Can we avoid tests for zero in fast elliptic-curve arithmetic?

Daniel J. Bernstein *

Department of Mathematics, Statistics, and Computer Science (M/C 249)
The University of Illinois at Chicago, Chicago, IL 60607–7045
djb@cr.yp.to

Abstract. This paper analyzes the exact extent to which 0 and ∞ cause trouble in Montgomery's fast branchless formulas for x-coordinate scalar multiplication on elliptic curves of the form $by^2 = x^3 + ax^2 + x$. The analysis shows that some multiplications and branches can be eliminated from elliptic-curve primality proofs and from elliptic-curve cryptography.

1 Introduction

Define sequences $(x_1, x_2, \ldots)$ and $(z_1, z_2, \ldots)$ recursively, starting from x_1, z_1, a, by the equations

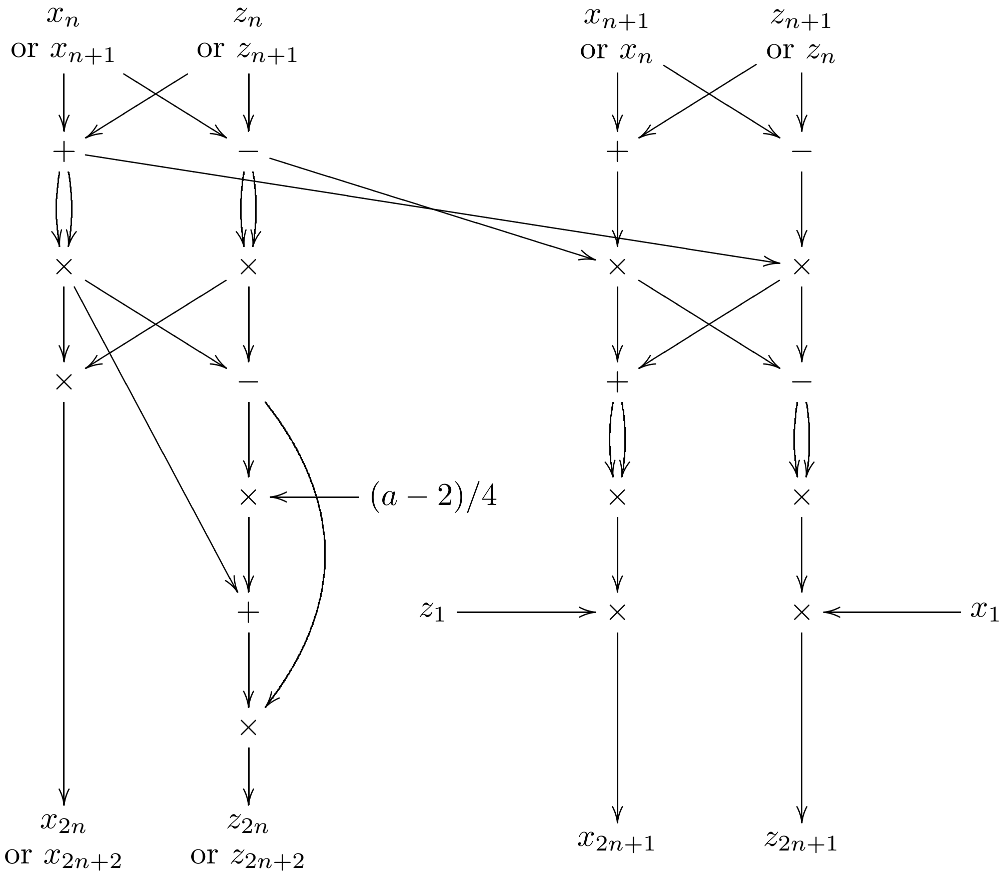
$$\begin{aligned}
x_{2n} &= (x_n^2 - z_n^2)^2 = (x_n - z_n)^2(x_n + z_n)^2, \\
z_{2n} &= 4x_n z_n(x_n^2 + ax_n z_n + z_n^2) \\
&= ((x_n + z_n)^2 - (x_n - z_n)^2) \\
&\qquad \cdot \left((x_n + z_n)^2 + \frac{a-2}{4}((x_n + z_n)^2 - (x_n - z_n)^2)\right), \\
x_{2n+1} &= 4(x_n x_{n+1} - z_n z_{n+1})^2 z_1 \\
&= ((x_n - z_n)(x_{n+1} + z_{n+1}) + (x_n + z_n)(x_{n+1} - z_{n+1}))^2 z_1, \\
z_{2n+1} &= 4(x_n z_{n+1} - z_n x_{n+1})^2 x_1 \\
&= ((x_n - z_n)(x_{n+1} + z_{n+1}) - (x_n + z_n)(x_{n+1} - z_{n+1}))^2 x_1.
\end{aligned}$$

It is well known—and, unfortunately, not always true—that these sequences compute scalar multiples on an elliptic curve: specifically, that $(x_n/z_n, \ldots)$ is the nth multiple of the point $(x_1/z_1, \ldots)$ on the curve $by^2 = x^3 + ax^2 + x$.

