Numerical analysis meets number theory: using rootfinding methods to calculate inverses mod p^n [*]

Michael Knapp[†] and Christos Xenophontos

Mathematical Sciences Department

Loyola College

4501 N. Charles Street

Baltimore, MD 21210

This article is dedicated to the memory of George Mackiw

June 17, 2004

1 Introduction

In this article we explore a very interesting application of tools from numerical analysis to number theory. As the title suggests, we will see how one can use classical rootfinding methods, such as Newton's method which is typically studied in an undergraduate numerical analysis course, to calculate the reciprocal of an integer modulo p^n, where p is

[*] Technical Report 2004-02, Mathematical Sciences Department, Loyola College, 4501 N. Charles Street, Baltimore, MD 21210.

[†] The first author is partially supported by NSF grant DMS-0344082.

a prime number. We first encountered this idea in [1], where Newton's method was used to find the reciprocal of a finite segment p-adic number (also referred to as Hensel code; see [1] for more details).

What do we mean by the reciprocal of an integer modulo p^n? We define a fraction modulo p^n as follows. If a, b and α are integers and a is not divisible by p, then we say that

$$\alpha \equiv \frac{b}{a} \pmod{p^n} \qquad \text{if} \qquad a\alpha \equiv b \pmod{p^n}.$$

Using this definition, the reciprocal $\frac{1}{a}$ of an integer a modulo p^n is a solution of the congruence $ax \equiv 1 \pmod{p^n}$. In other words, it is an inverse of a modulo p^n.

The idea of using Newton's method to perform division dates back to the early days of computing, since one can actually approximate the reciprocal of a number by performing only the operations of multiplication and addition. The idea behind iterative rootfinding methods such as Newton's method is as follows. Suppose that we have a function $f(x)$ for which we wish to find a zero in an interval $[a, b]$. To accomplish this, let $x_0 \in [a, b]$ be an initial guess for the zero, and let $g(x)$ be an iteration function. Then we calculate further approximations through the formula

$$x_{n+1} = g(x_n), \quad n = 0, 1, \ldots. \tag{1}$$

If the initial guess x_0 and the iteration function $g(x)$ are suitably chosen, then the sequence $x_0, x_1, x_2, \ldots$ should converge to a zero of $f(x)$ in $[a, b]$.

If this does in fact occur, then we can talk about the rate at which the sequence converges to a zero of $f(x)$. Roughly speaking, if the rate of convergence of a method is m (i.e. the method converges with order m), then after each iteration the number of correct significant digits in the approximation increases by a factor of approximately m.

For example, if our approximation converges quadratically (i.e. with order 2), then the number of correct significant digits approximately doubles with each iteration.

Now what does all this have to do with congruences modulo p^n? In this situation, the role of significant digits will be played by smaller powers of p. We will start with an inverse of a modulo p, and then find inverses of a modulo higher powers of p until we have an inverse modulo p^n. The connection between digits and powers of p can be found by looking at the base p representation of numbers. It is well-known that any positive integer α has a base p representation

$$\alpha = c_0 + c_1 p + c_2 p^2 + \cdots + c_k p^k,$$

where $0 \leq c_i \leq p-1$ for each i. Each coefficient c_i is a digit in the base p representation of α. To find a number which is congruent to α modulo p^r, we can simply truncate the base p expansion of α after the first r digits. Thus we may think of an inverse of a modulo p^r as giving the correct first r digits of an inverse of a modulo p^n.

We note here that our Theorems 1 and 3 ahead are not truly original. In fact, they are simple consequences of more powerful theorems about iterative methods. We will briefly touch upon this point in the last section.

Finally, we would like to thank Robert Benedetto and Lisa Oberbroeckling for some very helpful discussions about the rates of convergence of p-adic iterative methods.

2 Newton's method (for division mod p^n)

The iteration function for Newton's method is $g(x) = x - f(x)/f'(x)$, whence equation (1) becomes

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \ , \ \ n = 0, 1, \ldots . \qquad (2)$$

Under suitable assumptions on f, f' and x_0 (see for example Theorem 3.2 on page 100 of [2]), the above iteration converges to a zero of $f(x)$ in $[a, b]$ at a *quadratic* rate. So to calculate $\frac{1}{a}$, we let $f(x) = \frac{1}{x} - a$, and solve $f(x) = 0$ using Newton's Method. In this case iteration (2) becomes

$$x_{n+1} = x_n(2 - ax_n) \ , \ \ n = 0, 1, \ldots. \tag{3}$$

Like Newton's method for real numbers, we can show that Newton's method also converges quadratically for congruences. This is proven in the following theorem. Although we're using different language, this is essentially the same as the main theorem of [1]. We believe that our proof is simpler, however.

