HENSEL'S LEMMA

KEITH CONRAD

1. Introduction

In the p-adic integers, congruences are approximations: for a and b in $\mathbf{Z}_p$, $a \equiv b \bmod p^n$ is the same as $|a-b|_p \leq 1/p^n$. Turning information modulo one power of p into similar information modulo a higher power of p can be interpreted as improving an approximation.

Example 1.1. The number 7 is a square mod 3: $7 \equiv 1^2 \bmod 3$. Although $7 \not\equiv 1^2 \bmod 9$, we can write 7 as a square mod 9 by replacing 1 with $1+3$: $7 \equiv (1+3)^2 \bmod 9$. Here are expressions of 7 as a square modulo further powers of 3:

$$\begin{aligned}
7 &\equiv (1+3+3^2)^2 \bmod 3^3,\\
7 &\equiv (1+3+3^2)^2 \bmod 3^4,\\
7 &\equiv (1+3+3^2+2\cdot 3^4)^2 \bmod 3^5,\\
&\ \vdots\\
7 &\equiv (1+3+3^2+2\cdot 3^4+2\cdot 3^7+3^8+3^9+2\cdot 3^{10})^2 \bmod 3^{11}.
\end{aligned}$$

If we can keep going indefinitely then 7 is a perfect square in $\mathbf{Z}_3$ with square root

$$1+3+3^2+2\cdot 3^4+2\cdot 3^7+3^8+3^9+2\cdot 3^{10}+\cdots.$$

That we really can keep going indefinitely is justified by Hensel's lemma, which will provide conditions under which the root of a polynomial mod p can be lifted to a root in $\mathbf{Z}_p$, such as the polynomial X^2-7 with $p=3$: its two roots mod 3 can both be lifted to square roots of 7 in $\mathbf{Z}_3$.

We will first give a basic version of Hensel's lemma, illustrate it with examples, and then other versions that can be applied in cases where the basic version is inadequate.

2. A Basic Version of Hensel's lemma

Theorem 2.1 (Hensel's lemma)**.** *If $f(X) \in \mathbf{Z}_p[X]$ and $a \in \mathbf{Z}_p$ satisfies*

$$f(a) \equiv 0 \bmod p, \quad f'(a) \not\equiv 0 \bmod p$$

then there is a unique $\alpha \in \mathbf{Z}_p$ such that $f(\alpha)=0$ in $\mathbf{Z}_p$ and $\alpha \equiv a \bmod p$.

Example 2.2. Let $f(X) = X^2-7$. Then $f(1) = -6 \equiv 0 \bmod 3$ and $f'(1) = 2 \not\equiv 0 \bmod 3$, so Hensel's lemma tells us there is a unique 3-adic integer α such that $\alpha^2 = 7$ and $\alpha \equiv 1 \bmod 3$. We saw approximations to α in Example 1.1, *e.g.*, $\alpha \equiv 1+3+3^2+2\cdot 3^4 \bmod 3^5$.

Proof. We will prove by induction that for each $n \geq 1$ there is an $a_n \in \mathbf{Z}_p$ such that

- $f(a_n) \equiv 0 \bmod p^n$,
- $a_n \equiv a \bmod p$.

(After proving this, we'll show it leads to a solution of $f(\alpha)=0$ as a limit.) The case $n=1$ is trivial, using $a_1 = a$. If the inductive hypothesis holds for n, we seek $a_{n+1} \in \mathbf{Z}_p$ such that

- $f(a_{n+1}) \equiv 0 \bmod p^{n+1}$,
- $a_{n+1} \equiv a \bmod p$.

Since $f(a_{n+1}) \equiv 0 \bmod p^{n+1} \Rightarrow f(a_{n+1}) \equiv 0 \bmod p^n$, each root of $f(X) \bmod p^{n+1}$ reduces to a root of $f(X) \bmod p^n$. By the inductive hypothesis there is a root $a_n \bmod p^n$ such that $a_n \equiv a \bmod p$, so we seek a p-adic integer a_{n+1} such that $a_{n+1} \equiv a_n \bmod p^n$ and $f(a_{n+1}) \equiv 0 \bmod p^{n+1}$. Writing

$$a_{n+1} = a_n + p^n t_n$$

for some $t_n \in \mathbf{Z}_p$ to be determined, can we make $f(a_n + p^n t_n) \equiv 0 \bmod p^{n+1}$?

To compute $f(a_n + p^n t_n) \bmod p^{n+1}$, we use a polynomial identity:

$$f(X+Y) = f(X) + f'(X)Y + g(X,Y)Y^2 \tag{2.1}$$

for some polynomial $g(X,Y) \in \mathbf{Z}_p[X,Y]$. This formula comes from isolating the first two terms in the binomial theorem: writing $f(X) = \sum_{i=0}^{d} c_i X^i$ we have

$$f(X+Y) = \sum_{i=0}^{d} c_i (X+Y)^i = c_0 + \sum_{i=1}^{d} (c_i(X^i + iX^{i-1}Y) + g_i(X,Y)Y^2),$$

where $g_i(X,Y) \in \mathbf{Z}[X,Y]$. Thus

$$f(X+Y) = \sum_{i=0}^{d} c_i X^i + \sum_{i=1}^{d} ic_i X^{i-1}Y + \sum_{i=1}^{d} g_i(X,Y)Y^2 = f(X) + f'(X)Y + g(X,Y)Y^2,$$

where $g(X,Y) = \sum_{i=1}^{d} c_i g_i(X,Y) \in \mathbf{Z}_p[X,Y]$. This gives us the desired identity.[1]

To make (2.1) numerical, for all x and y in $\mathbf{Z}_p$ the number $z := g(x,y)$ is in $\mathbf{Z}_p$, so

$$x, y \in \mathbf{Z}_p \Longrightarrow \boxed{f(x+y) = f(x) + f'(x)y + zy^2, \text{ where } z \in \mathbf{Z}_p.} \tag{2.2}$$

In this formula set $x = a_n$ and $y = p^n t_n$:

$$f(a_n + p^n t_n) = f(a_n) + f'(a_n)p^n t_n + zp^{2n}t_n^2 \equiv f(a_n) + f'(a_n)p^n t_n \bmod p^{n+1} \tag{2.3}$$

since $2n \geq n+1$. In $f'(a_n)p^n t_n \bmod p^{n+1}$, the factors $f'(a_n)$ and t_n only matter mod p since there is already a factor of p^n present and the modulus is p^{n+1}. Recalling that $a_n \equiv a \bmod p$, we get $f'(a_n)p^n t_n \equiv f'(a)p^n t_n \bmod p^{n+1}$. Therefore from (2.3),

$$\begin{aligned} f(a_n + p^n t_n) \equiv 0 \bmod p^{n+1} &\Longleftrightarrow f(a_n) + f'(a)p^n t_n \equiv 0 \bmod p^{n+1} \\ &\Longleftrightarrow f'(a)t_n \equiv -f(a_n)/p^n \bmod p, \end{aligned} \tag{2.4}$$

where the ratio $f(a_n)/p^n$ is in $\mathbf{Z}_p$ since we assumed that $f(a_n) \equiv 0 \bmod p^n$. There is a solution for t_n in (2.4), a congruence mod p, since we assume $f'(a) \not\equiv 0 \bmod p$.

Armed with this choice of t_n and setting $a_{n+1} = a_n + p^n t_n$, we have $f(a_{n+1}) \equiv 0 \bmod p^{n+1}$ and $a_{n+1} \equiv a_n \bmod p^n$, so in particular $a_{n+1} \equiv a \bmod p$. This completes the induction.

Starting with $a_1 = a$, our inductive argument has constructed a sequence $a_1, a_2, a_3, \ldots$ in $\mathbf{Z}_p$ such that $f(a_n) \equiv 0 \bmod p^n$ and $a_{n+1} \equiv a_n \bmod p^n$ for all n. The second condition, $a_{n+1} \equiv a_n \bmod p^n$, implies that $\{a_n\}$ is a Cauchy sequence in $\mathbf{Z}_p$. Let α be its limit in $\mathbf{Z}_p$. We want to show $f(\alpha) = 0$ and $\alpha \equiv a \bmod p$.

From $a_{n+1} \equiv a_n \bmod p^n$ for all n we get $a_m \equiv a_n \bmod p^n$ for all $m > n$, so $\alpha \equiv a_n \bmod p^n$ by letting $m \to \infty$. At $n = 1$ we get $\alpha \equiv a \bmod p$. For general n,

$$\alpha \equiv a_n \bmod p^n \Longrightarrow f(\alpha) \equiv f(a_n) \equiv 0 \bmod p^n \Longrightarrow |f(\alpha)|_p \leq \frac{1}{p^n}.$$

---

[1]The identity (2.1) is similar to Taylor's formula: $f(x+h) = f(x) + f'(x)h + (f''(x)/2!)h^2 + \cdots$. The catch is that terms in Taylor's formula have factorials in the denominator, which can require some extra care when reducing modulo powers of p: think about $f''(x)/2! \bmod 2$, for instance. What (2.1) essentially does is extract the first two terms of Taylor's formula and say that what remains has *p-adic integral* coefficients, so (2.1) can be reduced mod p, or mod p^n for all $n \geq 1$.

Since this estimate holds for all n, $f(\alpha) = 0$.

It remains to show α is the unique root of $f(X)$ in $\mathbf{Z}_p$ that is congruent to $a \bmod p$. Suppose $f(\beta) = 0$ and $\beta \equiv a \bmod p$. To show $\beta = \alpha$ we will show $\beta \equiv \alpha \bmod p^n$ for all n. The case $n = 1$ is clear since α and β are both congruent to $a \bmod p$. Suppose $n \geq 1$ and we know that $\beta \equiv \alpha \bmod p^n$. Then $\beta = \alpha + p^n\gamma_n$ with $\gamma_n \in \mathbf{Z}_p$, so a calculation similar to (2.3) implies

$$f(\beta) = f(\alpha + p^n\gamma_n) \equiv f(\alpha) + f'(\alpha)p^n\gamma_n \bmod p^{n+1}.$$

Both α and β are roots of $f(X)$, so $0 \equiv f'(\alpha)p^n\gamma_n \bmod p^{n+1}$. Thus $f'(\alpha)\gamma_n \equiv 0 \bmod p$. Since $f'(\alpha) \equiv f'(a) \not\equiv 0 \bmod p$, we have $\gamma_n \equiv 0 \bmod p$, which implies $\beta \equiv \alpha \bmod p^{n+1}$. $\square$

Remark 2.3. Since $a_n + p^n t_n \bmod p^{n+1}$ only depends on $t_n \bmod p$, and t_n satisfying (2.4) is unique mod p, the $a_n \in \mathbf{Z}_p$ where (i) $f(a_n) \equiv 0 \bmod p^n$ and (ii) $a_n \equiv a \bmod p$ is unique modulo p^n. Or an argument similar to the last paragraph in the proof shows for each $n \geq 1$ that $f(X)$ has a unique root mod p^n that reduces to $a \bmod p$. Either way, in Theorem 2.1 we can think about the uniqueness of the lifting of the mod p root in two ways: it has a unique lifting to a root in $\mathbf{Z}_p$ or it has a unique lifting to a root in $\mathbf{Z}_p/(p^n)$ for all $n \geq 1$.

Remark 2.4. If $f(X) \in \mathbf{Z}[X]$ and $a \in \mathbf{Z}$, then $g(X, Y)$ in (2.1) is in $\mathbf{Z}[X, Y]$ and z in (2.2) is in $\mathbf{Z}$. Use this to check that the inductive reasoning used in the proof before passing to a p-adic limit near the end holds with $\mathbf{Z}$ in place of $\mathbf{Z}_p$, so for all $n \geq 1$ there is an $a_n \in \mathbf{Z}$ such that $f(a_n) \equiv 0 \bmod p^n$ and $a_n \equiv a \bmod p$, and such a_n is unique in $\mathbf{Z}/(p^n)$.

Here are six applications of Hensel's lemma.

Example 2.5. Let $f(X) = X^3 - 2$. We have $f(3) \equiv 0 \bmod 5$ and $f'(3) \not\equiv 0 \bmod 5$, Therefore Hensel's lemma with initial approximation $a = 3$ tells us there is a unique cube root of 2 in $\mathbf{Z}_5$ that is congruent to 3 mod 5. Explicitly, it is $3 + 2 \cdot 5^2 + 2 \cdot 5^3 + 3 \cdot 5^4 + \cdots$.

Example 2.6. Let $f(X) = X^3 - X - 2$. We have $f(0) \equiv 0 \bmod 2$ and $f(1) \equiv 0 \bmod 2$, while $f'(0) \equiv 1 \bmod 2$ and $f'(1) \equiv 0 \bmod 2$. Therefore Hensel's lemma with initial approximation $a = 0$ implies there is a unique $\alpha \in \mathbf{Z}_2$ such that $f(\alpha) = 0$ and $\alpha \equiv 0 \bmod 2$. Explicitly, $\alpha = 2 + 2^2 + 2^4 + 2^7 + \cdots$.

Although 1 is a root of $f(X)$ mod 2, it does *not* lift to a root in $\mathbf{Z}_2$ since it doesn't even lift to a root mod 4: $f(1) \equiv 2 \bmod 4$ and $f(3) \equiv 2 \bmod 4$, so if $\beta \in \mathbf{Z}_2$ and $\beta \equiv 1 \bmod 2$ then $\beta \equiv 1$ or $3 \bmod 4$ and therefore $f(\beta) \equiv 2 \not\equiv 0 \bmod 4$.

Example 2.7. For each positive integer n *not* divisible by p and each $u \equiv 1 \bmod p\mathbf{Z}_p$, u is an nth power in $\mathbf{Z}_p^\times$. Apply Hensel's lemma to $f(X) = X^n - u$ with initial approximation $a = 1$: $f(1) = 1 - u \equiv 0 \bmod p$ and $f'(1) = n \not\equiv 0 \bmod p$. Therefore there is a unique solution to $\alpha^n = u$ in $\mathbf{Z}_p$ such that $\alpha \equiv 1 \bmod p$. Example 1.1 is the case $u = 7$, $p = 3$, and $n = 2$: 7 has a unique 3-adic square root that is $\equiv 1 \bmod 3$.

Example 2.8. For an *odd* prime p, suppose $u \in \mathbf{Z}_p^\times$ is a square mod p. We will show u is a square in $\mathbf{Z}_p^\times$. For example, 2 is a square mod 7 since $2 \equiv 3^2 \bmod 7$, and it will follow that 2 is a square in $\mathbf{Z}_7^\times$.

