J. Math. Math. Sci. (2025) 1–30

On invariants of plane curve singularities in positive characteristic

Nguyen Hong Duc (Hanoi, Vietnam)

(Received Dec. 26, 2024; accepted Jan. 20, 2025)

Abstract. In this survey paper we give an overview on some aspects of singularities of algebraic plane curves over an algebraically closed field of arbitrary characteristic. We review, in particular, classical results and recent developments on invariants of plane curve singularities.

1. Introduction

The study of plane curve singularities started with fundamental work of Heisuke Hironaka on the resolution of singularities ([20] 1964), Oskar Zariski's studies in equisingularity ([30] 1965-1968), Michael Artin's paper on isolated rational singularities of surfaces ([4] 1966), and the work by René Thom, Bernard Malgrange, John Mather,... on singularities of differentiable mappings. It culminated in the 1970ties and 1980ties with the work of John Milnor and Pierre Deligne, who intorduced what is now called the Milnor fibration, Milnor number and the Milnor's formula ([23] 1968, [11] 1973), Egbert Brieskorn's discovery of exotic spheres as neighborhood boundaries of isolated hypersurface singularities (1966) and the connection to Lie groups (1971), Vladimir Arnold's classification of singularities ([1, 2, 3] 1972-1976), and many others, e.g. Andrei Gabrielov, Sabir Gusein-Zade, Ignaciao Luengo, Seidenberg, Walker, Antonio Campillo, C.T.C. Wall, A. Melle-Hernández, Johnatan Wahl, Le Dung Trang, Bernard Teissier, Dierk Siersma, Joseph Steenbrink, Gert-Martin Greuel, Yousra Boubakri, Thomas Markwig, Félix Delgado de la

Key words and phrases: Classification of singularities, Modality, Normal forms
2020 Mathematics Subject Classification: Primary 14B05; Secondary 14J10

Mata, P. Cassou-Noguès, E. Garcıa Barroso, Arkadiusz Ploski, Hefez Abramo, Olmedo Rodrigues, Rodrigo Salomão ... (see [8, 9, 28, 27, 12, 29, 24, 10, 7, 13, 14, 18]). In this survey paper we give an overview on some aspects of singularities of algebraic plane curves over an algebraically closed field of arbitrary characteristic. We review, in particular, classical results and recent developments on invariants of plane curve singularities that should serve as a quick guide to references.

In this note, by a plane curve singularity we mean a non-unit formal power series in $\mathbb{k}[\![x,y]\!]$. Invariants of a plane curve singularity f will be quantities (e.g. integers) associated to f which is stabe in right or contact equivalent classes. Recall that two plane curve singularities f and g are right equivalent if $f = \Phi(g)$ for some automorphism of local $\mathbb{k}$-algebra $\mathbb{k}[\![x,y]\!]$. They are called contact equivalent if $f = u \cdot \Phi(g)$ for some automorphism of local $\mathbb{k}$-algebra $\mathbb{k}[\![x,y]\!]$ and for some unit $u \in \mathbb{k}[\![x,y]\!]$. We denote by $f \sim_r g$ and by $f \sim_c g$ respectively.

We study classical invariants of plane curve singularities such as multiplicity, Milnor number ($\mu(f)$), delta and kappa invariants ($\delta(f), \kappa(f)$), semi-group ($S(f)$) and their relations. Especially we are interested in studying the Milnor formula in positive characteristic, which states, in characteristic zero, that for any reduced plane curve singularity f

$$\mu(f) = 2\delta(f) - r(f) + 1, \tag{1.1}$$

where $r(f)$ denotes the number of branches of f. More precisely, we give some partial answers to the following problem.

Problem 1. *Is there at least a “reasonable” characterization of those plane curve singularities such that Equation 1.1 holds?*

2. Preliminaries

2.1. Resolution of singularities

Let $0 \neq f \in \mathfrak{m} \subset \mathbb{k}[\![x,y]\!]$. Then $R := R_f = \mathbb{k}[\![x,y]\!]/(f)$ (or f) is called a *plane curve singularity*. There is a unique (up to multiplication with units) decomposition

$$f = f_1^{\rho_1} \cdot \ldots \cdot f_r^{\rho_r},$$

with $f_i \in \mathfrak{m}$ irreducible in $\mathbb{k}[\![x,y]\!]$ and $\rho_i \geq 1$ for all $i = 1, \ldots, r$. The series f_i resp. the rings R_{f_i} are called the *branches* of f resp. of R_f. The plane curve singularity f is said *reduced* if $\rho_i = 1$ for $i = 1, \ldots, r$. It is *irreducible* if it reduced and $r = 1$. Recall that the multiplicity of f, denoted by $\mathrm{mt}(f)$, is the minimal degree of the homogeneous part of f. So

$$f = \sum_{k \geq m := \mathrm{mt}(f)} f_k(x, y),$$

where f_k is is either zero or a homogeneous polynomial of degree k and $f_m \neq 0$. Then f_m is decomposed into linear factors,

$$f_m = \prod_{i=1}^{s} (\alpha_i x - \beta_i y)^{r_i},$$

with $(\beta_i : \alpha_i) \in \mathbb{P}^1$ pairwise distinct. We call f_m the *tangent cone* of f. The points $P_i := (\beta_i : \alpha_i), i = 1, \ldots, s$, are the *tangent directions* or the *infinitely near points in the* 1*-st neighbourhood of* 0 of f. For each i, the number r_i is called the multiplicity of P_i, and denoted by m_{P_i}. Note that $m = r_1 + \ldots + r_s$.

For each tangent direction $P := (\beta : \alpha)$ of f, we define a morphism $\iota_P \colon \Bbbk[\![x, y]\!] \to \Bbbk[\![x_P, y_P]\!]$ and a series $f_P \in \Bbbk[\![x_P, y_P]\!]$ as follows

- if $\alpha \neq 0$ then

$$\iota_P(x) = \frac{x_P y_P + \beta y_P}{\alpha}, \iota_P(y) = y_P, \text{ and } \iota_P(f) = y_P^m f_P$$

- if $\alpha = 0$ then

$$\iota_P(x) = x_P, \iota_P(y) = \frac{\alpha x_P + x_P y_P}{\beta}, \text{ and } \iota_P(f) = x_P^m f_P.$$

The series f_P is called the *local equation of the strict transform of f at P*. For each $n \geq 1$, if P is an infinitely near points in the n-th neighbourhood of 0, and if Q is a tangent direction of $f_P(x_P, y_P)$, then Q is called *an infinitely near point in the* $(n+1)$*-th neighbourhood of* 0, denoted by $Q \overset{n+1}{\to} 0$ or simply $Q \to 0$. We also denote by $f_Q(x_Q, y_Q)$ the local equation of the strict transform of f_P at Q. Note that, by definition, if $Q \to P$ then $\mathrm{mt}(f_Q) \leq \mathrm{mt}(f_P)$. The following lemma can be proved easily by using induction.

Lemma 2.1. *Let $f, g \in \Bbbk[\![x, y]\!]$ be plane curve singularities and let P be a tangent direction of f of multiplicity r. Then*

(i) If $m = \mathrm{mt}(f) = 1$, then $\mathrm{mt}(f_P) = 1$ for all $P \to 0$.

(ii) We have $\mathrm{mt}(f_P) \leq r \leq m$. In particular, if f has at least two tangent directions, then $\mathrm{mt}(f_P) < m$.

(iii) Assume g is an irreducible component of f. Then if $Q \overset{n}{\to} 0$ for g then $Q \overset{n}{\to} 0$ for f, and moreover, g_Q is an irreducible component of f_Q.

We denote by $R^{(n)}$ the ring

$$R^{(n)} := R_f^{(n)} := \bigoplus_{Q \overset{n}{\to} 0} \Bbbk[\![x_Q, y_Q]\!]/f_Q(x_Q, y_Q),$$

and call it the *n-th strict transform of f*. Then we have the following inclusions

$$R = R^{(0)} \hookrightarrow R^{(1)} \hookrightarrow \ldots \hookrightarrow R^{(n)} \to \ldots \tag{2.1}$$

defined inductively as

$$R = R^{(0)} \hookrightarrow R^{(1)} = \bigoplus_{P \overset{1}{\to} 0} \Bbbk[\![x_P, y_P]\!]/f_P(x_P, y_P), g \mapsto \oplus \iota_P(g).$$

Theorem 2.1. *Let $f \in \Bbbk[\![x, y]\!]$ be a reduced plane curve singularities. Then*

(i) the sequence of injective morphisms (2.1) stabilizes. More precisely, there exists $k \geq 1$ such that

$$R^{(n)} \cong \bigoplus_{i=1}^{r} \Bbbk[\![t]\!],$$

for all $n \geq k$;

(ii) the morphisms $R^{(i)} \hookrightarrow R^{(i+1)}$ are integral extensions in the quotient ring $Q(R)$ of R;

(iii) the ring

$$R^{(n)} \cong \bigoplus_{i=1}^{r} \Bbbk[\![t]\!],$$

for all $n \gg 1$ is the integral closure of R, is also called the normalization *of R and denoted by $\bar{R}$.*

Proposition 2.2. *Any plane curve singularity $f \in \Bbbk[\![x, y]\!]$ can be factorized as*

$$f = \prod_{P \overset{1}{\to} 0} \bar{f}_P$$

in $\Bbbk[\![x, y]\!]$ such that $\bar{f}_P$ has a unique tangent direction, and the $\bar{f}_P$ are pairwise coprime.

In particular, if $f \in \mathfrak{m} \subset \Bbbk[\![x, y]\!]$ is irreducible, then f has a unique tangent direction.

2.2. Parametrization equivalence

Definition 2.1. Let $0 \neq f \in \mathfrak{m} \subset \Bbbk[[x,y]]$ be reduced and $R \hookrightarrow \bar{R}$ be its normalization. A composition of the natural projection $\Bbbk[[x,y]] \twoheadrightarrow R$, the normalization $R \hookrightarrow \bar{R}$ and an isomorphism $\bar{R} \cong \bigoplus_{i=1}^{r} \Bbbk[[t]]$,

$$\psi \colon \Bbbk[[x,y]] \twoheadrightarrow R \hookrightarrow \bar{R} \cong \bigoplus_{i=1}^{r} \Bbbk[[t]]$$

is called a *(primitive) parametrization* of f (or of R). More precisely,

(a) if f is irreducible, then a parametrization of f is given by a map

$$\psi \colon \Bbbk[[x,y]] \longrightarrow \Bbbk[[t]], (x,y) \mapsto (x(t), y(t)),$$

(b) if f decomposes into several branches, then a parametrization of R is given by a set of parametrizations of the branches. More precisely, if $f = f_1 \cdot \ldots \cdot f_r$ is a decomposition of f into irreducible factors, then $\bar{R} \cong \bigoplus_{i=1}^{r} \Bbbk[[t]]$ is the normalization of R and a parametrization ψ of R can be represented as a matrix of the form:

$$\psi(t) = (\psi_1(t), \ldots, \psi_r(t)),$$

where for $i = 1, \ldots, r$, $(\psi_i(t) = (x_i(t), y_i(t))$ represents a parametrization of the i-th branch.

A parametrization of a reduced plane curve singularity has the following properties:

Proposition 2.3. *Let $0 \neq f \in \mathfrak{m} \subset \Bbbk[[x,y]]$ be reduced and $\psi \colon \Bbbk[[x,y]] \twoheadrightarrow R \hookrightarrow \bar{R} \cong \bigoplus_{i=1}^{r} \Bbbk[[t]]$ be its parametrization. Then*

(i) $ker(\psi) = (f)$,

(ii) ψ satisfies the following universal factorization property: Each $\psi' : \Bbbk[[x,y]] \to \bigoplus_{i=1}^{r} \Bbbk[[t]]$ such that $\psi'(f) = 0$, factorizes in a unique way through ψ, that is there exists the unique morphism $\phi : \bigoplus_{i=1}^{r} \Bbbk[[t]] \to \bigoplus_{i=1}^{r} \Bbbk[[t]]$ such that $\psi' = \phi \circ \psi$. Moreover, if ψ' is also a parametrization of f, then ϕ is an isomorphism.

Proposition 2.4. *Let $f \in \Bbbk[[x,y]]$ be irreducible such that $m := \mathrm{mt}(f) = \mathrm{ord} f(0,y)$. Assume that m is not divisible by $\mathrm{char}(\Bbbk)$, then f has a* **Puiseux parametrization**, *i.e. a parametrization of the form*

$$(x(t)|y(t)) := (t^m | \sum_{k \geq m} c_k t^k).$$

Moreover, there exists a unit $u \in \Bbbk[\![x,y]\!]$ *such that*

$$f = u \cdot \prod_{j=1}^{m} (y - y(\xi^j x^{1/m})),$$

where ξ *is a primitive* m*-th root of unity.*

Now we define the notion of parametrization equivalence.

Definition 2.2. Let $\psi, \psi' \colon \Bbbk[\![x,y]\!] \to \bar{R} = \bigoplus_{i=1}^{r} \Bbbk[\![t]\!]$. Then ψ is said to be *equivalent* to ψ', $\psi \sim \psi'$, if there exist a reparametrization $\phi \in Aut_{\Bbbk}(\bar{R})$ and a coordinate change $\Phi \in Aut_{\Bbbk}(\Bbbk[\![x,y]\!])$ such that $\psi' \circ \Phi = \phi \circ \psi$.

Let $f, g \in \Bbbk[\![x,y]\!]$ be reduced. Then f is said to be *parametrization equivalent* to g, $f \sim_p g$, if there exist a parametrization ψ of f and a parametrization ψ' of g such that $\psi \sim \psi'$.

Note that, if $f \sim_p g$, then for any parametrization ψ (resp. ψ') of f (resp. g) we have $\psi \sim \psi'$ by Proposition 2.3(ii).

Proposition 2.5. *Let* f, g *be two given plane curve singularities. Then*

$$f \sim_p g \Leftrightarrow f \sim_c g.$$

Proof.

cf. [25, Proposition 1.2.10]. ■

2.3. Intersection multiplicity and classical invariants

Definition 2.3. Let $f \in \Bbbk[\![x,y]\!]$ be reduced and let $\psi \colon \Bbbk[\![x,y]\!] \twoheadrightarrow R \hookrightarrow \bar{R} \cong \bigoplus_{i=1}^{r} \Bbbk[\![t]\!]$ be a parametrization of f.

(a) We call $\delta(f) := \dim_{\Bbbk} \bar{R}/R$ the δ*-invariant* of f.

(b) We introduce the *valuation map*

$$v := (v_1, ..., v_r) \colon R \to (\mathbb{Z}_{\geq 0} \cup \infty)^r, g \mapsto \mathrm{ord}(g(x_i(t), y_i(t)))_{i=1,\ldots,r}.$$

Its image $\Gamma(R) := \Gamma(f) := v(R)$ is a semigroup, called the *semigroup of values* of f.

(c) Let $\mathscr{C} := (R : \bar{R}) := \{u \in R \mid u\bar{R} \subset R\}$ be the *conductor ideal* of $\bar{R}$ in R (cf. [31]). Then $\mathscr{C}$ is an ideal of both R and $\bar{R}$. So one has $\mathscr{C} = (t^{c_1}) \times \cdots \times (t^{c_r})$ for some $\mathbf{c} := (c_1, \ldots, c_r) \in \mathbb{Z}^r_{\geq 0}$. We call $\mathbf{c}$ the *conductor (exponent)* of f. One obviously has $\mathbf{c} + \mathbb{Z}^r_{\geq 0} \subset S(f)$ and $\mathbf{c}$ is the minimum element in $S(f)$ with this property w.r.t. the product ordering on $\mathbb{Z}^r_{\geq 0}$, i.e. the partial ordering given by: if $\alpha = (\alpha_1, \ldots, \alpha_r), \beta = (\beta_1, \ldots, \beta_r) \in \mathbb{Z}^r_{\geq 0}$ the $\alpha \leq \beta$ if and only if $\alpha_i \leq \beta_i$ for every $i = 1, \ldots, r$.

Definition 2.4. Let $g \in \mathbb{k}[\![x, y]\!]$ be irreducible and $(x(t), y(t))$ its parametrization. Then the intersection multiplicity of any $f \in \mathbb{k}[\![x, y]\!]$ with g is given by

$$i_0(f, g) := \mathrm{ord} f(x(t), y(t)).$$

If u is a unit then we define $i_0(f, u) := 0$.

The intersection multiplicity of f with a plane curve singularity $g = g_1 \cdot \ldots \cdot g_s$, g_i irreducible, is defined to be the sum

$$i(f, g) := i_0(f, g_1) + \ldots + i(f, g_s).$$

The Milnor number $\mu(f)$ and kappa invariant $\kappa(f)$ of f are defined respectively as

$$\mu(f) := i_0(f_x, f_y);\ \kappa(f) := i_0(f, \alpha f_x + \beta f_y),$$

where $(\alpha : \beta) \in \mathbb{P}^1$ is generic.

Proposition 2.6. *Let $f, g \in \mathbb{k}[\![x, y]\!]$. Then*

$$i_0(f, g) = i_0(g, f) = \dim \mathbb{k}[\![x, y]\!]/(f, g).$$

Proof.

cf. [15, Proposition I.3.12] ■

Corollary 2.1. *Let $f \in \mathbb{k}[\![x, y]\!]$ be an irreducible plane curve singularity. A couple $(x(t), y(t))$ of two power series is a parametrization of f if and only if*

$$f\left(x(t), y(t)\right) = 0 \text{ and } \min\{\mathrm{ord}x(t), \mathrm{ord}y(t)\} = \mathrm{mt}(f).$$

Proposition 2.7. *Let $f, g \in \mathbb{k}[\![x, y]\!]$ be two reduced power series which have no factor in common. Then*

$$\delta(fg) = \delta(f) + \delta(g) + i_0(f, g)$$

and

$$\kappa(fg) = \kappa(f) + \kappa(g) + i_0(f, g).$$

Proof.

cf. [15, Proposition I.3.32, Corollary 3.39] ■

Proposition 2.8. *Let $f \in \mathbb{k}[\![x, y]\!]$ be a reduced plane curve singularity. Then*

$$\delta(f) = \sum_{Q \to 0} \frac{\mathrm{mt}(f_Q)\,(\mathrm{mt}(f_Q) - 1)}{2}.$$

Proof.

cf. [15, Proposition I.3.34] ∎

Proposition 2.9. *Let $f = f_1 \cdot \ldots \cdot f_r$ with f_i irreducible and let $\mathbf{c} = (c_1, \ldots, c_r)$ its conductor. Then, for any $i = 1, \ldots, r$ one has*

$$\begin{aligned} c_i &= 2\delta(f_i) + \sum_{j \neq i} i_0(f_i, f_j)) \\ &= c(f_i) + \sum_{j \neq i} i_0(f_i, f_j) \end{aligned}$$

and therefore $2\delta(f) = c(f) := c_1 + \cdots + c_r$.

Proof.

cf. [19]. ∎

Lemma 2.2 (Dedekind's formula)**.** *Suppose that $i_0(f, x) = \mathrm{ord}(f) \neq 0 \mod p$. Then*

$$i_0\left(f, \frac{\partial f}{\partial y}\right) = c(f) + \mathrm{ord}(f) - 1.$$

For more facts on the conductor see [15], [12], [19]. The following proposition says that the δ-invariant, the conductor and the maximal contact multiplicity are invariant under contact equivalence, and by Proposition 2.5, they are also invariant under parametrization equivalence.

Proposition 2.10. *Let $f, g \in \Bbbk[[x, y]]$, let $u, v \in \Bbbk[[x, y]]^*$ be unit and let $\Phi \in \mathrm{Aut}_{\Bbbk}(\Bbbk[[x, y]])$. Then $i(f, g) = i(u \cdot \Phi(f), v \cdot \Phi(g))$. Moreover, if $f \sim_c g$, then*

(i) $\delta(f) = \delta(g)$.

(ii) $\kappa(f) = \kappa(g)$.

(iii) $\mathbf{c}(f) = \mathbf{c}(g)$ *(up to a permutation of the indices* $\{1, \ldots, r\}$*).*

Proof.

cf. [25, Proposition 1.2.19]. ∎

For reduced plane curve $f = f_1 \cdot \ldots \cdot f_r$ with f_i irreducible we define

1. $\underline{\mathrm{mt}}(f) := (\mathrm{mt}(f_1), \ldots, \mathrm{mt}(f_r)) \in \mathbb{Z}^r$ the *multi-multiplicity* of f,

2. $\underline{c}(f) := (c(f_1), \ldots, c(f_r)) = (2\delta(f_1), \ldots, 2\delta(f_r)) \in \mathbb{Z}^r$ the *multi-conductor* of f.

These tuples are invariant under parametrization and contact equivalence as the following corollary shows.

Corollary 2.2. *If* $f \sim_c g$ *then* $\underline{\mathrm{mt}}(f) = \underline{\mathrm{mt}}(g)$ *and* $\underline{\mathrm{c}}(f) = \underline{\mathrm{c}}(g)$ *(up to a permutation of the indices* $\{1, \dots, r\}$*).*

Proof.

Follows from Proposition 2.10. ■

We recall that if f is a plane curve singularity then its Milnor number $\mu(f)$ is $\dim \Bbbk[[x, y]]/(f_x, f_y)$, where f_x, f_y be the partials of f. Proposition 2.6 yields that the Milnor number can be computed as an intersection multiplicity: $\mu(f) = i(f_x, f_y)$.

2.4. Newton diagrams and Newton factorizations

Let us recall the definition of the Newton diagram of a plane curve singularity. To each power series $f = \sum_{(\alpha,\beta)} c_{\alpha,\beta} x^\alpha y^\beta \in \Bbbk[[x, y]]$ we can associate its *Newton polyhedron* $\Gamma_+(f)$ as the convex hull of the set

$$\bigcup_{\alpha \in \mathrm{supp}(f)} ((\alpha, \beta) + \mathbb{R}^2_{\geq 0}).$$

where $\mathrm{supp}(f) = \{\alpha | c_{\alpha,\beta} \neq 0\}$ denotes the support of f. This is an unbounded polytope in $\mathbb{R}^n$. We call the union $\Gamma(f)$ of its compact faces the *Newton diagram* of f. By $\Gamma_-(f)$ we denote the union of all line segments joining the origin to a point on $\Gamma(f)$. For each subset Δ in $\mathbb{R}^2_{\geq 0}$ we denote

$$\mathrm{in}_\Delta(f) := \sum_{(\alpha,\beta) \in \Delta} c_{\alpha,\beta} x^\alpha y^\beta \in \Bbbk[[x, y]].$$

The initial part of f is defined to be

$$f_{in} := \mathrm{in}_{\Gamma(f)}(f).$$

Proposition 2.11. *Let* $f \in \mathfrak{m} \subset \Bbbk[[x, y]]$ *be an irreducible plane curve singularity such tha* $i_0(f, x) = n$ *and* $i_0(f, x) = m$*. Let* $(x(t), y(t))$ *be parametrization of* f*. Then*

(i) $\mathrm{ord}(x(t)) = n$ *and* $\mathrm{ord}(y(t)) = m$*.*

(ii) The Newton diagram of f *is a straight line segment.*

(iii) There exist $\xi, \lambda \in \Bbbk^*$ *such that*

$$f_{in}(x, y) = \xi \cdot (x^{m/q} - \lambda y^{n/q})^q,$$

where $q = (m, n)$*.*

Proof.

cf. [9, Lemma 3.4.3, 3.4.4, 3.4.5]. ■

Proposition 2.12. *[8, Lemma 3] Let $f \in \Bbbk[\![x,y]\!]$ and let $E_i, i = 1, \ldots, k$ be the edges of its Newton diagram. Then there is a factorization of f:*

$$f = \text{monomial} \cdot \bar{f}_1 \cdot \ldots \cdot \bar{f}_k$$

such that $\bar{f}_i$ is convenient, and $\mathrm{in}_{E_i}(f) = \text{monomial} \times (\bar{f}_i)_{in}$. In particular, if f is convenient then $f = \bar{f}_1 \cdot \ldots \cdot \bar{f}_k$.

3. Milnor numbers and delta invariants

3.1. Milnor numbers

We first introduce the different notions of non-degeneracy originated by Kouchnirenko and Wall. For this let

$$f = \sum_{\alpha} c_\alpha x^\alpha \in \mathfrak{m} \subset \Bbbk[\![\mathbf{x}]\!] := \Bbbk[\![x_1, \ldots, x_n]\!]$$

be a power series, let P be a C-polytope, i.e. a compact rational polytope P of dimension $n-1$ in the positive orthant $\mathbb{R}^n_{\geq 0}$ and the region above P is convex and every ray in the positive orthant emanating from the origin meets P in exactly one point. For each subset $\Delta \subset \mathbb{R}^n_{\geq 0}$ we denote by $f_\Delta := \mathrm{in}_\Delta(f) := \sum_{\alpha \in \Delta} c_\alpha x^\alpha$ the initial form or principal part of f along Δ. Following Kouchnirenko we call f *non-degenerate* (ND) *along* Δ if the Jacobian ideal* $\mathrm{j}(\mathrm{in}_\Delta(f))$ has no zero in the torus $(\Bbbk^*)^n$. f is then said to be *Newton non-degenerate* (NND) if f is non-degenerate along each face (of any dimension) of the Newton diagram $\Gamma(f)$. We do not require f to be convenient.

To define inner non-degeneracy we need to fix two more notions. The face Δ is an *inner face* of P if it is not contained in any coordinate hyperplane. Each point $q \in \Bbbk^n$ determines a coordinate hyperspace $H_q = \bigcap_{q_i = 0}\{x_i = 0\} \subset \mathbb{R}^n$ in $\mathbb{R}^n$. We call f *inner non-degenerate* (IND) *along* Δ if for each zero q of the Jacobian ideal $\mathrm{j}(\mathrm{in}_\Delta(f))$ the polytope Δ contains no point on H_q. f is called *inner Newton non-degenerate* (INND) *w.r.t. a C-polytope* P if no point of $\mathrm{supp}(f)$ lies below P and f is IND along each inner face of P. We call f simply *inner Newton non-degenerate* (INND) if it is INND w.r.t some C-polytope.

*The Jacobian ideal $j(f)$ denotes the ideal generated by all partials of $f \in \Bbbk[\![x]\!]$.

Finally, we call f *weakly non-degenerate* (WND) *along* Δ if the Tjurina ideal[†] $tj(\text{in}_\Delta(f))$ has no zero in the torus $(\Bbbk^*)^n$, and f is called *weakly Newton non-degenerate* (WNND) if f is weakly non-degenerate along each facet of $\Gamma(f)$. Note that NND implies WNND while NND does not imply INND. See [7, Remark 3.1] for facts on and relations between the different types of non-degeneracy.

For any compact polytope Q in $\mathbb{R}^n_{\geq 0}$ we denote by $V_k(Q)$ the sum of the k-dimensional Euclidean volumes of the intersections of Q with the k-dimensional coordinate subspaces of $\mathbb{R}^n$ and, following Kouchnirenko, we then call

$$\mu_N(Q) = \sum_{k=0}^{n} (-1)^{n-k} k! V_k(Q)$$

the Newton number of Q. For a power series $f \in \Bbbk[\![\mathbf{x}]\!]$ we define the *Newton number* of f to be

$$\mu_N(f) = \sup\{\mu_N(\Gamma_-(f_m)) | f_m := f + x_1^m + \ldots + x_n^m, m \geq 1\}.$$

If f is convenient then

$$\mu_N(f) = \mu_N(\Gamma_-(f)).$$

The following theorem was proved by Kouchnirenko in arbitrary characteristic. We recall that $\mu(f) := \dim \Bbbk[\![\mathbf{x}]\!]/j(f)$ is the *Milnor number* of f.

Theorem 3.1. *[21] For $f \in \Bbbk[\![\mathbf{x}]\!]$ we have $\mu_N(f) \leq \mu(f)$, and if f is NND and convenient then $\mu_N(f) = \mu(f) < \infty$.*

Since Theorem 3.1 does not cover all semi-quasihomogeneous singularities, Wall introduced the condition INND (denoted by NPND* in [29]). Using Theorem 3.1, Wall proved the following theorem for $\Bbbk = \mathbb{C}$ which was extended to arbitrary $\Bbbk$ in [7].

Theorem 3.2. *[29], [7] If $f \in \Bbbk[\![\mathbf{x}]\!]$ is INND, then*

$$\mu(f) = \mu_N(f) = \mu_N(\Gamma_-(f)) < \infty.$$

Kouchnirenko proved that the condition "convenient" is not necessary in Theorem 3.1. In the planar case, the authors in [7] show that Kouchnirenko's result holds in arbitrary characteristic without the assumption that f is convenient (allowing $\mu(f) = \infty$):

Proposition 3.3. *[7, Proposition 4.5] Suppose that $f \in \Bbbk[\![x, y]\!]$ is NND, then $\mu_N(f) = \mu(f)$.*

[†]For $f \in \Bbbk[\![x]\!]$ we call $tj(f) = (f) + j(f)$ the Tjurina ideal of f.

Theorem 3.4. *Let $f \in \mathfrak{m} \subset \Bbbk[\![x,y]\!]$. Then the following are equivalent*

(i) $\mu(f) = \mu_N(f) < \infty$.

(ii) f *is INND.*

Proof.

The theorem follows from the following lemmas (for proofs, see [16]):

Lemma 3.1. *Let $f, g \in \Bbbk[\![x,y]\!]$ be convenient such that $\Gamma_-(f) \subseteq \Gamma_-(g)$. Then*

(i) $\mu_N(f) \leq \mu_N(g)$.

(ii) The equality holds if and only if $\Gamma_-(f) \cap \mathbb{R}^2_{\geq 1} = \Gamma_-(g) \cap \mathbb{R}^2_{\geq 1}$, *where*

$$\mathbb{R}^2_{\geq 1} = \{(x,y) \in \mathbb{R}^2 | x \geq 1, y \geq 1\}.$$

Note that Part (i) of the lemma holds true in many variables by [6, Cor. 5.6]. Let us denote by $\Gamma_1(f)$ the cone joining the origin with $\Gamma(f) \cap \mathbb{R}^2_{\geq 1}$. (cf. Fig. 1).

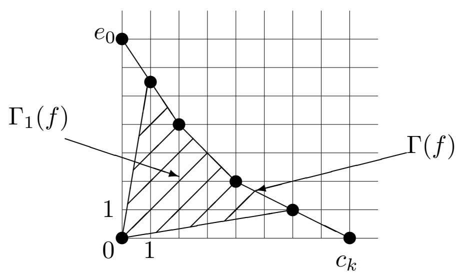


Fig. 1.

Lemma 3.2. *Let $f = \sum c_{ij} x^i y^j \in K[\![x,y]\!]$ be convenient and let $(0,n)$ (resp. $(m,0)$) be the vertex on the y-axis (resp. on the x-axis) of $\Gamma(f)$. Assume that $m = n = 0 \mod p$ then $\mu(f) > \mu_N(f)$.*

Lemma 3.3. *Let $f \in K[\![x,y]\!]$ be convenient. If f is degenerate along some edge or some inner vertex of $\Gamma(f)$ then $\mu(f) > \mu_N(f)$.*

■

Corollary 3.1. *Let $f \in \Bbbk[\![x,y]\!]$ and let $M \in \mathbb{N}$ such that $\Gamma(f) \subset \Gamma(f_M)$ with $f_M := f + x^M + y^M$. Then f is INND if and only if it is INND w.r.t. $\Gamma(f_m)$ for some (equivalently for all) $m > M$.*

3.2. Delta-invariants

We consider now another important invariant of plane curve singularities, the invariant δ and its combinatorial counterpart, the Newton invariant δ_N. We show that both coincide iff f is weighted homogeneous Newton non-degenerate (WHNND), a new non-degenerate condition introduced below.

Let $E_1, \ldots, E_k$ be the edges of the Newton diagram of f. We denote by $l(E_i)$ the lattice length of E_i, i.e. the number of integral points on E_i minus one and by $s(\mathrm{in}_{E_i}(f))$ the number of non-monomial irreducible (reduced) factors of $\mathrm{in}_{E_i}(f)$. We set

(a) If f is convenient, we define

$$\delta_N(f) := V_2(\Gamma_-(f)) - \frac{V_1(\Gamma_-(f))}{2} + \frac{\sum_{i=1}^k l(E_i)}{2},$$

and otherwise we set $\delta_N(f) := \sup\{\delta_N(f_m) | f_m := f + x^m + y^m, m \in \mathbb{N}\}$ and call it the *Newton δ-invariant* of f.

(b) $r_N(f) := \sum_{i=1}^k l(E_i) + \max\{j | x^j \text{ divides } f\} + \max\{l | y^l \text{ divides } f\}$.

(c) $s_N(f) := \sum_{i=1}^k s(\mathrm{in}_{E_i}(f)) + \max\{j | x^j \text{ divides } f\} + \max\{l | y^l \text{ divides } f\}$.

Note that all of these numbers depend on the Newton diagram of f and hence are coordinate-dependent.

Proposition 3.5. *For $0 \neq f \in (x, y)$ we have $r(f) \leq r_N(f)$, and if f is WNND then $r(f) = r_N(f)$.*

Proof.

cf. [7, Lemma 4.10] ■

Let E be an edge of the Newton diagram of f. Then we can write f_E as follows,

$$\mathrm{in}_E(f) = \prod_{i=1}^{s} (a_i x^{m_0} - b_i y^{n_0})^{r_i},$$

where $a_i, b_i \in K^*$, $(a_i : b_i)$ pairwise distinct; $m_0, n_0, r_i \in \mathbb{N}_{>0}$, $\gcd(m_0, n_0) = 1$. It easy to see that

$$s = s(\mathrm{in}_E(f)) \text{ and } l(E) = \sum_{i=1}^{s} r_i.$$

This implies $s(\mathrm{in}_E(f)) \leq l(E)$ and hence $s_N(f) \leq r_N(f)$.

Let $f = f_d^{\mathrm{w}} + f_{d+1}^{\mathrm{w}} + \ldots$ with $f_d^{\mathrm{w}} \neq 0$ be the (n_0, m_0)-weighted homogeneous decomposition of f.

Definition 3.1. We say that f is *weighted homogeneous non-degenerate* (WHND) *along* E if either $r_i = 1$ for all $i = 1, \dots, s$ or $(a_i x^{m_0} - b_i y^{n_0})$ does not divide f^{w}_{d+1} for each $r_i > 1$.

f is called *weighted homogeneous Newton non-degenerate* (WHNND) if its Newton diagram has no edge or if it is WHND along each edge of its Newton diagram.

Remark 3.1. (a) In [22] the author introduced superisolated singularities to study the μ-constant stratum. We recall that $f \in \Bbbk[\![x, y]\!]$ is *superisolated* if it becomes regular after only one blowing up. By ([22, Lemma 1]), this is equivalent to: $f_{m+1}(\beta_i, \alpha_i) \neq 0$ for all tangent directions $(\beta_i : \alpha_i)$ of f with $r_i > 1$, where $f = f_m + f_{m+1} + \dots$ is the homogeneous decomposition of f and

$$f_m = \prod_{i=1}^{s} (\alpha_i x - \beta_i y)^{r_i}.$$

Note that this condition concerns all factors of f_m including monomials. For WHNND singularities we require a similar condition, but for "all weights" and without any condition on the monomial factors of the first term of the weighted homogeneous decomposition of f.

(b) Since a plane curve singularity is superisolated iff it becomes regular after only one blowing up, we have $\delta(f) = m(m-1)/2$ and hence $\delta(f) = \delta_N(f) = m(m-1)/2$, by Proposition 4.1. It follows from Theorem 4.3 that

(c) A superisolated plane curve singularity is WHNND.

(d) The plane curve singularity $x^2 + y^5$ is WHNND but not superisolated.

Proposition 3.6. *With notations as above, f is WND along E if and only if $s(f_E) = l(E)$ or, equivalently, iff $r_i = 1$ for all $i = 1, \dots, s$. In particular, WNND implies WHNND.*

Proof.

cf. [16, Proposition 3.5] ∎

Proposition 3.7. *For $0 \neq f \in (x, y)$ we have $s_N(f) \leq r(f)$ and if f is WHNND then $s_N(f) = r(f)$.*

Proof.

cf. [16, Proposition 3.7] ∎

Proposition 3.8. *For $0 \neq f \in (x, y)$ we have $s_N(f) \leq r(f) \leq r_N(f)$, and both equalities hold if and only if f is WNND.*

Proof.

The inequalities follow from Proposition 3.5 and Proposition 3.7. For each edge E of $\Gamma(f)$, by Proposition 3.6, f is WND along E iff $s(f_E) = l(E)$. This implies that f is WNND if and only if $s_N(f) = r_N(f)$ since $s(f_E) \leq l(E)$ and both sides are additive with respect to edges of $\Gamma(f)$. ■

We investigate now the relations between $\nu(f)$, $\delta_N(f)$ and $\delta(f)$, which were studied in [5] and [7].

Proposition 3.9. *[7, Prop. 4.9] For $0 \neq f \in (x, y)$ we have $\delta_N(f) \leq \delta(f)$, and if f is WNND then $\delta_N(f) = \delta(f)$.*

Hence WNND is sufficient but, by the following example, not necessary for $\delta_N(f) = \delta(f)$.

Example 3.10. Let $f(x, y) = (x + y)^2 + y^3 \in \Bbbk[\![x, y]\!]$. Then f is not WNND but $\delta_N(f) = \delta(f) = 1$. This easy example shows also that WNND depends on the coordinates since $x^2 + y^3$ is WNND. Note that f is WHNND.

Theorem 3.11. *[16, Theorem 3.12] Let $f \in \Bbbk[\![x, y]\!]$ be reduced. Then $\delta(f) = \delta_N(f)$ if and only if f is WHNND.*

Proof.

Recall that, if E is an edge of the Newton diagram of f. Then we can write f_E as follows,

$$\mathrm{in}_E(f) = \prod_{i=1}^{s} (a_i x^{m_0} - b_i y^{n_0})^{r_i},$$

where $a_i, b_i \in K^*$, $(a_i : b_i)$ pairwise distinct; $m_0, n_0, r_i \in \mathbb{N}_{>0}$, $\gcd(m_0, n_0) = 1$. It easy to see that

$$s = s(\mathrm{in}_E(f)) \text{ and } l(E) = \sum_{i=1}^{s} r_i.$$

The theorem is then based on the following lemmas. We refer to [16] for detail proofs.

Lemma 3.4. *There exist an integer n and an infinitely near point $P \xrightarrow{n} 0$ in the n-th neighbourhood of 0, such that*

$$\mathrm{in}_{E_P}(f_P)(u, v) = \textit{monomial} \times \prod_{i=1}^{s} (a_i u - b_i v)^{r_i},$$

where f_P is the local equation of the strict transform of f at P and E_P is some edge of its Newton diagram $\Gamma(f_P)$. Moreover, f is WHND along E if and only if f_P is WHND along E_P.

Let us denote by Q_i the points $(a_i : b_i)$ and by m_{Q_i} the multiplicity of f_{Q_i}. Then

Lemma 3.5. *The following are equivalent*

(i) f *is WHND along* E.

(ii) $m_{Q_i} = 1$ *for all* i.

■

3.3. Milnor formula

We recall that if f is a plane curve singularity then its Milnor number $\mu(f)$ is $\dim \Bbbk[\![x,y]\!]/(f_x, f_y)$, where f_x, f_y be the partials of f. Proposition 2.6 yields that the Milnor number can be computed as the intersection multiplicity of f_x and f_y: $\mu(f) = i(f_x, f_y)$. Moreover if $\Bbbk = \mathbb{C}$, the Milnor's famous formula (see, [23, Thm 10.5], or also [15, Prop. 3.35]) gives a relation between the Milnor number, the δ-invariant:

$$\mu(f) = 2\delta(f) - r(f) + 1.$$

This also holds true in characteristic zero. But in positive characteristic, it is in general not true as the following example shows: $f = x^3 + x^4 + y^6 + y^7 \in \Bbbk[\![x,y]\!]$ with $\mathrm{char}(\Bbbk) = 3$. Then

$$r(f) = 1;\ \mu(f) = 18;\ \delta(f) = 6.$$

In positive characteristic the equality holds under certain conditions of the Newton diagram, e.g. NND ([7, Thm. 9]) or INND ([16, Cor. 3.16]). However without the assumption of Newton non-degeneracy one has at least an inequality as proven by Pierre Deligne [11], see also [24]:

$$\mu(f) \geq 2\delta(f) - r(f) + 1.$$

The difference of the two sides is measured by the so called Swan character, denoted by $\mathrm{Sw}(f)$, which counts wild vanishing cycles that can only occur in positive characteristic.

