f(x_2)\ldots f(x_{n+1})\in H$. Suppose that $f(x_1)f(x_2)\ldots f(x_{n+1})\in\phi_{R'}(H)$. Then $x_1x_2\ldots x_{n+1}\in\phi_{R'}(H)^c\subseteq\phi_{R'}(H^{ce})^c=\phi_R(H^c)$ which is a contradiction. So $f(x_1)f(x_2)\cdots f(x_{n+1})\in H\setminus\phi_{R'}(H)$. Since H is a $\phi_{R'}$-(n,J)-ideal of R', we get either $f(x_1)f(x_2)\cdots f(x_n)\in H$ or $f(x_1)\cdots\widehat{f(x_k)}\cdots f(x_{n+1})\in Jac(R')$ for some $1\le k\le n$. Thus $x_1x_2\cdots x_n\in H^c$ or $x_1\cdots\widehat{x_k}\cdots x_{n+1}\in f^{-1}(Jac(R'))=Jac(R)$ for some $1\le k\le n$. Therefore, H^c is a ϕ_R-(n,J)-ideal of R, which completes the proof. □

Let H be an ideal of a ring R and $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function. Define $\phi_H : \mathbb{I}(R/H) \to \mathbb{I}(R/H) \cup \{\emptyset\}$ by $\phi_H(I/H) = (\phi(I)+H)/H$ for every $I \in \mathbb{I}(R)$ such that $H \subseteq I$ and $\phi_H(I) = \emptyset$ if $\phi(I) = \emptyset$.

Theorem 2.10. Let $H \subseteq I$ be two proper ideals of a ring R, and $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function.

(1) If I is a ϕ-(n,J)-ideal of R, then I/H is a ϕ_H-(n,J)-ideal of R/H.

(2) If $\phi(I) \subseteq H$ and I is a ϕ-(n,J)-ideal of R, then I/H is a weakly (n,J)-ideal of R/H.

(3) If $H \subseteq \phi(I) \cap Jac(R)$, and I/H is a ϕ_H-(n,J)-ideal of R/H, then I is a ϕ-(n,J)-ideal of R.

Proof. (1) Let $(x_1+H),\ldots,(x_{n+1}+H) \in R/H$ such that $(x_1+H)\cdots(x_{n+1}+H) \in I/H \setminus \phi_H(I/H) = I/H \setminus (\phi(I)+H)/H$. So $x_1\cdots x_{n+1} \in I \setminus \phi(I)$. Since I is a ϕ-(n,J)-ideal, we then have $x_1\cdots x_n \in I$ or $x_1\cdots \widehat{x_k} \cdots x_{n+1} \in Jac(R)$ for some $k \in \{1,\ldots,n\}$. Hence $(x_1+H)\cdots(x_n+H) \in I/H$ or $(x_1+H)\cdots\widehat{(x_k+H)}\cdots(x_{n+1}+H) \in Jac(R/H)$. This proves that I/H is a ϕ_H-(n,J)-ideal of R/H.
(2) This follows immediately from (1).
(3) Suppose that $x_1\cdots x_{n+1} \in I \setminus \phi(I)$ for some $x_1,\ldots,x_{n+1} \in R$. Then, $(x_1+H)\cdots(x_{n+1}+H) \in I/H \setminus \phi(I)/H = I/H \setminus \phi_H(I/H)$. By hypothesis, we get that either

$$(x_1+H)\cdots(x_n+H) \in I/H \text{ or } (x_1+H)\cdots\widehat{(x_k+H)}\cdots(x_{n+1}+H) \in Jac(R/H) = Jac(R)/H$$

for some $k \in \{1,\ldots,n\}$. Thus, $x_1\cdots x_n \in I$ or $x_1\cdots\widehat{x_k}\cdots x_{n+1} \in Jac(R)$ for some $k \in \{1,\ldots,n\}$, as desired. □

Let R be a ring and S be a multiplicatively closed subset of R. Let $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function. We define $\phi_S : \mathbb{I}(R_S) \to \mathbb{I}(R_S) \cup \{\emptyset\}$ by $\phi_S(J) = (\phi(J \cap R))_S$ and $\phi_S(J) = \emptyset$ if $\phi(J \cap R) = \emptyset$ for every ideal J of R_S. Note that $\phi_S(J) \subseteq J$.

Proposition 2.11. *Let R be a ring, $\phi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ be a function and S be a multiplicatively closed subset of R. If I is a ϕ-(n,J)-ideal of R with $(\phi(I))_S \subseteq \phi_S(I)$, then I_S is a ϕ_S-(n,J)-ideal of R_S.*

Proof. Let $\frac{x_1}{s_1},\ldots,\frac{x_{n+1}}{s_{n+1}} \in R_S$ such that $\frac{x_1}{s_1}\cdots\frac{x_{n+1}}{s_{n+1}} \in I_S \setminus \phi_S(I_S)$ and $\frac{x_1}{s_1}\cdots\frac{x_n}{s_n} \in I_S$. Then, there exists an element $s \in S$ such that $sa_1\cdots a_{n+1} \in I$. Since $(\phi(I))_S \subseteq \phi_S(I)$, we must have $x_1\cdots x_n(sx_{n+1}) \in I \setminus \phi(I)$. If $x_1\cdots x_n \in I$, then $\frac{x_1}{s_1}\cdots\frac{x_n}{s_n} \in I_S$, which is a contradiction. So, there exists $k \in \{1,\ldots,n\}$ such that $x_1\cdots\widehat{x_k}\cdots x_n(sx_{n+1}) \in Jac(R)$. Hence $\frac{x_1}{s_1}\cdots\widehat{\frac{x_k}{s_k}}\cdots\frac{x_{n+1}}{s_{n+1}} \in (Jac(R))_S \subseteq Jac(R_S)$, as desired. □

Let R be a ring and M be an R-module. Then $R \ltimes M$, is called the *trivial* (*ring*) *extension of* R *by* M. We recall that $R \ltimes M$ is the ring whose additive structure is that of the external direct sum $R \oplus M$ and whose multiplication is defined by $(a,e)(b,f) = (ab, af+be)$ for all $a,b \in R$ and all $e,f \in M$. (This construction is also known as the *idealization* $R(+)M$). The basic properties of trivial ring extensions are summarized in the books [19], [18]. Trivial ring extensions have been studied and generalized by many authors (for example, cf. [4, 14, 15, 20]). We recall that if I is an ideal of R and F is a submodule of M, then $I \ltimes F$ is an ideal of $R \ltimes M$ if and only if $IE \subseteq F$. In the next result, we study some ϕ-(n,J)-ideals of trivial ring extensions.

Theorem 2.12. Let R be a ring and M be an R-module. Let $\psi : \mathbb{I}(R) \to \mathbb{I}(R) \cup \{\emptyset\}$ and $\phi : \mathbb{I}(R \ltimes M) \to \mathbb{I}(R \ltimes M) \cup \{\emptyset\}$ be two functions such that $\phi(I \ltimes M) = \psi(I) \ltimes M$. Then $I \ltimes M$ is a ϕ-(n,J)-ideal of $R \ltimes M$ if and only if I is a ψ-(n,J)-ideal of R.

Proof. Assume that $x_1\cdots x_{n+1} \in I \setminus \psi(I)$ for some $x_1,\ldots,x_{n+1} \in R$ such that $x_1\cdots x_n \notin I$. Hence $(x_1,0)\cdots(x_{n+1},0) = (x_1\cdots x_{n+1},0) \in I \ltimes M \setminus \phi(I \ltimes M)$. Since $I \ltimes M$ is a ϕ-(n,J)-ideal of $R \ltimes M$ and $x_1\cdots x_n \notin I$, there is $k \in \{1,\ldots,n\}$ such that $(x_1,0)\cdots\widehat{(x_k,0)}\cdots(x_{n+1},0) \in Jac(R \ltimes M)$. Thus $x_1\cdots\widehat{x_k}\cdots x_{n+1} \in Jac(R)$. Conversely, if $(x_1,m_1)\cdots(x_{n+1},m_{n+1}) \in I \ltimes M \setminus \phi(I \ltimes M)$ then $x_1\cdots x_{n+1} \in I \setminus \psi(I)$. The fact that I is a ψ-(n,J)-ideal of R implies that either $x_1\cdots x_n \in I$ or $x_1\cdots\widehat{x_k}\cdots x_{n+1} \in Jac(R)$. Therefore, we have either $(x_1,m_1)\cdots(x_n,m_n) \in I \ltimes M$ or $(x_1,m_1)\cdots\widehat{(x_k,m_k)}\cdots(x_{n+1},m_{n+1}) \in Jac(R \ltimes M)$, as desired. □

Let A and B be two rings with unity, let J be an ideal of B and let $f : A \longrightarrow B$ be a ring homomorphism. In this setting, we consider the following subring of $A \times B$:

$$A \bowtie^f J := \{(a, f(a) + j) \in A \times B | a \in A, j \in J\}$$

is called the *amalgamation of A and B along J with respect to f*. This construction is a generalization of the amalgamated duplication of a ring along an ideal denoted $A \bowtie I$ (introduced and studied by D'Anna and Fontana in [12]). In [10, 11], D'Anna, Finocchiaro and Fontana introduced the more general context of amalgamations. They have studied these constructions in the frame of pullbacks which allowed them to establish numerous results on the transfer of various ideal and ring-theoretic properties from A and $f(A) + J$ to $A \bowtie^f J$. The concept of amalgamation is an important and an interesting concept that received a considerable attention by well-known established algebraists. The interest of amalgamations resides in their ability to cover basic constructions in commutative algebra, including classical pullbacks and trivial ring extensions. Moreover, other classical constructions (such as $A + XB[X]$, $A + XB[[X]]$ and the $D + M$ constructions) can be studied as particular cases of the amalgamation ([10, Examples 2.5 and 2.6]) and other classical constructions, such as the CPI extensions (in the sense of Boisen and Sheldon [9]) are strictly related to it ([10, Example 2.7 and Remark 2.8]). In [10, 13, 17], the authors studied the basic properties of this construction (e.g., characterizations for $A \bowtie^f J$ to be a Noetherian ring, an integral domain, a reduced ring) and they characterized those distinguished pullbacks that can be expressed as an amalgamation.

Theorem 2.13. Let $f : A \to B$ be a ring homomorphism, J an ideal of B, and let $\psi : \mathfrak{I}(A) \to \mathfrak{I}(A) \cup \{\emptyset\}$ and $\phi : \mathfrak{I}(A \bowtie^f J) \to \mathfrak{I}(A \bowtie^f J) \cup \{\emptyset\}$ be two functions such that $\phi(I \bowtie^f J) = \psi(I) \bowtie^f J$ if $\psi(I) \neq \emptyset$ and $\phi(I \bowtie^f J) = \emptyset$ if $\psi(I) = \emptyset$. If $I \bowtie^f J$ is a ϕ-(n,J)-ideal of $A \bowtie^f J$, then I is a ψ-(n,J)-ideal of R. Moreover, the converse is true if $J \subseteq Jac(B)$.

Proof. If $I \bowtie^f J$ is a ϕ-(n,J)-ideal of $A \bowtie^f J$, then $(I \bowtie^f J)/(0 \bowtie^f J)$ is $\phi_{0 \bowtie^f J}$-(n,J)-ideal by Theorem 2.10. We conclude that I is a ψ-(n,J)-ideal of A. For the converse, suppose that I is a ψ-(n,J)-ideal of A and let $(x_1, f(x_1) + j_1), \ldots, (x_{n+1}, f(x_{n+1}) + j_{n+1}) \in A \bowtie^f J$ such that

$$(x_1, f(x_1) + j_1) \cdots (x_{n+1}, f(x_{n+1}) + j_{n+1}) \in I \bowtie^f J \setminus \phi(I \bowtie^f J).$$

So, $x_1 \cdots x_{n+1} \in I \setminus \psi(I)$. If $x_1 \cdots x_n \in I$, then $(x_1, f(x_1) + j_1) \cdots (x_n, f(x_n) + j_n) \in I \bowtie^f J$. Now, assume that there exists $k \in \{1, \ldots, n\}$ such that $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in Jac(A)$, then

$$(x_1, f(x_1) + j_1) \cdots (x_k, \widehat{f(x_k)} + j_k) \cdots (x_{n+1}, f(x_{n+1}) + j_{n+1}) = (x_1 \cdots \widehat{x_k} \cdots x_{n+1}, f(x_1 \cdots \widehat{x_k} \cdots x_{n+1}) + j)$$

for some $j \in J$. This implies that $(x_1, f(x_1) + j_1) \cdots (x_k, \widehat{f(x_k)} + j_k) \cdots (x_{n+1}, f(x_{n+1}) + j_{n+1}) \in Jac(A \bowtie^f J)$ beacuse $J \subseteq Jac(B)$. This completes the proof. □

Corollary 2.14. *Let A be a ring and I, P be proper ideals of A. let $\psi : \mathfrak{I}(A) \to \mathfrak{I}(A) \cup \{\emptyset\}$ and $\phi : \mathfrak{I}(A \bowtie I) \to \mathfrak{I}(A \bowtie I) \cup \{\emptyset\}$ be two functions such that $\phi(P \bowtie I) = \psi(P) \bowtie I$ if $\psi(P) \neq \emptyset$ and $\phi(P \bowtie I) = \emptyset$ if $\psi(P) = \emptyset$. Assume that $I \subseteq Jac(A)$. Then, $P \bowtie I$ is a ϕ-(n,J)-ideal of $A \bowtie I$ if and only if P is a ψ-(n,J)-ideal of A.*

Corollary 2.15. *Let M be a maximal ideal of an integral domain T and D a subring of T such that $D \cap M = \{0\}$. Assume that T and D are local and let I be an ideal of D. Then $I + M$ is an (n,J)-ideal of $D + M$ if and only if I is an (n,J)-ideal of D.*

Proof. Let $\sigma : D \hookrightarrow T$ be the natural embedding. Clearly, $D \bowtie^\sigma M$ is a local ring by [21, Corollary 5.5]. Now, since $D \bowtie^\sigma M \cong D + M$ (as $D \cap M = \{0\}$) and $I \bowtie^\sigma M \cong I + M$, then $I + M$ is an (n,J)-ideal of $D + M$ if and only if $I \bowtie^\sigma M$ is an (n,J)-ideal of $D \bowtie^\sigma M$. By Theorem 2.13, it follows that $I + M$ is an (n,J)-ideal of $D + M$ if and only if I is an (n,J)-ideal of D.

□

3 n-J-ideals of commutative rings

We start this section by the following definition.

