Hoàng Xuân Sính's thesis: Categorifying Group Theory

John C. Baez (California, USA)

(Received Aug. 8, 2023)

Abstract. During what Vietnamese call the American War, Alexander Grothendieck spent three weeks teaching mathematics in and near Hanoi. Hoàng Xuân Sính took notes on his lectures and later did her thesis work with him by correspondence. In her thesis she developed the theory of 'Gr-categories', which are monoidal categories in which all objects and morphisms have inverses. Now often called '2-groups', these structures allow the study of symmetries that themselves have symmetries. After a brief account of how Hoàng Xuân Sính wrote her thesis, we explain some of its main results, and its context in the history of mathematics.

1. Introduction

The story of Hoàng Xuân Sính is remarkable because it combines dramatic historical events with revolutionary mathematics. Some mathematicians make exciting discoveries while living peaceful lives. Many have their work disrupted or prematurely cut off by wars and revolutions. But some manage to carry out profound research on the fiery background of history.

In war-torn Hanoi, Hoàng Xuân Sính met the visionary mathematician Alexander Grothendieck, who had visited to give a series of lectures—in part as a protest against American aggression. After he returned to France, she did her thesis with him by correspondence, writing it by hand under the light of

Key words and phrases: 2-group, Gr-category, Picard category, monoidal category, symmetric monoidal category, crossed module.
2020 Mathematics Subject Classification: 18G45

a kerosene lamp as the bombing of Hanoi reached its peak. In her thesis she established the most fundamental properties of a novel mathematical structure that takes the concept of symmetry and pushes it to new heights, making precise the concept of *symmetries of symmetries*. She then traveled to Paris to defend her thesis before some of the most illustrious mathematicians of the era. After returning to Hanoi, she set up the first private university in Vietnam! But our story here is about her thesis.

In 1951, Hoàng Xuân Sính completed her bachelor's degree in Hanoi. She then traveled to Paris for a second baccalaureate in mathematics. She stayed in France to study for the competitive examination for civil service at the University of Toulouse, which she completed in 1959. Then she returned to Vietnam and taught mathematics at the Hanoi National University of Education.

In late 1967, when the U.S. attacks on Vietnam were escalating, something surprising happened. Grothendieck visited Vietnam and spent a month teaching algebraic geometry to the Hanoi University mathematics department staff. Hoàng Xuân Sính took the notes for these lectures. Because of the war, Grothendieck's lectures were held away from Hanoi, first in the nearby countryside and later in Đại Từ.

Hoàng Xuân Sính asked Grothendieck to be her thesis supervisor, and he accepted. After he returned to France, she continued to work with him by correspondence. She later wrote [27]:

> If I remember correctly, he wrote to me twice and I wrote to him three times. The first time he wrote to me was to give me the subject of the thesis and the work plan; the second is to tell me to drop the problem of inverting objects if I can't do it. As for me, I think I wrote to him three times: the first was to tell Grothendieck that I couldn't invert objects because of the non-strict commutativity; the second is to tell him that I succeeded in inverting objects; and the third is to tell him that I have finished the job. The letters had to be very short because we were in time of war, eight months for a letter to arrive at its destination between France and Vietnam. When I finished my work in writing, I sent it to my brother, who lives in France, and he brought it to Grothendieck.

According to the website of Thang Long University, her two main impressions from her contacts with Grothendieck were these:

1. A good teacher is a teacher who turns something difficult into something easy.

2. We should always avoid anything that is fictitious, live in accordance to our own feelings and value simple people.

She finished her thesis in 1972. Around Christmas that year, the United States dropped over 20,000 tons of bombs on North Vietnam, mainly Hanoi. So, it is not surprising that she defended her thesis three years later, when the North had almost won. But she mentions another reason for the delay [45]:

> I was a doctorate student during wartime, bombs and bullets. Back then, I was teaching at Hanoi National University of Education; the school did not provided a way to take leave to study for a doctorate. I taught during the day and worked on my thesis during the night under the kerosene lamp light. I wrote in French under my distant teacher's guidance. When I got the approval from France to come over to defend, there were disagreeable talks about not letting me because they was afraid I wasn't coming back. The most supportive person during the time was Lady Hà Thị Quế—President of the Vietnamese Women Coalescent organization. Madame Quế was a guerilla, without the conditions to get much education, but gave very convincing reasons to support me. She said, firstly, I was 40 years old, it is very difficult to get a job abroad at 40 years old, and without a job, how can I live? Second, my child is at home, no woman would ever leave her child... so comrades, let's not be worried, let her go. I finished my thesis in 1972, and three years later with the help and struggle of the women's organization, I was able to travel over to defend in 1975....

After finishing her thesis, she went to France to defend it at Paris Diderot University (also called Paris VII). Her thesis committee included not only Alexander Grothendieck but also some other excellent mathematicians:

- Henri Cartan (a member of Bourbaki famous for his work on sheaf theory, algebraic topology, and potential theory),
- Laurent Schwartz (famous for developing the theory of distributions),
- Michel Zisman (best known for his work on the "calculus of fractions" in category theory),
- Jean-Louis Verdier (a former student of Grothendieck who generalized Poincaré duality to algebraic geometry).

Her thesis defense lasted two and a half hours. Soon thereafter she defended a *second* thesis, entitled *The Embedding of a One-dimensional Complex in a Two-dimensional Differential Manifold.*

But it is her first thesis that is our concern here. Its title was *Gr-catégories*. But what are Gr-categories, and why are they so important? I will give a quick answer here, and a more detailed one in the rest of the paper.

Throughout science and mathematics, symmetry plays a powerful simplifying role. This is often formalized using group theory. However, group theory is just the tip of a larger subject that could be called 'higher-dimensional group theory'. For example, in many contexts where we might be tempted to use groups, we can use a richer setup where in addition to symmetries of some object, there are higher levels of structure: symmetries of symmetries, symmetries of symmetries of symmetries, and so on.

For example, if we are studying the rotational symmetries of a sphere, we usually describe a rotation as a special sort of 3×3 matrix. But we can also consider a *continuous path* of such rotation matrices, starting at one rotation matrix and ending at another. Such a path may be considered as a symmetry going between symmetries.

We can continue this pattern to higher levels by considering continuous 1-parameter families of continuous paths of rotation matrices, which describe symmetries of symmetries of symmetries, and so on. But there is already a fascinating world to discover if we stop at the second level. This is what the concept of Gr-category formalizes. Gr-categories got their name because they blend the concepts of *group* and *category*: a Gr-category is a category that resembles a group. The objects of a Gr-category can be thought of as symmetries, and the morphisms as symmetries of symmetries.

In Section 2 we describe a key prerequisite for the theory of Gr-categories, namely 'monoidal categories', and a bit of the history of this concept, which was still fairly new when Hoàng Xuân Sính wrote her thesis. In Section 3 we explain Gr-categories and Hoàng Xuân Sính's fundamental theorem classifying these. Then we turn to examples of Gr-categories. In Section 4 we discuss 'strict' Gr-categories, where the associative law and unit law hold as equations rather than merely up to isomorphism. We explain how to get strict Gr-categories from 'crossed modules' and use this to give many examples of Gr-categories. In Section 5 we give examples of what could be called 'abelian' Gr-categories, which Hoàng Xuân Sính called 'Pic-categories' after the mathematician Émile Picard. In Section 6 we describe how Gr-categories arise in topology, and give more examples using these ideas. In Section 7 we give a tiny taste of how Gr-categories—now called '2-groups'—have made an appearance in theoretical and mathematical physics. Finally, in Appendix A we give some definitions concerning monoidal and symmetric monoidal categories, allowing us to state some of Hoàng Xuân Sính's results more precisely.

2. Monoidal categories

Hoàng Xuân Sính's work on Gr-categories is part of the trend toward 'categorification': taking familiar structures built using sets and finding their analogues built using categories. As we shall see, a Gr-category is a category that resembles a group. But a crucial subtlety is that a Gr-category has inverses at two different levels: both the objects and morphisms have inverses! The objects describe symmetries, while the morphisms describe symmetries *of* symmetries.

To define Gr-categories we should start with monoidal categories. A monoidal category is a category that resembles a monoid. A bit more precisely, a monoidal category is a category M with a tensor product

$$\otimes\colon \mathsf{M} \times \mathsf{M} \to \mathsf{M}$$

and a unit object $I \in \mathsf{M}$, obeying the associative and unit laws up to specified isomorphisms, which in turn must obey some laws of their own. Given this, a Gr-category is a monoidal category in which every morphism and every object has an inverse. Here a morphism f has an inverse if there is a morphism f' such that $f \circ f'$ and $f' \circ f$ are identity morphisms. Similarly, an object g in a monoidal category has an inverse if there an object g' such that $g \otimes g' \cong I$ and $g' \otimes g \cong I$.

Now that monoidal categories are well-understood, the definition of Gr-category seems simple and natural. However, monoidal categories were still quite new when Hoàng Xuân Sính wrote her thesis, and she did even not call them that. In 1963, Bénabou published a short paper containing a preliminary version of monoidal categories under the name of 'catégories avec multiplication' [5]. Also in 1963, Mac Lane published a paper correcting an important problem with Bénabou's definition and giving the modern definition [30]. However, at this time Mac Lane called monoidal categories 'bicategories'. Ironically, this term is now used for a very different concept introduced later by Bénabou [6]. According to Mac Lane [34], it was Eilenberg who coined the term 'monoidal category'. It seems to have first appeared in a paper by Eilenberg and Kelly [15]. Hoàng Xuân Sính does not cite Mac Lane's 1963 paper. For her terminology on monoidal categories, she refers to the thesis of Neantro Saavedra-Rivano, another student of Grothendieck [40, 41].

But let us see how Bénabou handled the concept of 'category with multiplication', and how Mac Lane's important correction plays a key role in Hoàng Xuân Sính's work. First, Bénabou demanded that a category with multiplication have a functor

$$\otimes\colon \mathsf{M} \times \mathsf{M} \to \mathsf{M}.$$

We call this functor the **tensor product**, and write $\otimes(x, y) = x \otimes y$ and $\otimes(f, g) = f \otimes g$ for objects $x, y \in M$ and morphisms f, g in M. Second, Bénabou said that a category with multiplication have an object $I \in \mathsf{M}$ called the **unit**. For example, the unit for the tensor product in the category of vector spaces can be taken to be any one-dimensional vector space, while the unit for the Cartesian product in the category of sets can be taken to be any one-element set.

Bénabou did not demand that the tensor product be associative 'on the nose', in the obvious equational way:

$$(x \otimes y) \otimes z = x \otimes (y \otimes z).$$

Instead, he required that the tensor product be associative up to a natural isomorphism, which we now call the 'associator':

$$a_{x,y,z} \colon (x \otimes y) \otimes z \xrightarrow{\sim} x \otimes (y \otimes z).$$

Similarly, he did not demand that I act as the unit for the tensor product on the nose, but only up to natural isomorphisms which are now called the left and right 'unitors':

$$\ell_x \colon I \otimes x \xrightarrow{\sim} x, \qquad r_x \colon x \otimes I \xrightarrow{\sim} x.$$

The reason is that in applications, it is usually too much to expect equations between objects in a category: usually we just have isomorphisms, and this is good enough! Indeed this is a basic principle of categorification: equations between objects are bad; we should instead specify isomorphisms.

If we stopped the definition here, there would be a problem, since we could use the associator to build *several different* isomorphisms between a pair of objects constructed using tensor products. The simplest example occurs when we have four objects $w, x, y, z \in \mathsf{M}$. Then we can build two isomorphisms from

$((w \otimes x) \otimes y) \otimes z$ to $w \otimes (x \otimes (y \otimes z))$ using associators:

$$\begin{tikzcd}
 & ((w \otimes x) \otimes y) \otimes z \arrow[dl, "a_{w,x,y} \otimes 1_z"'] \arrow[dr, "a_{w \otimes x,y,z}"] & \\
(w \otimes (x \otimes y)) \otimes z \arrow[d, "a_{w,x \otimes y,z}"'] & & (w \otimes x) \otimes (y \otimes z) \arrow[d, "a_{w,x,y \otimes z}"] \\
w \otimes ((x \otimes y) \otimes z) \arrow[rr, "1_w \otimes a_{x,y,z}"'] & & w \otimes (x \otimes (y \otimes z))
\end{tikzcd} \tag{2.1}$$

If these isomorphisms were different, we would need to to say *how* we reparenthesized one expression to get the other, just to specify an isomorphism between them. This situation is theoretically possible, but doesn't seem to come up much in mathematics. So, it seems wise to require that the above pentagon commutes. It does in most examples we usually care about.

When we build isomorphisms using unitors, we get further ways to build multiple isomorphisms between objects constructed using tensor products and the unit object. The simplest example is this:

$$\begin{tikzcd}
I \otimes I \arrow[r, bend left, "\ell_I"] \arrow[r, bend right, "r_I"'] & I
\end{tikzcd} \tag{2.2}$$

When we combine unitors and associators, the three simplest examples are these triangles:

$$\begin{tikzcd}
(I \otimes x) \otimes y \arrow[rr, "a_{I,x,y}"] \arrow[dr, "\ell_x \otimes 1_y"'] & & I \otimes (x \otimes y) \arrow[dl, "\ell_{x \otimes y}"] \\
 & x \otimes y &
\end{tikzcd} \tag{2.3}$$

$$\begin{tikzcd}
(x \otimes I) \otimes y \arrow[rr, "a_{x,I,y}"] \arrow[dr, "r_x \otimes 1_y"'] & & x \otimes (I \otimes y) \arrow[dl, "1_x \otimes \ell_y"] \\
 & x \otimes y &
\end{tikzcd} \tag{2.4}$$

$$
\begin{tikzcd}
(x \otimes y) \otimes I \arrow[rr, "a_{x,y,I}"] \arrow[dr, "r_{x \otimes y}"'] & & x \otimes (y \otimes I) \arrow[dl, "1_x \otimes r_y"] \\
& x \otimes y &
\end{tikzcd}
\tag{2.5}
$$

So, we should require that these triangles commute too.

Of course, there are infinitely many other diagrams we can build using associators and unitors. Bénabou tried to handle all of them in one blow. He imposed an axiom saying essentially that *all* diagrams constructed from associators, left unitors and right unitors must commute. Unfortunately such an axiom fails to have the desired effect unless one states it very carefully—which Bénabou, alas, did not. The problem is that even in a perfectly nice monoidal category, there can be diagrams built from the associator that fail to commute due to 'accidental' equations between tensor products. For instance, suppose in some monoidal category we just happen to have $(x \otimes y) \otimes z = b \otimes (c \otimes d)$ and $(b \otimes c) \otimes d = (u \otimes v) \otimes w$ and $x \otimes (y \otimes z) = u \otimes (v \otimes w)$. Then we can build this diagram:

$$
\begin{tikzcd}
(u \otimes v) \otimes w = (b \otimes c) \otimes d \arrow[rr, "a_{b,c,d}"] \arrow[dr, "a_{u,v,w}"'] & & b \otimes (c \otimes d) = (x \otimes y) \otimes z \arrow[dl, "a_{x,y,z}"] \\
& u \otimes (v \otimes w) = x \otimes (y \otimes z) &
\end{tikzcd}
$$

According to Bénabou's axiom, this diagram must commute. But this does not hold in many of the examples we are interested in. So, Bénabou's axiom is too strong.

Mac Lane's solution [30] was to show that if diagrams (2.1), (2.2), (2.3), (2.4) and (2.5) commute, all diagrams built from associators and unitors that we really *want* to commute actually *do* commute. Making this precise and proving it was a major feat: it is called Mac Lane's coherence theorem. In 1964, Kelly [28] improved Mac Lane's result by showing that the pentagon identity (2.1) and the middle triangle identity (2.4) imply the rest. Thus, the modern definition of monoidal category is as follows:

Definition 2.1. *A* ***monoidal category*** *is a category M together with a* ***tensor product*** *functor $\otimes\colon M \times M \to M$, a* ***unit object*** *$I \in M$, and natural isomorphisms called the* ***associator****:*

$$a_{x,y,z}\colon (x \otimes y) \otimes z \xrightarrow{\sim} x \otimes (y \otimes z),$$

the ***left unitor****:*

$$\ell_x\colon I \otimes x \xrightarrow{\sim} x,$$

and the ***right unitor****:*

$$r_x \colon x \otimes I \xrightarrow{\sim} x.$$

such that the pentagon (2.1) *and the triangle* (2.4) *commute.*

As noted by Mac Lane, the pentagon was also discussed in a 1963 paper by James Stasheff [44], as part of an infinite sequence of polytopes called 'associahedra'. Stasheff defined a concept of 'A_∞-space', which is roughly a topological space having a product that is associative up to homotopy, where this homotopy satisfies the pentagon identity up homotopy, that homotopy satisfies yet another identity up to homotopy, and so on, *ad infinitum*. The nth of these identities is described by the n-dimensional associahedron. The first identity is just the associative law, which plays a crucial role in the definition of monoid. Mac Lane realized that the second, the pentagon identity, should play a similar role in the definition of monoidal category. The higher ones show up, at least implicitly, in Mac Lane's proof of his coherence theorem. They become even more important when we consider monoidal bicategories, monoidal tricategories and so on.

Less glamorous than the associator, the unitors and the laws they obey are also crucial to Mac Lane's coherence theorem. Together with the associahedra, they are described by a sequence of cell complexes sometimes called 'monoidahedra' [46] or 'unital associahedra' [36].

3. Gr-categories

We have already stated Hoàng Xuân Sính's definition of Gr-category, but now we are ready to examine her work on this concept more carefully, so let us repeat the definition:

Definition 3.1. *A* ***Gr-category*** *is a monoidal category such that every object has an inverse and every morphism has an inverse.*

She not only defined Gr-categories; she also defined when two Gr-categories are 'equivalent'. The idea is that equivalent Gr-categories are the same for all practical purposes. The definition of equivalence is a bit lengthy, so we relegate it to Appendix A. But the idea, even in a rough form, is very interesting.

We can define maps between Gr-categories. We say two Gr-categories G and G' are isomorphic if there are maps $F \colon \mathsf{G} \to G'$ and $G \colon \mathsf{G}' \to \mathsf{G}$ that are inverses, so

$$GF = 1_{\mathsf{G}} \quad \text{and} \quad FG = 1_{\mathsf{G}'}.$$

This is of course a typical definition of isomorphism for any sort of mathematical gadget. But because Gr-categories are categories, it turns out we can also define isomorphisms *between* maps between Gr-categories. We then say two Gr-categories G and G' are 'equivalent' if there are maps $F\colon \mathsf{G} \to G'$ and $G\colon \mathsf{G}' \to \mathsf{G}$ that are inverses up to isomorphism:

$$GF \cong 1_{\mathsf{G}} \quad \text{and} \quad FG \cong 1_{\mathsf{G}'}.$$

Two groups are isomorphic if renaming the elements of one group gives the other. Similarly, two Gr-categories are isomorphic if and only if we can get from one to the other by renaming the objects and morphisms. But equivalence is more flexible! For example, suppose we have a Gr-category G with a collection of isomorphic objects. Then we can *remove* all but one of these and get a smaller Gr-category that is equivalent to G. The idea is that it counts as 'redundant' to have multiple isomorphic copies of an object in a Gr-category. We can remove these redundant copies and get an equivalent Gr-category.

It is worth spelling this out in a bit more detail in an example. Suppose a Gr-category G has an isomorphism between two distinct objects, say $f\colon g \xrightarrow{\sim} g'$. Then we can remove g from G, along with all the isomorphisms going to and from it, and get a smaller category G'. This category is not immediately a Gr-category, since the tensor product of two objects in G' may happen to equal g. But we can redefine such tensor products to equal g', and suitably 'reroute' the associator and unitors of G' using the isomorphism f, to make G' into a Gr-category. Furthermore, this new Gr-category G' will be equivalent to G.

More generally, we can show any Gr-category is equivalent to a **skeletal** Gr-category: one where any two isomorphic objects must be equal. To classify Gr-categories up to equivalence, it thus suffices to classify skeletal Gr-categories up to equivalence.

Working with skeletal Gr-categories simplifies the classification problem. In a general Gr-category, the group laws hold up to isomorphism, since

$$(g \otimes g') \otimes g'' \cong g \otimes (g' \otimes g'')$$

$$I \otimes g \cong g \cong g \otimes I$$

and for any object g there is an object g' with

$$g \otimes g' \cong I \cong g' \otimes g.$$

But in a skeletal Gr-category, isomorphic objects are equal, so these laws become equations. Thus in a skeletal Gr-category the set of objects forms a group!

Indeed, from a skeletal Gr-category G we easily get three pieces of data:

- the group G of objects,
- the group A of automorphisms of the unit object $I \in \mathsf{G}$, which is an abelian group,
- an action ρ of the group G as automorphisms of the abelian group A, given as follows:
$$\rho(g)a := 1_g \otimes a \otimes 1_{g^{-1}}.$$

Clearly, the group G tells everything about tensoring objects in G. More subtly, the group A tells us everything we need to know about composing morphisms in G. Why is this? First, in a skeletal Gr-category every morphism is an isomorphism, and isomorphic objects are equal. Thus every morphism in G is an automorphism $f\colon g \to g$ for some $g \in \mathsf{G}$. Second, the group of automorphisms of any $g \in \mathsf{G}$ is isomorphic to the group of automorphisms of $I \in \mathsf{G}$, via the map sending $f\colon g \to g$ to $1_{g^{-1}} \otimes f\colon I \to I$.

The group A is abelian because given $a, b\colon I \to I$ we have

$$ab = (a1_I)\otimes(1_I b) = (a\otimes 1_I)(1_I \otimes b) = a\otimes b = (1_I\otimes a)(b\otimes 1_I) = (1_I b)\otimes(a1_I) = ba.$$

This marvelous calculation, later called the 'Eckmann–Hilton argument' [11], also shows that tensoring automorphisms of $I \in \mathsf{G}$ is the same as composing them! A further argument shows that the rest of the information about tensoring morphisms in G is contained in the action ρ of G on A.

The elephant in the room is the associator. To describe the associator of a Gr-category we need a fourth piece of data—the subtlest and most interesting piece. As we have seen, in any monoidal category the tensor product is associative up to a natural transformation called the associator:

$$\alpha_{g_1,g_2,g_3}\colon (g_1 \otimes g_2) \otimes g_3 \to g_1 \otimes (g_2 \otimes g_3).$$

Since this takes three objects and gives a morphism, we should try to encode this into a map from G^3 to A. How can we do this?

In a skeletal Gr-category, isomorphic objects are equal, so we can drop the parentheses and write

$$\alpha_{g_1,g_2,g_3}\colon g_1 \otimes g_2 \otimes g_3 \to g_1 \otimes g_2 \otimes g_3.$$

But beware: α_{g_1,g_2,g_3} may not be an identity morphism! It is just some automorphism of $g_1 \otimes g_2 \otimes g_3$. We can turn this into an automorphism of the unit object I and define a map

$$a\colon G^3 \to A$$

as follows:

$$a(g_1, g_2, g_3) = 1_{(g_1\otimes g_2\otimes g_3)^{-1}} \otimes \alpha_{g_1,g_2,g_3}.$$

Not every map $a\colon G^3 \to A$ arises from some Gr-category. After all, the associator must obey the pentagon identity (2.1). In terms of the map a, this says that for all $g_1, g_2, g_3, g_4 \in G$ we have

$$\rho(g_0)a(g_1,g_2,g_3) - a(g_0g_1,g_2,g_3) + a(g_0,g_1g_2,g_3) - a(g_0,g_1,g_2g_3) + a(g_0,g_1,g_2) = 0. \tag{3.1}$$

Here we have switched to writing the group operation in A additively, since this group is abelian.

Now, the precise value of the map $a\colon G^3 \to A$ depends on our choice of a skeletal Gr-category equivalent to G. Thus a is not uniquely determined by G. With some calculation, we can check that by changing our choice we can change the map $a\colon G^3 \to A$ to any map $a'\colon G^3 \to A$ with

$$a' - a = df$$

where $f\colon G^2 \to A$ is an arbitrary map and

$$(df)(g_1,g_2,g_3) = \rho(g_1)f(g_2,g_3) - f(g_1g_2,g_3) + f(g_1,g_2g_3) - f(g_1,g_2). \tag{3.2}$$

Equations (3.1) and (3.2) look intimidating at first sight, but by the time Hoàng Xuân Sính wrote her thesis these equations were already familiar in the subject of group cohomology. For any group G equipped with an action on an abelian group A, there is a sequence of groups $H^n(G,A)$ called the 'cohomology groups' of G with coefficients in A [39]. For $n = 1$ and $n = 2$ these groups first arose in Galois theory and the study of group extensions, and they are fairly easy to interpret. For $n = 3$ and above, these groups were independently discovered in 1943–45 by teams in different countries who found it difficult to communicate during the chaos of World War II. For the complex history of these developments see the accounts of Hilton [22], Mac Lane [31, 32, 33] and Weibel [47].

The significance of these higher cohomology groups, even $H^3(G,A)$, was at first rather obscure. Around 1945, Eilenberg and Mac Lane [12] found a powerful topological explanation of these groups. Ultimately, the idea is that that for any group G there is a topological space called the 'Eilenberg–Mac Lane space' $K(G,1)$ whose cohomology groups with a suitable system of 'local coefficients' depending on A and the action ρ are the groups $H^n(G,A)$. But the meaning of $H^3(G,A)$ became even more vivid thanks to Hoàng Xuân Sính's work.

How does this come about? First, the resemblance of equations (3.1) and (3.2) is no coincidence! For any map $h\colon G^n \to A$ there is a similar-looking formula for a map $dh\colon G^{n+1} \to A$. A map $h\colon G^n \to A$ with $dh = 0$ is called an ***n*-cocycle**, and a map $h\colon G^n \to A$ of the form $h = df$ for some f is called an ***n*-coboundary**. Since

$$ddf = 0$$

for all f, every n-boundary is automatically an n-cocycle. The group of n-cocycles modulo n-coboundaries is called the **nth cohomology group** of G with coefficients in A, and denoted $H^n(G, A)$. Elements of this group are called **cohomology classes**.

In this language, (3.1) says that the associator of a skeletal Gr-category gives a 3-cocycle $a\colon G^3 \to A$. Similarly, (3.2) says that the 3-cocycles a and a' differ by a 3-coboundary—and thus represent the same cohomology class.

Thus, Hoàng Xuân Sính's work shows that any Gr-category gives, not only two groups G and A and an action of the first on the second, but also an element of $H^3(G, A)$. More importantly, she proved that two Gr-categories with the same groups G and A and the same action ρ of G on A are equivalent *if and only if* their associators define the same element of $H^3(G, A)$. We state this result more carefully in Theorem A.1 of Appendix A.

This gave a new explanation of the meaning of the cohomology group $H^3(G, A)$. In simple terms, this group classifies the possible associators that a Gr-category can have when the rest of its structure is held fixed. The element of $H^3(G, A)$ determined by the associator of a Gr-category G is now called the **Sính invariant** of G.

4. Strict Gr-categories

We have emphasized that in a Gr-category, the associative and unit laws hold only up to natural isomorphism. Nonetheless it is interesting to consider Gr-categories where these natural isomorphisms are identity morphisms. In a skeletal Gr-category we have

$$(g \otimes g') \otimes g'' = g \otimes (g' \otimes g'')$$

and

$$I \otimes g = g = g \otimes I.$$

However, this is somewhat deceptive, because the associator and unitors may still not be identity morphisms! Thus we say a Gr-category is **strict** if its associator and unitors are identity morphisms.

In 1978, Hoàng Xuân Sính published a paper about strict Gr-categories in which she proved that every Gr-category is equivalent to a strict one [25]. At a first glance this might seem to contradict that fact that every Gr-category G is equivalent to a skeletal one, which is then characterized up to equivalence by four pieces of data:

- its group G of objects,
- the abelian group A of automorphisms of the unit object,
- the action ρ of G on A,
- the cohomology class $[a] \in H^3(G, A)$ arising from its associator.

Since any Gr-category is equivalent to a strict one, where the associator is the identity, isn't $[a] \in H^3(G, A)$ always zero? No, because that strict Gr-category is not generally skeletal! In fact, it follows from what we have said that $[a] = 0$ if and only if G is equivalent to a strict and skeletal Gr-category.

This fact bears repeating, since it trips up beginners so often: *every Gr-category is equivalent to a strict one, and also equivalent to a skeletal one, but it is equivalent to a strict and skeletal one if and only if its Sính invariant* $[a]$ *is zero.*

Before Gr-categories were introduced, strict Gr-categories were already known, though of course not under that name. Their early history is quite obscure, at least to this author. Strict Gr-categories are now often called 'categorical groups', but the earliest paper I have found using that term dates only to 1981, and uses it to mean something else [43]. In 1976, Brown and Spencer [8] published a proof that strict Gr-categories, which they called '$\mathcal{G}$-groupoids', are equivalent to 'crossed modules', a structure introduced by J. H. C. Whitehead in 1946 for use in topology [48, 49]. But they write that this result was known to Verdier in 1965, used by Duskin in some unpublished notes in 1969, and discovered independently by them in 1972.

Crossed modules are still a useful way to get examples of Gr-categories. For a modern proof that crossed modules are equivalent to strict Gr-categories, see Forrester-Barker [16]. But what exactly is a crossed module?

Definition 4.1. *A* ***crossed module*** *is a quadruple* (G, H, t, ρ) *where* G *and* H *are groups,* $t\colon H \to G$ *is a homomorphism, and* ρ *is an action of* G *as automorphisms of* H *such that* t *is* ***G-equivariant****:*

$$t(\rho(g)h) = g\, t(h)\, g^{-1}$$

and t *satisfies the so-called* ***Peiffer identity****:*

$$\rho(t(h))h' = hh'h^{-1}.$$

We can obtain a crossed module from a strict Gr-category G as follows:

- Let G be the set of objects of G, made into a group by tensor product.

- Let H be the set of all morphisms from $I \in \mathsf{G}$ to arbitrary objects of G, made into a group by tensor product.
- Let $t\colon H \to G$ map any morphism $f\colon 1 \to g$ to $g \in \mathsf{G}$.
- Let $\rho(g)h = 1_g \otimes h \otimes 1_{g^{-1}}$.

Conversely, we can build a strict Gr-category G from a crossed module (G, H, t, ρ) as follows. We take the set of objects of G to be G, and we define the tensor product of objects using multiplication in G. We take the set of morphisms of G to be the semidirect product $H \rtimes G$ in which multiplication is given by

$$(h, g)(h', g') = (h\rho(g)h', gg').$$

We define the tensor product of morphisms using multiplication in this semidirect product. We treat the pair (h, g) as a morphism from $g \in \mathsf{G}$ to $t(h)g \in \mathsf{G}$, and we define the composite of morphisms

$$(h, g)\colon g \to g', \qquad (h', g')\colon g' \to g''$$

to be

$$(hh', g)\colon g \to g''.$$

Using this correspondence we can now get many examples of strict Gr-categories. For example: any action of a group on an abelian group, any normal subgroup of a group, or any central extension of a group gives a crossed module, and thus a strict Gr-category. So, strict Gr-categories unify a number of important concepts in group theory!

To see this, first suppose ρ is any action of a group G on an abelian group H. Then defining $t\colon H \to G$ by $t(h) = 1$, the G-equivariance of t and the Peiffer identity are easy to check, so we get a crossed module. Because $t(h) = 1$, the resulting strict Gr-category is skeletal. In fact we can turn this argument around, and show that any skeletal strict Gr-category arises this way (up to isomorphism).

Second, suppose H is any normal subgroup of G. Let $t\colon H \to G$ be the inclusion of H in G and let

$$\rho(g)h = ghg^{-1}$$

Then clearly t is G-equivariant and the Peiffer identity holds, so we get a crossed module. Here the resulting strict Gr-category is usually *not* skeletal, since typically we do not have $t(h) = 1$ for all $h \in H$.

Third, suppose we have a central extension of a group G by a group C, i.e. a short exact sequence

$$1 \xrightarrow{i} C \longrightarrow H \xrightarrow{t} G \longrightarrow 1$$

where the image of C lies in the center of H. Then we can define an action ρ of G on H by choosing any function $j\colon G \to H$ with $t(j(g)) = g$ for all $g \in G$ and letting

$$\rho(g)h = j(g)hj(g)^{-1}.$$

We can check that ρ does not depend on the choice of j. With more work, we can check that (G, H, t, ρ) is a crossed module. The resulting strict Gr-category is again usually not skeletal. On the contrary, since t is surjective, in this Gr-category all objects are isomorphic.

If we build a strict Gr-category from a crossed module, and it is not skeletal, it may have a nontrivial Sính invariant. For the story of how mathematicians discovered the relation of the 3rd cohomology of groups to crossed modules, see Mac Lane's historical remarks [32, 33]. The connection was certainly known well before Hoàng Xuân Sính wrote her thesis.

5. Pic-categories

The subject of Gr-categories is interesting because there are examples arising from many different branches of mathematics, and the relationships between these examples exposes links between these branches. One important and tractable class of Gr-categories studied by Hoàng Xuân Sính are what she called 'Pic-categories'. Let us take a look at these.

Any monoidal category M gives a Gr-category called its **core**, namely the subcategory consisting of only the invertible objects and only the invertible morphisms. For example, if R is any commutative ring, there is a category $R\mathsf{Mod}$ where

- objects are R-modules,
- morphisms are homomorphisms of R-modules.

This category becomes monoidal with the usual tensor product of R-modules, and the unit for this tensor product is R itself, considered as an R-module. The core of $R\mathsf{Mod}$ is a Gr-category that could be called $\mathsf{Pic}(R)$, the **Picard category** of R. By Hoàng Xuân Sính's classification of Gr-categories, all the information in this Gr-category is contained in four pieces of data:

- The group of isomorphism classes of invertible R-modules. This is called the **Picard group** $\mathrm{Pic}(R)$. This group is always abelian.

- The abelian group of invertible R-module homomorphsms $f\colon R \to R$. Since these are all given by multiplication by invertible elements of R, this is called the **units group** $R^\times$.
- The action of the Picard group on the units group. This action is always trivial.
- The Sính invariant of $\mathsf{Pic}(R)$. This is always trivial.

It is evident from this that the Picard category of a commutative ring is a specially simple sort of Gr-category.

The Picard group and units group are important invariants of a commutative ring. For example, when R is the ring of algebraic integers of some algebraic number field, the Picard group $\mathrm{Pic}(R)$ is isomorphic to the 'ideal class group' of R, familiar in number theory. The reason is that in this case, every invertible R-module is isomorphic to an ideal of R. Even better, it turns out that for a ring of algebraic integers $\mathrm{Pic}(R)$ is trivial if and only if unique prime factorization holds in R—at least up to reordering and units. When $\mathrm{Pic}(R)$ is nontrivial, unique factorization fails, and this is one reason Dedekind, building on earlier work of Kummer, introduced the ideal class group. We can compare two examples:

- When $R = \mathbb{Z}[i]$ is the ring of Gaussian integers, the Picard group $\mathsf{Pic}(R)$ is trivial and the units group $\mathbb{R}^\times$ is $\mathbb{Z}/4$, consisting of $1, i, -1$ and $-i$. We have unique prime factorization in $\mathbb{Z}[i]$.
- When $R = \mathbb{Z}[\sqrt{-5}]$, the Picard group $\mathsf{Pic}(R)$ is $\mathbb{Z}/2$ and the units group $R^\times$ is $\mathbb{Z}/2$, consisting of 1 and -1. We do not have unique prime factorization in $\mathbb{Z}[\sqrt{-5}]$.

It is worth comparing an example from topology. Suppose X is a compact Hausdorff space. Then we can let R be the ring of continuous complex-valued functions on X, with pointwise addition and multiplication. In this case $\mathsf{Pic}(R)$ is equivalent to the Gr-category where

- objects are complex line bundles over X,
- morphisms are isomorphisms between complex line bundles,

and the tensor product is the usual tensor product of line bundles. It follows that we can identify the Picard group $\mathrm{Pic}(R)$ with the set of isomorphism classes of complex line bundles over X, made into a group using tensor products. This is famously isomorphic to the cohomology group $H^2(X, \mathbb{Z})$. The units group $R^\times$ is simply the group of continuous functions from X to $\mathbb{C}^\times$, the multiplicative group of invertible complex numbers.

Line bundles also show up in another example of a Gr-category. Suppose X is a complex projective algebraic variety. Then the category of holomorphic vector bundles over X, with its usual tensor product, is a monoidal category. Its core is a Gr-category where

- objects are holomorphic line bundles over X,
- morphisms are isomorphisms between holomorphic line bundles,

and the tensor product is the usual tensor product of holomorphic line bundles. In this case the group of isomorphism classes of objects is called the **Picard group** of X. This Picard group is much more interesting than the previously mentioned purely topological example. It depends not only on the topology of X, but on its holomorphic structure. Moreover, instead of a discrete group, this Picard group is best thought of as a topological group whose connected components are themselves projective algebraic varieties!

All the examples of Gr-categories mentioned in this section so far are not only monoidal categories, but 'symmetric' monoidal categories. These were introduced by Mac Lane in his 1963 paper [30], though not under that name. Just as monoidal categories are a categorification of monoids, symmetric monoidal categories are a categorification of *commutative* monoids. But instead of requiring that the tensor product commute 'on the nose', we demand that it commute up to a natural isomorphism, which must obey a law of its own:

Definition 5.1. *A* ***symmetric monoidal category*** *is a monoidal category* M *equipped with a natural isomorphism called the* ***symmetry***

$$S_{x,y} \colon x \otimes y \to y \otimes x$$

such that

$$S_{y,x} \circ S_{x,y} = 1_{x \otimes y}$$

for all objects $x, y \in \mathsf{M}$ and the following diagram commutes for all $x, y, z \in \mathsf{M}$:

$$
\begin{tikzcd}
& (x \otimes y) \otimes z \arrow[r, "S_{x,y} \otimes z"] & (y \otimes x) \otimes z \arrow[dr, "a_{y,x,z}"] & \\
x \otimes (y \otimes z) \arrow[ur, "a^{-1}_{x,y,z}"] \arrow[dr, "S_{x,y \otimes z}"'] & & & y \otimes (x \otimes z) \arrow[dl, "y \otimes S_{x,z}"] \\
& (y \otimes z) \otimes x \arrow[r, "a_{y,z,x}"'] & y \otimes (z \otimes x) &
\end{tikzcd}
\tag{5.1}
$$

This commuting hexagon says that switching the object x past $y \otimes z$ all at once is the same as switching it first past y and then past z (with some associators thrown in to move the parentheses).

Hoàng Xuân Sính made the following definition:

Definition 5.2. *A* ***Pic-category*** *is a symmetric monoidal Gr-category.*

In fact, all the examples of Gr-categories we have seen so far are even better than Pic-categories: they are what Hoàng Xuân Sính called 'restreintes'.

Definition 5.3. *A Pic-category is* ***restrained*** *if the symmetry $S_{g,g} \colon g \otimes g \to g \otimes g$ is the identity for every object $g \in \mathsf{G}$.*

In particular, what we are calling the Picard category of a commutative ring and the Picard category of a complex projective algebraic variety are restrained Pic-categories.

Among Gr-categories, restrained Pic-categories are especially simple. Recall that Gr-categories are classified up to equivalence by four pieces of data:

- the group G of isomorphism classes of objects,
- the abelian group A of automorphisms of the unit object,
- an action ρ of G as automorphisms of A,
- an element $[a] \in H^3(G, A)$.

A restrained Pic-category always has these properties:

- the group G is abelian,
- the action ρ is trivial,
- the element $[a]$ is zero.

In fact, Hoàng Xuân Sính showed in her thesis that any restrained Pic-category is characterized by a pair of abelian groups G and A. To do this, she first defined an appropriate concept of 'equivalence' for Pic-categories, which is more fine-grained than equivalence of Gr-categories since the symmetry also plays a role. Then, she classified Pic-categories and also restrained Pic-categories up to this notion of equivalence. We explain the latter classification in Theorem A.4 of Appendix A.

We can also get restrained Pic-categories from chain complexes. A **chain complex of abelian groups** is a sequence of abelian groups $C_0, C_1, \ldots$, together with homomorphisms

$$C_0 \xleftarrow{\partial_1} C_1 \xleftarrow{\partial_2} C_2 \xleftarrow{\partial_3} \cdots$$

such that $\partial_n \circ \partial_{n+1} = 0$. A **2-term** chain complex of abelian groups is one where $C_n = 0$ except for C_0 and C_1. Thus a 2-term chain complex of abelian groups is just an elaborate way of thinking about two abelian groups and a homomorphism between them. However, this way of thinking is useful because it paves the way for generalizations.

Given a 2-term chain complex C of abelian groups, say $C_0 \xleftarrow{\partial} C_1$, we can construct a category G_C where:

- objects are elements $g \in C_0$,
- a morphism $h\colon g \to g'$ is an element $h \in C_1$ with $dh = g' - g$,
- to compose morphisms $h\colon g \to g'$ and $h'\colon g' \to g''$ we add them, obtaining $h' + h\colon g \to g''$.

We can make this category G_C into a Gr-category by using addition in C_0 as the tensor product of objects and addition in C_1 as the tensor product of morphisms. Since addition in these abelian groups is commutative we can make G_C into a symmetric monoidal category where the symmetry $S_{g,g'}\colon g + g' \to g' + g$ is the identity. Then G_C becomes a strict Pic-category.

In a 1982 paper [26] Hoàng Xuân Sính showed that in a certain sense *all* strict Pic-categories arise from this simple construction. More precisely, she proved that every strict Pic-category is equivalent to one of the form G_C for some 2-term chain complex C of abelian groups. We state this result in Theorem A.5 of Appendix A. As a consequence, the Pic-category of a commutative ring, which we defined using invertible R-modules and isomorphisms between these, can also be expressed in terms of a 2-term chain complex.

6. Gr-categories and topology

For more examples of Gr-categories it pays to exploit the connection between Gr-categories and topology. Our brief discussion of group cohomology touched on this theme, but did not do justice to either the history or the mathematics.