Write $u \equiv a^2 \bmod p$, so $a \not\equiv 0 \bmod p$. For the polynomial $f(X) = X^2 - u$ we have $f(a) \equiv 0 \bmod p$ and $f'(a) = 2a \not\equiv 0 \bmod p$, since p is not 2, so Hensel's lemma tells us that $f(X)$ has a root in $\mathbf{Z}_p$ that reduces to $a \bmod p$, which means u is a square in $\mathbf{Z}_p^\times$. Conversely, if $u \in \mathbf{Z}_p^\times$ is a p-adic square, say $u = v^2$, then $1 = |v|_p^2$, so $v \in \mathbf{Z}_p^\times$ and $u \equiv v^2 \bmod p$. Thus the elements of $\mathbf{Z}_p^\times$ that are squares in $\mathbf{Q}_p$ are precisely those that reduce to squares mod p. For example, the nonzero squares mod 7 are 1, 2, and 4, so $u \in \mathbf{Z}_7^\times$ is a 7-adic square if and only if $u \equiv 1, 2, \text{or } 4 \bmod 7$.

This result can have problems when $p = 2$ because $2a \equiv 0 \bmod 2$. In fact the lifting of a square root mod 2 to a 2-adic square root really does have a problem: $3 \equiv 1^2 \bmod 2$ but 3 is not a square in $\mathbf{Z}_2$ since 3 is not a square mod 4 (the squares mod 4 are 0 and 1). And 3 is not a square in $\mathbf{Q}_2$ either because a hypothetical square root in $\mathbf{Q}_2$ would have to be in $\mathbf{Z}_2$: if $\alpha^2 = 3$ in $\mathbf{Q}_2$ then $|\alpha|_2^2 = |3|_2 = 1$, so $|\alpha|_2 = 1$, and thus $\alpha \in \mathbf{Z}_2^\times \subset \mathbf{Z}_2$.

Example 2.9. For each integer k between 0 and $p-1$, $k^p \equiv k \bmod p$. Letting $f(X) = X^p - X$, we have $f(k) \equiv 0 \bmod p$ and $f'(k) = pk^{p-1} - 1 \equiv -1 \not\equiv 0 \bmod p$. Hensel's lemma implies that there is a unique $\omega_k \in \mathbf{Z}_p$ such that $\omega_k^p = \omega_k$ and $\omega_k \equiv k \bmod p$. For instance, $\omega_0 = 0$ and $\omega_1 = 1$. When $p > 2$, $\omega_{p-1} = -1$. Other ω_k for $p > 2$ are more interesting. For $p = 5$, ω_k is a root of $X^5 - X = X(X^4 - 1) = X(X-1)(X+1)(X^2+1)$. Thus ω_2 and ω_3 are square roots of -1 in $\mathbf{Z}_5$:

$$\omega_2 = 2 + 5 + 2 \cdot 5^2 + 5^3 + 3 \cdot 5^4 + 4 \cdot 5^5 + 2 \cdot 5^6 + 3 \cdot 5^7 + \cdots,$$
$$\omega_3 = 3 + 3 \cdot 5 + 2 \cdot 5^2 + 3 \cdot 5^3 + 5^4 + 2 \cdot 5^6 + 5^7 + \cdots.$$

The numbers ω_k for $0 \le k \le p-1$ are distinct since they are already distinct when reduced mod p, so $X^p - X = X(X^{p-1} - 1)$ splits completely in $\mathbf{Z}_p[X]$. Its roots in $\mathbf{Z}_p$ are 0 and $(p-1)$th roots of unity. The number ω_k is called the *Teichmuller representative* for k.

Example 2.10. As an application to solving equations in $\mathbf{Q}_p$, not just $\mathbf{Z}_p$, for $p \ne 2$ and $a \in \mathbf{Q}_p$ we'll show the equation $y^2 = x^4 + a$ has a p-adic solution (x, y) whenever $|x|_p$ is sufficiently large compared to $|a|_p$.

Rewrite the equation as $y^2 = x^4(1 + a/x^4)$. Since the first term x^4 is a square for all x, it suffices to find a condition on x that implies $1 + a/x^4$ is a square in $\mathbf{Q}_p$. When $p \ne 2$, such a condition is $|a/x^4|_p < 1$, *i.e.*, $|x|_p > |a|_p^{1/4}$: for such x, $a/x^4 \in p\mathbf{Z}_p$, so $1 + a/x^4 \equiv 1 \bmod p\mathbf{Z}_p$, and therefore $1 + a/x^4$ is a square in $\mathbf{Z}_p^\times$ by Examples 2.7 or 2.8.

We'll extend this to the case $p = 2$ in Example 4.5.

3. Roots of unity in $\mathbf{Q}_p$ via Hensel's Lemma

Hensel's lemma is often considered to be a method of finding roots to polynomials, but that is just one aspect: the existence of a root. There is also a uniqueness part to Hensel's lemma: it tells us there is a unique root within a certain distance of an approximate root. We'll use the uniqueness to find all of the roots of unity in $\mathbf{Q}_p$.

Theorem 3.1. *The roots of unity in $\mathbf{Q}_p$ are the $(p-1)$th roots of unity for p odd and ± 1 for $p = 2$.*

Proof. If $x^n = 1$ in $\mathbf{Q}_p$ then $|x|_p^n = 1$, so $|x|_p = 1$. This means every root of unity in $\mathbf{Q}_p$ lies in $\mathbf{Z}_p^\times$. Therefore we work in $\mathbf{Z}_p^\times$ right from the start.

First let's consider roots of unity of order relatively prime to p. Assume ζ_1 and ζ_2 are roots of unity in $\mathbf{Z}_p^\times$ with order prime to p. Letting m be the product of the orders of these roots of unity, they are both roots of $f(X) = X^m - 1$ and m is prime to p. Since $|f'(\zeta_i)|_p = |m\zeta_i^{m-1}|_p = 1$, the uniqueness aspect of Hensel's lemma implies that the only root α of $X^m - 1$ satisfying $|\alpha - \zeta_1|_p < 1$ is ζ_1. So if $\zeta_2 \equiv \zeta_1 \bmod p\mathbf{Z}_p$ then $\zeta_2 = \zeta_1$: distinct roots of unity in $\mathbf{Z}_p^\times$ having order prime to p must be incongruent mod p. In Example 2.9 we found in each nonzero coset mod $p\mathbf{Z}_p$ a root of $X^{p-1} - 1$, and $p-1$ is prime to p. Therefore each congruence class mod $p\mathbf{Z}_p$ contains a $(p-1)$th root of unity, so the only roots of unity of order prime to p in $\mathbf{Q}_p$ are the roots of $X^{p-1} - 1$.

Now we consider roots of unity of p-power order. We will show the only pth root of unity in $\mathbf{Z}_p^\times$ is 1 for odd p and the only 4th roots of unity in $\mathbf{Z}_2^\times$ are ± 1. This implies the only pth power roots of unity in $\mathbf{Z}_p^\times$ are 1 for odd p and ± 1 for $p = 2$. (For instance, if there were a nontrivial p-th power root of unity in $\mathbf{Q}_p$ for $p \neq 2$ then there would be a root of unity in $\mathbf{Q}_p$ of order p, but we're going to show there aren't any of those.) This part of the proof will not use Hensel's lemma.

We first consider odd p and suppose $\zeta^p = 1$ in $\mathbf{Z}_p^\times$ with $\zeta \neq 1$. Since $\zeta^p \equiv \zeta \bmod p\mathbf{Z}_p$, we have $\zeta \equiv 1 \bmod p\mathbf{Z}_p$. Therefore $\zeta = 1 + py$ with $y \in \mathbf{Z}_p$. Since ζ is not 1, ζ is a root of $(X^p - 1)/(X - 1) = 1 + X + X^2 + \cdots + X^{p-1}$. For all integers $k \geq 0$, $\zeta^k = (1 + py)^k \equiv 1 + kpy \bmod p^2\mathbf{Z}_p$ by the binomial theorem, so

$$\begin{aligned} 0 &= 1 + \zeta + \zeta^2 + \cdots + \zeta^{p-1} \\ &= \sum_{k=0}^{p-1} \zeta^k \\ &\equiv \sum_{k=0}^{p-1} (1 + kpy) \bmod p^2\mathbf{Z}_p \\ &\equiv p + \frac{p(p-1)}{2} py \bmod p^2\mathbf{Z}_p. \end{aligned} \tag{3.1}$$

Since p is odd, $(p-1)/2 \in \mathbf{Z}$, so (3.1) implies $0 \equiv p \bmod p^2$, which is a contradiction. Therefore there is no pth root of unity in $\mathbf{Q}_p$ other than 1.

Now we turn to $p = 2$. We want to show the only 4th roots of unity in $\mathbf{Z}_2^\times$ are ± 1. If $\zeta \in \mathbf{Z}_2^\times$ is a 4th root of unity and $\zeta \neq \pm 1$ then $\zeta^2 = -1$, so $\zeta^2 \equiv -1 \bmod 4\mathbf{Z}_2$. However,

$$\zeta \in \mathbf{Z}_2^\times \Longrightarrow \zeta \equiv 1 \text{ or } 3 \bmod 4\mathbf{Z}_2 \Longrightarrow \zeta^2 \equiv 1 \bmod 4\mathbf{Z}_2$$

and $1 \not\equiv -1 \bmod 4\mathbf{Z}_2$. Thus there is no 4th root of unity in $\mathbf{Q}_2$ besides ± 1.

For prime p, a root of unity is a (unique) product of a root of unity of p-power order and a root of unity of order prime to p, so the only roots of unity in $\mathbf{Q}_p$ are the roots of $X^{p-1} - 1$ for $p \neq 2$ and ± 1 for $p = 2$. $\square$

4. A stronger version of Hensel's Lemma

The hypotheses of Theorem 2.1 are $f(a) \equiv 0 \bmod p$ and $f'(a) \not\equiv 0 \bmod p$. This means $a \bmod p$ is a *simple root* of $f(X) \bmod p$. We will now discuss a more general version of Hensel's lemma than Theorem 2.1. It can be applied to cases where $a \bmod p$ is a multiple root of $f(X) \bmod p$: $f(a) \equiv 0 \bmod p$ and $f'(a) \equiv 0 \bmod p$. This will allow us to describe squares in $\mathbf{Z}_2^\times$ and, more generally, pth powers in $\mathbf{Z}_p^\times$.

Theorem 4.1 (Hensel's lemma)**.** *Let $f(X) \in \mathbf{Z}_p[X]$ and $a \in \mathbf{Z}_p$ satisfy*

$$|f(a)|_p < |f'(a)|_p^2.$$

There is a unique $\alpha \in \mathbf{Z}_p$ such that $f(\alpha) = 0$ in $\mathbf{Z}_p$ and $|\alpha - a|_p < |f'(a)|_p$. Moreover,

(1) $|\alpha - a|_p = |f(a)/f'(a)|_p < |f'(a)|_p$,
(2) $|f'(\alpha)|_p = |f'(a)|_p$.

Since $f'(a) \in \mathbf{Z}_p$, $|f'(a)|_p \leq 1$. If $|f'(a)|_p = 1$ then the hypotheses of Theorem 4.1 reduce to those of Theorem 2.1: saying $|f(a)|_p < 1$ and $|f'(a)|_p = 1$ means $f(a) \equiv 0 \bmod p$ and $f'(a) \not\equiv 0 \bmod p$. Theorem 4.1 actually goes beyond the conclusions of Theorem 2.1 when the hypotheses of Theorem 2.1 hold, since we learn in Theorem 4.1 exactly how far away the root α is from the approximate root a. But the main point of Theorem 4.1 is that it allows for the possibility that $|f'(a)|_p < 1$, which isn't covered by Theorem 2.1 at all.

We will prove Theorem 4.1 by two methods, in Sections 5 and 6. Here are some applications where the polynomial has a multiple root mod p.

Example 4.2. Let $f(X) = X^4 - 7X^3 + 2X^2 + 2X + 1$. Then $f(X) \equiv (X+1)^2(X^2+1) \bmod 3$ and we notice 2 mod 3 is a double root. Since $|f(2)|_3 = 1/27$ and $|f'(2)|_3 = |-42|_3 = 1/3$, the condition $|f(2)|_3 < |f'(2)|_3^2$ holds, so there is a unique root α of $f(X)$ in $\mathbf{Z}_3$ such that $|\alpha - 2|_3 < 1/3$, *i.e.*, $\alpha \equiv 2 \bmod 9$.

In fact there are two roots of $f(X)$ in $\mathbf{Z}_3$:

$$2 + 3 + 2 \cdot 3^2 + 2 \cdot 3^3 + 2 \cdot 3^4 + 2 \cdot 3^6 + \cdots \text{ and } 2 + 3^2 + 3^4 + 2 \cdot 3^5 + 2 \cdot 3^6 + \cdots .$$

The second root reduces to 2 mod 9, and is α. The first root reduces to 5 mod 9, and its existence can be verified by checking $|f(5)|_3 = 1/27 < |f'(5)|_3^2 = 1/9$.

Example 4.3. Let $f(X) = X^3 - 10$ and $g(X) = X^3 - 5$. We have $f(X) \equiv (X-1)^3 \bmod 3$ and $g(X) \equiv (X-2)^3 \bmod 3$: 1 is an approximate 3-adic root of $f(X)$ and 2 is an approximate 3-adic root of $g(X)$. We want to see if they can be refined to genuine 3-adic roots. The basic form of Hensel's lemma in Theorem 2.1 can't be used since the polynomials do not have simple roots mod 3. Instead we will try to use the stronger form of Hensel's lemma in Theorem 4.1.

Since $|f(1)|_3 = 1/9$ and $|f'(1)|_3 = 1/3$, we don't have $|f(1)|_3 < |f'(1)|_3^2$, so Theorem 4.1 can't be used on $f(X)$ with $a = 1$. However, $|f(4)|_3 = 1/27$ and $|f'(4)|_3 = 1/3$, so we can use Theorem 4.1 on $f(X)$ with $a = 4$: there is a unique root α of $X^3 - 10$ in $\mathbf{Z}_3$ satisfying $|\alpha - 4|_3 < 1/3$, so $\alpha \equiv 4 \bmod 9$. The expansion of α begins as $1 + 3 + 3^2 + 2 \cdot 3^6 + 3^7 + \cdots$.