Theorem 1 *Suppose that x_n is an inverse of a modulo p^r. Then x_{n+1} given by (3) is an inverse of a modulo p^{2r}.*

To prove this, we know that $ax_n \equiv 1 \pmod{p^r}$, and therefore can write $ax_n = sp^r + 1$ for some integer s. Then we have

$$\begin{aligned} ax_{n+1} &= ax_n(2 - ax_n) \\ &= 2ax_n - (ax_n)^2 \\ &= 2(sp^r + 1) - (sp^r + 1)^2 \\ &= -s^2p^{2r} + 1 \\ &\equiv 1 \pmod{p^{2r}}. \end{aligned}$$

Hence x_{n+1} is an inverse of a modulo p^{2r}, as desired.

So if we can find an inverse of a modulo p to use as an initial guess, then we can use Newton's method to find an inverse of a modulo p^n. But how can we find this initial guess? If the prime p is small, then we can often find an inverse of a modulo p by inspection. If p is larger, then we can use the following theorem, known as Euler's theorem.

Theorem 2 (Euler) *If $(a, m) = 1$, then*

$$a^{\phi(m)} \equiv 1 \pmod{m}.$$

In this theorem, (a, m) represents the greatest common divisor of a and m, and $\phi(m)$ is Euler's phi-function, defined as the number of integers t such that $1 \leq t \leq m$ and $(t, m) = 1$. If p is a prime number and $(a, p) = 1$, then a consequence of Euler's theorem is that a^{p-2} is an inverse of a modulo p.

We can evaluate a^{p-2} modulo p by the technique of repeated squaring. For example, to find the inverse of 29 modulo 53, we need to evaluate 29^{51} modulo 53. To do this, we see that

$$\begin{aligned} 29^1 \equiv & & 29 & \pmod{53} \\ 29^2 \equiv & & 46 & \pmod{53} \\ 29^4 \equiv & \; 46^2 \equiv & 49 & \pmod{53} \\ 29^8 \equiv & \; 49^2 \equiv & 16 & \pmod{53} \\ 29^{16} \equiv & \; 16^2 \equiv & 44 & \pmod{53} \\ 29^{32} \equiv & \; 44^2 \equiv & 28 & \pmod{53}. \end{aligned}$$

Thus we have

$$29^{51} = 29^{32} \cdot 29^{16} \cdot 29^2 \cdot 29^1 \equiv (28)(44)(46)(29) \equiv 11 \pmod{53},$$

and so $\frac{1}{29} \equiv 11 \pmod{53}$.

Exercise 1 *Use Euler's theorem to find the reciprocal of 12 modulo 41.*

Exercise 2 *Use Euler's Theorem to find the reciprocal of 3 modulo 7. Note that for a small prime like 7 it is just as easy to find an answer by inspection as it is to use Euler's theorem.*

Let us now illustrate the use of Newton's method in the context of the present paper via two examples.

Example 1 *Let* $p = 5$*,* $a = 3$ *and* $n = 8$*. We wish to find an integer congruent to* $\frac{1}{3}$ *modulo* 5^8 *using the Newton iteration (3). As our initial guess, we choose* $x_0 = 2$ *since* $3(2) \equiv 1 \pmod 5$ *and so* $2 \equiv \frac{1}{3} \pmod 5$*. Then we have from (3) that*

$$x_1 = 2\,(2 - 3 \cdot 2) = -8 \equiv 17 = 2 + 3(5) \pmod{5^2}.$$

Note that $3(17) = 51 \equiv 1 \pmod{5^2}$*, and so* $17 \equiv \frac{1}{3} \pmod{5^2}$ *as indicated by Theorem 1. Iterating twice more gives us*

$$\begin{aligned} x_2 &\equiv 417 = 2 + 3(5) + 1(5)^2 + 3(5)^3 \pmod{5^4} \\ x_3 &\equiv 260417 = 2 + 3(5) + 1(5)^2 + 3(5)^3 + 1(5)^4 + 3(5)^5 + 1(5)^6 + 3(5)^7 \pmod{5^8}, \end{aligned}$$