One of the dreams of topology is to classify topological spaces. Two such spaces X and Y are homeomorphic if there are maps going back and forth, say $f\colon X \to Y$ and $g\colon Y \to X$, that are inverses:

$$gf = 1_X \quad \text{and} \quad fg = 1_Y.$$

But it has long been known that classifying spaces up to homeomorphism is an absolutely unattainable goal. So, various lesser but still herculean tasks have been proposed as substitutes.

For example, instead of demanding that f and g are inverses 'on the nose', we can merely ask for them to be inverses up to homotopy. Given two maps $h, h'\colon A \to B$, a **homotopy** between them is a continuous 1-parameter family of maps interpolating between them: that is, map $H\colon [0,1] \times A \to B$ with

$$\begin{array}{rcl} H(0,a) & = & h(a), \\ H(1,a) & = & h'(a). \end{array}$$

If there is a homotopy from h to h' we write $h \simeq h'$. We then say two spaces X and Y are **homotopy equivalent** if there are maps $f\colon X \to Y$ and $g\colon Y \to X$ such that

$$gf \simeq 1_X \quad \text{and} \quad fg \simeq 1_Y.$$

Classifying spaces up to homotopy equivalence is still impossibly complicated unless we require that these spaces are locally nice in some sense. Manifolds, being locally homeomorphic to $\mathbb{R}^n$, are the paradigm of what we might mean by locally nice. But we do not need to go this far. A 'CW complex' is a space built by starting with a discrete set of points, or 0-balls, and iteratively gluing on 1-balls, 2-balls, and so on, by attaching their boundaries to the space built so far. Topologists have adopted the goal of classifying CW complexes up to homotopy equivalence as a kind of holy grail of the subject.

We can simplify this quest in a couple of ways. First, since every CW complex is a disjoint union of connected CW complexes, we can focus on connected one. Second, it is very useful to equip a connected CW complex X with a chosen point, called a 'basepoint' and denoted $* \in X$. Spaces with basepoint are called **pointed** spaces. When we working with these we demand that all maps send the basepoint of one space to that of another, and that homotopies do this for all values of the parameter t.

Now, instead of trying to classify all connected pointed CW complexes in one blow, which is still beyond our powers, it is more manageable to start with ones whose interesting features are concentrated in low dimensions and work our way up. For this we should recall the idea of homotopy groups. Suppose X is a pointed CW complex, and let S^n be the n-sphere with an arbitrarily chosen basepoint. Then we define the ***n*th homotopy group** $\pi_n(X)$ to be the set of homotopy classes of maps from S^n to X—where again, we demand that all maps and homotopies are compatible with the chosen basepoints. We can think of $\pi_n(X)$ as the set of holes in X that can be caught with an n-dimensional lasso. Despite its name $\pi_0(X)$ is a mere set, and a 1-element set when X is connected, so this gives no information at all. But $\pi_n(X)$ is a group for $n \geq 1$, and an abelian group for $n \geq 2$.

We say a CW complex X is an ***n*-type** if, regardless of how we choose a basepoint for it, $\pi_k(X)$ is trivial for $k > n$. Thus, intuitively, an n-type is a locally nice space whose interesting features live only in dimensions $\leq n$, at least as viewed through the eye of homotopy groups.

A remarkable fact, discovered by Eilenberg and Mac Lane around 1954, is that connected pointed 1-types are classified by groups [13, 14]. On the one hand, given a connected pointed 1-type X we get a group $\pi_1(X)$. More surprisingly, two connected pointed 1-types have isomorphic groups π_1 if and only if they are homotopy equivalent. On the other hand, given any group G, there is a concrete procedure for building a connected pointed 1-type X with $\pi_1(X) = G$. This space X is now called the **Eilenberg–Mac Lane space** $K(G, 1)$, and it is quite easy to describe.

Start with a point $*$ to serve as the basepoint. For each element $g \in G$, take an interval and glue both of its ends to this point:

$$* \xrightarrow{g} *$$

Here we draw the two ends of the interval separately for convenience, but they are really the same point $*$. Then, for any pair of elements $g, h \in G$, take a triangle and glue its edges to intervals for g, h and gh:

$$\begin{tikzcd} & * \arrow[dr, "h"] & \\ * \arrow[ur, "g"] \arrow[rr, "gh"'] & & * \end{tikzcd}$$

Next, for any triple of elements $g, h, k \in G$, glue a tetrahedron of this sort to the already present triangles:

$$
\begin{tikzcd}
* \arrow[r, "h"] \arrow[dr, "hk"'] & * \arrow[d, "k"] \\
* \arrow[u, "g"] \arrow[r, "ghk"'] \arrow[ur, dotted, "gh"] & *
\end{tikzcd}
$$

And so on: glue on an n-dimensional simplex for each n-tuple of elements of G. The result is the Eilenberg–Mac Lane space $K(G, 1)$.

And now we reach the main point: just as connected pointed 1-types are classified by groups, connected pointed 2-types are classified by Gr-categories!

This realization did not come out of the blue—far from it. A crucial first step was J. H. C. Whitehead's concept of crossed module, formulated around 1946 without the aid of category theory [48, 49]. In 1950, Mac Lane and Whitehead [35] proved that a crossed module captures all the information in a connected pointed 2-type. As mentioned in our discussion of strict Gr-categories, it seems Verdier knew by 1965 that a crossed module is another way of thinking about a strict Gr-category (though the latter term did not yet exist). A proof was first published by Brown and Spencer [8] in 1976. However, Grothendieck was familiar with some of these ideas before then, which must be one reason he proposed that Hoàng Xuân Sính write her thesis on Gr-categories.

Indeed, Grothendieck already had ideas for vast generalizations, which became more important in his later work, and especially his monumental text *Pursuing Stacks* [17]. He conjectured that n-types should be classified up to homotopy equivalence by algebraic structures called 'n-groupoids'. Very roughly speaking, an n-groupoid has objects, morphisms between objects, 2-morphisms between 1-morphisms, and so on up to level n, all invertible up to higher morphisms. Filling in the details and proving Grothendieck's conjecture—usually called the 'homotopy hypothesis'—continues to be a challenge [21].

An n-groupoid with one object is called an 'n-group'. A 1-group is simply a group, while a 2-group is a Gr-category. As a spinoff of the homotopy hypothesis, connected pointed n-types should be classified up to homotopy equivalence by n-groups. As we have seen, the case $n = 1$ was handled by Eilenberg and Mac Lane. While the case $n = 2$ was tackled by the authors listed above, it was still not completely clarified when Hoàng Xuân Sính wrote her thesis.

In fact, it is possible to mimic Eilenberg and Mac Lane's construction and build a connected pointed 2-type starting from any Gr-category G, much as they built a connected pointed 1-type starting from a group. For this it is convenient to start by replacing G with an equivalent strict Gr-category G. So, let us assume G is strict.

Start with a point $*$. For each object $g \in \mathsf{G}$, take an interval and glue both of its ends to this point:

$$* \xrightarrow{g} *$$

Then, for any triple of objects $g, h, k \in \mathsf{G}$ and any morphism $a\colon g \otimes h \to k$, take a triangle and glue its edges to the intervals for g, h and k:

$$\begin{tikzcd}
& * \arrow[dr, "h"] & \\
* \arrow[ur, "g"] \arrow[rr, "k"'] & \arrow[u, phantom, "\Downarrow a"] & *
\end{tikzcd}$$

Next, for any tetrahedron as shown below, glue this tetrahedron to the already present triangles if this tetrahedron commutes:

$$\begin{tikzcd}[row sep=huge, column sep=huge]
* \arrow[r, "h"] \arrow[dr, "m"'] \arrow[dr, phantom, shift left=4ex, "c"] \arrow[dr, phantom, shift right=4ex, "\Downarrow a"] & * \arrow[d, "k"] \\
* \arrow[u, "g"] \arrow[ur, dotted, "\ell"] \arrow[ur, phantom, shift left=4ex, "\Swarrow b"] \arrow[ur, phantom, shift right=4ex, "d"] \arrow[r, "n"'] & *
\end{tikzcd}$$

By saying that this tetrahedron 'commutes', we mean that the composite of the front two sides equals the composite of the back two sides, i.e. the following two morphisms in G are equal:

$$\begin{tikzcd}[row sep=huge, column sep=huge]
* \arrow[r, "h"] \arrow[dr, "m"'] \arrow[dr, phantom, shift left=4ex, "\Swarrow b"] \arrow[dr, phantom, shift right=4ex, "\Downarrow a"] & * \arrow[d, "k"] \\
* \arrow[u, "g"] \arrow[r, "n"'] & *
\end{tikzcd}
\quad = \quad
\begin{tikzcd}[row sep=huge, column sep=huge]
* \arrow[r, "h"] & * \arrow[d, "k"] \\
* \arrow[u, "g"] \arrow[ur, "\ell"] \arrow[ur, phantom, shift left=4ex, "\Searrow c"] \arrow[ur, phantom, shift right=4ex, "\Downarrow d"] \arrow[r, "n"'] & *
\end{tikzcd}$$

Next glue on a 4-simplex whenever all its tetrahedral faces are present, and so on in all higher dimensions. This gives a space $B\mathsf{G}$ called the **classifying space** of G. This is the connected pointed 2-type corresponding to the Gr-category G. For more details see Duskin [10], Bullejos and Cegarra [9], and Noori [38].

Remember, a Gr-category G is characterized up to equivalence by four pieces of data:

- a group G,
- an abelian group A,
- an action ρ of G on A,
- the Sính invariant $[a] \in H^3(G, A)$.

These are all visible from the classifying space $B\mathsf{G}$. This space has $\pi_1(B\mathsf{G}) = G$ and $\pi_2(B\mathsf{G}) = A$. For any pointed space the group π_1 has an action on the abelian group π_2, and for the space $B\mathsf{G}$ this action is ρ. More subtly, every space has a 'Postnikov invariant', an element of $H^3(\pi_1, \pi_2)$, and for $B\mathsf{G}$ this is the Sính invariant $[a] \in H^3(G, A)$.

We have seen how to turn a Gr-category into a space, which happens to be a connected pointed 2-type. But the relation between Gr-categories and topology runs deeper than that. There is also a way to turn a pointed space into a Gr-category!

Suppose X is a pointed space. First, remember that we can get a group from X, namely $\pi_1(X)$, usually called the **fundamental group** of X. Elements of $\pi_1(X)$ are homotopy classes of loops in X: that is, maps $g\colon S^1 \to X$. To multiply two loops we simply combine them into a single loop, starting the second where the first ended.

In Example 2.3.3 of her thesis [23], Hoàng Xuân Sính tersely described how to enhance this procedure to get a Gr-category from X. This Gr-category is now called the **fundamental 2-group** of X, and denoted $\Pi_2(X)$. An object of $\Pi_2(X)$ is a map $g\colon S^1 \to X$, and a morphism $f\colon g \to g'$ is a homotopy class of homotopies from g to g'. The details are carefully worked out in various later sources, e.g. the work of Hardie, Kamps and Kieboom [19, 20].

There are other ways to obtain the fundamental 2-group of a connected pointed space X. For example:

- let $G = \pi_1(X)$,
- let $A = \pi_2(X)$,
- let ρ be the action of $\pi_1(X)$ on $\pi_2(X)$,
- let $[a] \in H^3(\pi_1(X), \pi_2(X))$ be the Postnikov invariant of X.

Then, we can take $\Pi_2(X)$ to be the Gr-category associated to this quadruple $(G, A, \rho, [a])$.

As a result of all this, not only can we turn any Gr-category G into a connected pointed 2-type $B\mathsf{G}$, we can also turn any connected pointed 2-type X into a Gr-category $\Pi_2(X)$. Furthermore, with work one can show that:

- Any Gr-category G is equivalent to $\Pi_2(B\mathsf{G})$.
- Any connected pointed 2-type X is homotopy equivalent to $B(\Pi_2(X))$.

As a corollary, connected pointed 2-types are classified by Gr-categories. That is, two connected pointed 2-types are homotopy equivalent if and only if their fundamental 2-groups are equivalent as Gr-categories.

But one can equally well turn the logic around and say Gr-categories are classified by connected pointed 2-types. Two Gr-categories are equivalent if and only if their classifying spaces are homotopy equivalent. So, the wall separating algebra and topology has been completely knocked down, at least in this limited realm! This is a nice piece of evidence for Grothendieck's homotopy hypothesis.

7. Gr-categories in physics

While our story has been focused on Hoàng Xuân Sính's thesis and its immediate context, it would be a disservice not to mention that Gr-categories—now often called 2-groups [1]—later took on a new life in theoretical and mathematical physics. The reason is that 'gauge theory', the spectacularly successful approach to physics based on groups, can be generalized to 'higher gauge theory' using 2-groups [2, 3]. Just as gauge theory based on groups describes how point particles change state as they trace out 1-dimensional paths in spacetime, higher gauge theory based on 2-groups describes both particles and strings, the latter of which trace out 2-dimensional 'worldsheets' in spacetime. Furthermore, there is no reason to stop with 2-groups [37, 42]. While no theories of physics based on higher gauge theory have received any experimental confirmation, the mathematics behind these theories is magnificent, and there are hopes that someday physicists will synthesize phases of matter described by higher gauge theory [4, 7]. This would be a wonderful realization of Hoàng Xuân Sính's vision.

A. Equivalence for Gr-categories and Pic-categories

As we have seen, the concept of 'equivalence' is crucial for many of Hoàng Xuân Sính's results on Gr-categories and Pic-categories. This concept is subtler than isomorphism. We can see this already for categories. Two categories C and D are isomorphic if there are functors going back and forth, $F\colon \mathsf{C} \to \mathsf{D}$

and $G\colon \mathsf{D} \to \mathsf{C}$, that are inverses: $GF = 1_{\mathsf{C}}$ and $FG = 1_{\mathsf{D}}$. But experience has taught us than demanding *equations* between functors is too harsh: instead it suffices to have *natural isomorphisms* between them. So, we say C and D are **equivalent** if there are functors $F\colon \mathsf{C} \to \mathsf{D}$ and $G\colon \mathsf{D} \to \mathsf{C}$ for which there exist natural isomorphisms

$$\alpha\colon GF \stackrel{\sim}{\Longrightarrow} 1_{\mathsf{C}}, \quad \beta\colon FG \stackrel{\sim}{\Longrightarrow} 1_{\mathsf{D}}.$$

(We use a single arrow for functors and a double arrow for natural transformations.) For Gr-categories and Pic-categories, we need to enhance this concept of equivalence to take into account the extra structure that these categories carry.

As we have seen, Gr-categories are simply monoidal categories with extra properties. Similarly Pic-categories are symmetric monoidal categories with extra properties. So, it is enough to define notions of equivalence for monoidal and symmetric monoidal categories. We begin by defining monoidal functors. These do not need to preserve the tensor product and unit object strictly, but rather only up to natural isomorphism. However, these natural isomorphisms need to obey some laws of their own!

Definition A.1. *A functor $F\colon \mathsf{M} \to \mathsf{N}$ between monoidal categories is* ***monoidal*** *if it is equipped with:*

- *a natural isomorphism $\Phi_{x,y}\colon F(x) \otimes F(y) \xrightarrow{\sim} F(x \otimes y)$ and*
- *an isomorphism $\phi\colon I_{\mathsf{N}} \xrightarrow{\sim} F(I_{\mathsf{M}})$*

such that:

- *the following diagram commutes for any objects $x, y, z \in \mathsf{M}$:*

$$\begin{tikzcd}
(F(x) \otimes F(y)) \otimes F(z) \arrow[r, "\Phi_{x,y} \otimes 1_{F(z)}"] \arrow[d, "a_{F(x),F(y),F(z)}"'] & F(x \otimes y) \otimes F(z) \arrow[r, "\Phi_{x \otimes y, z}"] & F((x \otimes y) \otimes z) \arrow[d, "F(a_{x,y,z})"] \\
F(x) \otimes (F(y) \otimes F(z)) \arrow[r, "1_{F(x)} \otimes \Phi_{y,z}"'] & F(x) \otimes F(y \otimes z) \arrow[r, "\Phi_{x, y \otimes z}"'] & F(x \otimes (y \otimes z)),
\end{tikzcd}$$

- *the following diagrams commute for any object $x \in \mathsf{M}$:*

$$\begin{tikzcd}
I_{\mathsf{N}} \otimes F(x) \arrow[r, "\ell_{F(x)}"] \arrow[d, "\phi \otimes 1_{F(x)}"'] & F(x) \\
F(I_{\mathsf{M}}) \otimes F(x) \arrow[r, "\Phi_{I_{\mathsf{M}}, x}"'] & F(I_{\mathsf{M}} \otimes x) \arrow[u, "F(\ell_x)"']
\end{tikzcd}
\qquad
\begin{tikzcd}
F(x) \otimes I_{\mathsf{N}} \arrow[r, "r_{F(x)}"] \arrow[d, "1_{F(x)} \otimes \phi"'] & F(x) \\
F(x) \otimes F(I_{\mathsf{M}}) \arrow[r, "\Phi_{x, I_{\mathsf{M}}}"'] & F(x \otimes I_{\mathsf{M}}). \arrow[u, "F(r_x)"']
\end{tikzcd}$$

We similarly have a concept of 'symmetric monoidal functor'. Here the natural isomorphism Φ must get along with the symmetry:

Definition A.2. *A functor $F\colon \mathsf{M} \to \mathsf{N}$ between symmetric monoidal categories is **symmetric monoidal** if it is monoidal and it makes the following diagram commute for all $x, y \in \mathsf{M}$:*

$$\begin{tikzcd}
F(x) \otimes F(y) \arrow[r, "S_{F(x),F(y)}"] \arrow[d, "\Phi_{x,y}"'] & F(y) \otimes F(x) \arrow[d, "\Phi_{y,x}"] \\
F(x \otimes y) \arrow[r, "F(S_{x,y})"'] & F(y \otimes x)
\end{tikzcd}$$

Next we need monoidal *natural transformations*. Recall that a monoidal functor $F\colon \mathsf{M} \to \mathsf{N}$ is really a triple (F, Φ, ϕ) consisting of a functor, a natural isomorphism Φ, and an isomorphism ϕ. A 'monoidal natural transformation' must get along with these extra isomorphisms:

Definition A.3. *Suppose that (F, Φ, ϕ) and (G, Γ, γ) are monoidal functors from the monoidal category M to the monoidal category N. Then a natural transformation $\alpha\colon F \Rightarrow G$ is **monoidal** if the diagrams*

$$\begin{tikzcd}
F(x) \otimes F(y) \arrow[r, "\alpha_x \otimes \alpha_y"] \arrow[d, "\Phi_{x,y}"'] & G(x) \otimes G(y) \arrow[d, "\Gamma_{x,y}"] \\
F(x \otimes y) \arrow[r, "\alpha_{x \otimes y}"'] & G(x \otimes y)
\end{tikzcd}$$

and

$$\begin{tikzcd}
& I_{\mathsf{N}} \arrow[dl, "\phi"'] \arrow[dr, "\gamma"] & \\
F(I_{\mathsf{M}}) \arrow[rr, "\alpha_{I_{\mathsf{M}}}"'] & & G(I_{\mathsf{M}})
\end{tikzcd}$$

commute.

There are no extra conditions required of **symmetric monoidal** natural transformations: they are simply monoidal natural transformations between symmetric monoidal functors.

We can compose two monoidal natural transformations and get another monoidal natural transformation. Also, any monoidal functor F has an identity monoidal natural transformation $1_F\colon F \Rightarrow F$. This allows us to make the following definition:

Definition A.4. *A monoidal natural transformation $\alpha\colon F \Rightarrow G$ is a **monoidal natural isomorphism** if there is a monoidal natural transformation $\beta\colon G \Rightarrow F$ that is an inverse to α:*

$$\beta\alpha = 1_F \quad \text{and} \quad \alpha\beta = 1_G.$$

Now at last we are ready to define equivalence for Gr-categories and Pic-categories, so we can state some of Hoàng Xuân Sính's major results in a precise way. For this we need the fact that we can compose two monoidal functors and get a monoidal functor, and similarly in the symmetric monoidal case.

Definition A.5. *Two monoidal categories M and N are **equivalent** if there exist monoidal functors $F\colon \mathsf{M} \to \mathsf{N}$ and $G\colon \mathsf{N} \to \mathsf{M}$ such that there exist monoidal natural isomorphisms $\alpha\colon GF \Rightarrow 1_{\mathsf{N}}$ and $\beta\colon FG \Rightarrow 1_{\mathsf{N}}$. Two Gr-categories G and H are equivalent if they are equivalent as monoidal categories.*

Definition A.6. *Two symmetric monoidal categories M and N are **equivalent** if there exist symmetric monoidal functors $F\colon \mathsf{M} \to \mathsf{N}$ and $G\colon \mathsf{N} \to \mathsf{M}$ such that there exist symmetric monoidal natural isomorphisms $\alpha\colon GF \Rightarrow 1_{\mathsf{N}}$ and $\beta\colon FG \Rightarrow 1_{\mathsf{N}}$. Two Pic-categories G and H are equivalent if they are equivalent as symmetric monoidal categories.*

With these concepts in hand, we can state some results on the classification of Gr-categories and Pic-categories. We have already mentioned the following pair of results. The former is Prop. 3.22 in Hoàng Xuân Sính's thesis [23], while the latter is Prop. 5.3 in her 1978 paper on strict Gr-categories [25].

Theorem A.1. *Any Gr-category is equivalent to a skeletal Gr-category, i.e. one for which isomorphic objects are necessarily equal.*

Theorem A.2. *Any Gr-category is equivalent to a strict Gr-category, i.e. one for which the associator, left unitor and right unitor are all identity natural transformations.*

We have also seen how any skeletal Gr-category G gives a 4-tuple (G, A, ρ, α). The following result gives a classification of skeletal Gr-categories in these terms. It is a special case of Proposition 3.47 of Hoàng Xuân Sính's thesis [23]:

Theorem A.3. *Suppose G and G' are skeletal Gr-categories giving 4-tuples (G, A, ρ, α) and (G', A', ρ', α'), respectively. Then G and G' are equivalent if and only if there exist isomorphisms $\phi\colon G \to G'$ and $\psi\colon A \to A'$ such that:*

- *the actions ρ and ρ' are related as follows:*

$$\rho'(\phi(g))(\psi(a)) = \psi(\rho(g)(a))$$

for all $g \in G$ and $a \in A$,

- *the cohomology classes α and α' are related as follows:*

$$\alpha'(\phi(g_1), \phi(g_2), \phi(g_3)) - \psi(\alpha(g_1, g_2, g_3)) = df$$

  for some $f\colon G'^2 \to A$.

She stated a more general result that does not require G and G' to be skeletal, but since every Gr-category is equivalent to a skeletal one, a classification of skeletal Gr-categories up to equivalence can serve as a classification of all Gr-categories. For a deeper treatment of the relation between Gr-categories and group cohomology, see Section 6 of Joyal and Street's unpublished paper [29].

Interestingly, the unitors do not provide any extra information required for the classification of Gr-categories up to equivalence. The reason is that given any Gr-category G, we can change the tensor product by setting $I \otimes g = g = g \otimes I$ for every $g \in \mathsf{G}$, change the left and right unitors to identity morphisms, and adjust the associators to obtain a new Gr-category G' that equivalent to G. For a self-contained proof see [1, Prop. 39].

The following classification of restrained Pic-categories is Corollary 3.61 in Hoàng Xuân Sính's thesis [23]. She derived it as a consequence of a subtler classification of *all* Pic-categories, her Proposition 3.59.

Theorem A.4. *Suppose G and H are restrained Pic-categories. Then G and H are equivalent if and only if both these conditions hold:*

- *The abelian group of isomorphism classes of objects in G is isomorphic to the abelian group of isomorphism classes of objects in H.*
- *The abelian group of automorphisms of I_{G} is isomorphic to the abelian group of automorphisms of I_{H}.*

Later, in her 1982 paper "Catégories de Picard restreintes" [26], Hoàng Xuân Sính showed that every restrained Gr-category arises from a 2-term chain complex of abelian groups in the manner described at the end of Section 5:

Theorem A.5. *Suppose G is a restrained Pic-category. Then there is some 2-term chain complex C of abelian groups such that G is equivalent, as a Pic-category, to G_C.*

Acknowledgements

I thank Hoàng Xuân Sính for answering my questions about her thesis, Lisa Raphals for translating my questions into French, and David Egolf and David Michael Roberts for pointing out mistakes and ways in which this paper could be improved.

References

[1] **John C. Baez and Aaron D. Lauda,** Higher-dimensional algebra V: 2-groups, *Theor. Appl. Categ.* **12** (2004), 423–491. Also available as arXiv:math.QA/0307200.

[2] **John C. Baez and John Huerta,** An invitation to higher gauge theory, *Gen. Rel. Grav.* **43** (2011), 2335–2392. Also available as arXiv:1003.4485.

[3] **John C. Baez and Urs Schreiber,** Higher gauge theory, in *Categories in Algebra, Geometry and Mathematical Physics*, eds. Alexei Davydov, Michael Batanin, Michael Johnson, Stephen Lack and Amnon Neeman, Contemp. Math. 431, AMS, Providence, 2007, pp. 7–30. Also available as arXiv:math/0511710.

[4] **Maissam Barkeshli, Yu-An Chen, Po-Shen Hsin and Ryohei Kobayashi,** Higher-group symmetry in finite gauge theory and stabilizer codes, 2022. Available as arXiv:2211.11764.

[5] **Jean Bénabou,** Catégories avec multiplication, *Comptes Rendus des Séances de l'Académie des Sciences* **256** (1963), 1887–1890. Also available as https://gallica.bnf.fr/ark:/12148/bpt6k3208j/f1965.item.

[6] **Jean Bénabou,** Introduction to bicategories, in *Reports of the Midwest Category Seminar*, Springer, Berlin, 1967, pp. 1–77.

[7] **Arkadiusz Bochniak, Leszek Hadasz, Piotr Korcyl and Błażej Ruba,** Study of a lattice 2-group gauge model, in *Proceedings of the 38th International Symposium on Lattice Field Theory, LATTICE2021*, Proceedings of Science, 2022. Also available as arXiv:2109.12097.

[8] **Ronald Brown and C. B. Spencer,** $\mathcal{G}$-groupoids, crossed modules, and the classifying space of a topological group, *Proc. Kon. Akad. v. Wet.* **79** (1976), 296–302. Also available at https://core.ac.uk/download/pdf/82096733.pdf.

[9] **Manuel Bullejos and Antonio M. Cegarra,** On the geometry of 2-categories and their classifying spaces, *K-Theory* **29** (2003), 211-229. Also available at http://www.ugr.es/%7Ebullejos/geometryampl.pdf.

[10] **Jack Duskin,** Simplicial matrices and the nerves of weak n-categories I: nerves of bicategories, *Theory and Applications of Categories*, **9** (2001), 198–308. Available at http://www.tac.mta.ca/tac/volumes/9/n10/9-10abs.html.

[11] **Benno Eckmann and Peter H. Hilton,** Group-like structures in general categories. I. Multiplications and comultiplications, *Math. Ann.* **145** (3) (1962), 227–255.

[12] **Samuel Eilenberg and Saunders Mac Lane,** Relations between homology and homotopy groups of spaces,*Ann. Math.* **46** (2) (1945), 480–509.

[13] **Samuel Eilenberg and Saunders Mac Lane,** On the groups $H(\Pi, n)$, I, *Ann. Math.* **58** (1) (1953), 55–106.

[14] **Samuel Eilenberg and Saunders Mac Lane,** On the Groups $H(\Pi, n)$, II: Methods of computation, *Ann. Math.* **50** (1) (1954), 49–139.

[15] **Samuel Eilenberg and G. Max Kelly,** Closed categories, *Proceedings of the Conference on Categorical Algebra: La Jolla 1965*, Springer, 1966.

[16] **Magnus Forrester-Barker,** Group objects and internal categories. Available as math.CT/0212065.

[17] **Alexander Grothendieck,** *Pursuing Stacks*, 1984. Available as arXiv:2111.01000.

[18] **Alexander Grothendieck,** *Récoltes et Semailles: Réflexions et témoignages sur un passé de mathématicien*, Gallimard, Paris, 2022. Also available at https://webusers.imj-prg.fr/ leila.schneps/grothendieckcircle/writings.php.

[19] **K. A. Hardie, K. H. Kamps and R. W. Kieboom,** A homotopy 2-groupoid of a topological space, *Appl. Cat. Str.* **8** (2000) 209–234.

[20] **K. A. Hardie, K. H. Kamps and R. W. Kieboom,** A homotopy bigroupoid of a topological space, *Appl. Cat. Str.* **9** (2001) 311–327.

[21] **Simon Henry and Edoardo Lanari,** On the homotopy hypothesis in dimension 3. Available as arXiv:1905.05625.

[22] **Peter Hilton,** The birth of homological algebra, *Rocky Mountain J. Math.* **32** (4) (2002), 1101–1116. Also available at https://projecteuclid.org/journals/rocky-mountain-journal-of-mathematics/volume-32/issue-4/The-Birth-of-Homological-Algebra/10.1216/rmjm/1181070011.full.

[23] **Hoàng Xuân Sính,** *Gr-catégories*, Ph.D. thesis, Université Paris VII, 1973. Handwritten version available at https://pnp.mathematik.uni-stuttgart.de/lexmath/kuenzer/sinh.html. Version typeset by Cristian David Gonzalez Avilés available at https://agrothendieck.github.io/divers/GCS.pdf.

[24] **Hoàng Xuân Sính,** *Gr-categories: summary*. Handwritten version available at https://pnp.mathematik.uni-stuttgart.de/lexmath/kuenzer/sinh.html.

[25] **Hoàng Xuân Sính,** Gr-catégories strictes, *Acta Math. Vietnam.* **3** (2) (1978), 47–59. Also available at https://pnp.mathematik.uni-stuttgart.de/lexmath/kuenzer/Hoang_Xuan_Sinh_Gr_categories_strictes.pdf.

[26] **Hoàng Xuân Sính,** Catégories de Picard restreintes, *Acta Math. Vietnam.* **7** (1) (1983), 117–122. Also available at https://pnp.mathematik.uni-stuttgart.de/lexmath/kuenzer/Hoang_Xuan_Sinh_Categories_de_Picard_restreintes.pdf.

[27] **Hoàng Xuân Sính,** personal communication (translated from French), July 7, 2023.

[28] **G. Max Kelly,** On MacLane's conditions for coherence of natural associativities, commutativities, etc., *J. Alg.* **1** (1964), 397–402. Also available at https://www.sciencedirect.com/science/article/pii/0021869364900183.

[29] **André Joyal and Ross Street,** Braided monoidal categories, Macquarie Mathematics Report No. 860081, November 1986. Available at http://web.science.mq.edu.au/∼street/JS1.pdf.

[30] **Saunders Mac Lane,** Natural associativity and commutativity, *Rice University Studies* **49** (4) (1963), 28–46. Available as https://scholarship.rice.edu/handle/1911/62865.

[31] **Saunders Mac Lane,** Origins of the cohomology of groups, *Enseign. Math.* **24** (1978), 191–219.

[32] **Saunders Mac Lane,** Historical note, *Jour. Alg.* **60** (2) (1979), 319–320. Also available at https://www.sciencedirect.com/science/article/pii/0021869379900851.

[33] **Saunders Mac Lane,** Group extensions for 45 years, *Math. Intell.* **10** (2) (1988), 29–35.

[34] **Saunders Mac Lane,** *Categories for the Working Mathematician*, Springer, Berlin, 2013.

[35] **Saunders Mac Lane and J. H. C. Whitehead,** On the 3-type of a complex, *Proc. Nat. Acad. Sci.* **36** (1950), 41–48. Also available at https://www.pnas.org/doi/pdf/10.1073/pnas.36.1.41.

[36] **Fernando Muro and Andrew Tonks,** Unital associahedra, *Forum Math.* **26** (2) (2014), 593–620. Also available as arXiv:1110.1959.

[37] **Thomas Nikolaus, Urs Schreiber and Danny Stevenson,** Principal ∞-bundles: general theory, *J. Homotopy Relat. Struct.* **10** (4) (2015), 749–801. Also available as arXiv:1207.0248.

[38] **Behrang Noohi,** Notes on 2-groupoids, 2-groups and crossed-modules, *Homology Homotopy Appl.* **9** (1) 2008, 75–106. Also available as arXiv:math/0512106.

[39] **Joseph J. Rotman,** *Introduction to Homological Algebra*, Academic Press, New York, 1979.

[40] **Neantro Saavedra-Rivano,** *Catégories Tannakienes*, Ph.D. thesis, Université Paris VII, 1970.

[41] **Neantro Saavedra-Rivano,** *Catégories Tannakienes*, Springer Lecture Notes in Mathematics 265, Springer, Berlin, 1972. Related version available at https://eudml.org/doc/87193.

[42] **Urs Schreiber,** Differential cohomology in a cohesive infinity-topos, 2013. Available as arXiv:1310.7930.

[43] **Alexandru Solian,** Coherence in categorical groups, *Comm. Alg.* **9** (1981), 1039–1057.

[44] **James Stasheff,** Homotopy associative H-spaces I, II, *Trans. Amer. Math. Soc.* **108** (1963), 275–312.

[45] **An Thanh, Hoàng Xuân Sính với "Ý tưởng lãng mạn nhất cuộc đời",** *Thông Tin Đối Ngoại*, November 21, 2019. Available at https://web.archive.org/web/20190926174449/http:/tapchithongtindoingoai.vn/kieu-bao-huong-ve-to-quoc/gs-hoang-xuan-sinh-voi-y-tuong-lang-man-nhat-cuoc-doi-19467.

[46] **Todd Trimble,** Combinatorics of polyhedra for n-categories, September 5, 1999. Available as https://math.ucr.edu/home/baez/trimble/polyhedra.html.

[47] **Charles A. Weibel,** History of homological algebra, in *The History of Topology*, ed. Ioan M. James, Elsevier, 1999. Also available as https://conf.math.illinois.edu/K-theory/0245/.

[48] **J. H. C. Whitehead,** Note on a previous paper entitled 'On adding relations to homotopy groups', *Ann. Math.* **47** (1946), 806–810.

[49] **J. H. C. Whitehead,** Combinatorial homotopy II, *Bull. Amer. Math. Soc.* **55** (1949), 453–496.
Also available at https://projecteuclid.org/journals/bulletin-of-the-american-mathematical-society/volume-55/issue-5/Combinatorial-homotopy-II/bams/1183513797.full.

John C. Baez
Department of Mathematics, University of California
Riverside CA, 92521
USA
`Baez@math.ucr.edu`
and
Centre for Quantum Technologies, National University of Singapore
117543, Singapore

J. Math. Math. Sci. (2023) 37–55

A survey on solution approaches for the equilibrium problem defined by the Nikaido-Isoda-Fan inequality

Le Dung Muu (Hanoi, Vietnam)

Jean Jacques Strodiot (Namur, Belgium)

Xuan Thanh Le (Hanoi, Vietnam)

(Received Jul. 18, 2022)

Abstract. We provide a brief survey on basic solution approaches for solving the equilibirum problem defined by the Nikaido-Isoda-Fan inequality. Namely, first we state the problem and consider its most important special cases including the optimization, inverse optimization, Kakutani fixed point, variational inequality, Nash equilibrium problems. Next, we present some basic solution approaches for the problem. Finally, as an application, we consider the famous Cournot-Nash oligopolistic equilibrium model and discuss algorithms for solving it.

1. Introduction

Thoughout the paper let $\mathbb{H}$ be a real Hilbert space. In what follows we mainly work on the weak topology of $\mathbb{H}$. Let $C \subseteq \mathbb{H}$ be a closed convex set and $f : \mathbb{H} \times \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$. We suppose that $f(x,y) \in \mathbb{R}$ for every $x, y \in C$. As usual, we call f an equilibirum bifunction if $f(x,x) = 0$ for every $x \in C$. The problem to be considered in this paper is formulated as follows.

$$(EP) \qquad \text{Find } x^* \in C \text{ such that } f(x^*, y) \geq 0 \text{ for all } y \in C.$$

The inequality appeared in problem (EP) was first used by Nikaido and Isoda in 1955 [37] in a non-cooperative convex game. In the seminal paper [19] in 1972, Fan called

Key words and phrases: Nikaido-Isoda-Fan equilibrium problem, variational inequality, Cournot-Nash equilibrium model

2020 Mathematics Subject Classification: 65J15, 47H05, 47J25, 47J20

problem (EP) a minimax inequality and established solution existence results for it when C is convex, compact and f is quasiconvex on C. To our best knowledge, up to now there does not exist an algorithm for finding a solution of the problem considered in [19]. This result by Fan was extended by Brezis, Nirenberg, and Stampachia in [11]. In 1984, Muu [29] called (EP) a variational inequality and studied its some stability properties. In 1992, Muu and Oettli [32] called problem (EP) an equilibrium one, and a penalty algorithm was proposed for finding a solution of (EP) when f possesses certain monotonicity properties. After the appearance of the paper [10] by Blum and Oettli in 1994, problem (EP) attracted much attention of many authors, see e.g. the interesting monographs by Bigi et al. [8], mainly for solution technique issues in Hilbert spaces, and by Kassay et al. [24], mainly for theoretical aspects in vector topological spaces.

It worth mentioning that when $f(x,\cdot)$ is convex and subdifferentiable on C, the equilibrium problem (EP) can be reformulated as the following multivalued variational inequality.

$$(MultiVI) \qquad \text{Find } x^* \in C, v^* \in F(x^*) \text{ such that } \langle v^*, x - x^* \rangle \geq 0 \text{ for all } x \in C,$$

where $F(x^*) = \partial_2 f(x^*, x^*)$ with $\partial_2 f(x^*, x^*)$ being the diagonal subdifferential of f at x^*, that is the subdifferential of the convex function $f(x^*,\cdot)$ at x^*. In the case $f(x,\cdot)$ is semi-strictly quasiconvex rather than convex, problem (EP) can take the form of $(MultiVI)$ with $F(x) := Na_{f(x,x)} \setminus \{0\}$, where $Na_{f(x,x)}$ is the normal cone of the adjusted sublevel set of the function $f(x,\cdot)$ at the level $f(x,x)$, see [5]. More details about the links between equilibrium problems and variational inequalities can be found in [6].

In this paper we provide a brief survey on solution approaches for problem (EP) in real Hilbert spaces. Namely, in the next section we present some important special cases of problem (EP) such as optimization, reverse optimization, multivalued variational inequality, the Kakutani fixed point, the Walras and Nash equilibrium problems. The third section is devoted to the discussion of some basic solution approaches for problem (EP) involving bifunctions having certain monotonicity properties. We close the paper by showing a formulation of the Cournot-Nash oligopolistic equilibrium model in the form of equilibrium problem (EP) and discuss algorithms for solving the model for the convex and quasiconcave cost functions.

2. Special cases

Although the formulation of problem (EP) is very simple, it contains a lot number of important problems as special cases. In [32] it has been shown that the opimization, Kakutani fixed point and multivalued variational inequality problems can be formulated in the form of (EP). In [10] the Nash equilibrium problem has been formulated equivalently as a problem of the form (EP). Some other problems such as inverse optimization, vector optimization have been converted equivalently into problems of the form (EP), see e.g. [8]. For more detail, below we present the form (EP) for the Kakutani fixed point, multivalued variational inequality, Nash equilibrium and inverse optimization problems. The other ones can be found in e.g. [8, 10, 24].

- *Optimization.* Consider the minimization problem

$$(OP) \qquad \min_{x \in C} g(x)$$

in which $g : \mathbb{H} \to \mathbb{R}$ is a single-valued function and $C \subset \mathbb{R}$ is a closed convex set. Let $f(x,y) := g(y) - g(x)$ and consider the equilibrium problem

$$(EP_1) \qquad \text{Find } x^* \in C \text{ such that } f(x^*,y) \geq 0 \text{ for all } y \in C.$$

It is clear that $x^* \in C$ a global minimizer of (OP) if and only if

$$g(x^*) \leq g(y) \quad \forall y \in C,$$

or equivalently,

$$f(x^*,y) = g(y) - g(x^*) \geq 0 \quad \forall y \in C,$$

i.e., x^* is a solution to (EP_1). It means that the minimization problem (OP) is equivalent to the equilibrium problem (EP_1) in the sense that their solution sets coincide.

- *Kakutani fixed point*. Let $C \subset \mathbb{R}^n$ be a compact convex set, $F : C \rightrightarrows C$ an upper semicontinuous multi-valued mapping with convex, compact values. The Kakutani fixed point problem asks:

$$(KP) \qquad \text{Find } x^* \in C \text{ such that } x^* \in F(x^*).$$

The Kakutani fixed point theorem, which is one of famous fixed point ones, states that such a point x^* exists. In a special case, when F is single valued, this theorem becomes the Brouwer theorem that was proved in 1910. Up to now there does not exist an efficient algorithm for finding a fixed point for the Brouwer mapping. In 1967 Scarf, an economist, first developed an algorithm for finding a Brouwer fixed point in $\mathbb{R}^n$ (see [43]), but many computational experiments show that this algorithm and its modifications can only solve the problem with moderate dimension n. In order to formulate the Kakutani fixed point problem in the form of problem (EP), we define the bifunction $f : C \times C \to \mathbb{R}$ by taking

$$f_2(x,y) := \max_{u \in F(x)} \langle x - u, y - x \rangle$$

for each $x, y \in C$. It is shown in [32] that a point x^* solves the problem (KP), i.e. $x^* \in F(x^*)$, if and only if it is a solution to the following equilibrium problem

$$(EP_2) \qquad \text{Find } x^* \in C \text{ such that } f_2(x^*, y) \geq 0 \text{ for all } y \in C.$$

In other words, the Kakutani fixed point problem (KP) is equivalent to the equilibrium problem (EP_2) in the sense that their solution sets coincide.