However it still holds true if f is NND by [7, Thm. 4.13]. Using the general inequality

$$\mu_N(f) = 2\delta_N(f) - r_N(f) + 1 \leq 2\delta(f) - r(f) + 1 \leq \mu(f)$$

from [7], then Theorem 3.4, Proposition 3.5 and Proposition 4.2 imply.

Although we can compute the number of wild vanishing cycles, it seems hard to understand them. In [16] we have posed the following problems.

Problem 3.12. *Is there any "geometric" way to understand the wild vanishing cycles, distinguishing them from the ordinary vanishing cycles counted by* $2\delta - r + 1$*? Is there at least a "reasonable" characterization of those singularities without wild vanishing cycles?*

Problem 3.13. *Find an "elementary proof" for the inequality*

$$\mu(f) \geq 2\delta(f) - r(f) + 1.$$

We will discuss more carefully about this topic in the last two sections.

4. Gamma and kappa invariants

The results in this section are borrowed from [26].

4.1. Gamma invariants

Following [26, Section 2] we introduce and study new (gamma) invariants $\gamma, \tilde{\gamma}$ of plane curve singularities which have not been considered before. In characteristic zero, these invariants coincide and are equal to the Milnor number (see Remark 4.1). So they may be considered as generalizations of the Milnor number in positive characteristic and are believed to be useful in studying classical invariants. In this section we use them to connect the delta and kappa invariant. We will show, in Proposition 4.1, that

$$\kappa(f) \geq \gamma(f) + \mathrm{mt}(f) - 1$$

and in Theorem 4.4, that

$$\gamma(f) \geq 2\delta(f) - r(f) + 1$$

and obtain the inequality in the main result (Theorem 4.5) of the section:

$$\kappa(f) \geq 2\delta(f) + \mathrm{mt}(f) - 1$$

with equalitity if and only if p is m-good for f (see, Definition 4.3 for the notion of m-goodness).

Definition 4.1. Let $f \in \Bbbk[\![x, y]\!]$ be reduced. The number $\tilde{\gamma}_{x,y}(f)$ (or $\tilde{\gamma}(f)$, if the coordinate $\{x, y\}$ is fixed) of f, is defined as follows:

(a) $\tilde{\gamma}(x) := 0$, $\tilde{\gamma}(y) := 0$.

(b) If f is irreducible and *convenient* (i.e. $i_0(f,x), i_0(f,y) < \infty$), then

$$\tilde{\gamma}(f) := \min\{i_0(f,f_x) - i_0(f,y) + 1, i_0(f,f_y) - i_0(f,x) + 1\}.$$

(c) If $f = f_1 \cdot \ldots \cdot f_r$, then

$$\tilde{\gamma}(f) := \sum_{i=1}^{r} \Big(\tilde{\gamma}(f_i) + \sum_{j \neq i} i_0(f_i, f_j)\Big) - r + 1.$$

Definition 4.2. The *gamma invariant* of a reduced plane curve singularity f, denoted by $\gamma(f)$, is the minimum of $\tilde{\gamma}_{X,Y}(f)$ for all coordinates X, Y.

Remark 4.1. (a) In characteristic zero, $\gamma(f) = \tilde{\gamma}(f) = \mu(f)$ due to Theorems 4.3, 4.4 and the Milnor formula.

(b) In general we have, by definition, that $\gamma(f) \leq \tilde{\gamma}(f)$ (with equality if p is im-good for f, see Definition 4.3 and Corollary 4.1) and that $\gamma(f) = \tilde{\gamma}(g)$ for some g right equivalent to f (f is called *right equivalent* to g, denoted by $f \sim_r g$, if there is an automorphism $\Phi \in \mathrm{Aut}_{\Bbbk}(\Bbbk[[x,y]])$ such that $f = \Phi(g)$).

(c) The number $\tilde{\gamma}$ depends on the choice of coordinates, i.e. it is not invariant under right equivalence. E.g. $f = x^3 + x^4 + y^5$ and $g = (x+y)^3 + (x+y)^4 + y^5$ in $\Bbbk[[x,y]]$ with $\mathrm{char}(\Bbbk) = 3$ and then $f \sim_r g$, but $\tilde{\gamma}(f) = 8$, $\tilde{\gamma}(g) = 10$. However, as we will see in Proposition 4.2, if the characteristic p is multiplicity good for f then $\tilde{\gamma}(f) = \tilde{\gamma}(g)$ for all g contact equivalent to f. Recall that f, g are *contact equivalent* if there is an automorphism $\Phi \in \mathrm{Aut}_{\Bbbk}(\Bbbk[[x,y]])$ and a unit $u \in \Bbbk[[x,y]]$ such that $f = u \cdot \Phi(g)$, and we denote this by $f \sim_c g$.

(d) It follows from the definition that $\tilde{\gamma}(u) = 1$ and $\tilde{\gamma}(u \cdot f) = \tilde{\gamma}(f)$ for every unit u and therefore γ is invariant under contact equivalence.

(e) The Milnor number μ is invariant under right equivalence. The numbers $\delta, \kappa, \mathrm{mt}, r, i$ are invariant under contact equivalence (see, for instance [25], Prop. 1.2.19 for the invariance of δ). This means that, if $f \sim_c g$ then

$$\delta(f) = \delta(g),\ \kappa(f) = \kappa(g),\ \mathrm{mt}(f) = \mathrm{mt}(g) \text{ and } r(f) = r(g).$$

Moreover, for any $\Phi \in \mathrm{Aut}_{\Bbbk}(\Bbbk[[x,y]])$ and units u, v, one has

$$i_0(f,h) = i_0\left(u \cdot \Phi(f), v \cdot \Phi(h)\right).$$

Before studying in detail gamma invariants, we collect several facts on invariants of plane curve singularities which we use later. For proofs, we refer to [15] and [25].

Remark 4.2. (a) If f is irreducible, then

$$\kappa(f) = \min\{i_0(f, f_x), i_0(f, f_y)\}.$$

Indeed, taking a parametrization $(x(t), y(t))$ of f we obtain that

$$\kappa(f) = \text{ord}\ \left(\alpha f_x(x(t), y(t)) + \beta f_y(x(t), y(t))\right),$$

which equals to the minimum of $i(f, f_x)$ and $i(f, f_y)$ since $(\alpha : \beta)$ is generic.

(b) If f is convenient, then

$$\tilde{\gamma}(f) \quad = \quad i_0(f, \alpha x f_x + \beta y f_y) - i_0(f, x) - i_0(f, y) + 1,$$

where $(\alpha : \beta) \in \mathbb{P}^1$ is generic.

Definition 4.3. Let $\text{char}(\Bbbk) = p \geq 0$ and let $f = f_1 \cdot \ldots \cdot f_r \in \Bbbk[\![x, y]\!]$ be reduced with f_i irreducible. The characteristic p is said to be

(a) *multiplicity good* (*m-good*) for f if the multiplicities $\text{mt}(f_i) \neq 0 \pmod p$ for all $i = 1, \ldots, r$;

(b) *intersection multiplicity good* (*im-good*) for f if for all $i = 1, \ldots, r$, either $i(f_i, x) \neq 0 \pmod p$ or $i(f_i, y) \neq 0 \pmod p$;

(c) *right intersection multiplicity good* (*right im-good*) for f if it is im-good for f after some change of coordinate. That is, it is im-good for some g right equivalent to f.

Note that these notions are trivial in characteristic zero, i.e. if $p = 0$ then it is always m-good, im-good and right im-good for f. In general we have

$$\text{“m-good”} \Longrightarrow \text{“im-good”} \Longrightarrow \text{“right im-good”}.$$

The following proposition gives us the first relations between the gamma invariants and classical invariants.

Proposition 4.1. *Let $f \in \Bbbk[\![x, y]\!]$ be reduced. Then*

$$\gamma(f) \leq \tilde{\gamma}(f) \leq \kappa(f) - \text{mt}(f) + 1$$

with equality if p is m-good for f.

Proof.

cf. [26, Proposition 2.6]. ■

The following proposition says that the number $\tilde{\gamma}$ is invariant under contact equivalence in the class of singularities for which p is m-good. It will be shown in Corollary 4.1 that $\tilde{\gamma}$ is invariant under contact equivalence in the class of singularities for which p is im-good.

Proposition 4.2. *Let $f \in \Bbbk[\![x,y]\!]$ be reduced such that p is m-good for f and let $g \sim_c f$. Then $\tilde{\gamma}(g) = \tilde{\gamma}(f)$. In particular, $\gamma(f) = \tilde{\gamma}(f)$.*

Proof.

This follows from Proposition 4.1 and Remark 4.1(e). See [25, Lemma 2.3.4] for a direct proof. ∎

Theorem 4.3. *Let $f \in \Bbbk[\![x,y]\!]$ be reduced. Then*

$$\tilde{\gamma}(f) \geq 2\delta(f) - r(f) + 1.$$

Equality holds if and only if the characteristic p is im-good for f.

Proof.

cf. [26, Theorem 2.11]. ∎

Corollary 4.1. *Assume that p is im-good for f. Then*

$$\gamma(f) = \tilde{\gamma}(f).$$

The following simple corollary should be useful in computation, since the number in the left side is easily computed.

Corollary 4.2. *Assume that $p > \mathrm{mt}(f)$. Then*

$$\mu(f) - \tilde{\gamma}(f) = \mathrm{Sw}(f).$$

Theorem 4.4. *Let $f \in \Bbbk[\![x,y]\!]$ be reduced. Then*

$$\gamma(f) \geq 2\delta(f) - r(f) + 1.$$

Equality holds if and only if the characteristic p is right im-good for f.

Proof.

Taking g right equivalent to f such that $\gamma(f) = \tilde{\gamma}(g)$ and combining Theorem 4.3 and Remark 4.1 we get

$$\gamma(f) = \tilde{\gamma}(g) \geq 2\delta(g) - r(g) + 1 = 2\delta(f) - r(f) + 1$$

with equality if and only if p is im-good for g. It remains to show that if p is right im-good for f, then

$$\gamma(f) = 2\delta(f) - r(f) + 1.$$

Indeed, by definition, p is im-good for some h right equivalent to f. Again combining Theorem 4.3 and Remark 4.1 we get

$$\gamma(f) = \gamma(h) \leq \tilde{\gamma}(h) = 2\delta(h) - r(h) + 1 = 2\delta(f) - r(f) + 1 \leq \gamma(f).$$

This implies that

$$\gamma(f) = 2\delta(f) - r(f) + 1,$$

which completes the theorem. ■

4.2. Kappa invariants and Plücker formulas

We prove in this section the main result (Theorem 4.5) and apply it to Plücker formulas (Corollaries 4.4, 4.5). Furthermore we show, in Corollary 4.3 (resp. Corollary 4.5), that if p is "big" for f (resp. for a plane curve C), then f (resp. C) has no wild vanishing cycle.

Theorem 4.5. *Let $f \in \Bbbk[[x, y]]$ be reduced. One has*

$$\kappa(f) \geq 2\delta(f) + \mathrm{mt}(f) - r(f)$$

with equality if and only if p is m-good for f.

The following interesting corollary says that if the characteristic p is "big" for f, then f has no wild vanishing cycle.

Corollary 4.3. *Assume that $p > \kappa(f)$. Then f has no wild vanishing cycle, i.e.* $\mathrm{Sw}(f) = 0$*. Moreover one has*

$$\begin{aligned} \kappa(f) &= 2\delta(f) + \mathrm{mt}(f) - r(f) \\ &= \mu(f) + \mathrm{mt}(f) - 1. \end{aligned}$$

Let C be a irreducible curve of degree d in $\mathbb{P}^2$ defined by a homogeneous polynomial $F \in \Bbbk[x, y, z]$. Let $\mathrm{Sing}(C)$ resp. $C^* := C \setminus \mathrm{Sing}(C)$ the singular resp. smooth locus of C, and let $s(C) := \sharp\mathrm{Sing}(C)$ the number of singular points. Let $\rho\colon C^* \to \check{\mathbb{P}}^2, P = (x\colon y\colon z) \mapsto (F_x(P)\colon F_y(P)\colon F_z(P))$ the dual (Gauss) map and $\deg(\rho)$ its degree. We call the closure of the image of ρ in $\check{\mathbb{P}}^2$ the *dual curve* of C denoted by $\check{C}$. We denote by $\check{d}$ the degree of $\check{C}$. For each singular point $P \in \mathrm{Sing}(C)$ take a local equation $f_P = 0$ of C at P, and define

$$\begin{aligned} \delta(C) &:= \sum \delta(f_P), & \mathrm{mt}(C) &:= \sum \mathrm{mt}(f_P), \\ \mu(C) &:= \sum \mu(f_P), & r(C) &:= \sum r(f_P), \\ \mathrm{Sw}(C) &:= \sum \mathrm{Sw}(f_P). \end{aligned}$$

where all the sums are taken over $P \in \mathrm{Sing}(C)$.

Corollary 4.4. *Using the above notions, we have*

$$\begin{aligned}\deg(\rho)\cdot\check{d} &\leq d(d-1)-2\delta(C)+r(C)-\mathrm{mt}(C)\\ &= d(d-1)-\mu(C)-\mathrm{mt}(C)+s(C)+\mathrm{Sw}(C),\end{aligned}$$

with equality if and only if p is multiplicity good (m-good) *for C, i.e. p is m-good for all the f_P.*

Combining Corollaries 4.3 and 4.3 we obtain

Corollary 4.5. *With the above notions, assume that*

$$\max_{P\in\mathrm{Sing}(C)}\{\kappa(f_P)\}<p,$$

(for example, $d(d-1)<p$). Then C has no wild vanishing cycle, i.e. $\mathrm{Sw}(C)=0$. *Moreover one has*

$$\begin{aligned}\deg(\rho)\cdot\check{d} &= d(d-1)-2\delta(C)+r(C)-\mathrm{mt}(C)\\ &= d(d-1)-\mu(C)-\mathrm{mt}(C)+s(C).\end{aligned}$$

5. Semigroup of a plane algebroid branch

In this section we study the semigroup of a given irreducible plane curve singularity and apply it to study Problem 3.12 proposed in Section 3. The proofs can be found in [13, 14].

5.1. Semigroups

We say that a subset G of $\mathbb{N}$ is a *semigroup* if it contains 0 and if it is closed under addition. Let $n>0$ be an integer. A sequence of positive integers $(v_0,\dots,v_h)$ is said to be a *Seidenberg n-characteristic sequence* or *n-characteristic sequence* if $v_0=n$ and it satisfies the following two axioms

(a) Set $d_i=\gcd(v_0,\dots,v_i)$ for $0\leq i\leq h$ and $n_i=\frac{d_{i-1}}{d_i}$ for $1\leq i\leq h$. Then $d_h=1$ and $n_i>1$ for $1\leq i\leq h$.

(b) $n_{i-1}v_{i-1}<v_i$ for $2\leq i\leq h$.

Note that condition (b) implies that the sequence $(v_1,\dots,v_h)$ is strictly increasing. If $n>1$ then $h\geq 1$. If $h=1$ then the sequence (v_0,v_1) is a Seidenberg

n-characteristic sequence if and only if $v_0 = n$ and $\gcd(v_0, v_1) = 1$. There is exactly one 1-sequence which is (1).

Let G be a nonzero semigroup and let $n \in G$, $n > 0$. Then there exists (cf. [17], Chapter 6, Proposition 6.1) a unique sequence $v_0, \dots, v_h$ such that $v_0 = n$, $v_k = \min(G \backslash v_0\mathbb{N} + \cdots + v_{k-1}\mathbb{N})$ for $k \in \{1, \dots, h\}$ and $G = v_0\mathbb{N} + \cdots + v_h\mathbb{N}$. We call the sequence $(v_0, \dots, v_h)$ the *n-minimal system of generators of G*. If $n = \min(G\backslash\{0\})$ then we say that $(v_0, \dots, v_h)$ is the *minimal set* of generators of G. We will study semigroups generated by n-characteristic sequences.

Proposition 5.1. *Let $G = v_0\mathbb{N}+\cdots+v_h\mathbb{N}$ where $(v_0, \dots, v_h)$ is an n-characteristic sequence. Then*

(i) The sequence $(v_0, \dots, v_h)$ is the n-minimal system of generators of G.

(ii) $\min(G\backslash\{0\}) = \min(v_0, v_1)$.

(iii) The minimal system of generators of G is $(v_0, v_1, \dots, v_h)$ if $v_0 < v_1$, $(v_1, v_0, \dots, v_h)$ if $v_1 < v_0$ and $v_0 \not\equiv 0$ (mod v_1) and $(v_1, v_2, \dots, v_h)$ if $v_0 \equiv 0$ (mod v_1). Moreover, the minimal system of generators of G is a $\min(G\backslash\{0\})$-characteristic sequence.

(iv) Let $c = \sum_{k=1}^{h}(n_k - 1)v_k - v_0 + 1$. Then for every $a, b \in \mathbb{Z}$: if $a + b = c - 1$ then exactly one element of the pair (a, b) belongs to G. Consequently c is the smallest element of G such that all integers bigger than or equal to it are in G.

(v) c is an even number and $\sharp(\mathbb{N}\backslash G) = \dfrac{c}{2}$.

The number c is called the conductor *of the semigroup G.*

5.2. Polar factorization theorems

The aim of this section is to study the structure of the semigroup associated with a plane branch and its relation to the factorization theorems.

Let $f = f(x, y) \in \mathbb{k}[[x, y]]$ be an irreducible power series and let $S(f)$ be the semigroup associated with the branch $\{f = 0\}$. Suppose that $\{f = 0\} \neq \{x = 0\}$ and put $n = i_0(f, x)$. That is,

$$S(f) = \{i(f, g) \mid g \in \mathbb{k}[[x, y]] \setminus (f)\mathbb{k}[[x, y]]\}.$$

Let $(\bar{b}_0, \dots, \bar{b}_h)$, $\bar{b}_0 = n$ be the n-minimal system of generators of $S(f)$. We define

$$e_0 = n, e_k = \gcd(e_{k-1}, \bar{b}_k) \text{ and } n_k = \frac{e_{k-1}}{e_k} \text{ for } k \in \{1, \dots, h\}.$$

Lemma 5.1. *We have $e_h = 1$.*

Proof.

It follows from Theorem 2.1 that $\pi\colon \Bbbk[\![x,y]\!]/(f) \to \Bbbk[\![t]\!]$ is the normalization, and hence

$$Q\left(\Bbbk[\![x,y]\!]/(f)\right) \cong \Bbbk(\!(t)\!).$$

This implies that, there exist $p, q \in \Bbbk[\![x,y]\!]/(f)$ such that

$$\frac{p\left(x(t), y(t)\right)}{q\left(x(t), y(t)\right)} = t,$$

where $x(t) = \pi(x)$ and $y(t) = \pi(y)$. Taking order of both sides we get

$$\operatorname{ord}\, p\left(x(t), y(t)\right) - \operatorname{ord}\, q\left(x(t), y(t)\right) = 1.$$

Since $\operatorname{ord}\, p\left(x(t), y(t)\right)$ and $\operatorname{ord}\, q\left(x(t), y(t)\right)$ are elements in $S(f)$, it follows that $\gcd(S(f)) = 1$ and hence $e_h = 1$. ∎

Corollary 5.1 (Conductor formula)**.** *One has*

$$c(f) = \sum_{k=1}^{h}(n_k - 1)\bar{b}_k - \bar{b}_0 + 1.$$

Theorem 5.2 (Semigroup Theorem)**.** *Let $\{f = 0\}$ be a branch such that $\{f = 0\} \neq \{x = 0\}$. Set $n = i_0(f, x)$ and let $\bar{b}_0, \ldots, \bar{b}_h$ be the n-minimal system of generators of the semigroup $S(f)$. There exists a sequence of monic polynomials $f_0, f_1, \ldots, f_{h-1} \in \Bbbk[\![x]\!][y]$ such that for $k \in \{1, \ldots, h\}$:*

- (a_k) $\deg_y(f_{k-1}) = \frac{n}{e_{k-1}}$,
- (b_k) $i_0(f, f_{k-1}) = \bar{b}_k$ *for* $k \in \{1, \ldots, h\}$,
- (c_k) *if* $k > 1$ *then* $n_{k-1}\bar{b}_{k-1} < \bar{b}_k$.

Moreover $n_k > 1$ for all $k \in \{1, \ldots, h\}$.

Theorem 5.3 (Merle-Granja's Factorization Theorem)**.** *Let $\{f = 0\}$ be a branch such that $\{f = 0\} \neq \{x = 0\}$. Set $n = i_0(f, x)$ and let $\bar{b}_0, \ldots, \bar{b}_h$ be the n-minimal system of generators of the semigroup $S(f)$. Fix k, $1 \leq k \leq h$. Let $g = g(x, y) \in \Bbbk[\![x,y]\!]$ be a power series such that*

- *(i)* $i_0(g, x) = \frac{n}{e_k} - 1$,
- *(ii)* $i_0(f, g) = \sum_{i=1}^{k}(n_i - 1)\bar{b}_i$.

Then there is a factorization $g = g_1 \cdots g_k \in \Bbbk[\![x,y]\!]$ *such that*

1. $i_0(g_i, x) = \frac{n}{e_i} - \frac{n}{e_{i-1}}$ *for* $i \in \{1, \dots, k\}$,
2. *if* $\phi \in \Bbbk[\![x,y]\!]$ *is an irreducible factor of* g_i, $i \in \{1, \dots, k\}$ *then*
   (a) $\frac{i_0(f,\phi)}{i_0(\phi,x)} = \frac{e_{i-1}\bar{b}_i}{n}$,
   (b) $i_0(\phi, x) \equiv 0 \mod \frac{n}{e_{i-1}}$.

Theorem 5.4 (Merle's factorization theorem)**.** *Let* $\{f = 0\}$ *be a branch such that* $\{f = 0\} \neq \{x = 0\}$. *Set* $n = i_0(f, x)$ *and let* $\bar{b}_0, \dots, \bar{b}_h$ *be the* n*-minimal system of generators of the semigroup* $S(f)$. *Suppose that* $n > 1$ *and* $n \not\equiv 0 \mod (\operatorname{char} \Bbbk)$. *Then* $\frac{\partial f}{\partial y} = g_1 \cdots g_h$ *in* $\Bbbk[\![x,y]\!]$, *where*

(i) $i_0(g_i, x) = \frac{n}{e_i} - \frac{n}{e_{i-1}}$ *for* $i \in \{1, \dots, h\}$.

(ii) *If* $\phi \in \Bbbk[\![x,y]\!]$ *is an irreducible factor of* g_i, $i \in \{1, \dots, h\}$, *then*

$$\frac{i_0(f,\phi)}{i_0(\phi,x)} = \frac{e_{i-1}\bar{b}_i}{n} \text{ and } i_0(\phi, x) \equiv 0 \mod \frac{n}{e_{i-1}}.$$

Proof. Since $n \not\equiv 0$ (mod char $\Bbbk$) we have $i_0\left(\frac{\partial f}{\partial y}, x\right) = n-1$. By the Dedekind formula and the Conductor formula we have $i_0\left(f, \frac{\partial f}{\partial y}\right) = c(f) + n - 1 = \sum_{k=1}^{h}(n_k - 1)\bar{b}_k$. The theorem is then proved by applying Theorem 5.3 to the series $g = \frac{\partial f}{\partial y}$. ■

5.3. García Barroso-Ploski's theorem

In this section, we give a proof of the following theorem.

Theorem 5.5. *[14, Theorem 1.1] Let* $f \in \Bbbk[\![x,y]\!]$ *be an irreducible singularity and let* $(\bar{b}_0, \dots, \bar{b}_h)$ *be the minimal system of generators of* $S(f)$. *Suppose that* $p = \operatorname{char} \Bbbk > \operatorname{ord} f$. *Then the following two conditions are equivalent:*

(i) $\bar{b}_k \not\equiv 0 \mod p$, *for* $k \in \{1, \dots, h\}$;

(ii) $\mu(f) = c(f)$.

Proof. Since $p > \operatorname{ord} f$, it follows from Dedekind's formula (Lemma 2.2) that

$$i_0(f, \frac{\partial f}{\partial y}) = c(f) + \operatorname{ord} f - 1.$$

It remains to prove that

$$i_0(f, \frac{\partial f}{\partial y}) = \mu(f) + \mathrm{ord} f - 1$$

if and only if $\bar{b}_k \not\equiv 0 \pmod{p}$ for $k \in \{1, \ldots, h\}$.

In fact, let ϕ be an irreducible factor of $\frac{\partial f}{\partial y}$ and let $(x(t), y(t))$ be a parametrization of $\phi = 0$. Then

$$\mathrm{ord} x(t) = i_0(x, \phi) = \mathrm{ord}\phi < \mathrm{ord} f < p$$

and, consequently, ord $x(t) \not\equiv 0 (\mathrm{mod} p)$, which implies ord $x'(t) = \mathrm{ord}\, x(t) - 1$. We have

$$\frac{d}{dt} f(x(t), y(t)) = \frac{\partial f}{\partial x}(x(t), y(t)) x'(t).$$

This yields

$$\mathrm{ord}\ f(x(t), y(t)) - 1 \le \mathrm{ord}\ \frac{\partial f}{\partial x}(x(t), y(t)) + \mathrm{ord}\, x(t) - 1$$

with equality if and only if $i_0(f, \phi) = \mathrm{ord} f(x(t), y(t)) \not\equiv 0 \mod p$. Taking the sum over all irreducible factors $\frac{\partial f}{\partial y}$ gives us

$$i_0(f, \frac{\partial f}{\partial y}) \le i_0(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) + \mathrm{ord} f - 1 = \mu(f) + \mathrm{ord} f - 1 \tag{5.1}$$

with equality if and only if $i_0(f, \phi) \not\equiv 0 \mod p$ for all ϕ.

Let $\frac{\partial f}{\partial y} = g_1 \cdots g_h$ be the Merle factorization of the polar $\frac{\partial f}{\partial y}$ and assume that ϕ is an irreducible factor of g_k. Then by Theorem 5.4, we can write ord $\phi = m_k \frac{n}{e_{k-1}}$, where $m_k \geqslant 1$ is an integer. Since ord $\phi < p$, it yields that $m_k < p$ and therefore $m_k \not\equiv 0 \mod p$. Again, by Theorem 5.4

$$i_0(f, \phi) = \frac{e_{k-1} \bar{b}_k}{n} \mathrm{ord}\ \phi = m_k \bar{b}_k.$$

Therefore, $i(f, \phi) \not\equiv 0 \pmod{p}$ if and only if $\bar{b}_k \not\equiv 0 \pmod{p}$. The theorem hence follows from (5.1). ■

Remark 5.1. *If $p <$ ord f, then the proof of Theorem 5.5 fails, even if* ord $f \not\equiv 0 \pmod{p}$. *Take $f = x^{p+2} + y^{p+1} + x^{p+1} y$.*

Conjecture 5.6. *Let $f \in \Bbbk[\![x, y]\!]$ be an irreducible singularity and let $\bar{b}_0, \ldots, \bar{b}_g$ be the minimal system of generators of $S(f)$. Then the following two conditions are equivalent:*

(i) $\bar{b}_k \not\equiv 0 \mod p$, *for $k \in \{1, \ldots, g\}$;*

(ii) $\mu(f) = c(f)$.

The conjecture is true if $S(f)$ is generated by $\bar{b}_0$ and $\bar{b}_1$. Recently, Hefez, Rodrigues and Salomao in [18] have proved that (i) implies (ii).

Theorem 5.7 (Hefez, Rodrigues and Salomao)**.** *Let* $f \in \mathbb{k}[[x, y]]$ *be an irreducible singularity and let* $\bar{b}_0, \ldots, \bar{b}_g$ *be the minimal system of generators of* $S(f)$*. If* $\bar{b}_k \not\equiv 0 \mod p$*, for* $k \in \{1, \ldots, g\}$ *then* $\mu(f) = c(f)$*.*

Acknowledgement

This article has been written during many visits of the author to Vietnam Institute for Advanced Studies in Mathematics. The author thanks sincerely the institute for their hospitality and valuable supports.

References

[1] **V. I. Arnol'd**, *Normal forms for functions near degenerate critical points, the Weyl groups of* A_k, D_k, E_k *and Lagrangian singularities,* Funct. Anal. Appl. 6 (1972), 254–272.

[2] **V. I. Arnol'd**, *Classification of unimodal critical points of functions,* Funct. Anal. Appl. 7 (1973), 230–231.

[3] **V. I. Arnol'd**, *Local normal form of functions,* Invent. Math. 35 (1976), 87–109.

[4] **M. Artin**, *On isolated rational singularities of surfaces,* Amer. J. Math. 88 (1966), 129–136.

[5] **P. Beelen; R. Pellikaan**, *The Newton polygon of plane curves with many rational points,* Designs, Codes and Cryptography 21(2000), 41-67.

[6] **C. Bivià-Ausina**, *Local Łojasiewicz exponents, Milnor numbers and mixed multiplicities of ideals,* Math. Z. 262 (2009), no. 2, 389–409.

[7] **Y. Boubakri, G.-M. Greuel, and T. Markwig**, *Invariants of hypersurface singularities in positive characteristic,* Rev. Mat. Complut. 25(2012), no. 1, 61–85.

[8] **E. Brieskorn; H. Knörrer**, *Plane Algebraic Curves,* Birkhaeuser (1986), 721 pages.

[9] **A. Campillo**, *Algebroid Curves in Positive Characteristic,* SLN 813, Springer-Verlag (1980).

[10] **P. Cassou-Noguès, A. Ploski**, *Invariants of plane curve singularities and Newton diagrams,* Univ. Iagel. Acta Math. 49 (2011), 9–34.

[11] **P. Deligne**, *La formule de Milnor, Sém. Géom. Algébrique du Bois-Marie,* 1967-1969, SGA 7 II, Lecture Notes in Math. 340, Expose XVI, (1973), 197–211.

[12] **F. Delgado**, *The semigroup of values of a curve singularity with several branches,* Manuscr. math. **59** (1987), 347–374.

[13] **E. García Barroso and A. Ploski**, *An approach to plane algebroid branches,* Rev. Mat. Complut. **28** (2015), no. 1, 227–252.

[14] **E. García Barroso and A. Ploski**, *The Milnor number of plane irreducible singularities in positive characteristic.* Bulletin of the London Mathematical Society (2016) 48 (1): 94-98.

[15] **G.-M. Greuel, C. Lossen and E. Shustin**, Introduction to Singularities and deformations, Math. Monographs, Springer-Verlag (2006).

[16] **G.-M. Greuel, H. D. Nguyen**, *Some remarks on the planar Kouchnirenko's theorem,* Rev. Mat. Complut. 25 (2012), no. 2, 557–579.

[17] **A. Hefez**, Irreducible Plane Curve Singularities. Real and Complex Singularities. Lecture Notes in Pure and Appl. Math., vol. 232, pp. 1–120. Dekker, New York (2003).

[18] **A. Hefez,J. H. O. Rodrigues, R. Salomão**, *The Milnor number of plane branches with tame semigroups of values*, Bull. Braz. Math. Soc. (N.S.) **49** (2018), no. 4, 789–809.

[19] **J. Herzog, E. Kunz**, *Die Wertehalbgruppe eines lokalen Rings der Dimension 1,* Springer (1971).

[20] **H. Hironaka**, *Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II,* Ann. of Math. (2) 79 (1964), 109–203; 79 (1964), 205–326.

[21] **A. G. Kouchnirenko**, *Polyèdres de Newton et nombres de Milnor,* Invent. Math. 32 (1976), 1-31.

[22] **I. Luengo**, *The μ-constant stratum is not smooth,* Invent. Math., 90 (1987), 139–152.

[23] **J. Milnor**, *Singular points of complex hypersurfaces,* Princeton Univ. Press (1968).

[24] **A. Melle-Hernández; C. T. C. Wall**, *Pencils of curves on smooth surfaces*, Proc. Lond. Math. Soc., III. Ser. 83 (2001), no. 2, 257–278.

[25] **H. D. Nguyen**, *Classification of singularities in positive characteristic,* Ph.D. thesis, TU Kaiserslautern (2013). `http://www.dr.hut-verlag.de/9783843911030.html`

[26] **H. D. Nguyen**, *Invariants of plane curve singularities and Plücker formulas in positive characteristic,* Ann. Inst. Fourier (Grenoble) 66(2016), no.5, 2047–2066.

[27] **A. Seidenberg**, *Elements of the Theory of Algebraic Curves.* Addison-Wesley Publishing Co., Reading, Mass.- London-Don Mills, Ont. (1968)

[28] **R.J. Walker**, *Algebraic Curves,* Dover Publ., New York, 1962.

[29] **C. T. C. Wall**, *Newton polytopes and non-degeneracy,* J. reine angew. Math. 509 (1999), 1–19.

[30] **O. Zariski**, *Studies in equisingularity . I. Equivalent singularities of plane algebroid curve,* Amer. J. Math. 87 (1960), 507–536.

[31] **Zariski, O.; Samuel, P.** *Commutative Algebra,* Vol. I, II, Springer (1960).

Nguyen Hong Duc
Thang Long Institute of Mathematics and Applied Sciences (TIMAS),
Thang Long University
Hanoi
Vietnam
duc.nh@thanglong.edu.vn

J. Math. Math. Sci. (2025) 31–54

On meromorphic solution of linear difference - differential equation via partially shared values of meromorphic functions and their growth

Ha Tran Phuong (Thai Nguyen, Vietnam)

Nguyen Van Thin* (Thai Nguyen, Vietnam)

(Received Jan. 5, 2025; accepted Mar. 28, 2025)

Abstract. In this paper, we investigate shared value problems related to a meromorphic function of hyper order less than one and its linear difference-differential polynomial. In general, under certain conditions of sharing values of the meromorphic functions and their difference-differential polynomial, a given meromorphic function must satisfy a difference-differential equation. Furthermore, we also study the order of meromorphic solutions of some classes of difference-differential equations.

1. Introduction

We use standard notations from Nevanlinna theory. Denote by $\sigma(f)$ the order of growth of a meromorphic function f on the complex plane $\mathbb{C}$, and also use the notation $\varsigma(f)$ to denote the hyper order of f,

$$\sigma(f) = \limsup_{r\to\infty} \frac{\log T(r,f)}{\log r}, \ \ \varsigma(f) = \limsup_{r\to\infty} \frac{\log\log T(r,f)}{\log r},$$

Key words and phrases: Meromorphic function, Nevanlinna theory.
2020 Mathematics Subject Classification: 30D35
Corresponding author*: Nguyen Van Thin
The research results are supported by the Ministry of Education and Training of Vietnam under the project with the name "The Nevanlinna - Cartan's second main theorem and some applications" and grant number B2024-TNA-19.

respectively, where $T(r,f)$ is the characteristic function of f.
A meromorphic function a is said to be small with respect to f if $T(r,a) = o(T(r,f))$, as $r \to +\infty$ possibly outside a set of finite Lebesgue measure. We denote $\mathcal{S}(f)$ by the set of small functions with respect to f and $\widehat{\mathcal{S}}(f) = \mathcal{S}(f) \cup \{\infty\}$. Let a be a small function with respect to f. The defect $\delta(f,a)$ of f at a is defined by

$$\delta(a,f) = 1 - \limsup_{r\to\infty} \frac{N(r, \frac{1}{f-a})}{T(r,f)},\ \Theta(a,f) = 1 - \limsup_{r\to\infty} \frac{\overline{N}(r, \frac{1}{f-a})}{T(r,f)}.$$

We can define another defect as follows:

$$\Theta(\infty,f) = 1 - \limsup_{r\to\infty} \frac{\overline{N}(r,f)}{T(r,f)},\ \delta(\infty,f) = 1 - \limsup_{r\to\infty} \frac{N(r,f)}{T(r,f)}.$$

The five-point theorem due to Nevanlinna states that if two non-constant meromorphic functions f and g in $\mathbb{C}$ share five distinct values ignoring multiplicities (IM), then $f \equiv g$. Recently, Halburd, Korhonen, and Tohge [7, 8, 10], Chiang and Feng [3] extended the Nevanlinna theory for difference operator. Difference Nevanlinna theory has emerged as a result of recent interest on value distribution and growth of meromorphic solutions of difference equations [3, 9], also uniqueness of meromorphic functions with difference polynomials.

Definition 1.1. *[15] Let l be a non-negative integer or infinite. Denote by $E_l(a,f)$ the set of all a-points of f where an a-point of multiplicity m is counted m times if $m \le l$ and $l+1$ times if $m > l$. If $E_l(a,f) = E_l(a,g)$, we say that f and g share (a,l). It is easy to see that if f and g share (a,l), then f and g share (a,p) for $0 \le p \le l$. Also we note that f and g share the value a - IM or CM if and only if f and g share $(a,0)$ or (a,∞), respectively.*

Let p be a positive integer and $a \in \mathbb{C} \cup \{\infty\}$. We use $N_{p)}(r, \frac{1}{f-a})$ to denote the counting function of the zeros of $f-a$, whose multiplicities are not greater than p, $N_{(p+1}(r, \frac{1}{f-a})$ to denote the counting function of the zeros of $f-a$ whose multiplicities are not less than $p+1$, and we use $\overline{N}_{p)}(r, \frac{1}{f-a})$ and $\overline{N}_{(p+1}(r, \frac{1}{f-a})$ to denote their corresponding reduced counting functions (ignoring multiplicities) respectively. We use $\overline{E}_{p)}(a,f)$ ($\overline{E}_{(p+1}(a,f)$) to denote the set of zeros of $f-a$ with multiplicities $\le p$ ($\ge p+1$) (ignoring multiplicity), respectively. We also denote $N_p(r, \frac{1}{f-a})$ by

$$N_p(r, \frac{1}{f-a}) = \overline{N}(r, \frac{1}{f-a}) + \cdots + \overline{N}_{(p}(r, \frac{1}{f-a}).$$

Then we define the truncated deficiency as

$$\delta_p(a, f) = 1 - \limsup_{r\to\infty} \frac{N_p(r, \dfrac{1}{f-a})}{T(r, f)}.$$

Let f be a nonconstant meromorphic function with hyper-order less than 1, we denote $L(f)$ by

$$L(f) := \sum_{j=1}^{k} a_j f(z + c_j),$$

where $a_j \neq 0, j = 1, \ldots, k$, $c_j \in \mathbb{C}$ $(j = 1, \ldots, k)$ are distinct complex numbers.

In 2015, Li, Korhonen and Yang [13] proved some results uniqueness for entire function f and its linear difference polynomial $L(f)$ which share partially values, and under some conditions about defect values. In 2020, X. Qi and L. Yang [18] investigated the uniqueness problem for derivative of transcendental entire function of finite order f and $f(z + c)$ share 0-CM and a-IM, where a is a nonzero complex. In 2022, S. Chen and A. Xu [2] extended the results of Qi-Yang [18] as follows: Let f be a non-constant meromorphic function of hyper order $\varsigma(f) < 1$, c be a non-zero finite complex number, and k be a positive integer. If $f^{(k)}(z)$ and $f(z + c)$ share $0, \infty$-CM and $1 - IM$, then $f^{(k)}(z) \equiv f(z + c)$. Motivate by the results of Li, Korhonen and Yang [13], in this paper, we first prove a result for uniqueness of meromorphic function and its linear difference-differential polynomial $(L(f))^{(n)}$ as follows.