Definition 3.1. Let R be a ring. A proper ideal I of R is said to be an n-J-ideal if whenever $x_1 \cdots x_{n+1} \in I$, for $x_1, \dots, x_{n+1} \in R$, then $x_1 \cdots x_n \in Jac(R)$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in I$, for some $k \in \{1, \dots, n\}$.

Example 3.2. For a prime number p, the ideal $I = p\mathbb{Z}$ is an n-J-ideal which is not an (n,J)-ideal. Indeed. Set $x_i = 1$ for $i = 1,\dots,n$ and $x_{n+1} = p$, we then have $x_1 \cdots x_{n+1} \in I$ but neither $x_1 \cdots x_n \in I$ nor $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in Jac(\mathbb{Z})$, for $k \in \{1, \dots, n\}$.

Theorem 3.3. Let $I \subseteq Jac(R)$ be an ideal of a ring R. If $Jac(R)$ is a divided prime ideal, then I is an n-J-ideal if and only if I is a J-ideal.

Proof. It is clear that if I is a J-ideal, then it is an n-J-ideal. Conversely, suppose that $ab \in I$ and $a \notin Jac(R)$ for some $a, b \in R$. Since $Jac(R)$ is prime, $ab \in Jac(R)$ and $a \notin Jac(R)$ implies that $b \in Jac(R)$. Now, since $Jac(R)$ is divided and $a^{n-1} \notin Jac(R)$, there exists an element $c \in R$ such that $b = a^{n-1}c$. Hence $ab = a^n c \in I$. Observe that $a^n \notin Jac(R)$. Since I is an n-J-ideal, it follows that $a^{n-1}c = b \in I$. This proves that I is a J-ideal. □

The next result is a characterization of 2-J-ideals.

Theorem 3.4. For an ideal I of a ring R, the following assertions are equivalent

1. I is a 2-J-ideal of R.
2. If $a, b \in R$ and $ab \notin Jac(R)$, then $(I : ab) \subseteq (I : a) \cup (I : b)$.
3. If $a, b \in R$ and $ab \notin Jac(R)$, then $(I : ab) \subseteq (I : a)$ or $(I : ab) \subseteq (I : b)$.
4. If $abJ \subseteq I$ for some $a, b \in R$ and an ideal J of R, then $ab \in Jac(R)$ or $aJ \subseteq I$ or $bJ \subseteq I$.
5. If $JKL \subseteq I$ for some ideals I, J, K of R, then $JK \subseteq Jac(R)$ or $KL \subseteq I$ or $JL \subseteq I$.

Proof. (1) $\Rightarrow$ (2) Let $c \in (I : ab)$. Then $abc \in I$ and $ab \notin Jac(R)$, so $ac \in I$ or $bc \in I$, hence $c \in (I : a)$ or $c \in (I : b)$, that is $(I : ab) \subseteq (I : a) \cup (I : b)$.
(2) $\Rightarrow$ (3) The result is clear.
(3) $\Rightarrow$ (4) Assume that $abJ \subseteq I$ but neither $ab \in Jac(R)$ nor $aJ \subseteq I$ nor $bJ \subseteq I$. Then there exist $j_1, j_2 \in J$ such that $aj_1 \notin I$ and $bj_2 \notin I$. Since $abj_1, abj_2 \in I, aj_1 \notin I$ and $bj_2 \notin I$, it means that $(I : ab) \not\subseteq (I : a)$ and $(I : ab) \not\subseteq (I : b)$ which is a contradiction to (3).
(4) $\Rightarrow$ (5) Suppose that $JKL \subseteq I$ but neither $JK \subseteq Jac(R)$ nor $KL \subseteq I$ nor $JL \subseteq I$. Then there exist $j_1, j_2 \in J, k_1, k_2 \in K$ satisfying $j_1k_1 \notin Jac(R), k_2L \not\subseteq I$ and $j_2L \not\subseteq I$. Since $j_2k_2L \subseteq I, j_2L \not\subseteq I$ and $k_2L \not\subseteq I$, we conclude that $j_2k_2 \in Jac(R)$. Since $j_1k_1L \subseteq I$ and $j_1k_1 \notin Jac(R)$, we have either $j_1L \subseteq I$ or $k_1L \subseteq I$ by (4). So there are three cases. Suppose that $j_1L \subseteq I$ but $k_1L \not\subseteq I$. Since $j_2k_1L \subseteq I, j_2L \not\subseteq I$ and $k_1L \not\subseteq I$, we have $j_2k_1 \in Jac(R)$. Since $j_1L \subseteq I$ but $j_2L \not\subseteq I$, we write $(j_1 + j_2)L \not\subseteq I$. Since $(j_1 + j_2)k_1L \subseteq I$ but $k_1L \not\subseteq I$, we conclude that $(j_1 + j_2)k_1 \in Jac(R)$. Thus $j_1k_1 \in Jac(R)$ which is a contradiction. If $k_1L \subseteq I$ but $j_1L \not\subseteq I$. Similarly, we obtain a contradiction. Now, suppose that $j_1L \subseteq I$ and $k_1L \subseteq I$. Since $k_1L \subseteq I$ and $k_2L \not\subseteq I$, we have that $(k_1 + k_2)L \not\subseteq I$. since $j_2(k_1 + k_2)L \subseteq I$, and $j_2L \not\subseteq I$, we conclude that $j_2(k_1 + k_2) \in Jac(R)$. Since $j_2k_2 \in Jac(R)$, we have $j_2k_1 \in Jac(R)$. On the other hand, since $j_1L \subseteq I$ and $j_2L \not\subseteq I$, we have $(j_1 + j_2)L \not\subseteq I$ and proceeding as above we obtain that $j_1k_2 \in Jac(R)$. Finally, since $(j_1 + j_2)(k_1 + k_2)L \subseteq I$ but $(j_1 + j_2)L \not\subseteq I$ and $(k_1 + k_2)L \not\subseteq I$, we deduce that $(j_1 + j_2)(k_1 + k_2) \in Jac(R)$. Since $j_2k_2, j_2k_1, j_1k_2 \in Jac(R)$, we conclude $j_1k_1 \in Jac(R)$, a contradiction.
(5) $\Rightarrow$ (1) Let $a, b, c \in R$ with $abc \in I$. Take $J = (a), K = (b)$ and $L = (c)$ in (5). □

Theorem 3.5. Let $f : R_1 \to R_2$ be a ring homomorphism.

1. If $f^{-1}(Jac(R_2)) \subseteq Jac(R_1)$ and I_2 is an n-J-idealof R_2, then $f^{-1}(I_2)$ is an n-J-ideal of R_1.

2. Assume that f is an epimorphism and $\mathrm{Ker}(f) \subseteq I_1$. If I_1 is an n-J-ideal of R_1, then $f(I_1)$ is an n-J-ideal of R_2.

Proof. (1) Take $a_1,...,a_{n+1} \in R_1$ and assume that $a_1a_2...a_{n+1} \in f^{-1}(I_2)$. Then $f(a_1a_2...a_{n+1}) = f(a_1)f(a_2)...f(a_{n+1}) \in I_2$, which implies $f(a_1)...f(a_n) = f(a_1...a_n) \in Jac(R_2)$ or $f(a_1\cdots\widehat{a_k}\cdots a_{n+1}) \in I_2$, for some $k \in \{1,\dots,n\}$. As $f^{-1}(Jac(R_2)) \subseteq Jac(R_1)$, it follows that $a_1...a_n \in Jac(R_1)$ or $a_1\cdots\widehat{a_k}\cdots a_{n+1} \in f^{-1}(I_2)$ for some $k \in \{1,\dots,n\}$. Thus $f^{-1}(I_2)$ is an n-J-ideal of R_1.
(2) Take $a_i = f(x_1)$ and assume that $a_1...a_{n+1} = f(x_1...x_{x+1}) \in f(I_1)$. It follows from $\mathrm{Ker}(f) \subseteq I_1$ that $x_1...x_{n+1} \in I_1$, and it follows that $x_1...x_n \in Jac(R_1)$ or $x_1\cdots\widehat{x_k}\cdots x_{n+1} \in I_1$, for some $k \in \{1,\dots,n\}$. Hence $a_1...a_{n+1} \in Jac(R_2)$ or $a_1\cdots\widehat{a_k}\cdots a_{n+1} \in f(I_1)$, for some $k \in \{1,\dots,n\}$, as required. □

In view of Theorem 3.5, we conclude the following result.

Corollary 3.6. *Let $J \subseteq I$ be ideals of R. If I is an n-J-ideal of R, then I/J is an n-J-ideal of R/J.*

Theorem 3.7. . Let $S \subset R$ be multiplicatively closed. If I is an n-J-ideal of R and $I \cap S = \emptyset$, then I_S is an n-J-ideal of R_S.

Proof. Suppose that $\frac{a_1}{s_1}...\frac{a_{n+1}}{s_{n+1}} \in I_S$. Then $ua_1...a_{n+1} \in I$ for some $u \in S$. Hence $a_1...a_n \in Jac(R)$ or $a_1\cdots\widehat{a_k}\cdots a_{n+1} \in I$, for some $k \in \{1,\dots,n\}$. This implies that $\frac{a_1}{s_1}...\frac{a_n}{s_n} \in (Jac(R))_S \subseteq Jac(R_S)$ or $\frac{a_1}{s_1}...\frac{\widehat{a_k}}{s_k}...\frac{a_{n+1}}{s_{n+1}} = \frac{ua_1...\widehat{a_k}...a_{n+1}}{us_1...\widehat{s_k}...s_{n+1}} \in I_S$, and it follows that I_S is an n-J-ideal of R_S. □

Proposition 3.8. *Let R_1 and R_2 be rings and let I be an ideal of R_1. Then the following statements are equivalent.*

1. *$I \times R_2$ is a 2-J-ideal of $R = R_1 \times R_2$.*

2. *I is a prime ideal of R_1.*

3. *$I \times R_2$ is a prime ideal of $R = R_1 \times R_2$.*

Proof. (1) $\Rightarrow$ (2) Let $a,b \in R_1$ and $ab \in I$. Thus we have $(a,1)(b,1)(1,1) \in I \times R_2$. Since $(a,1)(b,1) \notin Jac(R)$, we conclude $(a,1)(1,1) \in I$ or $(b,1)(1,1) \in I$, that is, $a \in I$ or $b \in I$, as needed.
(2) $\Rightarrow$ (3) $\Rightarrow$ (1) is clear. □

Theorem 3.9. Let I_1 and I_2 be proper ideals of R_1 and R_2 respectively. The following statements are equivalent.

1. $I = I_1 \times I_2$ is a 2-J-ideal of $R = R_1 \times R_2$.

2. $I_1 = Jac(R_1)$ and $I_2 = Jac(R_2)$ are prime ideals.

3. $I = I_1 \times I_2$ is a 2-absorbing ideal of R and $I \subseteq Jac(R)$.

Proof. (1) $\Rightarrow$ (2) Assume that $I_1 \neq Jac(R_1)$, and take $a \in I_1 \backslash Jac(R_1)$. Then $(a,1)(1,0)(1,1) \in I$ and $(a,1)(1,0) \notin Jac(R)$. Since I is a 2-J-ideal, we conclude that $(a,1)(1,1) \in I_1 \times I_2$ or $(1,0)(1,1) \in I_1 \times I_2$, a contradiction. Thus $I_1 = Jac(R_1)$. If $I_1 = Jac(R_1)$ is not prime, then there are elements $a, b \notin Jac(R_1) = I_1$ such that $ab \in I_1$. Then $(a,1)(b,1)(1,0) \in I, (a,1)(b,1) \notin Jac(R), (a,1)(1,0) \notin I$ and $(b,1)(1,0) \notin I$, a contradiction. Thus I_1 is prime in R_1. The same arguments show that $I_2 = Jac(R_2)$ is a prime ideal of R_2.
(2) $\Rightarrow$ (3) Suppose that $I_1 = Jac(R_1)$ and $I_2 = Jac(R_2)$ are prime ideals. Hence, $I_1 \times R_2$ and $R_1 \times I_2$ are prime ideals of $R = R_1 \times R_2$. Since intersection of two prime ideals is always 2-absorbing by [6], we conclude that $I = (I_1 \times R_2) \cap (R_1 \times I_2)$ is a 2-absorbing ideal of R.
(3) $\Rightarrow$ (1) It is clear that I is 2-absorbing, then I is a 2-J-ideal. □

Now, we study the n-J-ideals of trivial ring extensions.

Theorem 3.10. Let R be a ring and M be an R-module. Then $I \ltimes M$ is an n-J-ideal of $R \ltimes M$ if and only if I is an n-J-ideal of R.

Proof. Assume that $x_1 \cdots x_{n+1} \in I$ for some $x_1, \ldots, x_{n+1} \in R$ such that $x_1 \cdots x_n \notin Jac(R)$. Hence $(x_1, 0) \cdots (x_{n+1}, 0) = (x_1 \cdots x_{n+1}, 0) \in I \ltimes M$. Since $I \ltimes M$ is an n-J-ideal of $R \ltimes M$ and $x_1 \cdots x_n \notin Jac(R)$, there is $k \in \{1, \ldots, n\}$ such that $(x_1, 0) \cdots \widehat{(x_k, 0)} \cdots (x_{n+1}, 0) \in I \ltimes M$. Thus $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in I$. Conversely, if $(x_1, m_1) \cdots (x_{n+1}, m_{n+1}) \in I \ltimes M$ then $x_1 \cdots x_{n+1} \in I$. The fact that I is an n-J-ideal of R implies that either $x_1 \cdots x_n \in Jac(R)$ or $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in I$. Therefore, we have either $(x_1, m_1) \cdots (x_n, m_n) \in Jac(R \ltimes M)$ or $(x_1, m_1) \cdots \widehat{(x_k, m_k)} \cdots (x_{n+1}, m_{n+1}) \in I \ltimes M$, as desired. □

We close this section by the following result.

Theorem 3.11. Let $f : A \to B$ be a ring homomorphism and J an ideal of B. If $I \bowtie^f J$ is an n-J-ideal of $A \bowtie^f J$, then I is an n-J-ideal of A. Moreover, the converse is true if $J \subseteq Jac(B)$.