- *Variational inequality problem.* Let $F : C \rightrightarrows \mathbb{H}$ be a (multivalued) operator with convex, (weakly) compact values and $\varphi : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ such that $\varphi(x)$ is finite on C. The mixed variational inequality problem stated in [18] is formulated as $(MixedVI)$
  Find $x^* \in C$ such that $\exists u^* \in F(x^*) : \langle u^*, y - x^* \rangle + \varphi(y) - \varphi(x^*) \geq 0 \ \forall y \in C$.

  Clearly, when F is single valued, this problem is reduced to the following one:

  $$\text{Find } x^* \in C \text{ such that } \langle F(x^*), y - x^* \rangle + \varphi(y) - \varphi(x^*) \geq 0 \ \forall y \in C.$$

  It is worth noting that, when C is a convex cone and φ is a constant, problem $(MixedVI)$ becomes the following complementarity one.

  $$(MC) \qquad \text{Find } x^* \in C \text{ such that } \exists u^* \in F(x^*) : \ \langle u^*, x^* \rangle = 0.$$

  Let
  $$f_3(x, y) := \max_{u \in F(x)} \langle u, y - x \rangle + \varphi(y) - \varphi(x)$$
  and consider the following equilibrium problem

  $$(EP_3) \qquad \text{Find } x^* \in C \text{ such that } f_3(x^*, y) \geq 0 \text{ for all } y \in C.$$

  It was proved in [32] that a point x^* is a solution of problem $(MixedVI)$ if and only if it is also a solution to (EP_3). Therefore, $(MixedVI)$ is equivalent to the equilibrium problem (EP_3) in the sense that they share the same solution set.

- *Nash equilibria*. In a noncooperative game with N players, each player i has a set of possible strategies $C_i \subseteq \mathbb{R}^{n_i}$ and aims at minimizing a cost function $g_i : C \to \mathbb{R}$ with $C := C_1 \times \ldots \times C_N$. By definition, a *Nash equilibrium point* is any point in C such that no player can reduce her/his cost by unilaterally changing her/his strategy. The Nash equilibrium problem is to find such a Nash equilibrium point, i.e. a point $x^* \in C$ such that

  $$g_i(x^*) \leq g_i(x^*[y_i]) \quad \forall y_i \in C_i, i = 1, \ldots, N,$$

  where $x^*[y_i]$ stands for the vector obtained from x^* by replacing the component x_i^* with y_i. If we take $f_4 : C \times C \to \mathbb{R}$ defined as

  $$f_4(x, y) := \sum_{i=1}^{N} \Big[g_i(x[y_i]) - g_i(x) \Big],$$

and consider the following equilibrium problem

$$(EP_4) \qquad \text{Find } x^* \in C \text{ such that } f_4(x^*,y) \geq 0 \text{ for all } y \in C,$$

then it is not hard to see that x^* is a solution to the Nash equilibrium problem if and only if it is a solution to (EP_4).

- *An inverse optimization*. Let $C_1 \subseteq \mathbb{R}^n$ and $C_2, C_3 \subseteq \mathbb{R}^m$ be convex sets and $g_j : C_2 \to \mathbb{R} (j = 1, \dots, m)$. Let

$$h_2(p,q) := \sum_{j=1}^{m} p_j g_j(q).$$

The inverse problem reads as
$(InvP)$
Find $p^* = (p_1^*, \dots, p_n^*)^T \in C_1$ such that $\arg\min\{h_2(p^*,q) \mid q \in C_2\} \cap C_3 \neq \emptyset$.

In some economics models p^* plays the role of a price that is required to be found such that the latter inclusion is satisfied. Clearly, this inverse problem can be formulated as a noncooperative game with three players. The first player controls p by choosing a point $p^* \in C_1$, the second one solves the problem $\min_{q \in C_2} f_2(p^*,q)$, while the third player controls her/his strategy in C_3. Of course one can extend this model by assuming that the first and third players have more general lost functions, say, $h_1(p,q,r)$ and $h_3(p,q,r)$. Following the equivalent of the Nash equilibrium problem in noncooperative game with (EP_4) as discussed above, the inverse optimization problem $(InvP)$ in turn can take the form of an equilibrium problem (EP).

3. Solution approaches

3.1. Basic solution existence

Under the condition $f(x,x) = 0$ for every $x \in C$, it follows immediately that x^* is a solution to problem (EP) if and only if $x^* \in \text{argmin}\{f(x^*,y) \mid y \in C\}$, i.e. x^* is a fixed point of the mapping $S(\cdot)$ with $S(x) = \text{argmin}\{f(x,y) \mid y \in C\}$. The first result for solution existence of the equilibrium problem (EP) is due to Fan [19] in 1972. There, Fan called (EP) a minimax inequality and established the following theorem. His proof was based upon the KKM Lemma (a variant of a fixed point theorem).

Theorem 3.1. *(see [19]). Let C be a compact, convex set in a Hausdorff topological vector space. Let $f : C \times C \to \mathbb{R}$ be a continuous bifunction such that for every $x \in C$*

we have $f(x,x) = 0$ and $f(x,\cdot)$ is quasiconvex on C. Then, the equilibrium problem (EP) is solvable, i.e., there exists $x^* \in C$ such that $f(x^*,y) \geq 0$ for every $y \in C$.

To our best knowledge, up to now there does not exist an efficient algorithm for approximating the solution mentioned in this theorem.

3.2. Some solution approaches

A key assumption for equilibrium problem (EP), that we assume in what follows, is that the bifunction f is convex with respect to its second variable on the feasible convex set C, i.e., $f(x,\cdot)$ is convex on C for any fixed $x \in C$. Under this main assumption, we have the following auxiliary problem principle.

3.2.1. Auxiliary problem principle and fixed point

The auxiliary problem principle first was introduced by Cohen [12] for the optimization and variational inequality problems and extended to the equilibrium problem [28].

Theorem 3.2. *(Auxiliary problem principle). Suppose that $f(x,\cdot)$ is subdifferentiable on C for every $x \in C$. Then a point x^* is a solution of problem (EP) if and only if it is also a solution to the following regularized equilibrium one*

$$(REP) \qquad \text{Find } x^* \in C \text{ such that } f_\rho(x^*,y) := f(x^*,y) + \frac{1}{2\rho}\|y - x^*\|^2 \geq 0 \ \forall y \in C,$$

where $\rho > 0$.

A main advantage of the regularized problem is that the bifunction $f_\rho(x,\cdot)$ is strongly convex on C, which implies that the mathematical program $\min\{f_\rho(x,y) \mid y \in C\}$ always admits a unique solution. Thus x^* is a solution of (EP) if and only if $x^* = s(x^*)$, where $s(x^*)$ is the unique solution of the strongly convex mathematical programming problem $\min\{f_\rho(x^*,y) \mid y \in C\}$, that means x^* is a fixed point of $s(\cdot)$.

It is worth noting that, under some continuity property of the bifunction f, the solution-map $s : C \to C$ is continuous, and therefore, by the Brouwer fixed point theorem, it has a fixed point whenever C is compact. In order to find a fixed point of this mapping, one needs additional assumptions to ensure that the mapping has a certain Lipschitz property such as contractive or nonexpansive. Under the properties, one can derive iterative scheme for approximating a fixed point of the map s.

For this purpose the following monotonicity concepts for a bifunction are commonly used [10], see also [7] Section 20.

Definition 3.1. *Let $f : \mathbb{H} \times \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ and $D \subseteq \mathbb{H}$ such that f is finite on D. The bifunction f is said to be* *(i) strongly monotone on D with modulus $\mu > 0$ (shortly μ-strongly monotone) if*

$$f(x,y) + f(y,x) \leq -\mu\|x-y\|^2 \quad \forall x,y \in D;$$

(ii) μ-strongly pseudomonotone on C if

$$f(x,y) \geq 0 \quad \Rightarrow \quad f(y,x) \leq -\mu\|x-y\|^2;$$

(iii) monotone on D if

$$f(x,y) + f(y,x) \leq 0 \quad \forall x,y \in D;$$

(iv) pseudomonotone on D if for all $x,y \in D$ we have

$$f(x,y) \geq 0 \quad \Rightarrow \quad f(y,x) \leq 0.$$

More types of monotonicity can be found in e.g. [9].

The monotonicity notions of a bifunction are generalizations of those for a (multi-valued) operator. We recall from [7, 42] that a (multi-valued) operator F with compact values is said to be

(i) strongly monotone on C with modulus $\mu > 0$ (shortly μ-strongly monotone) if

$$\langle u-v, x-y\rangle \geq \mu\|x-y\|^2 \quad \forall x,y \in C, u \in F(x), v \in F(y);$$

(ii) μ-strongly pseudomonotone on C if for all $x,y \in C, u \in F(x), v \in F(y)$ we have

$$\langle u, y-x\rangle \geq 0 \quad \Rightarrow \quad \langle v, y-x\rangle \leq \mu\|x-y\|^2;$$

(iii) monotone on C if

$$\langle u-v, x-y\rangle \geq 0 \quad \forall x,y \in C, u \in F(x), v \in F(y);$$

(iv) pseudomonotone on C if for all $x,y \in C, u \in F(x), v \in F(y)$ we have

$$\langle u, y-x\rangle \geq 0 \quad \Rightarrow \quad \langle v, y-x\rangle \geq 0.$$

Clearly the strongly monotonicity implies the monotonicity, which in turn implies the pseudomonotonicity.

The following Lipschitz-type for a bifunction, which is an extension of the Lipschitz property of a map, is often used.

Definition 3.2. *(see [27]) The bifunction f is said to be Lipschitz-type on D with the constants L_1, L_2 if*

$$f(x,y) + f(y,z) \geq f(z,x) - L_1\|x-y\|^2 - L_2\|y-z\|^2 \quad \forall x,y,z \in D.$$

Clearly, in the optimization case, when $f(x,y) = g(y) - g(x)$, then f is monotone and Lipschitz-type for any function g.

For the statement of the next lemma, we need the following definition.

Definition 3.3. *We say that a multi-valued mapping $F : \mathbb{H} \rightrightarrows \mathbb{H}$ is Lipschitz with Hausdorff distance on a closed convex set $C \subset \mathbb{H}$ if there exists a so-called Lipschitz constant $L > 0$ such that*

$$d_H(F(x), F(y)) \leq L\|x - y\| \quad \forall x, y \in C,$$

in which

$$d_H(F(x), F(y)) = \max\{\sup_{u \in F(x)} \inf_{v \in F(y)} \|u - v\|, \sup_{v \in F(y)} \inf_{u \in F(x)} \|u - v\|\}$$

is the Hausdorff distance between two sets $F(x)$ and $F(y)$.

For the multi-valued mixed variational inequality problem $(MixedVI)$, by taking

$$f(x,y) := \max_{u \in F(x)} \langle u, y - x\rangle + \varphi(y) - \varphi(x)$$

we have the following relationships.

Lemma 3.1. *(see [41]) (i) If F is Lipschitz with Hausdorff distance on C with Lipschitz constant L, then f is Lipschitz-type on C with constants $L_1 = \frac{L\xi}{2}, L_2 = \frac{L}{2\xi}$ with any $\xi > 0$.*

(ii) If F is monotone (resp., strongly monotone, pseudomonotone), then f is strongly monotone (resp., monotone, pseudomonotone) on C.

The following problem is called Minty (or dual) problem for (EP).

(DEP) Find $y^* \in C$ such that $f(x, y^*) \leq 0 \; \forall x \in C$.

The following theorem provides a relationship between (EP) and (DEP).

Theorem 3.3. *(see [29]). (i) If f is pseudomonotone on C, then every solution (if exists) of problem (EP) is also a solution of problem (DEP).*

(ii) Conversely, if $f(x,\cdot)$ is lower semicontinuous and for any $y \in C$ the function $f(\cdot,y)$ is hemicontinuous at zero (i.e., for any $x' \in C$ one has $\lim_{t \to 0^+} f(tx + (1-t)x', y) = f(x', y)$ for all $y \in C$), then every solution of problem (DEP) is also a solution of (EP).

Note that, since $f(x,\cdot)$ is convex, the solution set of the Minty problem is a convex set as it is the intersection of a family of convex sets of the type $\{y \in C \mid f(x,y) \leq 0\}$.

3.2.2. Contraction fixed point method

This method is based upon the Banach contraction fixed point theorem. Namely, we have the following theorem.

Theorem 3.4. *(see [33]). Suppose that*

(i) for each $x \in C$, the function $f(x,\cdot)$ is convex, subdifferentiable on C;

(ii) f is μ-strongly monotone and Lipschitz type with constants L_1, L_2 on C. Then one can choose regularization $\rho > 0$ (depending on μ and the Lipschitz constants L_1, L_2) such that the mapping $s(\cdot) : C \to C$ defined by $s(x) = argmin\{f_\rho(x,y) \mid y \in C\}$ is contractive on C. Consequently, for any starting point $x^0 \in C$, the sequence $\{x^k\}$ is defined by $x^{k+1} = s(x^k)$ satisfying

$$\|x^{k+1} - x^*\| \leq \alpha \|x^k - x^*\| \quad \forall k \geq 0,$$

provided that $0 < \rho < 1/(2L_2)$ and $\alpha = 1 - 2\rho(\mu - L_1)$, where x^* is the unique solution of problem (EP).

Note that we can replace the regularization function $\|\cdot\|^2$ in (REP) by any strongly differentiable convex one (Bregman function, for instance). This contraction method can be extended to the case f is strongly pseudomonotone in e.g. [16]. Note furthermore that, under the assumption of the above theorem, problem (EP) always admits a unique solution even the feasible set C may not be compact (see [10]). Some other results for solution existence of problem (EP) can be found in e.g. [10] and the monographs [8, 24, 25].

For the variational inequality problem concerning single-valued mapping F:

$$(VI) \qquad \text{Find } x^* \in C \text{ such that } \langle F(x^*), y - x^* \rangle \geq 0 \ \forall y \in C,$$

we have $s(x) = P_C(x - \frac{1}{2\rho}F(x))$, where P_C stands for the metric projection onto the closed convex set C. In this case we have $x^{k+1} = P_C(x^k - \frac{1}{2\rho}F(x^k))$. It is well known that if F is merely monotone (not strongly monotone or not strongly pseudomonotone), the sequence of the iterates $\{x^k\}$ may not be convergent. For multivalued monotone variational inequality problems, an algorithm by coupling the Banach iterative scheme and the proximal point method was developed in [2].

3.2.3. Extragradient method

The extragradient method was introduced by Korpelevich [26] for optimization and saddle point problems. Then it has been extended to the equilibrium problem see e.g. [41]. Namely, we have the following results.

Theorem 3.5. *(see [41]). Suppose that the bifunction f is subdifferentiable, pseudomonotone, and Lipschitz-type on C with constants L_1, L_2, while $f(\cdot,y)$ is upper* *semicontinuous for each $y \in C$. The sequence $\{x_k\}$ of iterates defined by*

$$y^k = argmin\{f_\rho(x^k, y) := f(x^k, y) + \frac{1}{2\rho}\|y - x^k\|^2 \mid y \in C\},$$

$$x^{k+1} = argmin\{f_\rho(y^k, y) := f(y^k, y) + \frac{1}{2\rho}\|y - x^k\|^2 \mid y \in C\}$$

converges to a solution of (EP) provided $0 < \rho < \min\{1/(2L_1), 1/(2L_2)\}$.

Note that, as before, we can replace the regularization function $\|\cdot\|^2$ by any Bregman one.

In order to avoid the Lipschitz-type condition, a linesearch extragradient algorithm has been described in [41] and its convergence has been proved. Recently, in [21], an algorithm, where the stepsize is updated at each iteration (without linesearch), for solving pseudomonotone equilibrium problem has been developed.

As we have seen, equilibrium problem (EP) can be formulated equivalently as a fixed point problem. When the bifunction is strongly monotone, the fixed point map is contractive. The following results in [4] show that when f posseses certain monotonicity property, problem (EP) can be formulated as a fixed point problem with the map having certain nonexpansive or generalized nonexpansive property. For this purpose, let us define two mappings, the proximal mapping and the composited mapping. The proximal mapping is denoted by T_ρ and defined as the solution set of the regularized strongly monotone equilibrium problem

$$\text{Find } z \in C \text{ such that } f(z, y) + \frac{1}{2\rho}\langle y - z, z - x\rangle \geq 0 \ \forall y \in C.$$

For this mapping we have the following theorem

Theorem 3.6. *(see [7]) Suppose that*

(i) the solution set $S(EP)$ of problem (EP) is not empty;

(ii) $f(\cdot, y)$ is upper semicontinuous and $f(x, \cdot)$ is lower semicontinuous, convex on C for every $x, y \in C$.
Then for any $\rho > 0$, the mapping T_ρ is defined everywhere, single valued, and firmly nonexpansive, i.e.,

$$\|T_\rho(x) - T_\rho(y)\|^2 \leq \langle T_\rho(x) - T_\rho(y), y - x\rangle \quad \forall x, y \in C.$$

Moreover the solution set of (EP) coincides with the fixed point set of T_ρ.

The composited mapping is defined for each $x \in C$ by taking

$$C_\rho(x) := \text{argmin}\{f(B_\rho(x), y) + \frac{1}{2\rho}\|y - z\|^2 \mid y \in C\},$$

where

$$B_\rho(x) := \text{argmin}\{f(x, y) + \frac{1}{2\rho}\|y - x\|^2\}.$$

Theorem 3.7. *(see [4]). Assume that*

(i) The solution set $S(EP) \neq \emptyset$;

(ii) f is subdifferentiable and satisfies the Lipschitz-type with constants L_1, L_2.

(iii) f is jointly continuous on an open set containing $C \times C$.

Then C_ρ is quasinonexpansive and demiclosed at 0 provided that

$$0 < \rho < \min\{1/(2L_1), 1/(2L_2)\}.$$

A survey on the relationship between the fixed point and the equilibrium problem (EP) can be found in [30].

3.2.4. The proximal and Tikhonov regularization methods

The equilibrium problem (EP), in general, has many solutions, so it is a ill-posed one. The two main regularization methods commonly used to handle ill-posedness are the Tikhonov and proximal ones. The Tikhonov and proximal point regularization methods have been used to various problems in different fields of applied mathematics. These methods have been extended by Moudafi in [28]. The key idea of these methods is of the use of a suitable regularization bifunction to define regularized equilibrium problems depending on regularization parameters, thereby to obtain a trajector that converges to a solution of the original problem whenever the parameter tends to a suitable value.

In a regularization method a sequence of regularized equilibrium problems is defined, at each iteration k, as

$$\text{Find } x^{\rho_k} \in C \text{ such that } f_{\rho_k}(x^{\rho_k}, y) := f(x^{\rho_k}, y) + \frac{1}{2\rho_k} g_k(x^{\rho_k}, y) \geq 0 \ \forall y \in C, \tag{3.1}$$

where $\rho_k > 0$ (regularization parameter) and g_k is a strongly monotone bifunction.

First we consider the Tikhonov regularization method, where the regularized problem is defined with $\frac{1}{2\rho_k} = c_k$ and $g_k := \langle x - x^g, y - x\rangle$ (does not depend on k) and x^g is a guessed solution. Then the regularized problem takes the form

$$(TREP_k) \quad \text{Find } x^k \in C \text{ such that } f_{c_k}(x^k, y) := f(x^k, y) + c_k\langle x^k - x^g, y - x^k\rangle \geq 0 \ \forall y \in C.$$

We make use the following assumptions.

(A1) $f(\cdot, y)$ is (weakly) upper semicontinuous for each $y \in C$;

(A2) $f(x, \cdot)$ is lower semicontinuous and convex for each $x \in C$.

Then we have the following result.

Theorem 3.8. *(see [23]) Suppose that f is monotone on C. Then problem $(TREP_k)$ is strongly monotone (hence always admits a unique solution x^k) and x^k converges strongly to some x^* with $c_k \searrow 0$ as $k \to +\infty$.*

Unlike the Tikhonov regularization, in the proximal regularization the regularized bifunction at each iteration k depends on the previous iterate, such a bifunction often used is

$$f_k(x,y) := f(x,y) + c_k\langle x - x^{k-1}, y - x\rangle, \text{ where } c_k > 0.$$

For the proximal regularization method we have the following convergence result.

Theorem 3.9. *([23]) Suppose that f is monotone on C and Assumptions (A1), (A2) are satisfied. Then, for each k, the regularized problem*

$(PREP_k)$ *Find $x^k \in C$ such that $f_k(x^k,y) := f(x^k,y) + c_k\langle x^k - x^{k-1}, y - x^k\rangle \geq 0\ \forall y \in C$*

is strongly monotone (hence always admits a unique solution). Furthermore, x^k converges weakly to some x^* as $k \to +\infty$ and $c_k \to c < +\infty$.

In the case f is pseudomonotone, but not monotone, since the sum of a pseudomonotone and a strongly monotone bifunctions may not be monotone, even not pseudomonotone, the regularized problems for both Tikhonov and proximal regularization methods may have many solutions. However, any trajector converges to the same solution as shown in the following theorem.

Theorem 3.10. *(see [22]). Suppose that f is pseudomonotone on C and satisfies Assumptions (A1), (A2). Suppose furthermore that the solution sets of the original problem (EP) and each regularized problem $(TREP_k)$ are nonempty. Let x^k be any solution of the regularized problem $(TREP_k)$. Then the sequence $\{x^k\}$ converges strongly to the unique solution of the strongly monotone equilibrium problem*

(BEP) *Find $x \in S$ such that $g(x,y) \geq 0\ \forall y \in S$*

that is nearest to x^g, where S denotes the solution set of the original problem (EP) and $g(x,y) := \langle x - x^g, y - x\rangle$.

This result allows the bilevel level methods can be applied to the regularized pseudomonotone problem (EP) by solving the strongly monotone equilibrium problem (BEP). Since S is closed convex and g is strongly monotone, problem (BEP) always admits a unique solution. An algorithm for solving (BEP) was developed in [14].

For the proximal regularization method, we have a similar result, namely as follows.

Theorem 3.11. *(see [22]). Under the assumptions of Theorem 3.10, the sequence $\{x^k\}$ with x^k being any solution of the regularized problem $(PREP_k)$ converges weakly to a solution of problem (EP) provided $0 < c_k \to c < +\infty$.*

3.2.5. The gap function method

An important solution approach to equilibrium problem is based upon formulations of it in the form of a mathematical programming problem by using a gap function. We recall that $g : C \to \mathbb{R}$ is called a *gap function* for problem (EP) if $g(x) \geq 0$ for every $x \in C$ and $g(x) = 0$ if and only if x solves (EP). The first gap function is called the Auslender gap function which is defined as $g(x) := -\min\{f(x,y) \mid y \in C\}$. Clearly it is a gap function thanks to the condition $g(x,x) = 0$ for every $x \in C$. The main disadvantage of this gap function is that the problem defining it may not be solvable, and if yes, its solutions may not be unique. In order to overcome this disadvantage, Fukushima in [20] defined the following gap function that is called regularization gap function by taking for $x \in C, \rho > 0$, the function

$$g(x) := -\min\{f(x,y) + \frac{1}{2\rho}\|y-x\|^2 \mid y \in C\}.$$

Since the objective function of this optimization problem is strongly convex, it is always uniquely solvable. It is easy to see that it indeed is a gap function for (EP). The gap functions allow that methods of mathematical programming could be applied to solve equilibrium problems. However, since the gap function is not convex in general, finding its global minimum is a difficult task. Algorithms using a gap function for Minty equilibrium problem were proposed in [39, 40]. Other algorithms for Minty equilibrium problem can be found e.g. [15]. Some algorithms by coupling the extragradient method with the bundle, inertial (ball heavy), interior, ergodic and splitting techniques have been proposed in [1, 3, 17, 36, 38, 44] for solving pseudomonotone problem (EP).

4. Cournot-Nash oligopolistic equilibrium model

An important model for the Nash equilibria in economics is the Cournot-Nash oligopolistic model. This model was first introduced in [13] in 1838 by Cournot, a French economist, then it has been extended by using the famous Nash equilibrium concept.

An oligopolistic market model concerns with n firms (producers) that produce a common homogeneous commodity. Each firm has a profit function which is the difference between the income defined by the price and the cost. Each firm attempts to maximize its profit by choosing the corresponding production level on its strategy set.

In the classical model the price for all firms and the cost for each firm are given respectively as

$$p(x) := \alpha - \beta \sum_{j=1}^{n} x_j \text{ and } c_j(x_j) = \xi_j x_j + \eta_j,$$

where $\alpha > 0$ (in general is large) and $\beta > 0$ (often is small), $\xi_j > 0, \eta_j > 0$). So the price depends on the sum of the commodity, while the cost for each firm depends only on the amount of the commodity that it produces. Then the profit of each firm j is

$$f_j(x) := p(x)x_j - c_j(x_j) \quad (j = 1, \ldots, n). \tag{4.1}$$

Actually, each firm seeks to maximize its profit by choosing the corresponding production level under the presumption that the production of the other firms are parametric input. A commonly used approach to this model is based upon the famous Nash equilibrium concept. A point (strategy) $x^* = (x_1^*, \ldots, x_n^*)^T \in C$ is said to be a Nash equilibrium point of this Cournot-Nash oligopolistic market model if

$$f_j(x^*) \geq f_j(x^*[x_j]) \quad \forall j = 1, \ldots, n, \ \forall x_j \in C_j. \tag{4.2}$$

It has been shown [31] that, mathematically, the problem of finding a Nash equilibrium strategy for an oligopolistic market model with the profit function of each firm being given by (4.1) can be formulated in the following mixed equilibrium problem.

(MEP) Find $x^* \in C$ such that $f(x,y) := \langle B_1 x - \alpha, y - x\rangle + \beta \Big[\sum_{j=1}^{n} y_j^2 - x_j^2\Big] \geq 0 \ \forall y \in C,$

where B_1 is the $(n \times n)$-matrix whose every diagonal entry is zero and the others are all β. It is well known (see [31]) that for this classical model, problem (MEP) can be formulated equivalently in the form of a strongly convex quadratic program. In the case there is a convex cost function rather than all are affine, the model can be formulated as a monotone equilibrium problem (see e.g. [33]). In practice, since the cost for producing a unit commodity does decrease as the amount of the commodity gets larger, the cost is a concave increasingly function, the model then can be formulated as the mixed equilibrium problem
$(MEP1)$

Find $x^* \in C$ such that $f(x,y) := \langle B_1 x - \alpha, y - x\rangle + \beta \Big[\sum_{j=1}^{n} c_j(y_j) - c_j(x_j)\Big] \geq 0 \ \forall y \in C.$

When c_j is concave even for only one j, the function $f(x,\cdot)$, in general, is neither convex nor quasiconvex, and therefore a local equilibrium point may not be a global one. An algorithm for finding a stationary point of this nonconvex equilibrium problem was developed in [34], whereas a branch-and-bound algorithm using global optimization techniques for approximating the equilibrium problem $(MEP1)$ was proposed in [35]. Recently in [45] an algorithm for solving problem $(MEP1)$ with the bifunction f is quasiconvex and pseudo-paramonotone.

5. Conclusion

We have outlined some basic solution methods for solving equilibrium problems under the two main assumptions that the bifunction involved possesses certain monotonicity property and is convex in its second variable. Namely, we have shown how to formulate the problem in the form of a fixed point one that satisfies a suitable contraction property or its generalized nonexpansiveness. We have also presented the auxiliary problem principle, the regularization techniques as well as extragradient and gap function methods. Unfortunately, these solution methods may fail to apply directly to the problems, where the bifunction involved is quasiconvex rather than convex. In our opinion, research on efficient algorithms for finding a solution of the equilibrium problem whose solution existence has been proved by Ky Fan would be very interesting. Further research for the following subjects might be of interest.

- Development of new more efficient algorithms for problem (EP);
- Solution algorithms for convex split feasibility problem of finding $x^* \in C, F(x^*) \in Q$ with C and/or Q being given implicitly as the solution sets of certain equilibrium problems;
- Extensions of the above mentioned methods to vector and set-valued equilibrium problems;
- Applications of problem (EP) to study models in game theory and in optimal control;
- Solution algorithms for practical equilibrium models encounted in economics, environments, and other fields by using the form of problem (EP);
- Solution methods for equilibrium problems where the bifunction is monotone (not paramonotone) and quasiconvex with respect to its second variable.

References

[1] **Anh, P. K. and Hai, T. N.,** A splitting algorithm for equilibrium problem given by the difference of two bifunctions, *Journal of Fixed Point Theory and Applications*, **20** (2018), 53.

[2] **Anh, P. N., Muu, L. D., Nguyen, V. H., and Strodiot, J. J.,** Using the Banach contraction principle to implement the proximal point method for multivalued monotone variational inequalities, *Journal of Optimization Theory and Applications*, **124** (2003), 285–306.

[3] **Anh, P. N., Hai, T. N., and Tuan, P. M.,** On ergodic algorithms for equilibrium problems, *Journal of Global Optimization*, **64** (2016), 179–195.

[4] **Anh, T. V. and Muu, L. D.,** Quasi-nonexpansive mappings involving pseudomonotone bifunctions on convex sets, *Journal of Convex Analysis*, **25(4)** (2018), 1105–1119.

[5] **Aussel, D.,** Adjusted sublevel set, normal operator, and quasi-convex programming, *SIAM Journal on Optimization*, **16(2)** (2005), 358–367.

[6] **Aussel, D., Dutta, J., and Pandit, T.,** About the links between equilibrium problems and variational inequalities, Pages 115–130 in Neogy, S., Bapat, R., and Dubey, D. (editors), *Mathematical Programming and Game Theory*, Indian Statistical Institute Series, Springer, 2018.

[7] **Bauschke, H. and Combettes, P.,** *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[8] **Bigi, G., Castellani, M., Pappalardo, M., and Passacantando, M.,** *Nonlinear Programming Techniques for Equilibria*, Springer, 2019.

[9] **Bigi, G. and Passacantando, M.,** Descent and penalization techniques for equilibrium problems with nonlinear constraint, *Journal of Optimization Theory and Applications*, **164** (2015), 804–818.

[10] **Blum, E. and Oettli. W.,** From optimization and variational inequalities to equilibrium problems, *Mathematics Student*, **63(1)** (1994), 123–145.

[11] **Bresis, H., Nirenberg, L., and Stampachia, G.,** A remark on Ky Fan's minimax principle, *Bollletino della Unione Matematica Italiana*, **6** (1972), 293–300.

[12] **Cohen, G.,** Auxiliary problem principle extended to variational inequalities, *Journal of Optimization Theory and Applications*, **59** (1988), 325–333.

[13] **Cournot, A. A.,** *Recherches sur les principles mathematiques de la theorie des richeness*. (English translation by Bacon, N. T.: *Researches into the Mathematical Principles of the Theory of Wealth*), New York: Macmillan, 1927.

[14] **Dinh, B. V., Hung, P. G., and Muu, L. D.,** Bilevel optimization as a regularization approach to pseudomonotone equilibrium problems, *Numerical Functional Analysis and Optimization*, **35(5)** (2014), 539–563.

[15] **Dinh, B. V. and Kim, D. S.,** Projection algorithms for solving nonmonotone equilibrium problems in Hilbert space, *Journal of Computational and Applied Mathematics*, **302** (2016), 106–117.

[16] **Duc, P. M., Muu, L. D., and Quy, N. V.,** Solution existence and algorithms with their convergence rate for strongly pseudomonotone equilibrium problems, *Pacific Journal of Optimization*, **12(4)** (2016), 833–845.

[17] **Duc, P. M. and Le, X. T.,** A splitting subgradient algorithm for solving equilibrium problems involving the sum of two bifunctions and application to Cournot-Nash model, *RAIRO Operations Research*, **55** (2021), S1395–S1410.

[18] **Facchinei, F. and Pang, J. S.,** *Finite Dimensional Variational Inequalities and Complementarity Problems*, Springer, Berlin, 2002.

[19] **Fan, K.,** A minimax inequality and applications, pages 103–113 in Shisha, O. (editor), *Inequalities*, Academic Press, New York, 1972.

[20] **Fukushima, M.,** A class of gap functions for quasi-variational inequality problems, *Journal of Industrial and Management Optimization*, **3(2)**, (2007), 165–174.

[21] **Hieu, D. V., Strodiot, J. J., and Muu, L. D.,** Strongly convergent algorithms by using new adaptive regularization parameter for equilibrium problems, *Journal of Computational and Applied Mathematics*, **376** (2020), 112844.

[22] **Hung, P. G. and Muu, L. D.,** The Tikhonov regularization extended to equilibrium problem involving pseudomonotne bifunctions, *Nonlinear Analysis: Theory, Methods & Applications*, **74(17)** (2011), 6121–6129.

[23] **Hung, P. G. and Muu, L. D.,** On inexact Tikhonov and proximal point regularization methods for pseudomonotone equilibrium problem, *Vietnam Journal of Mathematics*, **40** (2012), 255–274.

[24] **Kassay, G. and Rădulescu, V. D.,** *Equilibrium Problems and Applications*, Academic Press, 2018.

[25] **Konnov, I.,** *Combined Relaxation Methods for Variational Inequalities*, Springer, New York, 2001.

[26] **Korpelevich, G. M.,** The extragradient method for finding saddle points and other problems, *Ekonomika i Matematicheskie Metody*, **12** (1976), 747–756.

[27] **Mastroeni, G.,** Gap functions for equilibrium problems, *Journal of Global Optimization*, **27** (2004), 411–426.

[28] **Moudafi, A.,** Proximal methods for a class of bilevel monotone equilibrium problems, *Journal of Global Optimization*, **47** (2010), 287–292.

[29] **Muu, L. D.,** Stability property of a class of variational inequalities, *Optimization*, **15(3)** (1984), 347–351.

[30] **Muu, L. D. and Le, X. T.,** On fixed point approach to equilibrium problem, *Journal of Fixed Point Theory and Applications*, **23** (2021), 50.

[31] **Muu, L. D., Nguyen, V. H., and Quy, N. V.,** On Nash-Cournot oligopolistic market models with concave cost functions, *Journal of Global Optimization*, **41** (2008), 351–364.

[32] **Muu, L. D. and Oettli, W.,** Convergence of an adaptive penalty scheme for finding constrained equilibria, *Nonlinear Analysis*, **18(12)** (1992), 1159–1166.

[33] **Muu, L. D. and Quoc, T. D.,** Regularization algorithms for solving monotone Ky Fan inequalities with application to a Nash-Cournot equilibrium model, *Journal of Optimization Theory and Applications*, **142** (2009), 185–204.

[34] **Muu, L. D. and Quoc, T. D.,** One step from DC optimization to DC mixed variational inequalities, *Optimization*, **59(1)** (2010), 63–76.

[35] **Muu, L. D. and Quy, N. V.,** Global optimization from concave minimization to concave mixed variational inequality, *Acta Mathematica Vietnamica*, **45** (2020), 449–462.

[36] **Nguyen, T. T. V., Strodiot, J. J., and Nguyen, V. H.,** A bundle method for solving equilibrium problems, *Mathematical Programming*, **116** (2009), 529–552.

[37] **Nikaido, H. and Isoda, K.,** Note on noncooperative convex games, *Pacific Journal of Mathematics*, **5(5)** (1955), 807–815.

[38] **Pham, K. A. and Hai, T. N.,** Splitting extragradient-like algorithms for strongly pseudomonotone equilibrium problems, *Numerical Algorithms*, **76** (2017), 67–91.

[39] **Quoc, T. D., Anh, P. N., and Muu, L. D.,** Dual extragradient algorithms extended to equilibrium problems, *Journal of Global Optimization*, **52** (2012), 139–159.

[40] **Quoc, T. D. and Muu, L. D.,** Iterative methods for solving monotone equilibrium problems via dual gap functions, *Computational Optimization and Applications*, **51** (2012), 709–718.

[41] **Quoc, T. D., Muu, L. D., and Nguyen, V. H.,** Extragradient algorithms extended to equilibrium problems, *Optimization*, **57(6)** (2008), 749–776.

[42] **Rockafellar, R. T.,** Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization*, **14(5)** (1976), 877–899.

[43] **Scarf, H. E.,** The approximation of fixed points of a continuous mapping, *SIAM Journal on Applied Mathematics*, **15(5)** (1967), 1328–1343.

[44] **Vinh, N. T. and Muu, L. D.,** Inertial extragradient algorithms for solving equilibrium problems, *Acta Mathematica Vietnamica*, **44** (2019), 639–663.

[45] **Yen, L. H. and Muu, L. D.,** A subgradient method for equilibrium problems involving quasiconvex bifunctions, *Operations Research Letters*, **48(5)** (2020), 579–583.

Le Dung Muu (Corresponding author)
Thang Long Institute of Mathematics and Applied Science, Thang Long University
Hanoi
Vietnam
`ldmuu@math.ac.vn`

Jean Jacques Strodiot
Mathematics Department, University of Namur
Namur
Belgium
`jean-jacques.strodiot@unamur.be`

Xuan Thanh Le
Institute of Mathematics, Vietnam Academy of Science and Technology
Hanoi
Vietnam
`lxthanh@math.ac.vn`

J. Math. Math. Sci. (2023) 57–65

A generalization of product inequality for the higher topological complexity

Nguyen Viet Dung (Hanoi, Vietnam)

Nguyen Van Ninh (Thai Nguyen, Vietnam)

(Received Oct. 22, 2022)

Abstract. In [4] M. Farber defined the topological complexity $TC(X)$ of a path-connected space X. Generalizing this notion, ten years later, Yu. Rudyak introduced a sequence of invariants, called the higher topological complexities $TC_n(X)$, for any path-connected space X in [7]. These invariants have their origin in the notion of the Schwarz genus of a fibration defined in [8]. One of the tools used to calculate these invariants is the product inequality for the Schwarz genus. In this paper, we will give a generalization of the product inequality of the higher topological complexity.

1. Introduction

Let X be a path-connected topological space, PX the space of all continuous paths $\gamma : I = [0,1] \to X$ with the usual compact-open topology.

Let's consider the map

$$\begin{array}{rccc} \pi : & PX & \longrightarrow & X \times X \\ & \gamma & \longmapsto & (\gamma(0), \gamma(1)) \end{array}.$$

Key words and phrases: higher topological complexity, configuration space, Schwarz genus, product inequality.

2020 Mathematics Subject Classification: 55R80, 55M30

The first author is supported by the grant NCVCC01.16/22-22

In [4], M. Farber defined the topological complexity $TC(X)$ of X as the smallest number k such that there exists an open covering $\{U_i, i = 1, ..., k\}$ of $X \times X$ with a continuous section $s_i : U_i \to PX$ of π on each U_i, i.e. $\pi \circ s_i = id_{U_i}$.

In 2010 Yu. Rudyak (see [7]) introduced a series of invariants, denoted by $TC_n(X)$, $n \geq 2$, for any path-connected space X. $TC_n(X)$ is called the higher topological complexity of X. It coincides with the topological complexity $TC(X)$ defined by M. Farber when $n = 2$.

Since then, the higher topological complexity has been computed for many topological spaces like spheres in [7] and product of spheres in [1], wedge sum of spheres of different dimensions in [3], configuration spaces on Euclidean spaces in [5], configuration space of distinct ordered points on compact Riemann surfaces of genus g in [6], the complement of some classes of complex hyperplane arrangements in [2].

Being closely related to the notion of the Schwarz genus and the Lusternik-Shnirelman category, the higher topological complexity inherited many interesting properties of these invariants. Some of these properties such as some evaluations from above or from below are used in computing the higher topological complexity TC_n.

One of the important evaluations of higher topological complexity is the product inequality saying that if X and Y are path-connected spaces, then

$$TC_n(X \times Y) \leq TC_n(X) + TC_n(Y) - 1.$$

In this paper, we give a generalization of this inequality. The paper is organized as follows. In section 2 we investigate the higher topological complexity and some of its properties. We formulate and prove our main result in section 3.

2. Higher topological complexity

For $n = 2, 3, \ldots$ let J_n denote the wedge sum of n closed unit intervals $[0, 1]_i$, $i = 1, \ldots, n$ with 0 as attached point. Suppose that X^{J_n} denotes the space of all continuous maps $\gamma : J_n \to X$ with compact-open topology. Consider the map

$$\begin{array}{rccc} e_n^X : & X^{J_n} & \longrightarrow & X^n , \\ & \gamma & \longmapsto & (\gamma(1_1), \ldots, \gamma(1_n)) \end{array}$$

where 1_i is the unit in $[0, 1]_i$ respectively.

Definition 2.1 (see [7]). *The higher topological complexity $TC_n(X)$ of the space X is the smallest number k such that there is an open covering $\{U_i, i = 1, \dots, k\}$ of X^n and there exists a continuous section $s_i : U_i \to X^{J_n}$ of e_n^X, on each U_i, i.e., $e_n^X \circ s_i = id_{U_i}$.*

Obviously, when $n = 2$, $TC_2(X)$ coincides the topological complexity $TC(X)$ defined by M.Faber.

Remark 2.1. *It is known that the map e_n^X is a fibration in the sense of Serre. By definition, the higher topological complexity $TC_n(X)$ of the space X is exactly the Schwarz genus of the fibration e_n^X (see [8]). Moreover, as it is indicated in [7], e_n^X is a fibrational substitute of the diagonal map d_n, i.e. there exists a homotopy equivalence $h : X \to X^{J_n}$ such that $d_n = e_n^X \circ h$. Therefore, $TC_n(X)$ is also called the Schwarz genus of d_n.*

Similar to the topological complexity, the higher topological complexity $TC_n(X)$ is a homotopy invariant. This property has been proved for the topological complexity $TC(X)$ in [4]. We present here proof of this important property for the $TC_n(X)$.

Proposition 2.1. *Suppose that X is homotopic to Y. Then $TC_n(X) = TC_n(Y)$.*

Proof. Assume that there exist continuous maps $f : X \to Y$ and $g : Y \to X$ such that $f \circ g \simeq id_Y$. We will prove that $TC_n(Y) \leq TC_n(X)$.

Let U be a open set in X^n such that there exists a continuous section $s : U \to X^{J_n}$ of e_n^X.