This paper explains exactly what *is* true. Section 2 reviews the standard definition of scalar multiplication on an elliptic curve; Section 4 analyzes the connections between x_n, z_n, and nth multiples; Theorem 4.3 explains how x_n and z_n actually relate to the nth multiple of $(x_1/z_1, \ldots)$ on the curve $by^2 = x^3 + ax^2 + x$. **WARNING: This is an early draft, not yet checked.**

---

* The author was supported by the National Science Foundation under grant DMS–0140542 and by the Alfred P. Sloan Foundation. Date of this document: 2006.07.26. Permanent ID of this document: 3a823a6593bf3c4e1ffa27186c6c3191.



These sequences are of interest in a wide variety of applications of elliptic-curve scalar multiplication, including elliptic-curve cryptography (ECC), elliptic-curve primality proving (ECPP), and the elliptic-curve factorization method (ECM). For example, my ECC speed records in [2] use these sequences. The sequences were introduced by Montgomery in [12, Section 10.3.1] twenty years ago to speed up ECM. The point is that computing $(x_n, z_n, x_{n+1}, z_{n+1})$ takes just

- 4 squarings,
- 1 multiplication by $(a-2)/4$, which is small in most applications,
- 1 multiplication by z_1, which is small in most applications,
- 1 multiplication by x_1, which is small in many applications,
- 4 more multiplications,
- 4 additions, and
- 4 subtractions

for each bit of n, as shown in the data-flow diagram above. These are not always the smallest known costs for elliptic-curve scalar multiplication—see, e.g., [3]—but they have never been improved upon by more than a small percentage.

Sections 5, 6, and 7 explain how to use x_n and z_n to replace nth multiples in various applications of elliptic-curve scalar multiplication. The bottom line is that this paper speeds up elliptic-curve primality proofs and elliptic-curve cryptography by eliminating various multiplications and branches.

2 Elliptic curves

Fix a field k not of characteristic 2, and fix $a, b \in k$ with $b(a^2 - 4) \neq 0$. This section reviews the standard definition of the group $E(k)$, where E is the elliptic curve $by^2 = x^3 + ax^2 + x$ over k.

Elliptic curves not of the form $by^2 = x^3 + ax^2 + x$ are outside the scope of this paper. The particular shape $by^2 = x^3 + ax^2 + x$ was highlighted by Montgomery in [12, Section 10.3.1], and is often called "Montgomery form."

Define $E(k)$ as the set $\{\infty\} \cup \{(x, y) \in k \times k : by^2 = x^3 + ax^2 + x\}$. Define a unary operation $-$ on $E(k)$ as follows: $-\infty = \infty$; $-(x, y) = (x, -y)$. Define a binary operation $+$ on $E(k)$ as follows:

- $\infty + \infty = \infty$.
- $\infty + (x, y) = (x, y)$.
- $(x, y) + \infty = (x, y)$.
- $(x, y) + (x, -y) = \infty$.
- If $y \neq 0$ then $(x, y) + (x, y) = (x'', y'')$ where $\lambda = (3x^2 + 2ax + 1)/2by$, $x'' = b\lambda^2 - a - 2x$, and $y'' = \lambda(x - x'') - y$.
- If $x' \neq x$ then $(x, y) + (x', y') = (x'', y'')$ where $\lambda = (y' - y)/(x' - x)$, $x'' = b\lambda^2 - a - x - x'$, and $y'' = \lambda(x - x'') - y$.

Standard (although lengthy) calculations show that $E(k)$ is a commutative group with ∞ as neutral element, $-$ as negation, and $+$ as addition.

3 Montgomery's x-coordinate formulas

Montgomery in [12, Section 10.3.1] presented some surprisingly simple formulas for the x-coordinates of sums of points on elliptic curves $by^2 = x^3 + ax^2 + x$. This section reviews Montgomery's formulas.

Define $E(k)$ as in Section 2. Define $X : E(k) \to \{\infty\} \cup k$ as follows: $X(x, y) = x$; $X(\infty) = \infty$.

Note that if $X(Q) = 0$ then $Q = (0, 0)$. Indeed, $Q = (0, y)$ for some $y \in k$ with $by^2 = 0^3 + a0^2 + 0 = 0$, i.e., with $y = 0$.

Theorem 3.1. *Let k be a field not of characteristic 2. Let a, b be elements of k with $b(a^2 - 4) \neq 0$. Define E as the elliptic curve $by^2 = x^3 + ax^2 + x$ over k. Let Q be an element of $E(k)$ with $2Q \neq \infty$. Then $X(Q)^3 + aX(Q)^2 + X(Q) \neq 0$ and*

$$X(2Q) = \frac{(X(Q)^2 - 1)^2}{4(X(Q)^3 + aX(Q)^2 + X(Q))}.$$

Proof. $Q \neq \infty$ so $Q = (x, y)$ for some $x, y \in k$ satisfying $by^2 = x^3 + ax^2 + x$. If $y = 0$ then $2Q = (x, 0) + (x, 0) = \infty$, contradiction. Thus $y \neq 0$, $x^3 + ax^2 + x \neq 0$, and $2Q = (b\lambda^2 - a - 2x, \ldots)$ where $\lambda = (3x^2 + 2ax + 1)/2by$. Consequently

$$\begin{aligned}
X(2Q) &= b\lambda^2 - a - 2x = b\frac{(3x^2 + 2ax + 1)^2}{4b^2y^2} - a - 2x \\
&= \frac{(3x^2 + 2ax + 1)^2}{4by^2} - a - 2x = \frac{(3x^2 + 2ax + 1)^2}{4(x^3 + ax^2 + x)} - a - 2x \\
&= \frac{(3x^2 + 2ax + 1)^2 - 4(x^3 + ax^2 + x)(2x + a)}{4(x^3 + ax^2 + x)} \\
&= \frac{9x^4{+}12ax^3{+}(4a^2{+}6)x^2{+}4ax{+}1 - 4(2x^4{+}3ax^3{+}(a^2{+}2)x^2{+}ax)}{4(x^3 + ax^2 + x)} \\
&= \frac{x^4 - 2x^2 + 1}{4(x^3 + ax^2 + x)} = \frac{(x^2 - 1)^2}{4(x^3 + ax^2 + x)}.
\end{aligned}$$