Turning to $g(X)$, we have $|g(2)|_3 = 1/3$ and $|g'(2)|_3 = 1/9$, so we can't apply Theorem 4.1 with $a = 2$. In fact there is no root of $g(X)$ in $\mathbf{Q}_3$. If there were a root α in $\mathbf{Q}_3$ then $\alpha^3 = 5$, so $|\alpha|_3 = 1$, and thus $\alpha \in \mathbf{Z}_3$. Then $\alpha^3 \equiv 5 \bmod 3^n$, so 5 would be a cube modulo every power of 3. But 5 is not a cube mod 9 (the only cubes mod 9 are 0, 1, and 8). Therefore the mod 3 root of $X^3 - 5$ does not lift to a 3-adic root of $X^3 - 5$.

Theorem 4.4. *If $u \in \mathbf{Z}_2^\times$ then u is a square in $\mathbf{Q}_2$ if and only if $u \equiv 1 \bmod 8\mathbf{Z}_2$.*

Proof. If $u = v^2$ in $\mathbf{Q}_2$ then $1 = |v|_2^2$, so $v \in \mathbf{Z}_2^\times$. In $\mathbf{Z}_2/8\mathbf{Z}_2 \cong \mathbf{Z}/8\mathbf{Z}$, the units are 1, 3, 5, and 7, whose squares are all congruent to 1 mod 8, so $u = v^2 \equiv 1 \bmod 8\mathbf{Z}_2$. To show, conversely, that all $u \in \mathbf{Z}_2^\times$ satisfying $u \equiv 1 \bmod 8\mathbf{Z}_2$ are squares in $\mathbf{Z}_2^\times$, let $f(X) = X^2 - u$ and use $a = 1$ in Theorem 4.1. We have $|f(1)|_2 = |1 - u|_2 \le 1/8$ and $|f'(1)|_2 = |2|_2 = 1/2$, so $|f(1)|_2 < |f'(1)|_2^2$. Therefore $X^2 - u$ has a root in $\mathbf{Z}_2$, so u is a square in $\mathbf{Z}_2$. ☐

Example 4.5. We saw in Example 2.10 that for $p \neq 2$ and $a \in \mathbf{Q}_p$, $y^2 = x^4 + a$ has a solution in $\mathbf{Q}_p$ when $|x|_p^4 > |a|_p$. With Theorem 4.4 we'll get an analogue of this for $p = 2$.

Rewriting the equation as $y^2 = x^4(1 + a/x^4)$, we want $1 + a/x^4$ to be a square in $\mathbf{Q}_2$, and by Theorem 4.4 such a condition is $|a/x^4|_2 \le 1/8$, since then $1 + a/x^4 \equiv 1 \bmod 8\mathbf{Z}_2$.

The same method, using the stronger version of Hensel's lemma, shows for prime p, $a \in \mathbf{Q}_p$, and $n \mid m$ in $\mathbf{Z}^+$ that the equation $y^n = x^m + a$ has a p-adic solution (x, y) whenever $|x|_p$ is sufficiently large in terms of $|a|_p$, m, and n. Details are left to the reader.

Theorem 4.6. *If $p \neq 2$ and $u \in \mathbf{Z}_p^\times$, then u is a pth power in $\mathbf{Q}_p$ if and only if u is a pth power modulo p^2.*

Theorem 4.6 is false for $p = 2$: the criterion for an element of $\mathbf{Z}_2^\times$ to be a 2-adic square needs modulus 2^3, not modulus 2^2. For instance, 5 is a square mod 4 but 5 is not a 2-adic square since $5 \not\equiv 1 \bmod 8$. Theorem 4.6 explains what we found in Example 4.3: 10 is a 3-adic cube and 5 is not, since 10 mod 9 is a cube and 5 mod 9 is not a cube.

Proof. If $u = v^p$ in $\mathbf{Q}_p$ then $1 = |v|_p^p$, so $v \in \mathbf{Z}_p^\times$: we only need to look for pth roots of u in $\mathbf{Z}_p^\times$. Let $f(X) = X^p - u$. In order to use Theorem 4.1 on $f(X)$, we seek an $a \in \mathbf{Z}_p^\times$ such that $|f(a)|_p < |f'(a)|_p^2$. This means $|a^p - u|_p < |pa^{p-1}|_p^2 = 1/p^2$, or equivalently $a^p \equiv u \bmod p^3$.

So provided u is a pth power modulo p^3, Theorem 4.1 tells us that $X^p - u$ has a root in $\mathbf{Z}_p$, so u is a pth power. The criterion in the theorem, however, has modulus p^2 rather p^3. We need to do some work to bootstrap an approximate pth root from modulus p^2 to modulus p^3 in order for Theorem 4.1 to apply.

Suppose $u \equiv a^p \bmod p^2$ for some $a \in \mathbf{Z}_p$. Then $a \in \mathbf{Z}_p^\times$ and $u/a^p \equiv 1 \bmod p^2$. Write $u/a^p \equiv 1 + p^2c \bmod p^3$, where $0 \leq c \leq p-1$. By the binomial theorem,

$$(1+pc)^p = 1 + p(pc) + \sum_{k=2}^{p} \binom{p}{k}(pc)^k.$$

The terms for $k \geq 3$ are obviously divisible by p^3 and the term at $k = 2$ is $\binom{p}{2}(pc)^2 = \frac{p-1}{2}p^3c^2$, which is also divisible by p^3 since $p > 2$.

Therefore $(1+pc)^p \equiv 1 + p^2c \bmod p^3$, so $u/a^p \equiv (1+pc)^p \bmod p^3$. Now we can write

$$\frac{u}{a^p(1+pc)^p} \equiv 1 \bmod p^3.$$

From Theorem 4.1, a p-adic integer that is congruent to $1 \bmod p^3$ is a pth power (see the first paragraph again). Thus $u/(a^p(1+pc)^p)$ is a pth power, so u is a pth power. □

Remark 4.7. Hensel's lemma in Theorem 4.1 says a root mod p lifts uniquely to a root in $\mathbf{Z}_p$ under weaker conditions than Hensel's lemma in Theorem 2.1, but it does not guarantee a unique lift mod p^n, unlike in Theorem 2.1 (Remark 2.3). Consider Example 4.3: in $\mathbf{Z}_3$, $X^3 = 10$ has one solution $1 + 3 + 3^2 + 2 \cdot 3^6 + \cdots$, but for $n \geq 2$, $X^3 \equiv 10 \bmod 3^n$ has 3 solutions, not 1. One solution mod 3^n lifts to modulus 3^{n+1} and two don't. (For instance, solutions mod 27 are 4, 13, and 22 and only 13 lifts to solutions mod 81, in fact to 13, 40, and 67 mod 27.) This is *consistent* with a unique root in $\mathbf{Z}_3$ since the 3 solutions mod 3^n are congruent modulo 3^{n-1} and thus are close: they have a common limit in $\mathbf{Z}_3$ as $n \to \infty$.

5. First Proof of Theorem 4.1: Newton's Method

Our first proof of Theorem 4.1 will use Newton's method and is a modification of [4, Chap. II, §2, Prop. 2].

Proof. As in Newton's method from real analysis, define a sequence $\{a_n\}$ in $\mathbf{Q}_p$ by $a_1 = a$ and

$$a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)} \tag{5.1}$$

for $n \geq 1$. Set $t = |f(a)/f'(a)^2|_p < 1$. We will show by induction on n that

(i) $|a_n|_p \leq 1$, *i.e.*, $a_n \in \mathbf{Z}_p$,
(ii) $|f'(a_n)|_p = |f'(a_1)|_p$,
(iii) $|f(a_n)|_p \leq |f'(a_1)|_p^2 t^{2^{n-1}}$.

For $n = 1$ these conditions are all clear. Note in particular that we have equality in (iii) and $f'(a_1) \neq 0$ since $|f(a_1)|_p < |f'(a_1)|_p^2$.

For the inductive step, we need two polynomial identities. The first one, for $F(X) \in \mathbf{Z}_p[X]$, is

$$F(X+Y) = F(X) + F'(X)Y + g(X,Y)Y^2 \tag{5.2}$$

for some $g(X,Y) \in \mathbf{Z}_p[X,Y]$. This is (2.1) from the proof of Theorem 2.1. So

$$x, y \in \mathbf{Z}_p \Longrightarrow \boxed{F(x+y) = F(x) + F'(x)y + zy^2, \text{ where } z \in \mathbf{Z}_p} \tag{5.3}$$

with $z = g(x, y)$. The second polynomial identity we need is that for $F(X) \in \mathbf{Z}_p[X]$,

$$F(X) - F(Y) = (X - Y)G(X, Y)$$

for some $G(X, Y) \in \mathbf{Z}_p[X, Y]$. This comes from $X - Y$ being a factor of $X^i - Y^i$ for all $i \geq 1$. Writing $F(X) = \sum_{i=0}^{m} b_i X^i$,

$$F(X) - F(Y) = \sum_{i=1}^{m} b_i (X^i - Y^i)$$

and we can factor $X - Y$ out of each term on the right. For x and y in $\mathbf{Z}_p$, $G(x, y) \in \mathbf{Z}_p$, so

$$x, y \in \mathbf{Z}_p \Longrightarrow \boxed{|F(x) - F(y)|_p = |x - y|_p |G(x, y)|_p \leq |x - y|_p.} \tag{5.4}$$

Assume (i), (ii), and (iii) are true for n. To prove (i) for $n + 1$, first note a_{n+1} is defined since $f'(a_n) \neq 0$ by (ii). To prove (i) it suffices to show $|f(a_n)/f'(a_n)|_p \leq 1$. Using (ii) and (iii) for n, we have $|f(a_n)/f'(a_n)|_p = |f(a_n)/f'(a_1)|_p \leq |f'(a_1)|_p t^{2^{n-1}} \leq 1$.

To prove (ii) for $n+1$, (iii) for n implies $|f(a_n)|_p < |f'(a_1)|_p^2$ since $t < 1$ (and $|f'(a_1)|_p \neq 0$), so by (5.4) with $F(X) = f'(X)$,

$$|f'(a_{n+1}) - f'(a_n)|_p \leq |a_{n+1} - a_n|_p = \frac{|f(a_n)|_p}{|f'(a_n)|_p} = \frac{|f(a_n)|_p}{|f'(a_1)|_p} < |f'(a_1)|_p = |f'(a_n)|_p,$$

so $|f'(a_{n+1})|_p = |f'(a_n)|_p = |f'(a_1)|_p$.

To prove (iii) for $n + 1$, we use (5.3) with $F(X) = f(X)$, $x = a_n$ and $y = -f(a_n)/f'(a_n)$:

$$f(a_{n+1}) = f(x + y) = f(a_n) + f'(a_n)\left(-\frac{f(a_n)}{f'(a_n)}\right) + z\left(\frac{f(a_n)}{f'(a_n)}\right)^2 = z\left(\frac{f(a_n)}{f'(a_n)}\right)^2,$$

where $z \in \mathbf{Z}_p$. Thus, by (iii) for n,

$$|f(a_{n+1})|_p \leq \left|\frac{f(a_n)}{f'(a_n)}\right|_p^2 = \frac{|f(a_n)|_p^2}{|f'(a_1)|_p^2} \leq \frac{|f'(a_1)|_p^4 t^{2^n}}{|f'(a_1)|_p^2} = |f'(a_1)|_p^2 t^{2^n}.$$

This completes the induction.

Now we show $\{a_n\}$ is Cauchy in $\mathbf{Q}_p$. From the recursive definition of this sequence,

$$|a_{n+1} - a_n|_p = \left|\frac{f(a_n)}{f'(a_n)}\right|_p = \frac{|f(a_n)|_p}{|f'(a_1)|_p} \leq |f'(a_1)|_p t^{2^{n-1}}, \tag{5.5}$$

where we used (ii) and (iii). Thus $\{a_n\}$ is Cauchy. Let α be its limit, so $|\alpha|_p \leq 1$ by (i), *i.e.*, $\alpha \in \mathbf{Z}_p$. Letting $n \to \infty$ in (ii) and (iii) we get $|f'(\alpha)|_p = |f'(a_1)|_p = |f'(a)|_p$ and $f(\alpha) = 0$.

To show $|\alpha - a|_p = |f(a)/f'(a)|_p$, it's true if $f(a) = 0$ since then $a_n = a_1 = a$ for all n, so $\alpha = a$.[2] If $f(a) \neq 0$ then we will show $|a_n - a|_p = |f(a)/f'(a)|_p$ for all $n \geq 2$ and let $n \to \infty$. When $n = 2$ use the definition of a_2 in terms of $a_1 = a$. For all $n \geq 2$, by (5.5)

$$|a_{n+1} - a_n|_p \leq |f'(a_1)|_p t^{2^{n-1}} \leq |f'(a_1)|_p t^2 < |f'(a_1)|_p t = |f'(a)|_p t = \left|\frac{f(a)}{f'(a)}\right|_p, \tag{5.6}$$

where $t \in (0, 1)$ since $f(a) \neq 0$. If $|a_n - a|_p = |f(a)/f'(a)|_p$ then $|a_{n+1} - a_n|_p < |a_n - a|_p$ by (5.6), so $|a_{n+1} - a|_p = |(a_{n+1} - a_n) + (a_n - a)|_p = |a_n - a|_p = |f(a)/f'(a)|_p$.

The last thing to do is show α is the only root of $f(X)$ in the ball $\{x \in \mathbf{Z}_p : |x - a|_p < |f'(a)|_p\}$. This will not use anything about Newton's method. Assume $f(\beta) = 0$ and

[2]The case $f(a) = 0$ was not covered in an earlier draft and this omission was found when the proof here was formalized using the Lean theorem prover [5, p. 8].

$|\beta - a|_p < |f'(a)|_p$. Since $|\alpha - a|_p < |f'(a)|_p$ we have $|\beta - \alpha|_p < |f'(a)|_p$. Write $\beta = \alpha + h$, so $h \in \mathbf{Z}_p$. Then by (5.3) with $F(X) = f(X)$,

$$0 = f(\beta) = f(\alpha + h) = f(\alpha) + f'(\alpha)h + zh^2 = f'(\alpha)h + zh^2$$

for some $z \in \mathbf{Z}_p$. If $h \neq 0$ then $f'(\alpha) = -zh$, so $|f'(\alpha)|_p \leq |h|_p = |\beta - \alpha|_p < |f'(a)|_p$. But $|f'(\alpha)|_p = |f'(a)|_p$, so we have a contradiction. Thus $h = 0$, so $\beta = \alpha$. $\square$

Before we give a second proof of Theorem 4.1, it's worth noting that the a_n's converge to α very rapidly. From the inequality $|a_{n+1} - a_n|_p \leq |f'(a_1)|_p t^{2^{n-1}}$ for all $n \geq 1$ we obtain by the strong triangle inequality $|a_m - a_n|_p \leq |f'(a_1)|_p t^{2^{n-1}}$ for all $m > n$. Letting $m \to \infty$,

$$\boxed{|\alpha - a_n|_p \leq |f'(a_1)|_p t^{2^{n-1}} = |f'(a)|_p t^{2^{n-1}} = |f'(a)|_p \left|\frac{f(a)}{f'(a)^2}\right|_p^{2^{n-1}}.} \tag{5.7}$$

Since $|f(a)/f'(a)^2|_p < 1$, the exponent 2^{n-1} tells us that the number of initial p-adic digits in a_n that agree with those in the limit α is at least doubling at each step.