For $(B_1, \dots, B_n) \in U$, we have $s(B_1, \dots, B_n)$ is a map $\gamma : J_n \to X$. For any $i = 1, \dots, n$, let γ_i be the path in X defined by $\gamma_i(t) = \gamma|_{[0,1]_i}(t)$, where $[0, 1]_i$ denotes the i^{th} unit interval in the wedge $J_n = [0, 1] \vee \dots \vee [0, 1]$.

Let's consider the set

$$V = (g \times \dots \times g)^{-1}(U) = \{(A_1, ..., A_n) \in Y^n | (g(A_1), ..., g(A_n)) \in U\}.$$

We are going to construct a continuous section $\sigma : V \to Y^{J_n}$ of e_n^Y on this open set V of Y^n.

Suppose that $H_t : Y \to Y$ is the homotopy $id_Y \simeq f \circ g$ with $H_0 = id_Y$, $H_1 = f \circ g$. For $(A_1, A_2, ..., A_n) \in V$ we have $(g(A_1), ..., g(A_n)) \in U$ and therefore there exists a continuous section s of e_n^X. As mentioned above, $s(g(A_1), ..., g(A_n))$ is a path $\gamma : J_n \to X^n$. Now, we define $[A_1, A_i]$ to be the path in Y connecting A_1 to A_i , $i = 1, \dots, n$, by

$$[A_1, A_i](t) = \begin{cases} H_{3t}(A_1), & \text{if } 0 \leq t < \frac{1}{3} \\ f(s_{(g(A_1),g(A_i))}(3t - 1)), & \text{if } \frac{1}{3} \leq t < \frac{2}{3} \\ H_{3(1-t)}(A_i), & \text{if } \frac{2}{3} \leq t \leq 1 \end{cases} .$$

Here $s_{(g(A_1),g(A_i))}$ denotes the path in X, connecting $g(A_1)$ to $g(A_i)$, defined from s by

$$s_{(g(A_1),g(A_i))}(t) = \begin{cases} \gamma_1(1-2t), & \text{if } 0 \le t < \dfrac{1}{2} \\ \gamma_i(2t-1), & \text{if } \dfrac{1}{2} \le t \le 1 \end{cases}.$$

The section σ of e_n^Y on V is defined by putting

$$\sigma(A_1, ..., A_n) = ([A_1, A_1], [A_1, A_2], ..., [A_1, A_n]).$$

Here the right hand side denotes a map from J^n to X^n having its restriction on the i^{th} summand $[0,1]_i$ of J^n to be the path $[A_1, A_i]$. It is easy to see that σ defined above is continuous and $e_n^Y \circ \sigma = id_V$. Thus, $TC_n(Y) \le TC_n(X)$.

Similar arguments will prove $TC_n(X) \le TC_n(Y)$. And these imply the Proposition.

The following property shows the relation between the higher topological complexity $TC_n(X)$ of a space X and homotopy properties of X.

Proposition 2.2. *If X is a finite r-connected polyhedron, then*

$$TC_n(X) < \frac{n dim X + 1}{r+1} + 1.$$

In particular $r = 0$ (i.e. X is path-connected) then

$$TC_n(X) \le n \dim X + 1.$$

Proof. As it is mentioned in the Remark after Definition 2.1, the higher topological complexity $TC_n(X)$ is nothing but the Schwarz genus of the fibration e_n^X. The proposition is a consequence of the similar property of the Schwarz genus (see [8, Theorem 5]).

A lower bound for the higher topological complexity of the space X is given in terms of its cohomology with coefficient in any field $\mathbb{K}$.

Definition 2.2. *Let X be a finite path-connected polyhedron. Suppose that n is an integer, d_n is the diagonal map $d_n : X \longrightarrow X^n$ and $\mathbb{K}$ is a field.*

1. *The kernel of the homomorphism $d_n^* : H^*(X^n; \mathbb{K}) \longrightarrow H^*(X; \mathbb{K})$ is called the n-zero divisor of X.*

2. *d_n-zero divisor cup length of X, denoted by $\mathrm{cl}(X, n)$ (see[1]), is the maximal number k such that there exists k elements of the n-zero divisor of X satisfying $u_1 \cup u_2 \ldots \cup u_k \neq 0$.*

Then $cl(X, n)$ will be a lower bound of $TC_n(X)$. Precisely, we have the following proposition, which follows from [8, Theorem 4]. Detailed proofs can be found in [7] and [1].

Proposition 2.3. *Suppose that n is an integer, $n \geq 2$.*

i) Let X be a path-connected topology space. Then

$$TC_n(X) \geq cl(X, n) + 1.$$

ii) For finite path-connected polyhedra X and Y we have

$$cl(X \times Y, n) \geq cl(X, n) + cl(Y, n).$$

The next property, usually called product inequality, gives us an estimate of the higher topological complexity of the product space by those of its factors. This inequality is very useful in computing the higher topological complexity of many spaces.

Proposition 2.4. *For path-connected spaces X and Y. If $(X \times Y)^n$ is normal, then we have*

$$TC_n(X \times Y) \leq TC_n(X) + TC_n(Y) - 1. \tag{2.1}$$

Similar property for the Schwarz genus is known in [8, proposition 22] and for the topological complexity $TC(X)$ in [4]. A detailed proof for the case of the higher topological complexity can be found in [1].

3. A generalization of inequality product

In this section, we will generalize the inequality product for the higher topological complexity.

Theorem 3.1. *Suppose that E, X are finite path-connected CW-complexes, (Y, y_0) is a pointed space and $p : E \to X$ is a continuous map such that the following conditions hold*

i) For all $x \in X$, the fiber $p^{-1}(x)$ is homotopic to Y.

ii) The map p accepts a section $s : X \to E$, i.e., $p \circ s = id_X$.

iii) There exists a family of homotopy equivalences $h_x : p^{-1}(x) \to Y$ depending continuously on $x \in X$ such that $h_x(s(x)) = y_0$.

Then,

$$TC_n(E) \leq TC_n(X) + TC_n(Y) - 1.$$

To prove our main result, we first need some technical lemmas.

Lemma 3.1. *Let Y be a finite path-connected CW-complex, U an open set in Y^n on which there is a continuous section $s_U : U \to Y^{J_n}$ of e_n^Y. Then, for any $y_0 \in Y$ there exists always a continuous section $s'_U : U \to Y^{J_n}$ of e_n^Y such that $s'_U(A_1, ..., A_n)(0) = y_0$ for any $(A_1, ..., A_n) \in U$.*

Proof. We first consider the case when $y_0 \in U$.

For $(A_1, ..., A_n) \in U$ we have $s_U(A_1, ..., A_n)$ is a map $\gamma : J^n \longrightarrow Y^n$. Denote $\gamma(0) = P \in Y$. Define γ_i to be the restriction of γ on the i^{th} unit interval of J^n. That is, γ_i is a path connecting P and A_i.

By means of the section s_U, it implies that for any point $P \in U$, there exists a path connecting y_0 to P and this path depends continuously on P. Let denote this path by ℓ_P.

Now the section s'_U is defined by putting $s'_U(A_1, ..., A_n)$ to be a map $\gamma' : J^n \longrightarrow Y^n$, where γ' has its restriction on the i^{th} unit interval of J^n as $\ell_P * \gamma_i$. It is easy to see that $s'_U(A_1, ..., A_n)$ is a continuous section of e_n^Y and $s'_U(A_1, ..., A_n)(0) = y_0$.

Now suppose that $y_0 \notin U$. Let fix a point $y_1 \in U$. By the above arguments, we can construct a continuous section $\tilde{s}_U$ of e_n^Y such that $\tilde{s}_U(A_1, ..., A_n)(0) = y_1$. Suppose that $\tilde{s}_U(A_1, ..., A_n)$ is a map $\tilde{\gamma} : J^n \longrightarrow Y^n$. Then its restriction $\tilde{\gamma}_i$ on the i^{th} unit interval of J^n is a path connecting y_1 to A_i. Fix a path ℓ in Y connecting y_0 and y_1. Now we define the section s' by putting $s'_U(A_1, ..., A_n)$ to be a map $J^n \longrightarrow Y^n$ having its restriction on the i^{th} unit interval of J^n to be $\ell * \tilde{\gamma}_i$. Obviously, the defined map σ' is a section of e_n^Y having $s'_U(A_1, ..., A_n)(0) = y_0$.

The following lemma is implied from Proposition 20 of [8] and the fact that $TC_n(X)$ coincides with the Schwarz genus of the fibration e_X^n.

Lemma 3.2. *Given the topological space X, let $C = \{C_1, ..., C_p\}$ and $D = \{D_1, ..., D_q\}$ be open covering of X^n such that on each $C_i \cap D_j$ there exists local section of e_n. Then*

$$TC_n(X) \leq p + q - 1.$$

Proof. [Proof of Theorem 3.1] Suppose that $TC_n(X) = p$. By definition, there is an open covering $U = \{U_1, \ldots, U_p\}$ of X^n such that there exists a section of e_n^X on each U_i. Put $C_i = \{(A_1, \ldots, A_n) \in E^n | (p(A_1), \ldots, p(A_n)) \in U_i\}$. Then $C = \{C_1, \ldots, C_p\}$ is an open covering of E^n

Suppose that $TC_n(Y) = q$, and $V = \{V_1, \ldots, V_q\}$ is an open covering of Y^n such that there exists a section of e_n^Y on each V_j. Let

$$D_j = \{(A_1, \ldots, A_n) \in E^n | (h_{p(A_1)}(A_1), \ldots, h_{p(A_n)}(A_n)) \in V_j\}.$$

Then $D = \{D_1, \ldots, D_q\}$ is an open covering of E^n. Let we fix a section of e_n^Y on each V_j as that in the Lemma 3.1.

We will now construct a section of e_n^E on each $C_i \cap D_j,\ i = 1, \ldots, p,\ j = 1, \ldots, q$. Suppose that $(A_1, \ldots, A_n) \in C_i \cap D_j \subset E^n$.

By definition $(A_1, \ldots, A_n) \in C_i$ means that $(p(A_1), \ldots, p(A_n)) \in U_i$. Since there exists a continuous section s_1 of e_n^X on U_i, there is a path going from $p(A_1)$ to $p(A_i)$ defined by this section s_1. Let denote γ_2 the image of this path by the section s of p.

Since $(A_1, \ldots, A_n) \in D_j$ it implies that $(h_{p(A_1)}(A_1), \ldots, h_{p(A_n)}(A_n)) \in V_j$. By assumption, there exists a section s_2 of e_n^Y on this V_j. Let's choose s_2 to be the one defined in the Lemma 3.1. This section s_2 defines a path in Y connecting $h_{p(A_1)}(A_1)$ to the point y_0. We denote by γ_1 the inverse image of this path by the homotopy equivalence $h_{p(A_1)}$.

Similarly, the section s_2 defines a path in Y connecting y_0 to the point $h_{p(A_i)}(A_i)$, $i = 1, \ldots, n$. We denote by γ_3 the inverse image of this path by the homotopy equivalence $h_{p(A_i)}$.

We denote by $[A_1, A_i]$ the path $\gamma_1 * \gamma_2 * \gamma_3$.

Now we can define a map $\sigma : C_i \cap D_j \longrightarrow E^n$ by putting $\sigma(A_1, ..., A_n) = ([A_1, A_1], ..., [A_1, A_n])$ for any $(A_1, ..., A_n) \in C_i \cap D_j$.

This map σ is obviously a continuous section of e_n^E. Applying the Lemma 3.2 we have $TC_n(E) \leq p + q - 1 = TC_n(X) + TC_n(Y) - 1$.

Remark 3.1. *Suppose that X, Y are path-connected CW-complexes. Put $E = X \times Y$ and the map $p : E \to X$ to be the projection in the first component. The map $s : X \to E$ defined by $s(x) = (x, y_0)$ is obviously a section of p and $h_x : p^{-1}(x) \to Y$, $h_x(x, y) = y$ is a homotopy equivalence. Then, all the assumptions of Theorem 3.1 are satisfied. And we get again the product inequality (2.1) as stated in Proposition 2.4.*

Theorem 3.1 has been used to compute the higher topological complexity of configuration spaces on some topological manifolds in [3]. We briefly recall the case of configuration space $F(\mathbb{T}, k)$ on the two-dimensional torus.

The configuration space of k distinct ordered points in $\mathbb{T}$ is a subset of $\mathbb{T}^k$, defined by

$$F(\mathbb{T}, k) = \{(x_1, \cdots, x_k) \in \mathbb{T}^k |\ x_i \neq x_j \text{ with } 1 \leq i \neq j \leq k\}.$$

Proposition 3.2. *Let k be an integer with $k \geq 2$. The higher topological complexity of the configuration space of k instinct ordered points on the 2-dimensional torus $\mathbb{T}$ is*

$$TC_n(F(\mathbb{T}, k)) = n(k+1) - 1.$$

Observe that the projection on the first $k-1$ coordinates $\pi_k : F(\mathbb{T}, k) \longrightarrow F(\mathbb{T}, k-1)$ is a fibration with the fiber homotopic to the bouquet of k circles $Y_k = \underbrace{\mathbb{S}^1 \vee \ldots \vee \mathbb{S}^1}_{k}$. It is proved in [3] that this fibration π_k admits a section σ_k for any $k = 1, \ldots$ and satisfies all assumptions of Theorem 3.1. Combining properties of the higher topological complexity mentioned in the previous section and Theorem 3.1 we get

$$TC_n(F(\mathbb{T}, k)) \leq TC_n(F(\mathbb{T}, k-1)) + TC_n(Y_k) - 1.$$

Moreover, it follows from [3, Theorem 3] that $TC_n(Y_k) \leq n+1$ for $k \geq 2$. So,

$$TC_n(F(\mathbb{T}, k)) \leq TC_n(F(\mathbb{T}, k-1)) + n \text{ for all } k \geq 2.$$

By induction on k, we get

$$TC_n(F_n(\mathbb{T}, k)) \leq TC_n(F(\mathbb{T}, 1)) + n(k-1) = 2n - 1 + n(k-1) = n(k+1) - 1.$$

To prove that $n(k+1) - 1$ is the lower bound of $TC_n(F(\mathbb{T}, k))$ we need to use the lower bound stated in Proposition 2.3 and spectral sequence arguments. We will skip it here. A detailed proof can be found in [3]

References

[1] **I. Basabe, J. González, Y.B. Rudyak, and D. Tamaki,** *Higher topological complexity and homotopy dimension of configuration spaces on spheres*, Algebr. Geom. Topol. 14 (2014), 2103-2124.

[2] **Nguyen Viet Dung and Nguyen Van Ninh,** *The Higher Topological Complexity of Complement of Fiber Type Arrangements*, Acta Math Vietnam 42, 2(2017), p.249-256.

[3] **Nguyen Viet Dung and Nguyen Van Ninh,** *The higher topological complexity of configuration spaces of odd-dimensional spheres*, Singularities - Kagoshima, 2017, World Scientific (2020). pp 167-183.

[4] **M.Farber,** *Topological complexity of motion planning*, Discrete Comput. Geom 29(2003) 211 - 221.

[5] **J.Gonzaléz and M. Grant,** *Sequential motion planning of non-colliding particles in Euclidean spaces*, Proc. Amer. Math. Soc. 143 (2015), 4503-4512.

[6] **J.Gonzaléz and B. Gutiérrez,** *Topological complexity of collision-free multi-tasking motion planning on orientable surfaces*, Contemp. Math., Amer. Math. Soc. 702 (2018), 151-163.

[7] **Y.B.Rudyak,** *On higher analogs of topological complexity*, Topology and its Application, 157(2010) 916-920.

[8] **A.S.Schwarz,** *The genus of fiber space*, Amer. Math. Sci, Transl 55(1966) 49 - 140.

Nguyen Viet Dung
Institute of Mathematics, VAST
Hanoi
Vietnam
`vietdung@math.ac.vn; nvdung@cmc-u.edu.vn`

Nguyen Van Ninh
Thai Nguyen University of Education
Thai Nguyen
Vietnam
`nguyenninhsp@gmail.com`

J. Math. Math. Sci. (2023) 67–85

A blockchain-based Certificate Management System using the Hyperledger Fabric Platform

Quang Duy Tran (Hanoi, Vietnam)
Trong Minh Hoang (Hanoi, Vietnam)

(Received Nov. 26, 2022)

Abstract. In recent years, blockchain technology, with its outstanding advantages in terms of security and decentralized data storage, has quickly been applied in education, especially in managing educational certificates. The traditional paper certificate management process has many limitations on security, storage, verification, ownership, and prevention of certificate fraud and tampering. With emerging blockchain technology, the traditional paper certificate management process has changed to a new stage with massive benefits and opportunities to overcome the above challenges. Following this approach, many blockchain-based applications have been proposed, and several universities worldwide have adopted various solutions based on blockchain technology. This paper presents implementing an educational certificate management system based on blockchain technology using the Hyperledger Fabric platform. Our proposed solution can be solved in almost storage and security aspects of certificate management problems, such as confidence, verification, ownership, and tamper avoid.

1. Introduction

Blockchain technology is the core technology used to create a cryptocurrency. In 2008, this technology was introduced by Satoshi Nakamoto along with Bitcoin - a cryptocurrency [1]. It has been considered part of the fourth

Key words and phrases: Blockchain, Hyperledger Fabric, Educational Certificate, Verification, Permissioned Blockchain.
2020 Mathematics Subject Classification: 68P25

industrial revolution and was quickly researched and applied in many fields such as finance, insurance, healthcare, the Internet of Things (IoT), supply chain management, intellectual property management, etc.

The development of blockchain technology could be divided into three main stages: Blockchain 1.0, 2.0, and 3.0. Blockchain 1.0 was used for cryptocurrencies. Blockchain 2.0 was introduced along with smart contracts for managing digital properties. A typical case is the Ethereum platform proposed by Vitalik Buterin [6]. In Blockchain 3.0, many applications were developed in various sectors such as government, education, finance, insurance, healthcare, and science [10].

The application of Blockchain to education is still in the early stages. Educational institutions and universities have started to study and utilize blockchain technology. Most institutions use it to secure, share, and verify academic achievement. However, this technology can be applied in many cases, such as issuing the Certificate, storing a portfolio, managing intellectual property, identification, etc. [2]. Many researchers in the field believe that blockchain technology has much more to offer and can revolutionize the field. In the case of certificate management, digital certificates will be stored and secured on a blockchain system. Then, they cannot tamper, and the accuracy of the Certificate can be easily verified by anyone who can access the blockchain system through an application without any third party. Because no intermediary is required to verify the Certificate, the Certificate can still be validated even if the organization that issued it no longer exists. The records of certificates on a blockchain can only be destroyed if all copies stored on network nodes are eliminated.

In this paper, we briefly introduce applications of blockchain technology in the education field, educational certificates, and certificate management process and exploit the advantages of certificate management based on blockchain technology. We also propose a digital academic certificate management solution based on the Hyperledger Fabric platform. The rest of the article is structured as follows. Section 2 presents mainly related work. In section 3, we illustrate a brief overview of blockchain technology. Section 4 outlines the certificate management process and the types of different certificates. In Section 5, we propose an educational certificate management solution. Last but not least, we conclude this article in Section 6.

2. Related Works

With its outstanding features, blockchain technology could benefit the education field significantly. Recently, numerous blockchain-based apps for educational purposes have been developed, and extensive studies and analyses on Blockchain in the field of education have been conducted. Blockchain-based applications were classified into 12 main categories, including certificates management, competencies, and learning outcomes management, evaluating students' professional ability, protecting learning objects, securing collaborative learning environment, fees and credits transfer, obtaining digital guardianship consent, competitions management, copyrights management, enhancing students' interactions in e-learning, examination review, and supporting lifelong learning. Each category addresses security, identity authentication, trust, and privacy issues within the education environment [15].

In the case of certificate management, it concerns handling all forms of academic certificates, transcripts, or other accomplishment records. Many applications that used blockchain technology for managing digital certificates were proposed. And many educational institutions and universities worldwide have applied this technology to manage and issue their digital education certificates. In 2017, The University of Nicosia (UNIC) issued and validated certificates using a public blockchain based on the open-source standard "Blockcerts". The Knowledge Media Institute (KMI) within Open University(OU), United Kingdom (UK), has performed a study on enhancing standards for badging, certification, and reputation on the Web with the use of blockchain technology. In addition, the Massachusetts Institute of Technology (MIT) has used the Learning Machine (LM) Certificates to issue digital diplomas for students at the MIT Media Lab [2]. Sony Global Education has also developed an internal certification management system and a system that applies to the education sector using blockchain technology [3].

Some innovative digital educational certification platforms using blockchain technology were also proposed. Nespor [16] proposed a blockchain-based certification platform that compensated for using the school as a certification agent. This platform would authorize universities or educational institutions to supply official certificates for students with a high level of privacy of their information. Thus, students could share it directly with anyone requesting their certificates. Authors in [17] introduced a novel blockchain-based education solution for issuing and verifying official transcripts or certificates. The individuals could have access to their data records and can easily share those records with others. However, only certified organizations can access and modify the stored data under some restricted conditions and rules. Srivastava et al. [18] proposed a globally trusted blockchain-based educational framework among various stakeholders like universities, companies, and other educational institutions that agree to collaborate as a part of the framework. This framework supports verifying the academic certificates and course credits of a learner registered in a university or an educational institution which can be digitally transferred among the stakeholders. All stakeholders know students' education records by achieving consistency among the local copies of educational certificates and credits.

3. Blockchain Technology Background

Blockchain technology combines cryptographic theory, game theory, and techniques in computer science. This section outlines the key concepts and components used in this technology.

Blockchain. Blockchain is a block sequence containing a complete list of cryptographically signed transaction records, called a digital ledger [1]. Each block is cryptographically linked to the previous block after the validation process. When new blocks are created, they are replicated across copies of the ledger within the network, and any conflicts are resolved automatically using previously established rules. The first blockchain block is called the genesis block, which has no parent block. A typical blockchain is shown in Figure 1.

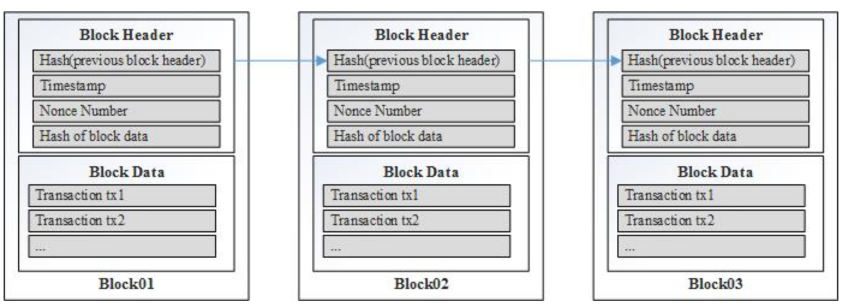

Figure 1. A typical Blockchain

Block. A block contains a block header and data (as shown in Figure 1). The block header contains metadata for this block. The block data includes a list of validated transactions published to the blockchain network. It should be noted that they can define their data fields in the different blockchain systems.

However, many blockchain systems have data fields like the following:
The block header includes data fields as

- Previous block header's Hash: a hash value that points to the previous block.
- Hash of block data: the hash value of all the transactions in the block
- Timestamp
- Nonce Number: used for every hash calculation

The block data is composed of a transaction list. The maximum number of transactions a block can contain depends on the block size and the size of each transaction.

Cryptographic Hash Functions. These functions are used for many operations in the Blockchain. Hashing applies a cryptographic hash function to input data, which calculates a relatively unique output (called a digest or a hash code). Input data can be nearly any type (e.g., a file, text, or image). Anyone can take input data, hash that data, and derive the same result. This proves that there was no change in the data. Even the smallest change to the input data (e.g., changing a single bit) will result in a completely different output digest [4]. Within a blockchain network, hash functions are used for many tasks, such as: creating unique identifiers, combining with the public key to derive addresses, and securing the block data and header.

Asymmetric-Key Cryptography. Blockchain uses asymmetric-key cryptography (public-key cryptography) [5]. Asymmetric-key cryptography uses a pair of keys: a public key and a private key. These keys are mathematically related to each other. The public key is made public, but the private key must be kept secret. Although there is a mathematical relationship between the two keys, the private key cannot be determined based on knowledge of the public key. Data can be encrypted with the public key and then decrypted with the private key.
Asymmetric-key cryptography enables a trusting relationship between users who do not know or trust one another by verifying the integrity and authenticity of transactions. A private key is used to encrypt transactions to create a digital signature of the transaction to do. The digital signature of the transaction is broadcasted throughout the whole network, and anyone also can use public keys to verify the digital signatures. Since the public key is always public, encrypting the transaction with the private key proves that the signer of the transaction has access to the private key. Alternatively, one can encrypt data with a user's public key such that only users with access to the private key can decrypt it.

In Blockchain, private keys create a digital signature by encrypting the transaction. Public keys verify digital signatures generated with private keys and derive addresses.

Consensus mechanism. Most blockchain systems use consensus protocols to reach consensus among untrustworthy nodes in decentralized environments. In the existing systems, there are many consensus mechanisms, such as Proof-of-Work (PoW) [1], Proof-of-Stake (PoS) [6], Practical Byzantine Fault Tolerance (PBFT) [7], and RAFT [8], etc.

PoW is a consensus protocol used in the Bitcoin network [1]. In a decentralized network, a node wants to publish a block of transactions and add this block to the Bitcoin blockchain to get rewards, and much work has to be done to prove that the node is not likely to attack the network. The nodes repeatedly run hashing functions to find a Nonce value, which is challenging but easy for others to validate. This enables all other nodes to validate any new blocks quickly, and any proposed block that is not satisfied will be rejected.

Proof-of-Stake (PoS) is used in Ethereum [6]. The PoS model is based on the idea that the more stake a user has invested into the system, the more likely they will want the system to succeed and the less likely they will want to subvert it. The stake is often an amount of coin the blockchain network user has invested into the system in various ways. PoS blockchain networks use the amount of stake as a determining factor for publishing new blocks. Thus, the capability of a blockchain network user to publish a new block is dependent on the ratio of their stake to the overall blockchain network amount of staked coin. There is no need to perform resource-intensive computations as found in POW with this consensus model. So, Compared to PoW, it saves more energy and is more effective.

Practical Byzantine Fault Tolerance (PBFT) [7] and RAFT [8] [9] are consensus protocols used in different versions of the Hyperledger Fabric platform. They are used in the ordering service, which receives endorsed transactions from the clients and emits a stream of blocks.

4. Certificate considerations

In this session, the main characteristics of a certificate and its management system will be presented to figure out how to apply blockchain technology to this area.

4.1. Certificate Concept

A certificate contains a certified statement, especially about the truth of something. In education, the Certificate is used as evidence for the achievement of learning outcomes, the teacher's competence, a learner's learning process, etc. Certification describes any process by which a certificate is issued to verify a claim [2].

4.2. Processes involved in the certification

There are three processes involved in certification including:

Issuing. This is the process of generating and recording certificate data such as certificated requests, certificate issuer, evidence, recipient, and signature onto a certificate. Usually, this data is stored in a centralized database and on a certificate issued to the user.

Sharing. The recipients will share their certificates with a third party in this process. There are three methods to sharing credentials: directly transferring the Certificate (or a copy of the Certificate) to the third party; storing the Certificate with an authorizer who is only allowed to share with specific people according to your request; publishing the Certificate by putting it in an online public store, where everyone may view it.

Verification. In this process, a third party will verify the Certificate's authenticity. There are three ways to do this

- Verification uses security features built into the Certificate, such as checking the seal's authenticity, special security paper, signature, etc.
- Verification of Certificate through the original issuer whereby a third party contacts the original issuer to ask whether they issued the Certificate.
- Verification utilizing an application that can access a centralized database provided by the issuer. Anyone can look up certificates in this database to see copies of all issued certificates. Level and compare the two versions with each other.

4.3. Limitations of Paper Certificates

Most certificates and records are still issued on paper and are widely used. However, paper certificates also have the following significant disadvantages [2]:

- Paper certificates are easy to forget. Therefore, the issuer must store a list of the issued certificates for verification. This verification is a manual process; hence, it is time-consuming and requires considerable human resources.

- While certificates may still be valid, the ability to verify them is lost if the issuer no longer exists.

- The more secure the Certificate, the Certificate cost is the more expensive.

- Issuers can cheat when issuing certificates without any restrictions.

- Once a certificate has been issued, there is no way to revoke the Certificate unless the owner relinquishes control of it.

4.4. Digital Certificates not using Blockchain Technology

Digital certificates are issued without using a blockchain system and have many advantages over paper certificates, such as:

- They use fewer resources to issue, maintain and use because the verification process can be done automatically. The security of a certificate is based on the security of cryptographic protocols, which assures that it is inexpensive to make but prohibitively expensive for anybody other than the issuer to reproduce.

- Issuers can revoke certificates.

- Some issuer fraud is not possible, such as changing the timestamp, changing the Certificate's serial number, etc.

However, this digital Certificate also has some significant disadvantages, such as:

- If not using digital signatures, they are still straightforward to forget.

- If digital signatures are used, the issuance and verification of certificates significantly depend on third parties who are digital signature providers.

- Keeping digital certificates safe can require complex backup systems.

4.5. Digital Certificates using Blockchain Technology

Blockchain technology is considered ideal for securing, sharing, and verifying certificates. Once certificate information is stored on a blockchain system, it cannot be altered or tampered with. As a result, blockchain-secured digital certificates hold significant advantages over paper certificates or other digital certificates (non-blockchain), such as:

- They cannot be tampered with.
- Verification of the Certificate is done easily by anyone through a piece of software without any intermediary parties.
- A certificate can still be validated even if the organization that issued it no longer exists or has an access system.
- Certificates can only be destroyed if every copy of the ledger on every node in the blockchain system is destroyed.
- Using a Hash of the document allows the document's signature to be published without needing to publish the document itself, thus ensuring its privacy.

5. The proposed platform for the education certificate management system

5.1. Scope and requirements

An educational certificate management system is built on a blockchain platform that focuses on solving problems in certificate management, such as security, authentication, and ownership, with specific requirements as follows:

- The issuers who want to participate as a system member must be predefined and licensed.
- The certificates stored on the blockchain system must be secure and cannot be tampered with. These certificates have to be verified easily and quickly by third parties who need to verify the certificates. The input to the solution is the certificate information. Certificate information includes the Certificate number, issuer's information, recipient's information, Certificate's content, expiration date, the status of the Certificate, etc.

- Build an application that helps issuers manage statistics certificates and the life cycle of certificates through the functions such as data input function, search, statistics, etc.
- Providing functions that allow recipients to share certificates and manage their certificates.
- Providing tools to help third parties(such as employers, educational institutions, etc.) verify certificates easily and quickly. Reduce time for validating a certificate.

5.2. Designing system

A. System architecture

We designed the education certificate management system based on the Hyperledger Fabric platform with the given requirements (Figure 2). Permissioned blockchain platforms are increasingly used in industry. To find the most suitable blockchain platform to design and implement an education certificate management system, we review and assess the three leading permissioned blockchain platforms: Hyperledger Fabric, Quorum, and R3 Corda, concerning performance, scalability, privacy, and other criteria.

Hyperledger Fabric is an open-source permissioned blockchain platform developed by the Linux Foundation community. It is designed for enterprise contexts and delivers critical differentiating capabilities over other popular distributed ledger or blockchain platforms [13]. The Fabric uses smart contracts (also called chaincodes) to implement the application logic. In this platform, consensus protocols can be plugged in (such as Practical Byzantine Fault Tolerance (PBFT), Raft or Kafka,...). With a highly modular and configurable architecture, Hyperledger Fabric is one of the permissioned blockchain platforms that can be applied in many fields, such as finance, banking, healthcare, human resources, supply chain, and even digital music delivery, etc. Walmart used Hyperledger Fabric to take on food traceability and safety. Sony Global Education chooses Hyperledger Fabric for its Next-Generation Credentials Platform [14].

R3 Corda is an open-source permissioned platform developed by R3 Corporation Corda's design was initially driven by the needs of regulated financial institutions but turned out to be far more broadly applicable [21]. Corda also uses smart contracts for implementing the application logic. It has two types of consensus: the validity of the transaction and the uniqueness of the transaction. Each signer checks validity before signing the transaction, and the Notary nodes check uniqueness. The consensus is reached by the nodes that carry out the transactions, not the entire system.

Quorum is an enterprise blockchain platform initially developed by J.P. Morgan for the financial sector but can be used for any industry. Quorum is a permissioned blockchain based on the Ethereum blockchain [23]. More precisely, it is a fork from go-Ethereum, with several better modifications regarding privacy, consensus protocol, performance, etc. It uses other consensus protocols for consortium blockchains, such as a Raft-based consensus protocol and Istanbul BFT [20].

Table 1 provides an essential characteristics summary of the three above blockchain frameworks [19] [20] [22].

Characteristics	**Hyperledger Fabric**	**R3 Corda**	**Quorum**
Governance/Support	Linux Foundation	R3 Corporation	J.P. Morgan Chase
Mode of operation	Permissioned	Permissioned	Permissioned
Consensus	Plug-in consensus mechanism	Voting-based/RAFT	Clique PoA of RAFT-based or Istanbul BFT
Smart contracts	Yes	Yes	Yes
Privacy	Channels and private data collections permit high customization	Transactions are private by default. Information shared on a need-to-know basis	Supports private and public transactions
Performance	Strongest latency and throughput values	Strongest latency, poor throughput	Poor latency, strong throughput

Table 1. Comparison of Hyperledger Fabric, R3 Corda, and Quorum.

Three types of users are involved in this proposed system: Recipients, Issuers, and Verifiers. Recipients are students or course participants. Each recipient will be assigned a username and password to access the Web application. Issuers are the educational institutions or universities that issue the Certificate to the recipients. Verifiers are the companies, employers, or educational institutions that need to verify the accuracy of the Certificate. The proposed system includes the following main components:

Blockchain system. Is a permission blockchain network designed based on the Hyperledger fabric platform. Details of this network are described in

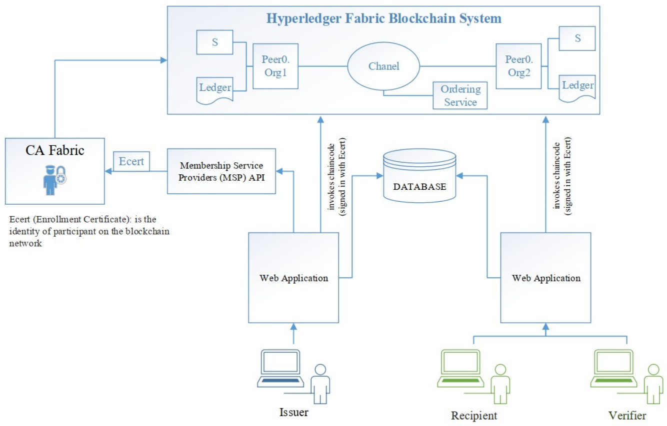

Figure 2. The proposed blockchain-based certificate management system architecture

the next section.

Web application for issuers. The application is built on the web platform. This application can connect to the Peer in the blockchain network and update query ledgers through smart contract calls. It includes the following functions: Register creates a certificate, update certificate information, view the Certificate in detail, search, certificate statistics, etc.

Web application for recipients and verifiers. This application is also built on a web platform and includes the recipient's and verifier's functions. The functions for recipients include sharing certificates, searching, view certificate in detail. The tasks for verifiers include: Looking up, verifying Certificates, etc.

Database. Is a database that stores attribute information, including information about the recipient, educational institution, Certificate, etc., as well as the user's login information.

B. Proposed blockchain model

Figure 3 depicts the hyper ledger fabric-based blockchain system architecture we designed and implemented for the certificate management solution.

The blockchain network consists of Peers (called Peer nodes), Endorsing Peers, the Ordering Service, and Certificate Authorities (CA). Each Peer hosts ledgers and smart contracts (in Hyperledger Fabric are called chain code). The endorsing peers are responsible for endorsing the transaction proposals before sending them to the ordering service. Peers work together through a channel called "EduChanel." The ledger stores the basic information of the Certificate. The Peers update and query certificate information on the ledger through preinstalled smart contracts. The ordering service includes three nodes. The Raft consensus protocol is used in the ordering service for creating new blocks.

When each organization joins the network, it owns two peers; one Peer is the endorsing Peer and Certificate Authority (CA). A user or a node wants to participate in the blockchain network to have a digital identity issued by a CA. Digital identities (or simply identities) have the form of cryptographically validated digital certificates that comply with the X.509 standard.

In this network, Fabric CA is used for each organization. As shown in Figure 3, organization Org1 will own Peer.O1, Endorsing Peer.O1, and CA1. CA1 issues digital identities for users of the Org1.

The blockchain network has two main operations: updating certificate information and querying certificate information. The update creates a new record of certificate information and updates it to the ledger on the Peers after the verification and consensus process. An information query is an action in which users can retrieve certificate information stored on the ledger at each Peer in the system.

The process of updating certificate information from the application to the ledger is briefly described in the following steps:

- Step 1: Initiates a transaction proposal
  When the user enters the Certificate's information (including the Certificate's attribute information and image) into the system through the web application's functions, the web application generates a transaction proposal. It sends it to the endorsing peers in the blockchain network for endorsement (through API functions provided by Fabric SDK). The proposal is a request to invoke a chain code function with input parameters such as certificate information, sender, etc., to read and update the ledger.

- Step 2: Endorsing a proposal
  Each of these endorsing peers validates the transaction proposal by independently executing a chain code using the transaction proposal to generate a transaction proposal response, signs it, and returns it to the application.

- Step 3: Receiving signed transaction proposal response
  The application receives a signed transaction proposal response from the endorsing peers. The application verifies the endorsing peer signatures and compares the proposal responses to determine if the proposal responses are the same. Then, the application generates a transaction with a transaction proposal and a signed transaction proposal response.

- Step 4: Submitting transaction

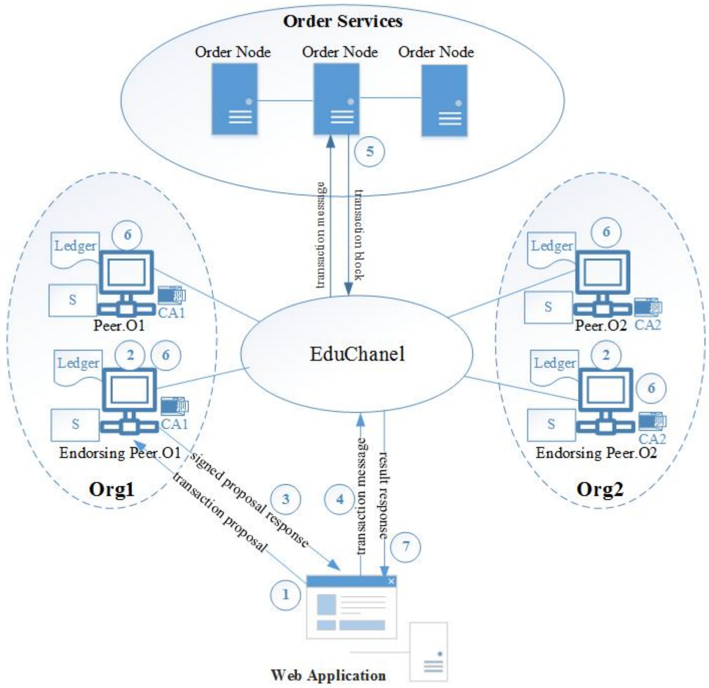

Figure 3. The proposed blockchain network architecture

The application sends the transaction to the ordering service node by "broadcasting" it through the " EduChanel " channel.

- Step 5: Packaging and Delivering blocks
  Since ordering service nodes receive the transaction, the order, and package transactions into a block, the ordering service nodes work together under the RAFT consensus mechanism for the publishing block. This is to ensure that a unique block is generated on the system. When the publishing block of transactions is completed. The block of transactions is distributed to all peers on the channel.

- Step 6: Ledger updated
  The peers receive a block of transactions from the ordering service; every transaction within a block is validated before it is committed to the ledger. Valid transactions are committed to the ledger. Invalid transactions are retained for audit but are not committed to the ledger.

- Step 7: The application receives the transaction result response
  When peers committed the transaction to their local ledger. Each Peer

emits an event of transaction result to notify the application that the transaction was validated or invalidated. The application receives the results of executing the transaction from the peers and displays them to a user.

Querying certificate information is much simpler than updating. The application sends a query proposal to the Peer for invoking the smart contracts. The Peer executes smart contracts to query information from its local ledger and returns the query result to the application.

5.3. Implementation

A. Blockchain System

We implement Hyperledger Fabric based blockchain system in Ubuntu 18.04 operating system. Docker is used for developing, implementing, and running peers, applications, and services. Therein, each Peer, application, or service is installed and running on a distinguished docker container.

The details of the deployment environment are as follows:

- Operating system: Ubuntu 18.04
- Hyperledger fabric version 2.2.2
- Hyperledger fabric CA Client version 1.4.9
- CouchDB Database version 3.1.

We deployed the blockchain system with assuming that there are two organizations in the system: Org1 and Org2. Each organization includes two peers: a normal Peer, an endorsing Peer, and a CA. Each Peer and CA are deployed and run on a distinguished container. In this system, we also use three ordering nodes for ordering service. They are also installed and run on a distinguished container.

Ledger. Ledger at a peer is stored on CouchDB database version 3.1. A few pieces of information are stored in the ledger, including the Name of the Issuer; ID of the Recipient; Hash of certificate data; Signature of the Issuer (The private key of the issuer signs the signature); Signature of the recipient (The private key of recipient signs signature), Certificate's date was created, etc.

Smart Contract. There are a few main smart contracts that are deployed and run in chain code at the Peer as follows:

- CreateNewCertificate: Issues and writes a new certificate to the ledger. The input information includes: Name of the Issuer; ID of the Recipient; Hash of certificate data; Signature of the Issuer (The private key of the issuer signs the signature); Signature of the recipient (the private key of the recipient signs the signature), Certificate's date was created, etc.

- CancelCertificate: revoke the issued Certificate for recipients. The input information includes the ID of the Certificate.

- GetCertificateByIssuer: Returns all the certificates issued by a specific Issuer. The input information includes the Public Key of the Issuer that issued the Certificate.

