TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 349, Number 4, April 1997, Pages 1437–1446
S 0002-9947(97)01514-6

p-ADIC POWER SERIES WHICH COMMUTE UNDER COMPOSITION

HUA-CHIEH LI

Abstract. When two noninvertible series commute to each other, they have same set of roots of iterates. Most of the results of this paper will be concerned with the problem of which series commute with a given noninvertible series. Our main theorem is a generalization of Lubin's result about isogenies of formal groups.

1. Introduction

Let $f(x) = \sum_{n=1}^{\infty} a_n x^n$ and $g(x) = \sum_{n=1}^{\infty} b_n x^n$ be formal power series without constant terms. We use the symbol $f \circ g$ to denote the formal power series $\sum_{n=1}^{\infty} c_n x^n$ obtained by substituting g for x in f and rearranging according to powers of x. We say g commutes with f in the sense that

$$f \circ g = g \circ f\,.$$

In complex analytic dynamics, when $0 < |f'(0)| = |a_1| < 1$ and f is convergent for $|z| < \rho$, where $\rho > 0$, the commuting family of this analytic function $f(z)$ are completely mastered by the Schröder's functional equation: $L\big(f(z)\big) = f'(0)L(z)$, where $L(z)$ is a unique function with $L'(0) = 1$ which satisfies this identity and converges in some neighbourhood of 0. Not only can one express all iterates $f^{\circ m}(z)$ as $L^{\circ -1}(a_1^m L(z))$, but with arbitrary b_1 the expression $L^{\circ -1}(b_1 L(z))$ gives us the unique power series $g(z)$ with $g'(0) = b_1$ which commutes with $f(z)$.

Let K be an algebraic extension of $\mathbf{Q}_p$ and let $\mathcal{O}$ be its integer ring with maximal ideal $\mathcal{M}$. If $\overline{K}$ is an algebraic closure of K, we denote by $\overline{\mathcal{O}}$ and $\overline{\mathcal{M}}$ the integral closure of $\mathcal{O}$ in $\overline{K}$ and the maximal ideal of $\overline{\mathcal{O}}$, respectively. The set of all power series over $\mathcal{O}$ without constant term is a monoid (noncommutative, associative, with unit) under composition. A series $u(x) \in \mathcal{O}[[x]]$ without constant term is called invertible if there exists a series $w(x) \in \mathcal{O}[[x]]$ such that $u \circ w(x) = x$. A necessary and sufficient condition for $u(x)$ to be invertible is that $u'(0) \in \mathcal{O}^*$.

If $f(x) \in \mathcal{O}[[x]]$ without constant term and $0 \neq f'(0) \in \mathcal{M}$, then we call $f(x)$ a noninvertible stable series. In this case, Lubin [2] shows a similar result as in complex dynamics: for any noninvertible series we can find a unique $L_f(x)$ with $L_f'(0) = 1$ such that $L_f\big(f(x)\big) = f'(0)L_f(x)$ and $L_f(x)$ converges in $\overline{\mathcal{M}}$. Therefore, for any $b_1 \in K$, $g(x) = L_f^{\circ -1}(b_1 L_f(x))$ is the unique power series in $K[[x]]$ with $g'(0) = b_1$ which commutes with $f(x)$. We define $\Lambda(f) = \{\alpha \in \overline{\mathcal{M}} |\, f^{\circ n}(\alpha) = 0,\ \text{for some } n\,\}$, the set of all roots of iterates of $f(x)$. The roots of iterates are of

Received by the editors June 24, 1994 and, in revised form, March 22, 1995.
1991 *Mathematics Subject Classification.* Primary 11S99; Secondary 11S31, 14L05.

serious interest. Lubin [2] also shows that if $g(x) \in \mathcal{O}[[x]]$ is a noninvertible series which commutes with $f(x)$, then $\Lambda(f) = \Lambda(g)$. One of the principal topics of this note is the question of which series in $\mathcal{O}[[x]]$ commutes with a given one.

In the commuting family of $f(x)$, there may exist a series $g(x) \in \mathcal{O}[[x]]$ such that $g = h \circ f$ but $h(x) \notin \mathcal{O}[[x]]$. One may ask under what conditions this can not happen. In a formal group case, Lubin [4] shows that if f, $g \in \mathrm{End}_{\mathcal{O}}(F)$ for some formal group F and $f|\,g$, then there exists $h \in \mathcal{O}[[x]]$ such that $g = h \circ f$. Our main theorem is a generalization of Lubin's result (without any formal group in the back ground) which says that the necessary and sufficient condition for a series $g(x)$ which commutes with $f(x)$ to satisfy $g(x) \in \mathcal{O}[[f(x)]]$ is that every root of $f(x)$ is also a root of $g(x)$.

The work presented here is part of the author's 1994 Brown Ph.D. thesis. Without Professor Rosen's continued help and encouragement, none of this work would have been possible. Professor Lubin was the one who introduced the author to the field of p-adic Dynamical Systems. His guidance in this research was indispensable.

2. Newton Copolygons and Valuation Functions

Recall that K is a field which is complete with respect to a valuation v. We normalize the valuation v such that $v(\pi) = 1$, where π is a generator of $\mathcal{M}$. There is a unique extension of v to $\overline{K}$, and this will likewise be denoted v.

Throughout this paper we deal exclusively with power series $f(x) \in \mathcal{O}[[x]]$ such that $f(0) = 0$ and $f'(0)$ is neither 0 nor any root of 1. We also suppose that not all coefficients of $f(x)$ are in $\mathcal{M}$. We denote this set by $\mathcal{S}_0(\mathcal{O})$. The lowest degree in which a unit coefficient appears will be called the Weierstrass degree of $f(x)$, denoted $\mathrm{wideg}(f)$. According to the Weierstrass Preparation Theorem there exist a unit power series $U(x) \in \mathcal{O}[[x]]$ and a distinguished polynomial $P(x) \in \mathcal{O}[[x]]$ ($P(x) = x^n + b_{n-1}x^{n-1} + \cdots + b_0$ where $b_i \in \mathcal{M}$) such that $f(x) = P(x)U(x)$ and $\deg(P) = \mathrm{wideg}(f)$. It is easy to see that $\mathrm{wideg}(f \circ g) = \mathrm{wideg}(f)\mathrm{wideg}(g)$.