and so we see that $260417 \equiv \frac{1}{3} \pmod{5^8}$*.*

Example 2 *Let* $p = 53, a = 29$ *and* $n = 16$*. From our discussion following Euler's theorem, we know that* $\frac{1}{29} \equiv 11 \pmod{53}$*, hence we choose as our initial guess* $x_0 = 11$*. Then, just like in the previous example, we have from (3) that*

$$\begin{aligned} x_1 &= 11(2 - 29 \cdot 11) = -3487 \equiv 2131 = 11 + 40(53) \pmod{53^2} \\ x_2 &\equiv 2448770 = 11 + 40(53) + 23(53)^2 + 16(53)^3 \pmod{53^4} \\ x_3 &\equiv 12881315257523 = \\ &= 11 + 40(53) + 23(53)^2 + 16(53)^3 + 7(53)^4 + 9(53)^5 + \\ &\quad +51(53)^6 + 10(53)^7 \pmod{53^8} \\ x_4 &\equiv 2940617900089909330991489347 = \\ &= 11 + 40(53) + 23(53)^2 + 16(53)^3 + 7(53)^4 + 9(53)^5 + 51(53)^6 + 10(53)^7 + \\ &\quad +40(53)^8 + 23(53)^9 + 16(53)^{10} + 7(53)^{11} + 9(53)^{12} + 51(53)^{13} + \\ &\quad +10(53)^{14} + 40(53)^{15} \pmod{53^{16}} \end{aligned}$$

and so we see that $2940617900089909330991489347 = \frac{1}{29} \pmod{53^{16}}$*.*

3 The secant method (for division mod p^n)

Another rootfinding method introduced in a typical numerical analysis course is the secant method, whose iteration is given by

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})} \ , \ \ n = 1, 2, \ldots . \tag{4}$$

Note that we now need *two* initial guesses x_0 and x_1, but we no longer need the derivative of $f(x)$. Since "there is no such thing as a free lunch", the trade-off here is that the order of convergence drops down to $\phi = (1 + \sqrt{5})/2 \approx 1.618$, which one easily recognizes as the *golden ratio.* In fact, we will show that after each iteration, instead of doubling (like in Newton's method), the number of correct digits increases by a factor of approximately ϕ. For our function $f(x) = \frac{1}{x} - a$, equation (4) becomes

$$x_{n+1} = x_n + x_{n-1} - a x_n x_{n-1}. \tag{5}$$

To establish the rate of convergence when using the secant method for congruences we have the following theorem.

Theorem 3 *Suppose that* $x_{n-1} \equiv \frac{1}{a} \pmod{p^\alpha}$ *and that* $x_n \equiv \frac{1}{a} \pmod{p^\beta}$*. Then, with* x_{n+1} *given by (5), we have* $x_{n+1} \equiv \frac{1}{a} \pmod{p^{\alpha+\beta}}$*.*

To prove this, note that since we have $a x_{n-1} \equiv 1 \pmod{p^\alpha}$ and $a x_n \equiv 1 \pmod{p^\beta}$, there exist integers s and t such that

$$a x_{n-1} = s p^\alpha + 1 \qquad \text{and} \qquad a x_n = t p^\beta + 1.$$

Then we have

$$\begin{aligned} a x_{n+1} &= a x_n + a x_{n-1} - (a x_n)(a x_{n-1}) \\ &= (s p^\alpha + 1) + (t p^\beta + 1) - (s p^\alpha + 1)(t p^\beta + 1) \\ &= -st p^{\alpha+\beta} + 1 \\ &\equiv 1 \pmod{p^{\alpha+\beta}}, \end{aligned}$$

as desired.

So if x_1 and x_2 are inverses of a modulo p^1, we can show by induction that x_n is the inverse of a modulo p^{F_n}, where F_n is the n^{th} Fibonacci number. Since it is well-known that F_n gets closer and closer to $\phi^n/\sqrt{5}$ as n gets large, we find that the secant method has order of convergence ϕ. Notice that although the general formula (4) for the iteration requires that our initial guesses x_0 and x_1 be different, this is not required in either formula (5) or the theorem. Thus we can take x_0 and x_1 to both be inverses of a modulo p, and in fact can even take them to be the same number. Let us illustrate the above ideas via an example.