Example 5.1. Let $f(X) = X^2 - 7$ in $\mathbf{Q}_3[X]$. It has two roots in $\mathbf{Z}_3$:

$$r = 1 + 3 + 3^2 + 2 \cdot 3^4 + 2 \cdot 3^7 + 3^8 + 3^9 + \cdots,$$
$$s = 2 + 3 + 3^2 + 2 \cdot 3^3 + 2 \cdot 3^5 + 2 \cdot 3^6 + 3^8 + 3^9 + \cdots.$$

Starting with $a_1 = 1$, for which $|f(a_1)/f'(a_1)^2|_3 = 1/3$, Newton's recursion (5.1) has limit α where $|\alpha - a_1|_3 < |f'(a_1)|_3 = 1$, so $\alpha \equiv a_1 \equiv 1 \bmod 3$. Thus $\alpha = r$. For example,

$$a_4 = \frac{977}{368} = 1 + 3 + 3^2 + 2 \cdot 3^4 + 2 \cdot 3^7 + 3^9 + 3^{10} + \cdots,$$

which has the same 3-adic digits as r up through terms including 3^7 (the first 8 digits). The estimate in (5.7) says $|r - a_n|_3 \leq |f'(a_1)|_3 (1/3)^{2^{n-1}} = (1/3)^{2^{n-1}}$ for all n. Using a computer, this inequality is an equality for $1 \leq n \leq 10$.

Example 5.2. Let $f(X) = X^2 - 17$ in $\mathbf{Q}_2[X]$. It has two roots in $\mathbf{Z}_2$:

$$r = 1 + 2^3 + 2^5 + 2^6 + 2^7 + 2^9 + \cdots,$$
$$s = 1 + 2 + 2^2 + 2^4 + 2^8 + \cdots.$$

Using Newton's recursion (5.1) for $f(X)$ with initial seed $a \in \mathbf{Z}_2^\times$, we need $|a^2 - 17|_2 < |2a|_2^2$, which is the same as $a^2 \equiv 17 \bmod 8$, and this congruence works for all $a \in \mathbf{Z}_2^\times$. Therefore (5.1) with $a_1 \in \mathbf{Z}_2^\times$ converges to r or s. Since $|f'(a)|_2 = 1/2$ for $a \in \mathbf{Z}_2^\times$, (5.1) with $a_1 = a$ has a limit α satisfying $|\alpha - a|_2 < |f'(a)|_2 = 1/2$, so $\alpha \equiv a \bmod 4$: if $a \equiv 1 \bmod 4$ then $\alpha = r$, and if $a \equiv 3 \bmod 4$ then $\alpha = s$. By (5.7), $|\alpha - a_n|_2 \leq |f'(a)|_2 (|f(a)/f'(a)^2|_2)^{2^{n-1}} = (1/2)(4|a^2 - 17|_2)^{2^{n-1}}$. For a few choices of a, this inequality is an equality for $1 \leq n \leq 10$:

- When $a = 1$, $|r - a_n|_2 = (1/2)^{2^n + 1}$ for $1 \leq n \leq 10$.
- When $a = 3$, $|s - a_n|_2 = (1/2)^{2^{n-1}+1}$ for $1 \leq n \leq 10$.
- When $a = 5$, $|r - a_n|_2 = (1/2)^{2^{n-1}+1}$ for $1 \leq n \leq 10$.

Example 5.3. Let's solve $X^2 - 1 = 0$ in $\mathbf{Q}_3$. This might seem silly, since we know the solutions are ± 1, but let's check how an initial approximation affects the 3-adic limit. We use $a_1 = 2$. (If we used $a_1 = 1$ then $a_n = 1$ for all n, which is not interesting.) When $f(X) = X^2 - 1$ the recursion for Newton's method is

$$a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)} = a_n - \frac{a_n^2 - 1}{2a_n} = \frac{a_n^2 + 1}{2a_n} = \frac{1}{2}\left(a_n + \frac{1}{a_n}\right).$$

Since $|f(2)|_3 = 1/3 < |f'(2)|_3^2$, Newton's recursion with $a_1 = 2$ converges in $\mathbf{Q}_3$. What is the limit? Since $a_1 \equiv -1 \bmod 3$, we have $a_n \equiv -1 \bmod 3$ for all n by induction: if $a_n \equiv -1 \bmod 3$ then $a_{n+1} = (1/2)(a_n + 1/a_n) \equiv (1/2)(-1 + 1/(-1)) \equiv (1/2)(-2) \equiv -1 \bmod 3$. Thus $\lim_{n\to\infty} a_n = -1$ in $\mathbf{Q}_3$. What makes this interesting is that in $\mathbf{R}$ all a_n are positive so the sequence of rational numbers $\{a_n\}$ converges to 1 in $\mathbf{R}$ and to -1 in $\mathbf{Q}_3$. The table below illustrates the rapid convergence in $\mathbf{R}$ and $\mathbf{Q}_3$ (the 3-adic expansion of -1 is $\overline{2} = 2222\ldots$).

n	a_n	Decimal approx.	3-adic approx.
1	2	2.00000000	20000000
2	5/4	1.25000000	22020202
3	41/40	1.02500000	22220222
4	3281/3280	1.00030487	22222222

Theorem 4.1 has Theorem 2.1 as the special case when $|f'(a)|_p = 1$. Using a change of variables, we will show Theorem 4.1 follows from Theorem 2.1, so the basic and strong versions of Hensel's lemma are in fact equivalent!

Theorem 5.4. *Theorem 2.1 implies Theorem 4.1.*

Proof. Suppose $f(X) \in \mathbf{Z}_p[X]$ and $a \in \mathbf{Z}_p$ satisfies $|f(a)|_p < |f'(a)|_p^2$. We will use Theorem 2.1 to show $f(X)$ has a unique root $\alpha \in \mathbf{Z}_p$ such that $|\alpha - a|_p < |f'(a)|_p$, and in fact $|\alpha - a|_p = |f(a)/f'(a)|_p$ and $|f'(\alpha)|_p = |f'(a)|_p$.

Since $|f(a)|_p < |f'(a)|_p^2$, set $b = f(a)/f'(a)^2$, so $f(a) = f'(a)^2 b$ and $|b|_p < 1$. A root $\alpha \in \mathbf{Z}_p$ of $f(X)$ where $|\alpha - a|_p < |f'(a)|_p$ is a root of the form $a + f'(a)s$ with $|s|_p < 1$. We thus want to show $f(a + f'(a)s) = 0$ for a unique $s \in \mathbf{Z}_p$ with $|s|_p < 1$ and also show that in fact $|s|_p = |f(a)/f'(a)^2|$. By the polynomial identity (5.2) with $F(X) = f(X)$, there is a polynomial $g(X, Y) \in \mathbf{Z}_p[X, Y]$ such that

$$f(X + Y) = f(X) + f'(X)Y + g(X, Y)Y^2,$$

so for all $s \in \mathbf{Z}_p$,

$$\begin{aligned} f(a + f'(a)s) &= f(a) + f'(a)(f'(a)s) + g(a, f'(a)s)(f'(a)s)^2 \\ &= f'(a)^2 b + f'(a)^2 s + g(a, f'(a)s)f'(a)^2 s^2 \\ &= f'(a)^2(b + s + g(a, f'(a)s)s^2). \end{aligned}$$

Set $h(X) = b + X + g(a, f'(a)X)X^2 \in \mathbf{Z}_p[X]$. Since $h(X)$ has constant term $b = f(a)/f'(a)^2$ and linear coefficient 1, $|h(0)|_p = |b|_p < 1$ and $|h'(0)|_p = |1|_p = 1$. Theorem 2.1 implies there is a unique $\beta \in \mathbf{Z}_p$ such that $h(\beta) = 0$ and $|\beta|_p < 1$, so $\alpha := a + f'(a)\beta$ is the unique root of $f(X)$ in $\mathbf{Z}_p$ such that $|\alpha - a|_p < |f'(a)|_p$.

To show $|\alpha - a|_p = |f(a)/f'(a)|_p$, rewrite this as $|\beta|_p = |f(a)/f'(a)^2|_p$. If $\beta = 0$ then $\alpha = a$, so $f(a) = 0$ and thus $|f(a)/f'(a)^2|_p = 0 = |\beta|_p$. If $\beta \neq 0$ then from

$$0 = h(\beta) = b + \beta + g(a, f'(a)\beta)\beta^2$$

we get $|b + \beta|_p = |g(a, f'(a)\beta)\beta^2|_p \leq |\beta|_p^2 < |\beta|_p$, so $|\beta|_p = |b|_p = |f(a)/f'(a)^2|_p$.

That $|f'(\alpha)|_p = |f(a)|_p$ follows from $|\alpha - a|_p < |f'(a)|_p$: by (5.4), $|f'(\alpha) - f'(a)|_p \leq |\alpha - a|_p < |f'(a)|_p$, so $|f'(\alpha)|_p = |f'(a)|_p$. □

Example 5.5. Let $f(X) = X^3 - 10$ with approximate root $a = 4$ in $\mathbf{Z}_3$. We can't apply Theorem 2.1 directly to $f(X)$ to show there is a solution to $f(\alpha) = 0$ in $\mathbf{Z}_3$ with α close to 4 since $f'(4) = 48 \equiv 0 \bmod 3$. From the proof of Theorem 5.4 we compute

$$f(X+Y) = (X+Y)^3 - 10 = X^3 + 3X^2Y + 3XY^2 + Y^3 - 10 = f(X) + f'(X)Y + (3X+Y)Y^2.$$

Then

$$f(4+48X) = f(4)+f'(4)48X+(12+48X)(48X)^2 = 54+48^2X+(12+48X)(48X)^2 = 48^2h(X)$$

where

$$h(X) = \frac{54}{48^2} + X + (12+48X)X^2 = \frac{3}{128} + X + 12X^2 + 48X^3.$$

Since $h(0) \equiv 0 \bmod 3$ and $h'(0) = 1 \not\equiv 0 \bmod 3$, $h(X)$ has a unique root $\beta \in 3\mathbf{Z}_3$ and that gives us a root $\alpha = 4 + 48\beta$ of $f(X)$.

By Theorem 2.1, each solution of $h(x) \equiv 0 \bmod 3^n$ with $x \equiv 0 \bmod 3$ lifts uniquely to a solution of $h(x) \equiv 0 \bmod 3^{n+1}$ but such uniqueness of lifting at each "finite level" is not true for solving $f(x) \equiv 0 \bmod 3^n$; see Remark 4.7. This different behavior is compatible because of the 48 (a multiple of 3) appearing on both sides of $f(4 + 48X) = 48^2h(X)$.

For example, suppose we want to solve $f(t) \equiv 0 \bmod 81$ in $\mathbf{Z}_3$ with $t \equiv 4 \bmod 3$. Then we can write $t = 4 + 48x$ with $x \in \mathbf{Z}_3$ since 48 is divisible by 3 just once, so

$$f(4 + 48x) \equiv 0 \bmod 81 \iff 48^2h(x) \equiv 0 \bmod 81 \iff h(x) \equiv 0 \bmod 9.$$

Note the modulus decreased from 81 to 9. The unique solution of $h(x) \equiv 0 \bmod 9$ is $3 \bmod 9$, and that is the same as $t = 4 + 48x \equiv 4 + 48 \cdot 3 \equiv 13 \bmod 27$. The modulus increased from 9 to 27 since congruences mod 9 are the same as congruences mod 27 when multiplying by 48. We have shown $f(t) \equiv 0 \bmod 81$ is the same as $t \equiv 13 \bmod 27$, and there are three liftings of 13 from mod 27 to mod 81: 13, 40, and 67.

6. Second Proof of Theorem 4.1: Contraction Mappings

Newton's method produces a sequence converging to a root of $f(X)$ by iterating the function $x - f(x)/f'(x)$ with initial value $a_1 = a$ where $|f(a)|_p < |f'(a)|_p^2$. A root can also be found with a different iteration $\varphi(x) = x - f(x)/f'(a)$, where again $|f(a)|_p < |f'(a)|_p^2$. The denominator is $f'(a)$, not $f'(x)$, so it doesn't change. We will show φ is a contraction mapping on a suitable ball around a. Then the contraction mapping theorem will imply φ has a (unique) fixed point α in that ball.[3] The condition $\varphi(\alpha) = \alpha$ says $\alpha - f(\alpha)/f'(a) = \alpha$, so $f(\alpha) = 0$ and we have a root of $f(X)$. Filling in the details leads to the following second proof of Theorem 4.1. If you're not interested in a second proof, go to the next section.

Proof. For $r \in [0, 1]$ to be determined, set $\overline{B}_a(r) = \{x \in \mathbf{Q}_p : |x - a|_p \leq r\}$, so $\overline{B}_a(r) \subset \mathbf{Z}_p$. Set

$$\varphi(x) = x - \frac{f(x)}{f'(a)}.$$

We seek an r such that φ maps $\overline{B}_a(r)$ back to itself and is a contraction on that ball.

To show φ is a contraction on some ball around a, we want to estimate

$$|\varphi(x) - \varphi(y)|_p = \left| x - y - \frac{f(x) - f(y)}{f'(a)} \right|$$

for all x and y near a in order to make this $\leq \lambda|x - y|_p$ for some $\lambda < 1$.