The *Newton polygon* is a natural tool to study the roots of p-adic power series (see Koblitz [1, pages 89–100]). Another geometric object, which contains the same information as the Newton polygon, is the *Newton copolygon*. Let $f(x) = \sum_{n=1}^{\infty} a_n x^n$. The Newton copolygon of $f(x)$, $\mathcal{N}^*(f)$, is defined to be the intersection in the Cartesian plane of all halfplanes defined by the inequalities $y \le ix + v(a_i)$. A part of the line $y = ix + v(a_i)$ is a segment of $\mathcal{N}^*(f)$ if and only if there exists $\alpha \in \mathbf{R}$ such that $i\alpha + v(a_i) < j\alpha + v(a_j)$ for all $j \ne i$. A point (ξ, η) is a vertex of $\mathcal{N}^*(f)$ if and only if there exist i and j such that $\eta = i\xi + v(a_i) = j\xi + v(a_j)$ and $j'\xi + v(a_{j'}) \ge \eta$ for all $j' \ne i$.

It is easy to see that two power series have the same Newton copolygon if and only if they have the same Newton polygon: indeed, the polygon and copolygon are essentially dual convex bodies. We have the following facts:

Lemma 2.1. *The vertices of $\mathcal{N}(f)$ (Newton polygon of $f(x)$) are in one-to-one correspondence with the segments of $\mathcal{N}^*(f)$; if (P, S^*) is a corresponding pair, the x-coordinate of P is the slope of S^* and the y-coordinate of P is the y-intercept of S^*.*

Proof. $P = \big(i, v(a_i)\big)$ is a vertex of $\mathcal{N}(f)$ if and only if there exists ξ such that for all $j > i$ and $j' < i$, $\Big(v(a_i) - v(a_j)\Big)/(i - j) > -\xi > \Big(v(a_i) - v(a_{j'})\Big)/(i - j')$. Thus

$i\xi + v(a_i) < j\xi + v(a_j)$, for all $j \neq i$. This is equivalent to the assertion that a part of $y = ix + v(a_i)$ is a segment of $\mathcal{N}^*(f)$. □

Lemma 2.2. *The nonvertical segments of $\mathcal{N}(f)$ are in one-to-one correspondence with the vertices of $\mathcal{N}^*(f)$; if (S, P^*) is a corresponding pair, the x-coordinate of P^* is the negative of the slope of S and the y-coordinate of P^* is the y-intercept of S.*

Proof. If $P^* = (\xi, \eta)$ is a vertex of $\mathcal{N}^*(f)$, then choose smallest i and biggest i' such that $\eta = i\xi + v(a_i) = i'\xi + v(a_{i'})$. This condition and $j\xi + v(a_j) \geq \eta$ tell us that the segment S which connects $\big(i, v(a_i)\big)$ and $\big(i', v(a_{i'})\big)$ is a segment of $\mathcal{N}(f)$. It is easy to check that the slope of S is $-\xi$ and the y-intercept of S is $i\xi + v(a_i) = \eta$.

The converse follows, if one reverses the above argument. □

The following is a consequence of the basic property of the Newton polygon (Koblitz [1, IV, Corollary of Theorem 14]) but is of sufficient importance to merit its own statement.

Proposition 2.3. *Let P^* be a vertex of the Newton copolygon of $f(x) \in K[[x]]$. If the x-coordinate of P^* is ξ and the change in slope at P^* is N, then there are, counting multiplicity, precisely N values of $x \in \overline{K}$ for which $f(x) = 0$ and $v(x) = \xi$.*

The valuation function of $f(x)$, denoted $\Psi_f(x)$, is a real-valued polygonal function defined for nonnegative values whose graph is the upper boundary of the Newton copolygon. We know that for any $\alpha \in \overline{\mathcal{M}}$ if $v(\alpha)$ is not the x-coordinate of any vertex of the Newton copolygon, then the relation $v(f(\alpha)) = \Psi_f(v(\alpha))$ holds. It follows from this that if $g(x)$ is another series without constant term, then

$$\Psi_f \circ \Psi_g = \Psi_{f \circ g}.$$

If $f(x) \in \mathcal{S}_0(\mathcal{O})$ is a noninvertible stable series with $\operatorname{wideg}(f) = d < \infty$, then the valuation function of $f(x)$, $\Psi_f(x)$, is a strictly increasing polygonal function with finitely many segments. The leftmost segment is the line $y = dx$ and the rightmost segment continuing to infinity is the line $y = x + v(f'(0))$. All the segments of $\Psi_f(x)$ lie entirely above the line $y = x$ (*i.e.* $\Psi_f(x) > x$). We want to know which series commute with f. One sees that if $f \circ g = g \circ f$, then $\Psi_f \circ \Psi_g = \Psi_{f \circ g} = \Psi_{g \circ f} = \Psi_g \circ \Psi_f$. Therefore, we first treat the problem of which polygonal functions commute with $\Psi_f(x)$. Since we are talking about valuation functions of some series over $\mathcal{O}$ with finite Weierstrass degree, so we just pay attention to increasing polygonal functions, $\Phi : \mathbf{R}^+ \to \mathbf{R}^+$ starting at the point (0, 0) (*i.e.* $\Phi(0) = 0$), and having finitely many segments.

Lemma 2.4. *Let $\Phi_1(x)$, $\Phi_2(x)$ be two polygonal functions which commute with $\Psi_f(x)$.*

(1) If the leftmost segment of $\Phi_1(x)$ and $\Phi_2(x)$ is equal, then $\Phi_1(x) = \Phi_2(x)$.

(2) If the rightmost segment of $\Phi_1(x)$ and $\Phi_2(x)$ is equal, then $\Phi_1(x) = \Phi_2(x)$.

Proof. (1) Suppose that $\Phi_1(x)$ and $\Phi_2(x)$ are two different polygonal functions commuting with $\Psi_f(x)$ such that their leftmost segments have same slope. We can find ξ and ϵ such that $\Phi_1(x) = \Phi_2(x)$ when $0 \leq x \leq \xi$, and $\Phi_1(x) = r_1 x + s_1$, $\Phi_2(x) = r_2 x + s_2$ when $x \in I = (\xi, \xi + \epsilon)$, where $r_1 \neq r_2$. Since $\Psi_f(x) > x$ and $\Psi_f(x)$ is a continuous increasing function, we can find $I' = (\xi', \xi' + \epsilon')$ such that $\xi' + \epsilon' < \xi$ and $\Psi_f(I') \subseteq I$. After adjusting ϵ', we can suppose that $\Psi_f(x) = rx + s$ when $x \in I'$. Hence $\Psi_f(\Phi_1(x)) = \Phi_1(\Psi_f(x)) = r_1(rx + s) + s_1$ and $\Psi_f(\Phi_2(x)) =$

$\Phi_2(\Psi_f(x)) = r_2(rx+s)+s_2$, if $x \in I'$. Since $\Phi_1(x) = \Phi_2(x)$, if $x \in I'$, we have $r_1(rx+s)+s_1 = r_2(rx+s)+s_2$ for all $x \in I'$. Thus $r_1 r = r_2 r$ and $r_1 s + s_1 = r_2 s + s_2$. Since $r \neq 0$, we have that $s_1 = s_2$ and $r_1 = r_2$. This is a contradiction.