Finally $X(Q) = x$. □

Theorem 3.2. *Let k be a field not of characteristic 2. Let a, b be elements of k with $b(a^2 - 4) \neq 0$. Define E as the elliptic curve $by^2 = x^3 + ax^2 + x$ over k. Let Q, R be elements of $E(k)$ with $Q \neq \infty$, $R \neq \infty$, $Q - R \neq \infty$, and $Q + R \neq \infty$. Then $X(Q) \neq X(R)$ and*

$$X(Q + R)X(Q - R) = \frac{(X(Q)X(R) - 1)^2}{(X(Q) - X(R))^2}.$$

Proof. $Q \neq \infty$ so $Q = (x, y)$ for some $x, y \in k$ satisfying $by^2 = x^3 + ax^2 + x$; and $R \neq \infty$ so $R = (x', y')$ for some $x', y' \in k$ satisfying $b(y')^2 = (x')^3 + a(x')^2 + x'$.

Suppose that $x = x'$. Then $by^2 = b(y')^2$ so $y = \pm y$. If $y = y'$ then $Q = R$ so $Q - R = \infty$, contradiction. If $y = -y'$ then $Q = -R$ so $Q + R = \infty$, contradiction.

Thus $x \neq x'$, and $Q + R = (b\lambda^2 - a - x - x', \ldots)$ where $\lambda = (y' - y)/(x' - x)$. Consequently

$$\begin{aligned}
X(Q + R) &= b\lambda^2 - a - x - x' = b\frac{(y' - y)^2}{(x' - x)^2} - a - x - x' \\
&= \frac{b(y')^2 + by^2 - 2byy'}{(x' - x)^2} - a - x - x' \\
&= \frac{(x')^3{+}a(x')^2{+}x' + x^3{+}ax^2{+}x - 2byy' - (a + x' + x)(x'{-}x)^2}{(x' - x)^2} \\
&= \frac{(x')^3 + x' + x^3 + x + 2axx' - 2byy' - (x' + x)(x' - x)^2}{(x' - x)^2} \\
&= \frac{(x' + x)(1 + xx') + 2axx' - 2byy'}{(x' - x)^2}.
\end{aligned}$$

Similarly $X(Q-R) = ((x'+x)(1+xx')+2axx'+2byy')/(x'-x)^2$. Thus

$$\begin{aligned}
&X(Q+R)X(Q-R)(x'-x)^4 \\
&\quad = ((x'+x)(1+xx')+2axx')^2 - (2byy')^2 \\
&\quad = ((x'+x)(1+xx')+2axx')^2 - 4by^2b(y')^2 \\
&\quad = ((x'+x)(1+xx')+2axx')^2 - 4(x^3+ax^2+x)((x')^3+a(x')^2+x') \\
&\quad = (x'+x)^2(1+xx')^2 + 4axx'(x'+x)(1+xx') + 4a^2x^2(x')^2 \\
&\qquad - 4(x^3{+}x)((x')^3{+}x') - 4a((x^3{+}x)(x')^2{+}((x')^3{+}x')x^2) - 4a^2x^2(x')^2 \\
&\quad = (x'+x)^2(1+xx')^2 + 4axx'(x'+x+(x')^2x+x^2x') \\
&\qquad - 4(x^3+x)((x')^3+x') - 4axx'(x^2x'+x'+(x')^2x+x) \\
&\quad = ((x')^2+2xx'+x^2)(1+2xx'+x^2(x')^2) \\
&\qquad - 4(x^3(x')^3+x^3x'+(x')^3x+xx') \\
&\quad = (x')^2+2xx'+x^2+2x(x')^3+4x^2(x')^2+2x^3x' \\
&\qquad + x^2(x')^4+2x^3(x')^3+x^4(x')^2 - 4(x^3(x')^3+x^3x'+(x')^3x+xx') \\
&\quad = (x')^2-2xx'+x^2-2x(x')^3+4x^2(x')^2-2x^3x' \\
&\qquad + x^2(x')^4-2x^3(x')^3+x^4(x')^2 \\
&\quad = ((x')^2-2xx'+x^2)(1-2xx'+x^2(x')^2) \\
&\quad = (x'-x)^2(xx'-1)^2
\end{aligned}$$

so $X(Q+R)X(Q-R) = (xx'-1)^2/(x'-x)^2$. Finally $X(Q) = x$ and $X(R) = x'$. □

4 Handling the exceptional cases

Consider the problem of efficiently computing nth multiples in the group $E(k)$ defined in Section 2.