Write $f(X)$ as a polynomial in $X - a$, say $f(X) = \sum_{i=0}^{d} b_i(X - a)^i$. Then $b_0 = f(a)$, $b_1 = f'(a)$, and $b_i \in \mathbf{Z}_p$ for $i > 1$. For all x and y in $\mathbf{Q}_p$,

$$f(x) - f(y) = \sum_{i=1}^{d} b_i((x - a)^i - (y - a)^i) = f'(a)(x - y) + \sum_{i=2}^{d} b_i((x - a)^i - (y - a)^i),$$

[3]There is an analogous method in real analysis to simplify Newton's method to the contraction mapping theorem by fixing the denominator in the recursion. See [2, Thm. 1.2, p. 164].

so

$$\varphi(x) - \varphi(y) = x - y - \frac{f(x) - f(y)}{f'(a)} = -\frac{1}{f'(a)} \sum_{i=2}^{d} b_i((x-a)^i - (y-a)^i). \tag{6.1}$$

(If $d \leq 1$ then the sum on the right is empty.) In the polynomial identity

$$X^i - Y^i = (X - Y) \sum_{j=0}^{i-1} X^{i-1-j} Y^j$$

for $i \geq 2$ set $X = x - a$ and $Y = y - a$. Then

$$\begin{aligned} |(x-a)^i - (y-a)^i|_p &= |x-y|_p \left| \sum_{j=0}^{i-1} (x-a)^{i-1-j}(y-a)^j \right|_p \\ &\leq |x-y|_p \max_{0 \leq j \leq i-1} |x-a|_p^{i-1-j} |y-a|_p^j \\ &\leq |x-y|_p \max(|x-a|_p, |y-a|_p)^{i-1} \\ &\leq |x-y|_p \max(|x-a|_p, |y-a|_p) \end{aligned}$$

when $|x-a|_p$ and $|y-a|_p$ are both at most 1. Therefore from (6.1),

$$|x-a|_p, |y-a|_p \leq 1 \Longrightarrow |\varphi(x) - \varphi(y)|_p \leq \frac{|x-y|_p \max(|x-a|_p, |y-a|_p)}{|f'(a)|_p},$$

so for $\lambda \in [0, 1)$,

$$|x-a|_p, |y-a|_p \leq \lambda |f'(a)|_p \Longrightarrow |\varphi(x) - \varphi(y)|_p \leq \lambda |x-y|_p. \tag{6.2}$$

If we can find $\lambda \in [0, 1)$, perhaps depending on a and $f(X)$, such that

$$|x-a|_p \leq \lambda |f'(a)|_p \Longrightarrow |\varphi(x) - a|_p \leq \lambda |f'(a)|_p \tag{6.3}$$

then (6.2) will tell us that φ is a contraction mapping on the closed ball around a of radius $\lambda |f'(a)|_p$. We will see that if $|f(a)|_p < |f'(a)|_p^2$ then a choice for λ is $|f(a)/f'(a)^2|_p$. In fact, we want to do more: show the condition $|f(a)|_p < |f'(a)|_p^2$ arises *naturally* by trying to make (6.3) work for some unknown λ.

For $\lambda \in (0, 1)$, when $|x-a|_p \leq \lambda |f'(a)|_p$ we have

$$|\varphi(x) - a|_p \leq \lambda |f'(a)|_p \Longleftrightarrow \left| x - a - \frac{f(x)}{f'(a)} \right|_p \leq \lambda |f'(a)|_p \Longleftrightarrow \left| \frac{f(x)}{f'(a)} \right|_p \leq \lambda |f'(a)|_p.$$

Returning to the formula $f(X) = \sum_{i=0}^{d} b_i (X-a)^i$, where $b_0 = f(a)$ and $b_1 = f'(a)$,

$$\frac{f(x)}{f'(a)} = \frac{f(a)}{f'(a)} + (x-a) + \sum_{i=2}^{d} \frac{b_i}{f'(a)} (x-a)^i. \tag{6.4}$$

When $|x-a|_p \leq \lambda |f'(a)|_p$, which is less than $|f'(a)|_p \leq 1$, we have for $i \geq 2$ that

$$\left| \frac{b_i}{f'(a)} (x-a)^i \right|_p \leq \frac{|x-a|_p^2}{|f'(a)|_p} \leq \lambda^2 |f'(a)|_p \leq \lambda |f'(a)|_p,$$

so by (6.4)

$$\left| \frac{f(x)}{f'(a)} \right|_p \leq \lambda |f'(a)|_p \Longleftrightarrow \left| \frac{f(a)}{f'(a)} \right|_p \leq \lambda |f'(a)|_p \Longleftrightarrow \left| \frac{f(a)}{f'(a)^2} \right|_p \leq \lambda.$$

To make this occur for some $\lambda < 1$ is equivalent to requiring $|f(a)/f'(a)^2|_p < 1$. Therefore if $|f(a)|_p < |f'(a)|_p^2$ and we set $\lambda = |f(a)/f'(a)^2|_p$, the mapping $\varphi(x) = x - f(x)/f'(a)$ is a contraction on the closed ball around a with radius $\lambda|f'(a)|_p = |f(a)/f'(a)|_p$ and contraction constant λ. Any closed ball in $\mathbf{Q}_p$ is complete, so the contraction mapping theorem implies that the sequence $\{a_n\}$ defined recursively by $a_1 = a$ and

$$a_{n+1} = \varphi(a_n) = a_n - \frac{f(a_n)}{f'(a)} \tag{6.5}$$

for $n \geq 1$ converges to the unique fixed point of φ in $\overline{B}_a(|f(a)/f'(a)|_p)$. By the definition of φ, a fixed point of φ is the same thing as a zero of $f(X)$, so there is a unique α in $\mathbf{Q}_p$ satisfying $f(\alpha) = 0$ and $|\alpha - a|_p \leq |f(a)/f'(a)|_p$.

To finish this proof of Theorem 4.1 we need to show $|\alpha - a|_p = |f(a)/f'(a)|_p$, α is the unique root of $f(X)$ in $\mathbf{Z}_p$ such that $|\alpha - a|_p < |f'(a)|_p$, and $|f'(\alpha)|_p = |f'(a)|_p$.

$\underline{|\alpha - a|_p = |f(a)/f'(a)|_p}$: We have $|a_2 - a|_p = |a_2 - a_1|_p = |f(a_1)/f'(a)|_p = |f(a)/f'(a)|_p$, and for all n

$$|a_{n+1} - a_n|_p = |\varphi^n(a) - \varphi^{n-1}(a)|_p = |\varphi^{n-1}(\varphi(a)) - \varphi^{n-1}(a)|_p \leq \lambda^{n-1}|a_2 - a_1|_p < \left|\frac{f(a)}{f'(a)}\right|_p.$$

Then $|a_n - a|_p = |f(a)/f'(a)|_p$ for all n by induction, so $|\alpha - a|_p = |f(a)/f'(a)|_p$ by letting $n \to \infty$.

$\underline{\alpha \text{ is the unique root of } f(X) \text{ such that } |\alpha - a|_p < |f'(a)|_p}$: This was proved at the end of the first proof of Theorem 4.1 without needing Newton's method. We won't rewrite the proof.

$\underline{|f'(\alpha)|_p = |f'(a)|_p}$: Since $a_1 = a$, of course $|f'(a_1)|_p = |f'(a)|_p$. If $|f'(a_n)|_p = |f'(a)|_p$ for some $n \geq 1$ then

$$|f'(a_{n+1}) - f'(a_n)|_p \leq |a_{n+1} - a_n|_p \leq \left|\frac{f(a)}{f'(a)}\right|_p$$

where the first inequality is from (5.4) and the second inequality is from a_n and a_{n+1} both lying in the closed ball around a of radius $|f(a)/f'(a)|_p$. Thus $|f'(a_{n+1}) - f'(a_n)|_p < |f'(a)|_p = |f'(a_n)|_p$, so $|f'(a_{n+1})|_p = |f'(a_n)|_p = |f'(a)|_p$. We've shown $|f'(a_n)|_p = |f'(a)|_p$ for all n, and letting $n \to \infty$ gives us $|f'(\alpha)|_p = |f'(a)|_p$. $\square$

It's worthwhile to compare the recursions from Newton's method and from the contraction mapping theorem:

- Newton's method: $a_{n+1} = a_n - \dfrac{f(a_n)}{f'(a_n)}$, with $a_1 = a$ and $|f(a)|_p < |f'(a)|_p^2$.
- Contraction mapping: $a_{n+1} = a_n - \dfrac{f(a_n)}{f'(a)}$, with $a_1 = a$ and $|f(a)|_p < |f'(a)|_p^2$.

The difference is the denominators $f'(a_n)$ and $f'(a)$, and this has a profound effect on the rate of convergence. In Newton's method, (5.7) tells us

$$|\alpha - a_n|_p \leq |f'(a)|_p \left|\frac{f(a)}{f'(a)^2}\right|_p^{2^{n-1}}. \tag{6.6}$$

To find an estimate on $|\alpha - a_n|_p$ from the contraction mapping theorem, the recursion given by (6.5) implies $|a_{n+1} - a_n|_p = |\varphi^{n-1}(a_2) - \varphi^{n-1}(a_1)|_p \leq |a_2 - a_1|_p\lambda^{n-1}$ for all $n \geq 1$, where $a_2 - a_1 = f(a)/f'(a)$ and $\lambda = |f(a)/f'(a)^2|_p$. If $m > n$ the strong triangle inequality implies

$$|a_m - a_n|_p \leq \max_{n \leq j \leq m-1} |a_{j+1} - a_j|_p \leq |a_2 - a_1|_p\lambda^{n-1} = \left|\frac{f(a)}{f'(a)}\right|_p \lambda^{n-1}.$$

Letting $m \to \infty$ yields

$$\boxed{|\alpha - a_n|_p \le \left|\frac{f(a)}{f'(a)}\right|_p \left|\frac{f(a)}{f'(a)^2}\right|^{n-1}.} \tag{6.7}$$

The upper bound in (6.6) goes to 0 much faster than the upper bound in (6.7), so we *anticipate* that $|\alpha - a_n|_p \to 0$ faster when $\{a_n\}$ is comes from Newton's recursion compared to the contraction mapping recursion. For instance, if $|f(a)/f'(a)^2|_p = 1/p$, then the bound in (6.6) goes to 0 like $(1/p)^{2^{n-1}}$ while in (6.7) the bound goes to 0 like $(1/p)^n$, so there is at least a doubling of correct p-adic digits in each step of Newton's recursion while we expect at most one new correct p-adic digit at each step by the contraction mapping recursion.

Example 6.1. Let $f(X) = X^2 - 7$ with $a = 1$, so $|f(a)/f'(a)|_3 = 1/3$. The sequence $\{a_n\}$ produced from the contraction mapping recursion (6.5) with $a_1 = 1$ has a limit α and with a computer we find $|\alpha - a_n|_3 = (1/3)^n$ for $1 \le n \le 10$. The sequence $\{a_n\}$ based on Newton's recursion in Example 5.1 with $a_1 = 1$ has the same limit α, but a computer tells us that $|\alpha - a_n|_3 = (1/3)^{2^{n-1}}$ for $1 \le n \le 10$, which is much smaller than $(1/3)^n$.

7. The inevitability of $|f(a)|_p < |f'(a)|_p^2$

In Theorem 4.1 we have $|f'(\alpha)|_p = |f'(a)|_p \ne 0$, so Hensel's lemma produces a simple root α of $f(X)$ in $\mathbf{Z}_p$ that is close to a. The criterion $|f(a)|_p < |f'(a)|_p^2$ in Theorem 4.1 is not just a sufficient condition for there to be a simple root of $f(X)$ near a but it is also necessary, as the next theorem makes precise.

Theorem 7.1. *If $f(X) \in \mathbf{Z}_p[X]$ has a simple root α in $\mathbf{Z}_p$, then for all $a \in \mathbf{Z}_p$ that are close enough to α we have $|f'(a)|_p = |f'(\alpha)|_p$ and $|f(a)|_p < |f'(a)|_p^2$. In particular, these conditions hold when $|a - \alpha|_p < |f'(\alpha)|_p$.*

Proof. By (5.4) with $F(X) = f'(X)$, $|f'(\alpha) - f'(a)|_p \le |\alpha - a|_p < |f'(\alpha)|_p$. Thus $|f'(a)|_p = |f'(\alpha)|_p$. By (5.3) with $F(X) = f(X)$,

$$f(a) = f(\alpha + (a - \alpha)) = f(\alpha) + f'(\alpha)(a - \alpha) + z(a - \alpha)^2 = f'(\alpha)(a - \alpha) + z(a - \alpha)^2$$

for some $z \in \mathbf{Z}_p$. Both terms on the right side have absolute value less than $|f'(\alpha)|_p^2$ since $|a - \alpha|_p < |f'(\alpha)|_p$, so $|f(a)|_p < |f'(\alpha)|_p^2 = |f'(a)|_p^2$. ∎

8. Hensel's lemma for power series

Hensel's lemma can be applied to p-adic power series, not just polynomials. By "p-adic power series" we have in mind those with $\mathbf{Z}_p$-coefficients, which we will call power series "over $\mathbf{Z}_p$". We will consider such series of two types: those that converge on $\mathbf{Z}_p$ and those that converge on $p\mathbf{Z}_p$ but not necessarily on $\mathbf{Z}_p$ (*e.g.*, e^X for $p > 2$). To prove Hensel's lemma for these two types of power series, we will use the following power series analogue of (5.3) and (5.4).

Lemma 8.1. *Let $F(X)$ be a power series with coefficients in $\mathbf{Z}_p$ that converges on $\mathbf{Z}_p$.*

(1) *For x and y in $\mathbf{Z}_p$, $F(x + y) = F(x) + F'(x)y + zy^2$ for some $z \in \mathbf{Z}_p$.*
(2) *For x and y in $\mathbf{Z}_p$, $|F(x) - F(y)|_p \le |x - y|_p$.*

If $F(X)$ converges on $p\mathbf{Z}_p$ but not necessarily on $\mathbf{Z}_p$, then (1) and (2) are true for $x, y \in p\mathbf{Z}_p$.

Proof. Let $F(X) = \sum_{i \geq 0} c_i X^i$, so $c_i \in \mathbf{Z}_p$. Convergence of $F(X)$ on $\mathbf{Z}_p$ implies $c_i \to 0$ as $i \to \infty$. To prove (1), the analogue of (5.3), for x and y in $\mathbf{Z}_p$ we have

$$\begin{aligned} F(x+y) &= \sum_{i \geq 0} c_i (x+y)^i \\ &= c_0 + \sum_{i \geq 1} c_i (x^i + ix^{i-1}y + g_i(x,y)y^2) \text{ where } g_i(x,y) \in \mathbf{Z}_p[x,y] \subset \mathbf{Z}_p \\ &= \sum_{i \geq 0} c_i x^i + \sum_{i \geq 1} ic_i x^{i-1} y + \sum_{i \geq 2} c_i g_i(x,y) y^2, \end{aligned}$$

where we can break apart the series since $|c_i|_p \to 0$ as $i \to \infty$ and the numbers x, y, and $g_i(x,y)$ are all in $\mathbf{Z}_p$ (all the series converge) with $g_1(x,y) = 0$. Thus

$$F(x+y) = F(x) + F'(x)y + zy^2$$

where $z = \sum_{i \geq 2} c_i g_i(x,y) \in \mathbf{Z}_p$.