Example 3 *Let $p = 7, a = 5$ and $n = 8$, i.e. we wish to find an integer congruent to $\frac{1}{5}$ modulo 7^8 using the iteration (5). We choose $x_0 = x_1 = 3$, since $5 \cdot 3 \equiv 1 \pmod{7}$. We have from (5)*

$$\begin{aligned}
x_2 &= 3 + 3 - 5 \cdot 3 \cdot 3 \equiv 10 = 3 + 1(7) \pmod{7^2} \\
x_3 &= 10 + 3 - 5 \cdot 10 \cdot 3 \equiv 206 = 3 + 1(7) + 4(7)^2 \pmod{7^3} \\
x_4 &\equiv 6723 = 3 + 1(7) + 4(7)^2 + 5(7)^3 + 2(7)^4 \pmod{7^5} \\
x_5 &\equiv 4611841 = \\
&= 3 + 1(7) + 4(7)^2 + 5(7)^3 + 2(7)^4 + 1(7)^5 + 4(7)^6 + 5(7)^7 \pmod{7^8}
\end{aligned}$$

and so we see that $4611841 \equiv \frac{1}{5} \pmod{7^8}$.

4 Fixed point iteration and high order convergent methods

It turns out that the general iteration formula (1) defines a larger class of iterative methods, called *fixed point* methods: instead of solving $f(x) = 0$ we solve $g(x) = x$

(for a suitably chosen $g(x)$); a point α such that $g(\alpha) = \alpha$ is called a *fixed point* of g. The advantage of this approach is that it can be easily generalized to higher dimensions and analyzed using a plethora of famous fixed point theorems. Newton's method is a special case of a fixed point iteration, as can be readily seen by equation (2). Under suitable assumptions on g (see Theorems 3.5 and 3.7 on pages 121–124 of [2]), iteration (1) converges to α for any initial guess x_0 sufficiently close to α, at a rate r such that

$$g'(\alpha) = g''(\alpha) = g'''(\alpha) = \cdots = g^{(r-1)}(\alpha) = 0 \quad \text{but} \quad g^{(r)}(\alpha) \neq 0. \tag{6}$$

With this in mind, one can construct iteration functions g such that (6) holds for some r, hence obtaining a method which converges at that rate.

In our case, Newton's method can be written as a fixed point iteration with $g(x) = x(2 - ax)$. Since we already know that this is a quadratically convergent method, we expect that (i) $g\left(\frac{1}{a}\right) = \frac{1}{a}$, (ii) $g'\left(\frac{1}{a}\right) = 0$ and (iii) $g''\left(\frac{1}{a}\right) \neq 0$.

Exercise 3 *Verify that* (i)–(iii) *above all hold.*

Now suppose we wanted to construct a method for finding the zero of $f(x) = \frac{1}{x} - a$, with a higher convergence rate. To this end, let

$$g(x) = x(1 + Y(x)) \tag{7}$$

and note that if $Y(x) = 1 - ax$ then we retrieve the fixed point iteration corresponding to Newton's Method (i.e. equation (3)). Note that with this choice of $Y(x)$ we have

$$Y\left(\frac{1}{a}\right) = 0 \quad \text{and} \quad Y'\left(\frac{1}{a}\right) = -a. \tag{8}$$

To construct a method with, say, a *cubic* convergence rate we need to choose an appropriate $Y(x)$ such that (8) holds, and in addition (6) holds with $r = 3$. The choice

$$Y(x) = y(x)(1 + y(x)) \tag{9}$$

where $y(x) = 1 - ax$ indeed satisfies these criteria.

Exercise 4 *Show that with $Y(x)$ given by* (9)*, the iteration function $g(x)$ defined by* (7) *satisfies $g\left(\frac{1}{a}\right) = \frac{1}{a}, g'\left(\frac{1}{a}\right) = g''\left(\frac{1}{a}\right) = 0$, and $g'''\left(\frac{1}{a}\right) \neq 0$.*

Combining (7) and (9), we see that the iteration now becomes

$$\begin{aligned} x_{n+1} &= x_n(1+Y(x_n)) \\ &= x_n\left[1+y(x_n)(1+y(x_n))\right] \\ &= x_n\left[1+(1-ax_n)(2-ax_n))\right]. \end{aligned} \tag{10}$$

The following example illustrates the use of iteration (10).