It is usually, but not always, true that any point of the form $(x_1/z_1, y_1)$ in $E(k)$ has nth multiple $(x_n/z_n, y_n)$ for some y_n, where x_n and z_n are defined recursively in Section 1. Indeed, put $Q = n(x_1/z_1, y_1)$ and $R = (n{+}1)(x_1/z_1, y_1)$, and assume inductively that $X(Q) = x_n/z_n$ and $X(R) = x_{n+1}/z_{n+1}$. Theorem 3.1 usually states that

$$X(2Q) = \frac{(X(Q)^2-1)^2}{4(X(Q)^3+aX(Q)^2+X(Q))} = \frac{(x_n^2-z_n^2)^2}{4(x_n^3z_n+ax_n^2z_n^2+x_nz_n^3)} = \frac{x_{2n}}{z_{2n}}$$

and Theorem 3.2 usually states that

$$X(Q{+}R) = \frac{(X(Q)X(R){-}1)^2}{X(Q{-}R)(X(Q){-}X(R))^2} = \frac{(x_nx_{n+1}-z_nz_{n+1})^2}{(x_1/z_1)(x_nz_{n+1}-x_{n+1}z_n)^2} = \frac{x_{2n+1}}{z_{2n+1}}.$$

But this logic breaks down if any of the hypotheses of Theorems 3.1 and 3.2 are violated: for example, if $2Q = \infty$.

Of course, if the nth multiple of $(x_1/z_1, \ldots)$ is ∞, then it is certainly not $(x_n/z_n, \ldots)$. But this is not the only case where the logic breaks down. For example, if the nth multiple of $(x_1/z_1, \ldots)$ is ∞, then the $(2n+1)$st multiple is the same as $(x_1/z_1, \ldots)$; is it true that $x_{2n+1}/z_{2n+1} = x_1/z_1$? The above induction does not reach x_n/z_n, so it also does not reach x_{2n+1}/z_{2n+1}.

Readers familiar with standard projective coordinates might guess that the complete story is as follows: (1) if the nth multiple of $(x_1/z_1, \ldots)$ is ∞ then $x_n = 0$ and $z_n = 0$; (2) if the nth multiple of $(x_1/z_1, \ldots)$ is not ∞ then $z_n \neq 0$ and the nth multiple is $(x_n/z_n, \ldots)$. But both parts of this guess turn out to be wrong. For example, take $x_1 = 0$ and $z_1 = 1$. The 2nd multiple of $(0, 0)$ is ∞, but $x_2 \neq 0$, contradicting the first part of the guess. Furthermore, the 3rd multiple of $(0, 0)$ is not ∞, but $z_3 = 0$, contradicting the second part of the guess.

One could check for ∞—equivalently, for various quantities being zero—during the recursive computation of x_n and z_n, and branch into a different computation when ∞ appears, falling back to various cases in the definition of elliptic-curve addition in Section 2. However, in some applications, checking for ∞ costs an extra multiplication for each bit of n, as discussed in Section 6. Furthermore, the complications are annoying for programmers who want a simple computation, and the branches are annoying for cryptographers who want to avoid leaking secrets through side channels.

Can we avoid these branches? The answer, in a nutshell, is yes. Theorem 4.3, the main theorem of this paper, shows exactly how x_n and z_n behave. Sections 5, 6, and 7 show how various applications can be efficiently adapted to the actual behavior of x_n and z_n.

The theorems

Define $/ : k \times k \to \{\infty\} \cup k$ as follows: x/z is the usual quotient in k for $z \neq 0$; $x/0 = \infty$. Note that if $x/z = x'/z'$ then $xz' = x'z$, but the converse is not necessarily true.

Theorem 4.1. *Let k be a field not of characteristic 2. Let a, b be elements of k with $b(a^2 - 4) \neq 0$. Define E as the elliptic curve $by^2 = x^3 + ax^2 + x$ over k. Let Q be an element of $E(k)$. Let x_1, z_1 be elements of k with $x_1/z_1 = X(Q)$ and $(x_1, z_1) \neq (0, 0)$. Define $x_2 = (x_1^2 - z_1^2)^2$ and $z_2 = 4x_1z_1(x_1^2 + ax_1z_1 + z_1^2)$. Then $x_2/z_2 = X(2Q)$ and $(x_2, z_2) \neq (0, 0)$.*

Proof. **Case 1:** $Q = \infty$. Then $2Q = \infty$; and $x_1/z_1 = X(Q) = \infty$ so $z_1 = 0$ so $z_2 = 0$ so $x_2/z_2 = \infty = X(2Q)$. Furthermore $x_1 \neq 0$ so $x_1^2 - z_1^2 \neq 0$ so $x_2 \neq 0$.

Case 2: $Q \neq \infty$ but $2Q = \infty$. Then $x_1/z_1 = X(Q) \neq \infty$ so $z_1 \neq 0$. Furthermore $Q = (X(Q), 0)$ by definition of doubling; so $X(Q)^3 + aX(Q)^2 + X(Q) = 0$; so $z_2 = 4z_1^4(X(Q)^3 + aX(Q)^2 + X(Q)) = 0$; so $x_2/z_2 = \infty = X(2Q)$.

If $(x_2, z_2) = (0, 0)$ then $x_1^2 - z_1^2 = 0$ so $x_1 = \pm z_1$ so $\pm 4z_1^2(z_1^2 \pm az_1^2 + z_1^2) = 0$ so $(a \pm 2)z_1^4 = 0$; but $a \pm 2 \neq 0$ since $a^2 \neq 4$, so $z_1 = 0$, so $(x_1, z_1) = (0, 0)$, contradiction. Hence $(x_2, z_2) \neq (0, 0)$.