Theorem 1.1. *Let k, n be positive integer numbers. Let $f(z)$ be a non-constant meromorphic function with hyper order less than* 1, *and assume that $(L(f))^{(n)}$ is not a constant function. Suppose that $f - 1$ and $(L(f))^{(n)} - 1$ share value $(0, l)$, f and $(L(f))^{(n)}$ share $\infty - IM$ and*

$$\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, (L(f))^{(n)}) \ (i \geq 2).$$

Then

$$(L(f))^{(n)} \equiv f \tag{1.1}$$

if one of the following assumptions holds:

(1) $l = 0$ *(i.e. $f - 1$ and $(L(f))^{(n)} - 1$ share the value* 0 *IM) and*

$$2\delta_2(0, f) + 3\Theta(0, f) + ((2n+4)k+3)\Theta(\infty, f) + 2(k-1)\delta(\infty, f) > (2n+6)k+5;$$

(2) $l = 1$ *and*

$$2\delta_2(0, f) + \frac{1}{2}\Theta(0, f) + ((n+2)k + \frac{5}{2})\Theta(\infty, f) + (k-1)\delta(\infty, f) > (n+3)k+3;$$

(3) $l \geq 2$ *and*

$$2\delta_2(0,f) + ((n+2)k+2)\Theta(\infty,f) + (k-1)\delta(\infty,f) > (n+3)k+2.$$

Remark 1.1. *In Theorem 1.1, the condition* $\overline{E}_{(i}(0,f) \subset \overline{E}_{(i}(0,(L(f))^{(n)})$ $(i \geq 2)$ *is weaker than condition* f *and* $(L(f))^{(n)}$ *share* $0-CM$. *If* $(L(f))^{(n)}$ *and* f *share* $0-CM$, *then* $\overline{E}_{(i}(0,f) = \overline{E}_{(i}(0,(L(f))^{(n)})$ $(i \geq 1)$. *Then Theorem 1.1 still holds when* $(L(f))^{(n)}$ *and* f *share* 0-*CM.*

From Theorem 1.1, when f is an entire function, we get the following result:

Corollary 1.1. *Let* k, n *be positive integer numbers. Let* $f(z)$ *be an nonconstant entire function with hyper order less than* 1, *and assume that* $(L(f))^{(n)}$ *is not a constant function. Suppose that* $f-1$ *and* $(L(f))^{(n)}-1$ *share value* $(0,l)$ *and*

$$\overline{E}_{(i}(0,f) \subset \overline{E}_{(i}(0,(L(f))^{(n)}) \ (i \geq 2).$$

Then

$$(L(f))^{(n)} \equiv f$$

if one of the following assumptions holds:

(1) $l = 0$ *(i.e.* $f-1$ *and* $(L(f))^{(n)}-1$ *share the value* 0 *IM) and*

$$2\delta_2(0,f) + 3\Theta(0,f) > 4;$$

(2) $l = 1$ *and*

$$2\delta_2(0,f) + \frac{1}{2}\Theta(0,f) > \frac{3}{2};$$

(3) $l \geq 2$ *and* $\delta_2(0,f) > \dfrac{1}{2}$.

The equation $(L(f))^{(n)} \equiv f$ implies also that f is a solution to a linear difference-differential equation with constant coefficients. Therefore, in the principle, we can give some properties of solutions by using the characteristic equation for linear difference-differential equations. Motivate by the works of X. Qi and L. Yang [18] and S. Chen and A. Xu [2], we prove the uniqueness result for derivative of meromorphic function and its difference polynomial as follows:

Theorem 1.2. *Let* k, n *be positive integer numbers. Let* $f(z)$ *be a nonconstant meromorphic function with hyper order less than* 1, *and assume that* $L(f)$ *and* $f^{(n)}$ *are not constant functions. Suppose that* $f^{(n)}-1$ *and* $L(f)-1$ *share value* $(0,l)$, $f^{(n)}$ *and* $L(f)$ *share* ∞-*IM, and*

$$\overline{E}_{(i}(0,f) \subset \overline{E}_{(i}(0,L(f)) \ (i \geq 2).$$

Then

$$L(f) \equiv f^{(n)} \tag{1.2}$$

if one of the following assumptions holds:
(1) $l = 0$ *(i.e.* $f^{(n)} - 1$ *and* $L(f) - 1$ *share the value* 0 *IM) and*

$$(4k + 2n + 3)\Theta(\infty, f) + 2(k - 1)\delta(\infty, f) + 2\Theta(0, f) + \delta_2(0, f) + 2\delta_{n+1}(0, f) + \delta_{n+2}(0, f) > 6k + 2n + 6;$$

(2) $l = 1$ *and*

$$\delta_2(0, f) + \delta_{n+2}(0, f) + \frac{1}{2}\Theta(0, f) + (2k + \frac{5}{2})\Theta(\infty, f) + (k - 1)\delta(\infty, f) > 3k + 3;$$

(3) $l \geq 2$ *and*

$$(2k + 2)\Theta(\infty, f) + (k - 1)\delta(\infty, f) + \delta_2(0, f) + \delta_{n+2}(0, f) > 3k + 2.$$

Since $f^{(n)}(z)$ and $f(z+c)$ share 0-CM implies that $\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, f(z+c))$ $(i \geq 2)$, then Theorem 1.2 still holds when $f^{(n)}(z)$ and $f(z+c)$ share 0-CM and $L(f) = f(z + c), k = 1$. The assumptions in Theorem 1.2 are weaker than those in Theorem D. Namely, we consider that $f^{(n)}$ and $f(z+c)$ share partially value 0 and ∞-IM, $f^{(n)}$ and $f(z + c)$ share $(1, l)$. We note that the method proving Theorem 1.2 is not the same to [2] and [18]. For more results about uniqueness of meromorphic functions and their shift share partially value, we recommend the readers to [4, 11, 12]. Outside that problem, the uniqueness of difference-differential of meromorphic functions sharing values or small functions which was considered by many authors, we refer the readers to [5, 17] for more details. From Theorem 1.2, we get the following result:

Corollary 1.2. *Let* n *be positive integer numbers. Let* $f(z)$ *be a nonconstant meromorphic function with hyper order less than* 1, *and assume that* $f(z + c)$ *and* $f^{(n)}$ *are not constant functions, where* c *is a nonzero complex number. Suppose that* $f^{(n)} - 1$ *and* $f(z+c)-1$ *share value* $(0, l)$, $f^{(n)}$ *and* $f(z+c)$ *share* ∞*-IM, and*

$$\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, f(z + c)) \ (i \geq 2).$$

Then

$$f(z + c) \equiv f^{(n)}(z)$$

if one of the following assumptions holds:
(1) $l = 0$ *(i.e.* $f^{(n)} - 1$ *and* $L(f) - 1$ *share the value* 0 *IM) and*

$$(2n+7)\Theta(\infty,f)+2\Theta(0,f)+\delta_2(0,f)+2\delta_{n+1}(0,f)\\ +\delta_{n+2}(0,f)>2n+12;$$

(2) *$l=1$ and*

$$\delta_2(0,f)+\delta_{n+2}(0,f)+\frac{1}{2}\Theta(0,f)+\frac{9}{2}\Theta(\infty,f)>6;$$

(3) *$l\geq 2$ and*

$$4\Theta(\infty,f)+\delta_2(0,f)+\delta_{n+2}(0,f)>5.$$

From Theorem 1.2, when $k=1$ and $L(f)=f(z+c)$, we get the following result for entire functions:

Corollary 1.3. *Let k,n be positive integer numbers. Let $f(z)$ be a nonconstant entire function with hyper order less than 1, and assume that $f(z+c)$ and $f^{(n)}$ are not constant functions. Suppose that $f^{(n)}-1$ and $f(z+c)-1$ share value $(0,l)$, and*

$$\overline{E}_{(i}(0,f)\subset\overline{E}_{(i}(0,f(z+c))\ (i\geq 2).$$

Then

$$f(z+c)\equiv f^{(n)}(z)$$

if one of the following assumptions holds:

(1) *$l=0$ (i.e. $f^{(n)}-1$ and $f(z+c)-1$ share the value 0 IM) and*

$$2\Theta(0,f)+\delta_2(0,f)+2\delta_{n+1}(0,f)+\delta_{n+2}(0,f)>5;$$

(2) *$l=1$ and*

$$\delta_2(0,f)+\delta_{n+2}(0,f)+\frac{1}{2}\Theta(0,f)>\frac{3}{2};$$

(3) *$l\geq 2$ and*

$$\delta_2(0,f)+\delta_{n+2}(0,f)>1.$$

Finally, we study the growth of solutions to equations (1.1) and (1.2).

Theorem 1.3. *The order of all transcendental meromorphic solutions f of equations (1.1) and (1.2) must satisfy $\sigma(f)\geq 1$.*

Example 1.4. *The function* $f(z) = \sin z$ *has order* $\sigma(f) = 1$ *and* f *is a solution of equation*

$$f'(z) = -2f(z+\pi) + f(z-\frac{\pi}{2}).$$

Here $L(f) = -2f(z+\pi) + f(z-\frac{\pi}{2})$. *We also have that* f *is a solution of*

$$f'(z+\pi) = f(z),$$

where $L(f) = f(z+\pi)$.

2. Some Lemmas

In order to prove our results, we need the following lemmas.

Lemma 2.1 (Halburd-Korhonen-Tohge [10])**.** *Let* $h : [0, +\infty) \to [0, +\infty)$ *be a non-decreasing continuous function, and let* $s \in (0, +\infty)$. *If the hyper order of* h *is strictly less than one, i.e.,*

$$\limsup_{r\to\infty} \frac{\log\log h(r)}{\log r} = \varsigma < 1,$$

then

$$h(r+s) = h(r) + o(\frac{h(r)}{r^{1-\varsigma-\varepsilon}}),$$

where $\varepsilon > 0$ *and* $r \to \infty$ *outside of a set of finite logarithmic measure.*

From Lemma 2.1, we get the following corollary.

Corollary 2.1. *[1, 10] Let* f *be a non-constant meromorphic function with* $\varsigma(f) = \varsigma < 1$, *and* $c \in \mathbb{C} \setminus \{0\}$. *Then*

$$N(r, f(z+c)) \leq N(r, f) + S(r, f), \qquad \overline{N}(r, f(z+c)) \leq \overline{N}(r, f) + S(r, f),$$

$$N(r, \frac{1}{f(z+c)}) \leq N(r, \frac{1}{f}) + S(r, f), \qquad \overline{N}(r, \frac{1}{f(z+c)}) \leq \overline{N}(r, \frac{1}{f}) + S(r, f),$$

$$T(r, f(z+c)) = T(r, f) + S(r, f).$$

Lemma 2.2. *[19] Let* n *be a postive integer number. Let* f *be a non-constant meromorphic function such that* $f^{(n)} \not\equiv 0$. *Then*

$$N(r, \frac{1}{f^{(n)}}) \leqslant T(r, f^{(n)}) - T(r, f) + N(r, \frac{1}{f}) + S(r, f);$$

$$N(r, \frac{1}{f^{(n)}}) \leqslant n\overline{N}(r, f) + N(r, \frac{1}{f}) + S(r, f).$$

Lemma 2.3. *[21] Let p and k be two positive integers. Let f be a non-constant meromorphic function such that $f^{(k)} \not\equiv 0$. Then*

$$N_p(r, \frac{1}{f^{(k)}}) \leqslant T(r, f^{(k)}) - T(r, f) + N_{p+k}(r, \frac{1}{f}) + S(r, f);$$
$$N_p(r, \frac{1}{f^{(k)}}) \leqslant k\overline{N}(r, f) + N_{p+k}(r, \frac{1}{f}) + S(r, f).$$

Lemma 2.4. *[20] Let f and g be two non-constant meromorphic functions, and let $a(z)$ $(a \not\equiv 0, \infty)$ be a small function of both f and g. If f and g share $(a(z), 0)$, then one of the following three cases holds:*

$$(i)\ \ T(r,f) \leqslant N_2(r,f) + N_2(r,\frac{1}{f}) + N_2(r,g) + N_2(r,\frac{1}{g}) + 2(\overline{N}(r,\frac{1}{f}) + \overline{N}(r,f)) + (\overline{N}(r,\frac{1}{g}) + \overline{N}(r,g)) + S(r,f) + S(r,g),$$

and the similar inequality holds for T(r,g);

(ii) $f \equiv g$;

(iii) $fg \equiv a^2$.

Lemma 2.5. *[20] Let f and g be two non-constant meromorphic functions, and let $a(z)$ $(a \not\equiv 0, \infty)$ be a small function of both f and g. If f and g share $(a(z), 1)$, then one of the following three cases holds:*

$$(i)\ \ T(r,f) \leqslant N_2(r,f) + N_2(r,\frac{1}{f}) + N_2(r,g) + N_2(r,\frac{1}{g}) + \frac{1}{2}(\overline{N}(r,\frac{1}{f}) + \overline{N}(r,f)) + S(r,f) + S(r,g),$$

and the similar inequality holds for T(r,g);

(ii) $f \equiv g$;

(iii) $fg \equiv a^2$.

Lemma 2.6. *[16, 20] Let f and g be two non-constant meromorphic functions, and let $a(z)$ $(a \not\equiv 0, \infty)$ be a small function of both f and g. If f and g share $(a(z), l), l \geq 2$, then one of the following three cases holds:*

$$(i)\ \ T(r,f) \leqslant N_2(r,f) + N_2(r,\frac{1}{f}) + N_2(r,g) + N_2(r,\frac{1}{g}) + S(r,f) + S(r,g)$$

and the similar inequality holds for T(r,g);

(ii) $f \equiv g$;

(iii) $fg \equiv a^2$.

Lemma 2.7. *[13] Let f be a non-constant meromorphic function with hyper-order less than* 1*, and $L(f) \not\equiv 0$ be defined as in Theorem A. Then*

$$N(r, \frac{1}{L(f)}) \leq T(r, L(f)) - T(r, f) + N(r, \frac{1}{f}) + S(r, f),$$
$$N(r, \frac{1}{L(f)}) \leq (k-1)N(r, f) + N(r, \frac{1}{f}) + S(r, f).$$

From Lemma 2.7, we get the following result:

Lemma 2.8. *Let n, p be integer numbers. Let f be a non-constant meromorphic function with hyper order less than* 1 *such that $L(f) \not\equiv 0$. Suppose $\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, L(f))$ (all $i \geq p+1$). Then*

$$N_p(r, \frac{1}{L(f)}) \leq T(r, L(f)) - T(r, f) + N_p(r, \frac{1}{f}) + S(r, f),$$
$$N_p(r, \frac{1}{L(f)}) \leq (k-1)N(r, f) + N_p(r, \frac{1}{f}) + S(r, f).$$

Proof. Apply to Lemma 2.7, we have

$$N(r, \frac{1}{L(f)}) \leqslant T(r, L(f)) - T(r, f) + N(r, \frac{1}{f}) + S(r, f). \tag{2.1}$$

We have

$$N(r, \frac{1}{L(f)}) = N_p(r, \frac{1}{L(f)}) + \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{L(f)}) \tag{2.2}$$

and

$$N(r, \frac{1}{f}) = N_p(r, \frac{1}{f}) + \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{f}). \tag{2.3}$$

Hence, combining (2.1) to (2.3) and by the assumption

$$\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, L(f)) \text{ (all } i \geq p+1),$$

we get $\overline{N}_{(j}(r, \frac{1}{f}) \leq \overline{N}_{(j}(r, \frac{1}{L(f)})$ for all $j \geq p+1$. Using Lemma 2.7 and (2.2), we have

$$N_p(r, \frac{1}{L(f)}) \leq T(r, L(f)) - T(r, f) - \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{L(f)}) + N(r, \frac{1}{f}) + S(r, f). \tag{2.4}$$

Combine (2.3) and (2.4) to get

$$\begin{aligned} N_p(r, \frac{1}{L(f)}) &\leq T(r, L(f)) - T(r, f) + N_p(r, \frac{1}{f}) \\ &\quad + \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{f}) - \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{L(f)}) + S(r, f) \\ &\leq T(r, L(f)) - T(r, f) + N_p(r, \frac{1}{f}) + S(r, f). \end{aligned}$$

The remain inequality is similarly proved. For convenience to readers, we write some steps as follows. From (2.1) and Lemma 2.7, we have

(2.5)
$$N_p(r, \frac{1}{L(f)}) \leq (k-1)N(r, f) + N(r, \frac{1}{f}) - \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{L(f)}) + S(r, f).$$

Then second statement comes from (2.3) and (2.5). ■

Next, we prove some results as following:

Lemma 2.9. *Let n be a integer number. Let f be a non-constant meromorphic function with hyper order less than 1 such that $(L(f))^{(n)} \not\equiv 0$. Then*

$$N(r, \frac{1}{(L(f))^{(n)}}) \leq T(r, (L(f))^{(n)}) - T(r, f) + N(r, \frac{1}{f}) + S(r, f),$$
$$N(r, \frac{1}{(L(f))^{(n)}}) \leq nk\overline{N}(r, f) + (k-1)N(r, f) + N(r, \frac{1}{f}) + S(r, f).$$

Proof. Apply Lemma 2.2, we have

$$N(r, \frac{1}{(L(f))^{(n)}}) \leqslant T(r, (L(f))^{(n)}) - T(r, L(f)) + N(r, \frac{1}{L(f)}) + S(r, f). \tag{2.6}$$

By Lemma 2.7, from (2.6), we get

$$N(r, \frac{1}{L(f)}) \leqslant T(r, L(f)) - T(r, f) + N(r, \frac{1}{f}) + S(r, f). \tag{2.7}$$

Combine (2.6) and (2.7), we get the first inequality. Next, we show the second inequality. By Lemma 2.2, we have

$$N(r, \frac{1}{(L(f))^{(n)}}) \leqslant n\overline{N}(r, L(f)) + N(r, \frac{1}{L(f)}) + S(r, f). \tag{2.8}$$

Combining (2.8), Lemma 2.7 and Corollary 2.1, we obtain

$$N(r, \frac{1}{(L(f))^{(n)}}) \leq nk\overline{N}(r, f) + (k-1)N(r, f) + N(r, \frac{1}{f}) + S(r, f).$$

■

From Lemma 2.9, we get the following result.

Corollary 2.2. *Let n be a integer number. Let f be a non-constant entire function with hyper order less than 1 such that $(L(f))^{(n)} \not\equiv 0$. Then*

$$N(r, \frac{1}{(L(f))^{(n)}}) \leq T(r, (L(f))^{(n)}) - T(r, f) + N(r, \frac{1}{f}) + S(r, f),$$
$$N(r, \frac{1}{(L(f))^{(n)}}) \leq N(r, \frac{1}{f}) + S(r, f).$$

Lemma 2.10. *Let n, p be integer numbers. Let f be a non-constant meromorphic function with hyper order less than 1 such that $(L(f))^{(n)} \not\equiv 0$. Suppose $\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, (L(f))^{(n)})$ (all $i \geq p+1$). Then*

$$N_p(r, \frac{1}{(L(f))^{(n)}}) \leq T(r, (L(f))^{(n)}) - T(r, f) + N_p(r, \frac{1}{f}) + S(r, f),$$
$$N_p(r, \frac{1}{(L(f))^{(n)}}) \leq nk\overline{N}(r, f) + (k-1)N(r, f) + N_p(r, \frac{1}{f}) + S(r, f).$$

Proof. Apply Lemma 2.9, we have

$$N(r, \frac{1}{(L(f))^{(n)}}) \leqslant T(r, (L(f))^{(n)}) - T(r, f) + N(r, \frac{1}{f}) + S(r, f). \tag{2.9}$$

We have

$$N(r, \frac{1}{(L(f))^{(n)}}) = N_p(r, \frac{1}{(L(f))^{(n)}}) + \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{(L(f))^{(n)}}) \tag{2.10}$$

and

$$N(r, \frac{1}{f}) = N_p(r, \frac{1}{f}) + \sum_{j=p+1}^{\infty} \overline{N}_{(j}(r, \frac{1}{f}). \tag{2.11}$$

Hence, combining (2.9) to (2.11) and by the assumption

$$\overline{E}_{(i}(0, f) \subset \overline{E}_{(i}(0, (L(f))^{(n)}) \text{ (all } i \geq p+1),$$

we get

$$\begin{aligned}N_p(r,\frac{1}{(L(f))^{(n)}}) &\le T(r,(L(f))^{(n)})-T(r,f)+N_p(r,\frac{1}{f})\\ &\quad+\sum_{j=p+1}^{\infty}\overline{N}_{(j}(r,\frac{1}{f})-\sum_{j=p+1}^{\infty}\overline{N}_{(j}(r,\frac{1}{(L(f))^{(n)}})+S(r,f)\\ &\le T(r,(L(f))^{(n)})-T(r,f)+N_p(r,\frac{1}{f})+S(r,f).\end{aligned}$$

By Lemma 2.9, we have

$$N(r,\frac{1}{(L(f))^{(n)}})\le nk\overline{N}(r,f)+(k-1)N(r,f)+N(r,\frac{1}{f})+S(r,f). \tag{2.12}$$

Hence, combining (2.9), (2.11) and (2.12), we obtain

$$\begin{aligned}N_p(r,\frac{1}{(L(f))^{(n)}}) &\le nk\overline{N}(r,f)+(k-1)N(r,f)+N_p(r,\frac{1}{f})\\ &\quad+\sum_{j=p+1}^{\infty}\overline{N}_{(j}(r,\frac{1}{f})-\sum_{j=p+1}^{\infty}\overline{N}_{(j}(r,\frac{1}{(L(f))^{(n)}})+S(r,f)\\ &\le nk\overline{N}(r,f)+(k-1)N(r,f)+N_p(r,\frac{1}{f})+S(r,f).\end{aligned}$$

■

From Lemma 2.10, we get the following result.

Corollary 2.3. *Let n,p be integer numbers. Let f be a non-constant entire function with hyper order less than 1 such that $(L(f))^{(n)}\not\equiv 0$. Suppose $\overline{E}_{(i}(0,f)\subset\overline{E}_{(i}(0,(L(f))^{(n)})$ (all $i\ge p+1$). Then*

$$\begin{aligned}N_p(r,\frac{1}{(L(f))^{(n)}}) &\le T(r,(L(f))^{(n)})-T(r,f)+N_p(r,\frac{1}{f})+S(r,f),\\ N_p(r,\frac{1}{(L(f))^{(n)}}) &\le N_p(r,\frac{1}{f})+S(r,f).\end{aligned}$$

Lemma 2.11. *Let c_1 and c_2 be two arbitrary complex numbers, and let f be a meromorphic function of finite order σ. Assume that $\varepsilon>0$, then there exists a subset $E\subset\mathbb{R}$ with finite logarithmic measure so that for all $|z|=r\notin E\cup[0,1]$, we have*

$$exp(-r^{\sigma-1+\varepsilon})\le\left|\frac{f(z+c_1)}{f(z+c_2)}\right|\le exp(r^{\sigma-1+\varepsilon}).$$

Lemma 2.12. *[6, Corollary 1] Assume that f is a transcendental meromorphic function of finite order $\sigma = \sigma(f)$. Let $\varepsilon > 0$, $k > j \geq 0$ be two integers. Then there exists a set $E \subset [0, 2\pi)$ with linear measure zero, so that if $\varphi \in [0, 2\pi) \setminus E$, then there is a constant $R_0 = R_0(\varphi) > 0$ so that for all z verifying $argz = \varphi$ and $|z| \geq R_0$, we have*

$$\left|\frac{f^{(k)}(z)}{f^{(j)}(z)}\right| \leq |z|^{(k-j)(\sigma-1+\varepsilon)}.$$

Lemma 2.13. *Assume that f is a transcendental meromorphic function of finite order $\sigma = \sigma(f)$. Let c_1 and c_2 be complex numbers and k is a positive integer and $\varepsilon > 0$. Then there is a subset $E_1 \subset \mathbb{R}$ with finite logarithmic measure and set $E \subset [0, 2\pi)$ with linear measure zero so that if $z = re^{i\varphi}$, $\varphi \in [0, 2\pi) \setminus E$, we have that*

$$\left|\frac{f^{(k)}(z+c_1)}{f(z+c_2)}\right| \leq |z|^{k(\sigma-1+\varepsilon)} exp(r^{\sigma-1+\varepsilon})$$

holds for all $|z| = r \geq r_0(\varphi) > 1$ and $|z| \notin E_1$.

Proof. Since f has finite order, then by Corollary 2.1, we have

$$T(r, f(z+c_1)) = T(r, f) + o(T(r, f)).$$

It implies that $f(z + c_1)$ has finite order and $\sigma f(z + c_1) = \sigma(f)$. By Lemma 2.12 for $g(z) = f(z + c_1)$, there is a set $E \subset [0, 2\pi)$ with linear measure zero, so that if $\varphi \in [0, 2\pi) \setminus E$, then there is a constant $R_0 = R_0(\varphi) > 1$ so that

$$\left|\frac{g^{(k)}(z)}{g(z)}\right| \leq |z|^{k(\sigma-1+\varepsilon)} \tag{2.13}$$

holds for all z satisfying $argz = \varphi$ and $|z| \geq R_0 > 1$. Using Lemma 2.11, there is a subset $E \subset \mathbb{R}$ with finite logarithmic measure so that for all $r \notin E_1 \cup [0, 1]$, we have

$$\exp(-r^{\sigma-1+\varepsilon}) \leq \left|\frac{f(z+c_1)}{f(z+c_2)}\right| \leq \exp(r^{\sigma-1+\varepsilon}). \tag{2.14}$$

Combine (2.13) and (2.13), we deduce that

$$\left|\frac{f^{(k)}(z+c_1)}{f(z+c_2)}\right| = \left|\frac{f^{(k)}(z+c_1)}{f(z+c_1)}\frac{f(z+c_1)}{f(z+c_2)}\right| \leq |z|^{k(\sigma-1+\varepsilon)}\exp(r^{\sigma-1+\varepsilon})$$

holds for all $z : argz = \varphi$ and $|z| \geq R_0 > 1$ and $|z| \notin E_1$. ∎

3. Proof of Theorems

3.1. Proof of Theorem 1.1

Proof. From the conditions of Theorem 1.1, we know that f and $(L(f))^{(n)}$ share $(1,l)$. We consider three cases as following of l.

Case 1: $l=0$. Apply Lemma 2.4, we may assume that two following inequalities hold:

$$\begin{aligned} T(r,(L(f))^{(n)}) \leqslant\ & N_2(r,(L(f))^{(n)}) + N_2(r,\frac{1}{(L(f))^{(n)}}) + N_2(r,f) + N_2(r,\frac{1}{f}) \\ & + 2(\overline{N}(r,\frac{1}{(L(f))^{(n)}}) + \overline{N}(r,(L(f))^{(n)})) + (\overline{N}(r,\frac{1}{f}) + \overline{N}(r,f)) + S(r,f), \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} T(r,f) \leqslant\ & N_2(r,f) + N_2(r,\frac{1}{f}) + N_2(r,(L(f))^{(n)}) + N_2(r,\frac{1}{(L(f))^{(n)}}) \\ & + 2(\overline{N}(r,\frac{1}{f}) + \overline{N}(r,f)) + (\overline{N}(r,\frac{1}{(L(f))^{(n)}}) + \overline{N}(r,(L(f))^{(n)})) + S(r,f). \end{aligned} \tag{3.2}$$

First, from Corollary 2.1, we have

$$N_2(r,(L(f))^{(n)}) \le 2\overline{N}(r,(L(f))^{(n)}) = 2\overline{N}(r,L(f)) \le 2k\overline{N}(r,f) + S(r,f). \tag{3.3}$$

By Lemma 2.10, we know

$$\begin{aligned} N_2(r,\frac{1}{(L(f))^{(n)}}) &\le nk\overline{N}(r,f) + (k-1)N(r,f) + N_2(r,\frac{1}{f}) + S(r,f), \\ \overline{N}(r,\frac{1}{(L(f))^{(n)}}) &\le nk\overline{N}(r,f) + (k-1)N(r,f) + \overline{N}(r,\frac{1}{f}) + S(r,f). \end{aligned} \tag{3.4}$$

Still using Lemma 2.10 and (3.1), (3.3)-(3.4), we get

$$\begin{aligned} T(r,(L(f))^{(n)}) \leqslant\ & T(r,(L(f))^{(n)}) - T(r,f) + 2N_2(r,\frac{1}{f}) + 3\overline{N}(r,\frac{1}{f}) \\ & + (k(2n+4)+3)\overline{N}(r,f) + 2(k-1)N(r,f) + S(r,f). \end{aligned}$$

This implies

$$T(r,f) \leqslant 2N_2(r,\frac{1}{f}) + 3\overline{N}(r,\frac{1}{f}) + (k(2n+4)+3)\overline{N}(r,f) + 2(k-1)N(r,f) + S(r,f). \tag{3.5}$$

Similarly, from Lemma 2.10 and (3.2), we obtain

$$\begin{aligned} T(r,f) &\leqslant 2N_2(r,\frac{1}{f}) + 3\overline{N}(r,\frac{1}{f}) + (k(2n+3)+4)\overline{N}(r,f) \\ &+ 2(k-1)N(r,f) + S(r,f) \\ &\leq 2N_2(r,\frac{1}{f}) + 3\overline{N}(r,\frac{1}{f}) + (k(2n+4)+3)\overline{N}(r,f) \\ &+ 2(k-1)N(r,f) + S(r,f). \end{aligned} \tag{3.6}$$

Therefore, combining (3.5) and (3.6), we get

$$\begin{aligned} T(r,f) &\leqslant 2(1-\delta_2(0,f))T(r,f) + 3(1-\Theta(0,f))T(r,f) \\ &+ (k(2n+4)+3)(1-\Theta(\infty,f))T(r,f) \\ &+ 2(k-1)(1-\delta(\infty,f))T(r,f) + S(r,f). \end{aligned}$$

This implies $(K_1 - ((2n+6)k+5))T(r,f) \leq S(r,f)$, where

$$\begin{aligned} K_1 &= 2\delta_2(0,f) + 3\Theta(0,f) + ((2n+4)k+3)\Theta(\infty,f) \\ &+ 2(k-1)\delta(\infty,f) - ((2n+6)k+5) > 0 \end{aligned}$$

since

$$2\delta_2(0,f)+3\Theta(0,f)+((2n+4)k+3)\Theta(\infty,f)+2(k-1)\delta(\infty,f) > (2n+6)k+5.$$

This is a contradiction. Thus, by Lemma 2.4, we must have $f \equiv (L(f))^{(n)}$ or $f.(L(f))^{(n)} \equiv 1$. We consider the case $f.(L(f))^{(n)} \equiv 1$. Since f and $(L(f))^{(n)}$ share $\infty-$ IM, then the case $f.(L(f))^{(n)} \equiv 1$ is impossible. Hence, we obtain

$$f \equiv (L(f))^{(n)}.$$

We have finished the proof of Theorem 1.1 in the case $l = 0$.

Case 2: $l = 1$. Apply to Lemma 2.5, we may assume that two inequality below hold:

$$\begin{aligned} T(r,(L(f))^{(n)}) &\leqslant N_2(r,(L(f))^{(n)}) + N_2(r,\frac{1}{(L(f))^{(n)}}) + N_2(r,f) + N_2(r,\frac{1}{f}) \\ &+ \frac{1}{2}(\overline{N}(r,\frac{1}{(L(f))^{(n)}}) + \overline{N}(r,(L(f))^{(n)})) + S(r,f), \end{aligned} \tag{3.7}$$

and

$$T(r,f) \leqslant N_2(r,f) + N_2(r,\frac{1}{f}) + N_2(r,(L(f))^{(n)}) + N_2(r,\frac{1}{(L(f))^{(n)}}) + \frac{1}{2}(\overline{N}(r,\frac{1}{f}) + \overline{N}(r,f)) + S(r,f). \tag{3.8}$$

Combine Lemma 2.10 and (3.7), we get

$$T(r,(L(f))^{(n)}) \leqslant T(r,(L(f))^{(n)}) - T(r,f) + 2N_2(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) + ((\frac{n+5}{2})k+2)\overline{N}(r,f) + \frac{k-1}{2}N(r,f) + S(r,f).$$

This implies

$$T(r,f) \leqslant 2N_2(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) + ((\frac{n+5}{2})k+2)\overline{N}(r,f) + \frac{k-1}{2}N(r,f) + S(r,f). \tag{3.9}$$

Similarly, from Lemma 2.10, (3.3)-(3.4) and (3.8), we obtain

$$T(r,f) \leqslant 2N_2(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) + ((n+2)k+\frac{5}{2})\overline{N}(r,f) + (k-1)N(r,f) + S(r,f). \tag{3.10}$$

Since

$$((\frac{n+5}{2}k+2)\overline{N}(r,f) + \frac{k-1}{2}N(r,f) \leq ((n+2)k+\frac{5}{2})\overline{N}(r,f) + (k-1)N(r,f),$$

then, combining (3.9) and (3.10), we get

$$T(r,f) \leqslant 2(1-\delta_2(0,f))T(r,f) + \frac{1}{2}(1-\Theta(0,f))T(r,f) + ((n+2)k+\frac{5}{2})(1-\Theta(\infty,f))T(r,f) + (k-1)(1-\delta(\infty,f))T(r,f) + S(r,f).$$

This implies

$$(K_2 - ((n+3)k+3))T(r,f) \leq S(r,f),$$

where

$$K_2 = 2\delta_2(0,f) + \frac{1}{2}\Theta(0,f) + ((n+2)k+\frac{5}{2})\Theta(\infty,f) + (k-1)\delta(\infty,f).$$

This is a contradiction with

$$2\delta_2(0,f) + \frac{1}{2}\Theta(0,f) + ((n+2)k+\frac{5}{2})\Theta(\infty,f) + (k-1)\delta(\infty,f) > (n+3)k+3.$$

By an argument as Case 1, we have

$$f \equiv (L(f))^{(n)}.$$

Case 3: $l \geq 2$. Apply Lemma 2.6, we may assume that two inequalities below hold.

$$T(r,(L(f))^{(n)}) \leqslant N_2(r,(L(f))^{(n)}) + N_2(r,\frac{1}{(L(f))^{(n)}}) + N_2(r,f) + N_2(r,\frac{1}{f}) + S(r,f), \tag{3.11}$$

and

$$T(r,f) \leqslant N_2(r,f) + N_2(r,\frac{1}{f}) + N_2(r,(L(f))^{(n)}) + N_2(r,\frac{1}{(L(f))^{(n)}}) + S(r,f). \tag{3.12}$$

Using Lemma 2.10, (3.3)-(3.4) and (3.11), (3.12) implies that

$$T(r,f) \leqslant 2N_2(r,\frac{1}{f}) + ((n+2)k+2)\overline{N}(r,f) + (k-1)N(r,f) + S(r,f). \tag{3.13}$$

Indeed, (3.11) implies

$$\begin{aligned} T(r,f) &\leq (2k+2)\overline{N}(r,f) + 2N_2(r,\frac{1}{f}) + S(r,f) \\ &\leq 2N_2(r,\frac{1}{f}) + ((n+2)k+2)\overline{N}(r,f) + (k-1)N(r,f) + S(r,f). \end{aligned}$$

Therefore, from (3.13) we deduce

$$\begin{aligned} T(r,f) \leqslant{}& 2(1-\delta_2(0,f))T(r,f) + ((n+2)k+2)(1-\Theta(\infty,f))T(r,f) \\ &+ (k-1)(1-\delta(\infty,f))T(r,f) + S(r,f). \end{aligned}$$

This implies $(K_3 - ((n+3)k+2))T(r,f) \leq S(r,f)$, where

$$K_3 = 2\delta_2(0,f) + ((n+2)k+2)\Theta(\infty,f) + (k-1)\delta(\infty,f).$$

This is a contradiction with

$$2\delta_2(0,f) + ((n+2)k+2)\Theta(\infty,f) + (k-1)\delta(\infty,f) > (n+3)k+2.$$

By an argument as Case 1, we have $f \equiv (L(f))^{(n)}$. ■

3.2. Proof of Theorem 1.2

Proof. From the conditions of Theorem 1.2, we know that $f^{(n)}$ and $L(f)$ share $(1, l)$. We consider three cases as following of l.

Case 1: $l = 0$. Apply Lemma 2.4, we may assume that two following inequalities hold:

(3.14)
$$T(r, L(f)) \leqslant N_2(r, L(f)) + N_2(r, \frac{1}{L(f)}) + N_2(r, f^{(n)}) + N_2(r, \frac{1}{f^{(n)}}) + 2(\overline{N}(r, \frac{1}{L(f)}) + \overline{N}(r, L(f))) + (\overline{N}(r, \frac{1}{f^{(n)}}) + \overline{N}(r, f^{(n)})) + S(r, f),$$

and

(3.15)
$$T(r, f^{(n)}) \leqslant N_2(r, f^{(n)}) + N_2(r, \frac{1}{f^{(n)}}) + N_2(r, L(f)) + N_2(r, \frac{1}{L(f)}) + 2(\overline{N}(r, \frac{1}{f^{(n)}}) + \overline{N}(r, f^{(n)})) + (\overline{N}(r, \frac{1}{L(f)}) + \overline{N}(r, L(f))) + S(r, f).$$

From Corrollary 2.1 and (3.14), we have

$$T(r, L(f)) \leqslant (2k+2)\overline{N}(r, f) + N_2(r, \frac{1}{L(f)}) + N_2(r, \frac{1}{f^{(n)}}) + (2k+1)\overline{N}(r, f) + 2\overline{N}(r, \frac{1}{L(f)}) + \overline{N}(r, \frac{1}{f^{(n)}}) + S(r, f), \tag{3.16}$$

Using Lemma 2.2 and Lemma 2.8, (3.16) implies that

$$T(r, L(f)) \leqslant (2k+2)\overline{N}(r, f) + T(r, L(f)) - T(r, f) + N_2(r, \frac{1}{f}) + n\overline{N}(r, f) + N_{n+2}(r, \frac{1}{f}) + (2k+1)\overline{N}(r, f) + 2((k-1)N(r, f) + \overline{N}(r, \frac{1}{f})) + n\overline{N}(r, f) + N_{n+1}(r, \frac{1}{f}) + S(r, f).$$

Hence, we deduce

$$T(r, f) \leq (4k+2n+3)\overline{N}(r, f) + 2(k-1)N(r, f) + 2\overline{N}(r, \frac{1}{f}) + N_2(r, \frac{1}{f}) + N_{n+2}(r, \frac{1}{f}) + 2N_{n+1}(r, \frac{1}{f}) + S(r, f). \tag{3.17}$$

From (3.15), using Lemma 2.2 and Lemma 2.8, we have

$$\begin{aligned} T(r,f) &\leqslant (2n+3k+4)\overline{N}(r,f) + 2(k-1)N(r,f) + \overline{N}(r,\frac{1}{f}) \\ &+ N_2(r,\frac{1}{f}) + 2N_{n+1}(r,\frac{1}{f}) + N_{n+2}(r,\frac{1}{f}) \\ &\leq (4k+2n+3)\overline{N}(r,f) + 2(k-1)N(r,f) + 2\overline{N}(r,\frac{1}{f}) \\ &+ N_2(r,\frac{1}{f}) + 2N_{n+1}(r,\frac{1}{f}) + N_{n+2}(r,\frac{1}{f}) + S(r,f). \end{aligned} \tag{3.18}$$

From (3.17) and (3.18), we have $K_4T(r,f) \leq S(r,f)$, where

$$\begin{aligned} K_4 = (4k+2n+3)\Theta(\infty,f) + 2(k-1)\delta(\infty,f) + 2\Theta(0,f) + \delta_2(0,f) \\ + 2\delta_{n+1}(0,f) + \delta_{n+2}(0,f) - (6k+2n+6). \end{aligned}$$

It is a contradiction since

$$\begin{aligned} (4k+2n+3)\Theta(\infty,f) + 2(k-1)\delta(\infty,f) + 2\Theta(0,f) + \delta_2(0,f) + 2\delta_{n+1}(0,f) \\ + \delta_{n+2}(0,f) > (6k+2n+6). \end{aligned}$$

Thus, by Lemma 2.4, we must have $f^{(n)} \equiv L(f)$ or $f^{(n)}.L(f) \equiv 1$. The equality $f^{(n)}.L(f) \equiv 1$ cannot occur since $f^{(n)}$ and $L(f)$ share ∞-IM. Hence, we obtain

$$f \equiv (L(f))^{(n)}.$$

We have finished the proof of Theorem 1.2 in the case $l = 0$.