Proof. If $I \bowtie^f J$ is an n-J-ideal of $A \bowtie^f J$, then $(I \bowtie^f J)/(0 \bowtie^f J)$ is an n-J-ideal by Theorem 2.10. We conclude that I is an n-J-ideal of A. For the converse, suppose that I is an n-J-ideal-ideal of A and let $(x_1, f(x_1) + j_1), \ldots, (x_{n+1}, f(x_{n+1}) + j_{n+1}) \in A \bowtie^f J$ such that

$$(x_1, f(x_1) + j_1) \cdots (x_{n+1}, f(x_{n+1}) + j_{n+1}) \in I \bowtie^f J.$$

So, $x_1 \cdots x_{n+1} \in I$. If there exists $k \in \{1, \ldots, n\}$ such that $x_1 \cdots \widehat{x_k} \cdots x_{n+1} \in I$, then

$$(x_1, f(x_1) + j_1) \cdots (x_k, \widehat{f(x_k)} + j_k) \cdots (x_{n+1}, f(x_{n+1}) + j_{n+1}) \in I \bowtie^f J.$$

Now, assume that $x_1 \cdots x_n \in Jac(A)$ then

$$(x_1, f(x_1) + j_1) \cdots \cdots (x_n, f(x_n) + j_n) = (x_1 \cdots \cdots x_n, f(x_1 \cdots \cdots x_n) + j)$$

for some $j \in J$. This implies that $(x_1, f(x_1) + j_1) \cdots \cdots (x_n, f(x_n) + j_n) \in Jac(A \bowtie^f J)$ because $J \subseteq Jac(B)$. This completes the proof. □

References

[1] D. F. Anderson and A. Badawi, *On n-absorbing ideals of commutative rings*, Comm. Algebra 39 (2011), no. 5, 1646–1672.

[2] D. D. Anderson and M. Bataineh, *Generalization of prime ideals*, Comm. Algebra 36 (2008), no. 2, 686-696.

[3] D. D. Anderson and E. Smith, *Weakly prime ideals*, Houston J. Math. 29 (2003), no. 4, 831-840.

[4] D. D. Anderson and M. Winders, *Idealization of a module*, J. Commut. Algebra 1 (2009), no. 1, 3–56.

[5] A. Anebri, N. Mahdou, E. Yıldız and Ü. Tekir, *On ϕ-(n,N)-ideals of commutative rings*, algebra colloquium, in press.

[6] A. Badawi, *On 2-absorbing ideals of commutative rings*, *Bull. Austral. Math. Soc.* 75(3)(2007), 417–429.

[7] A. Badawi, U. Tekir and E. Yetkin, *On 2-absorbing primary ideals in commutative rings*, Bull. Korean Math. Soc. 51(4) (2014) 1163–1173.

[8] S. M. Bhatwadekar and P. K. Sharma, *Unique factorization and birth of almost primes*, Comm. Algebra 33 (2005), no. 1, 43-49.

[9] M. Boisen and P.B. Sheldon, *CPI-extensions : overrings of integral domains with special prime spectrums*, Canad. J. Math. 29 (1977) 722-737.

[10] M. D'Anna, C. A. Finocchiaro and M. Fontana, *Amalgamated algebras along an ideal*, Commutative algebra and its applications, Walter de Gruyter, Berlin, (2009) 241-252.

[11] M. D'Anna, C. A. Finocchiaro and M. Fontana, *Properties of chains of prime ideals in an amalgamated algebra along an ideal*, J. Pure Appl. Algebra (2010), 1633-1641.

[12] M. D'Anna and M. Fontana, *The amalgamated duplication of ring along an ideal: the basic properties*, J. Algebra Appl. 6(3) (2007), 241-252.

[13] M. D'Anna, C.A. Finocchiaro, M. Fontana, *New algebraic properties of an amalgamated algebras along an ideal*, Comm. Algebra 44(5) (2016), 1836-1851.

[14] D. E. Dobbs, A. El Khalfi and N. Mahdou, *Trivial extensions satisfying certain valuation-like properties*, Comm. Algebra 47(5) (2019), 2060-2077.

[15] T. Dumitrescu, N. Mahdou and Y. Zahir, *Radical factorization for trivial extension and amalgamated duplication rings*, J. Algebra Appl. 20(2) (2021), 2150025, 10 pp.

[16] M. Ebrahimpour and R. Nekooei, *On generalizations of prime ideals*, Comm. Algebra 40(4) (2012), 1268-1279.

[17] A. El Khalfi, H. Kim and N. Mahdou, Amalgamation extension in commutative ring theory, a survey, Moroccan Journal of algebra and Geometry with applications, 1(1) (2022), 139-182.

[18] S. Glaz, *Commutative Coherent Rings*, Lecture Notes in Math. 1371, Springer-Verlag, Berlin, 1989.

[19] J. A. Huckaba, *Commutative Rings with Zero Divisors*, Dekker, New York, 1988.

[20] S. Kabbaj, *Matlis' semi-regularity and semi-coherence in trivial ring extensions : a survey*, Moroccan Journal of Algebra and Geometry with Applications, 1(1) (2022), 1-17.

[21] S. Kabbaj, K. Louartiti, and M. Tamekkante, *Bi-amalgamated algebras along ideals*, J. Commut. Algebra 9 (2017), no. 1, 65-87.

[22] I. Kaplansky, *Commutative Rings*, rev. ed., Univ. Chicago Press, Chicago, 1974.

[23] H. A. Khashan and B. B. Bani-Ata, *J-ideals of commutative rings*, Int. Electron. J. Algebra 29(1) (2021), 148-164.

[24] H. Mostafanasab and A. Y. Darani, *On ϕ-n-absorbing primary idals of commutative rings*, J. Korean Math. Soc. 53 (2016), 549-582.

[25] E. Yildiz, Ü. Tekir and S. Koç, $(2,J)$-*Ideals in commutative rings*, C. R. L'Acad. Emie Bulg. Sci., 73(9)(2020), 1201-1209.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco
Volume 2, Issue 1 (2023), pp 154-161
ISSN: 2820-7114

Title :

Note on *UN*-rings

Author(s):

Abdeslam Mimouni

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 154–161
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco **ISSN: 2820-7114**

Note on UN-rings

Abdeslam Mimouni
Department of Mathematics, KFUPM, Dhahran 31261, KSA
e-mail: *amimouni@kfupm.edu.sa*

Communicated by Ünsal Tekir
(Received 10 February 2023, Revised 03 April 2023, Accepted 07 April 2023)

Abstract. Let R be an associative ring with identity. Then R is said to be a UN-ring if every nonunit element of R can be written as a product of a unit and nilpotent elements. If in every UN-decomposition, the unit and nilpotent commute, the ring R is said to be a strongly UN-ring. In this paper, we study the notions of UN-rings in different contexts of commutative rings such us pullbacks, trivial ring extensions and amalgamations of algebras along ideals. Our aim is to generate new families of UN-rings. Namely, constructing UN-rings issued from pullbacks where the starting rings are not necessarily UN-ring and also to enrich the literature with such a rings. Examples illustrating the aims and scopes of our results are given.

Key Words: UN-ring, strongly UN-ring, pullbacks, trivial ring extension, amalgamation.
2010 MSC: 13A15, 13A18, 13F05, 13G05, 13C20.

1 Introduction

The study of associative rings whose elements can be written as a sum/product of units/idempotents/ nilpotent elements is an active areas of research in the theory of associative rings. In 1977, Nicholson introduced the class of clean rings, that is, rings R such that every element can be written as a sum of an idempotent and a unit (see [22], and [23] for more on strongly clean rings). This notion was intensively studied by many authors due to its close connection to the important notion of exchange rings. Since then, different classes of rings have generalized clean ring such as weakly clean rings (i.e., each $x \in R$ can be written as $x = u + e$ or $x = u - e$ where $u \in U(R)$ and $e \in Id(R)$, [1]), almost clean rings (i.e., each $x \in R$ can be written as $x = a + e$ where $a \in reg(R)$ and $e \in Id(R)$, [20]), semi-clean rings, uniquely clean rings etc. Recently, in [28], Zabavsky-Domsha-Romaniv introduced and studied the notions of clear elements and clear rings. An element x in a ring R is said to be a clear element if it can be written as $x = a + u$ where a is a unit-regular element and $u \in U(R)$; and the ring itself is a clear ring if all its elements are clear elements. Clearly a weakly clean element (resp. a weakly clean ring) is a clear element (resp. a clear ring).
While most of the works on rings whose elements have decompositions as a sum of units/idempotent/ nilpotent elements, associative rings whose elements are product of units/idempotent/nilpotent elements have received less much interest though it was an older notion. Recall that a unital associative ring R is said to be unit-regular provided that for any $a \in R$ there exists an invertible element $u \in R$ such that $a = aua$. Notice that ua is an idempotent element and $a = u^{-1}(ua)$, so that a unit regular ring is a ring where every element is a product of a unit and an idempotent element. Unit-regular rings were introduced by Ehrlich in 1968 (see [13]).

Recently, in [7], Calugareanu defined a ring R to be a UN-ring if every nonunit of R can be written as product of a unit and a nilpotent element. A nonunit element $x \in R$ is called strongly UN if in its UN-decomposition, the unit and nilpotent commute. In [7], the author showed that every simple Artinian ring is a UN-ring and that every noninvertible matrix over a division ring is a product of an invertible and a nilpotent matrix. In [25], the author gave a complete characterization of modules over a Dedekind domain and modules of finite length with a UN endomorphism ring. In 2021, Y. Zhou ([30]) answered two questions asked by CÄČlugÄČreanu. The first is whether there is a class of rings properly between the class of UN-rings with 2-good identity and the class of 2-good rings (recall that the notion of k-good rings, that is, rings R such that every element is the sum of k units, was introduced by P. Vamos in [26]). The second question is whether matrix rings over elementary divisor UN-rings are UN-rings. The author proved that matrix rings over right Hermite UN-rings are UN-ring, and consequently, matrix rings over elementary divisor UN-rings are UN-rings. In [24], D. Udar introduced and studied a new class of rings in which each nonunit element of a ring R is a product of a unit and a strongly nilpotent element. He obtained various properties and a complete characterization of these rings. He also investigate the subclass of these rings in which this multiplicative decomposition of nonunit elements is unique, and studied the group ring of these rings. Very recently (2023), in [19], the authors proved that for a commutative ring R, a matrix ring $\mathcal{M}_n(R)$ is UN if and only if R is UN-ring. Also they proved that for a ring R and for a nontrivial group G, if the group ring $R[G]$ is a UN-ring, then R is a UN-ring of characteristic p^α, G is a p-group and $p \in J(R)$. The converse holds if G is locally finite.

Pullbacks, trivial ring extensions of rings by modules and amalgamations of algebras along ideals are known to be important sources of examples and counter-examples, and have proven to be useful in solving many open problems and conjectures for various contexts in (commutative and non-commutative) ring theory. The purpose of this paper is to study the transfer of the notion of UN-rings to pullbacks, trivial ring extensions and amalgamations of algebras along ideals. Our aim is to generate new families of clear rings and to enrich the literature with such a rings. Section 2 deals with different types of pullbacks arising from commutative diagrams of canonical homomorphisms. First, we deal a general pullback called a diagram of type (Δ) arising from the following diagram, where T is a commutative ring, I is a nonzero ideal (but not necessarily maximal) of T and D is a subring of T/I:

$$(\Delta) \qquad \begin{array}{ccc} R := \phi^{-1}(D) & \longrightarrow & D \\ \downarrow & & \downarrow \\ T & \stackrel{\phi}{\longrightarrow} & T/I. \end{array}$$

We prove that R is a UN-ring if and only if D is a UN-ring and $I \subseteq Nil(R)$, (Theorem 2.4). As a first consequence, we prove that if R is the pullback of a diagram of type ($\square$), that is, a pullbcak issued from a diagram of type (T, M, D), where $I = M$ is a maximal ideal of T, then R is a UN-ring if and only $D = k$ is a field and $M = Nil(R)$ if and only if T is a UN-ring and $D = k$ is a field. In this case, R is local (so is T) with maximal ideal M and $M = Nil(R) = Nil(T)$ (Corollary 2.5).
A second consequence of Theorem 2.4 is a diagram of type (Δ) where we assume that I is an M-primary ideal of T for some maximal ideal M of T. We prove that for a such pullback, R is a UN-ring if and only if D and T are UN-rings (Corollary 2.6). A special case is when $I = J(T) \neq (0)$ is the Jacobson radical of T. We prove that if R is a UN-ring if and only if $D = k$ is a field and $Nil(R) = J(T)$ (Corollary 2.7).

Section 3 investigates the notion of UN-rings in trivial ring extensions and amalgamations of algebras along ideals. First, we prove that if A is a commutative ring and E is an A-module, then $R = A \ltimes E$ is a UN-ring if and only if so is A (Proposition 3.1). Next, we deal with amalgamations of algebras along ideals. We give a complete characterization of when the amalgamation of an algebra along an ideal is a UN-ring. Precisely, we prove that if A and B are commutative rings, J is an ideal of B, $f : A \longrightarrow B$ is a ring homomorphism and $R = A \bowtie^f J$, then R is a UN-ring if and only if A is a UN-ring, $J \subseteq Nil(B)$ and $f(U(A)) + J \subseteq U(f(A) + J)$, that is, for every $u \in U(A)$ and for every $j \in J$,

$(u, f(u)+j) \in U(f(A)+J)$ (Theorem 3.2).

We close the paper with a fourth section with examples illustrating the results obtained in Sections 2 and 3.

2 pullback constructions

Let T be a commutative ring, I a nonzero ideal (not necessarily maximal) of T, $\phi : T \longrightarrow T/I$ the canonical surjection, and D a proper subring of T/I. Let R be the pullback issued from the following diagram of canonical homomorphisms:

$$(\Delta) \qquad \begin{array}{ccc} R := \phi^{-1}(D) & \longrightarrow & D \\ \downarrow & & \downarrow \\ T & \stackrel{\phi}{\longrightarrow} & T/I. \end{array}$$

We refer to this diagram as a diagram of type (Δ), and if $I = M$ is a maximal ideal of T, we refer to this as a diagram of type $(\square)$. Clearly, $I = (R : T)$ and $D \cong R/I$. For ample details on the ideal structure of R and its ring-theoretic properties, we refer the reader to [2, 3, 4, 5, 16, 17, 18]. The case where $T = V$ is a valuation domain is crucial and we will refer to this case as a classical diagram of type $(\square)$.

We start this section with two trivial propositions that will be often used for constructing examples of UN-commutative rings. The first one shows that a commutative ring R is a UN-ring if and only if R is a local ring with maximal ideal $Nil(R)$. The second one deals with a particular quotient of polynomial rings and power series rings over a commutative ring. Notice that over a UN-ring R, neither the polynomial ring $R[X]$ nor the power series ring $R[[X]]$ is a UN-ring (as X cannot be a product of a unit and a nilpotent element). However, some particular cases of their quotients are UN-rings.