Case 3: $2Q \neq \infty$. Then $Q \neq \infty$ so $x_1/z_1 = X(Q) \neq \infty$ so $z_1 \neq 0$. Apply Theorem 3.1 to see that $z_2 = 4z_1^4(X(Q)^3 + aX(Q)^2 + X(Q)) \neq 0$ and $X(2Q) =$

$(X(Q)^2-1)^2/4(X(Q)^3+aX(Q)^2+X(Q)) = (x_1^2-z_1^2)^2/4(x_1^3z_1+ax_1^2z_1^2+x_1z_1^3) = x_2/z_2$. $\square$

Theorem 4.2. *Let k be a field not of characteristic 2. Let a, b be elements of k with $b(a^2-4) \neq 0$. Define E as the elliptic curve $by^2 = x^3+ax^2+x$ over k. Let Q and R be elements of $E(k)$. Let*

- *x_1, z_1 be elements of k with $x_1/z_1 = X(Q-R)$, $x_1 \neq 0$, and $z_1 \neq 0$;*
- *x_2, z_2 be elements of k with $x_2/z_2 = X(Q)$ and $(x_2, z_2) \neq (0, 0)$; and*
- *x_3, z_3 be elements of k with $x_3/z_3 = X(R)$ and $(x_3, z_3) \neq (0, 0)$.*

Define $x_5 = 4(x_2x_3 - z_2z_3)^2z_1$ and $z_5 = 4(x_2z_3 - z_2x_3)^2x_1$. Then $x_5/z_5 = X(Q+R)$ and $(x_5, z_5) \neq (0, 0)$.

Note that both x_1 and z_1 are assumed to be nonzero.

Proof. **Case 1:** $Q = R$. Then $x_1/z_1 = X(Q-R) = X(\infty) = \infty$ so $z_1 = 0$, contradiction.

Case 2: $Q \neq R$ and $Q = \infty$. Then $x_2/z_2 = X(Q) = \infty$ so $z_2 = 0$. Furthermore $x_1/z_1 = X(Q-R) = X(R) = x_3/z_3$ so $z_1x_3 = x_1z_3$. Hence $x_5 = 4(x_2x_3)^2z_1 = 4x_2^2x_1z_3x_3$ and $z_5 = 4(x_2z_3)^2x_1 = 4x_2^2x_1z_3^2$.

Observe that $z_5 \neq 0$. Indeed, $x_1 \neq 0$; $x_2 \neq 0$ since $z_2 = 0$; and $z_3 \neq 0$ since $x_3/z_3 = X(R) \neq \infty$. Thus $x_5/z_5 = x_3/z_3 = X(R) = X(Q+R)$ and $(x_5, z_5) \neq (0, 0)$.

Case 3: $Q \neq R$ and $R = \infty$. Then $x_3/z_3 = X(R) = \infty$ so $z_3 = 0$. Furthermore $x_1/z_1 = X(Q-R) = X(Q) = x_2/z_2$ so $z_1x_2 = x_1z_2$. Hence $x_5 = 4(x_2x_3)^2z_1 = 4x_3^2x_1z_2x_2$ and $z_5 = 4(z_2x_3)^2x_1 = 4x_3^2x_1z_2^2$.

Observe that $z_5 \neq 0$. Indeed, $x_1 \neq 0$; $x_3 \neq 0$ since $z_3 = 0$; and $z_2 \neq 0$ since $x_2/z_2 = X(Q) \neq \infty$. Thus $x_5/z_5 = x_2/z_2 = X(Q) = X(Q+R)$ and $(x_5, z_5) \neq (0, 0)$.

Case 4: $Q \neq R$ and $Q+R = \infty$. Then $x_2/z_2 = X(Q) = X(-R) = X(R) = x_3/z_3$ so $x_2z_3 = z_2x_3$ so $z_5 = 0$. Hence $x_5/z_5 = \infty = X(\infty) = X(Q+R)$. I will show that $x_5 \neq 0$; hence $(x_5, z_5) \neq (0, 0)$.

Note that $x_2 \neq 0$: if $x_2 = 0$ then $z_2 \neq 0$ so $X(Q) = x_2/z_2 = 0$ so $Q = (0, 0)$ so $R = -Q = -(0, 0) = (0, 0) = Q$, contradiction. Similarly $x_3 \neq 0$.

Suppose that $x_5 = 0$. Then $4(x_2x_3 - z_2z_3)^2z_1 = 0$, but $z_1 \neq 0$, so $x_2x_3 = z_2z_3$. Consequently $(x_2 - z_2)(x_3 + z_3) = x_2x_3 - z_2x_3 + x_2z_3 - z_2z_3 = 0$ and $(x_2 + z_2)(x_3 - z_3) = x_2x_3 + z_2x_3 - x_2z_3 - z_2z_3 = 0$. If $x_2 + z_2 \neq 0$ then $x_3 - z_3 = 0$ so $x_3 + z_3 = 2x_3 \neq 0$ so $x_2 - z_2 = 0$; i.e., $X(Q) = x_2/x_2 = 1$ and $X(R) = x_3/x_3 = 1$. Otherwise $x_2 = -z_2$ so $x_2 - z_2 = 2x_2 \neq 0$ so $x_3 = -z_3$; i.e., $X(Q) = -1$ and $X(R) = -1$. Either way $X(Q)^2 - 1 = 0$. Now $2Q \neq Q + R = \infty$ so $X(2Q) = (X(Q)^2-1)^2/\cdots = 0$ by Theorem 3.1. Thus $x_1/z_1 = X(Q-R) = X(2Q) = 0$ so $x_1 = 0$, contradiction.

