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 give a stronger version 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 $f(\alpha) = 0$ 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

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

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 $f(a_{n+1}) \equiv 0 \bmod p^{n+1} \Rightarrow f(a_{n+1}) \equiv 0 \bmod p^n$, any 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 $a_n \bmod p^n$, 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

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

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 $y = p^n t_n$:

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

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 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),

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

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 the congruence mod p since we assumed that $f'(a) \not\equiv 0 \bmod p$.

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,

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

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

[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 any $n \geq 1$.

It remains to show α is the unique root of $f(X)$ in $\mathbf{Z}_p$ that is congruent to a mod 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 mod 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 $\beta \equiv \alpha \bmod p^{n+1}$. □

Remark 2.3. An argument similar to the last paragraph shows for all $n \geq 1$ that $f(X)$ has a unique root mod p^n that reduces to a mod p. So 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 $\mathbf{Z}/(p^n)$ for all $n \geq 1$.

Here are five applications of Hensel's lemma.

Example 2.4. 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 2.5. 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: if $f(\beta) = 0$ and $\beta \equiv 1 \bmod 2$ then $f(\beta) \equiv 0 \bmod 4$ and $\beta \equiv 1$ or $3 \bmod 4$, but $f(1) \equiv 2 \bmod 4$ and $f(3) \equiv 2 \bmod 4$.

Example 2.6. For any positive integer n that is *not* divisible by p and any $u \equiv 1 \bmod p\mathbf{Z}_p$, 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 2.7. 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 $\mathbf{Z}_p$. 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 $\mathbf{Z}_7$.

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 mod p, which means u is a square in $\mathbf{Q}_p$. 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 2.8. 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 $\mathbf{Z}_5$:

$$\begin{aligned} \omega_2 &= 2 + 5 + 2\cdot 5^2 + 5^3 + 3\cdot 5^4 + 4\cdot 5^5 + 2\cdot 5^6 + 3\cdot 5^7 + \cdots, \\ \omega_3 &= 3 + 3\cdot 5 + 2\cdot 5^2 + 3\cdot 5^3 + 5^4 + 2\cdot 5^6 + 5^7 + \cdots. \end{aligned}$$

The numbers ω_k for $0 \leq k \leq p-1$ 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-adic $(p-1)$th roots of unity.

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 any 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 2.8 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 any nontrivial p-th power roots 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.)

We first consider odd p and suppose $\zeta^p = 1$ in $\mathbf{Z}_p^\times$. Since $\zeta^p \equiv \zeta \bmod p\mathbf{Z}_p$, we have $\zeta \equiv 1 \bmod p\mathbf{Z}_p$. For the polynomial $f(X) = X^p - 1$ we have $|f'(\zeta)|_p = |p\zeta^{p-1}|_p = 1/p$, so the uniqueness in Hensel's lemma implies that the ball

$$\{x \in \mathbf{Q}_p : |x - \zeta|_p < |f'(\zeta)|_p\} = \{x \in \mathbf{Q}_p : |x - \zeta|_p \leq 1/p^2\} = \zeta + p^2\mathbf{Z}_p$$

contains no pth root of unity except for ζ. We will now show that $\zeta \equiv 1 \bmod p^2\mathbf{Z}_p$, so 1 is in that ball and therefore $\zeta = 1$.

Write $\zeta = 1 + py$, where $y \in \mathbf{Z}_p$. Then

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

For $2 \leq k \leq p-1$, $\binom{p}{k}$ is divisible by p, so all terms in the sum over $2 \leq k \leq p-1$ are divisible by p^3. The last term $(py)^p$ is also divisible by p^3 (since $p \geq 3$). Therefore if we reduce the above equation modulo p^3 we get

$$1 \equiv 1 + p^2 y \bmod p^3 \Longrightarrow 0 \equiv p^2 y \bmod p^3.$$

Therefore y is divisible by p, so $\zeta \equiv 1 \bmod p^2$ and this forces $\zeta = 1$.

Now we turn to $p = 2$. We want to show the only 4th roots of unity in $\mathbf{Z}_2^\times$ are ± 1. This won't use Hensel's lemma. 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 $1 \not\equiv -1 \bmod 4\mathbf{Z}_2$.