Proposition 2.1. *A commutative ring R is a UN-ring if and only if $R = U(R)\dot{\cup}Nil(R)$ if and only if R is local with maximal ideal $Nil(R)$.*

Proof. Assume that R is a UN-ring and let $x \in R \setminus U(R)$. Then $x = uy$ where $u \in U(R)$ and $y \in Nil(R)$, and so $x^n = u^n y^n = 0$ when $y^n = 0$ for some positive integer n. Thus $x \in Nil(R)$ and so $R = U(R)\dot{\cup}Nil(R)$, as desired. □

Corollary 2.2. *Let $\mathbb{Z}_n$ be the ring of integers modulo $n \geq 2$. Then $\mathbb{Z}_n$ is a UN-ring if and only if n is a prime power, that is $n = p^r$ for some prime integer p and positive integer r.*

Proof. Assume that $\mathbb{Z}_n$ is a UN-ring and let $n = p_1^{r_1} \dots p_s^{r_s}$ be the prime decomposition of n. By Proposition 2.1, $\mathbb{Z}_n$ is local with maximal ideal $Nil(\mathbb{Z}_n)$. If $s \geq 2$, then $\mathbb{Z}_n$ have at least two maximal ideals ($p_1\mathbb{Z}_n$ and $p_2\mathbb{Z}_n$), which is absurd. Thus $s = 1$ and so $n = p_1^{s_1}$, as desired.
Conversely, clearly $\mathbb{Z}_{p^r}$ is a local ring with maximal ideal $p\mathbb{Z}_{p^r} = Nil(\mathbb{Z}_{p^r})$. Thus $\mathbb{Z}_{p^r}$ is a UN-ring by Proposition 2.1. □

In [19, Theorem 2.2], the authors proved that if I is an ideal of a ring R, then R/I is a UN-ring if and only if R/I^n is a UN-ring for all $n \in \mathbb{N}$ if and only if R/I^n is a UN-ring for some $n \in \mathbb{N}$. It is worth to notice the next proposition is an immediate consequence of [19, Theorem 2.2]. However, we present it here with a direct proof; and the given form of units in $A[X]/(X^n)$ (resp. $A[[X]]/(X^n)$) is needed in constructing examples in the amalgamations of algebras along ideals section (see Example 4.3).

Proposition 2.3. *Let A be a commutative ring, X an indeterminate over A and $n \geq 2$ an integer. Then $R = A[X]/(X^n)$ (resp. $R = A[[X]]/(X^n)$) is a UN-ring if and only if A is a UN-ring.*

Proof. Let x be the class of X modulo X^n. Then every element of R is of the form $a_0+a_1x+\cdots+a_{n-1}x^{n-1}$ for some $a_0,\ldots,a_{n-1} \in A$. Also notice that $U(R) = U(A) + Ax + \cdots + Ax^{n-1}$. Indeed, let $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \in U(R)$ and let $g(x) = b_0 + b_1x + \cdots + b_{n-1}x^{n-1} \in R$ such that $f(x)g(x) = 1$. Then $a_0b_0 = 1$ and so $a_0 \in U(A)$. Conversely, let $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \in R$ with $a_0 \in U(A)$, and let $g(x) = b_0 + b_1x + \cdots + b_{n-1}x^{n-1} \in R$ such that $f(x)g(x) = 1$. Then such a polynomial $g(x)$ exists provided that the coefficients b_i exist. Whence all b_i are recrucively given by: $a_0b_0 = 1$, so $b_0 = a_0^{-1}$, $a_0b_1 + a_1b_0 = 0$, so $b_1 = -a_0^{-2}a_1$, $a_0b_2 + a_1b_1 + a_2b_0 = 0$, so $b_2 = a_0^{-3}a_1^2 - a_0^{-2}a_2$. Iterating this process, and assuming that $b_0, b_1, \ldots, b_{n-2}$ are given, the coefficient b_{n-1} is given by: $a_0b_{n-1}+a_1b_{n-2}+\cdots+a_{n-1}b_0 = 0$, so $b_{n-1} = -a_0^{-1}(a_1b_{n-2} + \cdots + a_{n-1}b_0)$. It follows that $f(x) \in U(R)$, as desired.
Now, assume that A is a UN-ring and let $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} = a_0 + xg(x) \in R \setminus U(R)$. Then $a_0 \in A \setminus U(A)$, and so $a_0 \in Nil(A)$ by Proposition 2.1. Since $x^n = 0$, $(xg(x))^n = 0$ and so $xg(x) \in Nil(R)$. Thus $f(x) = a_0 + xg(x) \in Nil(R)$, and therefore R is a UN-ring by Proposition 2.1.
The converse is straightforward. □

The next theorem characterizes UN-pullback rings.

Theorem 2.4. For the diagram (Δ), R is a UN-ring if and only if D is a UN-ring and $I \subseteq Nil(R)$.

Proof. Assume that R is a UN-ring. Then $R = U(R)\dot{\cup}Nil(R)$, and thus $I \subseteq Nil(R)$. Now D is a UN-ring follows from [7, Proposition 3.1], as a homomorphic image of a UN-ring.
Conversely, assume that D is a UN-ring and $I \subseteq Nil(R)$. Then $D = U(D)\dot{\cup}Nil(D)$. Let $x \in R\setminus U(R)$ and set $d = \phi(x)$. If $d \in U(D)$, then $d^{-1} = \phi(y)$ for some $y \in R$. Thus $1 = dd^{-1} = \phi(xy)$ and so $1 - xy = a \in I$. Since $I \subseteq Nil(R)$, $(1 - xy)^r = a^r = 0$ for some positive integer r. Thus $1 - xz = 0$ and so $1 = xz$ for some $z \in R$, which is absurd. Hence $d \in D \setminus U(D) = Nil(D)$. Thus $\phi(x^s) = (\phi(x))^s = d^s = 0$ for some positive integer s. Hence $x^s \in I \subseteq Nil(R)$ and therefore $x \in Nil(R)$. Thus $R = U(R)\dot{\cup}Nil(R)$ and therefore R is a UN-ring. □

Corollary 2.5. *For the diagram of type $(\square)$, the following conditions are equivalent.*

1. *R is a UN-ring.*
2. *$D = k$ is a field and $M = Nil(R)$.*
3. *T is a UN-ring and $D = k$ is a field.*

In this case, R is local (so is T) with maximal ideal M and $M = Nil(R) = Nil(T)$.

Proof. (1) $\Longleftrightarrow$ (2) follows immediately from Theorem 2.4.
(1) $\Longrightarrow$ (3). Assume that R is a UN-ring. Clearly $M = Nil(R) \subseteq Nil(T) \subseteq M$ and so $M = Nil(R) = Nil(T)$. Let $x \in T \setminus M$. Then there is $\alpha \in T$ and $m \in M$ such that $1 = x\alpha + m$. Thus $1 - \alpha x = m \in M = Nil(R)$ and so $(1 - \alpha x)^r = m^r = 0$ for some positive integer r. Thus $1 = xz$ for some $z \in T$ and therefore $x \in U(T)$. Hence $T = M\dot{\cup}U(T) = Nil(T)\dot{\cup}U(T)$. Thus T is a UN local ring with maximal ideal M.
(3) $\Longrightarrow$ (2). Assume that T is a UN-ring and $D = k$ is a field. Then $T = U(T)\dot{\cup}Nil(T)$ and so $M \subseteq Nil(T)$. By maximality of M in T, $M = Nil(T)$ and so $M = M \cap R = Nil(T) \cap R = Nil(R)$, as desired. □

Corollary 2.6. *For the diagram of type (Δ), assume that I is an M-primary ideal of T where M is a maximal ideal of T. Then R is a UN-ring if and only if so are D and T.*

Proof. $\Longrightarrow$) Assume that R is a UN-ring. By Theorem 2.4, D is a UN-ring and $I \subsetneq Nil(R)$. Let $m \in M$. Then $\sqrt{I} = M$ implies that $m^s \in I$ for some positive integer s. Thus $m^s \in Nil(R)$ and so $m \in Nil(T)$. Therefore $M \subseteq Nil(T)$ and by maximality, $M = Nil(T)$. Now, let $x \in T \setminus M$. Then there is $a \in T$ and $m \in M$ such that $1 = ax + m$. So $1 - ax = m \in M = Nil(T)$ and so $x \in U(T)$. Thus $T = U(T)\dot{\cup}M = U(T)\dot{\cup}Nil(T)$ and therefore T is a UN-ring.
$\Longleftarrow$) Assume that D and T are UN-rings. Then $M = Nil(T)$ and so $I \subseteq M \cap R = Nil(T) \cap R = Nil(R)$. By Theorem 2.4, R is a UN-ring. □

Corollary 2.7. *For the diagram of type* (Δ)*, assume that* $I = J(T) \neq 0$*. Then* R *is a* UN*-ring if and only* $D = k$ *is a field and* $Nil(R) = J(T)$*.*

Proof. Assume that R is a UN-ring. Then $J(T) = I \subseteq Nil(R) \subseteq Nil(T) \subseteq J(T)$. Thus $J(T) = Nil(T) = Nil(R)$ and so R is local with maximal ideal $Nil(R) = J(T) = I$. Hence $D \cong R/I$ is a field.
The converse follows immediately from Theorem 2.4. □

3 Trivial ring extensions and amalgamations

Let A be a commutative ring and E an A-module. The trivial ring extension of A by E (also called the idealization of E over A) is the ring $R = A \ltimes E$ whose underlying group is $A \times E$ with multiplication given by $(a,e)(a',e') = (aa', ae' + a'e)$. It was introduced by Nagata [21] in order to facilitate interaction between rings and their modules as well as provide diverse contexts of rings with zero-divisors.

Proposition 3.1. *Let* A *be a commutative ring,* E *an* A*-module and* $R = A \ltimes E$*. Then* R *is a* UN*-ring if and only if so is* A*.*

Proof. Assume that A is a UN-ring and let $(x,e) \in R \setminus U(R)$. Then $a \in A \setminus U(A)$ and so $a = ub$ where $u \in U(A)$ and $b \in Nil(A)$. Set $f = u^{-1}(1-b)e \in E$. Then it is easy to see that $(a,e) = (u,e)(b,f)$ and clearly $(u,e) \in U(R)$ and if $b^n = 0$, then $(b,f)^{n+1} = (b^{n+1}, (n+1)b^n f) = (0,0)$ is a nilpotent element of R. Thus R is a UN-ring.
The converse is straightforward. □

Let A and B be two rings with unity, let J be an ideal of B and let $f : A \to B$ be a ring homomorphism. In this setting, we can consider the following subring of $A \times B$:

$$A \bowtie^f J := \{(a, f(a) + j) \mid a \in A, j \in J\}$$

called *the amalgamation of* A *and* B *along* J *with respect to* f. This construction is a generalization of *the amalgamated duplication of a ring along an ideal* (introduced and studied by D'Anna and Fontana in [9, 11, 12]). The interest of amalgamation resides, partly, in its ability to cover several basic constructions in commutative algebra, including pullbacks and trivial ring extensions (also called Nagata's idealizations)(cf. [21, page 2]). Moreover, other classical constructions (such as the $A + XB[X]$, $A + XB[[X]]$, and the $D + M$ constructions) can be studied as particular cases of the amalgamation ([10, Examples 2.5 and 2.6]) and other classical constructions, such as the CPI extensions (in the sense of Boisen and Sheldon [6]) are strictly related to it ([10, Example 2.7 and Remark 2.8]). In [10], the authors studied the basic properties of this construction (e.g., characterizations for $A \bowtie^f J$ to be a Noetherian ring, an integral domain, a reduced ring) and they characterized those distinguished pullbacks that can be expressed as an amalgamation. Moreover, in [12], they pursued the investigation on the structure of the rings of the form $A \bowtie^f J$, with particular attention to the prime spectrum, chain properties and Krull dimension. For more details on amalgamations, see [15]

Theorem 3.2. Let A and B be commutative rings, J an ideal of B, $f : A \longrightarrow B$ a ring homomorphism and $R = A \bowtie^f J$. Then R is a UN-ring if and only if A is a UN-ring, $J \subseteq Nil(B)$ and $f(U(A)) + J \subseteq U(f(A)+J)$, that is, for every $u \in U(A)$ and for every $j \in J$, $(u, f(u)+j) \in U(f(A)+J)$.

Proof. Assume that R is a UN-ring and let $u \in U(A)$ and $j \in J$. If $f(u)+j$ is not a unit in $f(A)+J$, then $(u, f(u)+j) \in R \setminus U(R) = Nil(R)$ by Proposition 2.1, and so $u \in Nil(A)$, which is absurd. Thus $f(u)+j \in U(f(A)+J)$. Now, let $a \in A \setminus U(A)$. Then $(a, f(a)) \in R \setminus U(R)$ and so $(a, f(a)) \in Nil(R)$. Thus $a \in Nil(A)$ and therefore $A = U(A) \dot{\cup} Nil(A)$. Again by Proposition 2.1, A is a UN-ring. Similarly, for every $j \in J$, $(0, j) \in R \setminus U(R)$ and so $(0, j) \in Nil(R)$. Thus $j \in Nil(B)$ and therefore $J \subseteq Nil(B)$, as desired. Conversely, let $(a, f(a)+j) \in R \setminus U(R)$. By hypothesis, $a \notin U(A)$ and so $a \in Nil(A)$. Also since $J \subseteq Nil(B)$, $f(a)+j \in Nil(f(A)+J)$. Thus $(a, f(a)+j) \in Nil(R)$, and therefore R is a UN-ring by Proposition 2.1. □

4 Examples

This section is devoted to examples illustrating the results in Sections 2 and 3. The first example illustrates Theorem 2.4. It provides a way on how to construct UN-rings issued from pullbacks where T is not necessarily a UN-ring.

Example 4.1. Let $\mathbb{Z}$ be the ring of integers. For a positive integer n, let $\mathbb{Z}_n$ be the ring on integers modulo n and X an indeterminate over $\mathbb{Z}_n$. Set $T = \mathbb{Z}_8[X]$, $I = 4\mathbb{Z}_8[X]$, $D = \mathbb{Z}_4$ and let R be the pullback arising from the diagram:

$$(\Delta) \quad \begin{array}{ccc} R := \phi^{-1}(\mathbb{Z}_4) & \longrightarrow & \mathbb{Z}_4 \\ \downarrow & & \downarrow \\ T = \mathbb{Z}_8[X] & \stackrel{\phi}{\longrightarrow} & \mathbb{Z}_4[X] = T/I. \end{array}$$

Clearly D is a UN-ring and $I^2 = 0$. Then, by Theorem 2.4, R is a UN-ring. However, T is not a UN-ring.