Case 2: $l = 1$. Apply Lemma 2.5, we may assume that two inequalities below hold:

$$\begin{aligned} T(r,L(f)) &\leqslant N_2(r,L(f)) + N_2(r,\frac{1}{L(f)}) + N_2(r,f) + N_2(r,\frac{1}{f}) \\ &+ \frac{1}{2}(\overline{N}(r,\frac{1}{L(f)}) + \overline{N}(r,L(f))) + S(r,f), \end{aligned} \tag{3.19}$$

and

$$\begin{aligned} T(r,f^{(n)}) &\leqslant N_2(r,f^{(n)}) + N_2(r,\frac{1}{f^{(n)}}) + N_2(r,L(f)) + N_2(r,\frac{1}{L(f)}) \\ &+ \frac{1}{2}(\overline{N}(r,\frac{1}{f}) + \overline{N}(r,f)) + S(r,f). \end{aligned} \tag{3.20}$$

Combine Lemma 2.8 and (3.19), we get

$$T(r,L(f)) \leqslant (2k+2)\overline{N}(r,f) + T(r,L(f)) - T(r,f) + 2N_2(r,\frac{1}{f})$$
$$+ \frac{1}{2}((k-1)N(r,f) + \overline{N}(r,\frac{1}{f})) + \frac{k}{2}\overline{N}(r,f) + S(r,f).$$

It implies that

$$\begin{aligned} T(r,f) &\leq 2N_2(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) \\ &\quad + (\frac{5k}{2}+2)\overline{N}(r,f) + \frac{k-1}{2}N(r,f) + S(r,f) \\ &\leq N_2(r,\frac{1}{f}) + N_{n+2}(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) \\ &\quad + (2k+\frac{5}{2})\overline{N}(r,f) + (k-1)N(r,f) + S(r,f). \end{aligned} \tag{3.21}$$

Similarly, from Lemma 2.3, Lemma 2.8 and (3.20), we obtain

$$T(r,f) \leqslant T(r,f^{(n)}) - T(r,f) + (2k+\frac{5}{2})\overline{N}(r,f) + (k-1)N(r,f) + N_2(r,\frac{1}{f})$$
$$+ N_{n+2}(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) + S(r,f).$$

Hence, we deduce

$$\begin{aligned} T(r,f) &\leqslant (2k+\frac{5}{2})\overline{N}(r,f) + (k-1)N(r,f) + N_2(r,\frac{1}{f}) \\ &\quad + N_{n+2}(r,\frac{1}{f}) + \frac{1}{2}\overline{N}(r,\frac{1}{f}) + S(r,f). \end{aligned} \tag{3.22}$$

From (3.21) and (3.22), we get $(K_5 - ((3k+3))T(r,f) \leq S(r,f)$, where

$$K_5 = \delta_2(0,f) + \delta_{n+2}(0,f) + \frac{1}{2}\Theta(0,f) + (2k+\frac{5}{2})\Theta(\infty,f) + (k-1)\delta(\infty,f).$$

It is a contradiction with

$$\delta_2(0,f) + \delta_{n+2}(0,f) + \frac{1}{2}\Theta(0,f) + ((2k+\frac{5}{2})\Theta(\infty,f) + (k-1)\delta(\infty,f) > 3k+3.$$

By an argument as Case 1 of Theorem 1.1, we have

$$f^{(n)} \equiv L(f).$$

Case 3: $l \geq 2$. Apply Lemma 2.6, we may assume that two inequalities below hold.

$$T(r, L(f)) \leqslant N_2(r, L(f)) + N_2(r, \frac{1}{L(f)}) + N_2(r, f) + N_2(r, \frac{1}{f}) + S(r, f), \tag{3.23}$$

and

$$T(r, f^{(n)}) \leqslant N_2(r, f^{(n)}) + N_2(r, \frac{1}{f^{(n)}}) + N_2(r, L(f)) + N_2(r, \frac{1}{L(f)}) + S(r, f). \tag{3.24}$$

Combine Lemma 2.8 and (3.23), we get

$$T(r, L(f)) \leqslant T(r, L(f)) - T(r, f) + (2k+2)\overline{N}(r, f) + 2N_2(r, \frac{1}{f}) + S(r, f).$$

This implies

$$\begin{aligned} T(r, f) &\leqslant 2N_2(r, \frac{1}{f}) + (2k+2)\overline{N}(r, f) + S(r, f) \\ &\leq N_2(r, \frac{1}{f}) + N_{n+2}(r, \frac{1}{f}) + (k-1)N(r, f) + (2k+2)\overline{N}(r, f) + S(r, f). \end{aligned} \tag{3.25}$$

Using Lemma 2.3, Lemma 2.8 and (3.24), we deduce

$$\begin{aligned} T(r, f^{(n)}) \leq{}& (2k+2)\overline{N}(r, f) + (k-1)N(r, f) + N_2(r, \frac{1}{f}) + N_{n+2}(r, \frac{1}{f}) \\ &+ T(r, f^{(n)}) - T(r, f) + S(r, f). \end{aligned}$$

It implies that

$$T(r, f) \leq (2k+2)\overline{N}(r, f) + (k-1)N(r, f) + N_2(r, \frac{1}{f}) + N_{n+2}(r, \frac{1}{f}) + S(r, f). \tag{3.26}$$

From (3.25) and (3.26), we get $(K_6 - (3k+2))T(r, f) \leqslant S(r, f)$, where

$$K_6 = (2k+2)\Theta(\infty, f) + (k-1)\delta(\infty, f) + \delta_2(0, f) + \delta_{n+2}(0, f).$$

This is a contradiction with

$$(2k+2)\Theta(\infty, f) + (k-1)\delta(\infty, f) + \delta_2(0, f) + \delta_{n+2}(0, f) > 3k+2.$$

By an argument as Case 1, we have

$$f^{(n)} \equiv L(f).$$

■

3.3. Proof of Theorem 1.3

Proof. First, we assume that f is a transcendental meromorphic solution of (1.1). It means that

$$(\sum_{j=1}^{k} a_j f(z+c_j))^{(n)} = f. \tag{3.27}$$

Assume that the solution of (3.27) has order $\sigma(f) < 1$, then we can choose $\varepsilon > 0$ such that $0 < \varepsilon < 1 - \sigma$. Apply Lemma 2.13, there is a subset $E_1^j \subset \mathbb{R}$ with finite logarithmic measure and set $E_j \subset [0, 2\pi)$ with linear measure zero so that if $z = re^{i\varphi}$, $\varphi \in [0, 2\pi) \setminus E_j$, we have that

$$\left|\frac{f^{(n)}(z+c_j)}{f(z)}\right| \le |z|^{n(\sigma-1+\varepsilon)}\exp(r^{\sigma-1+\varepsilon}),\ j = 1, \ldots, k, \tag{3.28}$$

hold for all $|z| = r \ge r_j(\varphi) > 1$ and $|z| \notin E_1^j$. We denote $E_1 = \cup_{j=1}^k E_1^j$ and $E = \cup_{j=1}^k E_j$, then E has measure zero in $[0, 2\pi)$ and E_1 has finite logarithmic measure. Denote $r_0 = \max_{j=1,\ldots,k} r_j(\varphi)$, then (3.28) holds for all $j = 1, \ldots, k$ and $z = re^{i\varphi}, \varphi \in [0, 2\pi) \setminus E$ and $|z| > r_0$, $|z| \notin E_1$. Thus, from (3.27) and (3.28), we get

$$1 \le \sum_{j=1}^{k} |a_j| r^{n(\sigma-1+\varepsilon)}\exp(r^{\sigma-1+\varepsilon}). \tag{3.29}$$

Since $\sigma - 1 + \varepsilon < 0$, let $r \to \infty, r \notin E_1$ in (3.29), the right side tends to zero and we get a contradiction. Hence we get $\sigma(f) \ge 1$. If f is a solution of (1.2), using Lemma 2.13 and by arguments as previous computing, we obtain $\sigma(f) \ge 1$.

■

References

[1] **T. B. Cao and L. Xu,** Logarithmic difference lemma in several complex variables and partial difference equations, *Annali di Matematica Pura ed Applicata (1923-)*, **199** (2020), 767–794.

[2] **S. Chen and A. Xu,** Uniqueness of Derivatives and Shifts of Meromorphic Functions, *Comput. Methods. Funct. Theory*, 22 (2022), 197–205.

[3] **Y. M. Chiang, S. J. Feng**, On the Nevanlinna characteristic $f(z+\eta)$ and difference equation in the complex plane, *Ramanujan. J*, **16** (2008), 105-129.

[4] **K. S. Charak, R. J. Korhonen, G. Kumar,** A note on partial sharing of values of meromorphic functions with their shifts, *J. Math. Anal. Appl*, **435** (2016), 1241-1248.

[5] **R. S. Dyavanal and M. M. Mathai**, Uniqueness of Difference-Differential Polynomials of Meromorphic Functions, *Ukr Math J*, 71 (2019), 1032–1042.

[6] **G. G. Gundersen**, Estimates for the logarithmic derivative of a meromorphic function, plus similar estimates, *J. London. Math. Soc*, 37(1) (1998), 88-104 .

[7] **R. G. Halburd, R.J. Korhonen,** Nevanlinna theory for the difference operator, *Ann. Acad. Sci. Fenn. Math*, **31**(2006), 463-478.

[8] **R. G. Hulburd, R. J. Korhonen**, Difference analogue of the Lemma on the logarithmic derivative with applications to difference equations, *J. Math. Anl. Appl*, **314** (2006), 477-487.

[9] **R. G. Halburd and R. J. Korhonen**, Meromorphic solutions of difference equations, integrability and the discrete Painlev'e equations, *J. Phys. A: Math. Theor*, **40** (2007), R1–R38.

[10] **R. Halburd, R. Korhonen, K. Tohge**, Holomorphic curves with shift-invariant hyperplane preimages, *Transactions of the American Mathematical Society,* **366** (2014), 4267-4298.

[11] **J. Heittokangas, R. Korhonen, I. Laine, J. Rieppo,** Uniqueness of meromorphic functions sharing values with their shift, *Complex Variables and Elliptic Equations*, **56** (2011), 81-92.

[12] **J. Heittokangas, R. Korhonen, I. Laine, J. Rieppo, J. Zhang**, Value sharing results for shifts of meromorphic functions, and sufficient conditions for periodicity, *J. Math. Anl. Appl*, **355** (2009), 352-363.

[13] **N. Li, R. Korhonen and L. Yang**, Nevanlinna uniqueness of linear difference polynomials, *Rocky Mountain J. Math*, **47** (2017), 905-926.

[14] **W. Lin, H. Yi,** Uniqueness theorems for meromorphic functions, *Indian. J. Pure Appl. Math*, **35** (2004), 121-132.

[15] **I. Lahiri,** Weighted sharing and uniqueness of meromorphic functions, *Nagoya. J. Math*, **161** (2001), 193-206.

[16] **I. Lahiri,** Weighted value sharing and uniqueness of meromorphic functions, *Complex Variables Theory Appl*, **46** (2001), 241-253.

[17] **S. Majumder and S. Saha**, A Note on the Uniqueness of Certain Types of Differential-Difference Polynomials, *Ukr Math J*, 73 (2021), 791–810.

[18] **X. Qi and L. Yang,** Uniqueness of Meromorphic Functions Concerning their Shifts and Derivatives, *Comput. Methods. Funct. Theory*, 20 (2020), 159-178.

[19] **H. X. YI,** Uniqueness of meromorphic functions and a question of C.C. Yang, *Complex Variables*, **14** (1990), 169-176.

[20] **X. Zhang,** Value sharing of meromorphic functions and some questions of Dyavanal, *Front. Math. China*, **7** (2012), 161-176.

[21] **J. Zhang, L. Z. Yang,** Some results related to a conjecture of R. Brück, *J. Inequal. Pure Appl. Math*, 8 (2007), Article 18.

Ha Tran Phuong
Thai Nguyen University of Education
Thai Nguyen
Vietnam
`hatranphuong@yahoo.com` and `phuonght@tnue.edu.vn`

Nguyen Van Thin
Thai Nguyen University of Education
Thai Nguyen
Vietnam
`thinmath@gmail.com` and `thinnv@tnue.edu.vn`

J. Math. Math. Sci. (2025) 55–76

An algorithm for solving the variational inequality problem over the solution set of the split variational inequality and fixed point problem

Tran Viet Anh (Hanoi, Vietnam)

(Received Mar. 17, 2025; accepted Apr. 16, 2025)

Abstract. In this paper, we introduce a new algorithm for solving strongly monotone variational inequality problem, where the constraint set is the solution set of the split variational inequality and fixed point problem. Our method uses dynamic step sizes selected based on information of the previous step, which gives strong convergence result without the prior knowledge of the given bounded linear operator's norm. In addition, using our method, we do not require any information of the Lipschitz and strongly monotone constants of the mappings. Several corollaries of our main result are also presented. Finally, a numerical example has been given to illustrate the effectiveness of our proposed algorithm.

1. Introduction

Consider two real Hilbert spaces, denoted as $\mathcal{H}_1$ and $\mathcal{H}_2$, with a bounded linear operator $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$. Let C be a nonempty closed convex subset of $\mathcal{H}_1$. Additionally, let $F : \mathcal{H}_1 \longrightarrow \mathcal{H}_1$ and $T : \mathcal{H}_2 \longrightarrow \mathcal{H}_2$ be given mappings. The Split Variational Inequality and Fixed Point Problem (SVIFPP) aim to find a solution x^* in the space $\mathcal{H}_1$ for which the image $A(x^*)$, under the operator A, serves as a fixed point for another mapping in $\mathcal{H}_2$.

To be more specific, the SVIFPP can be formulated as follows:

$$\text{Find } x^* \in C : \langle F(x^*), x - x^* \rangle \geq 0 \ \ \forall x \in C \tag{1.1}$$

Key words and phrases: Variational inequality, split variational inequality and fixed point problem, pseudomonotone mapping, demicontractive mapping.
2020 Mathematics Subject Classification: 49J40, 90C33, 47H17.

such that

$$T(A(x^*)) = A(x^*). \tag{1.2}$$

A particular instance of the SVIFPP, denoted by equations (1.1)-(1.2) with $F = 0$ and $T = P_Q$, corresponds to the Split Feasibility Problem (SFP). In short, the SFP can be stated as follows:

$$\text{Find } x^* \in C \text{ such that } A(x^*) \in Q, \tag{1.3}$$

where C and Q are two nonempty closed convex subsets of real Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Recently, it has been demonstrated that the SFP can serve as a practical model in intensity-modulated radiation therapy [10, 11, 13] and in various other real-world applications. To solve the SFP and their generalizations, numerous iterative projection methods have been developed. For more details, see [1–9, 12–16, 18, 21, 23, 24] and the references therein.

To find a specific solution to the SVIFPP, Hai et al. [14] investigated the following variational inequality problem

$$\text{Find } x^* \in \Omega_{\text{SVIFPP}} \text{ such that } \langle S(x^*), x - x^* \rangle \geq 0 \ \forall x \in \Omega_{\text{SVIFPP}}, \tag{1.4}$$

where $S : \mathcal{H}_1 \longrightarrow \mathcal{H}_1$ is η-strongly monotone and κ-Lipschitz continuous on $\mathcal{H}_1$, $F : \mathcal{H}_1 \longrightarrow \mathcal{H}_1$ is pseudomonotone on C and L-Lipschitz continuous on $\mathcal{H}_1$, $T : \mathcal{H}_2 \longrightarrow \mathcal{H}_2$ is γ-demicontractive and demi-closed at zero, $\Omega_{\text{SVIFPP}} = \{x^* \in \text{Sol}(C, F) : A(x^*) \in \text{Fix}(T)\}$ defines the solution set of the SVIFPP. As detailed in [14], the authors recommended the subgradient extragradient method to solve problem (1.4) (refer to Algorithm 1 in [14])

$$\begin{cases} x^0 \in \mathcal{H}_1, \\ u^n = A(x^n), \\ v^n = T(u^n), \\ y^n = x^n + \delta_n A^*(v^n - u^n), \\ z^n = P_C(y^n - \mu_n F(y^n)), \\ t^n = P_{C_n}(y^n - \mu_n F(z^n)), \\ x^{n+1} = t^n - \varepsilon_n S(t^n) \end{cases} \tag{1.5}$$

where $C_n = \{\omega \in \mathcal{H}_1 : \langle y^n - \mu_n F(y^n) - z^n, \omega - z^n \rangle \leq 0\}$, $\{\delta_n\} \subset [\underline{\delta}, \overline{\delta}] \subset \Big(0, \dfrac{1-\gamma}{\|A\|^2 + 1}\Big)$, $\{\mu_n\} \subset [a, b] \subset \Big(0, \dfrac{1}{L}\Big)$, $\{\varepsilon_n\} \subset (0, 1)$, $\lim\limits_{n\to\infty} \varepsilon_n = 0$ and $\sum\limits_{n=0}^{\infty} \varepsilon_n = \infty$. In [14], the authors proved that the sequence $\{x^n\}$, generated by (1.5), converges strongly to the unique solution x^* of the variational inequality problem (1.4), assuming the solution set Ω_{SVIFPP} of the SVIFPP is nonempty.

In extragradient methods, performing two projections onto the constrained set C per iteration can hinder the algorithm's efficiency. To overcome this challenge, Tseng's extragradient method [20] reduces the computational burden by performing only one projection onto C in each iteration. The formulation of Tseng's extragradient method is outlined as follows:

$$\begin{cases} x^0 \in \mathcal{H}, \\ y^n = P_C(x^n - \mu F(x^n)), \\ x^{n+1} = y^n - \mu(F(y^n) - F(x^n)), \end{cases} \tag{1.6}$$

where F is L-Lipschitz continuous, and $\mu \in \left(0, \dfrac{1}{L}\right)$. It is important to highlight that the main drawback of Algorithms (1.5) and (1.6) is the need to know the Lipschitz constants of the operator F, or at the very least, to have estimates of this parameter.

In this paper, motivated by the previously discussed works, we propose a novel algorithm designed to solve the variational inequality problem over the solution set of the split variational inequality and fixed point problem (1.4). The main contribution of the algorithm is the replacement of the subgradient extragradient method in Algorithm (1.5) with a modified version of Tseng's extragradient methods, which use self-adaptive step sizes. By implementing this modification, the need for the Lipschitz constant of the cost operator F is removed, resulting in a faster convergence rate. Additionally, our method does not require any prior information regarding the norm of the operator A.

The paper is structured as follows. Section 2 presents key definitions and preliminary results, which are utilized in Section 3, where the algorithm is introduced, its strong convergence is established, and several corollaries are discussed. In the final section, a numerical example is provided to compare the performance of the proposed algorithm with that of Hai et al. [14].

2. Preliminaries

Let C be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$. It is well-known that for all point $x \in \mathcal{H}$, there exists a unique point $P_C(x) \in C$ such that

$$\|x - P_C(x)\| = \min\{\|x - y\| : y \in C\}. \tag{2.1}$$

The mapping $P_C : \mathcal{H} \longrightarrow C$ defined by (2.1) is called the metric projection of $\mathcal{H}$ onto C. Notably, P_C is nonexpansive. Additionally, the following inequality holds for all for all $x \in \mathcal{H}$ and $y \in C$:

$$\langle x - P_C(x), y - P_C(x)\rangle \leq 0.$$

Definition 2.1. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces and let $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ be a bounded linear operator. An operator $A^* : \mathcal{H}_2 \longrightarrow \mathcal{H}_1$ with the property $\langle A(x), y\rangle = \langle x, A^*(y)\rangle$ for all $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$, is called an adjoint operator.*

The adjoint operator of a bounded linear operator A between Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$ always exists and is uniquely determined. Additionally, A^* is a bounded linear operator and the equality $\|A^*\| = \|A\|$ holds true.

Definition 2.2 (see [17])**.** *A mapping $S : \mathcal{H} \longrightarrow \mathcal{H}$ is said to be*

(i) η-strongly monotone on $\mathcal{H}$ if there exists $\eta > 0$ such that

$$\langle S(x) - S(y), x - y\rangle \geq \eta\|x - y\|^2 \ \ \forall x, y \in \mathcal{H};$$

(ii) κ-Lipschitz continuous on $\mathcal{H}$ if

$$\|S(x) - S(y)\| \leq \kappa\|x - y\| \ \ \forall x, y \in \mathcal{H}.$$

Definition 2.3. *A mapping $T : \mathcal{H} \longrightarrow \mathcal{H}$ is said to be*

(i) γ-demicontractive if $\mathrm{Fix}(T) \neq \emptyset$ and there exists a constant $\gamma \in [0, 1)$ such that

$$\|T(x) - x^*\|^2 \leq \|x - x^*\|^2 + \gamma\|T(x) - x\|^2 \ \ \forall x \in \mathcal{H}, \forall x^* \in \mathrm{Fix}(T);$$

(ii) demi-closed at zero if, for every sequence $\{x^n\}$ in $\mathcal{H}$, the following implication holds

$$\begin{cases} x^n \rightharpoonup x \\ \lim\limits_{n\to\infty} \|T(x^n) - x^n\| = 0 \end{cases} \Rightarrow x \in \mathrm{Fix}(T).$$

The subsequent lemmas are essential for establishing the main result in our paper.

Lemma 2.1 (see [16])**.** *Let C be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$. Let $F : \mathcal{H} \longrightarrow \mathcal{H}$ be a mapping such that $\limsup\limits_{n\to\infty}\langle F(x^n), y - y^n\rangle \leq \langle F(\overline{x}), y - \overline{y}\rangle$ for every sequences $\{x^n\}$, $\{y^n\}$ in $\mathcal{H}$ converging weakly to $\overline{x}$ and $\overline{y}$, respectively. Assume that $\mu_n \geq a > 0$ for all n, $\{x^n\}$ is a sequence in $\mathcal{H}$ satisfying $x^n \rightharpoonup \overline{x}$ and $\lim\limits_{n\to\infty} \|x^n - y^n\| = 0$, where $y^n = P_C(x^n - \mu_n F(x^n))$ for all n. Then $\overline{x} \in \mathrm{Sol}(C, F)$.*

Lemma 2.2 (see [22]). *Let $\{u_n\}$ be a sequence of nonnegative real numbers, $\{\alpha_n\}$ be a sequence in $(0,1)$ such that $\sum_{n=0}^{\infty}\alpha_n=\infty$ and $\{v_n\}$ be a sequence of real numbers with $\limsup_{n\to\infty} v_n \le 0$. Suppose that*

$$u_{n+1} \le (1-\alpha_n)u_n + \alpha_n v_n \ \forall n \ge 0.$$

Then $\lim_{n\to\infty} u_n = 0$.

Lemma 2.3 (see [19]). *Let $\{a_n\}$ be a sequence of nonnegative real numbers such that for any integer m, there exists an integer p such that $p \ge m$ and $a_p \le a_{p+1}$. Let n_0 be an integer such that $a_{n_0} \le a_{n_0+1}$ and define, for all integer $n \ge n_0$, by*

$$\tau(n) = \max\{k \in \mathbb{N} : n_0 \le k \le n, a_k \le a_{k+1}\}.$$

Then $\{\tau(n)\}_{n\ge n_0}$ is a nondecreasing sequence satisfying $\lim_{n\to\infty}\tau(n)=\infty$ and the following inequalities are satisfied:

$$a_{\tau(n)} \le a_{\tau(n)+1},\ a_n \le a_{\tau(n)+1} \ \forall n \ge n_0.$$

3. The algorithm and convergence analysis

In this section, we propose an algorithm with strong convergence for solving the problem (1.4). We specify the following assumptions related to the mappings S, F and T involved in the formulation of the problem (1.4).

(A_1): $S : \mathcal{H}_1 \longrightarrow \mathcal{H}_1$ is η-strongly monotone and κ-Lipschitz continuous on $\mathcal{H}_1$.

(A_2): $F : \mathcal{H}_1 \longrightarrow \mathcal{H}_1$ is pseudomonotone on C and L-Lipschitz continuous on $\mathcal{H}_1$.

(A_3): $\limsup_{n\to\infty}\langle F(x^n), y - y^n\rangle \le \langle F(\overline{x}), y - \overline{y}\rangle$ for every sequence $\{x^n\}$, $\{y^n\}$ in $\mathcal{H}_1$ converging weakly to $\overline{x}$ and $\overline{y}$, respectively.

(A_4): $T : \mathcal{H}_2 \longrightarrow \mathcal{H}_2$ is γ-demicontractive and demi-closed at zero.

The algorithm is presented as follows.

Algorithm 3.1.

Step 0. Choose $\mu_0 > 0$, $\mu \in (0,1)$, $\{\rho_n\} \subset [a,b] \subset \big(0, 1-\gamma\big)$, $\{\varepsilon_n\} \subset (0,1)$ such that $\lim_{n\to\infty}\varepsilon_n = 0$ and $\sum_{n=0}^{\infty}\varepsilon_n = \infty$.

Step 1. Let $x^0 \in \mathcal{H}_1$. Set $n := 0$.
Step 2. Compute $u^n = A(x^n)$, $v^n = T(u^n)$ and

$$y^n = x^n + \delta_n A^*(v^n - u^n),$$

where the step size δ_n is chosen in such a way that

$$\delta_n = \begin{cases} \dfrac{\rho_n \|v^n - u^n\|^2}{\|A^*(v^n - u^n)\|^2} & \text{if } A^*(v^n - u^n) \neq 0, \\ 0 & \text{if } A^*(v^n - u^n) = 0. \end{cases}$$

Step 3. Compute

$$\begin{aligned} z^n &= P_C(y^n - \mu_n F(y^n)), \\ t^n &= z^n - \mu_n(F(z^n) - F(y^n)), \end{aligned}$$

where

$$\mu_{n+1} = \begin{cases} \min\left\{ \dfrac{\mu \|y^n - z^n\|}{\|F(y^n) - F(z^n)\|}, \mu_n \right\} & \text{if } F(y^n) \neq F(z^n), \\ \mu_n & \text{if } F(y^n) = F(z^n). \end{cases}$$

Step 4. Compute

$$x^{n+1} = t^n - \varepsilon_n S(t^n).$$

Step 5. Set $n := n + 1$, and go to **Step 2**.

The strong convergence of the sequence generated through Algorithm 3.1 is established by the following theorem.

Theorem 3.1. *Assuming that conditions* (A_1), (A_2), (A_3) *and* (A_4) *hold, the sequence* $\{x^n\}$ *generated by Algorithm 3.1 converges strongly to the unique solution of problem* (1.4)*, provided that the solution set* Ω_{SVIFPP} *of the SVIFPP is nonempty.*

Proof. Since $\Omega_{\text{SVIFPP}} \neq \emptyset$, the problem (1.4) has a unique solution, denoted by x^*. In particular, $x^* \in \Omega_{\text{SVIFPP}}$, which implies that $x^* \in \text{Sol}(C, F)$ and $A(x^*) \in \text{Fix}(T)$. The proof of the theorem is divided into several steps.
Step 1. For all $n \geq 0$, we show that

$$\Big(1 - \mu^2 \frac{\mu_n^2}{\mu_{n+1}^2}\Big) \|y^n - z^n\|^2 \leq \|y^n - x^*\|^2 - \|t^n - x^*\|^2. \tag{3.1}$$

Given that $z^n = P_C(y^n - \mu_n F(y^n))$ and $x^* \in C$, by utilizing the properties of the projection mapping, we have

$$\langle y^n - \mu_n F(y^n) - z^n, x^* - z^n \rangle \leq 0$$

or, equivalently

$$-\langle y^n - z^n, z^n - x^*\rangle \le -\mu_n\langle F(y^n), z^n - x^*\rangle. \tag{3.2}$$

By applying the equality

$$\|y\|^2 = \|x+y\|^2 - \|x\|^2 - 2\langle x, y\rangle \ \ \forall x, y \in \mathcal{H}_1$$

and taking (3.2) into consideration, we derive

$$\begin{aligned} \|z^n - x^*\|^2 &= \|(y^n - z^n) + (z^n - x^*)\|^2 - \|y^n - z^n\|^2 - 2\langle y^n - z^n, z^n - x^*\rangle \\ &\le \|y^n - x^*\|^2 - \|y^n - z^n\|^2 - 2\mu_n\langle F(y^n), z^n - x^*\rangle. \end{aligned} \tag{3.3}$$

Since $x^* \in \mathrm{Sol}(C, F)$, it follows that $\langle F(x^*), z - x^*\rangle \ge 0$ for all $z \in C$. By applying the pseudomonotonicity of F on C, we deduce that $\langle F(z), z - x^*\rangle \ge 0$ for all $z \in C$. Taking $z = z^n \in C$, we obtain

$$\langle F(z^n), z^n - x^*\rangle \ge 0. \tag{3.4}$$

From the definition of μ_{n+1}, it follows that

$$\|F(y^n) - F(z^n)\| \le \frac{\mu}{\mu_{n+1}}\|y^n - z^n\|. \tag{3.5}$$

Indeed, if $F(y^n) = F(z^n)$, then the inequality (3.5) is satisfied. Otherwise, we derive the following

$$\mu_{n+1} = \min\left\{\frac{\mu\|y^n - z^n\|}{\|F(y^n) - F(z^n)\|}, \mu_n\right\} \le \frac{\mu\|y^n - z^n\|}{\|F(y^n) - F(z^n)\|},$$

which implies (3.5).
From (3.3), (3.4) and (3.5), we obtain

$$\begin{aligned} \|t^n - x^*\|^2 &= \|z^n - x^* - \mu_n(F(z^n) - F(y^n))\|^2 \\ &= \|z^n - x^*\|^2 - 2\mu_n\langle F(z^n) - F(y^n), z^n - x^*\rangle \\ &\qquad + \mu_n^2\|F(z^n) - F(y^n)\|^2 \\ &\le \|y^n - x^*\|^2 - \|y^n - z^n\|^2 - 2\mu_n\langle F(z^n), z^n - x^*\rangle \\ &\qquad + \mu_n^2\|F(z^n) - F(y^n)\|^2 \\ &\le \|y^n - x^*\|^2 - \|y^n - z^n\|^2 + \mu_n^2\|F(z^n) - F(y^n)\|^2 \\ &\le \|y^n - x^*\|^2 - \Big(1 - \mu^2\frac{\mu_n^2}{\mu_{n+1}^2}\Big)\|y^n - z^n\|^2. \end{aligned}$$

As a result, we get

$$\Big(1 - \mu^2\frac{\mu_n^2}{\mu_{n+1}^2}\Big)\|y^n - z^n\|^2 \le \|y^n - x^*\|^2 - \|t^n - x^*\|^2 \ \ \forall n \ge 0.$$

Step 2. For all $n \geq 0$, we have

$$\langle x^n - x^*, A^*(v^n - u^n)\rangle \leq -\frac{1-\gamma}{2}\|v^n - u^n\|^2. \tag{3.6}$$

Thanks to the γ-demicontractivity of T, we get

$$\begin{aligned}
&\langle x^n - x^*, A^*(v^n - u^n)\rangle \\
&\quad = \langle A(x^n - x^*), v^n - u^n\rangle \\
&\quad = \langle v^n - A(x^*), v^n - u^n\rangle - \|v^n - u^n\|^2 \\
&\quad = \frac{1}{2}\left[\left(\|v^n - A(x^*)\|^2 - \|u^n - A(x^*)\|^2\right) - \|v^n - u^n\|^2\right] \\
&\quad = \frac{1}{2}\left[\left(\|T(u^n) - A(x^*)\|^2 - \|u^n - A(x^*)\|^2\right) - \|v^n - u^n\|^2\right] \\
&\quad \leq \frac{1}{2}\left[\gamma\|T(u^n) - u^n\|^2 - \|v^n - u^n\|^2\right] \\
&\quad = -\frac{1-\gamma}{2}\|v^n - u^n\|^2.
\end{aligned}$$

Step 3. We show that

$$\mu_{n+1} \leq \mu_n,\ \mu_n \geq \min\left(\frac{\mu}{L}, \mu_0\right) \quad \forall n \geq 0, \lim_{n\to\infty} \mu_n = \mu^* \geq \min\left(\frac{\mu}{L}, \mu_0\right). \tag{3.7}$$

Since F is L-Lipschitz continuous on $\mathcal{H}_1$, we have

$$\|F(y^n) - F(z^n)\| \leq L\|y^n - z^n\|.$$

Thus, when $F(y^n) \neq F(z^n)$, it follows that

$$\frac{\mu\|y^n - z^n\|}{\|F(y^n) - F(z^n)\|} \geq \frac{\mu}{L}.$$

By induction, we obtain

$$\mu_n \geq \min\left(\frac{\mu}{L}, \mu_0\right) \quad \forall n \geq 0.$$

From the definition of μ_{n+1}, it is clear that $\mu_{n+1} \leq \mu_n$ for all $n \geq 0$. Therefore, together with the fact that $\mu_n \geq \min\left(\frac{\mu}{L}, \mu_0\right)$ for all $n \geq 0$, it follows that the sequence $\{\mu_n\}$ has a limit, denoted by μ^*, and we conclude that $\lim_{n\to\infty} \mu_n = \mu^* \geq \min\left(\frac{\mu}{L}, \mu_0\right)$.

Step 4. We show that, for all $n \geq 0$

$$\begin{aligned}
&\frac{a^2}{(\|A\|+1)^2}\|v^n - u^n\|^2 \leq \|y^n - x^n\|^2, \\
&\|y^n - x^n\|^2 \leq \frac{b}{1-\gamma-b}\left(\|x^n - x^*\|^2 - \|y^n - x^*\|^2\right).
\end{aligned} \tag{3.8}$$

We now consider two distinct cases.

Case 1. $A^*(v^n - u^n) = 0$. From (3.6), we deduce that $\|v^n - u^n\| = 0$. Since $\delta_n = 0$, it follows that $y^n = x^n$. Therefore, (3.8) holds.

Case 2. $A^*(v^n - u^n) \neq 0$. It follows from (3.6) that

$$\begin{aligned}
\|y^n - x^*\|^2 &= \|(x^n - x^*) + \delta_n A^*(v^n - u^n)\|^2 \\
&= \|x^n - x^*\|^2 + \|\delta_n A^*(v^n - u^n)\|^2 + 2\delta_n \langle x^n - x^*, A^*(v^n - u^n)\rangle \\
&\le \|x^n - x^*\|^2 + \delta_n^2 \|A^*(v^n - u^n)\|^2 - \delta_n(1-\gamma)\|v^n - u^n\|^2 \\
&= \|x^n - x^*\|^2 - \frac{\rho_n^2 \|v^n - u^n\|^4}{\|A^*(v^n - u^n)\|^2} \cdot \frac{1-\gamma-\rho_n}{\rho_n} \\
&\le \|x^n - x^*\|^2 - \frac{\rho_n^2 \|v^n - u^n\|^4}{\|A^*(v^n - u^n)\|^2} \cdot \frac{1-\gamma-b}{b} \quad \forall n \ge 0.
\end{aligned} \tag{3.9}$$

By applying (3.9), we get

$$\begin{aligned}
\|y^n - x^n\|^2 &= \delta_n^2 \|A^*(v^n - u^n)\|^2 \\
&= \frac{\rho_n^2 \|v^n - u^n\|^4}{\|A^*(v^n - u^n)\|^4} \|A^*(v^n - u^n)\|^2 \\
&= \frac{\rho_n^2 \|v^n - u^n\|^4}{\|A^*(v^n - u^n)\|^2} \\
&\le \frac{b}{1-\gamma-b} \left(\|x^n - x^*\|^2 - \|y^n - x^*\|^2\right) \quad \forall n \ge 0.
\end{aligned} \tag{3.10}$$

On the other hand

$$\|A^*(v^n - u^n)\| \le \|A^*\|\|v^n - u^n\| = \|A\|\|v^n - u^n\| \le (\|A\|+1)\|v^n - u^n\|. \tag{3.11}$$

By using (3.10) and (3.11) together, we obtain

$$\|y^n - x^n\|^2 \ge \frac{\rho_n^2 \|v^n - u^n\|^4}{(\|A\|+1)^2 \|v^n - u^n\|^2} \ge \frac{a^2}{(\|A\|+1)^2} \|v^n - u^n\|^2 \quad \forall n \ge 0.$$

Therefore, the inequalities in (3.8) are proven.

Now, choose $\varepsilon \in \left(0, \frac{2\eta}{\kappa^2}\right)$. From $\lim_{n\to\infty} \varepsilon_n = 0$ and $\lim_{n\to\infty}\left(1-\mu^2 \frac{\mu_n^2}{\mu_{n+1}^2}\right) = 1-\mu^2 > 0$, there exists $n_0 \in \mathbb{N}$ such that

$$\varepsilon_n < \varepsilon \quad \forall n \ge n_0, \quad 1-\mu^2 \frac{\mu_n^2}{\mu_{n+1}^2} > \frac{1-\mu^2}{2} > 0 \quad \forall n \ge n_0. \tag{3.12}$$

Step 5. For all $n \ge n_0$, we show that

$$\|t^n - \varepsilon_n S(t^n) - x^* + \varepsilon_n S(x^*)\| \le \left(1 - \frac{\varepsilon_n \tau}{\varepsilon}\right)\|t^n - x^*\|, \tag{3.13}$$

where $\tau = 1 - \sqrt{1-\varepsilon(2\eta - \varepsilon\kappa^2)} \in (0,1]$.
Given the κ-Lipschitz continuity and η-strong monotonicity of S on $\mathcal{H}_1$, we deduce

$$\begin{aligned}
\|t^n - x^* - &\varepsilon(S(t^n) - S(x^*))\|^2 \\
&= \|t^n - x^*\|^2 - 2\varepsilon\langle t^n - x^*, S(t^n) - S(x^*)\rangle + \varepsilon^2\|S(t^n) - S(x^*)\|^2 \\
&\le \|t^n - x^*\|^2 - 2\varepsilon\eta\|t^n - x^*\|^2 + \varepsilon^2\kappa^2\|t^n - x^*\|^2 \\
&= \left[1 - \varepsilon(2\eta - \varepsilon\kappa^2)\right]\|t^n - x^*\|^2.
\end{aligned}$$

From (3.12) and the inequality above, it follows that

$$\begin{aligned}
\|t^n - &\varepsilon_n S(t^n) - x^* + \varepsilon_n S(x^*)\| \\
&= \left\| \left(1 - \frac{\varepsilon_n}{\varepsilon}\right)(t^n - x^*) + \frac{\varepsilon_n}{\varepsilon}\left[t^n - x^* - \varepsilon(S(t^n) - S(x^*))\right] \right\| \\
&\le \left(1 - \frac{\varepsilon_n}{\varepsilon}\right)\|t^n - x^*\| + \frac{\varepsilon_n}{\varepsilon}\|t^n - x^* - \varepsilon(S(t^n) - S(x^*))\| \\
&\le \left(1 - \frac{\varepsilon_n}{\varepsilon}\right)\|t^n - x^*\| + \frac{\varepsilon_n}{\varepsilon}\sqrt{1 - \varepsilon(2\eta - \varepsilon\kappa^2)}\|t^n - x^*\| \\
&= \left[1 - \frac{\varepsilon_n}{\varepsilon}\left(1 - \sqrt{1 - \varepsilon(2\eta - \varepsilon\kappa^2)}\right)\right]\|t^n - x^*\| \\
&= \left(1 - \frac{\varepsilon_n\tau}{\varepsilon}\right)\|t^n - x^*\| \;\; \forall n \ge n_0.
\end{aligned}$$

Step 6. The sequences $\{x^n\}$, $\{y^n\}$, $\{t^n\}$ and $\{S(t^n)\}$ are bounded.
From inequality (3.13), we obtain

$$\begin{aligned}
\|x^{n+1} - x^*\| &= \|t^n - \varepsilon_n S(t^n) - x^* + \varepsilon_n S(x^*) - \varepsilon_n S(x^*)\| \\
&\le \|t^n - \varepsilon_n S(t^n) - x^* + \varepsilon_n S(x^*)\| + \varepsilon_n\|S(x^*)\| \\
&\le \left(1 - \frac{\varepsilon_n\tau}{\varepsilon}\right)\|t^n - x^*\| + \varepsilon_n\|S(x^*)\| \;\; \forall n \ge n_0. \qquad (3.14)
\end{aligned}$$

Using (3.1), (3.8) and (3.12), we get

$$\|t^n - x^*\| \le \|y^n - x^*\| \le \|x^n - x^*\| \;\; \forall n \ge n_0. \qquad (3.15)$$

By applying (3.14) and (3.15), we derive

$$\begin{aligned}
\|x^{n+1} - x^*\| &\le \left(1 - \frac{\varepsilon_n\tau}{\varepsilon}\right)\|x^n - x^*\| + \varepsilon_n\|S(x^*)\| \\
&= \left(1 - \frac{\varepsilon_n\tau}{\varepsilon}\right)\|x^n - x^*\| + \frac{\varepsilon_n\tau}{\varepsilon} \cdot \frac{\varepsilon\|S(x^*)\|}{\tau} \;\; \forall n \ge n_0.
\end{aligned}$$

In particular,

$$\|x^{n+1} - x^*\| \le \max\left\{\|x^n - x^*\|, \frac{\varepsilon\|S(x^*)\|}{\tau}\right\} \;\; \forall n \ge n_0,$$

and thus, by induction, we have

$$\|x^n - x^*\| \le \max\left\{\|x^{n_0} - x^*\|, \frac{\varepsilon\|S(x^*)\|}{\tau}\right\} \ \forall n \ge n_0.$$

Therefore, the sequence $\{x^n\}$ is bounded and this is true for the sequences $\{y^n\}$, $\{t^n\}$ and $\{S(t^n)\}$ as well, thanks to (3.15) and the Lipschitz continuity of S.

Step 7. We prove that $\{x^n\}$ converges strongly to x^*.