- GetCertificateByRecipients: Public Key of Recipient

- QueryCertificateByID: Get a detail of the Certificate based on its ID.

B. Web application for issuers

This web application is developed using Javascript language on the NodeJS platform version 12.12.0. This application includes the main functions:

- Sign in: allows issuers to log in to the application.

- Sign out: exit the application.

- Create Certificate: issues a new certificate for recipients. Including information: Name, Date of birth, year of graduation, Degree classification, Date of issue, and Image of Certificate.

- Cancel Certificate: revokes the issued Certificate.

- View certificates: views issued certificates.

C. Web application for recipients and verifiers

This web application is also developed using Javascript language on the NodeJS platform version 12.12.0. The main functions for recipients:

- Register: allows recipients to sign up for receiving Certificates.

- Sign in: allows recipients to log into the application.

- Sharing certificates: allows recipients to share their Certificates.

- View certificate: allows recipients to view their Certificate in detail. The functions of verifiers:

- Verify Certificate: this function allows verification Certificate based on Certificate's image or Certificate Hash that the recipient supplies.

D. Database

This database stores data of Web applications, using MongoDB database system version 5.05; stored information includes:

- Certificate: ID of the certificate, recipient's name, recipient's major, issuer's name, etc
- Recipient: Recipient ID, name, email, password, etc.

6. Conclusion

In this paper, we briefly presented an overview of applying blockchain technology in the education field and its advantages in the management of digital educational certificates. In addition, we have proposed an educational certificate management system based on the Hyperledger Fabric platform. The blockchain network architecture and the process of updating and querying certificate information are also presented in detail. This solution meets the requirements for certificate management, such as security, sharing, ownership, and verification. We have also successfully implemented this solution in a test environment. In the future, we hope to continue this solution in transcripts management and implement it in practice.

References

[1] **Nakamoto.,** *Bitcoin: A peer-to-peer electronic cash system,* Decentralized Business Review, p.21260, (2008)

[2] **Alexander Grech, Anthony F. Camilleri, Andreia Inamorato dos Santos,** *Blockchain in Education,* European Commission JRC, 2017.

[3] **J. Russell,** *Sony wants to digitize education records using the Blockchain,* https://techcrunch.com/2017/08/09/sony-education-blockchain, last accessed 2022/11/12.

[4] **National Institute of Standards and Technology,** *Secure Hash Standard (SHS).* Federal Information Processing Standards (FIPS) Publication 180-4, August 2015, https://doi.org/10.6028/NIST.FIPS.180-4.

[5] **D. Johnson, A. Menezes, and S. Vanstone,** *The elliptic curve digital signature algorithm (ecdsa),* International Journal of Information Security 1(1), 36-63 (2001).

[6] **Vitalik Buterin,** *A Next-Generation smart contract and decentralized application platform.* Ethereum White Paper.

[7] **Christian Cachin**: Architecture of the Hyperledger blockchain fabric. IBM Research (2016).

[8] **Diego Ongaro and John Ousterhout,** *In Search of an Understandable Consensus Algorithm (Extended Version),* https://raft.github.io, last accessed 2022/04/12.

[9] **Artem Barger, Yacov Manevich, Hagar Meir, Yoav Tock,** *A Byzantine Fault-Tolerant Consensus Library for Hyperledger Fabric,* arXiv:2107.06922 [cs.DC], 2021.

[10] **Gatteschi V, Lamberti F, Demartini C, Pranteda C, Santamaría V.,** *Blockchain and Smart Contracts for Insurance: Is the Technology Mature Enough?,* Future Internet. 2018; 10(2):20. https://doi.org/10.3390/fi10020020

[11] **P. K. Sharma, M.-Y. Chen, and J. H. Park,** A software-defined fog node based distributed blockchain cloud architecture for IoT. IEEE Access, PP(99):1-1 (2017).

[12] **Protocol-Labs,** *Filecoin: A decentralized storage network,* https://filecoin.io/filecoin.pdf, last accessed 2022/11/22.

[13] **Hyperledger,** *Fabric,* https://www.hyperledger.org/, last accessed 2022/11/22

[14] **Hyperledger.Org,** *Browse various use cases powered by Hyperledger technologies,* https://www.hyperledger.org/learn/case-studies, last accessed 2022/11/22

[15] **Alammary A, Alhazmi S, Almasri M, Gillani S.,** *Blockchain-Based Applications in Education: A Systematic Review,* Applied Sciences. 2019; 9(12):2400. https://doi.org/10.3390/app9122400.

[16] **Nespor, J.,** *Cyber schooling and the accumulation of school time,* Pedag. Cult. Soc. 2018, 1–17. https://doi.org/10.1080/14681366.2018.1489888

[17] **Han, M.; Li, Z.; He, J.S.; Wu, D.; Xie, Y.; Baba, A.,** *A Novel Blockchain-based Education Records Verification Solution,* In Proceedings of the 19th Annual SIG Conference on Information Technology Education, Fort Lauderdale, FL, USA, 3–6 October 2018; pp. 178–183. https://doi.org/10.1145/3241815.3241870

[18] **Srivastava, A.; Bhattacharya, P.; Singh, A.; Mathur, A.; Prakash, O.; Pradhan, R.,** *A Distributed Credit Transfer Educational Framework based on Blockchain,* In Proceedings of the 2018 Second International Conference on Advances in Computing, Control and Communi-

cation Technology (IAC3T), Allahabad, India, 21–23 September 2018; pp. 54–59. https://doi.org/10.1109/IAC3T.2018.8674023

[19] **Valenta, Martin; Philipp G. Sandner,** *Comparison of Ethereum, Hyperledger Fabric and Corda,* 2017, http://explore-ip.com/2017_Comparison-of-Ethereum-Hyperledger-Corda.pdf, last accessed 2023/01/23.

[20] **20. Julien Polge; Jérémy Robert; Yves Le Traon,** *Permissioned blockchain frameworks in the industry: A comparison,* ICT Express, Volume 7, Issue 2, 2021, pp. 229-233,ISSN 2405-9595. https://doi.org/10.1016/j.icte.2020.09.002.

[21] **Richard Gendal Brown,** *The Corda Platform: An Introduction,* 2018, https://corda.net/content/corda-platform-whitepaper.pdf, last accessed 2023/01/23.

[22] **Fausto Martin,** *Permissioned Blockchain Platform Comparison,* https://www.tradeheader.com/blog/hyperledger-fabric-comparative, last accessed 2023/01/23.

[23] **Quorum,** *Quorum WhitePaper,* https://www.blocksg.com/single-post/2017/12/27/quorum-whitepaper, last accessed 2023/01/23.

Tran Quang Duy
Information Technology Faculty, Thang Long University
Hanoi
Vietnam
`duytq@thanglong.edu.vn`

Hoang Trong Minh
Telecommunication Faculty, Posts and Telecoms Institute of Technology
Hanoi
Vietnam
`hoangtrongminh@ptit.edu.vn`

J. Math. Math. Sci. (2023) 87–116

On Korenblum constants for some weighted function spaces

Wee JunJie (Singapore)
Le Hai Khoi (Hanoi, Vietnam)

(Received Feb. 9, 2023)

Abstract. In this paper, we survey the results on the Korenblum Maximum Principle for some weighted function spaces. Progress and results discussed include the upper bounds and lower bounds of Korenblum constants, as well as the failure of the principle for weighted Bergman spaces, weighted Hardy spaces, weighted Bloch spaces, weighted Fock spaces, and mixed norm spaces. Existing and new open questions are provided.

1. Introduction

The Korenblum Maximum Principle is an important open problem in complex analysis as it acts as one of the fundamental properties of complex function spaces that remains unsolved. First conjectured in 1991, the principle was introduced [15] by Boris Korenblum for the classical Bergman space $A^2(\mathbb{D})$ in the following way.

Conjecture 1.1. *There exists a numerical constant c, $0 < c < 1$, such that if f and g are holomorphic in $\mathbb{D}$ and $|f(z)| \leq |g(z)|$ $(c < |z| < 1)$, then $\|f\|_{A^2} \leq \|g\|_{A^2}$.*

Key words and phrases: Fock space, Hardy space, Bergman space, Bloch space, Korenblum constant, Ramanujan's Master Theorem, Mellin transform, Dirichlet series.
2020 Mathematics Subject Classification: Primary 30H20, 46E15
The second-named author is supported by University of Science and Technology of Hanoi (USTH) under grant number USTH.GED.01/22-23

In [15], Korenblum defined c as the *Korenblum constant* and κ as the largest possible value of c. The exact value of κ remains unknown. In the same paper, Korenblum also proved that $\kappa_{A^2} \leq \frac{1}{\sqrt{2}} \approx 0.7071$.

Initially, only a series of partial results were at first discovered by Korenblum, O'Neil, Richards and Zhu [17], Korenblum and Richards [16], Matero [18], Schwick [22], and others. The existence of Korenblum constant for $A^2(\mathbb{D})$ was first proved in 1999 by Hayman [11] with $\kappa_{A^2} = 0.04$. Thereafter, many results were published by improving the lower bounds and upper bounds of κ_{A^2} (see [21, 24-28, 30]). As time progresses, many interesting results have been obtained by several authors. The regained interest in this problem in the recent years have contributed to many fascinating results for families of function spaces as well as intersections of function spaces. Hence, this calls for a timely review to summarize the key important results concerning the Korenblum Maximum Principle. There are certainly several partial results or results related to modified versions of the Korenblum Maximum Principle, and we apologize to those authors as their work are not explicitly mentioned. Ultimately, we hope that this survey might be of interest to not just complex analysts but also mathematicians from other related fields, and that it might inspire new readers with the interesting results that have been obtained so far. At the same time, we hope both existing and new researchers in this problem can take on existing and new open questions from this survey.

We describe the outline of this survey. First of all, no proofs are provided in this paper. Readers should refer to the original articles for detailed proofs. References are provided for all the results. In next section, we recall all basic definitions and notations for weighted Bergman spaces, weighted Hardy spaces, weighted Bloch spaces and weighted Fock spaces. We also list down specific weight functions that will be discussed in our survey. In fact, different weight functions play an important role in many key results for later sections. The results are organised into four sections, namely Sections 3 to 6. Tables summarizing key results are presented at the end of the section where appropriate. As the original Korenblum constant is defined for Bergman spaces, Section 3 discusses the key results for the Korenblum constants for Bergman spaces first. The results are thus separated into two sections: upper bounds and lower bounds. In Section 4, the Korenblum constants are discussed for other function spaces, namely, weighted Hardy spaces and weighted Fock spaces. Following this, Section 5 is solely dedicated to discuss the extension of results from classical weighted Fock spaces to intersections of weighted Fock spaces. In particular, this section extends from the results for classical weighted Fock spaces in Section 4 by leveraging some preliminaries in the well-known Ramanujan's Master Theorem. Hence, we first recall several key important preliminaries such as the Gamma function, Mellin transform of Dirichlet series and Generalised Hypergeometric function in Section 5.1. Section 5.2 constructs new weighted Fock spaces and reviews the upper bounds of Korenblum constants in the finite and infinite intersections of those spaces. After discussing all the main results in Korenblum constants for the weighted function spaces, Section 6 discusses the remaining results pertaining to the failure of Korenblum Maximum Principle. As the failure of Korenblum Maximum Principle in most function spaces are found using similar methods, we survey all of them together in Section 6. Our final Section 7 describes a possible future direction for the Korenblum Maximum Principle and lists down all existing and new open questions.

2. Basic Notations for Weighted Function Spaces

Let $\mathbb{D}$ be the open unit disk in the complex plane $\mathbb{C}$. We denote by $\mathcal{O}(\mathbb{D})$ (resp. $\mathcal{O}(\mathbb{C})$) the space of holomorphic functions (resp. entire functions) on $\mathbb{D}$ (resp. $\mathbb{C}$), endowed with the compact-open topology.

For a domain G, a continuous function $\varphi : G \to [0, \infty)$ can be defined as a *weight function* for weighted function spaces. In this paper, we are only interested in radial weight functions defined on $\mathbb{D}$ or $\mathbb{C}$, i.e. $\varphi(z) = \varphi(|z|)$. To be more precise, we list down the weights that will be used in this paper.

For $G = \mathbb{D}$,

(i) $\varphi(z) = (1 - |z|)^\alpha$, $\alpha \geq 0$, are the standard weights on the disc,

(ii) $\varphi(z) = (\alpha + 1)(1 - |z|^2)^\alpha$, $\alpha > -1$, are the classical Bergman weights,

(iii) $\varphi(z) = e^{-\frac{p\alpha}{2}|z|^2}$, $\alpha > -1$, are the exponential weights defined in [32].

For $G = \mathbb{C}$,

(i) $\varphi(z) = \frac{\alpha}{2}|z|^2$, where $\alpha > 0$, are the classical Fock space weights.

(ii) $\varphi(z) = \frac{\alpha}{2}\lambda|z| - \frac{1}{p}\log|d|$, where $\alpha > 0$, $0 < p < \infty$, $d \in \mathbb{C}\backslash\{0\}$, $\lambda > 0$, are the generalised Fock space weights discussed in [33].

First, we recall the general weighted Hardy space $H^p_\varphi(\mathbb{D})$ where $\varphi : \mathbb{D} \to [0, \infty)$.

Definition 2.1. *Let $\varphi(z) = (1 - |z|)^\alpha$ for $\alpha \geq 0$. For $0 < p < \infty$, the general weighted Hardy space $H^p_\varphi(\mathbb{D})$ consists functions $f(z) \in \mathcal{O}(\mathbb{D})$, for which*

$$\|f\|_{H^p_\varphi} = \sup_{0 \leq r < 1} \left[\varphi(r) \left(\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right)^{\frac{1}{p}} \right] < \infty.$$

If $\varphi(z) = (1-|z|)^\alpha$ where $\alpha \geq 0$, we obtain the weighted Hardy space $H^p_\alpha(\mathbb{D})$. Further, if $\alpha = 0$, we have the Hardy space $H^p(\mathbb{D})$. In the case $p = \infty$, we have the space $H^\infty(\mathbb{D})$ of bounded holomorphic functions on $\mathbb{D}$, where $\|f\|_{H^\infty} = \sup_{z\in\mathbb{D}} |f(z)|$.

Definition 2.2. *Let $0 < p < \infty$ and $\varphi : \mathbb{D} \to [0,\infty)$. Then the weighted Bergman space $A^p_\varphi(\mathbb{D})$ is the space consisting of analytic functions $f(z) \in \mathcal{O}(\mathbb{D})$ for which*

$$\|f\|_{A^p_\varphi} = \left[\frac{1}{\pi}\int_{\mathbb{D}} |f(z)|^p \varphi(z)\, dA(z)\right]^{\frac{1}{p}} < \infty.$$

Here $dA(z) = dxdy = rdrd\theta$, $z = x+iy = re^{i\theta}$, is the Lebesgue measure on $\mathbb{C}$.

Let $\varphi(z) = (\alpha+1)(1-|z|^2)^\alpha$ for $\alpha > -1$. Then we have the classical weighted Bergman space $A^p_\alpha(\mathbb{D})$ which is a Banach space. Further, if $\alpha = 0$, for $0 < p < \infty$, the space becomes the standard Bergman space $A^p(\mathbb{D})$. In particular, for $p = 2$, we have the classical Bergman space $A^2(\mathbb{D})$.

Interestingly, a weighted Bergman space with exponential weights $\varphi(z) = e^{-\frac{p\gamma}{2}|z|^2}$ $(0 < p < \infty, \gamma > -1)$ is introduced in [32]. We shall denote this weighted Bergman space with exponential weights as $A^p_\gamma(\mathbb{D})$, that is, the space of holomorphic functions $f(z) \in \mathcal{O}(\mathbb{D})$ for which

$$\|f\|_{A^p_\gamma} = \left[\frac{1}{\pi}\int_{\mathbb{D}} |f(z)|^p e^{\frac{p\gamma}{2}|z|^2}\, dA(z)\right]^{\frac{1}{p}} < \infty.$$

Note that $e^{\frac{p\gamma}{2}|z|^2} \to 1$ as $|z| \to 0$ and $e^{\frac{p\gamma}{2}|z|^2}$ approaches to the constant $e^{\frac{p\gamma}{2}}$ as $|z|$ approaches the boundary of $\mathbb{D}$.

Next, we have the weighted Fock spaces.

Definition 2.3. *Let $\varphi(z) : \mathbb{C} \to [0,\infty)$ be a weight function. For $0 < p < \infty$, the general weighted Fock space $\mathcal{F}^p_\varphi(\mathbb{C})$ with weight $\varphi(z)$, consists of entire functions $f(z) \in \mathcal{O}(\mathbb{C})$ for which*

$$\|f\|^p_{\mathcal{F}^p_\varphi} = \eta \int_{\mathbb{C}} |f(z)|^p e^{-p\varphi(z)}\, dA(z) < \infty,$$

where the constant η is chosen so that $\|1\|_{\mathcal{F}^p_\varphi} = 1$.

If $\varphi(z) = \frac{\alpha}{2}|z|^2$, $\alpha > 0$, then we have the classical weighted Fock space $\mathcal{F}^p_\alpha(\mathbb{C})$ with norm $\|f\|^p_{\mathcal{F}^p_\alpha} = \dfrac{p\alpha}{2\pi}\displaystyle\int_{\mathbb{C}} |f(z)|^p e^{-\frac{p\alpha}{2}|z|^2}\, dA(z)$. for the case $p = \infty$ and $\alpha > 0$, we have the space $\mathcal{F}^\infty_\alpha$ with norm $\|f\|_{\mathcal{F}^\infty_\alpha} = \sup\limits_{z\in\mathbb{C}} |f(z)|e^{-\frac{\alpha}{2}|z|^2}$.

For $0 < p \le \infty$ and $\alpha = 1$, we have the Fock space $\mathcal{F}^p(\mathbb{C})$, which is a Banach space if $1 \le p \le \infty$ and is a complete metric space with distance $d(f,g) = \|f-g\|_p^p$ if $0 < p < 1$.

Lastly, we have the weighted Bloch spaces.

Definition 2.4. *The weighted Bloch space B_φ with weight $\varphi : \mathbb{D} \to [0,\infty)$, consists of holomorphic functions $f(z) \in \mathcal{O}(\mathbb{D})$ for which*

$$\|f\|_{B_\varphi} = |f(0)| + \sup_{z\in\mathbb{D}} \varphi(z)(1-|z|^2)|f'(z)| < \infty.$$

Note that $\|f\|_{B_\varphi}$ is the weighted Bloch norm and elements of B_φ are known as weighted Bloch functions. If $\varphi(z) \equiv 1$, then we have the classical Bloch space B.

3. Korenblum Constants for Bergman Spaces

In this section, we survey the main results for the Korenblum constants of Bergman spaces. To avoid confusion and provide greater clarity, the results are divided into two sections: Upper bounds and Lower bounds.

3.1. Development of Upper Bounds

Recall from the introduction that Korenblum first discovered the upper bound for κ_{A^2} to be $\frac{1}{\sqrt{2}}$.

Theorem 3.1 ([15]). *Let $c > \frac{1}{\sqrt{2}}$. There exist functions f and g in $A^2(\mathbb{D})$ such that $|f(z)| \le |g(z)|$ for all $c < |z| < 1$, but $\|f\|_{A^2} > \|g\|_{A^2}$. Therefore, $\kappa_{A^2} \le \frac{1}{\sqrt{2}}$.*

The next natural question is whether can κ_{A^2} be equal to $\frac{1}{\sqrt{2}}$. An example by Martin [15] shows that $\kappa_{A^2} < \frac{1}{\sqrt{2}}$.

Theorem 3.2 ([15]). *Suppose $c = \frac{1}{\sqrt{2}}$. Let*

$$f(z) = \frac{1+(\sqrt{2}-1)z^{20}}{1+(\sqrt{2}-1)z^{-20}}, \quad g(z) = \sqrt{2}z.$$

Then $|f(z)| \le |g(z)|$ for all $c < |z| < 1$, but

$$\|f\|_{A^2} = \frac{\sqrt{1+(\sqrt{2}-1)^2/21}}{1+(\sqrt{2}-1)2^{-10}} > 1 = \|g\|_{A^2}.$$

In a series of papers [24-28,30], Wang used different pairs of functions f and g to improve the upper bounds for the Korenblum constant of $A^2(\mathbb{D})$. In [24], Wang first used the singular inner function $S_a(z) = \exp(-a\frac{1+z}{1-z})$ in $A^2(\mathbb{D})$, $a \in \mathbb{R}^+$, to prove that the Korenblum constant must be less than 0.69472.

Theorem 3.3 ([24]). *Let*

$$f(z) = e^{-a} S_a(z^n) = e^{-a} \exp\left(-a\frac{1+z^n}{1-z^n}\right) = \exp\left(-\frac{2a}{1-z^n}\right),$$

where a is any positive constant and

$$g(z) = e^{-\frac{2a}{1+c^n}} \frac{z}{c},$$

where $0 < c < 1$, $a = -\frac{1+c^n}{1-c^n} \log c > 0$, $n \in \mathbb{N}$. Then $|f(z)| \le |g(z)|$ in $c < |z| < 1$. When $n = 14$ and $c = 0.69472$, we have $||f||_{A^2} > ||g||_{A^2}$. Therefore, $\kappa_{A^2} < 0.69472$.

Following that, Wang managed to find sharper upper bounds for κ_{A^2} through the results below.

Theorem 3.4 ([25]). *Let $0 < c < 1$, $\quad a = -\frac{1+c^n}{1-c^n} \log c > 0$ and $n \in \mathbb{N}$. Then define*

$$f(z) = S_{a+b}(z^n), \quad g(z) = zS_b(z^n)$$

and we have $|f(z)| \le |g(z)|$ in $c < |z| < 1$. Moreover, when $a = 0.3902$, $b = 0.3395$, $n = 11$ and $c = 0.685086$, we have $||f||_{A^2} > ||g||_{A^2}$. Therefore, $\kappa_{A^2} < 0.685086$.

Theorem 3.5 ([26]). *Let $a = \frac{3\sqrt{6}}{11}$ and $n = 10$. Then we define*

$$f(z) = a + z^n, \quad g(z) = z(1 + az^n)$$

and we have $||f||_{A^2} = ||g||_{A^2}$ and $|f(z)| \le |g(z)|$ in $c < |z| < 1$, where $c = 0.679501$ is the real root in $(0, 1)$ of the equation

$$a + z^{10} = \frac{3\sqrt{6}}{11} + z^{10} = z + \frac{3\sqrt{6}}{11} z^{11} = z(1 + az^{10}). \tag{3.1}$$

Therefore, $\kappa_{A^2} < 0.679501$.

As a result of Theorem 3.5, the upper bound of Korenblum constant is now 0.679501 but $||f||_{A^2} = ||g||_{A^2}$ for this upper bound. Hence, in [26], Wang noted that this bound is not sharp and proceeded to obtain a better upper bound, i.e. $\kappa_{A^2} < 0.67795$.

Theorem 3.6 ([26]). *Let $0 < a < 1$, $b \geq 0, n \in \mathbb{N}$. Then we define*

$$f(z) = \frac{a + z^n}{(1 - az^n)^b}, \quad g(z) = \frac{z(1 + az^n)}{(1 - az^n)^b}.$$

Hence, we have $|f(z)| \leq |g(z)|$ in $c < |z| < 1$, where c is the real root in $(0, 1)$ of the equation

$$a + z^n = z(1 + az^n),$$

and when $a = 0.666707, b = 0.4768$ and $n = 10$, we have $c = 0.67795$ and $||f||_{A^2} > ||g||_{A^2}$. Therefore, $\kappa_{A^2} < 0.67795$.

In 2008, Shen [23] modified the above example to obtain a slightly better upper bound $\kappa_{A^2} < 0.677905$.

Theorem 3.7 ([23]). *Let $0 < a < 1$ and $n \in \mathbb{N}$. Then we define*

$$f(z) = \frac{a + z^n}{2 - az^n}, \quad g(z) = \frac{z(1 + az^n)}{2 - az^n}.$$

Hence, we have $|f(z)| \leq |g(z)|$ in $c < |z| < 1$, where c is the real root in $(0, 1)$ of the equation

$$a + z^n = z(1 + az^n),$$

and when $a = 0.6666714$ and $n = 10$, we have $c = 0.677905$ and $||f||_{A^2} > ||g||_{A^2}$. Therefore, $\kappa_{A^2} < 0.677905$.

A final improvement was made by Wang [29] where he obtained $\kappa_{A^2} < 0.6778994$ with the following counter example.

Theorem 3.8 ([29]). *Let $0 < a < 1$, $b \geq 0, n \in \mathbb{N}$. Then we define*

$$f(z) = \frac{a + z^n}{(1 - bz^n)^2}, \quad g(z) = \frac{z(1 + az^n)}{(1 - bz^n)^2}.$$

Hence, we have $|f(z)| \leq |g(z)|$ in $c < |z| < 1$, where c is the real root in $(0, 1)$ of the equation

$$a + z^n = z(1 + az^n),$$

and when $a = \sqrt{\frac{n-2}{2n-2}}$, $b = \sqrt{\frac{2}{(n-1)(n-2)}}$ and $n = 10$, we have $c = 0.6778994$ and $||f||_{A^2} > ||g||_{A^2}$. Therefore, $\kappa_{A^2} < 0.6778994$.

In summary, $\kappa_{A^2} < 0.6778994$ is the best upper bound of Korenblum constant for $A^2(\mathbb{D})$ so far.

Note that the upper bounds by Wang were numerically sharper but it lacks generalisations for the weighted Bergman spaces. In recent years, we obtained explicit expression for the upper bounds in the weighted Bergman spaces with exponential weights, $A^p_\gamma(\mathbb{D})$, $p \geq 1$, $\gamma \geq 0$ [32].

Theorem 3.9 ([32]). *Let $p \geq 1$, $\gamma \geq 0$. Consider the Bergman space $A^p_\gamma(\mathbb{D})$.*

1) For $\gamma = 0$, suppose

$$\left(\frac{2}{p+2}\right)^{\frac{1}{p}} < c < 1.$$

2) For $\gamma > 0$, suppose

$$\sqrt[p]{\frac{\left(\frac{2}{p\gamma}\right)^{\frac{p}{2}} \int_0^{\frac{p\gamma}{2}} u^{\frac{p}{2}} e^{-u} du}{(1 - e^{-\frac{p\gamma}{2}})}} < c < 1.$$

There exist functions f and g in $A^p_\gamma(\mathbb{D})$ such that $|f(z)| < |g(z)|$ for all $c < |z| < 1$, but $\|f\|_{A^p_\gamma} > \|g\|_{A^p_\gamma}$.

Remark 3.1. *Clearly, in order to have the Korenblum Maximum Principle for $A^p_\gamma(\mathbb{D})$, $p \geq 1$, $\gamma \geq 0$, we must have*

$$\kappa_{A^p_\gamma} \leq \begin{cases} \left(\frac{2}{p+2}\right)^{\frac{1}{p}}, & \gamma = 0, \\ \sqrt[p]{\frac{\left(\frac{2}{p\gamma}\right)^{\frac{p}{2}} \int_0^{\frac{p\gamma}{2}} u^{\frac{p}{2}} e^{-u} du}{(1-e^{-\frac{p\gamma}{2}})}}, & \gamma > 0. \end{cases}$$

The above result acts as a generalization of the initial result $\kappa_{A^2} \leq \frac{1}{\sqrt{2}}$ by Korenblum in Theorem 3.1. Nevertheless, this generalization obtains the upper bounds for rest of the spaces $A^p_\gamma(\mathbb{D})$, $p \geq 1$.

3.2. Development in Lower Bounds

In this subsection, we survey the progress on lower bound of Korenblum constant for Bergman spaces ever since Hinkkanen's result in 1999 [12]. Hinkkanen proved in 1999 that $\kappa_{A^p} \geq 0.15724$ ($p \geq 1$), thereby showing that Korenblum constants exist for all Bergman spaces $A^p(\mathbb{D})$, $p \geq 1$. After that, improvements were made in 2006 when Schuster [21] showed that the Korenblum maximum principle holds for $\kappa_{A^2} = 0.21$, which progresses upon the works of both Hayman and Hinkkanen. In that paper, Schuster made use of the following identities.

Proposition 3.10. *For any z, $w \in \mathbb{C}$,*

$$|z|^2 - |w|^2 \leq 2|z^2 - zw|.$$

Proposition 3.11. *For any subharmonic function h and $0 < r_1 < r_2 < 1$,*

$$\int_0^{2\pi} h(r_1 e^{i\theta})\, d\theta \leq \int_0^{2\pi} h(r_2 e^{i\theta})\, d\theta.$$

Using Propositions 3.10 and 3.11, Schuster proved Theorem 3.12.

Theorem 3.12 ([21])**.** *Suppose that $c = 0.21$. Then for any functions $f(z)$ and $g(z)$ holomorphic in $\mathbb{D}$, if $|f(z)| \leq |g(z)|$ $(c < |z| < 1)$, then $||f||_{A^2} \leq ||g||_{A^2}$. Therefore, $\kappa_{A^2} \geq 0.21$.*

Also in 2006, Wang [27] managed to improve the lower bounds of κ_{A^2} to 0.25018 and κ_{A^p} to 0.1921. Wang used similar methods but a different inequality from Hinkkanen and Schuster. For instance, Wang used the following proposition instead of Proposition 3.10 in proving $\kappa_{A^2} \geq 0.25018$.

Proposition 3.13. *For any $a \in (-1, 1)$ and $z,\ w \in \mathbb{C}$,*

$$|z|^2 - |w|^2 = \frac{|z - aw|^2 - |az - w|^2}{1 - a^2} \leq \frac{|z - aw|^2}{1 - a^2}. \tag{3.2}$$

Wang then made his final improvements to the lower bounds in 2011, where he showed that $\kappa_{A^2} \geq 0.28185$ and $\kappa_{A^p} \geq 0.23917$ for $p \geq 1$.

Recently, the Korenblum maximum principle was extended to a large family of function spaces that contains the classical weighted Bergman space $A^p_\varphi(\mathbb{D})$ [4]. In particular, a failure of the principle was proven for the mixed norm space $H^{p,q,s}$ where $0 < p, q, s < \infty$ by Karapetrović [4]. The mixed norm space $H^{p,q,s}$ $(0 < p, q, s < \infty)$ consists of all holomorphic functions in $\mathcal{O}(\mathbb{D})$ for which

$$\|f\|_{H^{p,q,s}} = \left(2sq \int_0^1 r(1 - r^2)^{sq-1} M_p^q(r, f)\, dr\right)^{1/q} < \infty, \tag{3.3}$$

where

$$M_p(r, f) = \left(\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p\, d\theta\right)^{1/p}.$$

Note that when $q = p$ and $s = \frac{\alpha+1}{p}$, then $\|f\|_{H^{p,p,\frac{\alpha+1}{p}}} = \|f\|_{A^p_\alpha}$ which implies that $H^{p,p,\frac{\alpha+1}{p}} = A^p_\alpha(\mathbb{D})$.

Similar to earlier results in Korenblum constants, Karapetrović also proved that the Korenblum constants exist under the mixed norm spaces, $H^{p,q,s}$, $1 \leq p \leq q < \infty$ and $0 < s < \infty$.

Theorem 3.14 ([4])**.** *Let $1 \leq p \leq q < \infty$ and $0 < s < \infty$. Then there exists a constant $0 < c < 1$ with the following property: If f and g are holomorphic functions on $\mathbb{D}$ such that $|f(z)| \leq |g(z)|$ for all $c < |z| < 1$, then $\|f\|_{H^{p,q,s}} \leq \|g\|_{H^{p,q,s}}$.*

Corollary 3.1 ([4])**.** *Let $0 < p < \infty$ and $-1 < \alpha < \infty$. Then the Korenblum maximum principle holds in weighted Bergman space $A^p_\alpha(\mathbb{D})$ if and only if $p \geq 1$.*

In summary, the results for Bergman spaces can be summarised using the following tables.

<table>
<tr><th></th><th>α</th><th>p</th><th>Lower Bound</th><th>Upper Bound</th></tr>
<tr><td rowspan="4">$A^p_\alpha(\mathbb{D})$
$\alpha > -1$
$1 \leq p < \infty$</td><td rowspan="2">$\alpha = 0$</td><td>$p = 2$</td><td>1999: $\kappa_{A^2} \geq \frac{1}{25}$
2006: $\kappa_{A^2} \geq 0.21$
2006: $\kappa_{A^2} \geq 0.25018$
2011: $\kappa_{A^2} \geq 0.28185$</td><td>1991: $\kappa_{A^2} < \frac{1}{\sqrt{2}}$
2003: $\kappa_{A^2} < 0.69472$
2004: $\kappa_{A^2} < 0.685086$
2004: $\kappa_{A^2} < 0.67795$
2008: $\kappa_{A^2} < 0.677905$
2008: $\kappa_{A^2} < 0.6778994$</td></tr>
<tr><td>$1 \leq p < \infty$</td><td>1999: $\kappa_{A^p} \geq 0.15724$
2006: $\kappa_{A^p} \geq 0.1921$
2011: $\kappa_{A^p} \geq 0.23917$</td><td>2020: $\kappa_{A^p} \leq \left(\frac{2}{p+2}\right)^{1/p}$</td></tr>
<tr><td>$\alpha > 0$</td><td>$1 \leq p < \infty$</td><td rowspan="2">2022: Corollary 2.18</td><td rowspan="2">No Specific Development</td></tr>
<tr><td>$-1 < \alpha < 0$</td><td>$1 \leq p < \infty$</td></tr>
</table>

Table 1. Main Development on classical weighted Bergman spaces $A^p_\alpha(\mathbb{D})$

<table>
<tr><th></th><th>γ</th><th>p</th><th>Lower Bound</th><th>Upper Bound</th></tr>
<tr><td rowspan="3">$A^p_\gamma(\mathbb{D})$
$\gamma > -1$
$1 \leq p < \infty$</td><td>$\gamma = 0$</td><td>-</td><td colspan="2">See Table 1 for $A^p(\mathbb{D})$</td></tr>
<tr><td>$\gamma > 0$</td><td>$1 \leq p < \infty$</td><td rowspan="2">No Specific Development</td><td>2020: Theorem 3.9</td></tr>
<tr><td>$-1 < \gamma < 0$</td><td>$1 \leq p < \infty$</td><td>No Specific Development</td></tr>
</table>

Table 2. Main Development on Bergman spaces with exponential weights $A^p_\gamma(\mathbb{D})$.

Based on the summary, we make some minor progress in the areas with no development so far. In particular, we can apply similar methods from Theorem 3.9 to obtain explicit expressions for the upper bound of Korenblum constants in classical weighted Bergman spaces $A^p_\alpha(\mathbb{D})$, $\alpha \geq 0$ and $p \geq 1$.

Theorem 3.15. *Let $p \geq 1$, $\alpha \geq 0$. Consider the Bergman space $A^p_\alpha(\mathbb{D})$. Suppose*

$$\sqrt[p]{\frac{(\alpha+1)\Gamma(\frac{p}{2}+1)\Gamma(\alpha+1)}{\Gamma(\frac{p}{2}+\alpha+2)}} < c < 1.$$

There exist functions f and g in $A^p_\alpha(\mathbb{D})$ such that $|f(z)| < |g(z)|$ for all $c < |z| < 1$, but $\|f\|_{A^p_\alpha} > \|g\|_{A^p_\alpha}$.

Remark 3.2. *Clearly, in order to have the Korenblum Maximum Principle for $A^p_\alpha(\mathbb{D})$, $p \geq 1$, $\alpha \geq 0$, we must have*

$$\kappa_{A^p_\alpha} \leq \sqrt[p]{\frac{(\alpha+1)\Gamma(\frac{p}{2}+1)\Gamma(\alpha+1)}{\Gamma(\frac{p}{2}+\alpha+2)}}.$$

4. Korenblum Constants for Other Weighted Function Spaces

In this section, we survey the results of Korenblum constants for other weighted function spaces. In general, for a function space $\mathcal{L}$, the Korenblum conjecture is as follows,

Conjecture 4.1. *There exists a numerical constant c, $0 < c < 1$, such that for any functions f and g in $\mathcal{L}$, if*

$$|f(z)| \leq |g(z)|, \quad \forall z \in E, \tag{4.1}$$

then

$$\|f\|_{\mathcal{L}} \leq \|g\|_{\mathcal{L}}. \tag{4.2}$$

Here, E is a set of values of z in order for (4.1) to imply (4.2). In [7], the set E satisfying the above conjecture for the function space $\mathcal{L}$ is also known as a dominating set for $\mathcal{L}$.

Weighted Hardy Spaces. Korenblum stated in [15] that for the case where $\mathcal{L}$ is the Hardy-Hilbert space $H^2(\mathbb{D})$, then (4.1) implies (4.2) even for the case $E = \mathbb{D}$. Interestingly, it was only until 1998 that the following criteria was discovered for a dominating set in general Hardy spaces $H^p(\mathbb{D})$ $(0 < p \leq \infty)$.

Theorem 4.1 ([7])**.** *Let $0 < p \leq \infty$. Then E is non-tangentially dense if and only if E is a dominating set for the general Hardy spaces $H^p(\mathbb{D})$ $(0 < p \leq \infty)$.*

As there is no specific development on the set E for $H^p_\alpha(\mathbb{D})$ $(\alpha > 0, \quad 0 < p \leq \infty)$ and even for $H^p_\varphi(\mathbb{D})$ where $\varphi(z) \neq (1-|z|)^\alpha$, then these areas call for investigation. In summary, we have the following table.

		α	p	**Results**
$H^p_\varphi(\mathbb{D})$	$H^p_\alpha(\mathbb{D})$	$\alpha = 0$	$0 < p \leq \infty$	1998: Theorem 4.1
		$\alpha > 0$	$0 < p \leq \infty$	No Specific Development
	-	-	-	No Specific Development

Table 3. Development of Korenblum constants on $H^p_\varphi(\mathbb{D})$.

Weighted Fock Spaces. Throughout the years, some results also showed an extension to the Fock spaces (see [7, 34]). It first began in 2006 when Schuster modified the proof of Theorem 3.12 and obtained $\kappa_{\mathcal{F}^2} \geq 0.54$.

Proposition 4.2 ([21])**.** *Let $f(z)$ and $g(z)$ be entire functions in $\mathcal{F}^2(\mathbb{C})$. Suppose that $|f(z)| \leq |g(z)|$ for any z such that $|z| > c$, where $c_{\mathcal{F}^2} = 0.54$. Then $||f||_{\mathcal{F}^2} \leq ||g||_{\mathcal{F}^2}$.*

Wang [27] then modified his result in $A^2(\mathbb{D})$ for $\mathcal{F}^2(\mathbb{C})$ and obtained the following improved lower bound.

Theorem 4.3 ([27])**.** *Let $f(z)$ and $g(z)$ be entire functions in $\mathcal{F}^2(\mathbb{C})$. Suppose that $|f(z)| \leq |g(z)|$ for any z such that $|z| > c$, where $c_{\mathcal{F}^2} = 0.7248$. Then $||f||_{\mathcal{F}^2} \leq ||g||_{\mathcal{F}^2}$.*

Till today, the lower bound $\kappa_{\mathcal{F}^2} \geq 0.7248$ remains as the best lower bound so far. In 2012, Zhu [34] showed that Korenblum constants actually exist for the classical weighted Fock spaces $\mathcal{F}^p_\alpha(\mathbb{C})$ where $\alpha > 0$ and $p \geq 1$.

Theorem 4.4 ([34])**.** *Let $f(z)$ and $g(z)$ be entire functions in $\mathcal{F}^p_\alpha(\mathbb{C})$. Suppose that $|f(z)| \leq |g(z)|$ for any z such that $|z| > c$. Then $||f||_{\mathcal{F}^p_\alpha} \leq ||g||_{\mathcal{F}^p_\alpha}$.*

To be specific, Zhu proved that the Korenblum Maximum Principle can hold for $\mathcal{F}^p_\alpha(\mathbb{C})$ where $\alpha > 0$ and $p \geq 1$ as there exist a sufficiently small positive c to satisfy

$$2c(1 - e^{-\frac{p\alpha}{2}c^2})\left(\alpha \int_c^\infty e^{-\frac{p\alpha}{2}\rho^2}(\rho^2 - c^2)d\rho\right)^{-1} < 1. \tag{4.3}$$

In general, the outline of the proofs for lower bounds of Korenblum constants are similar to the works of Schuster [21]. The main workhorse for the improvements of the lower bound relies on changing the inequality to manipulate $|f|^p - |g|^p$ in the proof. Consequently, this will lead to changes in the numerical estimate of the lower bound. However, in order to obtain a particular lower bound for the Korenblum constant, several results have used Mathematica to provide a numerical estimate.

On the other hand, the following simple result involving the Gamma function, is proved in 2020 [32]. The result provides an upper bound for $\kappa_{\mathcal{F}^p_\alpha}$, where $\alpha > 0$ and $p \geq 1$.

Theorem 4.5 ([32])**.** *Let $p \geq 1$, $\alpha > 0$ and*

$$c > \sqrt[p]{\left(\frac{2}{p\alpha}\right)^{\frac{p}{2}}\Gamma\left(\frac{p}{2}+1\right)}.$$

There exist functions f and g in $\mathcal{F}^p_\alpha(\mathbb{C})$, such that $|f(z)| < |g(z)|$ for any $|z| > c$, but $\|f\|^p_{\mathcal{F}^p_\alpha} > \|g\|^p_{\mathcal{F}^p_\alpha}$. Therefore,

$$\kappa_{\mathcal{F}^p_\alpha} \leq \sqrt[p]{\Big(\frac{2}{p\alpha}\Big)^{\frac{p}{2}}\Gamma\Big(\frac{p}{2}+1\Big)}.$$

By setting $p = 2$ and $\alpha = 1$ in Theorem 4.5, the following special case is obtained for $\mathcal{F}^2(\mathbb{C})$.