For any 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 $p = 2$. □

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 $f(\alpha) = 0$ 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 any $u \in \mathbf{Z}_2^\times$ satisfying $u \equiv 1 \bmod 8\mathbf{Z}_2$ is a square in $\mathbf{Z}_2^\times$, let $f(X) = X^2 - u$ and use $a = 1$ in Theorem 4.1. We have $|f(1)|_2 = |1 - u|_2 \leq 1/8$ 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 $\mathbf{Z}_2$. □

Theorem 4.5. *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 4.5 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 4.5 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, any p-adic integer that is congruent to 1 mod p^3 is a pth power (see the first paragraph of this proof again). Therefore $u/(a^p(1 + pc)^p)$ is a pth power, so u is a pth power. □

Remark 4.6. Hensel's lemma in Theorem 4.1 guarantees a unique lift of a root mod p to a root in $\mathbf{Z}_p$ under weaker conditions than for Hensel's lemma in Theorem 2.1, but it does not guarantee a unique lift to a root mod p^n, unlike for Theorem 2.1 (Remark 2.3). For example, if $p > 2$ then 1 mod p lifts to just one root of $x^p = 1$ in $\mathbf{Z}_p$, namely $x = 1$, but for $n \geq 2$ it lifts to p roots of $x^p \equiv 1 \bmod p^n$: $x \equiv 1 + p^{n-1}b \bmod p^n$ for $0 \leq b \leq p - 1$. (What happens if $p = 2$?) This is *consistent* with a unique lift to a root of $x^p = 1$ in $\mathbf{Z}_p$ because the multiple roots converge in $\mathbf{Z}_p$ to the same number: $1 + p^{n-1}b \to 1$ in $\mathbf{Z}_p$ 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 [3, 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 is

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

for some $g(X,Y) \in \mathbf{Z}_p[X,Y]$, which is derived in exactly the same way as in the proof of Theorem 2.1. Then

$$x, y \in \mathbf{Z}_p \Longrightarrow \boxed{f(x+y) = f(x) + f'(x)y + zy^2, \text{ where } z \in \mathbf{Z}_p.} \tag{5.2}$$

The second polynomial identity we need is that for any $F(X) \in \mathbf{Z}_p[X]$,

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

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 any $i \geq 1$. 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

$$x, y \in \mathbf{Z}_p \Longrightarrow |F(x) - F(y)|_p = |x-y|_p |G(x,y)|_p \leq |x-y|_p. \tag{5.3}$$

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 $t < 1$, so by (5.3) with $F(X) = f'(X)$,

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

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

To prove (iii) for $n+1$, we use (5.2) with $x = a_n$ 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,

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

This completes the induction.

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

$$|a_{n+1} - a_n|_p = \left|\frac{f(a_n)}{f'(a_n)}\right|_p = \frac{|f(a_n)|_p}{|f'(a_1)|_p} \leq |f'(a_1)|_p t^{2^{n-1}}, \tag{5.4}$$

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 $|\alpha - a|_p = |f(a)/f'(a)|_p$, we will show $|a_n - a|_p = |f(a)/f'(a)|_p$ for all $n \geq 2$. (Then just let $n \to \infty$.) This is clear for $n = 2$ from the definition of a_2 in terms of $a_1 = a$. For any $n \geq 2$ we have from (5.4)

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

Therefore if $|a_n - a|_p = |f(a)/f'(a)|_p$ we have $|a_{n+1} - a_n|_p < |a_n - a|_p$, 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 $|\beta - a|_p < |f'(a)|_p$. 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 (5.2),

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

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 $h = 0$, *i.e.*, $\beta = \alpha$. □

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 $m \to \infty$,

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

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 $\mathbf{Z}_3$:

$$\begin{aligned} \alpha_1 &= 1 + 3 + 3^2 + 2 \cdot 3^4 + 2 \cdot 3^7 + 3^8 + 3^9 + \cdots, \\ \alpha_2 &= 2 + 3 + 3^2 + 2 \cdot 3^3 + 2 \cdot 3^5 + 2 \cdot 3^6 + 3^8 + 3^9 + \cdots. \end{aligned}$$

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 $\alpha \equiv a_1 \bmod 3$. Thus $a_n \to \alpha_1$. 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 α_1 up through terms including 3^7 (the first 8 digits). The estimate in (5.5) says $|\alpha_1 - a_n|_3 \leq |f'(a_1)|_3 (1/3)^{2^{n-1}} = (1/3)^{2^{n-1}}$ 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 $\mathbf{Z}_2$:

$$\begin{aligned} \alpha_1 &= 1 + 2^3 + 2^5 + 2^6 + 2^7 + 2^9 + \cdots, \\ \alpha_2 &= 1 + 2 + 2^2 + 2^4 + 2^8 + \cdots. \end{aligned}$$

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 α_1 or α_2. 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 $\alpha = \alpha_1$, and if $a \equiv 3 \bmod 4$ then $\alpha = \alpha_2$. By (5.5), $|\alpha - a_n|_2 \leq |f'(a)|_2(|f(a)/f'(a)^2|_2)^{2^{n-1}} = (1/2)(4|a^2 - 17|_2)^{2^{n-1}}$. For a few choices of a, this inequality is an equality for $1 \leq n \leq 10$:

- When $a = 1$, $|\alpha_1 - a_n|_2 = (1/2)^{2^n+1}$ for $1 \leq n \leq 10$.
- When $a = 3$, $|\alpha_2 - a_n|_2 = (1/2)^{2^{n-1}+1}$ for $1 \leq n \leq 10$.
- When $a = 5$, $|\alpha_1 - a_n|_2 = (1/2)^{2^{n-1}+1}$ 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 $|f(2)|_3 = 1/3 < |f'(2)|_3^2$, the Newton 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 *exact same* 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

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 $|f(a)|_p < |f'(a)|_p^2$. We will construct a different sequence converging to a root of $f(X)$ by iterating $\varphi(x) = x - f(x)/f'(a)$ instead, where again $|f(a)|_p < |f'(a)|_p^2$. The denominator here 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. The condition $\varphi(\alpha) = \alpha$ says $\alpha - f(\alpha)/f'(a) = \alpha$, so $f(\alpha) = 0$, and in this way we get 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 of Theorem 4.1, go to the next section.

Proof. Pick a closed ball $\overline{B}_a(r) = \{x \in \mathbf{Q}_p : |x - a|_p \leq r\}$ where $r \leq 1$ is to be determined. Since $r \leq 1$ and $a \in \mathbf{Z}_p$, $\overline{B}_a(r) \subset \mathbf{Z}_p$. 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 any x and y in $\mathbf{Q}_p$,

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

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 any positive $\lambda < 1$,

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

If we can find $\lambda < 1$, 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 any $\lambda \in (0,1)$, 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

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

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 $|f'(\alpha)|_p = |f'(a)|_p$.

$|\alpha - a|_p = |f(a)/f'(a)|_p$: We have $|a_2 - a|_p = |a_2 - a_1|_p = |f(a_1)/f'(a)|_p = |f(a)/f'(a)|_p$, and for any 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 $n \to \infty$.

α is the unique root of $f(X)$ 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.

$|f'(\alpha)|_p = |f'(a)|_p$: 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 (5.3) 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 $|f'(\alpha)|_p = |f'(a)|_p$. □

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, (5.5) 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

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

Since the upper bound in (6.6) goes to 0 much faster than the upper bound in (6.7), we can *anticipate* that $|\alpha - a_n|_p$ would shrink to 0 faster when $\{a_n\}$ is produced from Newton's recursion compared to the contraction mapping recursion. In particular, when $|f(a)/f'(a)^2|_p = 1/p$, the decay bound in (6.6) goes to 0 like $(1/p)^{2^{n-1}}$ while in (6.7) the decay bound goes to 0 like $(1/p)^n$, so we expect there to be a doubling of correct p-adic digits at each step in Newton's recursion while we get just one new correct p-adic digit at each step by the contraction mapping recursion. Let's see an example.

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 $1 \leq n \leq 10$. 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 $1 \leq n \leq 10$, which is much smaller than $(1/3)^n$.

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

In Theorem 4.1 we have $|f'(\alpha)|_p = |f'(a)|_p \neq 0$, so Hensel's lemma produces a simple root 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 more or less 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. We have $|f'(\alpha) - f'(a)|_p \leq |\alpha - a|_p < |f'(\alpha)|_p$, so $|f'(a)|_p = |f'(\alpha)|_p$. Also

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

for some $z \in \mathbf{Z}_p$. Each term on the right side has absolute value less than $|f'(\alpha)|_p^2$, so also $|f(a)|_p < |f'(\alpha)|_p^2 = |f'(a)|_p^2$. □

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

Both of our proofs of Hensel's lemma, as well as the proof of Theorem 7.1, work in any field complete with respect to an absolute value satisfying the strong triangle inequality.

Theorem 8.1. *Let K be a field that is complete with respect to an absolute value such that $|x + y| \leq \max(|x|, |y|)$ for all x and y in K. Set $\mathfrak{o} = \{x \in K : |x| \leq 1\}$. If a polynomial $f(X)$ has coefficients in $\mathfrak{o}$ and some $a \in \mathfrak{o}$ satisfies*

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

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

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

For a version of Hensel's lemma where $\mathbf{Z}_p$ is replaced by any complete local ring (possibly not an integral domain, so division is more delicate), see [2, Theorem 7.3]. For a form of Hensel's lemma dealing with zeros of several polynomials (or power series) in several variables, see [1, Chap. III, §4.3].

References

[1] N. Bourbaki, "Commutative Algebra," Springer-Verlag, New York, 1989.
[2] D. Eisenbud, "Commutative Algebra with a View to Algebraic Geometry," Springer-Verlag, New York, 1995.
[3] S. Lang, "Algebraic Number Theory," 3rd ed., Springer-Verlag, New York, 1994.