To prove (2), the analogue of (5.4), we use $|x^i - y^i|_p \leq |x-y|_p$ for all $x, y \in \mathbf{Z}_p$ and $i \geq 0$:

$$F(x) - F(y) = \sum_{i \geq 1} c_i (x^i - y^i) \Longrightarrow |F(x) - F(y)|_p \leq \max_{i \geq 1} |c_i|_p |x^i - y^i|_p \leq \max_{i \geq 1} |c_i|_p |x-y|_p,$$

which is at most $|x-y|_p$ since all $|c_i|_p$ are at most 1.

If $F(X)$ converges on $p\mathbf{Z}_p$ but not necessarily on $\mathbf{Z}_p$, so the coefficients c_i are in $\mathbf{Z}_p$ but need not tend to 0, the proofs of (1) and (2) remain valid for $x, y \in p\mathbf{Z}_p$ after checking some details: in the proof of (1), the series for $F(x)$ and $F'(x)$ converge when $x \in \mathbf{Z}_p$, while the series $\sum_{i \geq 1} c_i g_i(x,y)$ converges since $|g_i(x,y)|_p \leq \max(|x|_p, |y|_p)^{i-2}$ by the binomial theorem that gave rise to $g_i(x,y)$. The proof of (2) carries over with no changes at all provided $x, y \in p\mathbf{Z}_p$ so that $F(x)$ and $F(y)$ converge. □

Here is a power series version of Theorem 4.1 for series over $\mathbf{Z}_p$ that converge on $\mathbf{Z}_p$.

Theorem 8.2. *Let $f(X)$ be a power series with coefficients in $\mathbf{Z}_p$ that converges on $\mathbf{Z}_p$. If an $a \in \mathbf{Z}_p$ satisfies*

$$|f(a)|_p < |f'(a)|_p^2$$

then there is a unique $\alpha \in \mathbf{Z}_p$ such that $f(\alpha) = 0$ in $\mathbf{Z}_p$ and $|\alpha - a|_p < |f'(a)|_p$. Moreover,

(1) $|\alpha - a|_p = |f(a)/f'(a)|_p < |f'(a)|_p$,
(2) $|f'(\alpha)|_p = |f'(a)|_p$.

In this theorem, $f'(a)$ makes sense for $a \in \mathbf{Z}_p$ since $f'(X)$ converges on $\mathbf{Z}_p$: the coefficients of $f(X)$ tend to 0 since $f(1)$ converges so the coefficients of $f'(X)$ also tend to 0 and thus $f'(X)$ converges on $\mathbf{Z}_p$.

Proof. Using Lemma 8.1 in place of (5.3) and (5.4), the proof of Theorem 4.1 using Newton's method carries over word for word to the setting of power series with $\mathbf{Z}_p$-coefficients that converge on $\mathbf{Z}_p$. □

Remark 8.3. Theorem 8.2 can also be proved by turning the power series $f(X)$ into a polynomial, thereby reducing to the polynomial case. The p-adic Weierstrass preparation theorem says $f(X) = W(X)U(X)$ where $W(X)$ is a polynomial and $U(X)$ is a unit power series with coefficients in $\mathbf{Z}_p$ that converges on $\mathbf{Z}_p$ (there's a power series $V(X)$ with $\mathbf{Z}_p$-coefficients converging on $\mathbf{Z}_p$ such that $U(X)V(X) = 1$). From $f(a) = W(a)U(a)$ and $U(a)V(a) = 1$ for all $a \in \mathbf{Z}_p$, roots of $f(X)$ in $\mathbf{Z}_p$ are the same as roots of $W(X)$ in $\mathbf{Z}_p$. If $|f(a)|_p < |f'(a)|_p^2$ then $|f(a)|_p = |W(a)|_p$ since $|U(a)|_p = 1$, and also $|f'(a)|_p = |W'(a)|_p$ since from $f'(a) = W'(a)U(a) + W(a)U'(a)$ and $|W(a)U'(a)|_p \leq |W(a)|_p = |f(a)|_p < |f'(a)|_p$

we must have $|f'(a)|_p = |W'(a)U(a)|_p = |W'(a)|_p$. Thus $|W(a)|_p < |W'(a)|_p^2$, so there is a unique $\alpha \in \mathbf{Z}_p$ fitting the conclusion of Theorem 4.1 for $W(X)$. It's left to the reader to check α also fits the uniqueness conditions in Theorem 8.2 for $f(X)$.

The basic version of Hensel's lemma is true for power series as a special case of the strong version proved above: if $f(X)$ is a power series over $\mathbf{Z}_p$ that converges on $\mathbf{Z}_p$ and an $a \in \mathbf{Z}_p$ satisfies $f(a) \equiv 0 \bmod p$ and $f'(a) \not\equiv 0 \bmod p$ then there is a unique $\alpha \in \mathbf{Z}_p$ such that $f(\alpha) = 0$ and $\alpha \equiv a \bmod p$.

Like Theorem 5.4, the basic and strong form of Hensel's lemma for power series with $\mathbf{Z}_p$-coefficients that converge on $\mathbf{Z}_p$ are equivalent, with a proof similar to the polynomial case, but care is needed since part of the proof uses convergent power series in two variables. Details are left to the reader. Check also that Theorem 7.1 and its proof carry over to power series over $\mathbf{Z}_p$ that converge on $\mathbf{Z}_p$, where a simple root of a power series is a root at which the derivative is nonzero.

For power series over $\mathbf{Z}_p$ that converge on $p\mathbf{Z}_p$ but not necessarily on $\mathbf{Z}_p$, here is a version of Hensel's Lemma.

Theorem 8.4. *Let $f(X)$ be a power series with coefficients in $\mathbf{Z}_p$ that converges on $p\mathbf{Z}_p$. If an $a \in p\mathbf{Z}_p$ satisfies*

$$|f(a)|_p < |f'(a)|_p^2$$

then there is a unique $\alpha \in p\mathbf{Z}_p$ such that $f(\alpha) = 0$ in $\mathbf{Z}_p$ and $|\alpha - a|_p < |f'(a)|_p$. Moreover,

(1) *$|\alpha - a|_p = |f(a)/f'(a)|_p < |f'(a)|_p$,*
(2) *$|f'(\alpha)|_p = |f'(a)|_p$.*

Proof. The proof of Theorem 8.2 (including its use of Lemma 8.1) carries over with the complete space $p\mathbf{Z}_p$ replacing the complete space $\mathbf{Z}_p$ as the set containing the approximate roots a_n and the root α that is their limit. □

Remark 8.5. The reduction of Theorem 8.2 to the case of polynomials in Remark 8.3 applies to Theorem 8.4 too by using a p-adic Weierstrass preparation theorem for power series with $\mathbf{Z}_p$-coefficients that converge on $p\mathbf{Z}_p$ (but perhaps not on $\mathbf{Z}_p$).

In the proof of Hensel's lemma for polynomials by Newton's method, the bounds on $|f(a_n)|_p$ and $|a_{n+1} - a_n|_p$ carry over to Hensel's lemma for power series, whether in the form of Theorem 8.2 or 8.4: $|f(a_n)|_p \le |f'(a)|_p(|f(a)|_p/|f'(a)^2|_p)^{2^{n-1}}$ and $|a_{n+1} - a_n|_p \le |f'(a)|_p|f(a)/f'(a)^2|_p^{2^{n-1}}$. Using the strong triangle inequality in the second bound for $a_m - a_n = (a_m - a_{m-1}) + \cdots + (a_{n+1} - a_n)$ when $m > n$ and then letting $m \to \infty$, we get the bound from (5.7) all over again, but this time for power series:

$$\boxed{|\alpha - a_n|_p \le |f'(a)|_p \left|\frac{f(a)}{f'(a)^2}\right|_p^{2^{n-1}}.} \tag{8.1}$$

Example 8.6. For $c \in \mathbf{Z}_p$, set $(1+X)^c = \sum_{n\ge 0} \binom{c}{n} x^n$, where $\binom{c}{n} = \frac{c(c-1)\cdots(c-n-1)}{n!}$. Since c is a p-adic limit of positive integers (use truncations of the p-adic expansion of c), the number $\binom{c}{n}$ is a p-adic limit of integers and thus $\binom{c}{n} \in \mathbf{Z}_p$. For $c \in \mathbf{Z}_p - \{0, 1, 2, 3, \ldots\}$, $\binom{c}{n} \in \mathbf{Z}_p^\times$ infinitely often (for those n whose base p expansion is an initial part of the p-adic expansion of c), so $(1+X)^c$ has coefficients in $\mathbf{Z}_p$ that do not tend to 0 and thus $(1+X)^c$ converges on $p\mathbf{Z}_p$ but not on $\mathbf{Z}_p$.

Consider $f(X) = (1+X)^{\sqrt{7}} - 10$, where $\sqrt{7} = 1 + 3 + 3^2 + 2 \cdot 3^4 + \cdots$ in $\mathbf{Z}_3$. For each $a \in 3\mathbf{Z}_3$, $f(a) \equiv 0 \bmod 3$ and $f'(a) = \sqrt{7}(1+a)^{\sqrt{7}-1} \equiv \sqrt{7} \not\equiv 0 \bmod 3$, so by Theorem 8.4 there is a unique solution to $f(\alpha) = 0$ in $3\mathbf{Z}_3$ by using Newton's method with initial value $a_1 = a$. Taking $a = 0$, $f(a_1) = -9$ and $f'(a_1) = \sqrt{7} \not\equiv 0 \bmod 3$. Using (8.1) for $f(X) = (1+X)^{\sqrt{7}} - 10$, $|\alpha - a_n|_3 \leq (1/9)^{2^{n-1}} = (1/3)^{2^n}$. Taking $n = 3$,

$$\alpha \equiv a_3 \equiv 3^2 + 2 \cdot 3^3 + 2 \cdot 3^4 + 3^5 + 2 \cdot 3^6 \bmod 3^8. \tag{8.2}$$

The number on the right is 1926. As a reality check[4], let's confirm that $f(a_3)$ is 3-adically small. Since $|f(x) - f(y)|_3 \leq |x - y|_3$ for $x, y \in \mathbf{Z}_3$, $f(a_3) \equiv f(1926) \bmod 3^8$. Since 1926 is divisible by 9, so $1926^n \equiv 0 \bmod 3^8$ for $n \geq 4$, and the power series coefficients of $f(X)$ are all in $\mathbf{Z}_3$, $f(1926) \equiv -9 + \sum_{n=1}^{3} \binom{\sqrt{7}}{n} 1926^n \bmod 3^8$. With a computer, the right side is $0 \bmod 3^8$.

Example 8.7. For $p > 2$ and $c \in 1 + p\mathbf{Z}_p$, set $f(X) = e^X - c$. The p-adic disc of convergence for e^X (and thus for $e^X - c$) on $\mathbf{Q}_p$ is $p\mathbf{Z}_p$. Since $f(0) = 1 - c$ and $f'(0) = 1$, we have $|f(0)|_p \leq 1/p$ and $|f'(0)||_p = 1$. Therefore Theorem 8.4 tells us there is a unique $\alpha \in p\mathbf{Z}_p$ such that $e^\alpha = c$ and $|\alpha|_p < 1$, and in fact $|\alpha|_p = |f(0)|_p = |c-1|_p$. This α in $p\mathbf{Z}_p$ for which $e^\alpha = c$ is the p-adic logarithm of c.

By (8.1) with $a = 0$, $|\alpha - a_n|_p \leq |1 - c|_p^{2^{n-1}}$. For instance, if we want to solve $e^\alpha = -2$ for $\alpha \in 3\mathbf{Z}_3$ and compute $\alpha \bmod 3^5$, use Newton's method with $f(x) = e^x + 2$. Then $|\alpha - a_n|_3 \leq |1 - (-2)|_3^{2^{n-1}} = 1/3^{2^{n-1}}$. The upper bound is at most $1/3^5$ for $n \geq 4$, so take $n = 4$. The Newton's method recursion is $a_{n+1} = a_n - f(a_n)/f'(a_n) = a_n - (e^{a_n} + 2)/e^{a_n} = a_n - 1 - 2/e^{a_n}$. By computer, the term for a_4 in the table below satisfies $e^{a_4} \equiv -2 \bmod 3^{15}$.

n	a_n
1	0
2	-3
3	$2 \cdot 3 + 2 \cdot 3^2 + 3^5 + 2 \cdot 3^7 + 2 \cdot 3^{10} + \cdots$
4	$2 \cdot 3 + 2 \cdot 3^2 + 3^5 + 3^6 + 2 \cdot 3^8 + 3^9 + \cdots$

9. Hensel's lemma beyond $\mathbf{Q}_p$

The proofs of Hensel's lemma work for polynomials with coefficients in a field K complete with respect to an absolute value satisfying the strong triangle inequality: $|x + y| \leq \max(|x|, |y|)$ for all x and y in K. Set $\mathfrak{o} = \{x \in K : |x| \leq 1\}$.

Theorem 9.1. *For $\mathfrak{o}$ as above and a polynomial $f(X)$ with coefficients in $\mathfrak{o}$, assume some $a \in \mathfrak{o}$ satisfies*

$$|f(a)| < |f'(a)|^2.$$

Then there is a unique $\alpha \in \mathfrak{o}$ such that $f(\alpha) = 0$ in $\mathfrak{o}$ and $|\alpha - a| < |f'(a)|$. Moreover,

1. $|\alpha - a| = |f(a)/f'(a)| < |f'(a)|$,
2. $|f'(\alpha)| = |f'(a)| \neq 0$.

Conversely, if $f(X)$ has a simple root $\alpha \in \mathfrak{o}$, then for $a \in K$ such that $|a - \alpha| < |f'(\alpha)|$ then we have $|f'(a)| = |f'(\alpha)|$ and $|f(a)| < |f'(a)|^2$.