(2) Repeat the argument as above *mutatis mutandis.* □

Lemma 2.5. *Let* $\operatorname{wideg}(f) = d = p_1^{\alpha_1} \cdots p_\lambda^{\alpha_\lambda} = q^c$, *where* $c = \gcd(\alpha_1, \ldots, \alpha_\lambda)$ *and* p_i*'s are distinct primes. Then for any polygonal function with finitely many segments which commutes with* $\Psi_f(x)$, *its leftmost segment must have slope* q^t *for some* $t \in \mathbf{Z}$.

Proof. Let $\Phi(x)$ be any polygonal function which satisfies our assumption. Since the slope and y-intercept of every segment of $\Psi_f(x)$ are integers, it is easy to check that the leftmost segment of $\Phi(x)$ is $y = rx$ and the rightmost segment of $\Phi(x)$ is $y = x + a$, where r and a are rational numbers. Because a is a rational number, we can find $n, m \in \mathbf{Z}$ such that $nv(f'(0)) = ma$. The rightmost segment of $\Psi_f^{\circ n}(x)$ (*resp.* $\Phi^{\circ m}(x)$) is $y = x + nv(f'(0))$ (*resp.* $y = x + ma$). Lemma 2.4 tells us that $\Psi_f^{\circ n}(x) = \Phi^{\circ m}(x)$. The slope of the leftmost segment of $\Psi_f^{\circ n}(x)$ (*resp.* $\Phi^{\circ m}(x)$) is d^n (*resp.* r^m). We have that $p_1^{n\alpha_1} \cdots p_\lambda^{n\alpha_\lambda} = r^m$. Suppose that $r = p_1^{\beta_1} \cdots p_\lambda^{\beta_\lambda}$. The equality tell us that

$$m\beta_1 = nc\frac{\alpha_1}{c}\,;\; m\beta_2 = nc\frac{\alpha_2}{c}\,; \ldots ;\, m\beta_\lambda = nc\frac{\alpha_\lambda}{c}.$$

Since $\gcd(\alpha_1/c, \ldots, \alpha_\lambda/c) = 1$, we have that $m \,|\, nc$. Let $t = nc/m$. Then $r = q^t$. □

Proposition 2.6. *The set of polygonal functions which have finitely many segments and commute with* $\Psi_f(x)$ *is a cyclic group under composition.*

Proof. Since $\Psi_f(x)$ is a strictly increasing function, every polygonal function, $\Phi(x)$, which commutes with $\Psi_f(x)$ is also strictly increasing. Therefore there exists a unique polygonal function $\Phi^{\circ -1}(x)$ which commutes with $\Psi_f(x)$ and satisfies $\Phi \circ \Phi^{\circ -1} = \Phi^{\circ -1} \circ \Phi = I(x)$ (identity map).

For every polygonal function which satisfies our assumption the slope of its leftmost segment has the form q^t. Choose t_o to be the smallest positive integer among those t's and let $\Psi_0(x)$ be such a polygonal function whose leftmost segment is $y = q^{t_o}x$. Let $\Phi(x)$ be any polygonal function which has finitely many segments and commutes with $\Psi_f(x)$. Then the leftmost segment of $\Phi(x)$ has slope q^t, for some $t \in \mathbf{Z}$. Assume $t = mt_o + i$, for some $m \in \mathbf{Z}$ and $0 < i < t_o$. We have that $\Psi_0^{\circ -m} \circ \Phi(x)$ is a polygonal function which has finitely many segments and commutes with $\Psi_f(x)$. The slope of the leftmost segment of $\Psi_0^{\circ -m} \circ \Phi(x)$ is q^i. This contradicts the choice of t_o. Thus $t = mt_o$, so $\Phi(x) = \Psi_0^{\circ m}(x)$. This group is cyclic, generated by $\Psi_0(x)$. □

Using the fact that the commuting family of $\Psi_f(x)$ is a cyclic group, we have the following results

Corollary 2.6.1. *Suppose that* $\Phi_1(x)$ *and* $\Phi_2(x)$ *are two polygonal functions which have finitely many segments and commute with* Ψ_f *and both the leftmost segments of* Φ_1, Φ_2 *have slope greater than* 1. *Let the x-coordinates of the leftmost vertex and the rightmost vertex of* $\Phi_i(x)$ *be* ξ_i *and* δ_i, $(i = 1, 2)$, *respectively. Then*

(1) $\Phi_1(x), \Phi_2(x) > x$.

(2) $\Phi_1(\xi_1) = \Phi_2(\xi_2)$ *and* $\delta_1 = \delta_2$.

(3) If the slope of the leftmost segment of Φ_1 is less than the slope of the leftmost segment of Φ_2 and $\Phi_1(\xi') \geq \delta_1$, then for $x \geq \xi'$, $\Phi_2(x) = \Phi_1(x) + c$ for some c.

Proof. Let $\Phi(x)$ be the generater of the commutant group of $\Psi_f(x)$ with the slope of its leftmost segment greater than 1 and suppose that the x-coordinates of the leftmost vertex and the rightmost vertex are ξ and δ, respectively.

(1): Assume that $\Phi(x) \leq x$ for some x. Since the slope of the leftmost segment of $\Phi(x)$ is greater than 1, there exists $\varepsilon > 0$ such that $\Phi(x) > x$, if $0 \leq x < \varepsilon$. Because $\Phi(x)$ is a continuous function, there exists a first ζ such that $\Phi(\zeta) = \zeta$. Hence $\Phi^{\circ n}(\zeta) = \zeta$, for all $n \in \mathbf{N}$. This implies $\Psi_f(\zeta) = \zeta$. It contradicts $\Psi_f(x) > x$ for all x. Our first assertion follows.

(2): Suppose that the leftmost segment of $\Phi(x)$ is $y = rx$. Then it is easy to check that the leftmost segment of $\Phi^{\circ 2}(x)$ is $y = r^2x$ and the x-coordinate of the leftmost vertex of $\Phi^{\circ 2}(x)$ is ξ/r. By induction, we obtain that the leftmost segment of $\Phi^{\circ n}(x)$ is $y = r^n x$ and the x-coordinate of the leftmost segment of $\Phi^{\circ n}(x)$ is ξ/r^{n-1}, for all $n \in \mathbf{N}$. Since $\Phi^{\circ n}(\xi/r^{n-1}) = r\xi = \Phi(\xi)$, we have $\Phi_1(\xi_1) = \Phi_2(\xi_2) = \Phi(\xi)$.