Corollary 4.1 ([32])**.** *Let $c_{\mathcal{F}^2} > 1$. There exist functions f and g in $\mathcal{F}^2(\mathbb{C})$ such that $|f(z)| < |g(z)|$ for all $|z| > c_{\mathcal{F}^2}$, but $\|f\|^2_{\mathcal{F}^2} > \|g\|^2_{\mathcal{F}^2}$. Therefore, $c_{\mathcal{F}^2} \leq 1$.*

For instance, when $p = 1$ and $\alpha = \frac{1}{2}$, we can write the upper bound as

$$\kappa_{\mathcal{F}^1_{0.5}} \leq \sqrt{4}\cdot\Gamma\left(\frac{3}{2}\right) = \sqrt{\pi}.$$

Theorem 4.5 inspired us to study the Korenblum constants for other Fock spaces under general weights such as $\frac{\alpha}{2}\lambda|z| - \frac{1}{p}\log|d|$, where $\alpha > 0$, $0 < p < \infty$, $d \in \mathbb{C}\backslash\{0\}$, $\lambda > 0$ in [32]. The interest in these specific set of general weights is that it introduces various weighted Fock spaces where the upper bound of its Korenblum constants involves special functions such as Gamma function, Mellin transform of Dirichlet series and Generalized Hypergeometric function. For example, we applied the Gamma function and obtained a simple generalisation of Theorem 4.5.

For any complex number s with $\mathrm{Re}(s) > 0$, the Gamma function is defined as

$$\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\,dx.$$

Furthermore, with a change of variables, we have

$$\int_0^\infty x^a e^{-bx^c}\,dx = \frac{1}{c}\cdot\left(\frac{1}{b}\right)^{\frac{a+1}{c}}\Gamma\left(\frac{a+1}{c}\right), \quad a,b,c>0. \tag{4.4}$$

Now, define the weighted Fock spaces $\mathcal{F}^{p,\alpha}_m(\mathbb{C})$, that is,

Definition 4.1. *For $0 < p < \infty$, $\alpha > 0$, $m > 0$, the weighted Fock space $\mathcal{F}^{p,\alpha}_m(\mathbb{C})$ with weight $\frac{\alpha}{2}|z|^m$ consists of entire functions $f(z) \in \mathcal{O}(\mathbb{C})$ for which*

$$\|f\|^p_{\mathcal{F}^{p,\alpha}_m} = \int_{\mathbb{C}} |f(z)|^p e^{-\frac{p\alpha}{2}|z|^m}\,dA(z) < \infty.$$

Now, the following result [33] generalises Theorem 4.5 for the weighted Fock space $\mathcal{F}_m^{p,\alpha}(\mathbb{C})$.

Theorem 4.6 ([33])**.** *Let $0 < p < \infty$, $\alpha > 0$, $m > 0$ and let*

$$c > \sqrt[p]{\left(\frac{2}{p\alpha}\right)^{\frac{p}{m}} \Gamma\left(\frac{p+2}{m}\right)\left(\Gamma\left(\frac{2}{m}\right)\right)^{-1}}. \tag{4.5}$$

There exist functions f and g in $\mathcal{F}_m^{p,\alpha}(\mathbb{C})$, such that $|f(z)| < |g(z)|$ for any $|z| > c$, but $\|f\|^p_{\mathcal{F}_m^{p,\alpha}} > \|g\|^p_{\mathcal{F}_m^{p,\alpha}}$. Therefore,

$$\kappa_{\mathcal{F}_m^{p,\alpha}} \leq \sqrt[p]{\left(\frac{2}{p\alpha}\right)^{\frac{p}{m}} \Gamma\left(\frac{p+2}{m}\right)\left(\Gamma\left(\frac{2}{m}\right)\right)^{-1}}.$$

In summary, the results for classical weighted Fock spaces can be summarised using the following table.

<table>
<tr><th></th><th>α</th><th>p</th><th>Lower Bound</th><th>Upper Bound</th></tr>
<tr><td rowspan="3">$\mathcal{F}^p_\alpha(\mathbb{C})$
$\alpha > 0$
$1 \leq p < \infty$</td><td rowspan="2">$\alpha = 1$</td><td>$p = 2$</td><td>2006 : $\kappa_{\mathcal{F}^2} \geq 0.54$
2006 : $\kappa_{\mathcal{F}^2} \geq 0.7248$</td><td>2020 : $\kappa_{\mathcal{F}^2} \leq 1$</td></tr>
<tr><td>$1 \leq p < \infty$</td><td rowspan="2">2012: Theorem 4.6</td><td rowspan="2">2020: Theorem 4.7</td></tr>
<tr><td>$\alpha \neq 1$</td><td>$1 \leq p < \infty$</td></tr>
</table>

Table 4. Development of Korenblum constants on $\mathcal{F}^p_\alpha(\mathbb{C})$.

5. Extension to Intersections of Weighted Fock spaces

In this section, we survey our extension of results from classical weighted Fock spaces to other weighted Fock spaces. To do this, we directed our attention to the Gamma function, which is one of the many special functions which satisfies the well-known Ramanujan's Master theorem. The Ramanujan's Master Theorem was first reported in Ramanujan's Quarterly Reports [2] and can be satisfied by many special functions, such as Mellin transform of Dirichlet series and Generalized Hypergeometric functions. Further details can be found in [1,2].

5.1. Preliminaries

The Ramanujan's Master theorem was first rigorously treated by G.H. Hardy in [9] whose proof relied on the Cauchy residue theorem and Mellin inversion theorem. Hardy proved that the Ramanujan's Master theorem can be satisfied for a sufficiently large class of functions that satisfies certain growth condition. We shall recall the Ramanujan's Master Theorem for Hardy's class of functions below.

Proposition 5.1 (Ramanujan's Master Theorem [1])**.** *Let $\omega(z)$ be a holomorphic and single-valued function defined on the half-plane $H(\delta) = \{z \in \mathbb{C} : \mathrm{Re}(z) \geq -\delta\}$ for some $0 < \delta < 1$. Suppose that, there exist positive constants C, P and $A < \pi$ such that the growth condition*

$$|\omega(u + iv)| < Ce^{Pu+A|v|}, \tag{5.1}$$

holds for all $z = u + iv \in H(\delta)$. Then for all $0 < \mathrm{Re}(s) < \delta$,

$$\int_0^\infty x^{s-1}\{\omega(0) - x\omega(1) + x^2\omega(2) - \cdots\}\, dx = \frac{\pi}{\sin \pi s}\omega(-s). \tag{5.2}$$

Similar to many previous applications in Ramanujan's Master theorem, (5.2) is more commonly written as

$$\int_0^\infty x^{s-1} \sum_{k=0}^\infty \frac{\phi(k)}{k!}(-x)^k\, dx = \Gamma(s)\phi(-s). \tag{5.3}$$

Equation (5.3) acts as a valid integral identity for computing the Mellin transforms for particular functions ϕ, assuming that the series term on the left-hand side holds and the integral is convergent for some values of $\mathrm{Re}(s)$. Over the years, equation (5.3) has been applied conveniently and directly by Ramanujan and several authors.

It is also well-known that this integral transform has close relations to the theory of *Dirichlet series*. Let $0 < (\lambda_n) \uparrow \infty$ be a sequence of positive real numbers and (d_n) be a sequence of complex numbers. We now consider a Dirichlet series, with real frequencies (λ_n),

$$\sum_{n=1}^\infty d_n e^{-\lambda_n s}, \quad s \in \mathbb{C}.$$

It is well-known that if we let $L = \limsup_{n\to\infty} \dfrac{\log n}{\lambda_n}$, then in case $L < \infty$, the following inequalities must hold

$$\limsup_{n\to\infty} \frac{\log |d_n|}{\lambda_n} \leq \sigma_c \leq \sigma_u \leq \sigma_a \leq \limsup_{n\to\infty} \frac{\log |d_n|}{\lambda_n} + L,$$

where $\sigma_c, \sigma_a, \sigma_u$ are abscissa of convergence, absolute convergence, or uniform convergence respectively. Readers may refer to the book [10] for more information regarding Dirichlet series.

In the interest of constructing weighted Fock spaces using Dirichlet series, we now restrict the Dirichlet series with non-negative coefficients as follows,

$$g(s) = \sum_{n=1}^{\infty} |d_n| e^{-\lambda_n s}, \ s \in \mathbb{C}, \quad \text{with } \sigma_c \leq 0. \tag{5.4}$$

Then the Dirichlet series $g(s)$ represents a holomorphic function on the half-plane $\{z \in \mathbb{C} : \mathrm{Re}(z) > \sigma_c\}$. Naturally, this class of Dirichlet series can be characterized into two classes [10].

(I) $\displaystyle\sum_{n=1}^{\infty} |d_n| = \infty$ and $\sigma_c = \limsup_{n\to\infty} \frac{\log(|d_1| + |d_2| + \cdots + |d_n|)}{\lambda_n} = 0.$

(II) $\displaystyle\sum_{n=1}^{\infty} |d_n| < \infty$ and $\sigma_c = \limsup_{n\to\infty} \frac{\log(|d_{n+1}| + |d_{n+2}| + \cdots)}{\lambda_n} \leq 0.$

The Dirichlet series itself comprises of several special functions that satisfies the Ramanujan's Master Theorem. Here, we list down the particular cases discussed in [33]:

- If $\lambda_n = F_{n+1}$ for all $n \in \mathbb{N}$, then $g(s)$ becomes the Fibonacci zeta-function $\zeta_F(s) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{F_n^s}$.
- If $\lambda_n = n$ and $d_n = 1$ for all $n \in \mathbb{N}$, then $g(s)$ is the Riemann Zeta function $\zeta(s)$.
- If $\lambda_n = n$ and $d_n = \frac{1}{n!}$ for all $n \in \mathbb{N}$, then $g(s) = e^{e^{-s}} - 1$.

In addition, one can also consider (d_n) to be arithmetical functions such as the Euler Totient function or the partition function $p(n)$.

The main connection between Dirichlet series and the weighted Fock spaces discussed in [33] is due to the Mellin transform of the Dirichlet series. Historically, Cahen [6] and Perron [9] (see also [5, p. 327]) discovered the Mellin transform of Dirichlet series a long time ago. As an immediate consequence from their early works, we made full use of the fact that $g(s)$ satisfies the Mellin inversion theorem. In fact, the Mellin transform of $g(s)$ is a unique family of special functions satisfying (5.3) and also has many applications in theoretical computer science (see [8]).

Proposition 5.2 ([10]). *Let $s > 0$ and $g(x)$ be series* (5.4)*, $x \in \mathbb{R}$. Then the Mellin transform of $g(x)$ can be computed as*

$$\mathcal{M}(g; s) = \int_0^{\infty} x^{s-1} g(x)\, dx = \Gamma(s) \sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^s}, \tag{5.5}$$

provided that the series on the right-hand side converges.

By establishing this connection, we proved the following identity which will be used to deal with our infinite intersections of weighted Fock spaces.

Proposition 5.3 ([33])**.** *Let s be a complex constant and $m \geq 0$. Then*

$$\sum_{n=1}^{\infty} |d_n| \int_{\mathbb{C}} |sz|^m e^{-\lambda_n |z|}\, dA(z) = \begin{cases} \infty, & \lambda_1 = 0. \\ 2\pi |s|^m \Gamma(m+2) \displaystyle\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^{m+2}}, & \lambda_1 > 0, \end{cases}$$

provided that the series in the right-hand side above converges.

The following are immediate consequences of Proposition 5.3.

Corollary 5.1 ([33])**.** *Let s be a complex constant, (F_n) be the Fibonacci sequence, and $m \geq 0$. Then*

$$\sum_{n=1}^{\infty} \int_{\mathbb{C}} |sz|^m e^{-F_{n+1}} |z|\, dA(z) = 2\pi |s|^m \Gamma(m+2)[\zeta_F(m+2) - 1],$$

where $\zeta_F(s) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{F_n^s}$ is the Fibonacci zeta-function, provided that the series converges.

Corollary 5.2 ([33])**.** *Let s be a complex constant and $m \geq 0$. Then*

$$\sum_{n=1}^{\infty} |d_n| \int_{\mathbb{C}} |bz|^m e^{-n|z|}\, dA(z) = 2\pi |s|^m \Gamma(m+2) \sum_{n=1}^{\infty} \frac{|d_n|}{n^{m+2}}.$$

In particular, if $d_n = 1$ for all $n \in \mathbb{N}$, then

$$\sum_{n=1}^{\infty} \int_{\mathbb{C}} |sz|^m e^{-n|z|}\, dA(z) = 2\pi |s|^m \Gamma(m+2) \zeta(m+2),$$

where $\zeta(\cdot)$ is the Riemann zeta-function, provided that the series converges.

In [33], we presented an interesting connection between Proposition 5.2, Corollary 5.2, and Generalized Hypergeometric functions. To do this, recall that the definition of Generalized Hypergeometric functions ${}_pF_q(\mathbf{c};\mathbf{d};z)$, where $p, q \in \mathbb{N}$ (see, e.g., [20]).

Definition 5.1. *Let $\mathbf{c} = (c_1, \cdots, c_p)$, $\mathbf{d} = (d_1, \cdots, d_q)$ be two real p- and q-tuples, $a^{(k)} = a(a+1)\cdots(a+k-1)$ be a real rising factorial. The Generalized Hypergeometric function is*

$$ {}_pF_q(\mathbf{c};\mathbf{d};z) = \sum_{k=0}^{\infty} \frac{c_1^{(k)} c_2^{(k)} \cdots c_p^{(k)}}{d_1^{(k)} d_2^{(k)} \cdots d_q^{(k)}} \frac{z^k}{k!},\ z \in \mathbb{C}. \tag{5.6}$$

Proposition 5.4 ([33]). *Let s be a complex constant and $m \geq 0$. For the function $e^{e^{-r}} - 1$, by Proposition 5.2 and Corollary 5.2, we have*

$$\int_{\mathbb{C}} |sz|^m (e^{e^{-|z|}} - 1)\, dA(z) = \int_0^{2\pi} \int_0^{\infty} |s|^m r^{m+1} \sum_{n=1}^{\infty} \frac{1}{n!} e^{-nr}\, dr\, d\theta$$

$$= 2\pi |s|^m \Gamma(m+2) \sum_{n=1}^{\infty} \frac{1}{n! n^{m+2}} = \sum_{n=1}^{\infty} \int_{\mathbb{C}} |sz|^m e^{-n|z| + \log|\frac{1}{n!}|}\, dA(z).$$

In particular, if m is an integer, we have

$$\int_{\mathbb{C}} |sz|^m (e^{e^{-|z|}} - 1)\, dA(z) = 2\pi |s|^m \Gamma(m+2) \sum_{k=0}^{\infty} \frac{1}{(k+1)!} \left(\frac{1}{k+1} \right)^{m+2}$$

$$= 2\pi |s|^m \Gamma(m+2) \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{1^{(k)}}{2^{(k)}} \right)^{m+3} = 2\pi |s|^m \Gamma(m+2)_{m+3} F_{m+3}(\mathbf{1}; \mathbf{2}; 1).$$

5.2. Korenblum Constants for Intersections of Weighted Fock Spaces

Using Proposition 5.3 and its corollaries, we defined the following weighted Fock space in [33].

Definition 5.2. *Let $0 < p < \infty$ and $\alpha > 0$. For a positive real number λ and a non-zero complex number d, we define the weighted Fock space*

$$\mathcal{F}^{p,\alpha}_{\lambda,d}(\mathbb{C}) := \left\{ f(z) \in \mathcal{O}(\mathbb{C}) : \|f\|^p_{\mathcal{F}^{p,\alpha}_{\lambda,d}} = \int_{\mathbb{C}} |f(z)|^p e^{-\frac{p\alpha}{2}\lambda|z| + \log|d|}\, dA(z) < \infty \right\}.$$

Naturally, we proceeded to obtain an upper bound of Korenblum constants for the above weighted Fock space $\mathcal{F}^{p,\alpha}_{\lambda,d}(\mathbb{C})$.

Theorem 5.5 ([33]). *Let*

$$c > \frac{2}{p\alpha\lambda} \sqrt[p]{\Gamma(p+2)}. \tag{5.7}$$

Then there exist functions f and g in $\mathcal{F}^{p,\alpha}_{\lambda,d}(\mathbb{C})$, such that $|f(z)| < |g(z)|$ for any $|z| > c$, but $\|f\|^p_{\mathcal{F}^{p,\alpha}_{\lambda,d}} > \|g\|^p_{\mathcal{F}^{p,\alpha}_{\lambda,d}}$. Therefore,

$$\kappa_{\mathcal{F}^{p,\alpha}_{\lambda,d}} \leq \frac{2}{p\alpha\lambda} \sqrt[p]{\Gamma(p+2)}.$$

With particular values of parameters, Theorem 5.5 gives interesting estimates. For example, if p is a positive integer, then we have

$$\kappa_{\mathcal{F}^{p,\alpha}_{\lambda,d}} \leq \frac{2}{p\alpha\lambda} \sqrt[p]{(p+1)!}.$$

In the case when $\lambda = \pi^{-\frac{2}{3}}$, $p = \frac{3}{2}$ and $\alpha = \frac{\sqrt[3]{225}}{3}$, we also have

$$\kappa_{\mathcal{F}^{p,\alpha}_{\lambda,d}} \leq \pi.$$

Finally, we built a sequence for these weighted Fock spaces and consider intersections of these spaces, i.e., let $0 < (\lambda_n) \uparrow \infty$ and (d_n) be a sequence of non-zero complex numbers. Each pair (λ_n, d_n) defines a weighted Fock space $\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}(\mathbb{C})$, for which, by Theorem 5.5, the following is an upper bound estimate for its Korenblum's constant,

$$\kappa_{\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}} \leq \frac{2}{p\alpha\lambda_n} \sqrt[p]{\Gamma(p+2)}, \ n \in \mathbb{N}. \tag{5.8}$$

From here, we first considered the finite intersection of the above weighted Fock spaces. We presented the case for an intersection of two spaces which can be written as

$$\mathcal{F}^{p,\alpha}_{i,j} = \mathcal{F}^{p,\alpha}_{\lambda_i,d_i} \bigcap \mathcal{F}^{p,\alpha}_{\lambda_j,d_j} \quad (i < j),$$

endowed with the topology given by the norm

$$\|f\|_{\mathcal{F}^{p,\alpha}_{i,j}} := \max\left\{\|f\|_{\mathcal{F}^{p,\alpha}_{\lambda_i,d_i}}, \|f\|_{\mathcal{F}^{p,\alpha}_{\lambda_j,d_j}}\right\}. \tag{5.9}$$

Note that the space $\mathcal{F}^{p,\alpha}_{i,j}$ is a Banach space with the norm above.

With standard arguments, we obtained the following upper bounds for Korenblum constant of $\mathcal{F}^{p,\alpha}_{i,j}$.

Theorem 5.6 ([33]). *Let*

$$c > \begin{cases} \frac{2}{p\alpha\lambda_i} \sqrt[p]{\Gamma(p+2)}, & |d_i| \geq |d_j| \\ \frac{2}{p\alpha\lambda_i} \sqrt[p]{\Gamma(p+2)}, & |d_i| < |d_j| \quad \text{and} \quad \frac{|d_i|}{\lambda_i^{p+2}} \geq \frac{|d_j|}{\lambda_j^{p+2}} \\ \frac{2}{p\alpha\lambda_j} \sqrt[p]{\Gamma(p+2)}, & |d_i| < |d_j| \quad \text{and} \quad \frac{|d_i|}{\lambda_i^{p+2}} < \frac{|d_j|}{\lambda_j^{p+2}}. \end{cases} \tag{5.10}$$

Then there exist functions f and g in $\mathcal{F}^{p,\alpha}_{i,j}$, such that $|f(z)| < |g(z)|$ for any $|z| > c$, but

$$\|f\|_{\mathcal{F}^{p,\alpha}_{i,j}} > \|g\|_{\mathcal{F}^{p,\alpha}_{i,j}}.$$

Therefore,

$$\kappa_{\mathcal{F}^{p,\alpha}_{i,j}} \leq \begin{cases} \frac{2}{p\alpha\lambda_i} \sqrt[p]{\Gamma(p+2)}, & |d_i| \geq |d_j| \\ \frac{2}{p\alpha\lambda_i} \sqrt[p]{\Gamma(p+2)}, & |d_i| < |d_j| \quad \text{and} \quad \frac{|d_i|}{\lambda_i^{p+2}} \geq \frac{|d_j|}{\lambda_j^{p+2}} \\ \frac{2}{p\alpha\lambda_j} \sqrt[p]{\Gamma(p+2)}, & |d_i| < |d_j| \quad \text{and} \quad \frac{|d_i|}{\lambda_i^{p+2}} < \frac{|d_j|}{\lambda_j^{p+2}}. \end{cases}$$

Next, we considered the Korenblum constants for an infinite intersection of spaces $\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}$ $(n \in \mathbb{N})$,

$$\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}} = \left\{ f \in \mathcal{O}(\mathbb{C}) : \|f\|^p_{\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}} < \infty, \text{ for every } n \in \mathbb{N} \right\},$$

endowed with the topology given by the series of norms

$$\|f\|^p_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}} = \sum_{n=1}^{\infty} \|f\|^p_{\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}} < \infty.$$

For the above intersection, we have to ensure that it is non-empty as otherwise, it would be trivial to consider its Korenblum constants. Fortunately, if we want at least the simple constant function $f(z) = c \in \mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}$, we can have its norm

$$\sum_{n=1}^{\infty} \|f\|^p_{\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}} = 2\pi c^p \left(\frac{2}{p\alpha}\right)^2 \sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} < \infty,$$

by assuming that

$$\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} < \infty. \tag{5.11}$$

As mentioned previously, the pair of sequences (λ_n) and (d_n) contributes to a large class of special functions in Dirichlet series. Hence, there exist many pairs (λ_n, d_n) for which condition (5.11) may or may not hold. For example, if $\lambda_n = n$, we take $d_n = n^\rho$, then condition (5.11) is satisfied, if $\rho < 1$, and it is not satisfied, if $\rho \geq 1$.

Note also that since $\exists n_0$ such that $\lambda_{n_0} \geq 1$, then $\lambda_{n_0}^{kp} \geq 1$ for any $0 < p < \infty$ and $k \in \mathbb{N}$. Hence, $\lambda_{n_0}^{kp+2} \geq \lambda_{n_0}^2$. Then for all $n \geq n_0$, $\lambda_n \geq 1$ implies $\lambda_n^{kp+2} \geq \lambda_n^2$. As a result, condition (5.11) implies

$$\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^{kp+2}} < \infty. \tag{5.12}$$

This also allows all polynomials $z^k \in \mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}$.

In the rest of this section, we assume that condition (5.11) holds.

Then the space $\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}$ always contain a constant function $f(z) = c$ and all polynomial in z.

By using Proposition 5.3, we obtained an interesting upper bound for $\kappa_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}}$ in terms of Dirichlet series and Gamma functions.

Theorem 5.7 ([33]). *Let $0 < (\lambda_n) \uparrow \infty$ and (d_n) be a sequence of non-zero complex numbers satisfying condition* (5.11). *Suppose*

$$c > \frac{2}{p\alpha} \sqrt[p]{\Gamma(p+2) \sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^{p+2}} \left(\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} \right)^{-1}}, \tag{5.13}$$

then there exist functions f and g in $\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}$, such that $|f(z)| < |g(z)|$ for any $|z| > c$, but

$$\|f\|_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}} > \|g\|_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}}.$$

Therefore,

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}} \leq \frac{2}{p\alpha} \sqrt[p]{\Gamma(p+2) \sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^{p+2}} \left(\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} \right)^{-1}}.$$

Following the particular cases of Dirichlet series, we also obtained the following special cases of Theorem 5.7 in [33].

Corollary 5.3 ([33]). *Let $d_n = 1$ and (F_n) be the Fibonacci sequence. If the Korenblum Maximum Principle holds for the space $\mathcal{F}^{p,\alpha}_{\{F_{n+1},1\}}$, then*

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{F_{n+1},1\}}} \leq \frac{2}{p\alpha} \sqrt[p]{\Gamma(p+2) \frac{\zeta_F(p+2) - 1}{\zeta_F(2) - 1}},$$

where $\zeta_F(s) = \sum_{n=1}^{\infty} \frac{1}{F_n^s}$ is the Fibonacci zeta function.

Corollary 5.4 ([33]). *Let $\lambda_n = n$ for all $n \in \mathbb{N}$. If the Korenblum Maximum Principle holds for the space $\mathcal{F}^{p,\alpha}_{\{n,d_n\}}$, then*

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{n,d_n\}}} \leq \frac{2}{p\alpha} \sqrt[p]{\Gamma(p+2) \sum_{n=1}^{\infty} \frac{|d_n|}{n^{p+2}} \left(\sum_{n=1}^{\infty} \frac{|d_n|}{n^2} \right)^{-1}}, \tag{5.14}$$

provided that a series $\sum_{n=1}^{\infty} \frac{|d_n|}{n^2}$ converges.

For Corollary 5.4, if $d_n = 1$ for all $n \in \mathbb{N}$, we get a corollary which involves the Riemann zeta function $\zeta(z)$.

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{n,1\}}} \leq \frac{2}{p\alpha} \sqrt[p]{\frac{6}{\pi^2} \Gamma(p+2) \zeta(p+2)}. \tag{5.15}$$

On the other hand, if $d_n = \frac{1}{n!}$ for all $n \in \mathbb{N}$, we use Proposition 5.4 to get this corollary which involves the Generalized Hypergeometric functions.

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{n,\frac{1}{n!}\}}} \leq \frac{2}{p\alpha}\sqrt[p]{\frac{\Gamma(p+2)}{{}_3F_3(\mathbf{1};\mathbf{2};1)}\sum_{n=1}^{\infty}\frac{1}{n!n^{p+2}}}.$$

In addition, when p is an integer, we have

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{n,\frac{1}{n!}\}}} \leq \frac{2}{p\alpha}\sqrt[p]{\frac{\Gamma(p+2)_{p+3}F_{p+3}(\mathbf{1};\mathbf{2};1)}{{}_3F_3(\mathbf{1};\mathbf{2};1)}}.$$

Combining definitions 4.1 and 5.2 of two types of weighted Fock spaces together with Theorem 5.7 led us to generalise them to the following slightly more complicated weighted Fock spaces in [33].

Definition 5.3. *Let $0 < p < \infty$, $\alpha > 0$ and $m > 0$. For each $n \in \mathbb{N}$, the weighted Fock spaces $\mathcal{F}^{p,\alpha}_{m,\lambda_n,d_n}(\mathbb{C})$ consists of entire functions $f(z) \in \mathcal{O}(\mathbb{C})$ for which*

$$\|f\|^p_{\mathcal{F}^{p,\alpha}_{m,\lambda_n,d_n}} = \int_{\mathbb{C}} |f(z)|^p e^{-\frac{p\alpha}{2}\lambda_n|z|^m + \log|d_n|}\, dA(z) < \infty.$$

Similarly, we extended our results to the infinite intersections of weighted Fock spaces $\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}} = \bigcap_{n=1}^{\infty}\mathcal{F}^{p,\alpha}_{m,\lambda_n,d_n}(\mathbb{C})$. $\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}$ consists of entire functions $f(z) \in \mathcal{O}(\mathbb{C})$ such that $\|f\|^p_{\mathcal{F}^{p,\alpha}_{m,\lambda_n,d_n}} < \infty$ for all $n \in \mathbb{N}$, endowed with the topology given by the series of norms

$$\|f\|^p_{\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}} := \sum_{n=1}^{\infty}\|f\|^p_{\mathcal{F}^{p,\alpha}_{m,\lambda_n,d_n}} < \infty.$$

Following condition (5.11), we shall assume that the triple $\{m, d_n, \lambda_n\}$ must fulfill the condition

$$\sum_{n=1}^{\infty}\frac{|d_n|}{\lambda_n^{2/m}} < \infty, \tag{5.16}$$

which would imply that for any $k \in \mathbb{N}$ and $0 < p < \infty$,

$$\sum_{n=1}^{\infty}\frac{|d_n|}{\lambda_n^{\frac{kp+2}{m}}} < \infty.$$

This would mean that the space $\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}$ is also non-empty, as it contains the constant function c and all polynomial in z.

Our final upper bound in [33] is for the case $\lambda_n = n^m$ and $d_n = 1$ for all $n \in \mathbb{N}$. Clearly, for any $m > 0$, $0 < p < \infty$ and $\alpha > 0$, condition (5.16) must remain satisfied here. With a combination of proofs from both Theorem 4.6 and Theorem 5.7, the upper bound for the Korenblum constant of the space $\mathcal{F}^{p,\alpha}_{\{m,n^m,1\}}$ is

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{m,n^m,1\}}} \leq \left(\frac{2}{p\alpha}\right)^{\frac{1}{m}} \sqrt[p]{\frac{6}{\pi^2}\zeta(p+2)\Gamma\left(\frac{p+2}{m}\right)\Gamma\left(\frac{2}{m}\right)^{-1}}.$$

We summarize the results surveyed in this section. In this section, the following main results are surveyed.

(i) For $\mathcal{F}^{p,\alpha}_m(\mathbb{C})$,

$$\kappa_{\mathcal{F}^{p,\alpha}_m} \leq \sqrt[p]{\left(\frac{2}{p\alpha}\right)^{\frac{p}{m}}\Gamma\left(\frac{p+2}{m}\right)\left(\Gamma\left(\frac{2}{m}\right)\right)^{-1}}.$$

(ii) For $\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}(\mathbb{C})$,

$$\kappa_{\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}} \leq \frac{2}{p\alpha\lambda_n}\sqrt[p]{\Gamma(p+2)}.$$

(iii) For $\mathcal{F}^{p,\alpha}_{i,j}(\mathbb{C})$,

$$\kappa_{\mathcal{F}^{p,\alpha}_{i,j}} \leq \begin{cases} \frac{2}{p\alpha\lambda_i}\sqrt[p]{\Gamma(p+2)}, & |d_i| \geq |d_j| \\ \frac{2}{p\alpha\lambda_i}\sqrt[p]{\Gamma(p+2)}, & |d_i| < |d_j| \quad \text{and} \quad \frac{|d_i|}{\lambda_i^{p+2}} \geq \frac{|d_j|}{\lambda_j^{p+2}} \\ \frac{2}{p\alpha\lambda_j}\sqrt[p]{\Gamma(p+2)}, & |d_i| < |d_j| \quad \text{and} \quad \frac{|d_i|}{\lambda_i^{p+2}} < \frac{|d_j|}{\lambda_j^{p+2}}. \end{cases}$$

(iv) For $\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}$,

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}} \leq \frac{2}{p\alpha}\sqrt[p]{\Gamma(p+2)\sum_{n=1}^{\infty}\frac{|d_n|}{\lambda_n^{p+2}}\left(\sum_{n=1}^{\infty}\frac{|d_n|}{\lambda_n^2}\right)^{-1}}.$$

(v) Lastly, for $\mathcal{F}^{p,\alpha}_{\{m,n^m,1\}}$, $m > 0$ and $\alpha > 0$,

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{m,n^m,1\}}} \leq \left(\frac{2}{p\alpha}\right)^{\frac{1}{m}} \sqrt[p]{\frac{6}{\pi^2}\zeta(p+2)\Gamma\left(\frac{p+2}{m}\right)\Gamma\left(\frac{2}{m}\right)^{-1}}.$$

6. Failure of Korenblum Maximum Principle

In this section, we survey the remaining results from all function spaces having a failure of Korenblum Maximum Principle. The failure of Korenblum Maximum Principle was first discovered for the Bloch space B by Jiang, Prajitura and Zhao in [14]. In fact, the Korenblum constant does not exist even if E is the whole unit disc $\mathbb{D}$ in the Bloch space B. In other words, the whole unit disc $\mathbb{D}$ cannot even be a dominating set for the Bloch space B.

Theorem 6.1 ([14]). *Let $g(z) = z^3 + z$ and $f(z) = zg(z) = z^4 + z^2$. Then $|f(z)| \leq |g(z)|$ for all $z \in \mathbb{D}$, but $\|f\|_B > \|g\|_B$.*

In the recent years, there is a regained interest in Korenblum constants which should be largely attributed to the failure of Korenblum Maximum Principle for $A^p(\mathbb{D})$, $0 < p < 1$ reported in 2018. In 2018, Vladimir Božin and Karapetrović [3] discovered a complete failure in Korenblum Maximum Principle for Bergman space $A^p(\mathbb{D})$, $0 < p < 1$. In other words, no Korenblum constant exist to satisfy the Korenblum Maximum Principle for Bergman spaces as long as $0 < p < 1$. This result closes a research gap from Hinkkanen where he first proved the existence of κ_{A^p} for $A^p(\mathbb{D})$ but for $p \geq 1$. Note that Proposition 6.3 is a consequence of Proposition 6.2.

Proposition 6.2 ([3]). *Let $0 < p < 1$ and $0 < c < 1$. Then, there exist $n \in \mathbb{N}$ and $0 < \varepsilon < 1$ such that*

$$1 + \frac{np}{2}\varepsilon^{np+2} > \left(1 + \left(\frac{\varepsilon}{c}\right)^n\right)^p.$$

Theorem 6.3 ([3]). *Let $0 < p < 1$ and $0 < c < 1$. Then there exist functions f and g in $A^p(\mathbb{D})$ such that $|f(z)| < |g(z)|$ for all $c < |z| < 1$ and*

$$||f||_{A^p} > ||g||_{A^p}.$$

Shortly after, Lou and Hu [13] disproved the principle for classical weighted Fock spaces $\mathcal{F}^p_\alpha(\mathbb{C})$ where $0 < p < 1$, $\alpha > 0$ using similar methods. In particular, the following proposition is used to disprove the principle instead.

Proposition 6.4 ([13]). *Let $0 < p < 1$ and $\alpha > 0$. Suppose $c > 0$. Then there* *exist positive integer n and $0 < \rho < \infty$, such that*

$$2\rho^{np+2}\left(\int_0^1 ue^{-\frac{p\alpha}{2}\rho^2u^2}\,du + \int_1^\infty u^{np+1}e^{-\frac{p\alpha}{2}\rho^2u^2}\,du\right)$$
$$> \left(1+\left(\frac{\rho}{c}\right)^n\right)^p\left(\frac{2}{p\alpha}\right)^{\frac{np}{2}+1}\Gamma\left(\frac{np}{2}+1\right).$$

Lou and Hu [13] then obtained the following result.

Proposition 6.5 ([13]). *Let $0 < p < 1$ and $\alpha > 0$. Suppose $c > 0$. Then there exist functions f and g in $\mathcal{F}^p_\alpha(\mathbb{C})$ such that $|f(z)| < |g(z)|$ for all $|z| > c$ and*

$$||f||_{\mathcal{F}^p_\alpha} > ||g||_{\mathcal{F}^p_\alpha}.$$

The results about the failures of Korenblum Maximum Principle in $A^p(\mathbb{D})$, $0 < p < 1$ [3] and in $\mathcal{F}^p_\alpha(\mathbb{C})$, $0 < p < 1$, $\alpha > 0$ [13] inspired us to investigate whether there are any failures of the Korenblum Maximum Principle for the weighted Bergman space with exponential weights $A^p_\gamma(\mathbb{D})$, $0 < p < 1$, $\gamma \neq 0$. Similarly, we showed that such a failure exist and the result below not only proves this fact, but also generalizes Theorem 6.3 for any $\gamma > 0$ [32].

Proposition 6.6 ([32]). *Let $0 < p < 1$, $\gamma > 0$ and $0 < c < 1$. Then there exist positive integer n and $0 < \delta < 1$, such that*

$$2\delta^{np+2}\left(\int_0^1 ue^{-\frac{p\gamma}{2}\delta^2u^2}\,du + \int_1^{\frac{1}{\delta}} u^{np+1}e^{-\frac{p\gamma}{2}\delta^2u^2}\,du\right)$$
$$> \left(1+\left(\frac{\delta}{c}\right)^n\right)^p\left(\frac{2}{p\gamma}\right)^{\frac{np}{2}+1}\int_0^{\frac{p\gamma}{2}} u^{\frac{np}{2}}e^{-u}\,du.$$

Theorem 6.7 ([32]). *Let $0 < p < 1$ and $\gamma > 0$. Suppose $0 < c < 1$. Then there exist functions f and g in $A^p_\gamma(\mathbb{D})$ such that $|f(z)| < |g(z)|$ for any z with $c < |z| < 1$ and $\|f\|_{A^p_\gamma} > \|g\|_{A^p_\gamma}$.*

Recently, we turned our attention back to the weighted Fock spaces and answer whether the Korenblum constant exist even for small intersections of weighted Fock spaces. However, we found that the principle fails for the infinite intersection $\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}(\mathbb{C})$ when $0 < p < 1$ and $m \geq 2$ [33]. This also left a new open problem as to whether the principle fails for $0 < m < 2$.

The following proposition is a generalisation of Propostion 6.4.

Proposition 6.8 ([33]). *Let $0 < p < 1$, $\alpha > 0$, $m \geq 2$, $0 < (\lambda_n) \uparrow \infty$ and $c > 0$. Then there exist a positive integer k and $0 < \delta < 1$ such that for all*

$n \in \mathbb{N}$,

$$m\delta^{kp+2}\left(\int_0^1 ue^{-\frac{p\alpha}{2}\lambda_n\delta^m u^m}\,du + \int_1^\infty u^{kp+1}e^{-\frac{p\alpha}{2}\lambda_n\delta^m u^m}\,du\right) > \left(1+\left(\frac{\delta}{c}\right)^k\right)^p\left(\frac{2}{p\alpha\lambda_n}\right)^{(\frac{kp+2}{m})}\Gamma\left(\frac{kp+2}{m}\right).$$

Thus, we have the following result.

Theorem 6.9 ([33])**.** *Let $0 < p < 1$, $\alpha > 0$, $0 < (\lambda_n) \uparrow \infty$, $m \geq 2$ and (d_n) be a sequence of non-zero complex numbers. Suppose $c > 0$. Then there exist functions f and g in $\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}(\mathbb{C})$ such that $|f(z)| < |g(z)|$ for any z with $|z| > c$ and $\|f\|_{\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}} > \|g\|_{\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}}$.*

Remark 6.1. *In [33], we make note that Lemma 6.8 can be slightly modified or relaxed accordingly for finite intersection or individual spaces such as $\mathcal{F}^{p,\alpha}_{m,\lambda_i,d_i}(\mathbb{C}) \bigcap \mathcal{F}^{p,\alpha}_{m,\lambda_j,d_j}(\mathbb{C})$ and $\mathcal{F}^{p,\alpha}_{m,\lambda_n,d_n}(\mathbb{C})$ respectively. For these spaces, the proof for the failure of Korenblum constants when $0 < p < 1$ and $m \geq 2$ can be similarly proven just like Theorem 6.9. We refer the reader to [4] for further details.*

Instead of working towards smaller spaces, a recent failure of Korenblum maximum principle was also extended to the mixed norm space $H^{p,q,s}$ where $0 < p, q, s < \infty$ by Karapetrović [4]. Recall that the mixed norm space $H^{p,q,s}$ ($0 < p, q, s < \infty$) consists of all holomorphic functions in $\mathcal{O}(\mathbb{D})$ for which

$$\|f\|_{H^{p,q,s}} = \left(2sq\int_0^1 r(1-r^2)^{sq-1}M_p^q(r,f)\,dr\right)^{1/q} < \infty, \tag{6.1}$$

where

$$M_p(r,f) = \left(\frac{1}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|^p\,d\theta\right)^{1/p}.$$

Interestingly, this latest failure of Korenblum Maximum Principle was extended for the space $H^{p,q,s}$ where $0 < p, s < \infty$ and $0 < q < 1$.

Theorem 6.10 ([4])**.** *Let $0 < p, s < \infty$, $0 < q < 1$ and $0 < c < 1$. Then there exist functions f and g holomorphic on $\mathbb{D}$ such that $|f(z)| < |g(z)|$ for all $c < |z| < 1$, but $\|f\|_{H^{p,q,s}} > \|g\|_{H^{p,q,s}}$.*

7. Future Directions and Open Questions

7.1. Future Directions

In [31], a proposed future direction is to study the Korenblum constants for general function spaces with series norms. Recall that G is a domain and we shall write $\mathcal{O}(G)$ as the set of holomorphic functions defined on G. Then we define $\beta = (\beta_k)$ as a sequence of positive real numbers. This allows us to define the function space

$$\mathcal{H}(G,\beta) := \left\{ f(z) = \sum_{k=0}^{\infty} a_k z^k \in \mathcal{O}(G) : ||f||^2_{O(G)} = \sum_{k=0}^{\infty} |a_k|^2 \beta_k^2 < \infty \right\}.$$

When $G = \mathbb{C}$, we have the Hilbert space of entire functions. Further, if $\beta = (\sqrt{k!})_{k\in\mathbb{N}}$, we have the classical Fock space $\mathcal{F}^2(\mathbb{C})$. If $G = \mathbb{D}$, we have $H^2(\beta)$. In the case $\beta = (1)_{k\in\mathbb{N}}$, it becomes the classical Hardy space $H^2(\mathbb{D})$; if $\beta = (\frac{1}{\sqrt{k+1}})_{k\in\mathbb{N}}$, it becomes the classical Bergman space $A^2(\mathbb{D})$; if $\beta = (\sqrt{k+1})_{k\in\mathbb{N}}$, we then have the Dirichlet space $\mathcal{D}(\mathbb{D})$. If we consolidate the development of Korenblum constants under $\mathcal{H}(G,\beta)$, we obtain Table 5.

	G	-	$\beta = (\beta_k)$	**Function Space**	**Current Development**
$\mathcal{H}(G,\beta)$	$\mathbb{C}$	Hilbert Space of Entire Functions	$\sqrt{k!}$	$\mathcal{F}^2(\mathbb{C})$	$0.7248 \le \kappa_{\mathcal{F}^2} \le 1$
	$\mathbb{D}$	$H^2(\beta)$	1	$H^2(\mathbb{D})$	(see Table 4)
			$\frac{1}{\sqrt{k+1}}$	$A^2(\mathbb{D})$	$0.28185 \le \kappa_{A^2} < 0.67795$
			$\sqrt{k+1}$	$\mathcal{D}(\mathbb{D})$	No Specific Development but $\mathcal{D}(\mathbb{D}) \subset H^2(\mathbb{D})$

Table 5. Current Development on $\mathcal{H}(G,\beta)$

7.2. Open Questions

1. For the weighted Bergman spaces, the following open questions call for investigation. The open questions were previously mentioned in [32].