Theorems 5.4 and 7.1 also carry over to $\mathfrak{o}[X]$, with the same proofs (the proofs over $\mathbf{Z}_p$ were written specifically to make sure the proofs carry over to $\mathfrak{o}$ with no changes):

Theorem 9.2. *The following properties $\mathfrak{o}[X]$ are equivalent.*

1. *When $f(X) \in \mathfrak{o}[X]$ and there is an $a \in \mathfrak{o}$ such that $|f(a)| < 1$ and $|f'(a)| = 1$, there is a unique $\alpha \in \mathfrak{o}$ such that $f(\alpha) = 0$ in $\mathfrak{o}$ and $|\alpha - a| < 1$,*
2. *When $f(X) \in \mathfrak{o}[X]$ and there is an $a \in \mathfrak{o}$ such that $|f(a)| < |f'(a)|^2$, there is a unique $\alpha \in \mathfrak{o}$ such that $f(\alpha) = 0$ in $\mathfrak{o}$ and $|\alpha - a| < |f'(a)|$, and in fact $|\alpha - a|_p = |f(a)/f'(a)|_p$*

[4] A 3-adicity check, since we're not working over the reals?

Theorem 9.3. *If $f(X) \in \mathfrak{o}[X]$ has a simple root α in $\mathfrak{o}$, then for all $a \in \mathfrak{o}$ that are close enough to α we have $|f'(a)| = |f'(\alpha)|$ and $|f(a)| < |f'(a)|^2$. In particular, these conditions hold when $|a - \alpha| < |f'(\alpha)|$.*

For power series, everything written about power series with coefficients in $\mathbf{Z}_p$ in Section 8 carries over to power series with coefficients in $\mathfrak{o}$: all such power series converge on the maximal ideal $\mathfrak{m} = \{x \in \mathfrak{o} : |x| < 1\}$, and those with coefficients tending to 0 converge on $\mathfrak{o}$. We record below the statements of Hensel's lemma for both types of power series, leaving the proofs (similar to the polynomial case in most respects) to the reader.

Theorem 9.4. *Let $f(X)$ be a power series with coefficients in $\mathfrak{o}$ that converges on $\mathfrak{o}$. If an $a \in \mathfrak{o}$ satisfies*

$$|f(a)| < |f'(a)|^2$$

then there is a unique $\alpha \in \mathfrak{o}$ such that $f(\alpha) = 0$ in $\mathfrak{o}$ and $|\alpha - a| < |f'(a)|$. Moreover,

1. $|\alpha - a| = |f(a)/f'(a)| < |f'(a)|$,
2. $|f'(\alpha)| = |f'(a)|$.

Theorem 9.5. *Let $f(X)$ be a power series with coefficients in $\mathfrak{o}$, so it converges on $\mathfrak{m}$. If an $a \in \mathfrak{m}$ satisfies*

$$|f(a)| < |f'(a)|^2$$

then there is a unique $\alpha \in \mathfrak{m}$ such that $f(\alpha) = 0$ in $\mathfrak{o}$ and $|\alpha - a| < |f'(a)|$. Moreover,

1. $|\alpha - a| = |f(a)/f'(a)| < |f'(a)|$,
2. $|f'(\alpha)| = |f'(a)|$.

10. Hensel's lemma with submultiplicative absolute values

We will formulate a version of Hensel's lemma where $\mathbf{Z}_p$ is replaced by a ring A that is complete with respect to an absolute value similar to $|\cdot|_p$ where multiplicativity is replaced by submultiplicativity.

Definition 10.1. A *submultiplicative absolute value* $|\cdot|$ on a nonzero commutative ring A is a function $|\cdot| \colon A \to \mathbf{R}$ such that for all a and b in A,

(i) $0 \le |a| \le 1$ with $|a| = 0$ if and only if $a = 0$,
(ii) $|a + b| \le \max(|a|, |b|)$ (strong triangle inequality),
(iii) $|ab| \le |a||b|$ (submultiplicative).

For an ordinary absolute value, $|1| = 1$ by multiplicativity since $|1| = |1^2| = |1|^2$. When we have submultiplicativity, we can still get $|1| = 1$ when $|a| \le 1$ for all a: $|1| = |1^2| \le |1|^2$ and $|1| > 0$, so $1 \le |1|$. Since the reverse inequality $|1| \le 1$ holds by definition, we get $|1| = 1$. Therefore $|-a| = |a|$: $|-a| \le |-1||a| \le |a|$ and $|a| = |-(-a)| \le |-1||-a| \le |-a|$.

On A we have a metric $\rho(a, b) = |a - b|$; (that $\rho(a, b) = \rho(b, a)$ follows from $|a - b| = |b - a|$ because $|x| = |-x|$). Many basic properties of $|\cdot|_p$ on $\mathbf{Z}_p$ remain true for a submultiplicative absolute value, with the same or very similar proofs. For example, addition and multiplication on A are continuous and a sequence $\{a_n\}$ in A is Cauchy if and only if $|a_{n+1} - a_n| \to 0$.

Before describing a form of Hensel's lemma for polynomials and power series with coefficients in a ring complete for a submultiplicative absolute value, let's describe a source of examples of submultiplicative absolute values using powers of an ideal.

Let A be a nonzero commutative ring containing an ideal I such that $\bigcap_{n \ge 0} I^n = \{0\}$, where I^n is the nth power of I as an ideal[5]. For $a \ne 0$ in A, set $|a|_I = (1/2)^n$ if $a \in I^n$ for n as large as possible (there is a maximal n such that $a \in I^n$ because $\bigcap_{n \ge 0} I^n = \{0\}$), and

[5] I^n is the ideal generated by n-fold products $a_1 \cdots a_n$ for $a_i \in I$. It's *more* than all a^n for $a \in I$.

set $|0|_I = 0$. Then $0 \le |a|_I \le 1$ and $|a|_I = 0$ only when $a = 0$.[6] We have $|a|_I\rho = (1/2)^n \Rightarrow a \in I^n$ and $a \in I^n \Rightarrow |a|_I \le (1/2)^n$ (the exponent n in the second implication might not be maximal for the choice of a). That $|a+b|_I \le \max(|a|_I, |b|_I)$ is left to the reader. The function $\rho_I(a,b) = |a-b|_I$ is called the *I-adic metric* on A. To prove submultiplicativity of $|\cdot|_I$ on A, if a and b are nonzero in A with $|a|_I = (1/2)^n$ and $|b|_I = (1/2)^m$ then $ab \in I^{n+m}$, so $|ab|_I \le (1/2)^{n+m} = |a|_I|b|_I$. Obviously $|ab|_I \le |a|_I|b|_I$ if a or b is 0.

We have $A = I^0 \supset I \supset I^2 \supset I^3 \supset \cdots$ and elements of I^n are considered I-adically small if n is large. A sequence $\{a_j\}$ in A is *I-adically Cauchy* if for each $n \ge 1$ there is $j \ge 1$ such that $a_{j'} - a_{j''} \in I^n$ for all $j', j'' \ge j$. This is equivalent to $a_{j+1} - a_j \in I^n$ for all large j (depending on n) and is the same as $\{a_j\}$ being a Cauchy sequence for the metric ρ_I. We say A is *I-adically complete* if each I-adic Cauchy sequence in A converges in A.

An example of such a ring A and ideal I is $R[T]$ and $I = (T) = TR[T]$ where R is a commutative ring. Then $I^n = (T^n)$, which is the set of polynomials with no terms in degree less than n, so $\bigcap_{n\ge1} I^n = \{0\}$. The I-adic absolute value on $R[T]$ is called the T-adic absolute value and is denoted $|\cdot|_T$. We have $|f(T)|_T = (1/2)^n$ if and only if the lowest-degree nonzero term in $f(T)$ has degree n. The ring $R[T]$ is *not* T-adically complete since, for instance, the partial sums $s_n = 1 + T + T^2 + \cdots + T^n$ form a T-adic Cauchy sequence that has no limit in $R[T]$: $(1-T)s_n = 1 - T^{n+1}$, which tends to 1, so if $\{s_n\}$ has a limit s in $R[T]$ then $(1-T)s = 1$, but $1-T$ has no multiplicative inverse in $R[T]$.

Another example is the ring $R[[T]]$ of formal power series $\sum_{n\ge0} c_nT^n$ with $c_n \in R$ and the ideal $I = (T) = TR[[T]]$ consisting of multiples of T in $R[[T]]$ (the power series with constant term 0). The T-adic absolute value $|\cdot|_T$ on $R[[T]]$ is defined in the same way as on $R[T]$, and $R[T]$ is dense in $R[[T]]$: the sum of all terms in a power series with degree less than n is in $R[T]$ and differs from the power series by a series in I^n, and we can take n arbitrarily large. The ring $R[[T]]$ is T-adically complete: a T-adic Cauchy sequence has the coefficients in each degree eventually constant, leading to a candidate limit series in $R[[T]]$ that really is the limit of the sequence. So $R[[T]]$ is the T-adic completion of $R[T]$. For example, the sequence $s_n = 1 + T + T^2 + \cdots + T^n$ in $R[[T]]$ converges to $\sum_{n\ge0} T^n$, which is a multiplicative inverse of $1 - T$.

More generally, consider the ring $R[T_1, \ldots, T_d]$ and $I = (T_1, \ldots, T_d)$, so I^n is the polynomials in $R[T_1, \ldots, T_d]$ having no nonzero terms in degree below n. Then $|f(T_1, \ldots, T_d)|_I = (1/2)^n$ means the lowest-degree nonzero monomial in $f(T_1, \ldots, T_d)$ has degree n. This ring is not I-adically complete, but the ring $R[[T_1, \ldots, T_d]]$ of formal power series in d indeterminates is complete with respect to its ideal $I = (T_1, \ldots, T_d)$ of formal power series with constant term 0 and $R[T_1, \ldots, T_d]$ is dense in $R[[T_1, \ldots, T_d]]$.

Remark 10.2. On $\mathbf{Z}[T]$ we can consider the I-adic absolute value when I is (p), (T), and (p, T). What's the difference?

(1) Being p-adically small means all the coefficients are highly divisible by p.
(2) Being T-adically small means the polynomial has no low-degree terms.
(3) Being (p, T)-adically small means low-degree coefficients are highly divisible by p.

The p-adic completion of $\mathbf{Z}[T]$ is the set of power series $\sum_{n\ge0} c_nT^n$ where $c_n \in \mathbf{Z}_p$ and $c_n \to 0$, the T-adic completion of $\mathbf{Z}[T]$ is $\mathbf{Z}[[T]]$, and the (p, T)-adic completion of $\mathbf{Z}[T]$ is $\mathbf{Z}_p[[T]]$. The ring $\mathbf{Z}_p[[T]]$ is both T-adically and (p, T)-adically complete, but these notions of convergence are different. A sequence in $\mathbf{Z}_p[[T]]$ that converges T-adically also converges (p, T)-adically since $(T)^n \subset (p, T)^n$ as ideals in $\mathbf{Z}_p[[T]]$, but the converse is false: a sequence in $\mathbf{Z}_p$ that converges p-adically does not converge T-adically when viewed as constants in $\mathbf{Z}_p[[T]]$. And $\mathbf{Z}[T]$ is (p, T)-adically dense in $\mathbf{Z}_p[[T]]$ but is not T-adically dense in $\mathbf{Z}_p[[T]]$:

[6] The role of $1/2$ in the definition of $|a|_I$ when $a \ne 0$ could be replaced by an arbitrary number strictly between 0 and 1. The main point is that $|a|_I$ is very small when $a \in I^n$ for large n.

for $c \in \mathbf{Z}_p - \mathbf{Z}$, no $f(T) \in \mathbf{Z}[T]$ satisfies $|c - f(T)|_T < 1$ since such an inequality forces $f(0) = c$ but $f(0) \in \mathbf{Z}$ and $c \notin \mathbf{Z}$.

Lemma 10.3. *Let A be a nonzero commutative ring that is complete with respect to a submultiplicative absolute value $|\cdot|$.*

(1) $|ua| = |a|$ *for all* $u \in A^\times$ *and* $a \in A$.

(2) *If* $y \in A$ *satisfies* $|y| < 1$ *then* $1 + y \in A^\times$.

(3) *The units of A are closed in $A^\times$: if $\{u_n\}$ is a sequence in $A^\times$ with a limit in A, then the limit is in $A^\times$.*

Proof. (1) For $u \in A^\times$ and $a \in A$, $|ua| \leq |u||a| \leq |a|$. For the reverse inequality, let $uv = 1$ in A, so $|a| = |ua(v)| \leq |ua||v| \leq |ua|$.

(2) We use a geometric series: the partial sums $s_n = 1 - y + y^2 - y^3 + \cdots + (-y)^n$ are a Cauchy sequence in A by the strong triangle inequality since $|(-y)^n| \leq |y|^n \to 0$ as $n \to \infty$. Let $s = \lim_{n\to\infty} s_n$ in A, so $(1 + y)s = \lim_{n\to\infty}(1 + y)s_n = \lim_{n\to\infty} 1 - (-y)^n = 1$. Thus $1 + y \in A^\times$.

(3) Let $u_n \to L$ in A. Set $v_n = 1/u_n$, so $v_{n+1} - v_n = v_n v_{n+1}(u_n - u_{n+1})$. Thus $|v_{n+1} - v_n| \leq |u_n - u_{n+1}| \to 0$ as $n \to \infty$. Therefore $\{v_n\}$ is Cauchy in A. Let M be its limit, so $u_n v_n \to LM$ by continuity of multiplication. Also $u_n v_n = 1$ for all n, so $LM = 1$. Thus $L \in A^\times$. □

Example 10.4. A power series $f(T) \in R[[T]]$ is a unit if and only if its constant term $f(0)$ is a unit in R. Indeed, if $f(T)g(T) = 1$ then the constant terms have product 1, so $f(0) \in R^\times$. Conversely, if $f(0) \in R^\times$ then we can prove $f(T) \in R[[T]]^\times$ by writing $f(T) = f(0) + TF(T) = f(0)(1 + (1/f(0))TF(T)$. The first factor $f(0)$ is a unit in $R[[T]]$ since it is a unit in R and the second factor $1+(1/f(0))TF(T)$ is a unit in $R[[T]]$ by Lemma 10.3(2).