Suppose that $\Phi(x) = x + s$, if $\delta \leq x$. Since $\Phi(x) > x$, there exists $\delta' < \delta$ such that $\Phi(\delta') = \delta$. Hence $\delta \leq \Phi(x)$, if $\delta' \leq x$. Therefore $\Phi^{\circ 2}(x) = \Phi(x) + s$, if $\delta' \leq x$. Thus the x-coordinate of the rightmost vertex of $\Phi^{\circ 2}(x)$ is δ. By induction, we can obtain that $\Phi^{\circ n}(x) = \Phi(x) + (n-1)s$, if $\delta' \leq x$. Therefore $\delta_1 = \delta_2$.

(3): Since the slope of the leftmost segment of Φ_2 is greater than the slope of the leftmost segment of Φ_1, there exists $m \in \mathbf{N}$ such that $\Phi_2 = \Phi^{\circ m} \circ \Phi_1$. Suppose that $\Phi_1(\xi') \geq \delta_1$. Since $\delta = \delta_1$ and Φ_1 is a strictly increasing function, we have that $\Phi_1(x) > \delta$, if $x > \xi'$. Therefore $\Phi \circ \Phi_1(x) = \Phi_1(x) + s$, if $x \geq \xi'$. By induction, we have that $\Phi_2(x) = \Phi^{\circ m} \circ \Phi_1(x) = \Phi_1(x) + ms$, if $x \geq \xi'$. □

3. Main Results

Proposition 3.1. *Let f and g be noninvertible series in $\mathcal{S}_0(\mathcal{O})$ such that $f \circ g = g \circ f$. Then* $\operatorname{wideg}(f) = \operatorname{wideg}(g)$ *if and only if $v(f'(0)) = v(g'(0))$.*

Proof. Ψ_f and Ψ_g are two polygonal functions which commute with each other. Since the slope of the leftmost segment of Ψ_f (*resp.* Ψ_g) is $\operatorname{wideg}(f)$ (*resp.* $\operatorname{wideg}(g)$) and the rightmost segment of Ψ_f (*resp.* Ψ_g) is $y = x + v(f'(0))$ (*resp.* $y = x + v(g'(0))$), by Lemma 2.4, our claim follows. □

Proposition 3.2. *Let f be a noninvertible series with Weierstrass degree equal to $p_1^{\alpha_1} \cdots p_\lambda^{\alpha_\lambda} = (p_1^{\frac{\alpha_1}{c}} \cdots p_\lambda^{\frac{\alpha_\lambda}{c}})^c = q^c$, where $c = \gcd(\alpha_1, \alpha_2, \ldots, \alpha_\lambda)$ and p_i's are distinct primes. If g is another noninvertible series commuting with f, then* $\operatorname{wideg}(g) = q^t$ *for some $t \in \mathbf{N}$.*

Proof. First we check that $\operatorname{wideg}(g) \neq \infty$. If $\operatorname{wideg}(g) = \infty$, then $g(x) = \pi^m h(x)$ with $m > 0$ and $\operatorname{wideg}(h) = l$. Hence the first segment of Ψ_g is $y = lx + m$. But the first segment of Ψ_f is $y = q^c x$, we cannot have $\Psi_g \circ \Psi_f = \Psi_f \circ \Psi_g$. This contradicts our assumption. Since $\operatorname{wideg}(g) \neq \infty$, Lemma 2.5 tells us that $\operatorname{wideg}(g) = q^t$ for some $t \in \mathbf{N}$. □

Corollary 3.2.1. *If $f \circ g = g \circ f$, then*

$$\operatorname{wideg}(g)^{v(f'(0))} = \operatorname{wideg}(f)^{v(g'(0))}.$$

Proof. By Proposition 3.2, we can assume that $\operatorname{wideg}(f) = q^c$ and $\operatorname{wideg}(g) = q^t$. Since $f \circ g = g \circ f$, it implies $f^{\circ t} \circ g^{\circ c} = g^{\circ c} \circ f^{\circ t}$. Because $\operatorname{wideg}(f^{\circ t}) = (q^c)^t =$

$(q^t)^c$ =wideg$(g^{\circ c})$, by Proposition 3.1, we have $v\Big((f^{\circ t})'(0)\Big) = v\Big((g^{\circ c})'(0)\Big)$. Since $(f^{\circ t})'(0) = (f'(0))^t$ and $(g^{\circ c})'(0) = (g'(0))^c$, our assertion follows. □

Proposition 2.6 tells us that there exists a polygonal function Φ such that it generates all the polygonal functions which commute with Ψ_f. Actually, under some conditions Ψ_f itself can be a generator.

Corollary 3.2.2. *Let $f(x)$ be a noninvertible stable series with Weierstrass degree equal to $p_1^{\alpha_1}\cdots p_\lambda^{\alpha_\lambda} = q^c$, where $c = \gcd(\alpha_1, \dots, \alpha_\lambda)$. If $\gcd(c, v(f'(0))) = 1$, then every noninvertible stable series which commutes with $f(x)$ has the form $\mu \circ f^{\circ m}$ for some $m \in \mathbf{N}$ and $\mu(x) \in K[[x]]$ with $\mu(0) = 0$, $\mu'(0) \in \mathcal{O}^*$ and $\mu \circ f = f \circ \mu$.*

Proof. If $g(x)$ is noninvertible and commuting with $f(x)$, we have wideg$(g) = q^t$ for some $t \in \mathbf{N}$ and $t \cdot v(f'(0)) = c \cdot v(g'(0))$. Since $\gcd(c, v(f'(0))) = 1$, we have that $c \,|\, t$. Let $m = t/c$. Then wideg(g) =wideg$(f^{\circ m})$ and $v(g'(0)) = v\Big((f^{\circ m})'(0)\Big)$. Thus $g'(0) = \varpi \cdot (f^{\circ m})'(0)$ for some $\varpi \in \mathcal{O}^*$. There is a unique $\mu(x) \in K[[x]]$ such that $\mu \circ f = f \circ \mu$ and $\mu'(0) = \varpi$. (Recall that $\mu = L_f^{\circ -1}(\varpi L_f)$.) Since both g and $\mu \circ f^{\circ m}$ commute with f and have the same first degree coefficient, by uniqueness, $g = \mu \circ f^{\circ m}$. □

The power series $\mu(x)$ above, although it satisfies $\mu \circ f(x) \in \mathcal{O}[[x]]$ and $\mu'(0) \in \mathcal{O}$, may not be a series over $\mathcal{O}$. This leads us to check some integrality properties.

Definition. Let $f(x)$ and $g(x) \in \mathcal{O}[[x]]$, without constant terms.
Denote $f \,|\, g$, if there exists $h(x) \in \mathcal{O}[[x]]$, such that $g(x) = h(x) \cdot f(x)$.
Denote $f \parallel g$, if there exists $h(x) \in \mathcal{O}[[x]]$, such that $g(x) = h \circ f(x)$.