Based on (3.13), we deduce, for every $n \ge n_0$, that

$$\begin{aligned}
\|x^{n+1} - x^*\|^2 &\le \|x^{n+1} - x^*\|^2 + \varepsilon_n^2\|S(x^*)\|^2 \\
&= \|x^{n+1} - x^* + \varepsilon_n S(x^*)\|^2 - 2\langle\varepsilon_n S(x^*), x^{n+1} - x^*\rangle \\
&= \|t^n - \varepsilon_n S(t^n) - x^* + \varepsilon_n S(x^*)\|^2 - 2\varepsilon_n\langle S(x^*), x^{n+1} - x^*\rangle \\
&\le \left[\left(1 - \frac{\varepsilon_n\tau}{\varepsilon}\right)\|t^n - x^*\|\right]^2 - 2\varepsilon_n\langle S(x^*), x^{n+1} - x^*\rangle \\
&\le \left(1 - \frac{\varepsilon_n\tau}{\varepsilon}\right)\|t^n - x^*\|^2 - 2\varepsilon_n\langle S(x^*), x^{n+1} - x^*\rangle. \qquad (3.16)
\end{aligned}$$

We will consider two cases.

Case 1. Let us consider the case where there exists n_* such that $\{\|x^n - x^*\|\}$ is decreasing for $n \ge n_*$. As a result, the limit of $\{\|x^n - x^*\|\}$ exists. Consequently, from (3.15) and (3.16), we deduce, for all $n \ge n_0$, that

$$\begin{aligned}
0 &\le \|y^n - x^*\|^2 - \|t^n - x^*\|^2 \\
&\le \|x^n - x^*\|^2 - \|t^n - x^*\|^2 \\
&\le \left(\|x^n - x^*\|^2 - \|x^{n+1} - x^*\|^2\right) - 2\varepsilon_n\langle S(x^*), x^{n+1} - x^*\rangle.
\end{aligned}$$

Given that $\|x^n - x^*\|$ has a limit, with $\lim_{n\to\infty} \varepsilon_n = 0$, and the sequence $\{x^n\}$ is bounded, the above inequalities yield that

$$\lim_{n\to\infty} (\|y^n - x^*\|^2 - \|t^n - x^*\|^2) = 0, \qquad (3.17)$$

$$\lim_{n\to\infty} (\|x^n - x^*\|^2 - \|t^n - x^*\|^2) = 0. \qquad (3.18)$$

It follows from (3.1), (3.12) and (3.17) that

$$\lim_{n\to\infty} \|y^n - z^n\| = 0. \qquad (3.19)$$

From (3.17) and (3.18), we have

$$\lim_{n\to\infty} (\|x^n - x^*\|^2 - \|y^n - x^*\|^2) = 0. \qquad (3.20)$$

Consequently, from (3.8) and (3.20), we get

$$\lim_{n\to\infty} \|y^n - x^n\| = 0, \tag{3.21}$$

$$\lim_{n\to\infty} \|v^n - u^n\| \Rightarrow \lim_{n\to\infty} \|T(u^n) - u^n\| = 0. \tag{3.22}$$

Applying the triangle inequality along with the L-Lipschitz continuity of F on $\mathcal{H}_1$, we have

$$\begin{aligned}\|x^n - t^n\| &\le \|x^n - y^n\| + \|y^n - z^n\| + \|z^n - t^n\| \\ &= \|x^n - y^n\| + \|y^n - z^n\| + \|\mu_n(F(z^n) - F(y^n))\| \\ &\le \|x^n - y^n\| + \|y^n - z^n\| + \mu_n L\|z^n - y^n\| \\ &\le \|x^n - y^n\| + (1 + \mu_0 L)\|y^n - z^n\|.\end{aligned}$$

Therefore, using (3.19) and (3.21), it follows that

$$\lim_{n\to\infty} \|x^n - t^n\| = 0. \tag{3.23}$$

Now, we prove that

$$\limsup_{n\to\infty} \langle S(x^*), x^* - x^{n+1}\rangle \le 0. \tag{3.24}$$

Choose a subsequence $\{x^{n_\nu}\}$ from $\{x^n\}$ such that

$$\limsup_{n\to\infty} \langle S(x^*), x^* - x^{n+1}\rangle = \lim_{\nu\to\infty} \langle S(x^*), x^* - x^{n_\nu}\rangle.$$

As $\{x^{n_\nu}\}$ is bounded, we can assume without loss of generality that $x^{n_\nu} \rightharpoonup \overline{x}$. Hence

$$\limsup_{n\to\infty} \langle S(x^*), x^* - x^{n+1}\rangle = \langle S(x^*), x^* - \overline{x}\rangle. \tag{3.25}$$

Using the weak convergence $x^{n_\nu} \rightharpoonup \overline{x}$ and (3.21), we infer $y^{n_\nu} \rightharpoonup \overline{x}$. From (3.19), we have $\lim_{\nu\to\infty} \|y^{n_\nu} - z^{n_\nu}\| = 0$. Since $z^{n_\nu} = P_C(y^{n_\nu} - \mu_{n_\nu} F(y^{n_\nu}))$, $y^{n_\nu} \rightharpoonup \overline{x}$, $\mu_{n_\nu} \ge \min\left(\frac{\mu}{L}, \mu_0\right) > 0$. By Lemma 2.1, we obtain $\overline{x} \in \mathrm{Sol}(C, F)$. From $x^{n_\nu} \rightharpoonup \overline{x}$, we imply $u^{n_\nu} = A(x^{n_\nu}) \rightharpoonup A(\overline{x})$. Together with (3.22) and the demiclosedness of T, it follows that $A(\overline{x}) \in \mathrm{Fix}(T)$. Taking into account that $\overline{x} \in \mathrm{Sol}(C, F)$, we conclude that $\overline{x} \in \Omega_{\mathrm{SVIFPP}}$. Consequently, $\langle S(x^*), \overline{x} - x^*\rangle \ge 0$, and combined with (3.25), this gives (3.24).
By applying (3.15) and (3.16), we get

$$\|x^{n+1} - x^*\|^2 \le \Big(1 - \frac{\varepsilon_n \tau}{\varepsilon}\Big)\|x^n - x^*\|^2 + \frac{\varepsilon_n \tau}{\varepsilon} b_n \;\; \forall n \ge n_0, \tag{3.26}$$

where

$$b_n = \frac{2\varepsilon\langle S(x^*), x^* - x^{n+1}\rangle}{\tau}.$$

Using (3.24), we conclude that $\limsup_{n\to\infty} b_n \le 0$. Since $\varepsilon_n < \varepsilon$ $\forall n \ge n_0$ and $0 < \tau \le 1$, it follows that $\left\{\frac{\varepsilon_n \tau}{\varepsilon}\right\}_{n\ge n_0} \subset (0,1)$. As a result, from (3.26), $\sum_{n=0}^{\infty} \varepsilon_n = \infty$, $\limsup_{n\to\infty} b_n \le 0$ and Lemma 2.2, we deduce that $\lim_{n\to\infty} \|x^n - x^*\|^2 = 0$, which implies $x^n \to x^*$ as $n \to \infty$.

Case 2. Assume that for every integer m, there exists an integer n such that $n \ge m$ and $\|x^n - x^*\| \le \|x^{n+1} - x^*\|$. By applying Lemma 2.3, we can define a nondecreasing sequence $\{\tau(n)\}_{n\ge N}$ of $\mathbb{N}$ such that $\lim_{n\to\infty} \tau(n) = \infty$ and the following inequalities hold

$$\|x^{\tau(n)} - x^*\| \le \|x^{\tau(n)+1} - x^*\|, \quad \|x^n - x^*\| \le \|x^{\tau(n)+1} - x^*\| \quad \forall n \ge N. \tag{3.27}$$

Select $n_* \ge N$ such that $\tau(n) \ge n_0$ for all $n \ge n_*$. Using (3.27) and (3.14), we get

$$\begin{aligned}\|x^{\tau(n)} - x^*\| &\le \|x^{\tau(n)+1} - x^*\| \\ &\le \left(1 - \frac{\varepsilon_{\tau(n)}\tau}{\varepsilon}\right)\|t^{\tau(n)} - x^*\| + \varepsilon_{\tau(n)}\|S(x^*)\| \\ &\le \|t^{\tau(n)} - x^*\| + \varepsilon_{\tau(n)}\|S(x^*)\| \quad \forall n \ge n_*,\end{aligned}$$

which together with (3.15) implies, for all $n \ge n_*$, that

$$\begin{aligned}0 &\le \|y^{\tau(n)} - x^*\| - \|t^{\tau(n)} - x^*\| \\ &\le \|x^{\tau(n)} - x^*\| - \|t^{\tau(n)} - x^*\| \le \varepsilon_{\tau(n)}\|S(x^*)\|.\end{aligned} \tag{3.28}$$

Then, it follows from (3.28) and $\lim_{n\to\infty} \varepsilon_n = 0$ that

$$\begin{aligned}&\lim_{n\to\infty} (\|y^{\tau(n)} - x^*\| - \|t^{\tau(n)} - x^*\|) = 0, \\ &\lim_{n\to\infty} (\|x^{\tau(n)} - x^*\| - \|t^{\tau(n)} - x^*\|) = 0.\end{aligned} \tag{3.29}$$

Using (3.29) and the fact that the sequences $\{x^n\}$, $\{y^n\}$ and $\{t^n\}$ are bounded, we derive

$$\begin{aligned}&\lim_{n\to\infty} (\|y^{\tau(n)} - x^*\|^2 - \|t^{\tau(n)} - x^*\|^2) = 0, \\ &\lim_{n\to\infty} (\|x^{\tau(n)} - x^*\|^2 - \|t^{\tau(n)} - x^*\|^2) = 0.\end{aligned}$$

Applying the same reasoning as in the first case, it follows that

$$\lim_{n\to\infty} \|x^{\tau(n)} - t^{\tau(n)}\| = 0, \quad \limsup_{n\to\infty} \langle S(x^*), x^* - x^{\tau(n)}\rangle \le 0. \tag{3.30}$$

We now observe that

$$\begin{aligned}\|x^{\tau(n)+1} - x^{\tau(n)}\| &= \|t^{\tau(n)} - x^{\tau(n)} - \varepsilon_{\tau(n)} S(t^{\tau(n)})\| \\ &\leq \|t^{\tau(n)} - x^{\tau(n)}\| + \varepsilon_{\tau(n)} \|S(t^{\tau(n)})\|,\end{aligned}$$

which, in combination with (3.30), $\lim_{n\to\infty} \varepsilon_n = 0$ and the boundedness of $\{S(t^{\tau(n)})\}$, implies

$$\lim_{n\to\infty} \|x^{\tau(n)+1} - x^{\tau(n)}\| = 0. \tag{3.31}$$

Using (3.31) along with the Cauchy-Schwarz inequality, we get

$$\lim_{n\to\infty} \langle S(x^*), x^{\tau(n)} - x^{\tau(n)+1}\rangle = 0. \tag{3.32}$$

By combining (3.32) and (3.30), we conclude that

$$\begin{aligned}&\limsup_{n\to\infty} \langle S(x^*), x^* - x^{\tau(n)+1}\rangle \\ &\qquad= \limsup_{n\to\infty} \left[\langle S(x^*), x^* - x^{\tau(n)}\rangle + \langle S(x^*), x^{\tau(n)} - x^{\tau(n)+1}\rangle\right] \\ &\qquad= \limsup_{n\to\infty} \langle S(x^*), x^* - x^{\tau(n)}\rangle \leq 0. \qquad (3.33)\end{aligned}$$

Also, from (3.16) and (3.15), we get

$$\|x^{n+1} - x^*\|^2 \leq \Big(1 - \frac{\varepsilon_n \tau}{\varepsilon}\Big)\|x^n - x^*\|^2 + 2\varepsilon_n \langle S(x^*), x^* - x^{n+1}\rangle \ \ \forall n \geq n_0, \tag{3.34}$$

Since $\tau(n) \geq n_0$ holds for all $n \geq n_*$, we can conclude from (3.34) and (3.27) that for all $n \geq n_*$

$$\begin{aligned}\|x^{\tau(n)+1} - x^*\|^2 &\leq \Big(1 - \frac{\varepsilon_{\tau(n)}\tau}{\varepsilon}\Big)\|x^{\tau(n)} - x^*\|^2 + 2\varepsilon_{\tau(n)} \langle S(x^*), x^* - x^{\tau(n)+1}\rangle \\ &\leq \Big(1 - \frac{\varepsilon_{\tau(n)}\tau}{\varepsilon}\Big)\|x^{\tau(n)+1} - x^*\|^2 + 2\varepsilon_{\tau(n)} \langle S(x^*), x^* - x^{\tau(n)+1}\rangle.\end{aligned}$$

As a result, since $\varepsilon_{\tau(n)} > 0$

$$\|x^{\tau(n)+1} - x^*\|^2 \leq \frac{2\varepsilon}{\tau} \langle S(x^*), x^* - x^{\tau(n)+1}\rangle \ \ \forall n \geq n_*.$$

By combining this inequality with (3.27), given that $n_* \geq N$, we have

$$\|x^n - x^*\|^2 \leq \frac{2\varepsilon}{\tau} \langle S(x^*), x^* - x^{\tau(n)+1}\rangle \ \ \forall n \geq n_*. \tag{3.35}$$

Taking the limit in (3.35) as $n \to \infty$ and applying (3.33), we arrive at

$$\limsup_{n\to\infty} \|x^n - x^*\|^2 \leq 0.$$

Therefore, it follows that $x^n \to x^*$ as $n \to \infty$. This completes the proof of Theorem 3.1. ■

Remark 3.1. We highlight the advantages of Algorithm 3.1 compared to the algorithm of Hai et al. in [14, Algorithm 1].

i) In Algorithm 3.1, unlike the result in [14, Algorithm 1], the step size is selected in such a way that its implementation does not require any prior knowledge of the norms of the given bounded linear operators.

ii) Algorithm 1 in [14] requires computing or estimating the Lipschitz constant of the mapping F, which is generally a challenging task in practice. In contrast, our Algorithm 3.1 removes this restriction.

When F is set to zero and T is defined as P_Q, the SVIFPP described by equations (1.1)-(1.2) reduces to the SFP given in (1.3). Consequently, utilizing the results from Algorithm 1 and Theorem 3.1, we derive the following result for solving the variational inequality problem over the solution set of the SFP. It is important to note that the proposed algorithm requires only two projections per iteration, and notably, its implementation does not rely on any information about the norm of the operator A.

Algorithm 3.2.
Step 0. Choose $\{\rho_n\} \subset [a, b] \subset (0, 1)$, $\{\varepsilon_n\} \subset (0, 1)$ such that $\lim_{n\to\infty} \varepsilon_n = 0$ and $\sum_{n=0}^{\infty} \varepsilon_n = \infty$.
Step 1. Let $x^0 \in \mathcal{H}_1$. Set $n := 0$.
Step 2. Compute $u^n = A(x^n)$, $v^n = P_Q(u^n)$ and

$$y^n = x^n + \delta_n A^*(v^n - u^n),$$

where the stepsize δ_n is chosen in such a way that

$$\delta_n = \begin{cases} \dfrac{\rho_n \|v^n - u^n\|^2}{\|A^*(v^n - u^n)\|^2} & \text{if } A^*(v^n - u^n) \neq 0, \\ 0 & \text{if } A^*(v^n - u^n) = 0. \end{cases}$$

Step 3. Compute $z^n = P_C(y^n)$.
Step 4. Compute

$$x^{n+1} = z^n - \varepsilon_n S(z^n).$$

Step 5. Set $n := n + 1$, and go to **Step 2**.

Corollary 3.1. *Let C and Q be two nonempty closed convex subsets of two real Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, and let $S : \mathcal{H}_1 \longrightarrow \mathcal{H}_1$ be a strongly monotone and Lipschitz continuous mapping. Suppose that the solution set $\Omega_{SFP} = \{x^* \in C : A(x^*) \in Q\}$ of the SFP is nonempty. Then the sequence*

$\{x^n\}$ generated by Algorithm 3.2 converges strongly to $x^* \in \Omega_{SFP}$, which is the unique solution of the variational inequality problem

$$\langle S(x^*), x - x^* \rangle \geq 0 \ \forall x \in \Omega_{SFP}, \tag{3.36}$$

provided that the solution set Ω_{SFP} of the SFP is nonempty.

Assume the following conditions to be satisfied:
(B_1): $S : \mathcal{H} \longrightarrow \mathcal{H}$ is strongly monotone and Lipschitz continuous on $\mathcal{H}$.
(B_2): $F : \mathcal{H} \longrightarrow \mathcal{H}$ is pseudomonotone on C and Lipschitz continuous on $\mathcal{H}$.
(B_3): $\limsup_{n\to\infty} \langle F(x^n), y - y^n \rangle \leq \langle F(\overline{x}), y - \overline{y} \rangle$ for every sequence $\{x^n\}$, $\{y^n\}$ in $\mathcal{H}$ converging weakly to $\overline{x}$ and $\overline{y}$, respectively.

When $\mathcal{H}_1 = \mathcal{H}_2 := \mathcal{H}$, and both T and A are the identity mappings in $\mathcal{H}$, the SVIFPP reduces to the variational inequality problem (1.1). Consequently, by applying Algorithm 3.1 and utilizing Theorem 3.1, we obtain the following result for solving the variational inequality problem over the solution set of another VIP. It is important to emphasize that the proposed algorithm requires only one projection onto the feasible set at each iteration, and its implementation does not require any information about the Lipschitz constants of the mappings S and F, nor the modulus of strong monotonicity of S.

Algorithm 3.3.
Step 0. Choose $\mu_0 > 0$, $\mu \in (0,1)$ and $\{\varepsilon_n\} \subset (0,1)$ such that $\lim_{n\to\infty} \varepsilon_n = 0$, $\sum_{n=0}^{\infty} \varepsilon_n = \infty$.
Step 1. Let $x^0 \in \mathcal{H}$. Set $n := 0$.
Step 2. Compute

$$y^n = P_C(x^n - \mu_n F(x^n)),$$
$$z^n = y^n - \mu_n(F(y^n) - F(x^n)),$$

where

$$\mu_{n+1} = \begin{cases} \min\left\{ \dfrac{\mu\|x^n - y^n\|}{\|F(x^n) - F(y^n)\|}, \mu_n \right\} & \text{if } F(x^n) \neq F(y^n), \\ \mu_n & \text{if } F(x^n) = F(y^n). \end{cases}$$

Step 3. Compute

$$x^{n+1} = z^n - \varepsilon_n S(z^n).$$

Step 4. Set $n := n + 1$, and go to **Step 2**.

Corollary 3.2. *Under the assumption that conditions (B_1), (B_2) and (B_3) hold, the sequence $\{x^n\}$ generated by Algorithm 3.3 converges strongly to a* *point* $x^* \in \mathrm{Sol}(C, F)$*, which is the unique solution of the variational inequality*

$$\langle S(x^*), x - x^* \rangle \geq 0 \; \forall x \in \mathrm{Sol}(C, F), \tag{3.37}$$

provided that $\mathrm{Sol}(C, F) \neq \emptyset$.

4. Numerical illustrations

In this section, we present numerical experiments to demonstrate the effectiveness of the proposed algorithm. The Python scripts were run on a 2017 MacBook Pro, featuring a 2.3 GHz Intel Core i5 processor, an Intel Iris Plus Graphics 640 with 1536 MB of memory, and 8 GB of 2133 MHz LPDDR3 RAM. The experiments were conducted using Python version 3.11.

Example 4.1. ([14]) Let $\mathbb{R}^K$ be endowed with the standard Euclidean norm $\|x\| = \left(x_1^2 + x_2^2 + \cdots + x_K^2\right)^{\frac{1}{2}}$ for all $x = (x_1, x_2, \ldots, x_K)^T \in \mathbb{R}^K$. We consider the SVIFPP with the mapping $F : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ defined by $F(x) = (\sin \|x\| + 2)a^0$ for all $x \in \mathbb{R}^4$, where $a^0 = (12, -4, 4, -4)^T \in \mathbb{R}^4$. Additionally, let C be the set defined as

$$C = \{(x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4 : 12x_1 - 4x_2 + 4x_3 - 4x_4 \geq 9\}$$

and the bounded linear operator $A : \mathbb{R}^4 \longrightarrow \mathbb{R}^2$ defined by $A(x) = Mx$ for all $x \in \mathbb{R}^4$, where

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$$

Assume that $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is defined by, for all $y = (y_1, y_2)^T \in \mathbb{R}^2$

$$T(y) = \begin{cases} (y_1, y_2)^T & \text{if } y_1 \leq 0, \\ (-2y_1, y_2)^T & \text{if } y_1 > 0. \end{cases}$$

Then T is $\dfrac{1}{3}$-demicontractive and $\mathrm{Fix}(T) = (-\infty, 0] \times \mathbb{R}$.

Consider the mapping $S : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ be defined by $S(x) = x$ for all $x \in \mathbb{R}^4$. This mapping S is strongly monotone with $\eta = 1$ and Lipschitz continuous with $\kappa = 1$ on $\mathbb{R}^4$. In this situation, the problem (1.4) becomes the problem of finding the minimum-norm solution of the SVIFPP.
The solution set Ω_{SVIFPP} of the SVIFPP is given by

$$\begin{aligned} \Omega_{\mathrm{SVIFPP}} &= \{(x_1, x_2, x_3, x_4)^T \in \mathrm{Sol}(C, F) : A(x_1, x_2, x_3, x_4) \in \mathrm{Fix}(T)\} \\ &= \{(x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4 : 12x_1 - 4x_2 + 4x_3 - 4x_4 = 9, x_1 + x_3 + x_4 \leq 0\}. \end{aligned}$$

and the minimum-norm solution x^* of the SVIFPP is $x^* = \left(\frac{1}{2}, -\frac{1}{4}, 0, -\frac{1}{2}\right)^T$.

We now provide a comparison between Algorithm 3.1 and Algorithm 1 in [14]. Given that the exact solution of the problem is $x^* = \left(\frac{1}{2}, -\frac{1}{4}, 0, -\frac{1}{2}\right)^T$, we using $\|x^n - x^*\| \leq \varepsilon$ as the stopping condition. Both algorithms use the same initial point x^0, obtained by randomly generating values within the interval $[-10, 10]$. The parameters for each algorithm are chosen as follows:

- Algorithm 3.1: $\mu_0 = 2$, $\mu = 0.1$, $\rho_n = 1 - 10^{-2}$ and $\varepsilon_n = \frac{1}{n+2}$.
- Algorithm 1 in [14]: $\delta_n = \frac{n+1}{500n + 510}$, $\mu_n = \frac{n+1}{600n + 605}$ and $\varepsilon_n = \frac{1}{n+2}$.

Table 1. A comparison of Algorithm 3.1 and Algorithm 1 in [14] using various tolerances ε and the stopping criterion $\|x^n - x^*\| \leq \varepsilon$

	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$	
	Iter(n)	CPU time(s)	Iter(n)	CPU time(s)
Algorithm 3.1	4274	0.8088	89639	9.6047
Algorithm 1 in [14]	28470	2.1334	295407	18.5596

Table 1 shows that our Algorithm 3.1 outperforms Algorithm 1 in [14] in terms of both the number of iterations and CPU time.

Example 4.2. We consider the mapping $S : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ defined by $S(x) = (4x_1 + 16, 4x_2 - 4, 4x_3 + 3)^T$ for all $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$. It is straightforward to verify that S is both strongly monotone and Lipschitz continuous on $\mathbb{R}^3$. Define the sets $C = \{(x_1, x_2, x_3)^T \in \mathbb{R}^3 : x_1 - x_2 + 2x_3 = 4\}$, $Q = \{(u_1, u_2)^T \in \mathbb{R}^2 : 3u_1 - u_2 = 10\}$ and let the bounded linear operator $A : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be defined by $A(x) = Mx$, where

$$M = \begin{pmatrix} 1 & -4 & 2 \\ 2 & -9 & -4 \end{pmatrix}.$$

The solution set Ω_{SFP} of the SFP is given by

$$\begin{aligned} \Omega_{\text{SFP}} &= \begin{cases} x_1 - x_2 + 2x_3 = 4 \\ 3(x_1 - 4x_2 + 2x_3) - (2x_1 - 9x_2 - 4x_3) = 10 \end{cases} \\ &= \begin{cases} x_1 - x_2 + 2x_3 = 4 \\ x_1 - 3x_2 + 10x_3 = 10, \end{cases} \end{aligned}$$

which can be expressed in parametric form as:

$$\Omega_{\text{SFP}} = \left\{(2t + 1, 4t - 3, t)^T : t \in \mathbb{R}\right\}.$$

Assume that $x^* = (2t^* + 1, 4t^* - 3, t^*)^T \in \Omega_{\rm SFP}$ satisfies the variational inequality

$$\langle S(x^*), x - x^*\rangle \geq 0 \ \ \forall x \in \Omega_{\rm SFP}.$$

Given that $S(x^*) = (8t^* + 20, 16t^* - 16, 4t^* + 3)$, $x - x^* = (2t - 2t^*, 4t - 4t^*, t - t^*)^T$, the inequality becomes

$$(8t^* + 20)(2t - 2t^*) + (16t^* - 16)(4t - 4t^*) + (4t^* + 3)(t - t^*) \geq 0 \ \ \forall t \in \mathbb{R}.$$

This expression simplifies to $21(4t^* - 1)(t - t^*) \geq 0$ for all $t \in \mathbb{R}$. This inequality holds if and only if $t^* = \dfrac{1}{4}$. Therefore, the unique solution to the variational inequality problem (3.36) is $x^* = \left(\dfrac{3}{2}, -2, \dfrac{1}{4}\right)^T$.

We select an initial point $x^0 \in \mathbb{R}^3$, where each component of x^0 is randomly generated within the closed interval $[-10, 10]$. With $\varepsilon_n = \dfrac{1}{n+2}$ and the stopping criterion $\|x^n - x^*\| \leq \varepsilon$, we compute approximate solutions to the exact solution $x^* = \left(\dfrac{3}{2}, -2, \dfrac{1}{4}\right)^T$ for various tolerance levels ε, as presented in Table 2.

Table 2. Approximate solutions corresponding to various tolerance levels ε, obtained using Algorithm 3.2 with the stopping criterion $\|x^n - x^*\| \leq \varepsilon$

ε	Iter(n)	CPU time(s)	x^n
$\varepsilon = 10^{-2}$	19392	1.2879	$(1.491653, -1.996977, 0.254602)^T$
$\varepsilon = 10^{-3}$	194115	10.8199	$(1.499165, -1.999698, 0.250460)^T$
$\varepsilon = 10^{-4}$	1941349	109.1283	$(1.499917, -1.999970, 0.250046)^T$

Example 4.3. We consider the set $C \subset \mathbb{R}^3$ defined by

$$C = \{x = (x_1, x_2, x_3)^T \in \mathbb{R}^3 : 2x_1 - x_2 + 5x_3 \geq 6\}.$$

Next, define the mapping $F : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by $F(x) = (\sin\|x\| + 6)f^0$ for all $x \in \mathbb{R}^3$, where $f^0 = (2, -1, 5)^T \in \mathbb{R}^3$. It is easy to verify that F is pseudomonotone and Lipschitz continuous on $\mathbb{R}^3$. Furthermore, the solution set $\mathrm{Sol}(C, F)$ of the variational inequality problem $VIP(C, F)$ is given by

$$\mathrm{Sol}(C, F) = \{x = (x_1, x_2, x_3)^T \in \mathbb{R}^3 : 2x_1 - x_2 + 5x_3 = 6\}.$$

Now, consider the mapping $S : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ defined by $S(x) = x$ for all $x \in \mathbb{R}^3$. This mapping is strongly monotone with modulus $\eta = 1$ and Lipschitz continuous with constant $L = 1$ on $\mathbb{R}^3$. In this setting, Problem (3.37) reduces to finding the minimum-norm solution of the variational inequality problem $VIP(C,F)$. The resulting minimum-norm solution is given by $x^* = P_{\mathrm{Sol}(C,F)}(0) = (0.4, -0.2, 1)^T$.

We select an initial point $x^0 \in \mathbb{R}^3$, where each component of x^0 is randomly generated within the closed interval $[-10, 10]$. With parameters $\mu_0 = 4$, $\mu = 0.7$, $\varepsilon_n = \dfrac{1}{n+2}$ in Algorithm 3.3 and using the stopping criterion $\|x^{n+1} - x^n\| \leq \varepsilon$. With the tolerance $\varepsilon = 10^{-9}$, an approximate solution is obtained after 84027 iterations (with time 6.109 seconds), given by

$$x^{84027} = (0.400055, -0.199997, 0.999936)^T,$$

which serves as a good approximation to the exact solution $x^* = (0.4, -0.2, 1)^T$.

5. Conclusion

We propose a new algorithm for solving the strongly monotone variational inequality problem over the solution set of split variational inequality and fixed point problem in real Hilbert spaces. By placing suitable conditions on the parameters, we prove a strong convergence theorem for the algorithm, which avoids the need to compute or estimate the norms of the bounded linear operators. Importantly, the algorithm does not require prior knowledge of the Lipschitz or strongly monotone constants of the mappings. Additionally, we derive several corollaries from our main result and demonstrate the algorithm's performance with a basic numerical example.

References

[1] **Anh P.K., Anh T.V. and Muu, L.D.,** *On bilevel split pseudomonotone variational inequality problems with applications.* Acta Math, Vietnam. **42**, 413-429 (2017)

[2] **Anh T.V.,** *An extragradient method for finding minimum-norm solution of the split equilibrium problem.* Acta Math. Vietnam. **42**, 587-604 (2017).

[3] **Anh T.V.,** *A parallel method for variational inequalities with the multiple-sets split feasibility problem constraints.* J. Fixed Point Theory Appl. **19**, 2681-2696 (2017).

[4] **Anh T.V.,** *Linesearch methods for bilevel split pseudomonotone variational inequality problems.* Numer. Algorithms **81**, 1067-1087 (2019).

[5] **Anh T.V. and Muu L.D.,** *A projection-fixed point method for a class of bilevel variational inequalities with split fixed point constraints.* Optimization **65**, 1229-1243 (2016).

[6] **Buong N.,** *Iterative algorithms for the multiple-sets split feasibility problem in Hilbert spaces.* Numer. Algorithms **76**, 783-798 (2017).

[7] **Byrne C.,** *Iterative oblique projection onto convex sets and the split feasibility problem.* Inverse Probl. **18**, 441-453 (2002).

[8] **Byrne C., Censor Y., Gibali A. and Reich S.,** *The split common null point problem.* J. Nonlinear Convex Anal. **13**, 759-775 (2012).

[9] **Ceng L.C., Ansari Q.H. and Yao J.C.,** *Relaxed extragradient methods for finding minimum-norm solutions of the split feasibility problem.* Nonlinear Anal. **75**, 2116-2125 (2012).

[10] **Censor Y., Bortfeld T., Martin B. and Trofimov A.,** *A unified approach for inversion problems in intensity-modulated radiation therapy.* Phys. Med. Biol. **51**, 2353-2365 (2006).

[11] **Censor Y. and Elfving T.,** *A multiprojection algorithm using Bregman projections in a product space.* Numer. Algorithms **8**, 221-239 (1994).

[12] **Censor Y., Elfving T., Kopf N. and Bortfeld T.,** *The multiple-sets split feasibility problem and its applications for inverse problems.* Inverse Prob. **21**, 2071-2084 (2005).

[13] **Censor Y. and Segal A.,** *Iterative projection methods in biomedical inverse problems,* in: Y. Censor, M. Jiang, A.K. Louis (Eds.), Mathematical Methods in Biomedical Imaging and Intensity-Modulated Therapy, IMRT, Edizioni della Norale, Pisa, Italy, 2008, pp. 65-96.

[14] **Hai N.M., Van L.H.M. and Anh T.V.,** *An Algorithm for a Class of Bilevel Variational Inequalities with Split Variational Inequality and Fixed Point Problem Constraints.* Acta Math. Vietnam. **46**, 515–530 (2021).

[15] **Huy P.V., Hien N.D. and Anh T.V.,** *A strongly convergent modified Halpern subgradient extragradient method for solving the split variational inequality problem.* Vietnam J. Math. **48**, 187-204 (2020).

[16] **Huy P.V., Van L.H.M., Hien N.D. and Anh T.V.,** *Modified Tseng's extragradient methods with self-adaptive step size for solving bilevel split variational inequality problems.* Optimization **71**, 1721-1748 (2022).

[17] **Konnov I.V.,** *Combined Relaxation Methods for Variational Inequalities.* Springer, Berlin (2000).

[18] **Liu B., Qu B. and Zheng N.,** *A Successive Projection Algorithm for Solving the Multiple-Sets Split Feasibility Problem.* Numer. Funct. Anal. Optim. **35**, 1459-1466 (2014).

[19] **Maingé P.E.,** *A hybrid extragradient-viscosity method for monotone operators and fixed point problems.* SIAM J. Control Optim. **47**, 1499-1515 (2008).

[20] **Tseng P.,** *A modified forward–backward splitting method for maximal monotone mappings.* SIAM J. Control Optim. **38**, 431–446 (2000).

[21] **Wen M., Peng J.G. and Tang Y.C.,** *A cyclic and simultaneous iterative method for solving the multiple-sets split feasibility problem.* J. Optim. Theory Appl. **166**(3), 844-860 (2015).

[22] **Xu H.K.,** *Iterative algorithms for nonlinear operators.* J. London Math. Soc. **66**, 240–256 (2002).

[23] **Zhao J.L. and Yang Q.Z.,** *A simple projection method for solving the multiple-sets split feasibility problem.* Inverse Probl. Sci. Eng. **21**, 537-546 (2013).

[24] **Zhao J.L., Zhang Y.J. and Yang Q.Z.,** *Modified projection methods for the split feasibility problem and the multiple-sets split feasibility problem.* Appl. Math. Comput. **219**, 1644-1653 (2012).

Tran Viet Anh
Department of Scientific Fundamentals, Posts and Telecommunications
Institute of Technology
Hanoi
Vietnam
tranvietanh@outlook.com; tvanh@ptit.edu.vn

J. Math. Math. Sci. (2025) 77–94

Random dynamical systems generated by nonautonomous stochastic differential equations driven by fractional Brownian motions

Phan Thanh Hong (Hanoi, Vietnam)

(Received Mar. 25, 2025; accepted May 23, 2025)

Abstract. In this paper, we prove that a non-autonomous stochastic differential equation generates a continuous random dynamical system. The flow then possesses a random pullback attractor under the dissipativity condition(s) of the drift and smallness of diffusion part.

1. Introduction

This work is a follow up part of [7], [14] to study the the asymptotic qualitative behavior of the differential equation

$$dy_t = f(t, y_t)dt + g(t, y_t)dB^H_t, t \in \mathbb{R},\ y_0 \in \mathbb{R}^d. \tag{1.1}$$

in which B^H is a fractional Brownian motion with Hurst parameter H bigger than $\frac{1}{2}$; f and g are some continuous functions on $\mathbb{R} \times \mathbb{R}^d$.

When dealing with qualitative properties of (1.1), one important problem is the generation of *random dynamical system*, RDS in short ([1]). The concept of RDS is a combining idea of randomness and dynamical system. Theory

Key words and phrases: stochastic differential equations (SDE), Young integrals, Bebutov flow, random dynamical systems, random attractors.
2020 Mathematics Subject Classification: 37H10, 60H10

of RDS is the frame work to study the system's asymptotic for instance the random attractors, random manifolds, Lyapunov spectrum,...In general cases, when f, g are functions of $(t, y) \in \mathbb{R} \times \mathbb{R}^d$, the system generates a stochastic two-parameter flow $X(t_0, t, y_0, \omega)$ by mean of its Cauchy operators [4], the flow induces a random dynamical system (RDS) in case $f,\ g$ are time independent.

In [22, 23], a nonautonomous ordinary differential equations $dy(t) = f(t, y_t)dt$ is considered. By introducing the space "hull" of f, the solution can be viewed as a dynamical system. Motivated by these results, we establish conditions on $f,\ g$ to construct appropriate spaces for f, g which admit needed probability structures. The flow is then defined on the product spaces and possesses group property. Equation (1.1) then generates a RDS in the sense of Bebutov flow [22].

One another topic in this paper is study the existence of random pullback attractor of the system, see for instant [5] or [8], [10] for recent results established for stochastic differential equations driven by Hölder noises. We show in Section 3 that the generated RDS possesses a random pullback random attractor under dissipative assumption of f and point out that the attractor is singleton if dissipativity is strict and g is small in some sense.

2. Preliminaries

We briefly recall some notions used in the sequence.

- Let $\mathcal{C}([a,b], \mathbb{R}^r),\ r \geq 1$, denote the space of all continuous paths $x : [a,b] \to \mathbb{R}^r$ equipped with supremum norm $\|\cdot\|_{\infty,[a,b]}$ given by $\|x\|_{\infty,[a,b]} = \sup_{t\in[a,b]} |x_t|$.

- For $0 < \alpha < 1$, let x is a Hölder continuous function with exponent α on $[a,b]$. The semi norm $\alpha-$ Hölder of x is defined as

$$\|x\|_{\alpha-\mathrm{Hol},[a,b]} = \sup_{a\leq s<t\leq b} \frac{|x_t - x_s|}{(t-s)^\alpha}.$$

- For given $p \geq 1$, denote by $\mathcal{C}^{p-\mathrm{var}}([a,b], \mathbb{R}^r) \subset \mathcal{C}([a,b], \mathbb{R}^r)$ the space consists of all continuous paths x of finite $p-$variation, i.e.

$$\|x\|_{p-\mathrm{var},[a,b]} := \left(\sup_{a=t_0<t_1<\cdots<t_n=b} \sum_{i=1}^{n} |x_{t_{i+1}} - x_{t_i}|^p \right)^{1/p} < \infty.$$

The $p-$variation norm of x is defined by

$$\|x\|_{p-\mathrm{var},[a,b]} \quad := \quad |x_a| + \left|\!\left|\!\left| x \right|\!\right|\!\right|_{p-\mathrm{var},[a,b]}.$$

Then $(\mathcal{C}^{p-\mathrm{var}}([a,b],\mathbb{R}^r), \|\cdot\|_{p-\mathrm{var},[a,b]})$ is a (nonseparable) Banach space [11, Theorem 5.25, p. 92].

Young integral

Assume $y \in \mathcal{C}^{q-\mathrm{var}}([a,b],\mathbb{R}^{d\times m})$ and $x \in \mathcal{C}^{p-\mathrm{var}}([a,b],\mathbb{R}^m)$ with $\frac{1}{p}+\frac{1}{q}>1$, the Young integral $\int_a^b y_t dx_t$ is defined as the limitation of the Darboux sum

$$\int_a^b y_t dx_t := \lim_{|\Pi|\to 0} \sum_{t_i\in\Pi} y_{t_i}(x_{t_{i+1}} - x_{t_i}),$$

where the limit is taken over all the finite partitions $\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$ of $[a,b]$ with $|\Pi| := \max_i |t_{i+1} - t_i|$ (see [24]). The integral satisfies ([11, Theorem 6.8, p. 116])

$$\Big|\int_a^b y_u dx_u - y_a(x_b - x_a)\Big| \le (1 - 2^{1-\frac{1}{p}-\frac{1}{q}})^{-1} \left|\!\left|\!\left| y \right|\!\right|\!\right|_{q-\mathrm{var},[a,b]} \left|\!\left|\!\left| x \right|\!\right|\!\right|_{p-\mathrm{var},[a,b]}.$$

Fractional Brownian motions

A m-dimensional fractional Brownian motion index H, $B^H = (B^H_t),\ t\in\mathbb{R}$, is a vector consists of m independent one dimensional fractional Brownian motions index H which are centered continuous Gaussian processes with covariance function

$$R_H(s,t) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}),\ s,t\in\mathbb{R}.$$

For each $p\ge 1$ denote by $\mathcal{C}^{0,p-\mathrm{var}}([a,b],\mathbb{R}^m)$ the closure of set of smooth paths in $\mathcal{C}^{p-\mathrm{var}}([a,b],\mathbb{R}^m)$ and Ω the spaces of all continuous functions $\omega:\mathbb{R}\to\mathbb{R}^m$ vanish at 0 such that the restriction of ω on $[a,b]$ is in $\mathcal{C}^{0,p-\mathrm{var}}([a,b],\mathbb{R}^m)$ for all $[a,b]$. Then Ω is a separable metric space with the metric (see [2])

$$d(\omega^1,\omega^2) := \sum_{n=1}^{\infty} 2^{-n} \frac{\|\omega^1-\omega^2\|_{p\text{-var},[-n,n]}}{1+\|\omega^1-\omega^2\|_{p\text{-var},[-n,n]}}. \tag{2.1}$$

Follow [13], one can construct a canonical space for B^H on Ω for some $p > 1/H$ with Borel $\sigma-$algebra $\mathcal{F}$ and the law $\mathbb{P}$ of B^H. It is proved in [13] that together with Wiener shift (θ_t) defined as

$$\theta_t(\omega)(\cdot) := \omega(t+\cdot) - \omega(t), \quad \omega\in\Omega,$$

the space $(\Omega, \mathcal{F}, \mathbb{P}, (\theta_t))$ forms an ergodic dynamical system. From now on, we always work on the canonical space of B^H. We keep the old notation B^H and identify $B^H_\cdot(\omega) = \omega(\cdot), \omega \in \Omega$. Moreover, since we consider the case $H > 1/2$, p can be choosen in $(1/H, 2)$, the integral w.r.t. B^H can be defined by Young sense [24].