Case 5: $Q \neq R$; $Q \neq \infty$; $R \neq \infty$; and $Q+R \neq \infty$. Then $x_2/z_2 = X(Q) \neq \infty$ so $z_2 \neq 0$; $x_3/z_3 = X(R) \neq \infty$ so $z_3 \neq 0$; and $X(Q) \neq X(R)$ so $x_2/z_2 \neq x_3/z_3$ so $z_5 \neq 0$. Now

$$X(Q+R)\frac{x_1}{z_1} = X(Q+R)X(Q-R) = \frac{(X(Q)X(R)-1)^2}{(X(Q)-X(R))^2} = \frac{(x_2x_3 - z_2z_3)^2}{(x_2z_3 - x_3z_2)^2}$$

by Theorem 3.2; so $X(Q+R) = (x_2x_3 - z_2z_3)^2z_1/(x_2z_3 - x_3z_2)^2x_1 = x_5/z_5$. $\square$

Theorem 4.3. *Let k be a field not of characteristic 2. Let a, b be elements of k with $b(a^2-4) \neq 0$. Define E as the elliptic curve $by^2 = x^3 + ax^2 + x$ over k. Let Q be an element of $E(k)$. Let x_1, z_1 be elements of k with $x_1/z_1 = X(Q)$ and $(x_1, z_1) \neq (0,0)$. Recursively define $(x_2, x_3, \ldots)$ and $(z_2, z_3, \ldots)$ by*

$$\begin{aligned} x_{2n} &= (x_n^2 - z_n^2)^2 && \text{for } n \geq 1, \\ z_{2n} &= 4x_nz_n(x_n^2 + ax_nz_n + z_n^2) && \text{for } n \geq 1, \\ x_{2n+1} &= 4(x_nx_{n+1} - z_nz_{n+1})^2z_1 && \text{for } n \geq 1, \\ z_{2n+1} &= 4(x_nz_{n+1} - z_nx_{n+1})^2x_1 && \text{for } n \geq 1. \end{aligned}$$

Then $x_n/z_n = X(nQ)$ for each $n \geq 1$, except in the following case: if $x_1 = 0$, $z_1 \neq 0$, $n > 1$, and n is odd, then $Q = (0,0)$, $X(nQ) = 0$, and $x_n/z_n = \infty$. Furthermore, $(x_n, z_n) \neq (0,0)$ for each $n \geq 1$, except in the following cases: if $x_1 \neq 0$, $z_1 = 0$, and n is not a power of 2, then $Q = \infty$ and $(x_n, z_n) = (0,0)$; if $x_1 = 0$, $z_1 \neq 0$, and n is not a power of 2, then $Q = (0,0)$ and $(x_n, z_n) = (0,0)$.

Proof. **Case 1:** $z_1 = 0$. Then $X(Q) = x_1/z_1 = x_1/0 = \infty$ so $Q = \infty$ so $X(nQ) = X(n\infty) = X(\infty) = \infty$.

Observe that $z_n = 0$ for every $n \geq 1$; consequently $x_n/z_n = \infty = X(nQ)$ as claimed. Indeed, $z_2 = \cdots z_1 = 0$; $z_3 = \cdots(\cdots z_2 - \cdots z_1)^2 = 0$; $z_4 = \cdots z_2 = 0$; $z_5 = \cdots(\cdots z_3 - \cdots z_2)^2 = 0$; etc.

Next observe that $x_{2n} = x_n^4$ and $x_{2n+1} = \cdots z_1 = 0$. If $x_1 = 0$ then $x_n = 0$ for all $n \geq 1$ by induction. Otherwise $x_n \neq 0$ when n is a power of 2, while $x_n = 0$ for all other n by induction.

Case 2: $x_1 = 0$ and $z_1 \neq 0$. Then $X(Q) = x_1/z_1 = 0/z_1 = 0$ so $Q = (0,0)$ so $2Q = (0,0) + (0,0) = \infty$; so $X(nQ) = X(0,0) = 0$ for n odd, $X(nQ) = X(\infty) = \infty$ for n even.

Observe that $z_n = 0$ for every $n > 1$. Indeed, $z_2 = \cdots x_1 = 0$; $z_3 = \cdots x_1 = 0$; $z_4 = \cdots z_2 = 0$; $z_5 = \cdots x_1 = 0$; etc. Consequently each odd $n > 1$ has $x_n/z_n = \infty = X(nQ)$, while each even $n > 1$ has $x_n/z_n = \infty$ with $X(nQ) = 0$.

Next $x_n = 0$ for every odd n. Indeed, $x_{2n+1} = 4(x_nx_{n+1} - z_nz_{n+1})^2z_1$. One of $n, n+1$ is odd, so $x_nx_{n+1} = 0$ by induction; and $z_{n+1} = 0$.

Now $x_{2n} = x_n^4$. Thus $x_n = 0$ for every n that is not a power of 2, while $x_n \neq 0$ when n is a power of 2.