Proposition 3.3. *If $\hbar_1 \circ f = \hbar_2 \circ g$ with $\hbar_1$, $\hbar_2 \in \mathcal{O}[[x]]$ without constant terms and $\hbar_1'(0) \neq 0$, $\hbar_2'(0) \neq 0$, then any common root of $f(x)$ and $g(x)$ is a root of $f(x)$, $g(x)$ and $\hbar_1(f(x))$ of the same multiplicity.*

Proof. Let $\alpha \in \overline{\mathcal{M}}$ be a root of $f(x)$ and $g(x)$. $\hbar_1'(f(x))f'(x) = \hbar_2'(g(x))g'(x)$, so $\hbar_1'(0)f'(\alpha) = \hbar_2'(0)g'(\alpha)$. Hence $f'(\alpha) = 0$ if and only if $g'(\alpha) = 0$. Denote $f'^{(n)}$ as the n-th derivative of f.

$$(\hbar_1 \circ f)'^{(n)} = (\hbar_1' \circ f) \cdot f'^{(n)} + \sum_{i=2}^{n} \sum_{j_1+\cdots+j_i=n} C_{j_1\dots j_i}(\hbar_1'^{(i)} \circ f) \cdot f'^{(j_1)} \cdots f'^{(j_i)} \ ,$$

where $C_{j_1\dots j_i}$ is some constant. We use induction. Suppose that $f'^{(j)}(\alpha) = g'^{(j)}(\alpha) = 0$, for all $j < n$. Since $(\hbar_1 \circ f)'^{(n)}(\alpha) = \hbar_1'(0)f'^{(n)}(\alpha) = \hbar_2'(0)g'^{(n)}(\alpha)$, we have that $f'^{(n)}(\alpha) = 0$ if and only if $g'^{(n)}(\alpha) = 0$, if and only if $(\hbar_1 \circ f)'^{(n)}(\alpha) = 0$. □

Corollary 3.3.1. *Let $f(x)$ and $g(x)$ be noninvertible series in $\mathcal{S}_0(\mathcal{O})$, and $f \circ g = g \circ f$. If $f(\alpha) = g(\alpha) = 0$ then α is a zero of $f(x)$ and $g(x)$ of the same multiplicity.*

Proof. Since $g \circ f = f \circ g$ and $f'(0) \neq 0$, $g'(0) \neq 0$, our claim follows. □

Let $f, g \in \mathcal{O}[[x]]$ and $f \circ g = g \circ f$. If $f \parallel g$, then it is easy to see that $f \,|\, g$. Our goal is to prove that the converse is also true.

Consider two rings, $\mathcal{D} = \mathcal{O}[[x]]$ and $\mathcal{D}_f = \mathcal{O}[[f(x)]]$. $\mathcal{D}$ is a complete local ring with maximal ideal $(\mathcal{M},\, x)$. Since $\mathcal{D}_f \simeq \mathcal{D}$, $\mathcal{D}_f$ is also a complete local ring with maximal ideal $(\mathcal{M},\, f(x))$. Let $\mathcal{K}$ and $\mathcal{K}_f$ be the quotient fields of $\mathcal{D}$ and $\mathcal{D}_f$

respectively. Consider $f(T)-f(x)$ as a power series in T over $\mathcal{D}_f$, *i.e.* $f(T)-f(x) \in \mathcal{D}_f[[T]]$. According to the Weierstrass Preparation Theorem there exist a unit power series $U(T) \in \mathcal{D}_f[[T]]$ and a distinguished polynomial $F(T) \in \mathcal{D}_f[T]$ such that $f(T)-f(x) = F(T)U(T)$ and $\deg(F) =$wideg$(f) = d$. $F(T)$ is an Eisenstein polynomial over $\mathcal{D}_f$ and $\mathcal{D}_f$ is a UFD (unique factorization domain), so $F(T)$ is an irreducible polynomial over $\mathcal{K}_f$. Let $\mathcal{L}$ be the splitting field of $F(T)$. $\mathcal{L}$ is a Galois extension over $\mathcal{K}_f$ with Galois group Γ. $T = x$ is a root of $F(T)$, so for any $\tau \in \Gamma$, x^τ is also a root of $F(T)$ and then $f(x^\tau) = f(x) = (f(x))^\tau$. $\mathcal{D}$ is a free $\mathcal{D}_f$-module with basis $\{1, x, \ldots, x^{d-1}\}$. Thus, for all $g(x) \in \mathcal{D}$, we can write $g(x)$ as $g_0(f(x)) + g_1(f(x))x + \cdots + g_{d-1}(f(x))x^{d-1}$, where $g_i(x) \in \mathcal{O}[[x]]$. Hence

$$\begin{aligned}(g(x))^\tau &= g_0(f(x)) + g_1(f(x))x^\tau + \cdots + g_{d-1}(f(x))(x^\tau)^{d-1}\\ &= g_0(f(x^\tau)) + g_1(f(x^\tau))x^\tau + \cdots + g_{d-1}(f(x^\tau))(x^\tau)^{d-1} = g(x^\tau).\end{aligned}$$

If $g(x^\tau) = g(x)$ for all $\tau \in \Gamma$, then $g(x) \in \mathcal{K}_f$. Since $\mathcal{D}$ is integral over $\mathcal{D}_f$ and $\mathcal{D}_f$ is integrally closed, we have $g(x) \in \mathcal{O}[[f(x)]]$.

Theorem 3.4. *Let $f(x)$, $g(x)$, $\hbar_1(x)$ and $\hbar_2(x) \in \mathcal{S}_0(\mathcal{O})$ with $\hbar_1 \circ f = \hbar_2 \circ g$. Then $f \mid g$ if and only if $f \parallel g$.*

In particular, if $\hbar_1 = g$ and $\hbar_2 = f$, i.e. $f \circ g = g \circ f$, then $f \mid g$ if and only if $f \parallel g$.

Proof. We only have to prove that $f \mid g$ implies $f \parallel g$. Keep the notation as above. Suppose $f \mid g$. We will show that every root of $f(T) - f(x)$ is a root of $g(T) - g(x)$. For any $\tau \in \Gamma$, x^τ is a root of $f(T)-f(x)$. Therefore x^τ will be a root of $g(T)-g(x)$. Thus $(g(x))^\tau = g(x^\tau) = g(x)$, for all $\tau \in \Gamma$. Hence $g(x) \in \mathcal{O}[[f(x)]]$, *i.e.* $g(x) = h(f(x))$ for some $h(x) \in \mathcal{O}[[x]]$.