Finally, recall from [15, Proposition 2.1] that there exists random variable $\xi(\omega)$ and $\kappa > 0$ satisfying $\mathbb{E}e^{\kappa\xi^2} < \infty$ such that for some constant D, for almost all ω

$$\left\|\!\left| B^H_\cdot(\omega) \right|\!\right\|_{p-\mathrm{var},[0,1]} \leq D\xi(\omega).$$

It follows that for all $k > 0$, $\mathbb{E}\left\|\!\left| B^H_\cdot(\omega) \right|\!\right\|^k_{p-\mathrm{var},[0,1]} < \infty$.

3. Generation of random dynamical system

3.1. Bebutov flow

In this section we show that (1.1) generates a random dynamical system (RDS) in an extended space. A RDS on $\mathbb{R}^d$ over a metric dynamical system (see for instant [1]) $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*, (\theta^*_t))$ is a measurable mapping

$$\varphi : \mathbb{R}^+ \times \mathbb{R}^d \times \Omega^* \to \mathbb{R}^d, (t, x, \omega) \mapsto \varphi(t, \omega)x$$

satisfying
(i) $\varphi(0, \omega) = Id$ for all $\omega \in \Omega^*$,
(ii) $\varphi(t + s, \omega) = \varphi(t, \theta^*_s\omega) \circ \varphi(s, \omega)$ for all $s, t \in \mathbb{R}^+$, $\omega \in \Omega^*$.

If, in addition, $x \mapsto \varphi(t, \omega)x$ is continuous for all t, ω then φ is called continuous.

Recall from [22] that on $\mathcal{C} := \mathcal{C}(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}^d)$ the shift mapping $S = (S_t)_{t\in\mathbb{R}}$ is defined as

$$S_t h = S(t, h) =: h_t, \quad \forall h \in \mathcal{C},$$

h_t is called a translate of h given by $h_t(s, x) = h(t + s, x),\ (s, x) \in \mathbb{R} \times \mathbb{R}^d$.

Observe that if y is a solution to

$$dy_t = f(t, y_t)dt + g(t, y_t)d\omega_t, t \in \mathbb{R},\ y_0 \in \mathbb{R}^d, \tag{3.1}$$

where ω is a realization of B^H, then

$$\begin{aligned}(3.2)\quad & y_{s+t} \\ = & \int_0^{s+t} f(u,y_u)du + \int_0^{s+t} g(u,y_u)d\omega_u \\ = & \int_0^s f(u,y_u)du + \int_0^s g(u,y_u)d\omega_u + \int_s^{s+t} f(u,y_u)du + \int_s^{s+t} g(u,y_u)d\omega_u \\ = & y_s + \int_0^t S_s f(u,y_{s+u})du + \int_0^t S_s g(u,y_{s+u})d\theta_s\omega_u.\end{aligned}$$

Then $y_{s+\cdot}$ is the solution of (3.1) with coefficients $S_s f$, $S_s g$. This suggested using Krylov-Bogoliubov theorem [18, Chapter VI, §9] to construct probability structures on hull of f and g in appropriate metric spaces. To do this we consider (1.1) under the conditions as follows.

Assumptions

($\mathbf{H}_1$) $f(t,x)$ is uniformly continuous on $\mathbb{R}\times K$ for each K compact in $\mathbb{R}^d$, and there exists $C_f,\ f_0>0$ such that for all $x,y\in\mathbb{R}^d,\ s,t\in\mathbb{R}$

$$\begin{cases}(i)\ |f(t,x)-f(t,y)|\leq C_f|x-y|,\\ (ii)\ |f(t,0)|\leq f_0.\end{cases}$$

($\mathbf{H}_2$) $g(t,x)$ is bounded by $\|g\|_\infty$ and differentiable in x with $\partial_x g$ being locally Lipschitz in x uniformly in t. Moreover, there exists $C_g>0$ and $\beta\in(1-1/p,1)$ such that the following properties hold for all $x,y\in\mathbb{R}^d,\ s,t\in\mathbb{R}$

$$\begin{cases}(i)\ |g(t,x)-g(t,y)|\leq C_g|x-y|,\\ (ii)\ |g(t,x)-g(s,x)|+\|\partial_x g(t,x)-\partial_x g(s,x)\|\leq C_g|t-s|^\beta.\end{cases}$$

Under these conditions, system (1.1) possesses a unique solution $y_t = y(t,x_0,\omega)$, $t\in\mathbb{R}$ for each realization ω of B^H. Moreover, for all $[a,b]\subset\mathbb{R}$,

$$\|y\|_{p-\mathrm{var},[a,b]}\leq M(b-a)\,[|y_a|+1]\,\Lambda(\omega,[a,b]) \tag{3.3}$$

where M is a constant depend on $b-a$ and $\Lambda(\omega,[a,b])$ is a polynomial of $|\!|\!|\omega|\!|\!|_{p-\mathrm{var},[a,b]}$ (see [4],[8]).

3.1.1. Hull of f

In a similar manner of (2.1), define the metric d_0 in $\mathcal{C}$ - space of all continuous functions on $\mathbb{R}$ by replacing the $p-$variation norm $\|\cdot\|_{p-\mathrm{var},[a,b]}$ by supreme norm $\|\cdot\|_{\infty,[a,b]}$. For given f, the hull of f, denoted by $\mathcal{H}^f_{d_0}$ the closure of the sets $\{S_\tau f|\tau\in\mathbb{R}\}$ in $(\mathcal{C},d_0)$,

$$\mathcal{H}^f_{d_0} := \overline{\{S_\tau f|\tau\in\mathbb{R}\}}^{(\mathcal{C},d_0)}.$$

According to [22, Theorem 1, 14] S defines a dynamical system on $\mathcal{C}$. Moreover, by the assumptions, f is bounded and uniformly continuous on $\mathbb{R}\times K$ for each K compact in $\mathbb{R}^d$, $\mathcal{H}^f_{d_0}$ is compact in $\mathcal{C}$. We derive required properties for $\mathcal{H}^f_{d_0}$.

Note that, similar results apply for $\mathcal{C}^{1,0} = (\mathcal{C}^{1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^d),\rho)$-the space of continuous functions h with $\partial_x h\in\mathcal{C}$ with metric

$$\rho(h,k) = d_0(h,k) + d_0(\partial_x h,\partial_x k). \tag{3.4}$$

3.1.2. Hull of g

Next, we construct similar space for g. Firstly, consider the subspace $\mathcal{C}^{\alpha;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})\subset\mathcal{C}^{1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})$ containing functions h which is of local $\alpha-$Hölder w.r.t. t for each $x\in\mathbb{R}^d$ and moreover for each compact set K in $\mathbb{R}^d$

$$\sup_{x\in K}|\!|\!| h(\cdot,x)|\!|\!|_{\alpha-\mathrm{Hol},[a,b]} < \infty,\ \forall[a,b]\subset\mathbb{R}^d.$$

We consider the following metric on $\mathcal{C}^{\alpha;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})$ which is denoted by d_1

$$d_1(h^1,h^2) := \sum_{n=1}^{\infty}\frac{1}{2^n}\frac{\|h^1-h^2\|_{\alpha,1,0;K_n}}{1+\|h^1-h^2\|_{\alpha,1,0;K_n}}, \tag{3.5}$$

where

$$\begin{aligned}
\|h^1-h^2\|_{\alpha,1,0;K^1\times K^2} &:= \|h^1-h^2\|_{1,0;K^1\times K^2} + |\!|\!| h^1-h^2|\!|\!|_{\alpha,K^1\times K^2}\\
\|h^1-h^2\|_{1,0;K^1\times K^2} &:= \sup_{K^1\times K^2}|h^1-h^2| + \sup_{K^1\times K^2}\|\partial_x h^1-\partial_x h^2\|\\
|\!|\!| h^1-h^2|\!|\!|_{\alpha,K^1\times K^2} &:= \sup_{x\in K^2}|\!|\!| h^1(\cdot,x)-h^2(\cdot,x)|\!|\!|_{\alpha-\mathrm{Hol},K^1}
\end{aligned}$$

with K^1,K^2 are compact sets in $\mathbb{R}$, $\mathbb{R}^d$ respectively.

Proposition 3.1. *$(\mathcal{C}^{\alpha;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m}),d_1)$ is a complete metric space.*

Proof. See in the Appendix.

Next, we fix $1-\frac{1}{p}<\beta_0<\beta$, denoted by $(\mathcal{C}^{\beta_0;1,0},d_1)$ the space $(\mathcal{C}^{\beta_0;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m}),d_1)$. Put $\mathcal{H}^g_{d_1}$ the closure of $\{S_\tau g|\tau\in\mathbb{R}\}$ in $\mathcal{C}^{\beta_0;1,0}$, i.e.

$$\mathcal{H}^g_{d_1} := \overline{\{S_\tau g|\tau\in\mathbb{R}\}}^{(\mathcal{C}^{\beta_0;1,0},d_1)}.$$

The similar results hold for hull of g as stated below.

Lemma 3.1. *All $g^*\in\mathcal{H}^g_{d_1}$ satisfies $(\mathbf{H_2})$ and moreover, $\mathcal{H}^g_{d_1}$ is a compact set in $(\mathcal{C}^{\beta_0;1,0},d_1)$.*

Proof. See in the Appendix.

Since $\mathcal{C}^{\beta_0;1,0}$ is not separable, in the following we directly prove that S defines a dynamical system on $\mathcal{H}^g_{d_1}$.

Lemma 3.2. *S defines a dynamical system on $\mathcal{H}^g_{d_1}$.*

Proof. Due to [22, Theorem 12], S defined a dynamical system on $\mathcal{C}^{1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})$. We just need to check that for fixed $(t_0,h^0)\in\mathbb{R}\times\mathcal{H}^g_{d_1}$, if $t\in\mathbb{R}, h\in\mathcal{H}^g_{d_1}$ such that $|t-t_0|, d_1(h,h^0)\to 0$ then $\left|\!\left|\!\left| h_t(\cdot,x)-h^0_{t_0}(\cdot,x)\right|\!\right|\!\right|_{\beta^0-\mathrm{Hol},[a,b]\times K}\to 0$ for each a,b, each K compact in $\mathbb{R}^d$. Namely, by choosing appropriate $[a',b']$ we have

$$\begin{aligned}
& \left|\!\left|\!\left| h_t-h^0_{t_0}\right|\!\right|\!\right|_{\beta^0-\mathrm{Hol},[a,b]\times K}\\
\leq\ & \left|\!\left|\!\left| h_t-h^0_{t}\right|\!\right|\!\right|_{\beta^0-\mathrm{Hol},[a,b]\times K}+\left|\!\left|\!\left| h^0_t-h^0_{t_0}\right|\!\right|\!\right|_{\beta^0-\mathrm{Hol},[a,b]\times K}\\
\leq\ & \left|\!\left|\!\left| h-h^0\right|\!\right|\!\right|_{\beta^0-\mathrm{Hol},[a',b']\times K}+2\left|\!\left|\!\left| h^0\right|\!\right|\!\right|^{\beta_0/\beta}_{\beta-\mathrm{Hol},[a',b']\times K}\cdot\|h^0_t-h^0_{t_0}\|^{1-\beta_0/\beta}_{\infty,[a,b]\times K}\\
& \to 0,\ \text{ as } |t-t_0|\to 0,\ d_1(h,h^0)\to 0.
\end{aligned}$$

This shows the continuity of S on $\mathcal{H}^g_{d_1}$. Since $\mathcal{H}^g_{d_1}$ is compact, S is measurable w.r.t. the $\sigma-$algebra generated by d_1. The proof is completed.

■

3.2. Generation of RDS

Since $\mathcal{H}^f_{d_0},\mathcal{H}^g_{d_1}$ are compact sets with appropriate metrics constructed above, we deduce from Krylov-Bogoliubov theorem [18, Chapter VI, §9] that there are probability measures $\mathbb{P}^f,\mathbb{P}^g$ on measurable space $(\mathcal{H}^f_{d_0},\mathcal{B}^f),(\mathcal{H}^g_{d_1},\mathcal{B}^g)$ with Borel $\sigma-$algebras $\mathcal{B}^f,\mathcal{B}^g$, that are invariant under the shifts mapping S. Denote by $\bar\Omega$ the Catersian product $\mathcal{H}^f_{d_0}\times\mathcal{H}^g_{d_1}\times\Omega$ with the product Borel $\sigma-$field denoted by $\bar{\mathcal{B}}$ and the product measure $\bar{\mathbb{P}}=\mathbb{P}^f\times\mathbb{P}^g\times\mathbb{P}$ and consider the product dynamical system $\bar\theta:\mathbb{R}\times\bar\Omega\to\bar\Omega$ given by

$$\bar\theta(t,\tilde f,\tilde g,\omega)=(S_t\tilde f,S_t\tilde g,\theta_t\omega),\ (\tilde f,\tilde g,\omega)\in\bar\Omega.$$

It is evident that $(\bar\Omega,\bar{\mathcal{B}},\bar{\mathbb{P}},\bar\theta)$ forms a metric dynamical system.

Proposition 3.2. *For each $\bar\omega=(\bar f,\bar g,\omega)\in\bar\Omega$, equation*

$$dy_t=\bar f(t,y_t)dt+\bar g(t,y_t)d\omega_t,\ y_0\in\mathbb{R}^d,\ t\in\mathbb{R}^+ \tag{3.6}$$

possesses a unique solution $y(t,y_0,\bar\omega)$. The solution is continuous w.r.t. the initial condition y_0 and satisfies (3.3).

Proof. It is easy to check that all elements in $\mathcal{H}^f_{d_0}$ satisfies ($\mathbf{H_1}$). As stated in Lemma 3.1, $\bar{g}$ satisfies ($\mathbf{H_2}$). The statement is evident due to [4].

■

Theorem 3.3. *System*

$$dy_t = f(t, y_t)dt + g(t, y_t)dB^H_t \tag{3.7}$$

generates a continuous random dynamical system over $(\bar{\Omega}, \bar{\mathcal{B}}, \bar{P}, \bar{\theta})$.

Proof. For each $\bar{\omega} = (\bar{f}, \bar{g}, \omega) \in \bar{\Omega}$, consider (3.6). Define

$$\Phi^* : \mathbb{R}^+ \times \mathbb{R}^d \times \bar{\Omega} \to \mathbb{R}^d$$

where $\Phi^*(t, \bar{\omega})y_0$ is the value of the of the solution of (3.6) at the time $t \in \mathbb{R}^+$ with the initial time $s = 0$ and initial value y_0, i.e. $y(t, y_0, \bar{\omega})$. From (3.2), Φ^* satisfies cocycle property

$$\Phi^*(t + s, \bar{\omega})y_0 = \Phi^*(t, \bar{\theta}_s\bar{\omega}) \circ \Phi^*(s, \bar{\omega})y_0.$$

Next, to complete the proof we prove that the solution is continuous w.r.t. $\bar{\omega}$ as an element in the product of separable metric spaces $\mathcal{H}^f_{d_0}, \mathcal{H}^g_{d_1}, \Omega$. The measurability of the solution is obtained thank to [3, Lemma III. 14]. Namely, we fix t, x_0 and $[0, T]$ contains t and consider $\bar{\omega}^1 = (f^1, g^1, \omega^1)$, $\bar{\omega}^2 = (f^2, g^2, \omega^2)$ in $\bar{\Omega}$. Put $y^1_t := y(t, y_0, \bar{\omega}^1), y^2_t := y(t, y_0, \bar{\omega}^2)$ then we have

$$\begin{aligned} y^1_t &= x_0 + \int_0^t f^1(s, y^1_s)ds + \int_0^t g^1(s, y^1_s)d\omega^1_s, \\ y^2_t &= x_0 + \int_0^t f^2(s, y^2_s)ds + \int_0^t g^2(s, y^2_s)d\omega^2_s. \end{aligned}$$

Therefore, $z_t := y^1_t - y^2_t$ satisfies the equation

$$\begin{aligned} z_t &= y^1_t - y^2_t \\ &= \int_0^t [f^1(s, y^1_s) - f^2(s, y^2_s)]ds + \int_0^t [g^1(s, y^1_s)d\omega^1_s - \int_0^t g^2(s, y^2_s)]d\omega^2_s \\ &\quad + \int_0^t [f^2(s, y^1_s) - f^2(s, y^2_s)]ds + \int_0^t [f^1(s, y^1_s) - f^2(s, y^1_s)]ds \\ &\quad + \int_0^t g^1(s, y^1_s)d(\omega^1_s - \omega^2_s) + \int_0^t [g^1(s, y^1_s) - g^2(s, y^1_s)]d\omega^2_s \\ &\quad + \int_0^t [g^2(s, y^1_s) - g^2(s, y^2_s)]d\omega^2_s. \end{aligned}$$

Fixing $\bar{\omega}^1$, due to (3.3) one can find R depends on $\bar{\omega}^1$ such that $\|y(\cdot,y_0,\bar{\omega})\|_{p-\text{var},[0,T]} \leq R$ for all $\bar{\omega}$ lies in the neighbor of $\bar{\omega}^1$ of radius 1. We choose a upper bound for the norms of f^i, g^i, ω^i on $\bar{K} := [0,T]\times \bar{B}(0,R)$ and reuse the notation R for convenient. We will show that z is near 0 when $\|f^1-f^2\|_{\infty,\bar{K}}$, $\|g^1-g^2\|_{\infty,\bar{K}}$, $\|\partial_x g^1 - \partial_x g^2\|_{\infty,\bar{K}}$, and $\left\|\!\left|g^1-g^2\right|\!\right\|_{\beta_0,\bar{K}}$ less than ε small enough.

For $0\leq u<v\in[0,T]$ and $q:=1/\beta$

$$\begin{aligned}|z_u - z_v| &= \left|\int_u^v [f^2(s,y_s^1)-f^2(s,y_s^2)]ds\right| + \left|\int_u^v [f^1(s,y_s^1)-f^2(s,y_s^1)]ds\right| \\ &\quad + \left|\int_u^v [g^2(s,y_s^1)-g^2(s,y_s^2)]d\omega_s^2\right| + \left|\int_u^v g^1(s,y_s^1)d(\omega_s^1-\omega_s^2)\right| \\ &\quad + \left|\int_u^v [g^1(s,y_s^1)-g^2(s,y_s^1)]d\omega_s^2\right|\end{aligned}$$

in which

$$\begin{aligned}\left|\int_u^v [f^2(s,y_s^1)-f^2(s,y_s^2)]ds\right| &\leq C_f\int_u^v |z_s|ds, \\ \left|\int_u^v [g^2(s,y_s^1)-g^2(s,y_s^2)]d\omega_s^2\right| &\leq DC_g(1+\left\|\!\left|y^1\right|\!\right\|_{p-\text{var},[u,v]} + \left\|\!\left|y^2\right|\!\right\|_{p-\text{var},[u,v]})\times \\ &\qquad \times \left\|\!\left|\omega^2\right|\!\right\|_{p-\text{var},[u,v]} \|z\|_{q-\text{var},[u,v]}\end{aligned}$$

where the final estimate due to [4]. And

$$\begin{aligned}\left|\int_u^v [f^1(s,y_s^1)-f^2(s,y_s^1)]ds\right| &\leq \|f^1-f^2\|_{\infty,\bar{K}}(v-u), \\ \left|\int_u^v g^1(s,y_s^1)d(\omega_s^1-\omega_s^2)\right| &\leq D\left\|\!\left|\omega^1-\omega^2\right|\!\right\|_{p-\text{var},[u,v]}\left[\|y^1\|_{q-\text{var},[u,v]} + (v-u)^\beta + 1\right], \\ \left|\int_u^v [g^1(s,y_s^1)-g^2(s,y_s^1)]d\omega_s^2\right| &\leq D\left\|\!\left|\omega^2\right|\!\right\|_{p-\text{var},[u,v]}\left[\|g^1-g^2\|_{\infty,\bar{K}} + \left\|\!\left|g^1-g^2\right|\!\right\|_{q-\text{var},[u,v]}\right], \\ &\leq D\left\|\!\left|\omega^2\right|\!\right\|_{p-\text{var},[u,v]}\Big[\|g^1-g^2\|_{\infty,\bar{K}} \\ &\quad + \left\|\!\left|g^1-g^2\right|\!\right\|_{\beta_0,\bar{K}}(v-u)^\beta + \|\partial_x g^1-\partial_x g^2\|_{\infty,\bar{K}}\left\|\!\left|y^1\right|\!\right\|_{p-\text{var},[u,v]}\Big].\end{aligned}$$

In the final estimate we use the mean value theorem namely for $s,t\in[u,v]$

$$\begin{aligned}&|g^1(t,y_t^1)-g^2(t,y_t^1)-g^1(s,y_s^1)+g^2(s,y_s^1)| \\ &\leq |(g^1-g^2)(t,y_t^1)-|(g^1-g^2)(s,y_t^1)| + |(g^1-g^2)(s,y_t^1)-|(g^1-g^2)(s,y_s^1)| \\ &\leq \left\|\!\left|g^1-g^2\right|\!\right\|_{\beta_0,\bar{K}}(t-s)^{\beta_0} + \|\partial_x g^1-\partial_x g^2\|_{\infty,\bar{K}}|y_t^1-y_s^1|.\end{aligned}$$

Therefore

$$\|z\|_{q-\text{var},[u,v]} \leq D\left(\int_u^v |z_s|ds + \|z\|_{q,[u,v]} + A_{u,v}^{1/q}\right)$$

where D is a constant depending on R and A is a control function defined by

$$A^{1/q}_{u,v} := \varepsilon(v-u) + \left\|\omega^1-\omega^2\right\|_{p-\mathrm{var},[u,v]} + \varepsilon\left\|\omega^2\right\|_{p-\mathrm{var},[u,v]}.$$

Apply Lemma 4.1, since $z_0 = 0$ we obtain

$$\|z\|_{q-\mathrm{var},[0,T]} \le D(\|z_0\| + \varepsilon) = D\varepsilon \to 0 \;\; as \;\; \varepsilon \to 0.$$

This completes the proof. ∎

4. Random attractors

In what follows we recall the notion of the (global) random attractor. For a probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$, a set $\mathcal{M} \subset \mathbb{R}^d \times \Omega^*$ with closed $\omega-$section $\mathcal{M}(\omega) = \{x \in \mathbb{R}^d | (\omega, x) \in \mathcal{M}\}$ is called random set if the map $\omega \mapsto d(x, \mathcal{M}(\omega))$ is measurable for every $x \in \mathbb{R}^d$, where d is the Hausdorff semi-distance.

We work with the universe $\hat{\mathcal{D}}$- the family of *tempered* random sets $\hat{D}(\omega)$, i.e $\hat{D}(\omega)$ is contained in a ball $B(0, r(\omega))$ a.s., where the radius $r(\omega)$ is a tempered random variable, namely satisfies

$$\lim_{t\to\pm\infty} \frac{1}{t} \log^+ r(\theta^*_t\omega) = 0, \quad \text{a.s.} \tag{4.1}$$

Let φ be a continuous random dynamical system on $\mathbb{R}^d$ over a metric dynamical system $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*, (\theta^*_t))$. A random subset $\mathcal{A}$ is called invariant, if

$$\varphi(t,\omega)\mathcal{A}(\omega) = \mathcal{A}(\theta^*_t\omega) \;\; \forall t \in \mathbb{R}^+, \; a.s \;\; \omega \in \Omega^*.$$

It is called a *pullback random attractor* in $\hat{\mathcal{D}}$ if it is compact, invariant and attracts any $\hat{D} \in \hat{\mathcal{D}}$ in the pullback sense, i.e.

$$\lim_{t\to\infty} d(\varphi(t, \theta^*_{-t}\omega)\hat{D}(\theta^*_{-t}\omega) | \mathcal{A}(\omega)) = 0, \;\; \forall \hat{D} \in \hat{\mathcal{D}}, \; a.s. \;\; \omega \in \Omega^*. \tag{4.2}$$

A random set $\mathcal{B} \in \hat{\mathcal{D}}$ is called *pullback absorbing* in the universe $\hat{\mathcal{D}}$ if $\mathcal{B}$ absorbs all sets in $\hat{\mathcal{D}}$, i.e. for any $\hat{D} \in \hat{\mathcal{D}}$, there exists a time $t_0 = t_0(\omega, \hat{\mathcal{D}})$ such that

$$\varphi(t, \theta^*_{-t}\omega)\hat{D}(\theta^*_{-t}\omega) \subset \mathcal{B}(\omega), \text{ for all } t \ge t_0. \tag{4.3}$$

If there exists pullback absorbing set for φ, then it is proved that

$$\mathcal{A}(\omega) = \bigcap_{s\ge 0} \overline{\bigcup_{t\ge s} \varphi(t, \theta^*_{-t}\omega)\mathcal{B}(\theta^*_{-t}\omega)}. \tag{4.4}$$

is the random pullback attractor of φ. Moreover, it is unique in $\hat{\mathcal{D}}$ ([21]).

In the following, we assume that f is uniform dissipative ([6]), i.e. there exist $c, d > 0$ such that for all $t \in \mathbb{R},\ x \in \mathbb{R}^d$

$$\langle x, f(t,x)\rangle \leq c - d\|x\|^2. \tag{4.5}$$

We will prove that the RDS generated by (3.7) possesses a random attractor. The technique is followed from [8]. Here we sketch some main details.

Theorem 4.1. *In addition to* $(\mathbf{H_1}), (\mathbf{H_2})$ *if* f *satisfies* (4.5)*, then RDS generated by system* (3.7) *possesses a random pullback attractor almost sure.*

Proof. **Step 1**: First, fix $\bar{\omega} = (\bar{f}, \bar{g}, \omega) \in \bar{\Omega},\ [a,b] \subset \mathbb{R}^+$. We consider the corresponding ordinary differential equation

$$\dot{\mu_t} = \bar{f}(t, \mu_t),\ t \in [a,b],\ \mu_a = y_a. \tag{4.6}$$

where y is a solution of (3.6) on $[a,b]$.

Since f is dissipative,

$$\begin{aligned} \|\mu\|_{\infty,[a,b]} &\leq |\mu_a| + L, \\ \left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \mu \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{p-\mathrm{var},[a,b]} &\leq L\left(|\mu_a| + 1\right)(b-a) \end{aligned}$$

where L is a constant.

Define $k_t = y_t - \mu_t, t \in [a,b]$. Since k satisfies the equation

$$dk_t = d(y_t - \mu_t) = [\bar{f}(t, \mu_t + k_t) - \bar{f}(t, \mu_t)]dt + \bar{g}(t, \mu_t + k_t)d\omega_t$$

we have

$$k_t - k_s = \int_s^t \Big[\bar{f}(u, k_u + \mu_u) - \bar{f}(u, \mu_u)\Big]du + \int_s^t \bar{g}(u, k_u + \mu_u)d\omega_u.$$

It follows from $(\mathbf{H}_2)$ and the boundedness of g that

$$\begin{aligned} |k_t - k_s| \quad \leq \quad & \int_s^t C_f |k_u| du + \|\bar{g}\|_\infty \left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \omega \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{p-\mathrm{var},[s,t]} \\ & + K \left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \omega \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{p-\mathrm{var},[s,t]} \left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert \bar{g}(\cdot, k_\cdot + \mu_\cdot) \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{q-\mathrm{var},[s,t]}, \end{aligned} \tag{4.7}$$

where $q = 1/\beta$, $K = (1 - 2^{1-1/p-1/q})^{-1}$. Since

$$\begin{aligned} & |\bar{g}(t, k_t + \mu_t) - \bar{g}(s, k_s + \mu_s)| \\ & \quad \leq \quad |\bar{g}(t, k_t + \mu_t) - \bar{g}(t, k_s + \mu_s)| + |\bar{g}(t, k_s + \mu_s) - \bar{g}(s, k_s + \mu_s)| \\ & \quad \leq \quad C_g|k_t - k_s| + C_g|\mu_t - \mu_s| + C_g(t-s)^\beta \\ & \quad \leq \quad C_g|k_t - k_s| + M(1 + |\mu_a|^\beta)(t-s)^\beta, \qquad \forall a \leq s < t \leq b \end{aligned}$$

where $M = M(r)$ depends on $r = b - a$, we have

$$\left\| \bar{g}(\cdot, k_\cdot + \mu_\cdot) \right\|_{q-\mathrm{var},[s,t]} \leq C_g \left\| k \right\|_{p-\mathrm{var},[s,t]} + M(1 + |\mu_a|^\beta)(t-s)^\beta,$$

with a note that $q\beta \geq 1$ and $q \geq p$. Then

$$\begin{aligned} |k_t - k_s| &\leq \left[\|g\|_\infty + KM(1 + |\mu_a|^\beta) \right] \left\| \omega \right\|_{p-\mathrm{var},[s,t]} + \int_s^t C_f |k_u| du \\ &\quad + KC_g \left\| \omega \right\|_{p-\mathrm{var},[s,t]} \left\| k \right\|_{p-\mathrm{var},[s,t]}. \end{aligned}$$

Using Lemma 4.1 and Young inequality for product

$$\begin{aligned} \|k\|_{\infty,[a,b]} &\leq e^{2C_f r} \left[|k_a| + M(1 + |\mu_a|^\beta) \left\| \omega \right\|_{p-\mathrm{var},[a,b]} (1 + \left\| \omega \right\|^p_{p-\mathrm{var},[a,b]}) \right] \\ &\leq M(1 + |\mu_a|^\beta) \left\| \omega \right\|_{p-\mathrm{var},[a,b]} (1 + \left\| \omega \right\|^p_{p-\mathrm{var},[a,b]}) \\ &\leq \varepsilon |y_a| + \Lambda(\omega, [a,b]), \end{aligned} \tag{4.8}$$

where $\varepsilon > 0$ is chosen later and $\Lambda(\omega, [a,b])$ is a general polynomial of $\left\| \omega \right\|_{p-\mathrm{var},[a,b]}$.

Step 2: Next, we estimate the solution of (3.6) by discretization. By assumption of f, it can be seen that all $\tilde{f} \in \mathcal{H}^f_{d_0}$ satisfy (4.5). For each n, consider(4.6) with $[a,b]$ is replaced by $[n-1,n]$. By known result of (4.6) under condition (4.5), there exists $\eta \in (0,1), L > 0$ such that

$$|\mu_n| \leq \eta^* |y_n| + L.$$

Now in (4.8), we choose $0 < \varepsilon < 1 - \eta^*$ and $\eta = \eta^* + \varepsilon \in (0,1)$. Then,

$$\begin{aligned} |y_n| &\leq |k_n| + |\mu_n| \\ &\leq \eta |y_{n-1}| + \Lambda(\omega, [n-1,n]). \end{aligned}$$

Therefore,

$$\begin{aligned} |y_n| &\leq \eta |y_{n-1}| + \Lambda(\omega, [n-1,n]) \\ &\leq \eta^n |y_0| + \sum_{j=1}^n \eta^j \Lambda(\omega, [n-1-j, n-j]). \end{aligned} \tag{4.9}$$

Define $R(\bar{\omega}) := \sum_{j \geq 0} \eta^j \Lambda(\omega, [-j, -j+1])$, then as n large enough

$$|y(n, y_0, \theta_{-n}\bar{\omega})| \leq 1 + R(\bar{\omega}).$$

Step 3: Finally, we prove the existence of an absorbing set.

Using (3.3) the value of solution at arbitrary time is evaluated similarly. Namely, there exists a tempered random variable $\tilde{R}(\bar{\omega})$ (see [8]) such that

$$|y(t, y_0, \bar{\theta}_{-t}\bar{\omega})| \leq 1 + \tilde{R}(\bar{\omega})$$

as t large enough. It shows the existence of the absorbing set $\mathcal{B}(\bar{\omega}) = \bar{B}(0, \tilde{R}(\bar{\omega}))$. The proof of this step relies on the ergodicity of canonical space $(\Omega, \mathcal{F}, \mathbb{P}, \theta)$ and ergodic Birkhoff theorem.

Note that $\mathbb{E}\left\|\!\left| B^H \right|\!\right\|^m_{p-\mathrm{var},[0,1]} < \infty$ for all $m \in \mathbb{N}$. This deduces that $\Lambda(\omega)$ and then $\tilde{R}(\omega)$ is also integrable. Moreover, in (4.9), one can evaluate $|y_n|^m$ for any $m > 0$ and choose $\tilde{R}$ to be integrable at arbitrary order m.

The existence of random pullback attractor $\mathcal{A}(\bar{\omega})$ for Φ^* is proved.

■

Theorem 4.2. *If we assume f satisfies uniform one-sided dissipative condition*

$$\langle x - y, f(t,x) - f(t,y) \rangle \leq -L|x-y|^2, \ \forall t, x, y$$

for some $L > 0$. Then there exists $\epsilon > 0$ such that if $C_g < \epsilon$ the attractor is singleton.

Proof. Let y^1, y^2 be two solutions of (3.6) where the initial conditions lie in $\bar{B}(0,R)$. Put $\bar{y} = y^2 - y^1$ then

$$d\bar{y}_t = [\bar{f}(t, \bar{y}_t + y^1_t) - \bar{f}(t, y^1_t)]dt + [\bar{g}(t, y^2_t) - \bar{g}(t, y^1_t)]d\omega_t.$$

Once again, we consider the pure dt equation

$$d\bar{\mu}_t = [\bar{f}(t, \bar{\mu}_t + y^1_t) - \bar{f}(t, y^1_t)]dt, \quad \bar{\mu}_0 = \bar{y}_0.$$

By assumption of f, there exists $\eta \in (0,1)$ such that

$$|\bar{\mu}_1| \leq \eta |\bar{\mu}_0|.$$

Now, put $z = \bar{y} - \bar{\mu}$, we have

$$dz_t = [\bar{f}(t, \bar{y}_t + y^1_t) - \bar{f}(t, \bar{\mu}_t)]dt + [\bar{g}(t, y^2_t) - \bar{g}(t, y^1_t)]d\omega_t.$$

Computation leads to
(4.10)

$$|z_t - z_s| \leq \int_s^t C_f |z_u| du + DC_g \left\|\!\left| \omega \right|\!\right\|_{p-\mathrm{var},[s,t]} . \|z + \bar{\mu}\|_{p-\mathrm{var},[s,t]} (1 + \left\|\!\left| y^1 \right|\!\right\|_{p-\mathrm{var},[s,t]}).$$

By (3.3), for all $s, t \in [0,1]$

$$|z_t - z_s| \leq \int_s^t C_f |z_u| du + DRC_g \left\|\!\left| \omega \right|\!\right\|_{p-\mathrm{var},[s,t]} \Lambda(\omega, [0,1]). \|z + \bar{\mu}\|_{p-\mathrm{var},[s,t]},$$

then using Lemma 4.1,

$$\|z\|_{p-\mathrm{var},[0,1]} \leq DR|\bar{y}_0| C_g e^{RC_g \Lambda(\omega,[0,1])},$$

where $\Lambda(\omega,[a,b])$ is a general polynomial of $\|\omega\|_{p-\mathrm{var},[a,b]}$. We arrive at

$$|\bar{y}_1| \leq \eta|\bar{y}_0|\left[1+DRC_g e^{RC_g\Lambda(\omega,[0,1])}\right]. \tag{4.11}$$

The rest of the proof is followed step by step to [8, Theorem 3.11].

■

Appendix

Proof of Proposition 3.1

Proof. That d_1 is a metric on $\mathcal{C}^{\alpha;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^d\times m)$ is evident due to the seminorm properties of the Hölder norm. We only need to prove the completeness. Let h^n be a Cauchy sequence in $\mathcal{C}^{\alpha;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})$. Since $(\mathcal{C}^{1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^d\times m),\rho)$ is complete, there exists a subsequence, which we still use the notation h^n, converges to h in $\mathcal{C}^{1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})$, i.e.

$$\lim_{n\to\infty}\rho(h^n,h)=0.$$

We will prove that for each K^1,K^2 compact sets in $\mathbb{R},\mathbb{R}^d$, $\|h^n-h\|_{\alpha,K^1\times K^2}\to 0$ as $n\to\infty$. Fix $K\subset\mathbb{R}^d$ compact, we have for each $[a,b]\subset\mathbb{R}$ there exist a constant M such that

$$\sup_n\sup_{x\in K}\|h^n(\cdot,x)\|_{\alpha-\mathrm{Hol},[a,b]}\leq M.$$

For each $x\in K$

$$|h(t,x)-h(s,x)|=\lim_{n\to\infty}|h^n(t,x)-h^n(s,x)|\leq M|t-s|^\alpha,$$

this implies that $\sup_{x\in K}\|h(\cdot,x)\|_{\alpha-\mathrm{Hol},[a,b]}<\infty$ or $h\in\mathcal{C}^{\alpha;1,0}(\mathbb{R}\times\mathbb{R}^d,\mathbb{R}^{d\times m})$.

Now to complete the proof we show that h^n converges to h, in $\alpha-$Hölder norm on each K compact in $\mathbb{R}^d$. For each $s<t\in[a,b]$, $x\in K$

$$\begin{aligned}\frac{|(h^n-h)(t,x)-(h^n-h)(s,x)|}{|t-s|^\alpha} &= \lim_{m\to\infty}\frac{|(h^n-h^m)(t,x)-(h^n-h^m)(s,x)|}{|t-s|^\alpha}\\ &\leq \lim_{m\to\infty}\sup_{x\in K}\sup_{a\leq v<u\leq b}\frac{|(h^n-h^m)(u,x)-(h^n-h^m)(v,x)|}{|u-v|^\alpha}\\ &\leq \lim_{m\to\infty}\|h^n-h^m\|_{\alpha,[a,b]\times K},\end{aligned}$$

which implies

$$\|h^n - h\|_{\alpha,[a,b]\times K} \le \lim_{m\to\infty} \|h^n - h^m\|_{\alpha,[a,b]\times K} \to 0, \text{ as } n \to \infty.$$

The proof is completed.

Proof of Lemma 3.1

Proof.

It can be seen from the assumptions of g that g together with $\partial_x g$ satisfies the condition boundedness and equicontinuous on $\mathbb{R} \times K$ for each K compact in $\mathbb{R}^d$.

Due to [22, Theorem 16] $\mathcal{H}^g_{d_1}$ is compact in $(\mathcal{C}^{1,0}(\mathbb{R}\times\mathbb{R}^d, \mathbb{R}^{d\times m}), \rho)$. Hence, for $g^* \in \mathcal{H}^g_{d_1}$, $\partial_x g^*$ exists and is continuous. Moreover, and there exists t_n such that $\lim_{n\to\infty} d_1(g^*, g_{t_n}) = 0$.

It is evident that g^* is bounded by $\|g\|_\infty$, and

$$\begin{aligned}
|g^*(t,x) - g^*(t,y)| &= \lim_{n\to\infty} |g_{t_n}(t,x) - g_{t_n}(t,y)| \\
&= \lim_{n\to\infty} |g(t_n + t, x) - g(t_n + t, y)| \le C_g |x-y|, \\
|g^*(t,x) - g^*(s,x)| &+ \|\partial_x g^*(t,x) - \partial_x g^*(s,x)\| \\
&= \lim_{n\to\infty} |g_{t_n}(t,x) - g_{t_n}(s,x)| + \|\partial_x g_{t_n}(t,x) - \partial_x g_{t_n}(s,x)\| \\
&\le C_g |t-s|^\beta.
\end{aligned}$$

That $\partial_x g^*(t,x)$ is local Lipschitz in x uniformly in t is also obvious. The first statement is proved.

For the second one, since $\mathcal{H}^g_{d_1}$ is compact in $\mathcal{C}^{1,0}$, from a sequence $h^n \in \mathcal{H}^g_{d_1}$ there exists a subsequence h^{n_k} that converges (in ρ) to $h \in \mathcal{C}^{1,0}(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}^{d\times m})$. One may choose the subsequence in the form g_{t_n}. Applying the above arguments for $g^* = h$ and the sequence g_{t_n} we have $h \in \mathcal{C}^{\beta;1,0}(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}^{d\times m})$. Moreover, $\|h^{n_k}\|_{\beta,K^1\times K^2}$, $\|h\|_{\beta,K^1\times K^2}$ are less than C_g for K^1, K^2 are compact sets in $\mathbb{R}, \mathbb{R}^d$ respectively.