Example 4.2. The following example illustrates Corollary 2.5. Let $\mathbb{K} \subsetneq \mathbb{F}$ be an extension of fields, X an indeterminate over $\mathbb{F}$, $T = \mathbb{F}[X]/(X^2)$ and $M = X/(X^2)$. Then T is a local UN-ring (Proposition 2.1) with maximal ideal M. Let R be the pullback arising from the diagram:

$$(\Box) \quad \begin{array}{ccc} R := \phi^{-1}(\mathbb{K}) & \longrightarrow & \mathbb{K} \\ \downarrow & & \downarrow \\ \mathbb{F}[X]/(X^2) & \stackrel{\phi}{\longrightarrow} & T/M = \mathbb{F}. \end{array}$$

By Corollary 2.5, R is a UN-ring.

Example 4.3. The following example illustrates Corollary 2.6 Let A be a UN-ring, X an indeterminate over A, $T = A[[X]]/(X^3)$ and $I = X^2/(X^3)$. Then T is a local UN-ring (Proposition 2.1) with maximal ideal $M = X/(X^3)$, and clearly I is M-primary and $T/I = A[[X]]/(X^2)$. Now, let x be the isomorphy class of X modulo X^2 and R be the pullback arising from the diagram:

$$(\Delta) \quad \begin{array}{ccc} R := \phi^{-1}(A) & \longrightarrow & A \\ \downarrow & & \downarrow \\ A[[X]]/(X^3) & \stackrel{\phi}{\longrightarrow} & T/I = A[[X]]/(X^2) = A + Ax. \end{array}$$

By Corollary 2.6, R is a UN-ring.

Example 4.4. The following example illustrates Corollary 2.7. Let $T = \mathbb{Z}_4 \times \mathbb{Z}_4$. Then T is semi-local with maximal ideals $M = 2\mathbb{Z}_4 \times \mathbb{Z}_4$ and $N = \mathbb{Z}_4 \times 2\mathbb{Z}_4$ and $J(T) = M \cap N = 2\mathbb{Z}_4 \times 2\mathbb{Z}_4 = Nil(T)$. Clearly T is not a UN-ring and $T/J(T) = T/M \times T/N = \mathbb{Z}_2 \times \mathbb{Z}_2$. Let $D = \mathbb{Z}_2$ and $\iota : D \longrightarrow \mathbb{Z}_2 \times \mathbb{Z}_2$ defined by $\iota(x) = (x, x)$. and let R be the pullback arising from the diagram:

$$(\Delta) \quad \begin{array}{ccc} R & \longrightarrow & \mathbb{Z}_2 \\ \downarrow & & \downarrow \\ T = \mathbb{Z}_4 \times \mathbb{Z}_4 & \xrightarrow{\phi} & \mathbb{Z}_2 \times \mathbb{Z}_2. \end{array}$$

Since D is a field and $Nil(R) = Nil(T) \cap R = J(T) \cap R = J(T) = I$, by Corollary 2.7, R is a UN-ring.

Example 4.5. The following example illustrates Theorem 3.2. Let $A = \mathbb{Z}_4$, $B = \mathbb{Z}_2[[X]]/(X^5)$, $J = X^2/(X^5)$ and $f : A \longrightarrow B$ the homomorphism defined by $f(0) = f(2) = 0$ and $f(1) = f(3) = 1$. Clearly $J^2 \neq (0)$ but $J^3 = (0)$, so that $J \subseteq Nil(B) = X/(X^5)$. Now, let x denotes the class of X modulo X^5. Then every element of J is of the form $b_2x^2 + b_3x^3 + b_4x^4$ where $b_i \in \mathbb{Z}_2$ for $i = 2, 3, 4$. So for every $u \in U(A) = \{1, 3\}$ and every $f(x) = b_2x^2 + b_3x^3 + b_4x^4 \in J$, $f(u) + f(x) = 1 + b_2x^2 + b_3x^3 + b_4x^4 \in U(f(A) + J)$ (in fact it is easy to check that $(1 + f(x))^{-1} = 1 - b_2x^2 - b_3x^3 + (b_2^2 - b_4)x^4$). By Proposition 2.1, $A = \mathbb{Z}_{2^2}$ is a UN-ring. Thus, by Theorem 3.2, R is a UN-ring.

Acknowledgment. The author would like to express his sincere thanks to the referee for his/her helpful suggestions and comments, which have greatly improved this paper.

References

[1] M-S. Ahn and D. D. Anderson, Weakly clean rings and almost clean rings, Rocky Mountain J. 36 (3)(2006), 783–798.

[2] D. F. Anderson, A. Bouvier, D. E. Dobbs, M. Fontana, and S. Kabbaj, On Jaffard domains, Expo. Math. 6 (1988), 145–175.

[3] D. F. Anderson and D. E. Dobbs, Pairs of rings with the same prime ideals, Canad. J. Math. 32 (1980), 362–384.

[4] E. Bastida and R. Gilmer, Overrings and divisorial ideals of rings of the form D+M, Michigan Math. J. 20 (1992), 79–95.

[5] J.W. Brewer and E.A. Rutter, $D + M$ constructions with general overrings, Michigan Math. J. 23 (1976), 33–42.

[6] M. Boisen and P. B. Sheldon, CPI-extension: Overrings of integral domains with special prime spectrum, Canad. J. Math. 29 (1977), 722–737.

[7] G. Calugareanu, UN-rings, J. Algebra Appl. 15 (2016), no. 10, 1650182, 9 pp. https://doi.org/10.1142/S0219498816501826.

[8] V. P. Camillo and H-P. Yu, Exchange rings, units and idempotents, Comm. Algebra 22 (1994), 4737–4749.

[9] M. D'Anna, A construction of Gorenstein rings, J. Algebra 306 (2006), 507–519.

[10] M. D'Anna, C. A. Finocchiaro and M. Fontana, Amalgamated algebras along an ideal, Commutative algebra and its applications, Walter de Gruyter, Berlin, (2009), 241–252.

[11] M. D'Anna and M. Fontana, The amalgamated duplication of a ring along a multiplicative-canonical ideal, Ark. Mat. 45 (2007), 241–252.

[12] M. D'Anna and M. Fontana, An amalgamated duplication of a ring along an ideal, the basic properties, J. Algebra Appl., 6 (2007), 443–459.

[13] G. Ehrlich, Unit-regular rings, Portugal. Math. 27 (1968), 209–212.

[14] G. Ehrlich, Units and one-side units in regular rings, Trans. Amer. Math. Soc. 216 (1976), 81–90.

[15] A. El Khalfi, H. Kim and N. Mahdou, Amalgamation extension in commutative ring theory: a survey, Moroccan Journal of Algebra and Geometry with Applications 1 (1), (2022), 139–182.

[16] M. Fontana, Topologically defined classes of commutative rings, Ann. Mat. Pura Appl. 123 (1980) 331–355.

[17] M. Fontana and S. Gabelli, On the class group and the local class group of a pullback, J. Algebra 181 (3) (1996), 803–835.

[18] S. Gabelli and E. Houston, Coherentlike conditions in pullbacks, Michigan Math. J. 44 (1997), 99–123.

[19] K. Jangra and D. Udar, Un rings and group rings, Bull. Korean Math. Soc. 60 (2023), No. 1, pp. 83âĂŞ-91.

[20] W.W. McGovern, Clean semiprime f-rings with bounded inversion, Comm. Algebra 31 (2003), 3295–3304.

[21] M. Nagata, Local rings, Interscience Tracts in Pure and Applied Mathematics, No. 13, Interscience Publishers, New York-London, 1962.

[22] W. K. Nicholson, Lifting idempotents and exchange rings, Trans. Amer. Math. Soc. 229 (1977), 269–278.

[23] W. K. Nicholson, Strongly clean rings and FittingâĂŹs lemma, Comm. Algebra 27 (1999), 3583–3592.

[24] D. Udar, Rings whose nonunits are multiple of unit and strongly nilpotent element, Mathematical Modeling, Computational Intelligence Techniques and Renewable Energy, Springer, 25–30, 2022. https://doi.org/10.1007/978-981-16-5952-2_3.

[25] P. Vamos, On rings whose nonunits are a unit multiple of a nilpotent, J. Algebra Appl. 16 (2017), no. 7, 1750140, 13 pp. https://doi.org/10.1142/S0219498817501407

[26] P. Vamos, 2-good rings, Quart. J. Math. 56 (2005), 417–430.

[27] K. G. Wolfson, An ideal-theoretic characterization of the ring of all linear transformations, Amer. J. Math. 75 (1953), 358âĂŞ-386.

[28] B. Zabavsky, O. Domsha and O. Romaniv, Clear rings and clear elements, Mat. Stud. 55 (1) (2021), 3–9.

[29] D. Zelinsky, Every linear transformation is a sum of nonsingular ones, Proc Amer. Math. Soc. 5 (1954), 627–630.

[30] Y. Zhou, A note on UN-rings, Bull. Math. Soc. Sci. Math. Roumanie (N.S.) 64 (112) (2021), no. 1, 97–100.

Moroccan Journal of Algebra and Geometry with Applications
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco

Volume 2, Issue 1 (2023), pp 162-173

ISSN: 2820-7114

Title :

On *1*-absorbing prime submodules

Author(s):

Emel Aslankarayigit Ugurlu

Moroccan Journal of Algebra and Geometry with Applications
Vol. 2(1) (2023), 162–173
Supported by Sidi Mohamed Ben Abdellah University, Fez, Morocco ISSN: 2820-7114

On 1-absorbing prime submodules

Emel Aslankarayigit Ugurlu
Department of Mathematics, Marmara University, 34722, Istanbul, Turkey.
e-mail: *emel.aslankarayigit@marmara.edu.tr*

Communicated by Najib Mahdou
(Received 06 February 2023, Revised 12 April 2023, Accepted 17 April 2023)

Abstract. Our aim for this paper is to present the notion of *1-absorbing prime submodules* of A-module M. We display that the new notion is a generalization of prime submodules coinciding it is a sort of specialized 2-absorbing submodule. Along with some properties of them, we characterize the quasi local rings by the help of the new concept. Also, we investigate their behaviors under homomorphisms, in the localization of modules, and in a cartesian product of modules. After introducing the minimal 1-absorbing prime submodules, the radical$_1$ of ideals and submodules, we obtain some famous results for them. Furthermore, we obtain two characterizations of the concept in a multiplication module. Finally, we obtain a result for 1-absorbing prime submodules similar to the Prime Avoidance Theorem.
Key Words: prime ideal, prime submodule, 1-absorbing prime, prime avoidence theorem.
2010 MSC: Primary 13A15; Secondary 16P40, 16D60

1 Introduction

In abstract algebra, there are a great many publications addressing the structure of rings and modules, see [2, 4, 14, 15, 17, 18]. In this article, we focus on only commutative rings with a non-zero identity and non-zero unital modules. Let A always denote such a ring and let M denote such a A-module. The concept of prime ideals and its generalizations have a significant position in commutative algebra since they are used in understanding the structure of rings. Note that throughout this paper L (resp., Q) denotes a proper submodule (resp., ideal) of M (resp., A). Recall that Q is called a *prime ideal* if $ab \in Q$ yields $a \in Q$ or $b \in Q$, [3]. Some authors expanded the concept of prime ideals to modules, [8, 11, 13]. Also, L is called *prime* if whenever $ax \in L$ for any $a \in A, x \in M$, then $x \in L$ or $aM \subseteq L$, [13].

$Id(A)$ and $S(M)$ denote the lattice of all ideals of A and the lattice of all submodules in M, respectively. The *radical of* Q, denoted by $\sqrt{Q}$, is defined as the intersection of all prime ideals contain Q. Note that we have the equality $\sqrt{Q} = \{r \in A \mid r^k \in Q \text{ for some } k \in \mathbb{N}\}$, see [3]. For any $a \in A$, the principal ideal generated by a is denoted by (a). All unit elements of A is denoted by $U(A)$. For any element $m \in M$, the set $<m>= Am = \{rm : \forall r \in A\}$ is the cyclic submodule of M generated by m. If $M =< X >$, we say that M is a finitely generated A-module for any finite subset X of M. Note that the ideal $\{a \in A : aM \subseteq L\}$ is said to be the *residue* of L by M and we denote as $(L :_A M)$. If A is clear, it is written by only $(L : M)$. Especially, $Ann(M) := (0 :_A M)$ is said to be the *annihilator in* M. Whenever $Ann(M) = 0_A$, it is said to be a *faithful module*. Then for an element $m \in M$, the *annihilator of* m is defined as $Ann(m) := \{r \in A : rm = 0_M\}$ and it is an ideal of A. Moreover, we denote the *radical of* L as $rad(L)$.

In 2007, Badawi has introduced the concept of 2-absorbing ideals as a generalization of prime ideals: Q is called a *2-absorbing ideal* if whenever $r, s, t \in A$ and $rst \in Q$, then $rs \in Q$ or $rt \in Q$ or $st \in Q$, see [5]. Then in 2011, A. Y. Darani and F. Soheilnia defined the concept of 2-absorbing submodules as following: L is called *2-absorbing* if whenever $r, s \in A$; $x \in M$ with $rsx \in L$, either $rx \in L$ or $sx \in L$ or $rs \in (L : M)$, see [7]. Actually, the concept of 2-absorbing submodules is a generalization of prime submodules.