Example 4 *As in Example 1, let $p = 5$, $a = 3$ and $n = 8$. We wish to find an integer congruent to $\frac{1}{3}$ modulo 5^8 using the iteration (10). We expect that 2 iterations will suffice here, since the method converges cubically. Indeed, with $x_0 = 2$,*

$$\begin{aligned} x_1 &= 2\left[1+(1-3\cdot 2)(2-3\cdot 2)\right] = 42 \equiv 42 = 2+3(5)+1(5)^2 \pmod{5^3} \\ x_2 &= 42\left[1+(1-3\cdot 42)(2-3\cdot 42)\right] = 651042 \equiv 651042 \pmod{5^9}. \end{aligned}$$

Note that since 651042 is an inverse of 3 modulo 5^9, it is also an inverse of 3 modulo 5^8. Noting that $651042 \equiv 260417 \pmod{5^8}$, we see that

$$260417 = 2+3(5)+1(5)^2+3(5)^3+1(5)^4+3(5)^5+1(5)^6+3(5)^7$$

is the smallest inverse of 3 modulo 5^8.

Exercise 5 *For the cubically convergent iteration (10), prove an analog of Theorem 1, showing that if x_n is an inverse of a modulo p^r, then x_{n+1} given by (10) is an inverse of a modulo p^{3r}.*

It turns out that one can define a method with order of convergence r, for any r, if the choice

$$Y(x) = \underbrace{y(x)\left[1+y(x)\left[1+\cdots(1+y(x))\right]\right]}_{r-1 \text{ terms}} \tag{11}$$

is made in (7). For example, for order 4 convergence we should choose

$$Y(x) = y(x)\left[1 + y(x)(1 + y(x))\right].$$

5 Epilogue

In closing, we briefly mention some "big picture" issues which the reader may be curious about. First, we note that for all the rootfinding methods we have used, the rate of convergence for congruences has been the same as the rate of convergence over the real numbers. Is this an accident, or is something deeper going on? Also, for our problems about congruences, our methods are guaranteed to give us an exact inverse after a finite number of steps, whereas when used over the real numbers, these rootfinding methods will only give us approximations of the inverse without ever equaling the inverse exactly. Why does this difference occur? And what would happen in the congruence problem if we continued making more and more iterations resulting in inverses modulo higher and higher powers of p?

The answers to these questions lie in the realm of p-adic integers. Naively, p-adic integers are what you get if you look at congruences modulo p^n for higher and higher values of n, and in fact let n approach infinity. So we can naively think of p-adic integers as "congruences modulo p^∞." We've seen that when we write numbers in base p, everything is congruent modulo p^n to an n-digit number. By analogy, numbers "modulo p^∞" should have infinitely many digits. So a p-adic number looks like

$$\alpha = c_0 + c_1 p + c_2 p^2 + c_3 p^3 + \cdots = \sum_{i=0}^{\infty} c_i p^i.$$

Although numbers like this strike us as bizarre, the technical definition of p-adic numbers shows how we can make sense out of them. So we can think of the inverses of a modulo powers of p as approximations of the p-adic inverse. In this sense, our iterative methods are once again giving us better and better approximations rather than exact answers. A good place to begin learning about p-adic numbers is [3].

Finally, what about the order of convergence? If you look at the most general theorems about when iterative methods converge, they are not theorems specifically about real numbers but rather theorems about properties of complete metric spaces [4]. It turns out that because of the way the set of p-adic integers is defined, it forms a complete metric space. Thus, the same theorems as for real numbers guarantee us that our iterations will converge. Although it is beyond the scope of this article, theorems can be proven about the rates of convergence of the iteration. In particular, it can be proven that if the iteration function is a polynomial with integer coefficients and the iteration converges for both the real numbers and the p-adic integers, then the rates of convergence for the two number systems will always be the same.

References

[1] E. V. Krishnamurthy and V. K. Murthy, Fast Iterative Division of p-adic Numbers, *IEEE Transactions on Computers*, **32** (1983), 396–398.

[2] J. Epperson, *An Introduction to Numerical Methods and Analysis*, Wiley and Sons, 2002.

[3] F. Gouvêa, *P-adic Numbers: An Introduction*, Springer-Verlag, 1997.

[4] W. Rudin, *Principles of Mathematical Analysis*, McGraw Hill, 1976.