Case 3: $x_1 \neq 0$ and $z_1 \neq 0$. Replace x_1, z_1, x_2, z_2 in Theorem 4.1 with x_n, z_n, x_{2n}, z_{2n}: if $x_n/z_n = X(nQ)$ and $(x_n, z_n) \neq (0,0)$ then $x_{2n}/z_{2n} = X(2nQ)$ and $(x_{2n}, z_{2n}) \neq (0,0)$. Similarly, replace $x_2, z_2, x_3, z_3, x_5, z_5$ in Theorem 4.2 with $x_n, z_n, x_{n+1}, z_{n+1}, x_{2n+1}, z_{2n+1}$: if $x_n/z_n = X(nQ)$ and $(x_n, z_n) \neq (0,0)$ and $x_{n+1}/z_{n+1} = X((n+1)Q)$ and $(x_{n+1}, z_{n+1}) \neq (0,0)$ then $x_{2n+1}/z_{2n+1} = X((2n+1)Q)$ and $(x_{2n+1}, z_{2n+1}) \neq (0,0)$. By induction $x_n/z_n = X(nQ)$ and $(x_n, z_n) \neq (0,0)$ for every $n \geq 1$. $\square$

5 Elliptic-curve cryptography (ECC)

Miller in [11], and independently Koblitz in [8], proposed an elliptic-curve variant of the Diffie-Hellman secret-sharing system. Miller in [11, page 420] suggested using the standard "division-polynomials" recurrence to compute nth multiples using 26 multiplications per exponent bit. Miller in [11, page 425] suggested using x-coordinates instead of (x, y)-coordinates.

The secret-sharing system with x-coordinates works as follows. One user, say Alice, has a secret key s and a public key $X(sP)$, where P is a standard point on a standard elliptic curve. Another user, say Bob, has a secret key t and a public key $X(tP)$. Alice and Bob then both know a shared secret $X(stP)$, apparently quite difficult for an attacker to predict. The bottleneck here is elliptic-curve scalar multiplication: Alice has to compute the shared secret $X(stP)$ given her secret key s and Bob's public key $X(tP)$.

What happens if one uses Montgomery's x_n/z_n to replace $X(n \cdots)$? Can an attacker force ∞ to occur in the elliptic-curve secret-sharing system, or in other cryptographic protocols? Can an attacker thus obtain information about a secret n? It is worrisome to see unanalyzed discrepancies between the nth multiples in papers and the x_n and z_n in high-speed software; perhaps the discrepancies allow easy attacks on cryptographic protocols that would otherwise have been secure.

I suggest replacing X by a modified x-coordinate function $X_0 : E(k) \to k$ defined as follows: $X_0(x, y) = x$; $X_0(\infty) = 0$. Theorem 5.1, generalizing the results for nonsquare a^2-4 in my recent conference paper [2, Appendix B], shows that X_0 of an nth multiple is always very easy to compute via Montgomery's recurrence (x_n, z_n).

Theorem 5.1. *Let k be a field not of characteristic 2. Let a, b be elements of k with $b(a^2 - 4) \neq 0$. Define E as the elliptic curve $by^2 = x^3 + ax^2 + x$ over k. Let Q be an element of $E(k)$. Recursively define $(x_1, x_2, x_3, \ldots)$ and $(z_1, z_2, z_3, \ldots)$ by*

$$\begin{aligned}
x_1 &= X_0(Q), \\
z_1 &= 1, \\
x_{2n} &= (x_n^2 - z_n^2)^2 && \text{for } n \geq 1, \\
z_{2n} &= 4x_n z_n(x_n^2 + ax_n z_n + z_n^2) && \text{for } n \geq 1, \\
x_{2n+1} &= 4(x_n x_{n+1} - z_n z_{n+1})^2 z_1 && \text{for } n \geq 1, \\
z_{2n+1} &= 4(x_n z_{n+1} - z_n x_{n+1})^2 x_1 && \text{for } n \geq 1.
\end{aligned}$$

Then $X_0(nQ) = x_n/z_n$ if $z_n \neq 0$, and $X_0(nQ) = 0$ if $z_n = 0$.

In particular, if k is finite, then $X_0(nQ) = x_n z_n^{\#k-1}$ for every $n \geq 1$.

Proof. If $z_n \neq 0$ then $x_n/z_n \neq \infty$ so $X(nQ) = x_n/z_n \neq \infty$ by Theorem 4.3 so $X_0(nQ) = x_n/z_n$.

If $z_n = 0$ then $x_n/z_n = \infty$ so $X(nQ) = x_n/z_n = \infty$ or $X(nQ) = 0$ by Theorem 4.3. Either way $X_0(nQ) = 0$. □

6 Elliptic-curve primality proving (ECPP)

Goldwasser and Kilian in [5] suggested proving the primality of an integer p by exhibiting a point of order $q > (p^{1/4} + 1)^2$ on an elliptic curve over $\mathbf{Z}/p$. If p is not prime then there is a field quotient k of $\mathbf{Z}/p$ with $\#k \le \sqrt{p}$; but the same curve has a point of order q over k, so $q \le (\sqrt{\#k} + 1)^2 \le (p^{1/4} + 1)^2$ by Hasse's bounds in [7], contradiction.

This elliptic-curve primality-proving method has attracted interest for two reasons. First, there is a fast algorithm that is conjectured to always find an elliptic-curve primality proof—i.e., an appropriate elliptic curve, an appopriate point, an appropriate prime q, and a recursive proof of the primality of q. There have been many improvements in this algorithm; see [13] for the state of the art. Second, the resulting primality proofs are short: one can rather quickly verify, given a prime q and a point on an elliptic curve over $\mathbf{Z}/p$, that the point has order q on the curve.