Finally, put $h_k = h^{n_k} - h$. Since $\beta_0 < \beta$, for $s,t \in K^1, x \in K^2$

$$\begin{aligned}
\frac{|h_k(t,x) - h_k(s,x)|}{|t-s|^{\beta_0}} &= \left(\frac{|h_k(t,x) - h_k(s,x)|}{|t-s|^\beta}\right)^{\frac{\beta_0}{\beta}} .|h_k(t,x) - h_k(s,x)|^{1-\frac{\beta_0}{\beta}} \\
&\le \|h_k\|^{\frac{\beta_0}{\beta}}_{\beta,K^1\times K^2} \left(|h_k(t,x)| + |h_k(s,x)|\right)^{1-\frac{\beta_0}{\beta}}, \text{ hence} \\
\|h_k\|_{\beta_0,K^1\times K^2} &\le 4C_g^{\frac{\beta_0}{\beta}} \|h_k\|^{1-\frac{\beta_0}{\beta}}_{\infty,K^1\times K^2} \to 0 \;\; as \;\; k \to \infty.
\end{aligned}$$

To sum up, h^{n_k} converges to h in d_1. The proof is completed.

■

Lemma 4.1 (Gronwall-type Lemma)**.** *For $q \geq p$ so that $\frac{1}{p}+\frac{1}{q} > 1$, if y satisfies the following condition*

$$|y_t - y_s| \leq \hat{A}^{1/q}_{s,t} + a_1 \int_s^t |y_u| du + \|\!|\omega|\!\|_{p,[s,t]} \left(a_2|y_s| + a_3 \|\!|y|\!\|_{q-\mathrm{var},[s,t]}\right)$$

for all $s \leq t \in [a,b]$, where a_1, a_2, a_3 are positive real constants, $\hat{A}$ is a control function on $\{(s,t)|a \leq s \leq t \leq b\}$, then

$$\|y\|_{p,[a,b]} \leq \left[|y_a| + 2\hat{A}^{1/q}_{a,b} N_{[a,b]}\right] e^{2a_1(b-a)+\kappa N_{[a,b]}} N^{\frac{p-1}{p}}_{[a,b]}(\omega)$$

with $\kappa = \log \frac{a_3/a_2+2}{a_3/a_2+1}$, and

$$N_{[a,b]} \leq D(1 + \|\!|\omega|\!\|^p_{p,[a,b]})$$

for D depends on a_i. If $a_2 = 0$ one may take $\kappa = 0$.

Proof. See [8, Theorem 2.4].

■

Acknowledgments

This work is supported by the project ICRTM04_2024.02 of International Centre of Research and Postgraduate Training in Mathematics, Institute of Mathematics, Vietnam Academy of Science and Technology.

References

[1] **L. Arnold**, Random Dynamical Systems. Springer, Berlin Heidelberg New York, 1998.

[2] **I. Bailleul, S. Riedel, M. Scheutzow**, *Random dynamical systems, rough paths and rough flows.* J. Differential Equations, Vol. **262**, (2017), 5792–5823.

[3] **C. Castaing, M. Valadier** , *Convex analysis and measurable multifunctions. Lecture Notes in Math.* No 580, Springer-Verlag, Berlin, Heidelberg, New York, 1977

[4] **N. D. Cong, L. H. Duc, P. T. Hong**, *Young differential equations revisited.* J. Dyn. Diff. Equat., Vol. **30**, Iss. **4**, (2018), 1921–1943.

[5] **H. Crauel, P. Kloeden**, *Nonautonomous and random attractors.* Jahresber Dtsch. Math-Ver. **117** (2015), 173–206.

[6] **P. Kloeden, M. Rasmussen**, Nonautonomous Dynamical System. Mathematical Surveys and Monographs. https://doi.org/10.1090/surv/176 (2011).

[7] **L. H. Duc, P. T. Hong, N. D. Cong**, *Asymptotic stability for stochastic dissipative systems with a Hölder noise.* SIAM J. Control Optim.,Vol. 57, No. 4, pp. 3046 - 3071, (2019)

[8] **L. H. Duc, P. T. Hong**, *Asymptotic Dynamics of Young Differential Equations* . J. Dyn. Diff. Equat. 35, 1667–1692 (2023).

[9] **L. H. Duc, B. Schmalfuß, S. Siegmund**, *A note on the generation of random dynamical systems from fractional stochastic delay differential equations.* Stoch. Dyn., Vol. **15**(2015), No. 3, 1–13.

[10] **L. H. Duc**, *Random attractors for dissipative systems with rough noises* Discrete & Continuous Dynamical Systems 2022, 42(4): 1873-1902

[11] **P. Friz, N. Victoir**, Multidimensional stochastic processes as rough paths: theory and applications. Cambridge Unversity Press, Cambridge, 2010.

[12] **M. Garrido-Atienza, B. Maslowski, B. Schmalfuß**, *Random attractors for stochastic equations driven by a fractional Brownian motion.* International Journal of Bifurcation and Chaos, Vol.**20**, No. 9 (2010) 2761–2782.

[13] **M. Garrido-Atienza, B. Schmalfuss**, *Ergodicity of the infinite dimensional fractional Brownian motion.* J. Dyn. Diff. Equat., **23**, (2011), 671–681. DOI 10.1007/s10884-011-9222-5.

[14] **P. T. Hong**, *Nonautonomous attractors for Young differential equations driven by unbounded variation paths.* Journal of Mathematics and Science Thang Long University, Vol. **1**, No. 4 (2022).

[15] **Y. Hu**, Analysis on Gaussian Spaces. World scientific Publishing, (2016).

[16] **T. Lyons,** *Differential equations driven by rough signals, I, An extension of an inequality of L.C. Young.* Math. Res. Lett. **1**, (1994), 451–464.

[17] **B. Mandelbrot, J. van Ness,** *Fractional Brownian motion, fractional noises and applications.* SIAM Review, **4**, No. 10, (1968), 422–437.

[18] **V. V. Nemytskii, V. V. Stepanov**, Qualitative Theory of Differential Equations. GITTL, Moscow–Leningrad. (1949). English translation, Princeton University Press, (1960).

[19] **R. J. Sacker, G. Sell**, *Lifting properties in skew-product flows with applications to differential equations.* Memoirs of the American Mathematical Society, Vol. **11**, No. 190, (1977).

[20] **B. Schmalfuß**, *Attractors for the nonautonomous dynamical systems.* In K. Groger, B. Fiedler and J. Sprekels, editors, Proceedings EQUADIFF99, World Scientific, (2000), 684–690.

[21] **B. Schmalfuß.** *Attractors for the nonautonomous and random dynamical systems perturbed by impulses.* Discret. Contin. Dyn. Syst., **9**, No. 3, (2003), 727–744.

[22] **G. Sell**, *Nonautonomous differential equations and topological dynamics. I. The Basic theory* Transactions of the American Mathematical Society, Vol. **127**, No. 2 (1967), 241-262.

[23] **G. Sell**, *Nonautonomous Differential Equations and Topological Dynamics. II. Limiting Equations* Transactions of the American Mathematical Society, Vol. **127**, No. 2 (1967), 263-283

[24] **L.C. Young,** *An integration of Hölder type, connected with Stieltjes integration.* Acta Math. **67**, (1936), 251–282.

Phan Thanh Hong
Department of Mathematics, Thang Long University
Hanoi
Vietnam
and
International Centre of Research and Postgraduate Training in Mathematics,
Institute of Mathematics, Vietnam Academy of Science and Technology
Hanoi
Vietnam
hongpt@thanglong.edu.vn

J. Math. Math. Sci. (2025) 95–103

Images of the Singer transfers and their possibility to be injective

Nguyễn H. V. Hưng (Hanoi, Vietnam)

(Received Apr. 3, 2025; accepted May 4, 2025)

Abstract. This article is an attempt to investigate the possibility to be injective of the Singer transfer $\mathrm{Tr}_s^M : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes M_*) \to \mathrm{Ext}_{\mathscr{A}}^s(\Sigma^{-s}M, \mathbb{F}_2)$ for M being the $\mathscr{A}$-modules $\mathbb{F}_2 = \widetilde{H}^*S^0$ or $\widetilde{H}^*\mathbb{RP}^\infty$. The existence of a positive stem critical element of $\mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ in the image of the transfer $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ is equivalent to the existence of a positive stem critical element of $\mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$ in the image of the transfer Tr_{s+1}. If the existences happen, then $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ and Tr_{s+1} are not injective. We show that the critical element $\widehat{Ph_2}$ is not in the image of the fourth transfer, $\mathrm{Tr}_4^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_4} P(H_*\mathbb{V}_4 \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-4} \to \mathrm{Ext}_{\mathscr{A}}^{4,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$. Singer's conjecture is still open, as we have not known any critical element, which is in the image of the transfer.

1. Recollections on the Singer transfers and related topics

We sketch briefly the Singer transfer, which is the subject of this article.

Let $\mathscr{A}$ be the mod 2 Steenrod algebra. Singer defined in [10] the algebraic transfer for an $\mathscr{A}$-module M:

$$\mathrm{Tr}_s^M : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes M_*) \to \mathrm{Ext}_{\mathscr{A}}^s(\Sigma^{-s}M, \mathbb{F}_2),$$

Key words and phrases: Steenrod algebra, Adams spectral sequences, Singer transfer, Invariant theory.
2020 Mathematics Subject Classification: Primary 55T15, 55Q45, 55S10, 55P47.

where $\mathbb{V}_s$ denotes an elementary abelian 2-group of rank s, and $H_*\mathbb{V}_s$ is the mod 2 homology of a classifying space $B\mathbb{V}_s$ of $\mathbb{V}_s$, while M_* is the dual of the $\mathscr{A}$-module M, and $P(H_*\mathbb{V}_s \otimes M_*)$ denotes the primitive part of $H_*\mathbb{V}_s \otimes M_*$ under the action of $\mathscr{A}$. The Singer transfer is a useful tool in the study of $\mathrm{Ext}^s_{\mathscr{A}}(M, \mathbb{F}_2)$ by means of the Peterson hit problem and invariant theory.

Let $M = \widetilde{H}^*X$ be the reduced cohomology of X. Here X is a pointed CW-complex, whose mod 2 homology H_*X is finitely generated in each degree. Then the Singer transfer for $\widetilde{H}^*X$ is also called the Singer transfer for X, that is $\mathrm{Tr}^X_s := \mathrm{Tr}^{\widetilde{H}^*X}_s$. It is a remarkable tool to study $\mathrm{Ext}^s_{\mathscr{A}}(\widetilde{H}^*X, \mathbb{F}_2)$. The interest in studying $\mathrm{Ext}^s_{\mathscr{A}}(\widetilde{H}^*X, \mathbb{F}_2)$ is that this forms the E_2-page of the Adams spectral sequence converging to the 2-completion of the stable homotopy groups $\pi^S_*(X)$.

Let $t : \mathbb{RP}^\infty \to S^0$ be any map that induces an isomorphism in the first stable homotopy group π^S_1. The so-called algebraic Kahn-Priddy homomorphism $t_* : \mathrm{Ext}^s_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}^{s+1}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ is its "coboundary" one. This is an useful manner to attack the cohomology of the Steenrod algebra $\mathscr{A}$. The reason why the Kahn-Priddy map and particularly the infinite real projective space are taken into account is that this homomorphism is an epimorphism in positive stems and further it lowers the cohomology degree by relating $\mathrm{Ext}^{s+1}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ to $\mathrm{Ext}^s_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$.

It should be noted that the Singer map

$$\mathrm{Ext}^{i,j}_{\mathscr{A}}(H^*\mathbb{V}_s, \mathbb{F}_2) \to \mathrm{Ext}^{i+s,j+s}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2),$$

which becomes the Singer transfer for $i = 0$, is absolutely not the Kahn-Priddy map even for $s = 1$. The subtlety comes from the fact that the Kahn-Priddy map works with reduced cohomology, while the Singer map works with cohomology.

In this article, we follow all the notations of our preceding paper [4].

J. P. May proved in [8] that: As $\mathscr{A}$ is a cocommutative Hopf algebra, if M is a coalgebra in the category of $\mathscr{A}$-modules and N is an algebra in this category, then there exist Steenrod operations

$$Sq^i : \mathrm{Ext}^{s,t}_{\mathscr{A}}(M, N) \to \mathrm{Ext}^{s+i,2t}_{\mathscr{A}}(M, N).$$

In particular, for $i = 0$, $Sq^0 : \mathrm{Ext}^{s,t}_{\mathscr{A}}(M, N) \to \mathrm{Ext}^{s,2t}_{\mathscr{A}}(M, N)$.

By checking on bi-grading, $Sq^0 : \mathrm{Ext}^{s,t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}^{s,2t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ and $Sq^0 : \mathrm{Ext}^{s,t}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2) \to \mathrm{Ext}^{s,2t}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ can not commute with each other through the algebraic Kahn-Priddy homomorphism

$$t_* : \mathrm{Ext}^{s,t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}^{s+1,t+1}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2).$$

Indeed, on the one hand t_*Sq^0 sends the bi-degree (s,t) to $(s+1, 2t+1)$, on the other hand $Sq^0 t_*$ sends the bi-degree (s,t) to $(s+1, 2(t+1))$.

Similarly, $Sq^0 : \mathrm{Ext}^{s,t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}^{s,2t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ and the Kameko one $Sq^0 : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s} \to \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{2(t-s)+s+1}$ can not commute with each other through the Singer transfer $\mathrm{Tr}^{\mathbb{RP}^\infty}_s : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s} \to \mathrm{Ext}^{s,t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ because of bi-grading reason.

W. H. Lin defined in [7, p. 469] a map which is also denoted

$$Sq^0 : \mathrm{Ext}^{s,t}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}^{s,2t+1}_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$$

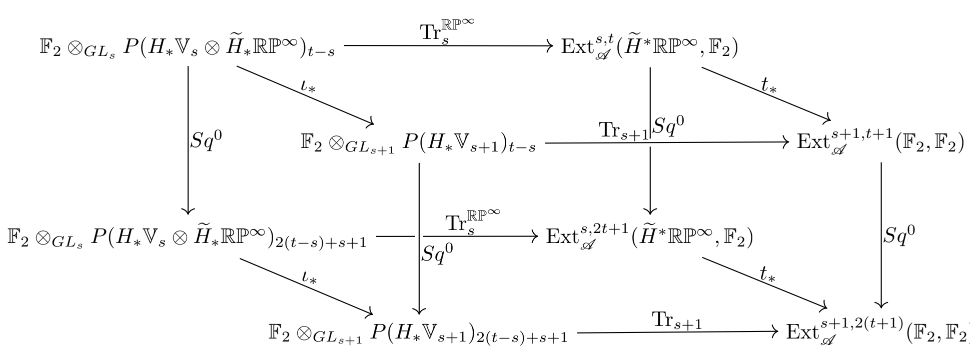
by ambiguity of notation. Remarkably, this Sq^0 commutes with the classical squaring operation $Sq^0 : \mathrm{Ext}^{s,t}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2) \to \mathrm{Ext}^{s,2t}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ through the algebraic Kahn-Priddy morphism, and also commutes with the Kameko one through the Singer transfer $\mathrm{Tr}^{\mathbb{RP}^\infty}_s$ (see [4, Prop. 4.1] or Proposition 1.1 below).

The operation Sq^0 defined by Lin on $\mathrm{Ext}^*_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ should have been called Lin's Sq^0. Note that, Lin's Sq^0 is not May's Sq^0. In the article we only use Lin's Sq^0 on $\mathrm{Ext}^*_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$. So, this is simply called Sq^0 for abbreviation.

We start with a commutative diagram and the concept of critical elements given in [2] and [4].

In the following diagram, the horizontal arrows are the Singer transfers, the two vertical right arrows are the squaring operations and the two vertical left arrows are the Kameko squaring ones, while t_* denotes the algebraic Kahn-Priddy morphism, and ι_* is the homomorphism induced from the canonical inclusion.

Proposition 1.1. ([4, Prop. 4.1]) *The diagram*



is commutative.

We created the concept of critical elements (in [2], [3], [4]) in order to show that, in general, at most of homological degrees, the Singer transfers Tr_s and $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ are not an isomorphism in infinitely many internal degrees. However, Singer's conjecture, which predicts that the Singer transfers is injective, is still open.

In particular, we defined on [2, page 2] the notion of s-spike as follows: An s-spike number is an one that can be written as $(2^{n_1}-1)+\cdots+(2^{n_s}-1)$, but cannot be written as a sum of less than s terms of the form (2^n-1).

Definition 1.1. ([2, Def. 5.2]) A nonzero element $z \in \mathrm{Ext}^s_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ is called *critical* if

(a) $Sq^0(z) = 0$,

(b) $2\mathrm{Stem}(z) + s$ is an s-spike.

By [2, Lemma 3.5], if $\mathrm{Stem}(z)$ is an s-spike, then so is $2\mathrm{Stem}(z) + s$.

Note that $Ph_1 \in \mathrm{Ext}^{5,14}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ is not a critical element, since $2\mathrm{Stem}(Ph_1) + 5 = 23$ is a 3-spike but not 5-spike. Actually, $23 = (16-1)+(8-1)+(2-1)$, and it is easy to see that 23 cannot be written as a sum of less than 3 terms of the form (2^n-1).

However, $Ph_2 \in \mathrm{Ext}^{5,16}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ is critical (see [2, Prop. 5.5]). Indeed,

(a) $Sq^0(Ph_2) = 0$, as is well known $\mathrm{Ext}^{5,32}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2) = 0$ (see e.g. Tangora [11]),

(b) $27 = 2\mathrm{Stem}(Ph_2) + 5$ is a 5-spike (see [2, Lemma 3.3]).

Definition 1.2. ([4, Def. 6.3]) A nonzero element $\widehat{z} \in \mathrm{Ext}^s_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ is called *critical* if

(a) $Sq^0(\widehat{z}) = 0$,

(b) $2\mathrm{Stem}(\widehat{z}) + (s+1)$ is an $(s+1)$-spike,

(c) $t_*(\widehat{z}) \neq 0$, where $t_* : \mathrm{Ext}^s_{\mathscr{A}}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}^{s+1}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ is the Kahn-Priddy homomorphism.

2. Results

The motivation for us to be interested in Proposition 2.3 is that the following theorem would probably give a negative answer to Singer's conjecture [10] on the transfer monomorphism. More precisely, we have

Theorem 2.1. *(i) If a critical element* $z \in \mathrm{Ext}_{\mathscr{A}}^{s,t}(\mathbb{F}_2, \mathbb{F}_2)$ *is in the image of the transfer* $\mathrm{Tr}_s : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s)_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s,t}(\mathbb{F}_2, \mathbb{F}_2)$*, then* Tr_s *is not a monomorphism.*

(ii) If a critical element $\widehat{z} \in \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ *is in the image of the transfer* $\mathrm{Tr}_s^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$*, then* $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ *is not a monomorphism.*

Proof. (i) See Case 2 of [2, Thm. 5.6] and [2, Thm. 5.9].

(ii) We are using the notation of Proposition 1.1.

We prove the following fact: If $\widehat{z} \in \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ is a critical element, which is in the image of the transfer $\mathrm{Tr}_s^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$, then $z = t_*(\widehat{z}) \in \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$ is also a critical element, which is in the image of the transfer $\mathrm{Tr}_{s+1} : \mathbb{F}_2 \otimes_{GL_{s+1}} P(H_*\mathbb{V}_{s+1})_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$. See Case 2 in Part (ii) of [4, Thm. 6.7].

Indeed, there is an element $\widehat{y} \in \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s}$ such that $\mathrm{Tr}_s^{\mathbb{RP}^\infty}(\widehat{y}) = \widehat{z}$. Then $z = t_*(\widehat{z}) \in \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$ is nonzero (by (c) of Definition 1.2) and it is a critical element, which is in the image of the transfer $\mathrm{Tr}_{s+1} : \mathbb{F}_2 \otimes_{GL_{s+1}} P(H_*\mathbb{V}_{s+1})_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$. (It should be noted that z satisfies Definition 1.1 with the number $s+1$ taken into account, instead of s.) Actually, we get

(a) $Sq^0(z) = Sq^0 t_*(\widehat{z}) = t_* Sq^0(\widehat{z}) = t_*(0) = 0$,

(b) $2\mathrm{Stem}(z) + (s+1) = 2\mathrm{Stem}(\widehat{z}) + (s+1)$ is an $(s+1)$-spike.

Denoting $y = \iota_*(\widehat{y})$, we have $\mathrm{Tr}_{s+1}(y) = \mathrm{Tr}_{s+1}\iota_*(\widehat{y}) = t_*\mathrm{Tr}_s^{\mathbb{RP}^\infty}(\widehat{y}) = t_*(\widehat{z}) = z$. That is, z is in the image of the transfer Tr_{s+1}. As z is nonzero and $z = \mathrm{Tr}_{s+1}(y)$, it implies that y is nonzero.

Note that, as $\mathrm{Tr}_{s+1}(y) = z$, by definition of the Singer transfer, it implies $\deg(y) = \mathrm{Stem}(z)$. From (b) of Definition 1.2, it concludes that $2\deg(y)+(s+1)$ is an $(s+1)$-spike. So, according to Kameko [5] (see also [2, Cor. 3.8]),

$$Sq^0 : \mathbb{F}_2 \otimes_{GL_{s+1}} P(H_*(\mathbb{V}_{s+1})_{t-s} \to \mathbb{F}_2 \otimes_{GL_{s+1}} P(H_*(\mathbb{V}_{s+1})_{2(t-s)+s+1}$$

is an isomorphism. In particular, from $y \neq 0$ it implies $Sq^0(y) \neq 0$. By the commutativity in Proposition 1.1, we have

$$\iota_* Sq^0(\widehat{y}) = Sq^0 \iota_*(\widehat{y}) = Sq^0(y) \neq 0.$$

Therefore $Sq^0(\widehat{y}) \neq 0$. Also, by the commutativity in Proposition 1.1,

$$\mathrm{Tr}_s^{\mathbb{RP}^\infty} Sq^0(\widehat{y}) = Sq^0 \mathrm{Tr}_s^{\mathbb{RP}^\infty}(\widehat{y}) = Sq^0(\widehat{z}) = 0,$$

(by (a) of Definition 1.2). That is, $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ sends a nonzero element to zero, so it is not a monomorphism.

The theorem is completely proved. ■

Theorem 2.2. *The existence of a positive stem critical element*

$$\widehat{z} \in \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$$

in the image of the transfer $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ *is equivalent to the existence of a positive stem critical element* $z \in \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$ *in the image of the transfer* Tr_{s+1}*. If the existences happen, then both* $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ *and* Tr_{s+1} *are not injective.*

Proof. From the beginning of the proof for Theorem 2.1, if there exists a critical element $\widehat{z} \in \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$, which is in the image of the transfer $\mathrm{Tr}_s^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$, then, by Definitions 1.1 and 1.2, $z = \widehat{z}$ is a critical element, which is in the image of the transfer $\mathrm{Tr}_{s+1} : \mathbb{F}_2 \otimes_{GL_{s+1}} P(H_*\mathbb{V}_{s+1})_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$.

Conversely, suppose there exists a positive stem critical element

$$z \in \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2),$$

which is in the image of the transfer

$$\mathrm{Tr}_{s+1} : \mathbb{F}_2 \otimes_{GL_{s+1}} P(H_*\mathbb{V}_{s+1})_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2).$$

From [4, Thm. 1.1], the algebraic Kahn-Priddy homomorphism

$$t_* : \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2) \to \mathrm{Ext}_{\mathscr{A}}^{s+1,t+1}(\mathbb{F}_2, \mathbb{F}_2)$$

is an epimorphism from $\mathrm{Im}\mathrm{Tr}_*^{\mathbb{RP}^\infty}$ onto $\mathrm{Im}\mathrm{Tr}_*$ in stem $t-s > 0$. So there exists $\widehat{z} \in \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ with $t_*(\widehat{z}) = z$, which is in the image of the transfer $\mathrm{Tr}_s^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_s} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-s} \to \mathrm{Ext}_{\mathscr{A}}^{s,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$. By combination of Definitions 1.1 and 1.2, $\widehat{z}$ is also a critical element.

According to Theorem 2.1, if the existences happen, then both $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ and Tr_{s+1} are not injective. The theorem is proved. ■

The following shows that Theorem 2.1(*ii*) can not be applied to $\widehat{Ph_2}$. Therefore, Singer's conjecture is still open, as we have not known any critical element, which is in the image of the transfer.

Proposition 2.3. *The elements* $\widehat{Ph_1}$ *and* $\widehat{Ph_2}$ *are not in the image of the fourth transfer for* $\mathbb{RP}^\infty$*:* $\mathrm{Tr}_4^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_4} P(H_*\mathbb{V}_4 \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-4} \to \mathrm{Ext}_{\mathscr{A}}^{4,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$.

Proof. According to Singer [10, Prop. 13.3],

$$(\mathbb{F}_2 \otimes_{\mathscr{A}} H^*(\mathbb{V}_5))_9^{GL_5} = 0.$$

By duality, Ph_1 is not in the image of the fifth transfer for S^0:

$$\mathrm{Tr}_5 : \mathbb{F}_2 \otimes_{GL_5} P(H_*\mathbb{V}_5)_9 \to \mathrm{Ext}_{\mathscr{A}}^{5,14}(\mathbb{F}_2, \mathbb{F}_2),$$

as the domain is zero in degree 9.

From Quỳnh [9, Prop. 1.3],

$$(\mathbb{F}_2 \otimes_{\mathscr{A}} H^*(\mathbb{V}_5))_{11}^{GL_5} = 0.$$

Passing to the duality, Ph_2 is not in the image of the fifth transfer for S^0:

$$\mathrm{Tr}_5 : \mathbb{F}_2 \otimes_{GL_5} P(H_*\mathbb{V}_5)_{11} \to \mathrm{Ext}_{\mathscr{A}}^{5,16}(\mathbb{F}_2, \mathbb{F}_2),$$

as the domain is zero in degree 11.

Lin [7] and Chen [1] constructed elements $\widehat{Ph_1}$ and $\widehat{Ph_2}$ in $\mathrm{Ext}_{\mathscr{A}}^{4,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2)$ for respectively $t = 14$ and $t = 16$, whose behaviors are given by the algebraic Kahn-Priddy homomorphism

$$t_*(\widehat{Ph_i}) = Ph_i, \quad (i = 1, 2).$$

Now we show that $\widehat{Ph_i}$ are not in the image of the fourth transfer for $\mathbb{RP}^\infty$

$$\mathrm{Tr}_4^{\mathbb{RP}^\infty} : \mathbb{F}_2 \otimes_{GL_4} P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-4} \to \mathrm{Ext}_{\mathscr{A}}^{4,t}(\widetilde{H}^*\mathbb{RP}^\infty, \mathbb{F}_2),$$

for respectively $t = 13$ and $t = 15$. Suppose the contrary that $\widehat{Ph_i}$ is in the image of the transfer. That is, there exists $\widehat{z}_i \in P(H_*\mathbb{V}_s \otimes \widetilde{H}_*\mathbb{RP}^\infty)_{t-4}$ such that $\mathrm{Tr}_4^{\mathbb{RP}^\infty}(\widehat{z}_i) = \widehat{Ph_i}$ Since the commutativity of the diagram in [4, Lemma 4.6], we have

$$\begin{aligned} \mathrm{Tr}_5 \iota_*(\widehat{z}_i) &= t_* \mathrm{Tr}_4^{\mathbb{RP}^\infty}(\widehat{z}_i) \\ &= t_*(\widehat{Ph_i}) = Ph_i. \end{aligned}$$

So, it concludes that Ph_i is in the image of the fifth transfer for S^0. This contradicts to the result by Singer for $i = 1$ or by Quỳnh for $i = 2$. This contradiction rejects the contrary hypothesis. The proposition is proved. ∎

Remark 2.4. In [4, Prop. 6.5], we actually prove that the set $\{h_{n_1} \cdots h_{n_k} \widehat{Ph_2}\}$ contains infinitely many critical elements, where h_n denotes the well-known Adams element. However, so far we have not known whether there is any element of the form $h_{n_1} \cdots h_{n_k} \widehat{Ph_2}$, which is in the image of the transfer $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$. Therefore, Singer's conjecture on the injectivity of $\mathrm{Tr}_s^{\mathbb{RP}^\infty}$ and Tr_{s+1} is still open.

Acknowledgment. This research was carried out when the author recently visited the Vietnam Institute for Advanced Study in Mathematics (VIASM), Hanoi. He would like to express his warmest thank to the VIASM for the hospitality and for the wonderful working condition.

Also, the author would like to thank Võ T. N. Quỳnh for drawing his attention to her article [9].

References

[1] **T. W. Chen**, *Determination of* $\mathrm{Ext}_{\mathscr{A}}^{5,*}(\mathbb{Z}/2, \mathbb{Z}/2)$, Topology and its Applications **158** (2011), 660–689.

[2] **Nguyễn H. V. Hưng**, *The cohomology of the Steenrod algebra and Representations of the general linear groups*, Trans. Amer. Math. Soc. **357** (2005), 4065–4089.

[3] **Nguyễn H. V. Hưng and Lưu X. Trường**, *The algebraic transfer for the infinite real projective space*, C. R. Acad. Sci. Paris, Series I, **357** (2019), 111–114.

[4] **Nguyễn H. V. Hưng and Lưu X. Trường**, *The Singer transfer for infinite real projective space*, Forum Math. **34** (2022), 1433–1462.

[5] **M. Kameko,** *Products of projective spaces as Steenrod modules*, Thesis, Johns Hopkins University 1990.

[6] **W. H. Lin**, *Algebraic Kahn-Priddy theorem*, Pacific J. Math. **96** (1981), 435–455.

[7] **W. H. Lin**, $\mathrm{Ext}_{\mathscr{A}}^{4,*}(\mathbb{Z}/2, \mathbb{Z}/2)$ *and* $\mathrm{Ext}_{\mathscr{A}}^{5,*}(\mathbb{Z}/2, \mathbb{Z}/2)$, Topology and its Applications, **155** (2008), 459–496.

[8] **J. P. May**, *A general algebraic approach to Steenrod operations*, Lect. Notes in Math. Vol. **168**, 153–231, Springer-Verlag, 1970.

[9] **Võ T. N. Quỳnh**, *On behavior of the fifth algebraic transfer*, Proceedings of the School and Conference in Algebraic Topology, 309–326, Geom. Topol. Monogr., 11, Geom. Topol. Publ., Coventry, 2007.

[10] **W. M. Singer**, *The transfer in homological algebra*, Math. Zeit. **202** (1989), 493–523.

[11] **M. C. Tangora**, *On the cohomology of the Steenrod algebra*, Math. Zeit. **116** (1970), 18–64.

Nguyễn H. V. Hưng
Thang Long Institute of Mathematics and Applied Sciences (TIMAS),
Thang Long University
Hanoi
Vietnam
hungnhv@thanglong.edu.vn

Balancing accuracy and interpretability in credit risk modeling: Evidence from peer-to-peer lending

Dinh Thuy Tien (Hanoi, Vietnam)

Nguyen Thi Tra My (Hanoi, Vietnam)

(Received Apr. 5, 2025; accepted May 22, 2025)

Abstract. Accurate credit risk assessment is crucial for the stability and growth of peer-to-peer (P2P) lending platforms. This study investigates the effectiveness of machine learning models in predicting loan defaults using historical Lending Club data. We evaluate logistic regression, decision tree, and random forest, employing feature engineering techniques like one-hot and weight of evidence encoding. Model performance is assessed using K-fold cross-validation and metrics such as accuracy and AUC. To enhance model interpretability, we utilize explainable AI techniques like LIME and SHAP, enabling lenders and borrowers to understand the factors driving loan defaults. Our findings demonstrate that while complex models offer higher predictive accuracy, simpler models like logistic regression with WoE encoding provide a suitable balance between accuracy and interpretability, fostering trust and responsible lending within the P2P lending ecosystem.

1. Introduction

The emergence of financial technology (Fintech) is widely recognized as one of the most significant innovations in the financial sector, reshaping the delivery and consumption of financial services at an unprecedented pace [1]. Broadly, Fintech solutions fall into two categories: those designed for individual consumers such as personal financial management, investment, and lending

Key words and phrases: Peer-to-Peer Lending, Credit Risk Assessment, Logistic Regression, Random Forest, Weight of Evidence Encodings, Explainable AI, LIME, SHAP, Model Interpretability, Lending Club

and those developed for financial institutions, offering solutions like customer identification and credit scoring (ISB, 2025).

Among consumer-facing Fintech innovations, Peer-to-Peer (P2P) lending has gained considerable traction as a disruptive force in traditional lending markets. In this model, individual lenders provide unsecured loans directly to borrowers through online platforms, bypassing conventional financial intermediaries. P2P lending platforms operate similarly to marketplace disruptors like Uber and Grab, facilitating connections between lenders and borrowers at scale.

Since the launch of the first P2P platform, Zopa, in the UK in 2005, the industry has grown rapidly across the globe [2]. The Credit Committee on the Global Financial System and the Financial Stability Board have identified China, the United States, and the United Kingdom as the largest P2P markets, with outstanding P2P credit in China reaching USD 99.7 billion, followed by USD 34.3 billion in the U.S. and USD 4.1 billion in the UK.

Several factors have contributed to this surge. The aftermath of the 2008 global financial crisis led to stricter regulatory capital requirements and constrained lending by traditional banks [3]. Simultaneously, declining interest rates on savings accounts pushed investors to seek higher-yield alternatives. P2P lending has also appealed to underserved segments, such as small and medium enterprises (SMEs) and rural populations who are often excluded from formal banking channels. In addition, the proliferation of mobile technology and the internet has created an infrastructure that supports the scalability of digital lending platforms.

Despite its rapid growth, the P2P lending industry faces significant challenges, particularly in ensuring robust credit risk assessment to maintain platform stability and protect stakeholders. The motivation for this study stems from the critical need to develop accurate and interpretable credit risk models that can effectively predict loan defaults while meeting regulatory and ethical standards. In P2P lending, where individual investors bear the financial risk of borrower defaults, inaccurate credit assessments can lead to substantial losses, erode investor confidence, and undermine the sustainability of lending platforms. Moreover, the lack of transparency in credit decision-making processes can exacerbate issues of trust and fairness, particularly for borrowers who may be denied loans without clear justifications. These challenges are compounded by the increasing complexity of modern credit datasets, which include diverse borrower attributes and require sophisticated analytical approaches to uncover meaningful patterns.

Traditional credit scoring methods, such as logistic regression and statistical scoring models, have long been valued for their simplicity and interpretability, making them suitable for regulatory compliance and stakeholder communication. However, these methods often struggle to capture the non-linear relationships and high-dimensional interactions present in large-scale P2P lending datasets. In contrast, machine learning (ML) techniques, such as decision trees and random forests, excel at modeling complex patterns and improving predictive accuracy, but their "black-box" nature poses challenges in regulated financial environments where explainability is paramount. For instance, regulations like the Fair Credit Reporting Act (1970) in the United States and the General Data Protection Regulation (GDPR) (2018) in the European Union mandate that lenders provide clear explanations for credit decisions, a requirement that complex ML models struggle to meet without additional interpretability tools. The integration of machine learning and mathematical models offers a promising solution to address these dual objectives of accuracy and interpretability. By combining the predictive power of advanced ML algorithms with the transparency of traditional mathematical frameworks, such as logistic regression with Weight of Evidence (WoE) encoding, this study aims to develop credit risk models that are both highly accurate and easily interpretable. Furthermore, the incorporation of explainable AI (XAI) techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), allows us to bridge the interpretability gap for complex ML models, enabling lenders to understand and communicate the factors driving credit decisions.

This study makes several key contributions to the field of credit risk assessment in P2P lending. First, we provide a comprehensive evaluation of machine learning models (logistic regression, decision trees, and random forests) under two preprocessing strategies-Weight of Evidence (WoE) encoding and one-hot encoding with min-max scaling-using a real-world dataset from Lending Club. This analysis identifies optimal modeling approaches that balance predictive accuracy with interpretability, offering practical guidance for P2P lending platforms. Second, we demonstrate the effectiveness of integrating traditional mathematical models with advanced ML techniques, showing that logistic regression with WoE encoding achieves a desirable trade-off between performance and transparency, while random forests enhanced with XAI tools deliver superior accuracy with actionable explanations. Third, we apply LIME and SHAP to interpret complex ML models, providing both local and global insights into the factors driving credit decisions, which supports regulatory compliance and enhances stakeholder trust. Finally, our findings contribute to the responsible deployment of ML in P2P lending by proposing a framework that aligns with regulatory expectations, promotes fair lending practices, and fosters transparency in credit scoring, thereby supporting the sustainable growth of the P2P lending ecosystem.

However, the growth of P2P lending raises concerns around regulatory oversight, consumer protection, and systemic risk. Countries like China and the U.S. have implemented regulatory safeguards-such as prohibiting platforms from holding client funds or disbursing loans directly in order to reduce risks such as fraud, mismanagement, and financial exclusion. Legal risks, including Ponzi-like schemes and investor discrimination, remain ongoing challenges, and concerns about transparency, liquidity, and platform viability persist [4] [5]. These developments underscore the need for robust credit risk models that can assess borrower quality and support the responsible expansion of P2P lending.

Given the scale and risks involved in digital lending ecosystems, credit default prediction has become an essential task in modern financial services. Building reliable models that can identify borrowers likely to default is crucial for mitigating financial loss, maintaining investor trust, and complying with regulatory standards. This paper focuses on building and interpreting machine learning models for predicting credit default risk using a real-world dataset from Lending Club, one of the largest P2P lending platforms in the United States.

2. Literature Review

Credit risk modeling has long been central to financial decision-making, with early work relying on traditional statistical techniques [6]. As credit markets expanded and digital lending platforms emerged, machine learning (ML) approaches have increasingly been used to enhance prediction accuracy and scale model deployment. Among these, logistic regression, decision trees, and random forests remain some of the most widely applied methods in consumer credit risk, including in peer-to-peer (P2P) lending [7] [8], mortgage default [9] [10], and credit card repayment modeling [11].

The rapid growth of digital lending platforms, particularly peer-to-peer (P2P) lending, has spurred research into advanced machine learning (ML) techniques for credit risk assessment, with a shared goal of improving predictive accuracy while addressing interpretability challenges in regulated financial environments. Ma et al. (2018) [7] made a significant contribution by addressing the critical problem of predicting loan defaults in P2P lending networks, aiming to enhance risk assessment for online platforms. They employed gradient boosting algorithms, specifically LightGBM and XGBoost, on a high-dimensional dataset from a Chinese P2P lending platform, incorporating borrower demographics, credit history, and transaction records, with preprocessing to handle missing values and outliers. Their results demonstrated superior performance, with LightGBM achieving an AUC of 0.85, underscoring the power of gradient boosting to capture complex, non-linear patterns in P2P lending data and setting a benchmark for predictive modeling in this domain, which directly informs our study's exploration of ML in P2P credit scoring.

Similarly, Duan (2019) [8] tackled credit default prediction across various lending contexts, focusing on modeling financial system risk. The study utilized deep neural networks (DNNs) on a proprietary consumer loan dataset, including features like credit scores, debt-to-income ratios, and payment histories. The DNNs achieved an accuracy of 0.82, surpassing logistic regression, but their complexity highlighted the need for interpretability in regulated settings, a challenge that aligns with our emphasis on explainable models for P2P lending.

In the mortgage sector, Sirignano et al. (2016) [9] addressed the problem of predicting default risk, a priority following the 2008 financial crisis. They applied recurrent neural networks (RNNs) to a large U.S. mortgage loan dataset, incorporating time-series data on payment behavior and macroeconomic indicators. Their model achieved an AUC of 0.78, demonstrating the ability of deep learning to model temporal dependencies, though the lack of interpretability posed limitations, reinforcing the need for explainable AI (XAI) methods in our work.

Kvamme et al. (2018) [10] also focused on mortgage default prediction, aiming to improve risk assessment for financial institutions. Using convolutional neural networks (CNNs) on a Norwegian mortgage dataset with borrower financials and loan characteristics, they achieved a recall of 0.71, outperforming traditional models. However, the black-box nature of CNNs underscored the importance of interpretability, a concern central to our study's use of XAI techniques like LIME and SHAP.

Butaru et al. (2016) [11] investigated credit card repayment risk, seeking to identify delinquency drivers in consumer credit. They applied logistic regression and random forests to a large dataset from a U.S. credit card issuer, including transaction and payment data. The random forest model yielded an AUC of 0.80, outperforming logistic regression, but regulatory demands for transparency favored the interpretable logistic regression, a finding that shapes our model selection strategy for balancing accuracy and explainability in P2P lending. These studies collectively highlight the potential of advanced ML to enhance credit risk prediction across diverse lending contexts, with Ma et al. (2018) [7] providing a particularly relevant framework for P2P lending through their high-performing gradient boosting approach. However, the recurring challenge of model interpretability, especially for complex models, underscores the need for integrating predictive power with transparency, a core objective of our study.