Let $\mathcal{D} = \mathcal{O}[[x]]$ with quotient field $\mathcal{K}$, $\mathcal{D}_f = \mathcal{O}[[f(x)]]$ with quotient field $\mathcal{K}_f$ and $\mathcal{D}_g = \mathcal{O}[[g(x)]]$ with quotient field $\mathcal{K}_g$. Consider $f(T) - f(x) \in \mathcal{D}_f[[T]]$. There exists $F(T) \in \mathcal{D}_f[T]$ a monic irreducible polynomial over $\mathcal{K}_f$ and a unit power series $U_1(T) \in \mathcal{D}_f[[T]]$, such that $f(T) - f(x) = F(T)U_1(T)$. Consider $\hbar_1(f(T)) - \hbar_1(f(x)) \in \mathcal{D}_f[[T]]$. We can also have $\hbar_1(f(T)) - \hbar_1(f(x)) = H_f(T)U_2(T)$, where $H_f(T)$ is a monic polynomial in $\mathcal{D}_f[T]$ and $U_2(T)$ is a unit of $\mathcal{D}_f[[T]]$. $T = x$ is a root of $f(T)-f(x)$ and $\hbar_1(f(T))-\hbar_1(f(x))$. Since $F(T)$ is the minimal polynomial of x in $\mathcal{K}_f[T]$, we have $F(T)|H_f(T)$ in $\mathcal{D}_f[T]$. Thus every root of $f(T) - f(x)$ is also a root of $\hbar_1(f(T)) - \hbar_1(f(x))$. Next, we consider $g(T) - g(x)$ as a power series in T over $\mathcal{D}_g[[x]]$. For the same reason as above, there exist $G(T) \in \mathcal{D}_g[T]$ a monic irreducible polynomial over $\mathcal{K}_g$ and a unit power series $U_3(T) \in \mathcal{D}_g[[T]]$, such that $g(T) - g(x) = G(T)U_3(T)$. Since $\hbar_2(g(T)) - \hbar_2(g(x)) \in \mathcal{D}_g[[T]]$, we have that $\hbar_2(g(T)) - \hbar_2(g(x)) = H_g(T)U_4(T)$ where $H_g(T)$ is a monic polynomial in $\mathcal{D}_g[T]$ and $U_4(T)$ is a unit of $\mathcal{D}_g[[T]]$. For the same reason as above, we have $G(T)|H_g(T)$ in $\mathcal{D}_g[T]$. Notice that $\mathcal{D}_f$, $\mathcal{D}_g \subset \mathcal{D}$. Therefore by the uniqueness of Weierstrass Preparation Theorem and $\hbar_1 \circ f = \hbar_2 \circ g$, we have $H_f(T) = H_g(T)$. For convenience, we let $H(T) = H_f(T) = H_g(T)$. Thus $F(T)|H(T)$ and $G(T)|H(T)$ in $\mathcal{D}[T]$.

Consider $F(T)$ as an element of $\mathcal{D}[T]$. Because $\mathcal{D}$ and $\mathcal{D}[T]$ are UFD, we can factorize $F(T) \in \mathcal{D}[T]$ into $F_1(T) \cdots F_s(T)$, where $F_1(T), \ldots, F_s(T)$ are monic irreducible polynomials over $\mathcal{K}$. Using Gauss' Lemma we have $F_1(T), \ldots, F_s(T) \in \mathcal{D}[T]$. Let $f_i(x) \in \mathcal{D} = \mathcal{O}[[x]]$ be the constant term of $F_i(T)$, *i.e.* $f_i(x) = F_i(0)$. Notice that $F(T) \equiv T^d$ mod the maximal ideal $(\mathcal{M}, x)$ of $\mathcal{D}$ where $d =$ wideg(f). Therefore $F_i(T) \equiv T^{d_i}$ (mod $\mathcal{M}$, x) for some $d_i > 0$. This implies that $f_i(x)$, the constant term of $F_i(T)$, is in $(\mathcal{M}, x)$. Hence $f_i(x)$ is not a unit in $\mathcal{D}$. We have $f_1(x)\cdots f_s(x) = f(x)u_1(x)$ where $u_1(x)$ is a unit of $\mathcal{D}$, because $F(0)U_1(0) = -f(x)$. Using same argument we can factorize $G(T) = G_1(T)\cdots G_t(T)$, where $G_i(T) \in \mathcal{D}[T]$ is a monic irreducible polynomial over $\mathcal{K}$ with constant term $g_i(x)$, and we have $g_1(x)\cdots\ g_t(x) = g(x)u_2(x)$ where $u_2(x)$ is a unit of $\mathcal{D}$. We can also factorize $H(T)$ into $H_1(T)\cdots H_r(T)$ in $\mathcal{D}[T]$. Since $F(T)|H(T)$ and $G(T)|H(T)$ in $\mathcal{D}[T]$, $\forall i \in \{1, \dots, s\}$ and $\forall j \in \{1, \dots, t\}$ there exist $i', j' \in \{1, \dots, r\}$ such that $F_i(T) = H_{i'}(T)$ and $G_j(T) = H_{j'}(T)$ respectively.

Assume T_1 is a root of $F(T)$ but not a root of $G(T)$. Thus there exists $i \in \{1, \dots, s\}$ such that $F_i(T_1) = 0$ but $F_i(T) \neq G_j(T)$, $\forall\, j \in \{1, \dots, t\}$. We reorder them such that $F_1(T_1) = 0$ and $F_1(T) = H_1(T), G_1(T) = H_2(T), \dots, G_t(T) = H_{t+1}(T)$. Since $F_1(T)G_1(T)\cdots G_t(T)|H_1(T)H_2(T)\cdots H_{t+1}(T)\cdots H_r(T) = H(T)$, by taking $T = 0$, we have $f_1(x)g_1(x)\cdots g_t(x)|\hbar_1(f(x))$. Thus $f_1(x)g(x)|\hbar_1(f(x))$. Since $f_1(x)$ is not a unit in $\mathcal{D} = \mathcal{O}[[x]]$ and $f \,|\, g$, there exists $\alpha \in \overline{\mathcal{M}}$ such that $f_1(\alpha) = g(\alpha) = 0$. This tells us that α is a root of $\hbar_1(f(x))$ with higher multiplicity than as a root of $g(x)$. This is contrary to our Proposition 3.3. Hence every root of $F(T)$ is a root of $G(T)$. Our claim follows. □

Remark. (1) This theorem is a generalization of Lubin's results [4, Theorems 1.4, 1.5].

(2) If $g(x) = h \circ f(x)$ and $f \circ g = g \circ f$, then $f \circ h \circ f = f \circ g = g \circ f = h \circ f \circ f$. Therefore $f \circ h = h \circ f$.