   Question 7.1 ([32])**.** *Let $p \ge 1$, let $\gamma \ge 0$ and let*

$$c = \begin{cases} \left(\frac{2}{p+2}\right)^{\frac{1}{p}}, & \gamma = 0, \\ \sqrt[p]{\dfrac{\left(\frac{2}{p\gamma}\right)^{\frac{p}{2}} \displaystyle\int_0^{\frac{p\gamma}{2}} u^{\frac{p}{2}} e^{-u} du}{(1-e^{-\frac{p\gamma}{2}})}}, & \gamma > 0. \end{cases}$$

Does there exist functions $f(z)$ and $g(z)$ in $A^p_\gamma(\mathbb{D})$ for which $|f(z)| < |g(z)|$ with $c < |z| < 1$ and $\|f\|_{A^p_\gamma} > \|g\|_{A^p_\gamma}$?

Question 7.2 ([32])**.** *Let $-1 < \gamma < 0$ and $1 \leq p < \infty$. Does there exist functions $f(z)$ and $g(z)$ in $A^p_\gamma(\mathbb{D})$ for which $|f(z)| < |g(z)|$ with $c < |z| < 1$ and $\|f\|_{A^p_\gamma} > \|g\|_{A^p_\gamma}$?*

2. For the weighted Hardy spaces, the following question calls for investigation.

Question 7.3. *Can Theorem 4.1 be generalised for $H^p_\alpha(\mathbb{D})$, $\alpha > 0$, $0 < p \leq \infty$? In addition, how does Theorem 4.1 change with respect to φ?*

3. For the intersection of weighted Fock spaces, we have the following open questions from [33].

Question 7.4. *It would be interesting to know whether the upper bound of $\kappa_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}}$ in Theorem 5.7 is always less than $\kappa_{\mathcal{F}^{p,\alpha}_{\lambda_1,d_1}}$. This is because the upper bound of $\kappa_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}}$ is always less than the upper bound of $\kappa_{\mathcal{F}^{p,\alpha}_{\lambda_1,d_1}}$. To see this, for $0 < \lambda_1 < \lambda_n$ for all $n \geq 2$, $\lambda_n^{p+2} > \lambda_1^p \lambda_n^2$ which shows that*

$$\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^{p+2}} \left(\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} \right)^{-1} < \frac{1}{\lambda_1^p} \sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} \left(\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} \right)^{-1} = \frac{1}{\lambda_1^p} < \infty.$$

Hence, we have

$$\kappa_{\mathcal{F}^{p,\alpha}_{\{\lambda_n,d_n\}}} \leq \frac{2}{p\alpha} \sqrt[p]{\Gamma(p+2) \sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^{p+2}} \left(\sum_{n=1}^{\infty} \frac{|d_n|}{\lambda_n^2} \right)^{-1}} < \frac{2}{p\alpha\lambda_1} \sqrt[p]{\Gamma(p+2)}.$$

Question 7.5. *Does there exist any relationship between $\kappa_{\mathcal{F}^{p,\alpha}_{\{m,n^m,1\}}}$ and $\kappa_{\mathcal{F}^{p,\alpha}_m}$?*

Question 7.6. *What is an upper bound for finite intersection of more than two spaces $\mathcal{F}^{p,\alpha}_{\lambda_n,d_n}(\mathbb{C})$, etc?*

Question 7.7. *Is it true that the principle still fails for $\mathcal{F}^{p,\alpha}_{\{m,\lambda_n,d_n\}}(\mathbb{C})$ when $0 < p < 1$ and $0 < m < 2$, in particular, $m = 1$?*

4. Following Theorem 6.1, we propose the following open question for the weighted Bloch spaces B_φ.

Question 7.8. *Does there exist functions $f(z)$ and $g(z)$ in B_φ such that $|f(z)| \leq |g(z)|$ for all $z \in \mathbb{D}$, but $\|f\|_{B_\varphi} > \|g\|_{B_\varphi}$?*

5. For the function space $\mathcal{H}(G,\beta)$, we propose the following open question.

   Question 7.9. *Is it possible to generalise the results for Korenblum constants under the spaces $H^2(\mathbb{D})$, $A^2(\mathbb{D})$ and $\mathcal{D}(\mathbb{D})$ under $H^2(\beta)$?*

   In general, the main challenge along this direction would be that the space $\mathcal{H}(G,\beta)$ involves norms in series notations while most of the function spaces discussed in this survey deals with integral norms.

References

[1] **Amdeberhan T., Espinosa O., Gonzalez I., Harrison M., Moll V. H. and Straub A.** *Ramanujan's Master Theorem,* Ramanujan J., **29** (2012), no. 1-3, 103–120.

[2] **Berndt Bruce C.** *The quarterly reports of S. Ramanujan,* Amer. Math. Monthly, **90** (1983), no. 8, 505–516.

[3] **Božin V. and Karapetrović B.,** *Failure of Korenblum's maximum principle in Bergman spaces with small exponents,* Proc. Amer. Math. Soc., **146** (2018), no. 6, 2577–2584.

[4] **Karapetrović B.,** *Korenblum maximum principle in mixed norm spaces,* Arch. Math. (Basel), **118** (2022), no. 5, 497–507.

[5] **Butzer Paul L. and Jansche S.,** *A direct approach to the Mellin transform,* J. Fourier Anal. Appl., **3** (1997), no. 4, 325–376.

[6] **Cahen E.,** *Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues,* Ann. Sci. École Norm. Sup. (3), **11** (1894), 75–164 (French).

[7] **Danikas N. and Hayman W. K.,** *Domination on sets and in H^p,* Results Math., **34** (1998), no. 1-2, 85–90, Dedicated to Paul Leo Butzer.

[8] **Flajolet P., Gourdon X. and Dumas P.,** *Mellin transforms and asymptotics: harmonic sums. Special volume on mathematical analysis of algorithms,* Theoret. Comput. Sci. , **144** (1995), no. 1-2, 3–58.

[9] **Hardy G. H.** *Ramanujan. Twelve lectures on subjects suggested by his life and work,* Cambridge University Press, Cambridge, England; Macmillan Company, New York, (1940).

[10] **Hardy G. H. and Riesz M.,** *The general theory of Dirichlet's series,* Cambridge Tracts in Mathematics and Mathematical Physics, No. 18, Stechert-Hafner, Inc., New York, (1964).

[11] **Hayman W. K.,** *On a conjecture of Korenblum,* Analysis (Munich), **19** (1999), no. 2, 195–205.

[12] **Hinkkanen A.,** *On a maximum principle in Bergman space,* J. Anal. Math., **79** (1999), 335–344.

[13] **Jianhui H. and Zengjian L.,** *The Korenblum's maximum principle in Fock spaces with small exponents,* J. Math. Anal. Appl., **470** (2019), no. 2, 770–776.

[14] **Liangying J., Gabriel T. P. and Ruhan Z.,** *On Korenblum's maximum principle for some function spaces,* The first NEAM, Theta Ser. Adv. Math., vol. 22, Editura Fundaţiei Theta, Bucharest, (2018), 59–80.

[15] **Korenblum B.,** *A maximum principle for the Bergman space,* Publ. Mat., **35** (1991), no. 2, 479–486.

[16] **Korenblum B. and Richards K.,** *Majorization and domination in the Bergman space,* Proc. Amer. Math. Soc., **117** (1993), no. 1, 153–158.

[17] **Korenblum B., O'Neil R., Richards K. and Zhu K.,** *Totally monotone functions with applications to the Bergman space,* Trans. Amer. Math. Soc., **337** (1993), no. 2, 795–806.

[18] **Matero J.,** *On Korenblum's maximum principle for the Bergman space,* Arch. Math. (Basel), **64** (1995), no. 4, 337–340.

[19] **Perron O.,** *Zur Theorie der Dirichletschen Reihen,* J. Reine Angew. Math., **134** (1908), 95–143 (German).

[20] **Rainville E. D.,** *Special functions,* Chelsea Publishing Co., Bronx, N.Y., (1971).

[21] **Schuster A.,** *The maximum principle for the Bergman space and the Möbius pseudodistance for the annulus,* Proc. Amer. Math. Soc., **134** (2006), no. 12, 3525–3530.

[22] **Schwick W.,** *On Korenblum's maximum principle,* Proc. Amer. Math. Soc., **125** (1997), no. 9, 2581–2587.

[23] **Shen C.,** *A slight improvement to Korenblum's constant,* J. Math. Anal. Appl., **337** (2008), no. 1, 464–465.

[24] **Chunjie W.,** *Refining the constant in a maximum principle for the Bergman space,* Proc. Amer. Math. Soc., **132** (2004), no. 3, 853–855.

[25] **Chunjie W.,** *An upper bound on Korenblum's constant,* Integral Equations Operator Theory, **49** (2004), no. 4, 561–563.

[26] **Chunjie W.,** *On Korenblum's constant,* J. Math. Anal. Appl., **296** (2004), no. 1, 262–264.

[27] **Chunjie W.,** *On Korenblum's maximum principle,* Proc. Amer. Math. Soc., **134** (2006), no. 7, 2061–2066.

[28] **Chunjie W.,** *Behavior of the constant in Korenblum's maximum principle,* Math. Nachr., **281** (2008), no. 3, 447–454.

[29] **Chunjie W.,** *Domination in the Bergman space and Korenblum's constant,* Integral Equations Operator Theory, **61** (2008), no. 3, 423–432.

[30] **Chunjie W.,** *Some results on Korenblum's maximum principle,* J. Math. Anal. Appl., **373** (2011), no. 2, 393–398.

[31] **JunJie W.,** *The Korenblum Maximum Principle for some Function Spaces,* Bachelor's Thesis, Nanyang Technological University, 2019. Available at URI `hdl.handle.net/10356/77142`.

[32] **JunJie W. and Khoi L. H.,** *Korenblum constants for some function spaces,* Proc. Amer. Math. Soc., **148** (2020), no. 3, 1175–1185.

[33] **JunJie W. and Khoi L. H.,** *Korenblum constants for various weighted Fock spaces,* Complex Variables and Elliptic Equations, posted on June 2022, 1-22, DOI 10.1080/17476933.2022.2052862.

[34] **Kehe Z.,** *Analysis on Fock spaces,* Graduate Texts in Mathematics, vol. 263, Springer, New York, (2012).

Wee JunJie, Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University (NTU)
637371 Singapore; `weej0016@e.ntu.edu.sg`

Le Hai Khoi, University of Science and Technology of Hanoi – USTH (Vietnam France University), Vietnam Academy of Science and Technology
Hanoi 10000, Vietnam; `le-hai.khoi@usth.edu.vn`

J. Math. Math. Sci. (2023) 117–128

A uniqueness theorem for meromorphic functions ignoring multiplicity

Nguyen Duy Phuong (Thai Nguyen, Vietnam)

(Received Mar. 13, 2023)

Abstract. In this paper, we give a uniqueness theorem for meromorphic functions ignoring multiplicity , which generalizes a An's theorem in [1].

1. Introduction. Main results

In this paper, by a meromorphic function we mean a meromorphic function on the complex plane $\mathbb{C}$.

Let f be a non-constant meromorphic function on $\mathbb{C}$. For every $a \in \mathbb{C}$, we define the function $\nu_f^a: \mathbb{C} \to \mathbb{N}$ by

$$\nu_f^a(z) = \begin{cases} 0 & \text{if } f(z) \neq a \\ d & \text{if } f(z) = a \text{ with multiplicity } d, \end{cases}$$

and set $\nu_f^\infty = \nu_{\frac{1}{f}}^0$, and define the function $\overline{\nu}_f^a: \mathbb{C} \to \mathbb{N}$ by $\overline{\nu}_f^a(z) = \min\ \{\nu_f^a(z), 1\}$, and set $\overline{\nu}_f^\infty = \overline{\nu}_{\frac{1}{f}}^0$. For $f \in \mathcal{M}(\mathbb{C})$ and a non-empty set $S \subset \mathbb{C} \cup \{\infty\}$, we define

$$E_f(S) = \bigcup_{a \in S} \{(z, \nu_f^a(z)) : z \in \mathbb{C}\}, \quad \overline{E}_f(S) = \bigcup_{a \in S} \{(z, \overline{\nu}_f^a(z)) : z \in \mathbb{C}\}.$$

Let $\mathcal{F}$ be a nonempty subset of $\mathcal{M}(\mathbb{C})$. Two functions f, g of $\mathcal{F}$ are said to *share* S, *counting multiplicity* (share S CM) if $E_f(S) = E_g(S)$, and to *share* S, *ignoring multiplicity* (share S IM) if $\overline{E}_f(S) = \overline{E}_g(S)$.

Key words and phrases: Meromorphic Function, uniqueness, ignoring multiplicity.
2020 Mathematics Subject Classification: 30D05

If the condition $E_f(S) = E_g(S)$ implies $f = g$ for any two non-constant meromorphic (entire) functions f, g, then S is called a unique range set for meromorphic (entire) functions counting multiplicity, or in brief, URSM (URSE). A set $S \subset \mathbb{C} \cup \{\infty\}$ is called a unique range set for meromorphic (entire) functions ignoring multiplicity, or in brief, URSM-IM (URSE-IM), if the condition $\overline{E}_f(S) = \overline{E}_g(S)$ implies $f = g$ for any pair of non-constant meromorphic (entire) functions.

In 1976 Gross ([10]) proved that there exist three finite sets S_j $(j = 1, 2, 3)$ such that any two entire functions f and g satisfying $E_f(S_j) = E_g(S_j)$, $j = 1, 2, 3$ must be identical. In the same paper Gross([10]) posed the following question:

Question A. *Can one find two (or possible even one) finite set S_j $(j = 1, 2)$ such that any two entire functions f and g satisfying $E_f(S_j) = E_g(S_j)$ $(j = 1, 2)$ must be identical?*

Yi ([18]-[20],[22]) first gave an affirmative answer to Question A. Since then, many results have been obtained for this and related topics (see ([1]-[15]), ([17]-[23])).

Concerning to Question A, a natural question is the following.

Question B. *What is the smallest cardinality for such a finite set S such that any two meromorphic functions f and g satisfying either $E_f(S) = E_g(S)$ or $\overline{E}_f(S) = \overline{E}_g(S)$ must be identical?*

So far, the best answer to Question B for the case of URSM was obtained by Frank and Reinders ([7]). They proved the following result.

Theorem C. *The set $\{z \in \mathbb{C}|\ P_{FR}(z) = \frac{(n-1)(n-2)}{2} z^n + n(n-2)z^{n-1} + \frac{(n-1)n}{2} z^{n-2} - c = 0\}$, where $n \geq 11$ and $c \neq 0, 1$, is a unique range set for meromorphic functions counting multiplicity.*

In 1997, H. X. Yi ([21]) first gave an answer to question B for the case of URSM-IM with 19 elements. Since then, many results have been obtained for this topic (see ([1]- [5])). So far, the best answer to Question B for the case of URSM-IM was obtained by Chakraborty([5]). He proved the following result.

Theorem D. *Let $S_{FR} = \{z \in \mathbb{C}|\ P_{FR}(z) = 0\}$. If $n \geq 15$, then S_{FR} is a URSM-IM.*

In 2022, An([1]) given a class of unique range sets for meromorphic functions ignoring multiplicity with 15 elements. He proved the following result.

Let $n \in \mathbb{N}^*, n \geq 3$. Consider polynomial $P(z)$ as follows:

$$P_A(z) = z^n - \frac{2na}{n-1} z^{n-1} + \frac{na^2}{n-2} z^{n-2} + 1 = Q_A(z) + 1, \tag{1.1}$$

where $a \in \mathbb{C},\ a \neq 0$. Suppose that

$$Q_A(a) \neq -1, \tag{1.2}$$

$$Q_A(a) \neq -2. \tag{1.3}$$

Theorem E. *Let $P_A(z)$ be defined by (1.1) with conditions (1.2) and (1.3), and let $S_A = \{z \in \mathbb{C} |\ P_A(z) = 0\}$. If $n \geq 15$, then S_A is a URSM-IM.*

Clearly, $P_A^{'}(z) = nz^{n-3}(z-a)^2$, and $P_{FR}^{'}(z) = \dfrac{n(n-1)(n-2)}{2} z^{n-3}(z-1)^2$. Therefore, this class is different from Chakraborty's Theorem D in([5]).

In this paper, we give a uniqueness theorem for meromorphic functions ignoring multiplicity, which generalizes Theorem E.

Now let us describe main results of the paper.

Let $q, k, m_1, m_2 \in \mathbb{N}^*$.

We will let $P(z)$ be polynomial having no multiple zeros of degree q in $\mathbb{C}[z]$:

$$P(z) = (m_1+m_1+1)\Big(\sum_{i=0}^{m_2} \tbinom{m_2}{i} \frac{(-1)^i}{m_1+m_2+1-i} z^{m_1+m_2+1-i} a^i\Big) + 1 = Q(z) + 1,$$

where

$$Q(z) = (m_1+m_2+1)\Big(\sum_{i=0}^{m_2} \tbinom{m_2}{i} \frac{(-1)^i}{m_1+m_2+1-i} z^{m_1+m_2+1-i} a^i\Big). \tag{1.4}$$

Suppose that

$$a \neq 0,\ Q(a) \neq -1,\ Q(a) \neq -2. \tag{1.5}$$

Clearly, $P^{'}(z) = (m_1+m_2+1)z^{m_1}(z-a)^{m_2}$, and has a zero at 0 of order m_1, and a zero at a of order m_2. Note that $q = m_1+m_2+1$.

We shall prove the following theorem.

Theorem 1. *Let $P(z)$ be defined in (1.4) with conditions (1.5), and let $S = \{z \in \mathbb{C} |\ P(z) = 0\}$. If $q \geq 15$, then S is a URSM-IM.*

Remark 2. *From proof of Theorem 1 we give a proof of Theorem D, which is different from Chakraborty's proof in([5])*(see section 3.).

Remark 3. *In Theorem 1, take $m_1 = n-3$ and $m_2 = 2$ we obtain Theorem E.*

Indeed, by $P_A^{'}(z) = nz^{n-3}(z-a)^2$ and $P^{'}(z) = (m_1+m_2+1)z^{m_1}(z-a)^{m_2}$, we obtain $P(z) = P_A(z)$ when $m_1 = n-3$, $m_2 = 2$.

2. Lemmas, Definitions

We assume that the reader is familiar with the notations of Nevanlinna theory (see, for example, ([6]), ([16])). We need some lemmas.

Lemma 2.1. ([6], paper 98;[16], paper 43) *Let f be a non-constant meromorphic function on $\mathbb{C}$ and let $a_1, a_2, ..., a_q$ be distinct points of $\mathbb{C} \cup \{\infty\}$. Then*

$$(q-2)T(r,f) \leq \sum_{i=1}^{q} \overline{N}(r, \frac{1}{f-a_i}) - N_0(r, \frac{1}{f'}) + S(r,f),$$

where $N_0(r, \frac{1}{f'})$ is the counting function of those zeros of f', which are not zeros of function $(f-a_1)...(f-a_q)$, and $S(r,f) = o(T(r,f))$ for all r, except for a set of finite Lebesgue measure.

Lemma 2.2. ([6, paper 99]) *For any non-constant meromorphic function f,*

$$T(r, \frac{1}{f'}) \leq 2T(r,f) + S(r,f).$$

Definition. Let f be a non-constant meromorphic function, and k be a positive integer. We denote by $\overline{N}_{(k}(r,f)$ the counting function of the poles of order $\geq k$ of f, where each pole is counted only once. If z is a zero of f, denote by $\nu_f(z)$ its multiplicity. We denote by $\overline{N}(r, \frac{1}{f'}; f \neq 0)$ the counting function of the zeros z of $f^{'}$ satisfying $f(z) \neq 0$, where each zero is counted only once.

Let be given two non-constant meromorphic functions f and g. For simplicity, denote by $\nu_1(z) = \nu_f(z)$ (resp.,$\nu_2(z) = \nu_g(z)$), if z is a zero of f(resp.,g). Let $f^{-1}(0) = g^{-1}(0)$. We denote by $N(r, \frac{1}{f}; \nu_1 = \nu_2 = 1)$(resp., $\overline{N}(r, \frac{1}{f}; \nu_1 > \nu_2 \geq 1)$) the counting function of the common zeros z, satisfying $\nu_1(z) = \nu_2(z) = 1$(resp., $\nu_1(z) > \nu_2(z) \geq 1$, where each zero is counted only once), and by $N(r, \frac{1}{f}; \nu_1 \geq 2)$ the counting function of the zeros z of f, satisfying $\nu_1(z) \geq 2$. Similarly, we define the counting functions $\overline{N}(r, \frac{1}{g}; \nu_2 > \nu_1 \geq 1)$, $N(r, \frac{1}{g}; \nu_2 \geq 2)$.

Lemma 2.3. ([1, Lemma 2.3])

Let f, g be two non-constant meromorphic functions and let $f^{-1}(0) = g^{-1}(0)$. Set

$$F = \frac{1}{f},\ G = \frac{1}{g},\ L = \frac{F^{''}}{F'} - \frac{G^{''}}{G'}.$$

Suppose that $L \not\equiv 0$. *Then*

$$1)\ N(r,L) \le \overline{N}_{(2}(r,f) + \overline{N}_{(2}(r,g) +$$
$$\overline{N}(r,\frac{1}{f};\nu_1 > \nu_2 \ge 1) + \overline{N}(r,\frac{1}{g};\nu_2 > \nu_1 \ge 1) + \overline{N}(r,\frac{1}{f'};f \ne 0) +$$
$$\overline{N}(r,\frac{1}{g'};g \ne 0).$$

Moreover, if a *is a common simple zero of* f *and* g, *then* $L(a) = 0$.

$$2)\ \overline{N}(r,\frac{1}{f}) + \overline{N}(r,\frac{1}{g}) + \overline{N}(r,\frac{1}{f};\nu_1 > \nu_2 \ge 1) + \overline{N}(r,\frac{1}{g};\nu_2 > \nu_1 \ge 1)$$
$$\le N(r,L) + \frac{1}{2}(N(r,\frac{1}{f}) + N(r,\frac{1}{g})) + N(r,\frac{1}{f};\nu_1 \ge 2) + N(r,\frac{1}{g};\nu_2 \ge 2)$$
$$+S(r,f) + S(r,g).$$

A polynomial $R(z)$ is called a *strong uniqueness polynomial for meromorphic (entire) functions* if for arbitrary two non-constant meromorphic (entire) functions f and g, and a nonzero constant c, the condition $R(f) = cR(g)$ implies $f = g$ (see ([2]), ([9]), ([13])). In this case we say $R(z)$ is a SUPM (SUPE). A polynomial $R(z)$ is called a *uniqueness polynomial for meromorphic (entire) functions* if for arbitrary two non-constant meromorphic (entire) functions f and g, the condition $R(f) = R(g)$ implies $f = g$ (see ([2]), ([9]), ([13])). In this case we say $R(z)$ is a UPM (UPE). Let $R(z)$ be a polynomial of the degree q. Assume that the derivative of $R(z)$ has mutually distinct k zeros $d_1, d_2, ..., d_k$ with multiplicities $q_1, q_2, ..., q_k$, respectively. We often consider polynomials satisfying the following condition introduced by Fujimoto ([8]):

$$R(d_i) \ne R(d_j), 1 \le i < j \le q. \tag{2.1}$$

The number k is called the *derivative index* of R.

H. Fujimoto ([8], Proposition 7.1)) proved the following:

Lemma 2.4. *Let* $R(z)$ *be a polynomial of degree* q *satisfying the condition (2.1), we assume furthermore that* $q \ge 5$ *and there are two non-constant meromorphic function* f *and* g *such that*

$$\frac{1}{R(f)} = \frac{c_0}{R(g)} + c_1$$

for two constants $c_0\ \ne 0$ *and* c_1. *If* $k \ge 3$ *or if* $k = 2, min\{q_1, q_2\} \ge 2$, *then* $c_1 = 0$.

Lemma 2.5. ([13], Theorem 1.1)

Let $P(z)$ be defined by (1.4) with conditions (1.5), and let $n \geq 6$. Then $P(z)$ is a strong uniqueness polynomial for meromorphic functions.

Lemma 2.6. ([3], Theorem 1.1)

Let $P_{FR_1}(z) = \frac{(n-1)(n-2)}{2} z^n + n(n-2)z^{n-1} + \frac{(n-1)n}{2} z^{n-2} - c = 0\}$, where $n \geq 8$ and $c \in \mathbb{C}$. Then $P_{FR_1}(z)$ is a strong uniqueness polynomial for meromorphic functions.

3. Proof of Theorems

Proof of Theorem 1

Recall that $P(z) = (z - a_1)...(z - a_q)$, $P^{'}(z) = qz^{m_1}(z - a)^{m_2}$, $q = m_1 + m_2 + 1$.

Suppose $q \geq 15$ and $\overline{E}_f(S) = \overline{E}_g(S)$, where $S = \{z \in \mathbb{C}| \ P(z) = 0\}$. Set

$$F = \frac{1}{P(f)},\ G = \frac{1}{P(g)}, L = \frac{F^{''}}{F'} - \frac{G^{''}}{G'},$$

$$T(r) = T(r, f) + T(r, g), S(r) = S(r, f) + S(r, g).$$

Then $T(r, P(f)) = qT(r, f) + S(r, f)$ and $T(r, P(g)) = qT(r, g) + S(r, g)$, and hence $S(r, P(f)) = S(r, f)$ and $S(r, P(g)) = S(r, g)$.

We consider two following cases:

Case 1. $L \equiv 0$. Then, we have $\frac{1}{P(f)} = \frac{c}{P(g)} + c_1$ for some constants $c \neq 0$ and c_1. By Lemma 2.4 we obtain $c_1 = 0$.

Therefore, there is a constant $C \neq 0$ such that $P(f) = CP(g)$. Then, applying Lemma 2.5 we obtain $f = g$.

Case 2. $L \not\equiv 0$.

Claim 1. We have

$$(q-2)T(r) \leq \overline{N}(r, \frac{1}{P(f)}) + \overline{N}(r, \frac{1}{P(g)}) - N_0(r, \frac{1}{f'}) - N_0(r, \frac{1}{g'}) + S(r), \quad (3.1)$$

where $N_0(r, \frac{1}{f'})$ ($N_0(r, \frac{1}{g'})$) is the counting function of those zeros of f', which are not zeros of function $(f - a_1)...(f - a_q)f(f - a)((g - a_1)...(g - a_q)g(g - a))$.

Indeed, applying the Lemma 2.1 to the functions f, g and the values $a_1, a_2, \ldots, a_q, 0, a, \infty$, and noting that

$$\sum_{i=1}^{q} \overline{N}(r, \frac{1}{f-a_i}) = \overline{N}(r, \frac{1}{P(f)}),\ \sum_{i=1}^{q} \overline{N}(r, \frac{1}{g-a_i}) = \overline{N}(r, \frac{1}{P(g)}),$$

we obtain

$$(q+1)T(r) \leq \overline{N}(r,f)+\overline{N}(r,g)+\overline{N}(r, \frac{1}{P(f)})+\overline{N}(r, \frac{1}{P(g)})+\overline{N}(r, \frac{1}{f})+\overline{N}(r, \frac{1}{g})+$$

$$\overline{N}(r, \frac{1}{f-a}) + \overline{N}(r, \frac{1}{g-a}) - N_0(r, \frac{1}{f'}) - N_0(r, \frac{1}{g'}) + S(r). \tag{3.2}$$

On the other hand,

$$\overline{N}(r,f) + \overline{N}(r,g) \leq (T(r,f) + T(r,g)) + S(r) = T(r) + S(r),$$

$$\overline{N}(r, \frac{1}{f}) + \overline{N}(r, \frac{1}{g}) \leq (T(r,f) + T(r,g)) + S(r) = T(r) + S(r),$$

$$\overline{N}(r, \frac{1}{f-a}) + \overline{N}(r, \frac{1}{g-a}) \leq (T(r,f) + T(r,g)) + S(r) = T(r) + S(r).$$

From this and (3.2) we obtain (3.1).

Claim 2. We have

$$\overline{N}(r, \frac{1}{P(f)}) + \overline{N}(r, \frac{1}{P(g)}) \leq$$

$$(\frac{q}{2} + 3)T(r) + \overline{N}(r, \frac{1}{[P(f)]'}; P(f) \neq 0) + \overline{N}(r, \frac{1}{[P(g)]'}; P(g) \neq 0) + S(r).$$

Indeed, by $\overline{E}_f(S) = \overline{E}_g(S)$ we get $(P(f))^{-1}(0) = (P(g))^{-1}(0)$. For simplicity, we set $\nu_1 = \nu_1(z)$, $\nu_2 = \nu_2(z)$, where $\nu_1(z) = \nu_{P(f)}(z)$, $\nu_2(z) = \nu_{P(g)}(z)$. Note that

$$\overline{N}_{(2}(r, P(f)) = \overline{N}(r,f),\ \overline{N}_{(2}(r, P(g)) = \overline{N}(r,g),$$

$$S(r, P(f)) = S(r,f),\ S(r, P(g)) = S(r,g), S(r) = S(r,f) + S(r,g).$$

Applying the Lemma 2.3 to the functions $P(f), P(g)$. Then we obtain

$$N(r, L) \leq \overline{N}(r,f)+\overline{N}(r,g)+\overline{N}(r, \frac{1}{P(f)}; \nu_1 > \nu_2 \geq 1)+\overline{N}(r, \frac{1}{P(g)}; \nu_2 > \nu_1 \geq 1)$$

$$+\overline{N}(r, \frac{1}{[P(f)]'}; P(f) \neq 0) + \overline{N}(r, \frac{1}{[P(g)]'}; P(g) \neq 0), \tag{3.3}$$

and

$$\overline{N}(r,\frac{1}{P(f)})+\overline{N}(r,\frac{1}{P(g)})+\overline{N}(r,\frac{1}{P(f)};\nu_1>\nu_2\geq 1)+$$

$$\overline{N}(r,\frac{1}{P(g)};\nu_2>\nu_1\geq 1)\leq N(r,L)+\frac{1}{2}(N(r,\frac{1}{P(f)})+N(r,\frac{1}{P(g)}))+$$

$$N(r,\frac{1}{P(f)};\nu_1\geq 2)+N(r,\frac{1}{P(g)};\nu_2\geq 2))+S(r). \tag{3.4}$$

Morover,

$$\overline{N}(r,f)+\overline{N}(r,g)\leq T(r)+S(r). \tag{3.5}$$

Obviously,

$$N(r,\frac{1}{P(f)})\leq qT(r,f)+S(r,f);N(r,\frac{1}{P(g)})\leq qT(r,g)+S(r,g),$$

$$N(r,\frac{1}{P(f)})+N(r,\frac{1}{P(g)})\leq qT(r)+S(r). \tag{3.6}$$

On the other hand, from $P(f)=(f-a_1)...(f-a_q)$ it follows that if z_0 zero is a zero of $P(f)$ with multiplicity ≥ 2, then z_0 is a zero of $f-a_i$ with multiplicity ≥ 2 for some $i\in\{1,2,...,q\}$, and therefore, it is a zero of $f^{'}$, so we have

$$N(r,\frac{1}{P(f)};\nu_1\geq 2)\leq N(r,\frac{1}{f'}).$$

From this and Lemma 2.2 we obtain

$$N(r,\frac{1}{P(f)};\nu_1\geq 2)\leq N(r,\frac{1}{f'})\leq T(r,f^{'})+S(r,f)\leq 2T(r,f)+S(r,f).$$

Similarly, we have

$$N(r,\frac{1}{P(g)};\nu_2\geq 2)\leq N(r,\frac{1}{g'})\leq T(r,g^{'})+S(r,g)\leq 2T(r,g)+S(r,g).$$

Therefore,

$$N(r,\frac{1}{P(f)};\nu_1\geq 2)+N(r,\frac{1}{P(g)};\nu_2\geq 2)\leq 2T(r)+S(r). \tag{3.7}$$

Combining (3.1)-(3.7) we get

$$\overline{N}(r,\frac{1}{P(f)})+\overline{N}(r,\frac{1}{P(g)})\leq$$

$$(\frac{q}{2}+3)T(r)+\overline{N}(r,\frac{1}{[P(f)]'};P(f)\neq 0)+\overline{N}(r,\frac{1}{[P(g)]'};P(g)\neq 0)+S(r).$$

Claim 2 is proved.

Claim 3. We have

$$\overline{N}(r, \frac{1}{[P(f)]'}; P(f) \neq 0) + \overline{N}(r, \frac{1}{[P(g)]'}; P(g) \neq 0) \leq 2T(r) + N_0(r, \frac{1}{f'}) +$$
$$N_0(r, \frac{1}{g'}) + S(r).$$

We have

$$\overline{N}(r, \frac{1}{[P(f)]'}; P(f) \neq 0) = \overline{N}(r, \frac{1}{f^{m_1}(f-a)^{m_2}f'}; P(f) \neq 0) \leq \overline{N}(r, \frac{1}{f}) +$$
$$\overline{N}(r, \frac{1}{f-a}) + \overline{N}_0(r, \frac{1}{f'}) \leq 2T(r, f) + \overline{N}_0(r, \frac{1}{f'}) + S(r, f). \tag{3.8}$$

Similarly,

$$\overline{N}(r, \frac{1}{[P(g)]'}; P(g) \neq 0) \leq 2T(r, g) + \overline{N}_0(r, \frac{1}{g'}) + S(r, g). \tag{3.9}$$

Inequalities (3.8) and (3.9) give us

$$\overline{N}(r, \frac{1}{[P(f)]'}; P(f) \neq 0) + \overline{N}(r, \frac{1}{[P(g)]'}; P(g) \neq 0) \leq$$
$$\leq 2T(r) + \overline{N}_0(r, \frac{1}{f'}) + \overline{N}_0(r, \frac{1}{g'}) + S(r).$$

Claim 3 is proved.

Claim 1, 2, 3 give us:

$$(q-2)T(r) \leq (\frac{q}{2} + 5)T(r) + S(r). \text{ So } (q-14)T(r) \leq S(r).$$

This is a contradiction to the assumption that $q \geq 15$. So $L \equiv 0$. Therefore $f = g$. Theorem 1 is proved.

A proof of Theorem D

By using the arguments similar in proof of Theorem 1 and Lemma 2.6 we give a proof of Theorem D, which is different from Chakraborty's proof of Theorem D in([5]).

Recall that $P_{FR}(z) = (z-a_1)...(z-a_n)$, $P'_{FR}(z) = \frac{n(n-1)(n-2)}{2} z^{n-3}(z-1)^2$.

Suppose $n \geq 15$ and $\overline{E}_f(S_{FR}) = \overline{E}_g(S_{FR})$, where $S_{FR} = \{z \in \mathbb{C}|\ P_{FR}(z) = 0\}$. Set

$$F = \frac{1}{P_{FR}(f)},\ G = \frac{1}{P_{FR}(g)}, L = \frac{F^{''}}{F'} - \frac{G^{''}}{G'},$$
$$T(r) = T(r,f) + T(r,g), S(r) = S(r,f) + S(r,g).$$

Then $T(r, P_{FR}(f)) = nT(r,f) + S(r,f)$ and $T(r, P_{FR}(g)) = nT(r,g) + S(r,g)$, and hence $S(r, P_{FR}(f)) = S(r,f)$ and $S(r, P_{FR}(g)) = S(r,g)$.

We consider two following cases:

Case 1. $L \equiv 0$. Then, we have $\frac{1}{P_{FR}(f)} = \frac{c}{P_{FR}(g)} + c_1$ for some constants $c \neq 0$ and c_1. By Lemma 2.4 we obtain $c_1 = 0$.

Therefore, there is a constant $C \neq 0$ such that $P_{FR}(f) = CP_{FR}(g)$. Then, applying Lemma 2.6 we obtain $f = g$.

Case 2. $L \not\equiv 0$. By using the arguments similar in proof of Theorem 1 we obtain

Claim 1. We have

$$(n-2)T(r) \leq \overline{N}(r, \frac{1}{P_{FR}(f)}) + \overline{N}(r, \frac{1}{P_{FR}(g)}) - N_0(r, \frac{1}{f'}) - N_0(r, \frac{1}{g'}) + S(r), \tag{3.10}$$

where $N_0(r, \frac{1}{f'})$ $(N_0(r, \frac{1}{g'}))$ is the counting function of those zeros of f', which are not zeros of function $(f-a_1)...(f-a_n)f(f-1)((g-a_1)...(g-a_n)g(g-1))$.

Claim 2. We have

$$\overline{N}(r, \frac{1}{P_{FR}(f)}) + \overline{N}(r, \frac{1}{P_{FR}(g)}) \leq$$
$$(\frac{n}{2}+3)T(r)+\overline{N}(r, \frac{1}{[P_{FR}(f)]'}; P_{FR}(f) \neq 0)+\overline{N}(r, \frac{1}{[P_{FR}(g)]'}; P_{FR}(g) \neq 0)+S(r).$$

Claim 3. We have

$$\overline{N}(r, \frac{1}{[P_{FR}(f)]'}; P_{FR}(f) \neq 0)+\overline{N}(r, \frac{1}{[P_{FR}(g)]'}; P_{FR}(g) \neq 0) \leq 2T(r)+N_0(r, \frac{1}{f'})+$$
$$N_0(r, \frac{1}{g'}) + S(r).$$

Claim 1, 2, 3 give us:

$$(n-2)T(r) \leq (\frac{n}{2} + 5)T(r) + S(r). \text{ So } (n-14)T(r) \leq S(r).$$

This is a contradiction to the assumption that $n \geq 15$. So $L \equiv 0$. Therefore $f = g$.

Theorem D is proved.

References

[1] Vu Hoai An, *A new class of unique range sets for meromorphic functions ignoring multiplicity with* 15 *elements,* Journal of Mathematics and Mathematical Sciences(2022), Vol. 1, 1-14.

[2] A. Banerjee, *A new class of strong uniqueness polynomial satisfying Fujimoto's conditions,* Ann. Acad. Sci. Fenn. Math. Vol. 40, 2015, 465-474.

[3] A. Banerjee, B. Chakraborty, S. Mallickc, *Further Investigations on Fujimoto Type Strong Uniqueness Polynomials*, Filomat 31:16 (2017), 5203-5216.

[4] S. Bartels, *Meromorphic functions sharing a set with 17 elements ignoring multiplicities,* Compl. Var. Theory Appl., 39, 85-92 (1999).

[5] B. Chakraborty, *On the Cardinality of a Reduced Unique-Range Set*, Ukr. Math. J., Vol. 72, No. 11, April, 2021, DOI 10.1007/s11253-021-01889-z.

[6] A. A. Goldberg and I. V. Ostrovskii, *Value Distribution of Meromorphic Functions*, Translations of Mathematical Monographs (2008), V.236.

[7] G.Frank and M. Reinders, *A unique range set for meromorphic functions with 11 elements,* Compl. Var. Theory Appl. 37:1, 1998, 185-193.

[8] H. Fujimoto, *On uniqueness of meromorphic functions sharing finite sets,* Amer. J. Math. 122, 2000, 1175-1203.

[9] H.Fujimoto, *On uniqueness polynomials for meromorphic functions,* Nagoya Math. J., 170, 33-46 (2003).

[10] F.Gross, *Factorization of meromorphic functions and some open problems*, Complex Analysis(Proc. Conf . Univ. Kentucky, Lexington, Ky. 1976), pp. 51-69, Lecture Notes in Math. Vol. 599, Springer, Berlin, 1977.

[11] Ha Huy Khoai, Some remarks on the genericity of unique range sets for meromorphic functions, Sci. China Ser. A Mathematics, Vol. 48, 2005, 262-267.

[12] Ha Huy Khoai, Vu Hoai An, and Pham Ngoc Hoa,*On functional equations for meromorphic functions and applications,* Arch. Math, DOI 10.1007/s00013- 017-1093-5, 2017.

[13] Ha Huy Khoai, Vu Hoai An and Nguyen Xuan Lai, *Strong uniqueness polynomials of degree 6 and unique range sets for powers of meromorphic functions,* Intern. J. Math., 2018, DOI:10.1142/S0129167X18500374.

[14] Ha Huy Khoai, Vu Hoai An and Le Quang Ninh, *Value-sharing and uniqueness for L-functions,* Ann. Polonici Math.,2021, 265-278.

[15] Ha Huy Khoai and Vu Hoai An, *Determining an L-function in the extended Selberg class by its preimages of subsets,* Ramanujan Journal, 58, 253-267 (2022).

[16] W.K.Hayman, *Meromorphic Functions*, Clarendon, Oxford(1964).

[17] P.Li and C.C. Yang, *Some further results on the unique range sets of meromorphic functions,* Kodai Math. J. 18, 1995, 437-450.

[18] H.X.Yi , *Uniqueness of Meromorphic Functions and question of Gross*, Sci. China (Ser. A), Vol.37 N0.7, July 1994, 802-813.

[19] H. X. Yi, *A question of Gross and the uniqueness of entire functions*, Nagoya Math. J. Vol. 138 (1995), 169-177.

[20] H. X. Yi, *Unicity theorems for meromorphic and entire functions III,* Bull. Austral. Math. Soc., 53, 71–82 (1996).

[21] H. X. Yi, *The reduced unique range sets for entire or meromorphic functions,* Compl. Var. Theory Appl., 32, 191–198 (1997).

[22] H. X. Yi, *On a question of Gross concerning uniqueness of entire functions*, Bull. Austral. Math. Soc. Vol. 57(1998), 343-349.

[23] H. X. Yi and W.C.Lin, *Uniqueness theorems concerning a question of Gross*, Proc. Japan Acad., Ser. A, 80, 2004, 136-140.