In contrast to $\mathbf{Z}_p$, where the condition $a \in \mathbf{Z}_p^\times$ means the same thing as $|a|_p = 1$, there is a difference in $R[[T]]$ between $|f(T)|_T = 1$ and $f(T) \in R[[T]]^\times$ unless R is a field: a power series $f(T) \in R[[T]]$ has $|f(T)|_T = 1$ when its constant term is in $R - \{0\}$, while $f(T)$ is a unit in $R[[T]]$ only when its constant term is in $R^\times$, and that is more restrictive than $|f(T)|_T = 1$ unless $R^\times = R - \{0\}$ (*i.e.*, unless R is a field). For instance, in $\mathbf{Z}[[T]]$ we have $|2 + T|_T = 1$ and $2 + T$ is not a unit in $\mathbf{Z}[[T]]$. More generally, if A is a ring and I is an ideal of A such that $\bigcap_{n\geq 0} I^n = \{0\}$, if $a \in A^\times$ then $|a|_I = (1/2)^0 = 1$ because no unit is contained in I (otherwise the powers of I could not have intersection $\{0\}$), but the converse might not be true.

Now we can present a basic form of Hensel's lemma in $A[X]$.

Theorem 10.5. *Let $f(X) \in A[X]$, where A is a nonzero commutative ring complete with respect to a submultiplicative absolute value $|\cdot|$.*

If an $a \in A$ satisfies $|f(a)| < 1$ and $f'(a) \in A^\times$ then then there is a unique $\alpha \in A$ such that $f(\alpha) = 0$ in A and $|\alpha - a| < 1$. Moreover,

(1) $|\alpha - a| = |f(a)| < 1$,

(2) $f'(\alpha) \in A^\times$.

Note the condition on $f'(a)$ is $f'(a) \in A^\times$, *not* $|f'(a)| = 1$; being a unit in A is potentially more precise than having absolute value 1.

Proof. The proof is like that of Theorem 4.1: set $a_1 = a$ and $a_{n+1} = a_n - f'(a_n)/f'(a_n)$ and check by induction that $f'(a_n) \in A^\times$ and $|f(a_n)| \leq |f(a)|^{2^{n-1}}$ for all $n \geq 1$. You need to be careful not to use multiplicativity on $|\cdot|$ on all products, but it can be used if one factor in a product is a unit in A, by Lemma 10.3(1).

The sequence $\{a_n\}$ is Cauchy in A since $|a_{n+1} - a_n| \to 0$, and letting $\alpha = \lim_{n\to\infty} a_n$, we have $f(\alpha) = 0$ since $|f(a_n)| \to 0$ and polynomials in $A[X]$ are continuous on A. We have $f'(\alpha) \in A^\times$ by Lemma 10.3(3) since $f'(a_n) \in A^\times$ for all n and $f'(X)$ is a polynomial.

The proof that $|\alpha - a| = |f(a)|$ is very similar to the case $A = \mathbf{Z}_p$. To prove α is the unique root of $f(X)$ such that $|\alpha - a| < 1$, we need to be a little more careful than in the case $A = \mathbf{Z}_p$ since $|\cdot|$ is submultiplicative rather than multiplicative. If $\beta \in A$ satisfies $f(\beta) = 0$ and $|\beta - a| < 1$ then $|\beta - \alpha| < 1$, so $\beta = \alpha + h$ where $|h| < 1$. Then

$$0 = f(\beta) = f(\alpha + h) = f(\alpha) + f'(\alpha)h + zh^2 = f'(\alpha)h + zh^2$$

for some $z \in A$. Then $f'(\alpha)h = -zh^2$. Taking absolute values of both sides and using Lemma 10.3(1), $|h| = |-zh^2| \le |h|^2$. Since $|h| < 1$, having $|h| \le |h|^2$ forces $|h| = 0$, so $h = 0$ and thus $\beta = \alpha$. □

Example 10.6. Let $A = \mathbf{Q}[[T]]$ and $f(X) = X^n - (1 + T) \in \mathbf{Q}[[T]][X] = A[X]$ where $n \ge 1$. Since $|f(1)|_T = |-T|_T < 1$ and $f'(1) = n \in A^\times$, there is a unique $g(T) \in \mathbf{Q}[[T]]$ such that $g(T)^n = 1 + T$ and $|g(T) - 1|_T < 1$, meaning $g(T) \in 1 + T\mathbf{Q}[[T]]$. This $g(T)$ begins as $1 + (1/n)T + \cdots$ and is the power series $(1 + T)^{1/n}$ in $\mathbf{Q}[[T]]$ with constant term 1. Its full formula is $\sum_{k\ge0} \binom{1/n}{k} T^k$.

Note that if we use $A = \mathbf{Z}[[T]]$ and $f(X) = X^n - (1 + T) \in A[X]$ (same as before), then $|f(1)|_T < 1$ and $f'(1) = n$, so if $n \ge 2$ then $f'(1) \notin A^\times$ (a unit in $\mathbf{Z}[[T]]$ has constant term ±1 by Example 10.4) and there is no solution to $g(T)^n = 1+T$ in $1+T\mathbf{Z}[[T]]$: such a solution would also be a solution in $1 + T\mathbf{Q}[[T]]$ and we saw the unique solution in $1 + T\mathbf{Q}[[T]]$ has linear coefficient $1/n$, which is not an integer. This illustrates why in Theorem 10.5 we can't always replace "$f'(a) \in A^\times$" with "$|f'(a)| = 1$".

We don't really need to replace $\mathbf{Z}[[T]]$ with $\mathbf{Q}[[T]]$ to find a power series solution to $X^n = 1 + T$: we just need n to be a unit in A, so we could use $A = \mathbf{Z}[1/n][[T]]$. For example, $\sqrt{1 + T} \in \mathbf{Z}[1/2][[T]]$. Indeed,

$$\sqrt{1+T} = 1 + \frac{1}{2}T - \frac{1}{2^3}T^2 + \frac{1}{2^4}T^3 - \frac{5}{2^7}T^4 + \frac{7}{2^8}T^5 - \frac{21}{2^{10}}T^6 + \cdots$$

and all coefficients have 2-power denominator. Similarly, $\sqrt[6]{1+T} \in \mathbf{Z}[1/6][[T]]$:

$$\sqrt[6]{1+T} = 1 + \frac{1}{6}T - \frac{15}{6^3}T^2 + \frac{55}{6^4}T^3 - \frac{8415}{6^7}T^4 + \frac{38709}{6^8}T^5 - \frac{1122561}{6^{10}}T^6 + \cdots.$$

Here is an analogue of Theorem 10.5 for roots of power series in $A[[X]]$. Every such series converges on $\{x \in A : |x| < 1\}$, and this is where we'll describe a criterion for the existence of a root.

Theorem 10.7. *Let $f(X) \in A[[X]]$, where A is a nonzero ring complete with respect to a submultiplicative absolute value $|\cdot|$.*

If some $a \in A$ such that $|a| < 1$ satisfies $|f(a)| < 1$ and $f'(a) \in A^\times$ then then there is a unique $\alpha \in A$ such that $f(\alpha) = 0$ in A and $|\alpha - a| < 1$. Moreover,

(1) $|\alpha - a| = |f(a)| < 1$,
(2) $f'(\alpha) \in A^\times$.

Proof. The proof proceeds in the same way as that of Theorem 10.5, relying on an extension of Lemma 8.1 to power series with coefficients in A rather than $\mathbf{Z}_p$. □

Example 10.8. Let $A = \mathbf{Q}[[T]]$ and

$$f(X) = e^X - (1 + T) = -T + X + \frac{1}{2}X^2 + \cdots \in A[[X]].$$

Since $|f(0)|_T = |-T|_T < 1$ and $f'(0) = 1$, there is a unique $g(T) \in T\mathbf{Q}[[T]]$ such that $f(g(T)) = 0$, meaning $g(T) \in T\mathbf{Q}[[T]]$ and $e^{g(T)} = 1 + T$. To find the coefficients of $g(T)$, differentiate both sides of the equation $e^{g(T)} = 1 + T$: $e^{g(T)}g'(T) = 1$, so

$$g'(T) = \frac{1}{e^{g(T)}} = \frac{1}{1+T} = \sum_{n \geq 0} (-1)^n T^n.$$

Therefore $g(T) = \sum_{n \geq 0}((-1)^n/(n+1))T^{n+1} = \sum_{n \geq 1}((-1)^{n-1}/n)T^n$, which is the formal power series for $\log(1+T)$ in $T\mathbf{Q}[[T]]$. The constant term of $g(T)$ is 0 since $g(T) \in T\mathbf{Q}[[T]]$.

Example 10.9. Let $A = \mathbf{Q}[[T]]$ and

$$f(X) = Xe^X - T = -T + X + X^2 + \frac{1}{2}X^3 + \cdots \in A[[X]].$$

Since $|f(0)|_T = |-T|_T < 1$ and $f'(0) = 1$, there is a unique $W(T) \in T\mathbf{Q}[[T]]$ such that $f(W(T)) = 0$, meaning $W(T)e^{W(T)} = T$. The series for $W(T)$ is

$$T - T^2 + \frac{3}{2}T^3 - \frac{8}{3}T^4 + \frac{125}{24}T^5 - \frac{54}{5}T^6 + \cdots = \sum_{n \geq 0} \frac{(-n)^{n-1}}{n!}T^n.$$

Classically, $W(T)$ as a function on $\mathbf{C}$ (initially near 0) is called Lambert's W-function.

Example 10.10. (Formal compositional inverse) Let $f(T) = c_1T + \cdots \in TR[[T]]$ for a commutative ring R. We'll show there is a $g(T) \in TR[[T]]$ such that $f(g(T)) = T$ if and only if $c_1 \in R^\times$, in which case $g(T)$ is unique and we also have $g(f(T)) = T$.

Suppose $g(T) = b_1T + \cdots \in TR[[T]]$ satisfies $f(g(T)) = T$. By a formal calculation, $f(g(T)) = c_1b_1T +$ higher order terms. That is, since $g(T) \equiv b_1T \bmod T^2$ we have $g(T)^n \equiv 0 \bmod T^2$ when $n \geq 2$, so $f(g(T)) \equiv c_1g(T) \equiv c_1b_1T \bmod T^2$. If $f(g(T)) = T$ then $T \equiv c_1b_1T \bmod T^2$, so $c_1b_1 = 1$ in R. Thus $c_1 \in R^\times$.

Conversely, suppose $c_1 \in R^\times$. We will use Theorem 10.7 with $A = R[[T]]$ and the power series $F(X) = f(X) - T \in A[[X]]$. Since $|F(0)|_T = |-T|_T < 1$ and $F'(0) = c_1 \in R^\times \subset R[[T]]^\times$, by Theorem 10.7 there is a unique $g(T) \in TR[[T]]$ such that $F(g(T)) = 0$, which means $f(g(T)) = T$.

Since $g(T)$ is in $TR[[T]]$, just like $f(T)$, we can run through the same reasoning with $g(T)$ in place of $f(T)$ by using $G(X) = g(X) - T$ in place of $F(X) = f(X) - T$ to see there is a unique $h(T) \in TR[[T]]$ such that $G(h(T)) = 0$, meaning $g(h(T)) = T$. Then $f(T) = f(g(h(T))) = h(T)$ since $f(g(T)) = T$. That proves $g(f(T)) = T$. For instance, in Example 10.8, rewriting $e^{\log(1+T)} = 1 + T$ as $F(\log(1 + T)) = T$ where $F(T) = e^T - 1$, we get $\log(1 + F(T)) = T$, or equivalently $\log(e^T) = T$.

The classical implicit function theorem says that if $f(x, y)$ is a C^1-function on an open set U in $\mathbf{R}^2$ and at a point (x_0, y_0) in U we have $f(x_0, y_0) = 0$ and $f_y(x_0, y_0) \neq 0$, then for all x close enough to x_0 there is a unique $g(x)$ near y_0 such that $f(x, g(x)) = 0$, and g is C^1 near x_0. Here is an analogue of this result for formal power series "near $(0, 0)$".

Example 10.11. (Formal implicit function theorem) Let $f(X, Y) = c_{10}X + c_{01}Y + \cdots$ in $R[[X, Y]]$, so $f(0, 0) = 0$ and $f_Y(0, 0) = c_{01}$. If $f_Y(0, 0) \in R^\times$, then there is a unique power series $g(X) \in XR[[X]]$ such that $f(X, g(X)) = 0$. To prove this, view $R[[X, Y]]$ as $R[[X]][[Y]] = A[[Y]]$ where $A = R[[X]]$, which is X-adically complete. Viewing $f(X, Y)$ in $R[[X]][[Y]]$ means thinking of $f(X, Y)$ as a power series in Y whose coefficients are power series in X. We have $|f(X, 0)|_X = |c_{10}X + \cdots|_X < 1$ and $f_Y(X, 0)$ is in $R[[X]]^\times$ since its constant term $f_Y(0, 0)$ is a unit in R (Example 10.4), so there is a unique $g(X) \in XR[[X]]$ such that $f(X, g(X)) = 0$ by Theorem 10.7 for power series in Y with $A = R[[X]]$.

In particular, if K is a field and $f(X,Y) \in K[[X,Y]]$ satisfies $f(0,0) = 0$ and $f_Y(0,0) \neq 0$ then there is a unique $g(X) \in XK[[X]]$ such that $f(X, g(X)) = 0$. An example of this is $f(X,Y) = \cos X - 1 + \sin Y = X^2/2 + Y + \cdots$ in $\mathbf{Q}[[X]][[Y]]$. The constant term is 0 and the coefficient of Y is nonzero in $\mathbf{Q}$, so there is a unique $g(X) \in X\mathbf{Q}[[X]]$ such that $\sin(g(X)) = 1 - \cos X$. The series $g(X)$ begins as $-(1/2)X^2 + (1/24)X^4 - (1/45)X^6 + \cdots$.

For a version of Hensel's lemma where $\mathbf{Z}_p$ is replaced by a complete local ring (possibly not an integral domain), see [3, Theorem 7.3]. For a version of Hensel's lemma dealing with zeros of several polynomials in several variables, see [1].

References

[1] K. Conrad, *A multivariable Hensel's lemma*, URL `https://kconrad.math.uconn.edu/blurbs/gradnumthy/multivarhensel.pdf`.

[2] C. H. Edwards, Jr., "Advanced Calculus of Several Variables," Academic Press, New York, 1973.

[3] D. Eisenbud, "Commutative Algebra with a View to Algebraic Geometry," Springer-Verlag, New York, 1995.

[4] S. Lang, "Algebraic Number Theory," 3rd ed., Springer-Verlag, New York, 1994.

[5] R. Y. Lewis, *A Formal Proof of Hensel's Lemma over the p-adic Integers*, URL `http://robertylewis.com/padics/padics.pdf`.