In 2021, Yassine, Nikmehr and Nikandish introduced a recent class of ideals that it is a class of ideals between 2-absorbing ideals and prime ideals: Q is called *1-absorbing prime* whenever for all non-units $r,s,t \in A$ with $rst \in Q$, $rs \in Q$ or $t \in Q$, see [19]. Afterwards, the authors introduced the notion of weakly 1-absorbing prime: whenever for all non-units $r,s,t \in A$ with $0 \neq rst \in Q$, $rs \in Q$ or $t \in Q$, see [10]. Note that every prime ideal is a 1-absorbing prime and every 1-absorbing prime ideal is a 2-absorbing ideal. Thus, we have a chain: prime ideals $\Rightarrow$ 1-absorbing prime ideals $\Rightarrow$ 2 absorbing ideals. On the other hand, we have a second chain: prime submodules $\Rightarrow$ 2 absorbing submodules. Thus we realize that there is a missing part in the second chain, which it is between prime submodules and 2-absorbing submodules. Then we define the missing part of the chain as *1-absorbing prime submodules.*

In Section 2, after introducing the notion of *1-absorbing prime submodules*, we examine the main properties of the new class. For all non-unit elements $r,s \in A$ and $x \in M$, if $rsx \in L$, either $rs \in (L:M)$ or $x \in L$, then L is said to be *1-absorbing prime*, see Definition 2.1. Firstly, we investigate in Proposition 2.3, the relation between 1-absorbing prime submodules and other special submodules, for example prime submodules, 2-absorbing submodules. We prove that every prime submodule is a 1-absorbing prime submodule, but the converse is not true: To see this, consider the cyclic submodule of $\mathbb{Z}_4$-module $\mathbb{Z}_4[X]$ generated by X, that is, $< X >$. Indeed, it is a 1-absorbing prime submodule, but is not a prime submodule of $\mathbb{Z}_4$-module $\mathbb{Z}_4[X]$, see Example 2.4. Also, we show that every 1-absorbing prime submodule is a 2-absorbing submodule. However, it is not true that every 2-absorbing submodule is 1-absorbing prime. Consider the cyclic submodule of $\mathbb{Z}$-module $\mathbb{Z}_{30}$ generated by $\overline{6}$, it is 2-absorbing but not 1-absorbing prime, see Example 2.5. Actually, by the help of Proposition 2.3, the second chain is completed. For the completed picture of these algebraic structures, see Figure 1. Among other results in this section, we give a characterization of 1-absorbing prime submodules, see Theorem 2.8. In Theorem 2.9, we characterize the quasi local rings by the help of our new concept. Furthermore, we examine the attitudes of the concept in a cartesian product of modules, in the localization of modules, and under homomorphisms. In Section 3, we introduce the *minimal 1-absorbing prime submodules, the radical*$_1$ of ideals and submodules. In Theorem 3.2 and Corollary 3.3, we obtain some famous results for 1-absorbing prime submodules. Afterwards, the Section 4 aims to give two characterizations of the concept in multiplication modules. By the help of some main proven results, we obtain two characterizations, see Theorem 4.1 and Theorem 4.10. Finally, the last section is dedicated to the Prime Avoidance Theorem for 1-absorbing prime submodules. After proving some propositions, we obtain 1-Prime Avoidance Theorem for submodules and 1-Prime Avoidance Theorem for cosets, see Theorem 5.2 and Theorem 5.6, respectively.

2 Properties of 1-absorbing prime submodules

Definition 2.1. For all non-units element $r,s \in A$ and $x \in M$, if $rsx \in L$, either $rs \in (L:M)$ or $x \in L$, then L is called **1-absorbing prime**.

Example 2.2. Assume $(A,\mathfrak{X})$ is a local ring with $\mathfrak{X}^2 = (0_A)$ and M is a A-module. Then every proper submodule in M is 1-absorbing prime. To see this, choose non-units $r,s \in A$ and $x \in M$ such that $rsx \in L$. Since $rs \in \mathfrak{X}^2 = (0_A)$, we have $rs \in (L:M)$, which implies L is 1-absorbing prime.

Proposition 2.3. *{Prime submodules} $\subseteq$ {1-absorbing prime submodules} $\subseteq$ {2-absorbing submodules}.*

Proof. Suppose L is a prime submodule of M. Take non-unit elements $r,s \in A$; $x \in M$ such that $rsx \in L$. Since L is prime, $rs \in (L:M)$ or $x \in L$, as desired. Suppose L is 1-absorbing prime. Take any $r,s \in A$ and $x \in M$ such that $rsx \in L$. Then we must obtain that $rs \in (L:M)$ or $rx \in L$ or $sx \in L$. If r,s are non-units, we have $rs \in (L:M)$ or $x \in L$, as required. Without loss generality, let r be unit. Thus one can see $rsx \in L$ yields $sx \in L$, as desired. □

Example 2.4. **(1-absorbing prime submodule that is not prime)** Consider the submodule $L =< X >$ of $\mathbb{Z}_4$-module $\mathbb{Z}_4[X]$. By previous example, L is a 1-absorbing prime submodule. However, L is not a prime submodule.

Example 2.5. **(2-absorbing submodule that is not 1-absorbing prime)** Let consider $\mathbb{Z}$-module $\mathbb{Z}_{30}$. Suppose that N is the cyclic submodule of $\mathbb{Z}$-module $\mathbb{Z}_{30}$ generated by $\overline{6}$, that is, $N =< \overline{6} >$. It is clear that $< \overline{6} >$ is a 2-absorbing submodule of $\mathbb{Z}$-module $\mathbb{Z}_{30}$, but $< \overline{6} >$ is not 1-absorbing prime. Indeed, $2 \cdot 2 \cdot \overline{3} \in < \overline{6} >$ but $\overline{4} \notin (N : \mathbb{Z}_{30})$ and $3 \notin < \overline{6} >$.

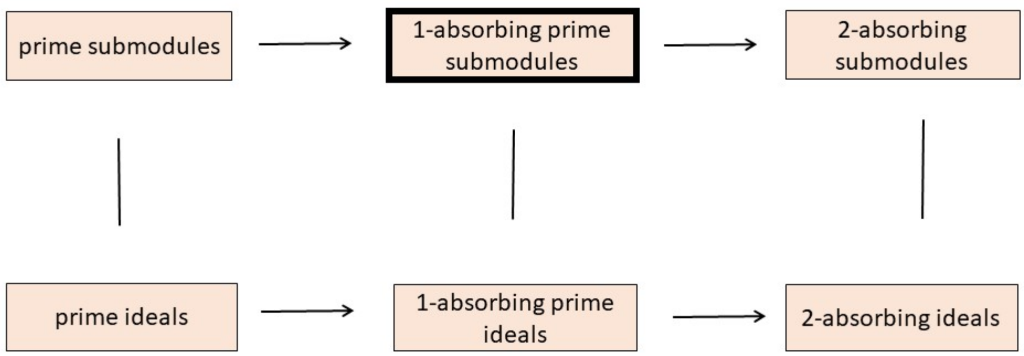


Figure 1: 1-absorbing prime submodules (ideals) vs other classical submodules (ideals)

Proposition 2.6. *Let L be a 1-absorbing prime submodule in M.*

1. *$(L : M)$ is a 1-absorbing prime ideal in A, hence $\sqrt{(L : M)}$ is prime.*

2. *$(L : x)$ is 1-absorbing prime, hence $\sqrt{(L : x)}$ is a prime ideal of A for every $x \in M \backslash L$.*

Proof. (1) Choose non-units $r, s, t \in A$ with $rst \in (L : M)$. For all $x \in M$ then $rstx \in L$. By our hypothesis, $rs \in (L : M)$ or $tx \in L$. This implies that $t \in (L : M)$ or $rs \in (L : M)$. Consequently, $(L : M)$ is 1-absorbing prime. Also, since $(L : M)$ is 1-absorbing prime, we conclude $\sqrt{(L : M)}$ is prime with the help of Theorem 2.3 in [19].
(2) Similar to (1). □

The next example displays that when $(L : M)$ is 1-absorbing prime, one can not say that L is 1-absorbing prime.

Example 2.7. Let $M = \mathbb{Z}\times\mathbb{Z}$ and $A = \mathbb{Z}$. Consider $L =< (3,0) >= \mathbb{Z}(3,0)$. Then it is clear that $(L : M) = (0)$. Then $(L : M)$ is a prime ideal of $\mathbb{Z}$, so 1-absorbing prime ideal in $\mathbb{Z}$. But $\mathbb{Z}(3,0)$ is not 1-absorbing prime. Indeed, choose $(1,0) \in \mathbb{Z}\times\mathbb{Z}$ and $3,2 \in \mathbb{Z}$, thus $3\cdot 2\cdot(1,0) \in \mathbb{Z}(3,0)$. However, $6 \notin (L : M) = (0)$ and $(1,0) \notin \mathbb{Z}(3,0)$. Thus $\mathbb{Z}(3,0)$ is not 1-absorbing prime.

Now, we give a characterization of the concept of 1-absorbing submodules of an A-module M.

Theorem 2.8. The items are equivalent:

1. L is a 1-absorbing prime submodule of M.
2. $(L : ab) \subseteq L$ for all non-units $a,b \in A$ such that $ab \notin (L : M)$.
3. For all non-units $a,b \in A$; a submodule K of M, $abK \subseteq L$ implies either $ab \in (L : M)$ or $K \subseteq L$.
4. If $IJK \subseteq L$, then $K \subseteq L$ or $IJ \subseteq (L : M)$ for any two proper ideals I,J and a proper submodule K of M.

Proof. (1) $\Rightarrow$ (2) Choose $x \in (L :_M ab)$, that is, $abx \in L$. By hypothesis, either $ab \in (L : M)$ or $x \in L$. The first one contradicts with our assumption, so that we conclude $(L :_M ab) \subseteq L$.

(2) $\Rightarrow$ (3) Let $a,b \in A$ be non-units and K be a submodule of M with $abK \subseteq L$, i.e., $K \subseteq (L :_M ab)$. Assume $ab \notin (L : M)$. By the item (2), we obtain $K \subseteq (L :_M ab) \subseteq L$, as needed.

(3) $\Rightarrow$ (4) Choose any two proper ideals I,J and a proper submodule K of M such that $IJK \subseteq L$. Suppose $IJ \nsubseteq (L : M)$. Then there are non-units $a,b \in A$ such that $a \in I$, $b \in J$ and $ab \in IJ\backslash(L : M)$. Also, $IJK \subseteq L$ implies that $abK \subseteq L$. Thus, we have $K \subseteq L$ by the item (3).

(4) $\Rightarrow$ (1) Choose non-unit $x,y \in A$ and $m \in M$ with $xym \in L$. Assume $m \notin L$. This means that $< m > \nsubseteq L$. Consider $I = (x), J = (y)$ and $K =< m >$. Since $IJK \subseteq L$ and $K \nsubseteq L$, by our hypothesis, $IJ \subseteq (L : M)$. Consequently, it means that $xy \in (L : M)$, as desired. □

Note that if A has exactly one maximal ideal, then A is called a *quasilocal ring*. In the following theorem, we prove a result on 1-absorbing prime submodules over quasilocal rings.

Theorem 2.9. If N is a 1-absorbing prime submodule in M which is not a prime submodule, then A is a quasilocal ring.

Proof. Assume that N is 1-absorbing prime that is not prime. Then there exist a non-unit $r \in A$; $m \in M$ which $rm \in N$ but $r \notin (N : M)$ and $m \notin N$. Choose a non-unit element $s \in A$. Hence we have that $rsm \in N$ and $m \notin N$. Because N is 1-absorbing prime, $rs \in (N : M)$. Let us take a unit element $u \in A$. We claim that $s + u$ is a unit element of A. To see this, assume $s + u$ is non-unit. Then $r(s + u)m \in N$. As N is 1-absorbing prime, $r(s + u) \in (N : M)$. This means that $ru \in (N : M)$, i.e., $r \in (N : M)$, which is a contradiction. Thus for any non-unit element s and unit element u in A, we have $s + u$ is a unit element. Similar to the proof of Theorem 2.4 in [19], we obtain A is a quasilocal ring. □

Corollary 2.10. *Assume M is a A-module, where A is not a quasilocal ring. L is 1-absorbing prime necessary and sufficient condition L is prime.*

Proof. It follows from previous theorem. □

Proposition 2.11. *Let $\{N_i\}_{i\in\Delta}$ be a chain of 1-absorbing prime submodules of A-module M. Then the followings hold:*

1. *$\bigcap_{i\in\Delta} N_i$ is 1-absorbing prime.*

2. *Assume that M is finitely generated. Then $\bigcup_{i\in\triangle} N_i$ is 1-absorbing prime.*

Proof. (1) Take non-unit $r,s \in A$ and $x \in M$ such that $rsx \in \bigcap_{i\in\triangle} N_i$. Assume that $x \notin \bigcap_{i\in\triangle} N_i$, so there exists $i \in\triangle$ such that $x \notin N_i$. Since N_i is 1-absorbing prime, we get $rs \in (N_i : M)$. For any $j \in\triangle$, we two cases.
Case 1: If $N_i \subseteq N_j$, then $(N_i : M) \subseteq (N_j : M)$, that is, $rs \in (N_j : M)$.
Case 2: If $N_j \subset N_i$, we obtain that $rs \in (N_j : M)$ since $x \notin N_j$ and N_j is 1-absorbing prime. As a consequence, we have $rs \in (\bigcap_{i\in\triangle} N_i : M)$.

(2) Since M is finitely generated, $\bigcup_{i\in\triangle} N_i$ is a proper submodule of M. Choose non-unit $r,s \in A$ and $x \in M$ such that $rsx \in \bigcup_{i\in\triangle} N_i$ and $x \notin \bigcup_{i\in\triangle} N_i$. Thus for $i \in\triangle$, $rsx \in N_i$ and $x \notin N_i$. This gives us $rs \in (N_i : M) \subseteq (\bigcup_{i\in\triangle} N_i : M)$, which completes the proof. □

Proposition 2.12. *Let $g : M \to M'$ be a homomorphism of A-module M and M'. Then the followings hold:*

1. *If L' is 1-absorbing prime in M' with $g^{-1}(L') \neq M$, $g^{-1}(L')$ is 1-absorbing prime in M.*

2. *Assume g is an epimorphism. If L is a 1-absorbing prime submodule of M with $Ker(g) \subseteq L$, $g(L)$ is a 1-absorbing prime submodule of M'.*

Proof. (1) Take non-units $r,s \in A$ and $x \in M$ such that $rsx \in g^{-1}(L')$. This means that $rsg(x) = g(rsx) \in L'$. Since L' is 1-absorbing prime, one can see $rs \in (L' : M')$ or $g(x) \in L'$. Then either $rs \in (g^{-1}(L') : M)$ or $x \in g^{-1}(L')$.

(2) Choose non-units $r,s \in A$ and $x' \in M'$ such that $rsx' \in g(L)$. By assumption there exists $x \in M$ such that $x' = g(x)$ and so $g(rsx) \in g(L)$. Then $rsx \in g^{-1}(g(L)) \subseteq L$, as $Ker(g) \subseteq L$. This implies that either $rs \in (L : M)$ or $x \in L$. If $rs \in (L : M)$, then $rsM \subseteq L$, that is, $rsg(M) = rsM' \subseteq g(L)$. Thus $rs \in (g(L) : M')$, it is done. If $x \in L$, then $x' = g(x) \in g(L)$, as required. □

One can easily obtain the following result by previous proposition.

Corollary 2.13. *Let $K \subset L$ be submodules of M. If L is a 1-absorbing prime submodule of M, then L/K is a 1-absorbing prime submodule of M/K.*

Theorem 2.14. Let $\emptyset \neq S \subseteq A$ be a multiplicatively closed subset and $S^{-1}L \neq S^{-1}M$. If L is 1-absorbing prime in M, $S^{-1}L$ is 1-absorbing prime in $S^{-1}A$-module $S^{-1}M$.