Logistic regression continues to be popular in the financial industry due to its simplicity and transparency. The model's coefficients can be directly interpreted as indicators of a feature's effect on the likelihood of default, making it highly suitable in regulated environments. Decision trees also offer transparency through their rule-based structure but are prone to instability when faced with noisy or imbalanced data. To improve predictive performance, many studies and industry applications turn to random forests, which aggregate predictions from multiple decision trees trained on randomized subsets of the data and features [12]. Although random forests generally outperform simpler models, they are less transparent, making them difficult to interpret a key concern in finance.

This lack of interpretability presents serious challenges in regulated credit environments. In the United States, the Fair Credit Reporting Act (1970) requires lenders to disclose the main reasons behind a loan rejection. In the European Union, the General Data Protection Regulation (GDPR) (2018) provides individuals with a "right to explanation" for algorithmic decisions [13]. In Vietnam, the regulatory landscape for digital lending is still evolving, but the State Bank of Vietnam's Fintech Sandbox Draft Decree (2021) has emphasized that platforms must provide clear disclosure of loan terms and decision criteria. However, there are no standardized guidelines yet for how credit risk scores should be calculated, especially when ML models are involved. This creates growing pressure on lenders to ensure their models are not only accurate but also explainable. To address these challenges, researchers have increasingly focused on interpretable machine learning. Traditional models such as logistic regression and decision trees are naturally interpretable, but may lack the flexibility to capture complex relationships in the data. In contrast, ensemble methods like random forests offer improved performance, but are considered black-box models. To bridge this gap, post hoc interpretability methods have been developed—most notably, LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations). LIME approximates a complex model locally using a linear surrogate [14], while SHAP attributes a model's prediction to individual features using cooperative game theory principles [15]. These methods have been applied to credit risk modeling, including work on Lending Club data [16], [17], [18].

In this study, we focus specifically on the lender's perspective, recognizing their need for both high-performing models and clear justifications for credit decisions. Using loan-level data from Lending Club, we evaluate the performance of logistic regression (with L1 and L2 regularization) and tree-based models (decision trees and random forests) under two different preprocessing strategies: (i) weight of evidence (WoE) and (ii) one-hot encoding with min-max scaling. Our results identify two models of interest: a logistic regression model using WoE, which is inherently interpretable, and a random forest model trained on one-hot encoded, scaled data, which achieves high accuracy but requires additional explanation tools. Therefore, we analyze the logistic model using standard coefficient interpretation, and apply LIME and SHAP exclusively to the random forest model to uncover the drivers behind its predictions.

By balancing predictive power with interpretability tailored for lenders, this work contributes to the responsible deployment of machine learning in P2P credit scoring, helping loan providers meet regulatory expectations while making informed, transparent lending decisions.

3. Methodology

3.1. Reviewing data

For this study, we utilized historical loan data from Lending Club, a leading U.S. peer-to-peer (P2P) lending platform, covering loans issued in 2018. The dataset is publicly available through Kaggle, specifically the "Lending Club Loan Data" dataset, which includes loans from 2007 to 2018 [23], licensed under CC0 1.0 Universal (Public Domain Dedication). This dataset contains hundreds of features per loan, including key financial attributes such as loan amount, interest rate, monthly installment, and borrower-related variables like homeownership type, annual income, monthly FICO score, debt-to-income ratio, and the number of open credit lines. The data represents loans actually funded through the platform, not loan applications, ensuring that the analysis reflects real lending outcomes.

To provide clarity on the dataset's structure, Table 1 presents an example of the training dataset, showcasing a subset of five loan records with selected features and their corresponding labels. This example illustrates the types of variables used and the binary classification labels derived for modeling.

Table 1. Example of Training Dataset from Lending Club 2018 Data

loan_amnt	**annual_inc**	**fico_range_low**	**dti**	**home_ownership**	**loan_status**
10000	60000	700	15.2	RENT	Fully Paid
15000	45000	665	22.5	MORTGAGE	Charged Off
20000	80000	720	18.7	OWN	Fully Paid
8000	35000	680	25.3	RENT	Default
12000	55000	695	20.1	MORTGAGE	Fully Paid

Note: loan_amnt (loan amount in USD), annual_inc (annual income in USD), fico_range_low (lower bound of FICO score), dti (debt-to-income ratio in %), home_ownership (borrower's homeownership status), loan_status (loan outcome).

Loan status serves as the outcome variable and reflects the borrower's repayment behavior. A loan is marked as "Current" if it is being repaid on time, "Late" if payment is between 16 and 120 days overdue, and "Default" if the delay exceeds 121 days. If Lending Club determines that a loan will not be repaid, it is labeled as "Charged-Off." To streamline the classification task, we limited our analysis to loans that were either Fully Paid, Default, or Charged-Off. We categorized Fully Paid loans as creditworthy, and those labeled as Default or Charged-Off as non-creditworthy. After filtering, the dataset comprised 8,323 non-creditworthy records and 47,384 creditworthy ones.

Following the definition of the classification labels, we examined the features available in the dataset. These features fall into three broad categories: borrower characteristics (such as FICO score, employment status, and annual income), platform-driven decisions (such as loan grade and interest rate), and loan performance outcomes (such as total payment). Because our objective is to develop a model that can be applied in real-world settings, we prioritized features that would be available to an investor at the time of loan issuance. This approach ensures that the model's predictions are not only accurate but also practical and actionable.

In doing so, we addressed two major concerns: data leakage and the use of platform-derived variables. Data leakage arises when a model incorporates information that would not be accessible at the time a prediction is made, especially when such information is strongly correlated with the target variable. For example, the total payment feature is highly predictive of loan outcome-loans that default or are paid off early typically have lower total payments. While including this variable may boost model performance during training, it undermines the model's applicability in real-time investment decisions, where such information is unavailable in advance.

Another concern involves variables that are not direct borrower attributes but instead are generated by Lending Club's internal risk models. The loan grade variable is a clear example, and features such as interest rate and installment amount are closely tied to this grade. Since these variables reflect Lending Club's proprietary assessment mechanisms rather than fundamental borrower characteristics, we excluded them from our analysis to ensure the model remains independent of platform-specific decisions and can generalize to other lending contexts.

3.2. Prediction models

The fundamental objective of credit scoring is to assess the creditworthiness of individual applicants, which is essentially a binary classification problem. A creditworthy applicant is expected to fulfill their financial obligations, whereas a non-creditworthy applicant is likely to default. Accordingly, we frame the consumer credit risk prediction task as estimating the probability of default for a borrower, based on a set of observed features.

Let $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ denote the feature vector for borrower i, which captures information such as credit history, income, debt-to-income ratio, and prior delinquencies. The target variable $y_i \in \{0, 1\}$ represents whether the borrower defaulted, where $y_i = 1$ indicates default and $y_i = 0$ indicates no default. The modeling tasks is then to estimate:

$$\hat{y}_i = \Pr(y_i = 1 \mid \mathbf{x}_i). \tag{1}$$

We apply two categories of machine learning models to this task: linear models-specifically L1- and L2-regularized logistic regression and tree-based models including decision trees and random forests.

3.2.1. Logistic Regression

Logistic regression is a statistical classification method that models the probability of a binary outcome as a function of a linear combination of input features. It was formally introduced in the context of binary response modeling by Cox (1958). The method models the log-odds of the probability of default as follows:

$$\log\left(\frac{1 - \Pr(y_i = 1)}{\Pr(y_i = 1)}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}. \tag{2}$$

The parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ are estimated by maximizing the likelihood function. To improve generalization and prevent overfitting, regularization is commonly applied.

L1-regularized logistic regression, also known as Lasso logistic regression, introduces a penalty term proportional to the absolute value of the coefficients:

$$\mathcal{L}_L = -\sum_{i=1}^{n} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] + \lambda \sum_{j=1}^{k} |\beta_j|. \tag{3}$$

This regularization induces sparsity, effectively performing feature selection by shrinking some coefficients to zero.

In contrast, L2-regularized logistic regression, or Ridge logistic regression, penalizes the squared magnitudes of the coefficients:

$$\mathcal{L}_L = -\sum_{i=1}^{n} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] + \lambda \sum_{j=1}^{k} \beta_j^2. \tag{4}$$

This penalty shrinks coefficients toward zero without eliminating them, which can be beneficial in the presence of multicollinearity.

3.2.2. Decision Tree

A decision tree is a non-parametric, supervised learning algorithm that predicts outcomes by recursively partitioning the input space based on feature thresholds. The Classification and Regression Tree (CART) algorithm, introduced by [12], constructs the tree by selecting feature-value splits that minimize an impurity criterion, typically Gini impurity:

$$\text{Gini}(t) = 1 - \sum_{c=1}^{C} p(c \mid t)^2 \tag{5}$$

where $p(c \mid t)$ is the proportion of class c in node t. This recursive partitioning continues until a stopping criterion is met (e.g., maximum depth or minimum node size), resulting in a tree structure where each leaf node represents a final prediction.

3.2.3. Random Forests

Random forest is an ensemble learning technique that aggregates predictions from multiple decision trees to improve classification performance and robustness [20]. Each decision tree in the ensemble is trained on a different bootstrap sample of the training data, and feature selection at each node is randomized. This combination of bootstrap aggregation (bagging) and random feature selection helps ensure low correlation among trees, which in turn reduces model variance.

To classify a new observation, each decision tree provides a prediction, and the random forest outputs the majority vote across all trees. While individual trees are relatively interpretable, the ensemble nature of random forests makes the model difficult to interpret as a whole. As such, random forests are often considered black-box models, despite their strong predictive performance and robustness to overfitting.

3.3. Preprocessing data

To prepare the dataset for modeling, we experimented with two distinct preprocessing strategies: weight of evidence (WoE) encoding and min-max scaling. Each strategy was applied independently, as WoE produces features that are already normalized, thereby eliminating the need for further scaling, while min-max scaling operates directly on the original continuous variables and does not require binning or WoE transformation.

Weight of evidence encoding is a widely used technique in credit risk modeling [21], particularly suitable for datasets containing special values, missing data, or outliers. We began by discretizing continuous features into bins, which allows special values to be grouped into separate categories and helps mitigate the influence of extreme values. WoE assigns each bin a numerical value based on the distribution of creditworthy and non-creditworthy borrowers. For bin i, the WoE value is defined as:

$$\text{WoE}_i = \ln\left(\frac{\frac{N_{\text{good},i}}{N_{\text{good}}}}{\frac{N_{\text{bad},i}}{N_{\text{bad}}}}\right) \tag{6}$$

where $N_{good,i}$ and $N_{bad,i}$ are the numbers of creditworthy and non-creditworthy observations in bin i, and N_{good} and N_{bad} are the total numbers of creditworthy and non-creditworthy observations in the dataset.

One advantage of WoE is that it standardizes feature values on a log-odds scale, which is especially useful for linear models like logistic regression. It also handles missing values and outliers effectively by assigning them to dedicated bins. However, because WoE relies on binning, it may introduce some loss of granularity and is less interpretable outside of credit modeling contexts.

As an alternative, we apply min-max scaling to the original continuous features without discretization. This method normalizes each feature x to a value x' in the range $[0, 1]$, according to the formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{7}$$

This transformation ensures that all features are on a comparable scale, which can help improve numerical stability and convergence in gradient-based models. While tree-based models such as decision trees and random forests are typically invariant to monotonic transformations, scaling can still be beneficial in controlling feature dominance and improving performance, especially when the features span very different numeric ranges.

Because we use both linear and non-linear models, our preprocessing strategy is designed to test which approach better supports each model type. In particular, we expect WoE to be more effective for linear models, where encoding categorical and binned variables in terms of log-odds enhances model interpretability and alignment with assumptions. For tree-based models, which naturally handle non-linear splits, we test whether simple min-max scaling provides sufficient normalization without the need for more domain-specific transformations like WoE.

3.4. Interpretability methods

In high-stakes domains such as credit scoring, interpretability is a key requirement for model adoption and trustworthiness. While complex models like random forests often deliver superior predictive performance, they are frequently criticized for their black-box nature. In this section, we explore three interpretability methods: coefficient analysis using Weight of Evidence (WoE), Local Interpretable Model-Agnostic Explanations (LIME), and SHapley Additive exPlanations (SHAP). Each method provides a different lens through which to understand model behavior and explain individual predictions.

3.4.1. Coefficient for WoE

When logistic regression is trained using features encoded with Weight of Evidence (WoE), model interpretability is naturally preserved. Since WoE transforms each variable into a continuous value representing the log-odds of creditworthiness, the estimated coefficients in the logistic regression model can be interpreted directly as the marginal effect of each feature on the log-odds of default. A positive coefficient indicates that an increase in the WoE-encoded feature increases the likelihood of default (i.e., reduces creditworthiness), while a negative coefficient implies the opposite.

This approach is particularly appealing for credit risk applications because it aligns with long-standing industry practices and produces additive, transparent risk contributions across features. Moreover, when features are pre-binned and monotonic WoE encodings are applied, the signs and magnitudes of the coefficients tend to be more stable and easier to interpret.

3.4.2. Local Interpretable Model-Agnostic Explanations (LIME)

LIME is a post hoc model-agnostic technique that provides local interpretability by approximating the decision boundary of any black-box model around a specific data point with a linear model [14]. This linear approximation is trained by sampling perturbed versions of the original input and fitting a locally weighted linear regression model. The weights are assigned based on the proximity of the perturbed samples to the original instance, typically measured using a kernel function.

The coefficients of the resulting local surrogate model serve as feature importance scores, highlighting how each input feature contributes to the model's prediction for that particular instance. The strength of LIME lies in its flexibility—it can be applied to any model and any type of data. However, its reliance on sampling introduces randomness, and explanations can vary slightly across different runs. Furthermore, the local linear approximation may not faithfully represent highly non-linear decision boundaries.

3.4.3. SHapley Additive exPlanations (SHAP)

SHAP is an explainable AI method based on cooperative game theory. It attributes a model's prediction for a specific data point to the contributions of each feature using the concept of Shapley values, originally developed to fairly distribute payouts among players in a coalition [22]. SHAP satisfies several desirable properties for interpretability, including local accuracy (the sum of the attributions equals the model output), missingness (features not present receive zero contribution), and consistency (if a feature's contribution increases in a model, its Shapley value will not decrease).

We employ the Kernel SHAP implementation introduced by [15], which approximates Shapley values using a weighted least squares regression. To explain a data point x_i , Kernel SHAP constructs a dataset of feature subsets sampled from x_i, with the remaining features replaced by background values from the training data. Each subset receives a weight based on the size of the subset, with smaller subsets (closer to the marginal contribution of a single feature) weighted more heavily. The regression solution yields the estimated Shapley values.

Despite its theoretical appeal, Kernel SHAP suffers from poor scalability: computing exact Shapley values requires evaluating all 2^k feature subsets, which becomes computationally infeasible for high-dimensional datasets. Approximate methods and sampling strategies are used in practice, but the method remains relatively expensive compared to alternatives like LIME.

3.4.4. Comparative Discussion

Both LIME and SHAP provide complementary perspectives for model interpretability. LIME excels in computational efficiency and model-agnostic flexibility, offering quick local approximations that are especially useful when working with large feature sets or real-time explanations. However, LIME may lack fidelity in capturing true feature interactions and does not guarantee consistency or local accuracy.

SHAP, on the other hand, offers theoretically grounded explanations that reflect both individual feature contributions and their interactions. It provides robust, additive attributions that sum to the model's prediction, but it is computationally more intensive and may be less suitable for real-time applications.

In practice, the choice between LIME and SHAP should be guided by the specific use case. For instance, when interpretability is paramount for auditability or regulatory compliance, SHAP may be more appropriate despite its computational cost. In contrast, when speed is essential and the model is used in a dynamic setting with frequent queries, LIME may offer a more practical solution. By applying both techniques judiciously, practitioners can better understand and validate complex machine learning models, particularly in sensitive domains like credit risk assessment.

4. Results

4.1. Model performance

In this section, we evaluate the performance of various machine learning models under two different preprocessing strategies: Weight of Evidence (WoE) encoding and one-hot encoding with min-max scaling. The models under consideration include L1- and L2-regularized logistic regression, decision tree, and random forest classifiers.

We employ 5-fold cross-validation to assess the generalization performance of each model. In this setup, the dataset is randomly partitioned into five equal subsets. In each fold, one subset is held out as the test set, while the remaining four subsets are used for training. Thus, in each iteration, the training set consists of 80% of the data and the test set comprises the remaining 20%. Performance metrics are computed on the test set and then averaged across all five folds to ensure robustness.

We utilized a dataset of 55,707 records (8,323 non-creditworthy and 47,384 creditworthy) and applied a 5-fold cross-validation approach, which inherently combines training and testing phases without a separate validation set. Specifically, in each fold, the dataset was split into a training set of approximately 44,566 records (80% of the data) and a test set of 11,141 records (20%), totaling 38,994 records for training and 16,713 for testing across all folds, as derived from the provided split (38,994 training + 16,713 test = 55,707). We did not use a distinct validation set because the 5-fold cross-validation process effectively validates the model by rotating the test set across folds, optimizing performance metrics like recall for the non-creditworthy class, as detailed in our hyperparameter tuning with GridSearchCV. This approach ensures that the model is evaluated on multiple subsets, providing a robust estimate of generalization performance without requiring a separate validation set.

To address the class imbalance in our dataset, we implemented a combination of Synthetic Minority Oversampling Technique (SMOTE), class weighting, and a recall-focused scoring metric to enhance the performance of our machine learning models, particularly for the non-creditworthy class. SMOTE was applied to the training data within each fold of our 5-fold cross-validation to generate synthetic non-creditworthy samples, balancing the class distribution while preserving the original test set for realistic evaluation. Additionally, we incorporated class weighting (e.g., balanced or a 1:5 ratio favoring non-creditworthy) in our logistic regression and random forest models to penalize misclassifications of the minority class more heavily. By using recall as the primary scoring metric in hyperparameter tuning via GridSearchCV, we prioritized the identification of non-creditworthy loans, minimizing false negatives critical to credit risk assessment. These strategies collectively mitigated the risk of overfitting and improved the models' ability to accurately classify non-creditworthy records, as evidenced by enhanced recall scores in our results.

To evaluate classification performance, we use accuracy, area under the receiver operating characteristic curve (AUC), and recall, with particular emphasis on recall due to the cost-sensitive nature of credit risk. Since our goal is to minimize the number of high-risk borrowers who are incorrectly classified as low-risk (i.e., false negatives), recall-defined as the proportion of true defaulters correctly identified is of primary importance. The AUC reflects the model's ability to distinguish between creditworthy and non-creditworthy applicants, while accuracy captures the overall proportion of correctly classified samples. Given that our dataset is relatively balanced, accuracy remains a meaningful metric alongside AUC and recall. For all models, the predicted probability of default is converted to a binary classification using a threshold of 0.5.

Tables 1 and 2 summarize the results of the 5-fold cross-validation for the two preprocessing pipelines. Among the models trained on WoE-encoded data, logistic regression models perform best in terms of both AUC and recall. Specifically, L2-penalized logistic regression achieves an accuracy of 0.66, an AUC of 0.71, and a recall of 0.67. Although the random forest achieves the highest accuracy (0.71), its recall is considerably lower (0.50), which limits its effectiveness for detecting defaulters. This supports the view that WoE encoding, when combined with interpretable linear models, offers strong predictive performance while maintaining transparency.

Table 2. 5-fold cross-validation performance of ML models using WoE encoding

Model	**Accuracy**	**AUC**	**Recall**
L1 Logistic Regression	0.65	0.70	0.65
L2 Logistic Regression	0.66	0.71	0.67
Decision Tree	0.61	0.67	0.61
Random Forest	0.71	0.69	0.50

In contrast, when models are trained on min-max scaled data with one-hot encoding, the random forest outperforms the other models across all metrics. It achieves an accuracy of 0.78, an AUC of 0.69, and a recall of 0.38. However, its recall remains relatively low, suggesting that even with higher accuracy, it may not be optimal for identifying risky applicants. Meanwhile, L1-penalized logistic regression achieves a competitive recall of 0.66, though its accuracy (0.59) and AUC (0.67) are lower than those of the random forest.

Table 3. 5-fold cross-validation performance of ML models using min-max scaling

Model	**Accuracy**	**AUC**	**Recall**
L1 Logistic Regression	0.59	0.67	0.66
L2 Logistic Regression	0.71	0.65	0.45
Decision Tree	0.57	0.66	0.67
Random Forest	0.78	0.69	0.38

From a practical standpoint, logistic regression trained with WoE features presents an attractive option for credit scoring applications. It offers interpretable coefficients that align with industry standards and regulatory requirements. However, implementing WoE encoding requires feature binning, monotonicity constraints, and careful calibration, which increases preprocessing complexity.

On the other hand, random forests, though superior in raw predictive power when trained on scaled features, suffer from limited interpretability. The ensemble nature of the model, which aggregates hundreds of decision paths, makes it difficult to explain individual predictions—an issue particularly relevant in regulated domains like consumer finance.

In the context of peer-to-peer lending, both types of misclassification-false negatives (defaulters misclassified as creditworthy) and false positives (creditworthy applicants denied loans)-have important business implications. Misclassifying defaulters results in financial losses, while rejecting potentially reliable borrowers leads to lost revenue. Given the scale of the lending industry, even small gains in recall or precision can translate into substantial economic impact. Despite this, regulatory constraints and the need for explainability often prevent lenders from adopting more complex but opaque models. This tradeoff motivates our deeper investigation into model interpretability in the next section.

4.2. Explaining model results

In this section, we analyze the interpretability of the models to support their practical adoption in loan decision-making systems. Machine learning models are increasingly used by lending institutions to assess the creditworthiness of borrowers. However, regulatory frameworks such as the Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) in the United States require that lenders provide specific reasons for loan denial. This has created a strong demand for interpretable models and reliable post hoc explanation techniques.

Interpretability is not only essential for regulatory compliance but also for building trust with applicants and improving internal risk assessment procedures. In this context, lenders seek to identify the key factors driving creditworthiness and to generate understandable explanations for individual decisions made by the model. Transparent models also allow companies to identify representative historical borrowers whose profiles are similar to new applicants, thus supporting a case-based reasoning approach.

4.2.1. WoE Coefficients for Logistic Regression

One effective approach to achieving model interpretability is through the use of logistic regression trained on Weight of Evidence (WoE)-encoded features. In this setting, each feature represents the log-odds of being creditworthy within a given bin, and the coefficients of the logistic regression model quantify the contribution of each feature to the log-odds of default.

The intercept of the trained L2-penalized logistic regression model is –1.713, and the coefficients for each WoE-transformed feature are presented in Table 3. A positive coefficient implies that higher values of the corresponding WoE feature (i.e., riskier bins) increase the likelihood of default, while negative coefficients imply the opposite. Because WoE encoding aligns feature values with the probability of default, the resulting coefficients can be directly interpreted as directional indicators of credit risk.

As shown in the table, the loan amount (loan_amnt_woe) is the most influential variable in determining creditworthiness, with a coefficient of 1.247. This suggests that larger loan amounts are associated with a higher probability of default. Other important predictors include recent credit inquiries (inq_last_6mths_woe, 0.971), annual income (annual_inc_woe, 0.908), and FICO score range (fico_range_low_woe, 0.725). These features align well with common industry understanding of credit risk factors.

The simplicity and transparency of this model make it particularly suitable for lending environments where interpretability and regulatory reporting are as important as predictive accuracy. In contrast to complex models such as random forests, logistic regression with WoE encoding provides clear justifications for both individual and population-level decisions.

Table 4. Coefficients of L2-Penalized Logistic Regression (WoE Model)

Feature	**Coefficient**
loan_amnt_woe	1.247
inq_last_6mths_woe	0.971
annual_inc_woe	0.908
fico_range_low_woe	0.725
verification_status_woe	0.677
home_ownership_woe	0.637
num_il_tl_woe	0.487
revol_util_woe	0.473
mort_acc_woe	0.423
mths_since_rcnt_il_woe	0.315
mo_sin_old_rev_tl_op_woe	0.255

4.2.2. Local Interpretable Model-Agnostic Explanations (LIME) for Random Forest

LIME (Local Interpretable Model-Agnostic Explanations) is a post hoc interpretability technique that approximates complex models by training a local, interpretable surrogate model around a prediction of interest [14]. In our case, LIME is used to interpret predictions made by the random forest model trained on one-hot encoded features with min-max scaling.

For each individual data point, LIME perturbs the instance to generate a synthetic neighborhood and fits a locally weighted linear regression to approximate the black-box model's behavior in that region. The coefficients of this surrogate model represent the impact of each feature on the prediction and can be interpreted as the change in the predicted probability resulting from a unit change in the feature value, holding other features constant.

Figure 1 shows a LIME explanation for a single borrower. The model assigns a 91% predicted probability of default, indicating high credit risk. The bar charts in the figure break down this prediction by feature contribution. Features shown in red increase the probability of default (Class 1), while those in green support a prediction of no default (Class 0).

In this example, the most influential features increasing the likelihood of default are verification_status_Verified = 0.00, indicating unverified income, which contributes approximately +0.05 to the prediction; application_type_Joint App = 0.00, adding +0.04; and both home_ownership_MORTGAGE = 0.00 and home_ownership_RENT = 0.00, each contributing +0.03. Additionally, fico_range_low = 0.35 and mort_acc = 0.25 fall into intervals that the model associates with higher risk, each contributing approximately +0.02 to the probability of default.

These features push the prediction strongly toward the "Default" class. On the other hand, features such as loan_amnt = 0.49, num_actv_bc_tl = 0.17, and inq_last_12m = 0.01 slightly counterbalance the default risk, with negative contributions shown in green. However, the mitigating effect of these features is not sufficient to override the cumulative positive influence of the others, resulting in a high default prediction.

LIME's local explanation highlights which features the model relied on for this specific decision and allows decision-makers to trace the rationale behind a prediction. While LIME does not guarantee global consistency or faithfulness to the underlying model, it is computationally efficient and flexible across model types and data structures. In production environments where transparency is critical for instance, when rejecting a loan application LIME can help generate individualized explanations in real time. These explanations satisfy regulatory requirements and help build trust with customers by providing a human-understandable rationale for each prediction.

4.2.3. SHapley Additive exPlanations (SHAP) for Random Forest

To interpret the predictions of the random forest model trained on one-hot encoded features with min-max scaling, we apply SHapley Additive exPlanations (SHAP). SHAP is an explainability technique rooted in cooperative game theory that decomposes a model's prediction into the contributions of each input feature [22]. For this task, we use the Tree SHAP algorithm, which is optimized for ensemble models such as random forests [15].

SHAP provides global explanations by measuring the average magnitude of each feature's contribution across all instances in the dataset. These values represent how much each feature, on average, influences the model's prediction toward either class: "Default" (Class 1) or "No Default" (Class 0). Figure 2 presents the SHAP summary plot based on mean absolute SHAP values, with red bars representing contributions toward predicting default, and blue bars representing contributions toward predicting no default.

The results highlight loan_amnt, fico_range_low, and verification_status_Verified as the most influential predictors in the model's decision-making. This aligns well with financial intuition: larger loan amounts and lower FICO scores are commonly associated with higher credit risk, while income verification status reflects the reliability of the reported income, which can strongly influence repayment behavior. Other important features include mort_acc (number of mortgage accounts), home_ownership_MORTGAGE, and inq_last_6mths (number of credit inquiries in the last six months), all of which are standard indicators in credit risk evaluation.

Unlike the WoE-based logistic regression model described in Section 4.2.1, which offers interpretability through linear coefficients aligned with log-odds,

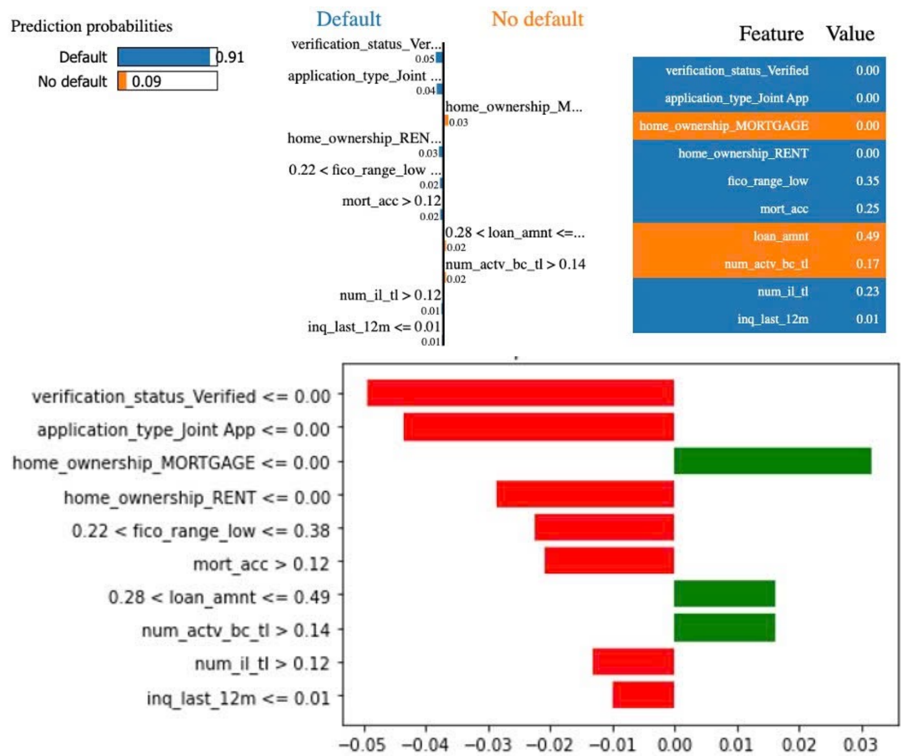


Figure 1. LIME explanation for one borrower classified as "Default" by the random forest model (predicted probability = 0.91). Red bars indicate features contributing to the "Default" prediction, while green bars indicate features supporting "No Default."

the random forest model requires post hoc interpretability tools like SHAP due to its non-linear and ensemble nature. While the random forest model achieves higher accuracy, its lack of transparency can be a barrier to deployment in regulated financial contexts. SHAP mitigates this by revealing how each feature contributes to predictions at both the global and individual levels.

Despite its advantages, SHAP also has limitations. While Tree SHAP is computationally efficient compared to the original Shapley value formulation, it can still be resource-intensive for very large models or datasets. Moreover, SHAP's explanations, while grounded in strong theory, may still be difficult to communicate to non-technical stakeholders, particularly when many features are involved.

Nonetheless, SHAP serves as a powerful bridge between predictive performance and interpretability. It allows stakeholders to audit the behavior of complex models like random forests and to gain trust in model predictions by understanding the most influential drivers of credit decisions.

We compare the interpretability results of our Random Forest model (min-max scaling) in Figures 1 and 2 with similar existing works to assess their quality. Our SHAP analysis (Figure 2) identifies loan amount, FICO score, and verification status as top predictors, aligning closely with [16] and [17], who also highlight FICO score and loan amount using SHAP on Lending Club data, and [18], who emphasize credit history on a Colombian P2P dataset. Similarly, our LIME explanation (Figure 1) provides detailed contributions (e.g., unverified income: +0.05, FICO score: +0.02) for a borrower predicted as "Default" (probability 0.91), offering more granularity than LIME results in Hadji Misheva et al. and Ariza-Garzón et al., enhancing individual decision explanations. We argue that these results are "good enough" for P2P lending credit risk assessment, as they provide actionable local (LIME) and global (SHAP) insights, meeting regulatory requirements for transparency and aligning with traditional risk factors like FICO score and loan amount. Despite the Random Forest's lower recall (0.38) compared to L2 Logistic Regression (0.67), the interpretability results are sufficiently detailed and relevant, supporting stakeholder trust and responsible lending practices though future work could improve predictive performance for non-creditworthy detection.

5. Conclusion

This study investigated the trade-off between predictive performance and model interpretability in the context of credit risk assessment for peer-to-peer (P2P) lending. Using historical data from Lending Club, we compared the effectiveness of logistic regression, decision tree, and random forest models under two preprocessing pipelines: Weight of Evidence (WoE) encoding and one-hot encoding with min-max scaling. Our results show that while random forest models trained on scaled one-hot features achieve the highest accuracy, logistic regression models using WoE encoding strike a more desirable balance between predictive power and interoperability.

From a practical perspective, the choice of model should reflect the priorities of the lending platform. When the primary objective is regulatory compliance and transparency, as is often the case in highly regulated environments, logistic regression with WoE encoding offers a clear advantage. This approach enables lenders to trace the contribution of each feature to the predicted probability of default, making the model easier to audit and justify. On the other hand, when prediction accuracy is paramount, especially in settings where regulatory

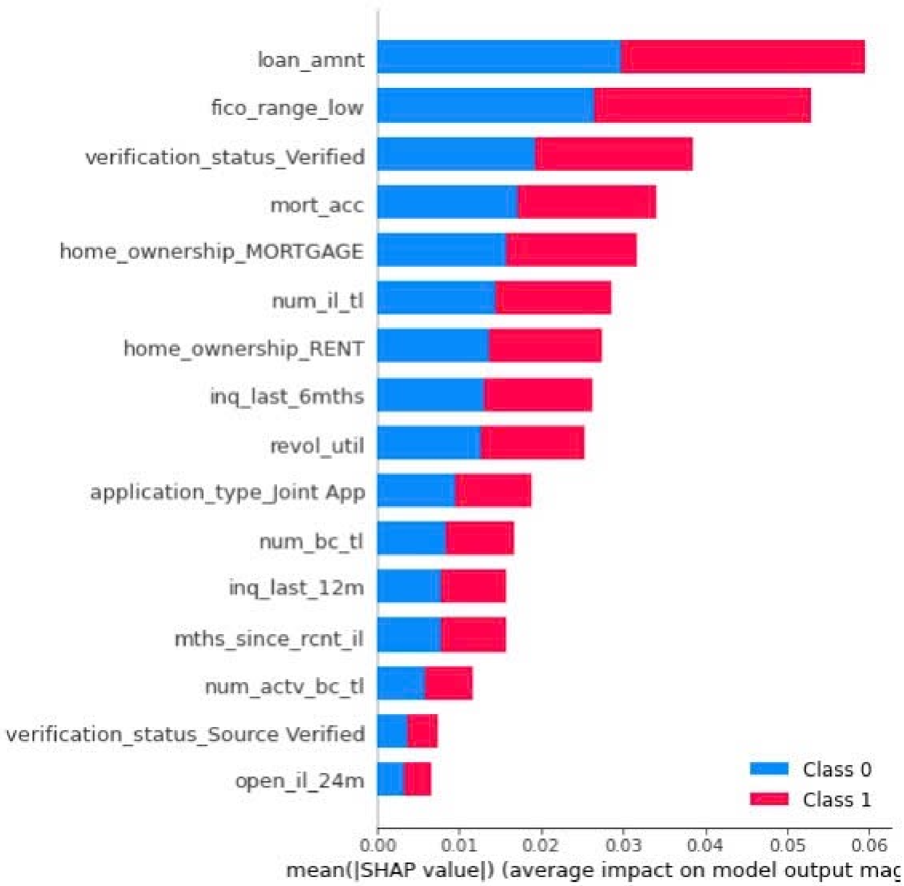


Figure 2. SHAP summary plot showing the global feature importance for the random forest model trained using one-hot encoding and min-max scaling. Red bars indicate contributions toward predicting "Default" (Class 1), while blue bars indicate contributions toward predicting "No Default" (Class 0).

constraints are less strict, random forests can provide superior performance.

To address the opacity of ensemble models, we employed two post hoc explainability techniques-LIME and SHAP-to interpret the predictions of the random forest. These tools revealed that key drivers of credit decisions include loan amount, FICO score, verification status, and the number of mortgage accounts factors that are consistent with traditional credit risk evaluation. The application of LIME enabled localized, instance-specific explanations, which are useful for generating individualized decision justifications. SHAP, in contrast, offered a global perspective on feature importance, contributing to broader model understanding and policy refinement.

Ultimately, our findings highlight that interpretability and accuracy need not be mutually exclusive. By selecting modeling techniques and explanation methods aligned with institutional goals and regulatory expectations, lenders can build trustworthy, effective credit scoring systems. As the P2P lending industry continues to evolve, integrating transparent machine learning models will be essential for promoting responsible lending practices, enhancing borrower trust, and maintaining regulatory compliance.

References

[1] **Lee, I., Shin, Y. J.** *Fintech: Ecosystem, business models, investment decisions, and challenges.* (Business Horizons, 2018), 61(1), 35–46.

[2] **Liu, H., Qiao, H., Wang, S., Li, Y.** *Platform competition in peer-to-peer lending considering risk control ability.* European Journal of Operational Research, 2018, 274(1), 280–290.

[3] **Turner, A.** *After the crisis, the banks are safer but the debt is a danger. Financial Times.* (2025, March 1). https://www.ft.com/content/9f481d3c-b4de-11e8-a1d8-15c2dd1280ff

[4] **Huang, R. H.** *Online P2P lending and regulatory responses in China: Opportunities and challenges.* European Business Organization Law Review (2018), 19(1), 78.

[5] **Havrylchyk, O.** *Regulatory framework for the loan-based crowdfunding platforms.* OECD Economics Department Working Papers. (2021)

[6] **Chapman, J. M.** *Factors affecting credit risk in personal lending.* In Commercial Banks and Consumer Installment Credit (1940), 109–139. NBER.

[7] **Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., Niu, X.** *Study on a prediction of P2P network loan default based on the machine learning*

lightGBM and XGBoost algorithms according to different high dimensional data cleaning. Electronic Commerce Research and Applications (2018), 31, 24–39.

[8] **Duan, J.** *Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction.* Journal of the Franklin Institute (2019), 356, 4716–4731.

[9] **Sirignano, J., Sadhwani, A., Giesecke, K.** *Deep learning for mortgage risk* (2016) arXiv preprint arXiv:1607.02470.

[10] **Kvamme, H., Sellereite, N., Aas, K., Sjursen, S.** *Predicting mortgage default using convolutional neural networks.* Expert Systems with Applications (2018), 102, 207–217.

[11] **Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., Siddique, A.** *Risk and risk management in the credit card industry.* (2016) Journal of Banking & Finance, 72, 218–239.

[12] **Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.** Classification and regression trees. Monterey, CA: Wadsworth and Brooks. (1984).

[13] **Goodman, B., & Flaxman, S.** *European Union regulations on algorithmic decision-making and a "right to explanation".* AI Magazine (2017), 38(3), 50–57.

[14] **Ribeiro, M. T., Singh, S., & Guestrin, C.** *"Why should I trust you?": Explaining the predictions of any classifier.* In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016), pp. 1135–1144.

[15] **Lundberg, S. M., & Lee, S.-I.** *A unified approach to interpreting model predictions.* In Advances in Neural Information Processing Systems (2017), 30, 4765–4774.

[16] **Hadji Misheva, B., Hirsa, A., Osterrieder, J., Kulkarni, O., & Fung Lin, S.** *Explainable AI in credit risk management.* Credit Risk Management (2021). https://ssrn.com/abstract=3768437

[17] **Albanesi, S., & Vamossy, D. F.** *Predicting consumer default: A deep learning approach.* National Bureau of Economic Research Working Paper. (2019)

[18] **Ariza-Garzón, M. J., Arroyo, J., Caparrini, A., & Segovia-Vargas, M.-J.** *Explainability of a machine learning granting scoring model in peer-to-peer lending.* IEEE Access (2020), 8, 64873–64890.

[19] **Breiman, L.** *Random Forests* Machine Learning (2021), 45, 5-32. http://dx.doi.org/10.1023/A:1010933404324

[20] **Cox, D. R.** *The regression analysis of binary sequences.* Journal of the Royal Statistical Society: Series B (Methodological) (1958), 20(2), 215–242.

[21] **Siddiqi, N.** *Credit risk scorecards: Developing and implementing intelligent credit scoring (Vol. 3)* John Wiley & Sons. (2012)

[22] **Shapley, L.** *A Value for n-Person Games. In: Kuhn, H. and Tucker, A., Eds., Contributions to the Theory of Games II* Princeton University Press, Princeton (1958), 307-317. https://doi.org/10.1515/9781400881970-018

[23] Link dataset (https://www.kaggle.com/datasets/wordsforthewise/lending-club)

Dinh Thuy Tien
Thang Long University
Hanoi
Vietnam
dinhthuytien@thanglong.edu.vn

Nguyen Thi Tra My
Thang Long University
Hanoi
Vietnam
myntt@thanglong.edu.vn