From now on, we denote $\mathrm{Comm}_{\mathcal{O}}(f)$ to be the set of all $g \in \mathcal{O}[[x]]$ with $g \circ f = f \circ g$. Proposition 3.3 and Theorem 3.4 say that if $f \circ g = g \circ f$, then the necessary and sufficient condition for $f \parallel g$ is that every root of f is also a root of g. One may ask under what circumstance this may happen. Let us return to polygonal functions. Let $f(x)$ be a noninvertible series in $\mathcal{S}_0(\mathcal{O})$ and let ξ and δ be the x-coordinates of the leftmost and rightmost vertices of $\mathcal{N}^*(f)$, respectively. If $\Psi_f(\xi) > \delta$, then there is $\xi' < \xi$ such that $\Psi_f(\xi) \geq \delta$. For any $h(x) \in \mathcal{O}[[x]]$ with $h \circ f = f \circ h$ and $\mathrm{wideg}(h) \geq \mathrm{wideg}(f)$, by Corollary 2.6.1, $\Psi_h(x) = \Psi_f(x) + c$, $\forall x \geq \xi'$. Thus if $P^* = (\xi_o, \Psi_f(\xi_o))$ is a vertex of $\mathcal{N}^*(f)$ and the change in slope at P^* is N, then $P'^* = (\xi_o, \Psi_h(\xi_o))$ is also a vertex of $\mathcal{N}^*(h)$ with the change in slope at P'^* equal to N. Suppose that $g \circ f = f \circ g$ and $\mathrm{wideg}(g) \geq \mathrm{wideg}(f)$. We claim that every root of f is also a root of g. Assume there is $\alpha \in \overline{\mathcal{M}}$ such that $f(\alpha) = 0$ but $g(\alpha) \neq 0$. Then since $g \circ f = f \circ g$, there exists n such that $g^{\circ n}(\alpha) = 0$. This says that the number of roots of $g^{\circ n}(x)$ with valuation equal to $v(\alpha)$ is greater than the number of roots of $g(x)$ with valuation equal to $v(\alpha)$. Since $f(x)$ and $g(x)$ have the same number of roots with valuation equal to $v(\alpha)$, it implies that the change in slope of $\mathcal{N}^*(g^{\circ n})$ at the vertex $\Big(v(\alpha), \Psi_{g^{\circ n}}(v(\alpha))\Big)$ is greater than the change in slope of $\mathcal{N}^*(f)$ at the vertex $\Big(v(\alpha), \Psi_f(v(\alpha))\Big)$. This is a contradiction. Therefore we have $f \,|\, g$. On the other side, suppose that $f \circ g = g \circ f$ and $\mathrm{wideg}(g) < \mathrm{wideg}(f)$ with the x-coordinates of the leftmost and rightmost vertices of $\mathcal{N}^*(g)$ equal to ξ_1 and δ_1, respectively. Then by Corollary 2.6.1, $\Psi_g(\xi_1) = \Psi_f(\xi) > \delta = \delta_1$. Using the same reason as above, we can get $g \,|\, f$. Therefore, if $\mathcal{N}^*(f)$ satisfies the condition mentioned above, then for all g_1, $g_2 \in \mathrm{Comm}_{\mathcal{O}}(f)$, we have that $g_1 \,|\, g_2$ if and only if $\mathrm{wideg}(g_1) \leq \mathrm{wideg}(g_2)$.

Corollary 3.4.1. *Let $f(x)$ be a noninvertible series in $\mathcal{S}_0(\mathcal{O})$ with Weierstrass degree d. Let the x-coordinates of the leftmost vertex and the rightmost vertex of $\mathcal{N}^*(f)$ be ξ and δ respectively. If $d\xi > \delta$, then there exists $h(x) \in \mathcal{O}[[x]]$ with $h \circ f = f \circ h$ such that $\forall\, g \in \mathrm{Comm}_{\mathcal{O}}(\mathrm{f})$, $g = u \circ h^{\circ m}$ for some $m \in \mathbf{N}$ and $u(x)$ an invertible series in $\mathrm{Comm}_{\mathcal{O}}(f)$.*

Proof. For any noninvertible series $g_1(x)$, $g_2(x) \in \mathrm{Comm}_{\mathcal{O}}(f)$, we have that $g_1 \,|\, g_2$ if and only if $\mathrm{wideg}(g_1) \leq \mathrm{wideg}(g_2)$. Let $h(x) \in \mathrm{Comm}_{\mathcal{O}}(f)$ be a noninvertible series with smallest Weierstrass degree q^t. For any $g(x) \in \mathrm{Comm}_{\mathcal{O}}(f)$, $\mathrm{wideg}(g) = q^s$. Let $s = mt + r$ with $0 \leq r < t$. Then $\mathrm{wideg}(g) \geq \mathrm{wideg}(h^{\circ m})$. Hence $h^{\circ m} \,|\, g$ and then $g = h_1 \circ h^{\circ m}$ for some $h_1(x) \in \mathcal{O}[[x]]$, by Theorem 3.4. $h_1(x) \in \mathrm{Comm}_{\mathcal{O}}(f)$ and $\mathrm{wideg}(h_1) = q^r < \mathrm{wideg}(h)$. Hence $\mathrm{wideg}(h_1) = 1$. Thus $h_1(x)$ is an invertible series. □

Example. Let $f(x) = \sum_{i=1}^{\infty} a_i x^i \in \mathcal{O}[[x]]$ with $\mathrm{wideg}(f) = d$ and $\upsilon(a_i) \geq \frac{d-i}{d-1}\upsilon(a_1)$, $\forall\, i < d$. Thus the Newton polygon of $f(x)$ has only one segment. (The simplest example is $f(x) = \pi x + x^d$.) It is easy to check that there is no polygonal function which has finitely many segments with the slope of its leftmost segment less than d that can commute with $\Psi_f(x)$. Since the Newton copolygon of $f(x)$ has only one vertex with x-coordinates ξ. The leftmost vertex is the same as the rightmost vertex. $d\xi > \xi$ tells us that for every $g(x) \in \mathrm{Comm}_{\mathcal{O}}(f)$, $g = u \circ f^{\circ m}$ where $u(x)$ is an invertible series in $\mathrm{Comm}_{\mathcal{O}}(f)$.

Remark. This Example is a generalization of Lubin's result [3, Theorem 3.3.1].