Proof. Choose two non-units $\frac{a}{x}, \frac{b}{y} \in S^{-1}A$ and $\frac{m}{z} \in S^{-1}M$ such that $\frac{a}{x}\frac{b}{y}\frac{m}{z} \in S^{-1}L$. Then there is $v \in S$ with $vabm \in L$. As L is 1-absorbing prime, one can see either $ab \in (L : M)$ or $vm \in L$. This result implies that either $\frac{ab}{xy} \in S^{-1}(L : M) \subseteq (S^{-1}L : S^{-1}M)$ or $\frac{vm}{vz} = \frac{m}{z} \in S^{-1}L$, which completes the proof. □

Theorem 2.15. Let $M_1 \times M_2$ be a module over $A_1 \times A_2$, where A_1 and A_2 are two commutative rings with nonzero identities. For two proper submodules L_1 of M_1 and L_2 of M_2, if $L_1 \times L_2$ is a 1-absorbing prime submodule in $M_1 \times M_2$, L_1 and L_2 are 1-absorbing prime.

Proof. Take two non-units $r,s \in A_1$; $x \in M_1$ with $rsx \in L_1$. Then consider $(r,0)(s,0)(x,0) \in L_1 \times L_2$. As $L_1 \times L_2$ is 1-absorbing prime, either $(rs,0) \in (L_1 \times L_2 : M_1 \times M_2)$ or $(x,0) \in L_1 \times L_2$. This implies that $rs \in (L_1 : M_1)$ or $x \in L_1$, that is, L_1 is 1-absorbing prime. Similarly, one can obtain L_2 is 1-absorbing prime. □

3 The radical$_1$ of ideals and submodules

Definition 3.1. Let P be a 1-absorbing prime submodule of M which $L \subsetneq P$. If there isn't a 1-absorbing prime P' such that $L \subseteq P' \subset P$, then P is called a **minimal 1-absorbing prime submodule of** L.

Theorem 3.2. If P is a 1-absorbing prime submodule of M with $L \subsetneq P$, there exists a minimal 1-absorbing prime submodule in L that it is contained in P.

Proof. Let define $\Lambda := \{P_i \in S(M) : P_i$ is a 1-absorbing prime submodule in M with $L \subseteq P_i \subsetneq P\}$. Since $L \subsetneq P$, we get $\Lambda \neq \emptyset$. Consider $(\Lambda, \supseteq)$. Let us take a chain $\{P_i\}_{i\in\triangle}$ in Λ. By Proposition 2.11(1), since $\bigcap_{i\in\triangle} P_i$ is a 1-absorbing prime, we can use Zorn's Lemma. Thus, there exists a maximal element $K \in \Lambda$. Then K is 1-absorbing prime and $L \subseteq K \subsetneq P$. Now, we shall prove K is minimal 1-absorbing prime. For the contrary, assume that there exists a 1-absorbing prime submodule K' which $L \subseteq K' \subsetneq K$. Then $K' \in \Lambda$ and $K \subsetneq K'$. This implies $K = K'$. It is done. □

Corollary 3.3. *For a proper submodule L in M, the statements hold:*

1. *Each 1-absorbing prime submodule contains at least one minimal 1-absorbing prime submodule of M.*
2. *Suppose that M is finitely generated. Every proper submodule of M has at least one minimal 1-absorbing prime submodule of M.*
3. *If M is finitely generated, then there exists a 1-absorbing prime submodule of M which contains L.*

Proof. (1) Obvious by Theorem 3.2.
(2) Let M be finitely generated. Then there exists a prime submodule P such that $L \subseteq P$, see [13]. Then P is 1-absorbing prime. Thus, by (1), it is done.
(3) By the claim in (2), it is clear. □

Definition 3.4. For any $I \in Id(A)$, we define

$$\Omega := \{I_i \in Id(A) : \ I_i \text{ is a 1-absorbing prime ideal with } I \subseteq I_i\}.$$

The intersection of all elements in Ω is called the **radical$_1$ of I**, and we denote it as

$$rad_1(I) := \bigcap_{I_i\in\Omega} I_i \text{ and if } \Omega = \emptyset \text{ or } I = A, \text{ we define } rad_1(I) := A.$$

Remark 3.5. One can easily see $rad_1(I) \subseteq \sqrt{I} = rad(I)$, since every prime ideal is a 1-absorbing prime ideal.

Definition 3.6. For any $N \in S(M)$, we define

$$\Omega := \{P_i \in S(M) : \ P_i \text{ is a 1-absorbing prime submodule such that } N \subseteq P_i\}.$$

Then the intersection of all elements in Ω is called the **radical$_1$ of N**, and we denote it as

$$rad_1(N) := \bigcap_{P_i\in\Omega} P_i \text{ and if } \Omega = \emptyset \text{ or } N = M, \text{ we define } rad_1(N) := M.$$

Remark 3.7. It is clear that $rad_1(N) \subseteq rad(N)$, since every prime submodule is a 1-absorbing prime submodule.

Proposition 3.8. *Let N, L be two submodules of M. Then the following statements hold:*

1. $N \subseteq rad_1(N)$.
2. $rad_1(rad_1(N)) \subseteq rad_1(N)$.
3. $rad_1(N \cap L) \subseteq rad_1(N) \cap rad_1(L)$.
4. $rad_1(IM) \subseteq rad_1(\sqrt{I}M)$.
5. $rad_1(L : M) \subseteq (rad_1(L) : M)$.

Proof. The first four items are elementary.

(5) If $rad_1(L) = M$, it is trivial. Let $rad_1(L) \neq M$. Then there is a 1-absorbing prime submodule K such that $L \subseteq K$. Thus $(K : M)$ is 1-absorbing prime with $(L : M) \subseteq (K : M)$. Hence $rad_1(L : M) \subseteq (K : M)$, that is, $rad_1(L : M)M \subseteq K$. The containment is held for all 1-absorbing prime submodules K_i with $L \subseteq K_i$. This implies that $rad_1(L : M)M \subseteq rad_1(L)$, i.e., $rad_1(L : M) \subseteq (rad_1(L) : M)$. □

Proposition 3.9. *Let M be finitely generated. Then $rad_1(L) = M$ if and only if $L = M$.*

Proof. Let $rad_1(L) = M$. Suppose that $L \neq M$. By Corollary 3.3(3), there exists a 1-absorbing prime submodule L' of M which $L \subseteq L'$. Hence, we conclude that $rad_1(L) = M \subseteq L'$, a contradiction. The other way of the claim is clear. □

Theorem 3.10. Let M be finitely generated and K, L be two submodules in M. Then $K + L = M$ if and only if $rad_1(K) + rad_1(L) = M$.

Proof. Assume that $K + L = M$. We know that $K \subseteq rad_1(K)$ and $L \subseteq rad_1(L)$. Thus $M = K + L \subseteq rad_1(K) + rad_1(L)$, it is done. Conversely, suppose that $K + L \neq M$. Since M is finitely generated, there is a maximal submodule T of M such that $K + L \subseteq T$, see [13]. Furthermore, T is prime (thus, 1-absorbing prime). As $rad_1(K) \subseteq T$ and $rad_1(L) \subseteq T$, we have $rad_1(K) + rad_1(L) \subseteq T$, that is, $M \subseteq T$. This contradicts with our assumption. Consequently, it must be $K + L = M$. □

4 Characterizations of 1-absorbing prime submodules of multiplication modules

Now, our aim is to give two characterizations of the new concept in multiplication modules. For the integrity of our study, we will remind some knowledge about multiplication modules. An A-module M is said to be *multiplication* if each submodule L of M has the form JM for an ideal J of A, see [9]. It is clear that one can write $L = JM \subseteq (L : M)M \subseteq L$. Then if M is multiplication, $L = (L : M)M$.

For the first characterization examine the following result:

Theorem 4.1. Assume M is a faithful finitely generated multiplication A-module. Then L is 1-absorbing prime in M if and only if for each submodules K_1, K_2, K_3 of M if $K_1K_2K_3 \subseteq L$, either $K_1K_2 \subseteq L$ or $K_3 \subseteq L$.

Proof. Let L be a 1-absorbing prime submodule of M. Suppose $K_1K_2K_3 \subseteq L$. Since M is a multiplication module, there are some ideals I_1, I_2, I_3 of A such that $K_1 = I_1M, K_2 = I_2M$, and $K_3 = I_3M$. Consider $I_1I_2I_3M \subseteq L$. By Theorem 2.8, it must be either $I_1I_2 \subseteq (L : M)$ or $I_3M \subseteq L$. This implies that $K_1K_2 \subseteq L$ or $K_3 \subseteq L$ because M is multiplication. For the converse, let I_1, I_2 be two ideals in A and choose a proper submodule K of M which $I_1I_2K \subseteq L$. As M is a multiplication module, there is an ideal I_3 of A such that $K = I_3M$. Then $I_1I_2I_3M \subseteq L$, by our hypothesis, $I_1I_2M \subseteq L$ or $I_3M \subseteq L$. The second option means $K \subseteq L$, it is done by Theorem 2.8. If $I_1 I_2 M \subseteq L = (L : M)M$, we conclude $I_1 I_2 \subseteq (L : M) + Ann(M)$ by Corollary in page 231 of [16]. Since M is faithful, $I_1 I_2 \subseteq (L : M)$, which completes the proof with Theorem 2.8. □

Recall from [9], a multiplication module can be represented by a maximal ideal. Assume P is a maximal ideal of A. To give the characterization, let us define the submodule $T_P(M) = \{x \in M :$ there is a $p \in P$ such that $(1-p)x = 0_M\}$. Whenever $T_P(M) = M$, M is said to be a *P-torsion module*. Also, if there is $p \in P$; $x \in M$ with $(1-p)M \subseteq Ax$, M is called a *P-cyclic module*. In Theorem 1.2 of [9], the authors proved that M is a multiplication A-module if and only if for any maximal ideal P of A, M is a P-cyclic or M is a P-torsion. For more information about multiplication modules, we refer the reader to [1], [6].

To obtain the second characterization, firstly we need the following results.

Theorem 4.2. Let M be a faithful multiplication A-module. Let J be a 1-absorbing prime ideal of A. Then $abx \in JM$ implies $ab \in J$ or $x \in JM$ for all non-units $a, b \in A$; $x \in M$.

Proof. Take $x \in M$; non-units $a, b \in A$ with $abx \in JM$. Suppose $ab \notin J$. Let us define $J' := \{r \in A : rx \in JM\}$. In case $J' = A$, there is nothing to prove. So, $J' \neq A$. Then there is a maximal ideal P of A, which $J' \subseteq P$. Now, we will prove $x \notin T_P(M)$. If $x \in T_P(M)$, there is $p \in P$ with $(1-p)x = 0_M$. This yields $1 - p \in J' \subseteq P$, a contradiction. Hence, $T_P(M) \neq M$. As M is multiplication, M is P-cyclic by the help of Theorem 1.2 in [9]. So, there is $p' \in P$ and $x' \in M$ which $(1-p')M \subseteq Ax'$. Then $(1-p')x \in Ax'$, so that there exists $s \in A$ with $(1-p')x = sx'$. Then $(1-p')abx = sabx' \in JM$ and $(1-p')abx \in Ax'$. Thus, there are $a' \in J$ such that $(1-p')abx = a'x'$. Since $sabx' = a'x'$, we obtain $abs - a' \in Ann(x')$. Moreover, $(1-p')M \subseteq Ax'$ gives us $(1-p')Ann(x')M \subseteq AAnn(x')x' = 0_M$, i.e., $(1-p')Ann(x') \subseteq Ann(M)$. Then $(1-p')Ann(x') = 0_A$, because M is faithful. This implies $(1-p')(abs - a') = 0_A$. Hence, one can see $abs(1-p') = a'(1-p') \in J$. Then $abs(1-p') \in J$. Now, there are two cases for $s \in A$:

Case 1: Assume s is unit. Then $ab(1-p') \in J$. If $1 - p'$ is unit, then $ab \in J$. This contradicts our assumption $ab \notin J$. Suppose $1 - p'$ is non-unit. As J is 1-absorbing prime, $ab \in J$ (which gives a contradiction) or $1 - p' \in J$. If $1 - p' \in J$, then we have $(1-p')x \in JM$, that is, $1 - p' \in J' \subseteq P$, a contradiction.

Case 2: Assume s is non-unit. Now, we have two possibilities for $1-p'$. If $1-p'$ is a unit element of A, then $sab \in J$. Since J is 1-absorbing prime, $ab \in J$ (again, it is not possible) or $s \in J$. Then $sx' \in JM$. Since $sx' = (1-p')x$, we have $(1-p')x \in JM$. So, $1 - p' \in J' \subseteq P$, which is not possible. If $1 - p'$ is non-unit, because J is 1-absorbing prime, either $abs \in J$ or $1 - p' \in J$. Again, since it is 1-absorbing prime, we have $ab \in J$ or $s \in J$ or $1 - p' \in J$. Every probability concludes a contradiction by the help of the previous explications.

Consequently, $J' = A$, i.e., $x \in JM$. □

Corollary 4.3. *Suppose M is a faithful multiplication A-module. Let J be an ideal of A such that $JM \neq M$. If J is a 1-absorbing prime ideal of A, then JM is 1-absorbing prime.*

Proof. Take non-unit elements $x, y \in A$ and $m \in M$ such that $xym \in JM$. Suppose $xy \notin (JM : M)$. Then $xy \notin J$. By Theorem 4.2, it must be $m \in JM$, as required. □

Theorem 4.4. Let M be a faithful finitely generated multiplication A-module. Let I be an ideal of A such that $IM \neq M$. Then IM is a 1-absorbing prime submodule of M if and only if I is a 1-absorbing prime of A.

Proof. Let IM be 1-absorbing prime. Take non-unit $x, y, z \in A$ such that $xyz \in I$. Suppose $xy \notin I$. By Theorem 10 in [16], we have $I = (IM : M)$. This implies that $xy \notin (IM : M)$ and $xyzM \subseteq IM$. Since IM is 1-absorbing prime, it must be either $xy \in (IM : M)$ or $zM \subseteq IM$. The first one contradicts with

$xy \notin I$. Thus, the second one implies $z \in (IM : M) = I$, as needed. The other way of the claim is obvious from Corollary 4.3. □

The proof of the next result is omitted since it is straightforward by Theorem 4.4.

Corollary 4.5. *Let M be a faithful finitely generated multiplication A-module and N be a proper submodule of M. Then*

1. *N is a 1-absorbing prime submodule of M if and only if $(N : M)$ is 1-absorbing prime ideal of A.*
2. *N is a 1-absorbing prime submodule of M if and only if $N = IM$ for some 1-absorbing prime ideal I of A.*

Proposition 4.6. *Let T, Q be some ideals in A with $T \subseteq Q$. If Q is 1-absorbing prime, Q/T is 1-absorbing prime.*

Proof. Choose non-unit $x+T, y+T, z+T$ in A/T such that $xyz+T \in Q/T$. This implies that $xyz \in Q$. Since $\{r+T : r \in U(A)\} \subseteq U(A/T)$, one can see x, y, z are non-units. As Q is 1-absorbing prime, either $xy \in Q$ or $z \in Q$. This means $xy+T \in Q/T$ or $z+T \in Q/T$. □

Definition 4.7. Let Q be an ideal of A. If the following equation $U(A/Q) = \{r+Q : r \in U(A)\}$ holds, then we say A satisfies the **good unit element property for** Q.