Nguyen Duy Phuong
Thai Nguyen University of Education
Thai Nguyen
Vietnam
phuongnd@tnu.edu.vn

J. Math. Math. Sci. (2023) 129–143

Truncated second main theorem for non-Archimedean meromorphic maps

Si Duc Quang (Hanoi, Vietnam)

(Received Dec. 10, 2022)

Abstract. Let $\mathbb{F}$ be an algebraically closed field of characteristic $p \geq 0$, which is complete with respect to a non-Archimedean absolute value. Let V be a projective subvariety of $\mathbb{P}^M(\mathbb{F})$. In this paper, we will prove some second main theorems for non-Archimedean meromorphic maps of $\mathbb{F}^m$ into V intersecting a family of hypersurfaces in $N-$subgeneral position with truncated counting functions.

1. Introduction and Main results

Let $\mathbb{F}$ be an algebraically closed field of characteristic $p \geq 0$, which is complete with respect to a non-Archimedean absolute value. Let $N \geq n$ and $q \geq N+1$. Let $H_1, \dots, H_q$ be hyperplanes in $\mathbb{P}^n(\mathbb{F})$. The family of hyperplanes $\{H_1\}_{i=1}^q$ is said to be in N-subgeneral position in $\mathbb{P}^n(\mathbb{F})$ if $H_{j_0} \cap \cdots \cap H_{j_N} = \varnothing$ for every $1 \leq j_0 < \cdots < j_N \leq q$.

In 2017, Yan [6] proved a truncated second main theorem for a non - Archimedean meromorphic map into $\mathbb{P}^n(\mathbb{F})$ with a family of hyperplanes in subgeneral position. With the standart notations on the Nevanlinna theory for non-Archimedean meromorphic maps, his result is stated as follows.

Theorem A (cf. [6, Theorem 4.6]) *Let $\mathbb{F}$ be an algebraically closed field of characteristic $p \geq 0$, which is complete with respect to a non-Archimedean absolute value. Let $f : \mathbb{F}^m \to \mathbb{P}^n(\mathbb{F})$ be a linearly non-degenerate non-Archimedean*

Key words and phrases: non-Archimedean, second main theorem, meromorphic mapping, Nevanlinna, hypersurface, subgeneral position.
2020 Mathematics Subject Classification: Primary 11S80, 11J97; Secondary 32H30

meromorphic map with index of independence s and $\mathrm{rank} f = k$. *Let* $H_1, \ldots, H_q$ *be hyperplanes in* $\mathbb{P}^n(\mathbb{F})$ *in N-subgeneral position* $(N \geq n)$. *Then, for all* $r \geq 1$,

$$(q - 2N + n - 1)T_f(r) \leq \sum_{i=1}^{q} N_f^{(a)}(H_i, r) - \frac{N+1}{n+1} \log r + O(1),$$

where

$$a = \begin{cases} p^{s-1}(n-k+1) & \text{if } p > 0, \\ n-k+1 & \text{if } p = 0. \end{cases}$$

Here, the index of independence s and the $\mathrm{rank} f$ are defined in Section 2 (Definition 2.1).

Also, in 2017, An and Quang [2] proved a truncated second main theorem for meromorphic mappings from $\mathbb{C}^m$ into a projective variety $V \subset \mathbb{P}^M(\mathbb{C})$ with hypersurfaces. Motivated by the methods of Yan [6] and An-Quang [2], our aim in this article is to generalize Theorem A to the case where the map f is from $\mathbb{F}^m$ into an arbitrary projective variety V of dimension n in $\mathbb{P}^M(\mathbb{F})$ and the hyperplanes are replaced by hypersurfaces of $\mathbb{P}^M(\mathbb{F})$ in N-subgeneral position with respect to V.

Firstly, we give the following definitions.

Definition B. *Let V be a projective subvariety of $\mathbb{P}^M(\mathbb{F})$ of dimension n $(n \leq M)$. Let $Q_1, \ldots, Q_q$ $(q \geq n+1)$ be q hypersurfaces in $\mathbb{P}^M(\mathbb{F})$. The family of hypersurfaces $\{Q_i\}_{i=1}^q$ is said to be in N-subgeneral position with respect to V if*

$$V \cap \Big(\bigcap_{j=1}^{N+1} Q_{i_j}\Big) = \varnothing \quad \text{for any } 1 \leq i_1 < \cdots < i_{N+1} \leq q.$$

If $N = n$, we just say $\{Q_i\}_{i=1}^q$ is in general position with respect to V.

Now, let V be as above and let d be a positive integer. We denote by $I(V)$ the ideal of homogeneous polynomials in $\mathbb{F}[x_0, \ldots, x_M]$ defining V and by H_d the $\mathbb{F}$-vector space of all homogeneous polynomials in $\mathbb{F}[x_0, \ldots, x_M]$ of degree d. Define

$$I_d(V) := \frac{H_d}{I(V) \cap H_d} \text{ and } H_V(d) := \dim_{\mathbb{F}} I_d(V).$$

Then $H_V(d)$ is called the Hilbert function of V. Each element of $I_d(V)$ which is an equivalent class of an element $Q \in H_d$, will be denoted by $[Q]$,

Definition C. *Let $f : \mathbb{F}^m \to V$ be a non-Archimedean meromorphic map with a reduced representation $\mathbf{f} = (f_0, \ldots, f_M)$. We say that f is degenerate over $I_d(V)$ if there is $[Q] \in I_d(V) \setminus \{0\}$ such that $Q(\mathbf{f}) \equiv 0$. Otherwise, we say that f is non-degenerate over $I_d(V)$.*

We will generalize Theorem A to the following.

Theorem 1.1. *Let V be a projective subvariety of $\mathbb{P}^M(\mathbb{F})$ of dimension n ($n \le M$). Let $\{Q_i\}_{i=1}^q$ be hypersurfaces of $\mathbb{P}^M(\mathbb{F})$ in N-subgeneral position with respect to V with $\deg Q_i = d_i$ ($1 \le i \le q$). Let d be the least common multiple of d_i's. Let f be a non-Archimedean meromorphic map of $\mathbb{F}^m$ into V, which is non-degenerate over $I_d(V)$ with the d^{th}-index of non-degeneracy s and $\mathrm{rank} f = k$. Then, for all $r \ge 1$,*

$$\left(q - \frac{(2N+n-1)H_d(V)}{n+1}\right) T_f(r) \le \sum_{i=1}^{q} \frac{1}{d_i} N_f^{(\kappa_0)}(Q_i, r) - \frac{N(H_d(V)-1)}{nd} \log r + O(1),$$

where

$$\kappa_0 = \begin{cases} p^{s-1}(H_d(V) - k) & \text{if } p > 0, \\ H_d(V) - k & \text{if } p = 0. \end{cases}$$

Here, the d^{th}-index of non-degeneracy s is defined in Section 2 (Definition 2.1). Note that, in the case where $V = \mathbb{P}^n(\mathbb{C}), d = 1, H_d(V) = n+1$, our result will give back Theorem A.

For the case of counting function without truncation level, we will prove the following.

Theorem 1.2. *Let V be a arbitrary projective subvariety of $\mathbb{P}^M(\mathbb{F})$. Let $\{Q_i\}_{i=1}^q$ be hypersurfaces of $\mathbb{P}^M(\mathbb{F})$ in N-subgeneral position with respect to V. Let f be a non-constant non-Archimedean meromorphic map of $\mathbb{F}^m$ into V. Then, for any $r > 0$,*

$$(q - N)T_f(r) \le \sum_{i=1}^{q} \frac{1}{\deg Q_i} N_f(Q_i, r) + O(1),$$

where the quantity $O(1)$ depends only on $\{Q_i\}_{i=1}^q$.

We see that, the above result is a generalization of the previous results in [1, 5].

2. Basic notions and auxiliary results

In this section, we will recall some basic notions from Nevanlinna theory for non-Archimedean meromorphic maps due to Cherry-Ye [3] and Yan [6].

2.1. Non-Archimedean meromorphic function. Let $\mathbb{F}$ be an algebraically closed field of characteristic p, complete with respect to a non-Archimedean absolute value $|\ |$. We set $\|z\| = \max_{1\le i\le m} |z_i|$ for $z = (z_1, \ldots, z_m) \in \mathbb{F}^m$ and define

$$\mathbb{B}^m(r) := \{z \in \mathbb{F}^m; \|z\| < r\}.$$

For a multi-index $\gamma = (\gamma_1, \ldots, \gamma_m) \in \mathbb{Z}^m_{\ge 0}$, define

$$z^\gamma = z_1^{\gamma_1} \cdots z_m^{\gamma_m},\ |\gamma| = \gamma_1 + \cdots + \gamma_m,\ \gamma! = \gamma! \cdots \gamma_m!.$$

For an analytic function f on $\mathbb{F}^m$ (i.e., entire function) given by a formal power series

$$f = \sum_\gamma a_\gamma z^\gamma$$

with $a_\gamma \in \mathbb{F}$ such that $\lim_{|\gamma|\to\infty} |a_\gamma| r^{|\gamma|} = 0$ ($\forall r \in \mathbb{F}^* = \mathbb{F} \setminus \{0\}$), define

$$|f|_r = \sup_\gamma |a_\gamma| r^{|\gamma|}.$$

We denote by $\mathcal{E}_m$ the ring of all analytic functions on $\mathbb{F}^m$.

We define a meromorphic function f on $\mathbb{F}^m$ to be the quotient of two analytic functions $g, h \in \mathcal{E}_m$ such that g and h have no common factors in $\mathcal{E}_m$, i.e., $f = \frac{g}{h}$. We define

$$|f|_r = \frac{|g|_r}{|h|_r}.$$

We denote by $\mathcal{M}_m$ the field of all meromorphic functions on $\mathbb{F}^m$, which is the fractional field of $\mathcal{E}_m$.

2.2. Derivatives and Hasse derivatives. For a meromorphic function $f \in \mathcal{M}_m$ and a multi-index $\gamma = (\gamma_1, \ldots, \gamma_m)$, we set

$$\partial^\gamma f = \frac{\partial^{|\gamma|} f}{\partial z_1^{\gamma_1} \cdots \partial z_m^{\gamma_m}}.$$

Let $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $\beta = (\beta_1, \ldots, \beta_m)$ be multi-indices. We say that $\alpha \ge \beta$ if $\alpha_i \ge \beta_i$ for all $i = 1, \ldots, m$. If $\alpha \ge \beta$, we define

$$\alpha - \beta = (\alpha_1 - \beta_1, \ldots, \alpha_m - \beta_m), \binom{\alpha}{\beta} = \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_m}{\beta_m}.$$

For an analytic function $f = \sum_\alpha a_\alpha z^\alpha$ and a multi-index γ, we define the Hasse derivative of multi-index γ of f by

$$D^\gamma f = \sum_{\alpha \ge \gamma} \binom{\alpha}{\gamma} a_\alpha z^{\alpha - \gamma}.$$

We may verify that $D^\alpha D^\beta f = \binom{\alpha+\beta}{\beta} D^{\alpha+\beta}$ for all $f \in \mathcal{E}_m$. Therefore, the Hasse derivative D can be extended to meromorphic functions in the following way:

- For a multi-index $e_i = (0, \dots, 0, \underset{j^{th}-position}{1}, 0, \dots, 0)$, we set $D_j^k f := D^{ke_i}(f)$.
- For a meromorphic function $f = \frac{g}{h}$ $(g, h \in \mathcal{E}_m)$, we define

$$D^{e_i} = D_j^1 f := \frac{hD_i^1 g - gD_i^1 h}{h^2},\ j = 1, \dots, m.$$

- For $\gamma = (\gamma_1, \dots, \gamma_m)$, we may choose a sequence of multi-indices $\gamma = \alpha^1 > \alpha^2 > \cdots > \alpha^{|\gamma|}$ such that $\alpha^i = \alpha^{i+1} + e_{j_i}$ $(j_i \in \{1, \dots, m\})$ for $1 \le i \le |\gamma| - 1$ and $\alpha^{|\gamma|} = e_{j_{|\gamma|}}$ $(j_{|\gamma|} \in \{1, \dots, m\})$ and define

$$D^{\alpha_i} h = \frac{1}{\binom{\alpha_{i+1}+e_{j_i}}{\alpha_{i+1}}} D^{e_{j_i}} D^{\alpha_{i+1}} h, \forall i = |\gamma| - 1, |\gamma| - 2, \dots, 1.$$

We summarize here the fundamental properties of the Hasse derivative from [6] as follows:

(i) $D^\gamma(f+g) = D^\gamma f + D^\gamma g,\ f, g \in \mathcal{M}_m$.

(ii) $D^\gamma(fg) = \sum_{\alpha,\beta} D^\alpha f D^\beta g,\ f, g \in \mathcal{M}_m$.

(iii) $D^\alpha D^\beta f = \binom{\alpha+\beta}{\beta} D^{\alpha+\beta} f,\ f \in \mathcal{M}_m$

(iv) (Lemma on the logarithmic derivative) For $f \in \mathcal{E}_m$,

$$|D^\gamma f|_r \le \frac{|f|_r}{r^{|\gamma|}},\ |\partial^\gamma f|_r \le \frac{|f|_r}{r^{|\gamma|}}.$$

(v) For $f \in \mathcal{E}_m$ and a multi-index γ, let P be an irreducible element of $\mathcal{E}_m$ that divides f with exact multiplicity e. If $e > |\gamma|$, then $P^{e-|\gamma|}$ divides $D^\gamma f$.

For each integer $k \ge 2$, let

$$\mathcal{M}_m[k] = \{Q \in \mathcal{M}_m : D_j^i Q \equiv 0 \text{ for all } 0 < i < k \text{ and } 1 \le j \le m\}.$$

If F has characteristic 0, then $\mathcal{M}_m[k] = \mathbb{F}$ for all $k \ge 2$. If $\mathbb{F}$ has characteristic $p > 0$ and if $s \ge 1$ is an integer, then $\mathcal{M}_m[p^s]$ is the fraction field of $\mathcal{E}_m$, where $\mathcal{E}_m[p^s] = \{g^{p^s} : g \in \mathcal{E}_m\}$ is a subring of $\mathcal{E}_m$. Moreover,

$$\mathcal{M}_m[p^{s-1}+1] = \mathcal{M}_m[p^s].$$

2.3. Non-Archimedean Nevanlinna's function.

Let $f = \sum_\gamma a_\gamma z^\gamma \in \mathcal{E}_m$ be an holomorphic function. The counting function of zeros of f is defined as follows:

$$N_f(0, r) = n_f(0, 0) \log r + \int_0^r (n_f(0, t) - n_f(0, 0)) \frac{dt}{t}\ (r > 0),$$

where

$$n_f(0,r) = \sup\{|\gamma|; |a_\gamma| r^{|\gamma|} = |f|_r\} \text{ and } n_f(0,0) = \min\{|\gamma|; a_\gamma \neq 0\}.$$

Let f be a meromorphic function on $\mathbb{F}^m$. Assume that $f = \frac{g}{h}$, where g, h are holomorphic functions without common factors. We define

$$N_f(0,r) = N_g(0,r) \text{ and } N_f(\infty, r) = N_h(0,r).$$

The Poisson-Jensen-Green formula (see [3, Theorem 3.1]) states that

$$N_f(0,r) - N_f(\infty, r) = \log|f|_r + C_f \quad \text{for all } r > 0,$$

where C_f is a constant depending on f but not on r.

Suppose that $f \not\equiv a$ for $a \in \mathbb{F}$. The counting function of f with respect to the point a is defined by

$$N_f(a,r) = N_{f-a}(0,r).$$

The proximity functions of f with respect to ∞ and a are defined respectively as follows

$$m_f(\infty, r) = \max\{0, \log|f|_r\} = \log^+|f|_r \text{ and } m_f(a,r) = m_{1/(f-a)}(\infty, r).$$

The characteristic function of f is defined by

$$T_f(r) = m_f(\infty, r) + N_f(\infty, r).$$

Note that, if $f = \frac{g}{h}$ as above then $T_f(r) = \max\{\log|g|_r, \log|h|_r\} + O(1)$.

The first main theorem is stated as follows:

$$T_f(r) = m_f(a,r) + N_f(a,r) + O(1) \ (\forall r > 0).$$

2.4. Truncated counting function.

Let $f \in \mathcal{E}_m$. For $j = 1, \ldots, m$, define

$$g_j = \gcd(f, D_j^1(f)) \text{ and } h_j = \frac{f}{g_j}.$$

The radical $R(f)$ of f is defined to be the least common multiple of h_j's.

Case 1: $\mathbb{F}$ has characteristic $p = 0$. The truncated counting function of zeros of f is defined by

$$N_f^{(l)}(0,r) = N_{\gcd(f, R(f)^l)}(0,r).$$

In particular,

$$N_f^{(1)}(0,r) = N_{R(f)}(0,r).$$

Case 2: $\mathbb{F}$ *has characteristic* $p > 0$. We define $R_{p^s}(f)$ by induction in $s = 0, 1, \dots$ For $s = 0$, set $R_{p^0}(f) = R(f)$. For $s \geq 1$, assume that $R_{p^{s-1}}(f)$ has been defined. We set

$$\overline{f} = \frac{f}{\gcd(f, R_{p^{s-1}}(f)^{p^s})},\ g_i = \gcd(\overline{f}, D_i^{p^s}\overline{f}),\ h_i = \frac{\overline{f}}{g_i}$$

for $i = 1, \dots, m$. Let H be the least common multiple of h_i's, and set

$$G = \frac{H}{\gcd(H, R_{p^{s-1}}(H)^{p^{s-1}})},$$

which is a p^sth power. Let R be the p^sth root of G and define the higher p^s-radical $R_{p^s}(f)$ of f to be the least common multiple of $R_{p^{s-1}}(f)$ and R.

Take a sequence $\{r_j\}_{i\in\mathbb{N}} \subset |\mathbb{F}^*|$ such that $r_j \to \infty$. Take s_j such that if $P \in \mathcal{E}_m$ is irreducible such that $P|f$ and P is not unit on $\mathbb{B}^m(r_j)$ then $P|R_{p^s}(f)$ for $s > s_j$. Let u_j be a unit on $\mathbb{B}^m(r_j)$ such that

$$R_{p^{s_j}}(f) = u_j R_{p^{s_{j+1}}}(f).$$

Define $v_j = \prod_{l=j}^{\infty} u_j$, which is unit on $\mathbb{B}^m(r_j)$, and

$$S(f) = \lim_{j\to\infty} \frac{R_{p^{s_j}}(f)}{v_j} \in \mathcal{E}_m,$$

which is called the square free part of f. The truncated (to level l) counting function of zeros of f is defined by

$$N_f^{(l)}(0,r) = N_{\gcd(f, S(f)^l)}(0,r).$$

2.5. Non-Archimedean meromorphic maps and family of hypersurfaces.

Let V be a projective subvariety of $\mathbb{P}^M(\mathbb{F})$ of dimension n ($n \leq M$). For a positive integer d, take a basis $\{[A_1], \dots, [A_{H_d(V)}]\}$ of $I_d(V)$, where $A_i \in \mathcal{H}_d[x_0, \dots, x_M]$. Let $f : \mathbb{F}^m \to \mathbb{P}^M(\mathbb{F})$ be a non-Archimedean meromorphic map with a reduced representation $\mathbf{f} = (f_0, \dots, f_M)$, which is non-degenerate over $I_d(V)$. We have the following definition.

Definition 2.1. *Assume that* $\mathbb{F}$ *has the character* $p > 0$. *Denote by* s *the smallest integer such that any subset of* $\{A_1(\mathbf{f}), \dots, A_{H_d(V)}(\mathbf{f})\}$ *linearly independent over* $\mathbb{F}$ *remains linearly independent over* $\mathcal{M}_m[p^s]$. *We call* s *is the* d^{th}*-index of non-degeneracy of* f.

We see that the above definition does not depend on the choice of the basis $\{[A_i]; 1 \le i \le H_d(V)\}$ and the choice of the reduced representation $\mathbf{f}$. If $V = \mathbb{P}^M(\mathbb{F})$ and $d = 1$ then s is also called the index of independence of f (see [6, Definition 4.1]).

The following three lemmas are proved in [2] for the case of $\mathbb{F} = \mathbb{C}$ and the canonical absolute value. However, with the same proof, they also hold for arbitrary algebraic closed field $\mathbb{F}$ of character $p \ge 0$ and complete with an arbitrary absolute value. We state them here without the proofs.

Throughout this paper, we sometimes identify each hypersurface in a projective variety with its defining homogeneous polynomial. The following lemma of An-Quang [2] may be considered as a generalization of the lemma on Nochka weights in [4].

Lemma 2.1 (cf. [2, Lemma 3])**.** *Let V be a projective subvariety of $\mathbb{P}^M(\mathbb{F})$ of dimension n $(n \le M)$. Let $Q_1, \ldots, Q_q$ be q $(q > 2N - k + 1)$ hypersurfaces in $\mathbb{P}^M(\mathbb{F})$ in N-subgeneral position with respect to V of the common degree d. Then there are positive rational constants ω_i $(1 \le i \le q)$ satisfying the following:*

i) $0 < \omega_i \le 1,\ \forall i \in \{1, \ldots, q\}$,

ii) Setting $\tilde{\omega} = \max_{j \in Q} \omega_j$, *one gets*

$$\sum_{j=1}^{q} \omega_j = \tilde{\omega}(q - 2N + n - 1) + n + 1.$$

iii) $\dfrac{n+1}{2N - n + 1} \le \tilde{\omega} \le \dfrac{n}{N}$.

iv) For $R \subset \{1, \ldots, q\}$ *with* $\sharp R = N + 1$, *then* $\sum_{i \in R} \omega_i \le n + 1$.

v) Let $E_i \ge 1$ $(1 \le i \le q)$ *be arbitrarily given numbers. For* $R \subset \{1, \ldots, q\}$ *with* $\sharp R = N + 1$, *there is a subset* $R^o \subset R$ *such that* $\sharp R^o = \mathrm{rank}_{\mathbb{F}}\{[Q_i]; i \in R^o\} = n + 1$ *and*

$$\prod_{i \in R} E_i^{\omega_i} \le \prod_{i \in R^o} E_i.$$

Let Q be a hypersurface in $\mathbb{P}^n(\mathbb{F})$ of degree d defined by $\sum_{I \in \mathcal{I}_d} a_I x^I = 0$, where $\mathcal{I}_d = \{(i_0, \ldots, i_M) \in \mathbb{N}_0^{M+1}\ :\ i_0 + \cdots + i_M = d\}$, $I = (i_0, \ldots, i_M) \in \mathcal{I}_d$, $x^I = x_0^{i_0} \cdots x_M^{i_M}$ and $(x_0 : \cdots : x_M)$ is homogeneous coordinates of $\mathbb{P}^M(\mathbb{F})$. Let f be an non-Archimedean meromorphic map from $\mathbb{F}^m$ into a projective subvariety V of $\mathbb{P}^M(\mathbb{F})$ with a reduced representation $\mathbf{f} = (f_0, \ldots, f_M)$. We define

$$Q(\mathbf{f}) = \sum_{I \in \mathcal{I}_d} a_I f^I,$$

where $f^I = f_0^{i_0} \cdots f_n^{i_n}$ for $I = (i_0, \ldots, i_n)$. We have the following lemma.

Lemma 2.2 (cf. [2, Lemma 4]). *Let $\{Q_i\}_{i\in R}$ be a set of hypersurfaces in $\mathbb{P}^n(\mathbb{F})$ of the common degree d and let f be a meromorphic mapping of $\mathbb{F}^m$ into $\mathbb{P}^n(\mathbb{F})$ with a reduced representation $\mathbf{f} = (f_0, \ldots, f_M)$. Assume that $\bigcap_{i\in R} Q_i \cap V = \varnothing$. Then, there exist positive constants α and β such that*

$$\alpha \|\mathbf{f}\|_r^d \leq \max_{i\in R} |Q_i(\mathbf{f})|_r \leq \beta \|\mathbf{f}\|_r^d \text{ for any } r > 0.$$

Lemma 2.3 (cf. [2, Lemma 5]). *Let $\{Q_i\}_{i=1}^q$ be a set of q hypersurfaces in $\mathbb{P}^M(\mathbb{F})$ of the common degree d. Then there exist $(H_V(d)-n-1)$ hypersurfaces $\{T_i\}_{i=1}^{H_V(d)-n-1}$ in $\mathbb{P}^M(\mathbb{F})$ such that for any subset $R \in \{1, \ldots, q\}$ with $\sharp R = \operatorname{rank}_{\mathbb{F}}\{[Q_i]; i \in R\} = n+1$, we get $\operatorname{rank}_{\mathbb{F}}\{\{[Q_i]; i \in R\} \cup \{[T_i]; 1 \leq i \leq H_d(V) - n - 1\}\} = H_V(d)$.*

2.5. Value distribution theory for non-Archimedean meromorphic maps.

Let $f : \mathbb{F}^m \to V \subset \mathbb{P}^M(\mathbb{F})$ be a non-Archimedean meromorphic map with a reduced representation $\mathbf{f} = (f_0, \ldots, f_N)$. The characteristic function of f is defined by

$$T_f(r) = \log \|\mathbf{f}\|_r,$$

where $\|\mathbf{f}\|_r = \max_{1\leq 0 \leq n} |f_i|_r$. This definition is well-defined upto a constant.

Let Q be a hypersurface in $\mathbb{P}^n(\mathbb{F})$ of degree d defined by $\sum_{I\in\mathcal{I}_d} a_I x^I = 0$, where $a_I \in \mathbb{F}$ $(I \in \mathcal{I}_d)$ and are not all zeros. If $Q(\mathbf{f}) \not\equiv 0$ then we define the proximity function of f with respect to Q by

$$m_f(Q, r) = \log \frac{\|\mathbf{f}\|_r^d \cdot \|Q\|}{|Q(\mathbf{f})|_r},$$

where $\|Q\| := \max_{I\in\mathcal{I}_d} |a_I|$. We see that the definition of $m_f(Q, r)$ does not depend on the choices of the presentations of f and Q.

The truncated (to level l) counting function of f with respect to Q is defined by

$$N_f^{(l)}(Q, r) := N_{Q(\mathbf{f})}^{(l)}(0, r).$$

For simplicity, we will omit the character $^{(l)}$ if $l = \infty$.

The first main theorem for non-Archimedean meromorphic maps states that

$$dT_f(r) = m_f(Q, r) + N_f(Q, r) + O(1).$$

Proposition 2.1 (cf. [6, Propositions 4.3, 4.4]). *Let p be the character of $\mathbb{F}$. Assume that $f : \mathbb{F}_m \to \mathbb{P}^n(\mathbb{F})$ is a non-Achimedean meromorphic map, which is* *linearly non-degenerate over* $\mathbb{F}$*, with a reduced representation* $\mathbf{f} = (f_0, \ldots, f_n)$*. Then there exist multi-indices* $\gamma^0 = (0, \ldots, 0), \gamma^1, \ldots, \gamma^n$ *with*

$$|\gamma^0| \leq \cdots \leq |\gamma^n| \leq \kappa_0 \leq \begin{cases} p^{s-1}(n-k+1) & \text{if } p > 0, \\ n-k+1 & \text{if } p = 0 \end{cases}$$

where s *is the index of independence of* f *and* $k = \operatorname{rank} f$*, such that the generalized Wronskian*

$$W_{\gamma^0, \ldots, \gamma^n}(f_0, \ldots, f_n) = \det \left(D^{\gamma^i} f_j \right)_{0 \leq i,j \leq n} \not\equiv 0.$$

Here $\operatorname{rank} f$ is defined by

$$\operatorname{rank} f = \operatorname{rank}_{\mathcal{M}_m} \{ (D^\gamma f_0, \ldots, D^\gamma f_n); |\gamma| \leq 1 \} - 1.$$

3. Proof of main theorems

Proof. [Proof of Theorem 1.1] By replacing Q_i with Q_i^{d/d_i} if necessary, we may assume that all Q_i $(i = 1, \ldots, q)$ do have the same degree d. It is easy to see that there is a positive constant β such that $\beta \|\mathbf{f}\|^d \geq |Q_i(\mathbf{f})|$ for every $1 \leq i \leq q$. Set $Q := \{1, \cdots, q\}$. Let $\{\omega_i\}_{i=1}^q$ be as in Lemma 2.1 for the family $\{Q_i\}_{i=1}^q$. Let $\{T_i\}_{i=1}^{H_d(V)-n-1}$ be $(H_d(V)-n-1)$ hypersurfaces in $\mathbb{P}^M(\mathbb{F})$, which satisfy Lemma 2.3.

Take a $\mathbb{F}$-basis $\{[A_i]\}_{i=1}^{H_V(d)}$ of $I_d(V)$, where $A_i \in H_d$. Since f is non-degenerate over $I_d(V)$, it implies that $\{A_i(\mathbf{f}); 1 \leq i \leq H_V(d)\}$ is linearly independent over $\mathbb{F}$. By Proposition 2.1, there multi-indices $\{\gamma^1 = (0, \ldots, 0), \gamma^2 \cdots, \gamma^{H_V(d)}\} \subset \mathbb{Z}_+^m$ such that $|\gamma^0| \leq \cdots \leq |\gamma^{H_d(V)}| \leq \kappa_0$, where

$$\kappa_0 \leq \begin{cases} p^{s-1}(H_V(d) - k) & \text{if } p > 0, \\ H_d(V) - k & \text{if } p = 0 \end{cases}$$

and the generalized Wronskian

$$W = \det \left(D^{\gamma^i} A_j(\mathbf{f}) \right)_{1 \leq i,j \leq H_d(V)} \not\equiv 0.$$

Here, we note that

$$
\begin{aligned}
k &= \operatorname{rank}_{\mathcal{M}_m}\{(D^\gamma f_0,\dots,D^\gamma f_M);|\gamma|\le 1\}-1\\
&= \operatorname{rank}_{\mathcal{M}_m}\left\{\left(D^\gamma(\frac{f_1}{f_0}),\dots,D^\gamma(\frac{f_M}{f_0})\right);|\gamma|\le 1\right\}\\
&\le \operatorname{rank}_{\mathcal{M}_m}\left\{\left(D^\gamma(\frac{A_2(\mathbf{f})}{A_1(\mathbf{f})}),\dots,D^\gamma(\frac{A_{H_d(V)}(\mathbf{f})}{A_1(\mathbf{f})})\right);|\gamma|\le 1\right\}\\
&= \operatorname{rank}_{\mathcal{M}_m}\left\{(D^\gamma(A_1(\mathbf{f})),\dots,D^\gamma(A_{H_d(V)}(\mathbf{f})));|\gamma|\le 1\right\}-1.
\end{aligned}
$$

For each $R^o=\{r_1^0,\dots,r_{n+1}^0\}\subset\{1,\dots,q\}$ with $\operatorname{rank}_{\mathbb{F}}\{Q_i\}_{i\in R^o}=\sharp R^o=n+1$, set

$$W_{R^o}\equiv\det\big(D^{\gamma^j}Q_{r_v^0}(\mathbf{f})(1\le v\le n{+}1),D^{\gamma^j}T_l(\mathbf{f})(1\le l\le H_V(d){-}n{-}1)\big)_{1\le j\le H_V(d)}.$$

Since $\operatorname{rank}_{\mathbb{F}}\{[Q_{r_v^0}](1\le v\le n+1),[T_l](1\le l\le H_V(d)-n-1)\}=H_V(d)$, there exists a nonzero constant $C_{R^o}\in\mathbb{F}$ such that $W_{R^o}=C_{R^o}\cdot W$.

We denote by $\mathcal{R}^o$ the family of all subsets R^o of $\{1,\dots,q\}$ satisfying

$$\operatorname{rank}_{\mathbb{F}}\{[Q_i];i\in R^o\}=\sharp R^o=n+1.$$

For each $r>0$, there exists $\bar R\subset Q$ with $\sharp\bar R=N+1$ such that $|Q_i(\mathbf{f})|_r\le |Q_j(\mathbf{f})|_r,\forall i\in\bar R,j\notin\bar R$. We choose $R^o\subset R$ such that $R^o\in\mathcal{R}^o$ and R^o satisfies Lemma 2.1(v) with respect to numbers $\{\dfrac{\beta\|\mathbf{f}\|_r^d}{|Q_i(\mathbf{f})|_r}\}_{i=1}^q$. Since $\bigcap_{i\in\bar R}Q_i=\varnothing$, by Lemma 2.2, there exists a positive constant $\alpha^{\bar R}$ such that

$$\alpha^{\bar R}\|\mathbf{f}\|_r^d\le\max_{i\in\bar R}|Q_i(\mathbf{f})|_r.$$

Then, we get

$$
\begin{aligned}
\frac{\|\mathbf{f}\|_r^{d(\sum_{i=1}^q\omega_i)}|W|_r}{|Q_1(\mathbf{f})|_r^{\omega_1}\cdots|Q_q(\mathbf{f})|_r^{\omega_q}} &\le \frac{|W|_r}{\alpha_{\bar R}^{q-N-1}\beta^{N+1}}\prod_{i\in\bar R}\left(\frac{\beta\|\mathbf{f}\|_r^d}{|Q_i(\mathbf{f})|_r}\right)^{\omega_i}\\
&\le A_{\bar R}\frac{|W|_r\cdot\|\mathbf{f}\|_r^{d(n+1)}}{\prod_{i\in\bar R^o}|Q_i(\mathbf{f})|_r}\\
&\le B_{\bar R}\frac{|W_{\bar R^o}|_r\cdot\|\mathbf{f}\|_r^{dH_V(d)}}{\prod_{i\in\bar R^o}|Q_i(\mathbf{f})|_r\prod_{i=1}^{H_V(d)-n-1}|T_i(\mathbf{f})|_r},
\end{aligned}
$$

where $A_{\bar R},B_{\bar R}$ are positive constants.

Therefore, for every $r>0$,

$$\log\frac{\|\mathbf{f}\|_r^{d(\sum_{i=1}^q\omega_i-H_d(V)}|W|_r}{|Q_1(\mathbf{f})|_r^{\omega_1}\cdots|Q_q(\mathbf{f})|_r^{\omega_q}}\le\max_R\log\frac{|W_R|_r}{\prod_{i\in R}|Q_i(\mathbf{f})|_r\prod_{i=1}^{H_V(d)-n-1}|T_i(\mathbf{f})|_r}+O(1)$$
$$\le-\sum_{j=1}^{H_d(V)}|\gamma^j|\log r+O(1),$$

where the maximum is taken over all subsets $R\subset\{1,\dots,q\}$ such that $\sharp R=n+1$ and $\mathrm{rank}_{\mathbb{F}}\{[Q_i];i\in R\}=n+1$. Here, the last inequality comes from the lemma on logarithmic derivative. By the Poisson-Jensen-Green formula, the definitions of the approximation function and the characteristic function, we have

$$\sum_{i=1}^q\omega_i m_f(Q_i,r)-dH_d(V)T_f(r)-N_W(0,r)\le-(H_d(V)-1)\log r+O(1),$$

(note that $\sum_{i=1}^{H_d(V)}|\gamma^i|\le H_d(V)-1$). Then, by the first main theorem, we obtain

$$(\sum_{i=1}^q\omega_i-H_d(V))dT_f(r)\le\sum_{i=1}^q\omega_iN_f(Q_i,r)-N_W(0,r)-(H_d(V)-1)\log r+O(1). \tag{3.1}$$

Claim. $\sum_{i=1}^q\omega_iN_f(Q_i,r)-N_W(0,r)\le\sum_{i=1}^q\omega_iN_f^{(\kappa_0)}(Q_i,r)+O(1)$.

Indeed, set $\tilde G_j=\gcd(Q_j(\mathbf{f}),S(Q_j(\mathbf{f}))^{\kappa_0})$. Since ω_i $(1\le i\le q)$ are rational numbers, there exists an integer A such that $\tilde\omega_i=A\omega_i$ $(1\le i\le q)$ are integers.

Let $P\in\mathcal{E}_m$ be an irreducible element with $P|\prod_{i=1}^qQ_i(\mathbf{f})^{\tilde\omega_i}$. There exists a subset R of $\{1,\dots,q\}$ with $\sharp R=N+1$ such that P is not a division of $Q_i(\mathbf{f})$ for any $i\notin R$. Denote by e_i the largest integer such that $P^{e_i}|Q_i(\mathbf{f})$ for each $i\in R$. Then, there is a subset $R^o\subset R$ with $\sharp R^o=n+1$, $W_{R^o}\not\equiv0$ and

$$\sum_{i\in R}\omega_i\max\{0,e_i-\kappa_0\}\le\sum_{i\in R^o}\max\{0,e_i-\kappa_0\}.$$

Also, since $W=C_{R^o}\cdot W_{R^o}$, it clear that P divides W with multiplicity at least

$$\min_{\{j_1,\dots,j_{n+1}\}\subset\{1,\dots,H_d(V)\}}\sum_{i\in R^0}\min\{0,e_i-|\gamma^{j_i}|\}\ge\sum_{i\in R^0}\min\{0,e_i-\kappa_0\}$$
$$\ge\sum_{i\in R}\omega_i\max\{0,e_i-\kappa_0\}$$
$$=\sum_{i\in R}\omega_i(e_i-\min\{e_i,\kappa_0\}).$$

This implies that

$$P^{\sum_{i\in R}\tilde{\omega}_i e_i}|W^A\cdot P^{\sum_{i\in R}\tilde{\omega}_i\min\{e_i,\kappa_0\}}.$$

We note that $P^{\tilde{\omega}_i\min\{e_i,\kappa_0\}}|G_i^{\tilde{\omega}_i}$. Therefore,

$$P^{\sum_{i\in R}\tilde{\omega}_i e_i}|W^A\cdot\prod_{i\in R}G_i^{\tilde{\omega}_i}.$$

This holds for every such irreducible element P. Then it yields that

$$\prod_{i=1}^{q}Q_i(\mathbf{f})^{\tilde{\omega}_i}|W^A\cdot\prod_{i=1}^{q}G_i^{\tilde{\omega}_i}.$$

Hence,

$$\sum_{i=1}^{q}N_f(Q_i,r)\le N_W(0,r)+\sum_{i=1}^{q}N_f^{(\kappa_0)}(Q_i,r).$$

The claim is proved.

From the claim, Lemma 2.1(ii) and the inequality (3.1), we obtain

$$\begin{aligned}(\tilde{\omega}(q-2N+n-1)&-H_d(V)+n+1)dT_f(r)\\&\le\sum_{i=1}^{q}\omega_i N_f^{(\kappa_0)}(Q_i,r)-(H_d(V)-1)\log r+O(1).\end{aligned}$$

Note that, $\omega_i\le\tilde{\omega}(1\le i\le q)$ and $\dfrac{n+1}{2N-n+1}\le\tilde{\omega}\le\dfrac{n}{N}$. Then, the above inequality implies that

$$\left(q-\frac{(2N-n+1)H_d(V)}{n+1}\right)\le\sum_{i=1}^{q}\frac{1}{d}N_f^{(\kappa_0)}(Q_i,r)-\frac{N(H_d(V)-1)}{nd}\log r+O(1).$$

The theorem is proved.

Proof. [Proof of Theorem 1.2] For $r>0$, without loss of generality, we may assume that

$$|Q_1(\mathbf{f})|_r^{1/\deg Q_1}\le|Q_2(\mathbf{f})|_r^{1/\deg Q_2}\le\cdots\le|Q_q(\mathbf{f})|_r^{1/\deg Q_{N+1}}.$$

Since $\bigcap_{i=1}^{N+1}Q_i=\varnothing$, by Lemma 2.2, there exists a positive constant C such that

$$C\|\mathbf{f}\|_r\le\max_{1\le i\le N+1}|Q_i(\mathbf{f})|_r^{1/\deg Q_i}=|Q_{N+1}(\mathbf{f})|_r^{1/\deg Q_{N+1}}.$$

Then, we get

$$\begin{aligned}\sum_{i=1}^{q}\frac{m_f(Q_i,r)}{\deg Q_i} &= \log\frac{\|\mathbf{f}\|_r^q}{|Q_1(\mathbf{f})|_r^{1/\deg Q_1}\cdots|Q_q(\mathbf{f})|_r^{1/\deg Q_q}}+O(1)\\ &\le \log\prod_{i=1}^{N}\frac{\|\mathbf{f}\|_r}{|Q_i(\mathbf{f})|_r^{1/\deg Q_i}}+O(1)\\ &= \sum_{i=1}^{N}\frac{m_f(Q_i,r)}{\deg Q_i}+O(1)\\ &\le N\cdot T_f(r)+O(1).\end{aligned}$$

Therefore,

$$(q-N)T_f(r)\le\sum_{i=1}^{q}\frac{1}{\deg Q_i}N_f(Q_i,r)+O(1)\quad (r>0).$$

The theorem is proved.

References

[1] T. T. H. An *A defect relation for non-Archimedean analytic curves in arbitrary projective varieties*, Proc. Amer. Math. Soc. **135** (2007), 1255–1261.

[2] D. P. An, S. D. Quang *Second main theorem and unicity of meromorphic mappings for hypersurfaces in projective varieties*, Acta Math. Vietnamica **42** (2017), 455–470.

[3] W. Cherry and Z. Ye, *Non-Archimedean Nevanlinna theory in several variables and the non-Archimedean Nevanlinna inverse problem*, Trans. Amer. Math. Soc. **349** (1997), 5043–5071.

[4] E. I. Nochka, *On the theory of meromorphic functions*, Sov. Math. Dokl. **27** (1983), 377–381.

[5] M. Ru, *A note on p-adic Nevanlinna theory*, Proc. Amer. Math. Soc. **129** (2001), 1263–1269.

[6] Q. Yan, *Truncated second main theorems and uniqueness theorems for non-Archimedean meromorphic maps*, Ann. Polon. Math. **119** (2017), 165–193

Si Duc Quang
Department of Mathematics
Hanoi National University of Education
136 Xuan Thuy, Cau Giay, Hanoi,
Vietnam
quangsd@hnue.edu.vn