Let us return to polygonal functions again. Keep all the notations as above. We can not always have $\Psi_f(\xi) > \delta$, but there exists n such that $\Psi_f^{\circ n}(\xi) = \Psi_{f^{\circ n}}(\xi) > \delta$. Again, by Corollary 2.6.1, for any $g \in \mathrm{Comm}_{\mathcal{O}}(f)$ with $\mathrm{wideg}(g) \geq \mathrm{wideg}(f^{\circ n})$, we have $\Psi_g(x) = \Psi_{f^{\circ n}}(x) + c$ for $x \geq \xi'$. This means that for any $\alpha \in \Lambda(f)$ with $\upsilon(\alpha) \geq \xi$, $g(\alpha) = 0$. For if $g(\alpha) \neq 0$ and $g^{\circ m}(\alpha) = 0$, then the change in slope of $\mathcal{N}^*(g^{\circ m})$ at $\Big(\upsilon(\alpha), \Psi_{g^{\circ m}}(\upsilon(\alpha))\Big)$ is greater than the change in slope of $\mathcal{N}^*(f^{\circ n})$ at $\Big(\upsilon(\alpha), \Psi_{f^{\circ n}}(\upsilon(\alpha))\Big)$. Since all roots of $f(x)$ have valuation greater than or equal to ξ, we have $f \,|\, g$.

Corollary 3.4.2. *Let $g(x) = u \circ f(x) \in \mathrm{Comm}_{\mathcal{O}}(f)$, where $u(x) \in k[[x]]$ with $u'(0) \in \mathcal{O}^*$. Then there exists $n \in \mathbf{N}$ such that $u^{\circ n}(x) \in \mathcal{O}[[x]]$.*

Proof. By the argument above, there exists $m \in \mathbf{N}$ such that for every $h \in \mathrm{Comm}_{\mathcal{O}}(f)$ if $\mathrm{wideg}(h) \geq \mathrm{wideg}(f^{\circ m})$, then $f \,|\, h$. Since $g^{\circ m} = u^{\circ m} \circ f^{\circ m}$ and $\mathrm{wideg}(g^{\circ m}) = \mathrm{wideg}(f^{\circ m})$, we have $f\,|u^{\circ m} \circ f^{\circ m}$. By Theorem 3.4,

$$u^{\circ m} \circ f^{\circ(m-1)}(x) \in \mathcal{O}[[x]].$$

Because $u^{\circ m} \circ f^{\circ(m-1)} \circ g = u^{\circ(m+1)} \circ f^{\circ m} \in \mathrm{Comm}_{\mathcal{O}}(f)$ and $\mathrm{wideg}(u^{\circ(m+1)} \circ f^{\circ m}) = \mathrm{wideg}(f^{\circ m})$, $f \,|\, u^{\circ(m+1)} \circ f^{\circ m}$. Again, by Theorem 3.4, $u^{\circ(m+1)} \circ f^{\circ(m-1)} \in \mathcal{O}[[x]]$. Using induction we have $u^{\circ r} \circ f^{\circ(m-1)}(x) \in \mathcal{O}[[x]]$, for all $r \geq m$. Since in the same commuting family with the same Weierstrass degree, up to composing with a commuting invertible series, we can only have finitely many series with different roots, so we can find $t \geq m$ and $n > 0$ such that $u^{\circ t} \circ f^{\circ(m-1)} | u^{\circ(t+n)} \circ f^{\circ(m-1)}$. Thus $u^{\circ n} \in \mathcal{O}[[x]]$. □

Remark. Since $\forall\, r \geq m$, $u^{\circ r} \circ f^{\circ(m-1)} \in \mathcal{O}[[x]]$ and $u^{\circ n} \in \mathcal{O}[[x]]$, we have that $u^{\circ z} \circ f^{\circ(m-1)} \in \mathcal{O}[[x]]$ for all $z \in \mathbf{Z}$.

Corollary 3.4.3. *Let $g(x) = h \circ f(x) \in \mathrm{Comm}_{\mathcal{O}}(\mathrm{f})$, where $h(x) \in K[[x]]$ with $h'(0) \in \mathcal{M}$. Then there exists $n \in \mathbf{N}$ such that $h^{\circ n}(x) \in \mathcal{O}[[x]]$.*

Proof. Since $\upsilon(g'(0)) = \upsilon(h'(0)) + \upsilon(f'(0)) > \upsilon(f'(0))$, by Corollary 3.2.1, $\mathrm{wideg}(g) > \mathrm{wideg}(f)$. There exists $m \in \mathbf{N}$ such that $f \,|\, g^{\circ m}$, so we have $h^{\circ m} \circ f^{\circ(m-1)}(x) \in \mathcal{O}[[x]]$. Because

$$\mathrm{wideg}(h^{\circ m} \circ f^{\circ(m-1)} \circ g) = \mathrm{wideg}(h^{\circ(m+1)} \circ f^{\circ m}) > \mathrm{wideg}(h^{\circ m} \circ f^{\circ m}),$$

we have $f \,|\, h^{\circ(m+1)} \circ f^{\circ m}$. Hence $h^{\circ(m+1)} \circ f^{\circ(m-1)}(x) \in \mathcal{O}[[x]]$. Using induction, we can get $h^{\circ r} \circ f^{\circ(m-1)}(x) \in \mathcal{O}[[x]]$ for all $r \geq m$. Since $\mathrm{wideg}(h^{\circ(r+1)} \circ f^{\circ(m-1)}) > \mathrm{wideg}(h^{\circ r} \circ f^{\circ(m-1)})$, there exists $s > m$ such that

$$\mathrm{wideg}(h^{\circ s} \circ f^{\circ(m-1)}) \geq \mathrm{wideg}(h^{\circ m} \circ f^{\circ m}).$$

Hence we have $f \,|\, h^{\circ(s-m)} \circ f^{\circ(m-1)}$. Thus $h^{\circ(s-m)} \circ f^{\circ(m-2)}(x) \in \mathcal{O}[[x]]$. Using induction again, we can get $h^{\circ n}(x) \in \mathcal{O}[[x]]$ for some $n \in \mathbf{N}$. □

Remark. In this proof, actually we can get $h^{\circ t}(x) \in \mathcal{O}[[x]]$ for all $t \geq n$.

References

[1] N. Koblitz, *p-adic Numbers, p-adic Analysis, and Zeta-Functions*, Springer-Verlag, New York, 1977. MR **57:**5964

[2] J. Lubin, Nonarchimedean Dynamical Systems, *Comp. Math.* 94 (1994), pp. 321–346. MR **96g:**11140

[3] J. Lubin, One-parameter Formal Lie Groups over p-adic Integer Rings, *Ann. of Math.* 80 (1964), pp. 464–484. MR **29:**5827

[4] J. Lubin, Finite Subgroups and Isogenies of One-parameter Formal Lie Groups, *Ann. of Math.* 85 (1967), pp. 296–302. MR **35:**189

Department of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.
E-mail address: li@math.nthu.edu.tw