For the other way of Proposition 4.6, we need to the good unit element property:

Remark 4.8. (Corollary 2.17 in [19]) Let Q be an ideal of A such that $T \subseteq Q$ and $U(A/T) = \{r+T : r \in U(A)\}$. Then Q is 1-absorbing prime if and only if Q/T is a 1-absorbing prime ideal in A/T.

Proposition 4.9. *Let A satisfy the good unit element property for $Ann(M)$. When L is a 1-absorbing prime submodule in M over $A/Ann(M)$, L is a 1-absorbing prime submodule in M over A.*

Proof. Choose $x \in M$; non-unit $a, b \in A$ with $abx \in L$. We must show that either $ab \in (L :_A M)$ or $x \in L$. Consider $(a+Ann(M))(b+Ann(M))x = abx + Ann(M)x \in L$. By our hypothesis, one can say $a+Ann(M)$ and $b+Ann(M)$ are non-unit elements in $A/Ann(M)$. Since L is a 1-absorbing prime submodule over the ring $A/Ann(M)$, we obtain $ab+Ann(M) \in (L :_{A/Ann(M)} M)$ or $x \in L$. If the second one holds, it is done. The first one implies $abM \subseteq L$, i.e., $ab \in (L :_A M)$, as required. □

As a final result in this section, we give the second characterization of 1-absorbing prime submodules of multiplication modules in the next theorem.

Theorem 4.10. Let A satisfy the good unit element property for $Ann(M)$, where M is a multiplication A-module. Then the followings are equivalent:

1. L is a 1-absorbing prime submodule of M.
2. $(L : M)$ is a 1-absorbing prime ideal of A.
3. For a proper ideal P of A such that $Ann(M) \subseteq P$ and P is 1-absorbing prime, then $L = PM$.

Proof. $(1) \Rightarrow (2)$ By Proposition 2.6.

$(2) \Rightarrow (3)$ Consider $P = (L : M)$.

$(3) \Rightarrow (1)$ By page 759 of [9], as M is a multiplication A-module, it is also a faithful multiplication $A/Ann(M)$-module. Moreover, because P is a 1-absorbing prime ideal in A, $P/Ann(M)$ is a 1-absorbing prime ideal in $A/Ann(M)$ by Proposition 4.6. Then $[P/Ann(M)]M$ is a 1-absorbing prime submodule in $A/Ann(M)$-module M with the help of Corollary 4.3. Then Proposition 4.9 implies that $[P/Ann(M)]M$ is 1-absorbing prime. As $[P/Ann(M)]M = L$, it is done. □

5 The 1-absorbing Prime Avoidance Theorem

Our aim for the part is to demonstrate the Prime Avoidance Theorem for 1-absorbing prime submodules of M. Firstly, we need the following proposition.

Now, recall from [12], a covering $N \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$ is efficient if N is not contained in the union of any $(n-1)$ of the submodules $N_1, N_2, \ldots, N_n$. Similary, $N = N_1 \cup N_2 \cup \cdots \cup N_n$ is said to be an efficient union if none of the N_i may be excluded, where $i = 1, 2, \ldots, n$. Also, note that a covering consists of two submodules can not be efficient.

Proposition 5.1. *Let $N, N_1, N_2, \ldots, N_n$ be submodules of a A-module M such that $N \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$ is an efficient covering ($n > 2$). If $\sqrt{(N_j : M)} \not\subseteq \sqrt{(N_k : m)}$ for $\forall\ m \notin N_k$ and $\forall\ j \neq k$, then no N_k is a 1-absorbing prime submodule of M, where $k = 1, 2, \ldots, n$.*

Proof. It is clear that $N = (N_1 \cap N) \cup (N_2 \cap N) \cup \cdots \cup (N_n \cap N)$ is an efficient union. Hence, for every $k \leq n$, there is $e_k \in N - N_k$. Note that $\bigcap_{j \neq k}^{n} N_j \cap N \subseteq N \cap N_k$ by Lemma 2.1 in [12]. Now, without losing the generality assume that N_1 is 1-absorbing prime. It must be $\sqrt{(N_j : M)} \not\subseteq \sqrt{(N_1 : m)}$ for $\forall\ m \notin N_1$ and $\forall\ j = 2, 3, \ldots, n$, by our hypothesis. Thus there is $s_j \in \sqrt{(N_j : M)}$ and $s_j \notin \sqrt{(N_1 : m)}$. This implies that there exists $n_j \in \mathbb{N}$ such that $s_j^{n_j} \in (N_j : M)$. Let $\beta = max\ \{n_j\}_{j=2,3,\ldots,n}$. Consider $s = (s_2 s_3 \cdots s_n)^{\beta} \in (N_j : M)$. Then clearly $se_1 \in sM \subseteq N_j$ for $\forall\ j = 2, 3, \ldots, n$. This implies that $se_1 \in \bigcap_{j=2}^{n} N \cap N_j$. Here, note that since $e_1 \notin N_1$ and N_1 is 1-absorbing prime, by Proposition 2.6(2), $(N_1 : e_1)$ is a 1-absorbing prime ideal of R. Moreover, $\sqrt{(N_1 : e_1)}$ is a prime ideal by Theorem 2.3 in [19]. Now, we claim that $se_1 \notin N_1$. Indeed, if $s \in (N_1 : e_1)$, we would have $s_2 s_3 \cdots s_n \in \sqrt{(N_1 : e_1)}$, this gives us $s_j \in \sqrt{(N_1 : e_1)}$ for some j. Since $s_j \notin \sqrt{(N_1 : m)}$ for $\forall\ m \notin N_1$ and $e_1 \notin N_1$, we would obtain a contradiction. Hence, we conclude $se_1 \in (\bigcap_{j=2}^{n} N \cap N_j) - N \cap N_1$, a contradiction. Consequently, N_1 is not 1-absorbing prime. □

Theorem 5.2. **(1-absorbing Prime Avoidance Theorem for Submodules)** Let $N_1, N_2, \ldots, N_n$ be a finite number of submodules of a A-module M and N be a submodule in M which $N \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$. Suppose that at most two of the N_i's are not 1-absorbing prime and $\sqrt{(N_j : M)} \not\subseteq \sqrt{(N_k : m)}$ for $\forall$ $m \notin N_k$ and $\forall\ j \neq k$. Then $N \subseteq N_k$ for some $k = 1, 2, \ldots, n$.

Proof. By using the containment $N \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$, we can find $N \subseteq N_{j_1} \cup N_{j_2} \cup \cdots \cup N_{j_t}$, which is an efficient covering. Then $1 \leq t \leq n$ and $t \neq 2$. If $t > 2$, there is at least one L_{j_i}, which is 1-absorbing prime. On the other hand, by the help of Proposition 5.1, we conclude a contradiction with $\sqrt{(N_j : M)} \not\subseteq \sqrt{(N_k : m)}$ for $\forall\ m \notin N_k$ and $\forall\ j \neq k$. Thus $t = 1$, that is, $N \subseteq N_k$ for some $k = 1, 2, \ldots, n$. □

As a final conclusion of our study, we will present "1-absorbing Prime Avoidance Theorem for Cosets". For this reason, we need the followings.

Let $N, N_1, N_2, \ldots, N_n$ be submodules of a A-module M and $N_1 + m_1, N_2 + m_2, \ldots, N_n + m_n$ be cosets in M. Then a covering $N \subseteq (N_1 + m_1) \cup (N_2 + m_2) \cup \cdots \cup (N_n + m_n)$ is said to be efficient if N is not contained in the union of any $(n-1)$ of the cosets, see [12].

Remark 5.3. Consider the above efficient covering. If $m_j = m$ for every $j \in \{1, 2, \ldots, n\}$, then the covering equals to $N - m \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$. Thus, $N - m$ is a coset efficiently covered by a union of submodules, see [12].

In order not to lose the entireness of the article, let us notice the following:

Lemma 5.4. *(Lemma 2.4 in [12]) Let $N \subseteq (N_1 + m_1) \cup (N_2 + m_2) \cup \cdots \cup (N_n + m_n)$ be an efficient covering of a submodule of N by cosets, where $n \geq 2$. Then $N \cap (\bigcap_{j \neq k}^{n} N_j) \subseteq N_k$ and $N \nsubseteq N_k$ for all k.*

Proposition 5.5. *Let $N + m \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$ be an efficient covering for $m \in M$ with $n \geq 2$. If $\sqrt{(N_j : M)} \nsubseteq \sqrt{(N_k : M)}$ for $\forall\, j \neq k$. Then no N_k is 1-absorbing prime in M, where $k = 1, 2, \ldots, n$.*

Proof. Assume $N + m \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$ is an efficient covering for $m \in M$ with $n \geq 2$. Then by Remark 5.3, we can apply Lemma 5.4. Thus, we conclude that $N \cap (\bigcap_{j \neq k}^{n} N_j) \subseteq N_k$ and $N \nsubseteq N_k$ for all k. Without losing the generality, let $k = 1$. For the contradictory, suppose N_1 is 1-absorbing prime. Consider the ideal $I = (\bigcap_{j=2}^{n} N_j : M)$. This implies that $I^2 N \subseteq IN \subseteq N \cap (\bigcap_{j=2}^{n} N_j) \subseteq N_1$. Since N_1 is 1-absorbing prime, either $I^2 \subseteq (N_1 : M)$ or $N \subseteq N_1$ by Theorem 2.8. The second one gives us a contradiction. Assume $I^2 \subseteq (N_1 : M)$. Then $\sqrt{I} = \sqrt{I^2} \subseteq \sqrt{(N_1 : M)}$. As $\sqrt{(\bigcap_{j=2}^{n} N_j : M)} = \sqrt{\bigcap_{j=2}^{n} (N_j : M)} = \bigcap_{j=2}^{n} \sqrt{(N_j : M)}$, then $\bigcap_{j=2}^{n} \sqrt{(N_j : M)} \subseteq \sqrt{(N_1 : M)}$. Note that $\sqrt{(N_1 : M)}$ is a prime ideal by Proposition 2.6(1). This result gives us $\sqrt{(N_j : M)} \subseteq \sqrt{(N_1 : M)}$ for some j, which contradicts with our assumption. Consequently, N_1 is not a 1-absorbing prime submodule. □

Finally, by the help of Remark 5.3, we can express the following result.

Theorem 5.6. **(1-absorbing Prime Avoidance Theorem for Cosets)** Let $N + m \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$ be a covering for $m \in M$. Suppose that at most one submodule N_i is not 1-absorbing prime. If $\sqrt{(N_j : M)} \nsubseteq \sqrt{(N_k : M)}$ for $\forall\, j \neq k$, then the submodule $N + \langle m \rangle \subseteq N_k$ for some $k = 1, 2, \ldots, n$.

Proof. By using the covering $N + m \subseteq N_1 \cup N_2 \cup \cdots \cup N_n$, we can find an efficient covering, $N + m \subseteq N_{j_1} \cup N_{j_2} \cup \cdots \cup N_{j_t}$. Then $1 \leq t \leq n$. It follows from Proposition 5.5 that $t = 1$. Thus, we conclude $N + m \subseteq N_k$ for some $k = 1, 2, \ldots, n$. It is clear that $N + \langle m \rangle \subseteq N_k$ since $m \in N + m \subseteq N_k$. □

References

[1] D. D. Anderson, T. Arabaci, U. Tekir and S. Koc, *On S-multiplication modules*, Comm. Algebra 48(8), 3398–3407, 2020.

[2] D. F. Anderson and A. Badawi, *On (m,n)-closed ideals of commutative rings*, J. Algebra Appl. 16(1), 1750013(1–21), 2017.

[3] M. F. Atiyah and I. G. MacDonald, *Introduction to commutative algebra*, CRC Press, 1969.

[4] A. Badawi, M. Issoual and N. Mahdou, *On n-absorbing ideals and (m,n)-closed ideals in trivial ring extensions of commutative rings*, J. Algebra Appl. 18(7), 1950123(1–19), 2019.

[5] A. Badawi, *On 2-absorbing ideals of commutative rings*, Bull. Austral. Math. Soc. 75, 417–429, 2007.

[6] A. Barnard, *Multiplication modules*, J. Algebra 71, 174–178, 1981.

[7] A. Y. Darani and F. Soheilnia, *On 2-absorbing and weakly 2-absorbing submodules*, Thai J. Math. **9**, 577–584, 2011.

[8] J. Dauns, *Prime modules*, J. reine Angew. Math. 2, 156–181, 1978.

[9] Z. A. El-Bast and P. F. Smith, *Multiplication modules*, Comm. Algebra 16(4), 755–779, 1988.

[10] S. Koc, U. Tekir and E. Yildiz, *On weakly 1-absorbing prime ideals*, Ricerche mat 1–16, 2021.

[11] C. P. Lu, *Prime submodules of modules*, Comm. Math. Univ. Sancti Pauli. 33, 61–69, 1984.

[12] C. P. Lu, *Unions of prime submodules*, Houston J. Math. 23(2), 203–213, 1997.

[13] R. L. McCasland and M. E. Moore, *Prime submodules*, Comm. Algebra 20, 1803–1817, 1992.

[14] H. Mostafanasab, U. Tekir, E. Y. Celikel, E. A. Ugurlu, G. Ulucak and A. Y. Darani, *Generalizations of 2-absorbing and 2-absorbing primary submodules* Hacet. J. Math. Stat. 48(4), 1001–1016, 2019.

[15] E. S. Sevim, U. Tekir and S. Koc, *S-Artinian rings and finitely S-cogenerated rings*, J. Algebra Appl. 19(3), 2050051(1–16), 2020.

[16] P. F. Smith, *Some remarks on Multiplication modules*, Arch Math. 50, 223–235, 1988.

[17] U. Tekir, *On the sheaf of modules*, Comm. Algebra 33(8), 2557–2562, 2005.

[18] G. Ulucak, U. Tekir and S. Koc, *On S-2-absorbing submodules and vn-regular modules*, An. St. Univ. Ovidius Constanta 28(2), 239–257, 2020.

[19] A. Yassine and M. J. Nikmehr, R. Nikandish, *On 1-absorbing prime ideals of commutative rings*, J. Algebra Appl. 20, 2150175(1–12), 2021.