The standard verification algorithm—see [6, Section 2.3]—works with affine coordinates and performs a division modulo p for each elliptic-curve addition. The division might fail, proving that p is actually composite. (With some effort one can define elliptic-curve addition in this case, as explained by Lenstra in [10, Section 3]; but this effort is unnecessary for elliptic-curve primality proving.) In the absence of such failures, the elliptic-curve operations over $\mathbf{Z}/p$ are consistent with the elliptic-curve operations over every field quotient k of $\mathbf{Z}/p$, as required for the Goldwasser-Kilian logic.

An obvious speedup here, as in other applications of elliptic curves, is to work with projective coordinates; i.e., to represent intermediate quantities as fractions, delaying all divisions until the last possible moment. But there is no guarantee that the simplest projective-coordinate algorithm produces the right results! The affine-coordinate algorithm checks invertibility of each denominator in $\mathbf{Z}/p$, either proving that p is composite or proving that the results are consistent with results over every field quotient k. The simplest projective-coordinate algorithm never checks invertibility, so it does not produce a complete proof of primality of p.

A corrected projective-coordinate algorithm checks invertibility of all the denominators, for example by computing $\gcd\{p, \text{product of denominators}\}$. This takes an extra multiplication modulo p for each elliptic-curve addition. Can these multiplications be eliminated?

The standard "division-polynomials" recurrence does not need intermediate invertibility tests. See, e.g., [13, Proposition 3.1]. But it is nevertheless a step backwards in efficiency.

I suggest instead using Montgomery's efficient recurrence. Theorem 6.1 shows that intermediate invertibility tests are not required here. The computation in Theorem 6.1 costs at most 10 multiplications per exponent bit. Normally c will be very small, saving 1 multiplication per exponent bit. One can also—at the expense of substantially more effort in finding q—force a to be small, saving another multiplication per exponent bit.

Beware that not all elliptic curves are isomorphic to curves of Montgomery form. In particular, the number of points on a Montgomery-form elliptic curve over a finite field is always in $4\mathbf{Z}$. An easy calculation suggests that, out of all elliptic curves over a prime field,

- $1/3+o(1)$ have odd order and thus are not isomorphic to Montgomery form;
- $1/4+o(1)$ have exactly one power of 2, same conclusion;
- for primes in $3+4\mathbf{Z}$, an additional $1/24+o(1)$ have exactly two powers of 2 without being isomorphic to Montgomery form; and
- the remaining curves are all isomorphic to Montgomery form.

Consequently one might speculate that my easy-to-verify primality proofs take either $8/3+o(1)$ or $12/5+o(1)$ times as long to find as traditional elliptic-curve primality proofs. But the actual slowdown is less severe: the curves generated in traditional elliptic-curve primality proofs typically have more factors of 2 than random curves, and thus are more likely to be isomorphic to Montgomery form.

Theorem 6.1. *Let q be a prime. Let f be a positive integer. Let p be an integer larger than 1. Let a, c be integers. Assume that $\gcd\{2(a^2-4)(c^3+ac^2+c), p\} = 1$. Recursively define $(x_1, x_2, x_3, \ldots)$ and $(z_1, z_2, z_3, \ldots)$ by*

$$\begin{aligned} x_1 &= c, \\ z_1 &= 1, \\ x_{2n} &= (x_n^2 - z_n^2)^2 && \text{for } n \geq 1, \\ z_{2n} &= 4x_n z_n(x_n^2 + a x_n z_n + z_n^2) && \text{for } n \geq 1, \\ x_{2n+1} &= 4(x_n x_{n+1} - z_n z_{n+1})^2 z_1 && \text{for } n \geq 1, \\ z_{2n+1} &= 4(x_n z_{n+1} - z_n x_{n+1})^2 x_1 && \text{for } n \geq 1. \end{aligned}$$

If $\gcd\{z_f, p\} = 1$ and $z_{qf} \bmod p = 0$ and $q > (\lceil p^{1/4} \rceil + 1)^2$ then p is prime.

If q is already proven prime then the other conditions here can be checked efficiently, proving the primality of p. I used $q > (\lceil p^{1/4} \rceil + 1)^2$ rather than $q > (p^{1/4}+1)^2$ because the latter condition is not as easy to check.

One can somewhat simplify Theorem 6.1 by taking $f = 1$, but the simplified theorem often requires large c, slowing down the computation.

Proof. Define k as the smallest field quotient of $\mathbf{Z}/p$, and define $b = c^3 + ac^2 + c$. Then $2 \neq 0$ in k; $c \neq 0$ in k; $b(a^2-4) \neq 0$ in k; and $z_q = 0$ in k.

Define $Q \in k \times k$ as the pair $(c, 1)$. Then $Q \in E(k)$ where E is the elliptic curve $by^2 = x^3 + ax^2 + x$ over k. Furthermore $Q \neq (0,0)$; $(x_1, z_1) \neq (0,0)$ in k; and $x_1/z_1 = c = X(Q)$ in k.

By Theorem 4.3, $X(qfQ) = x_{qf}/z_{qf} = x_{qf}/0 = \infty$ in k, so $qfQ = \infty$. By hypothesis q is prime, so fQ has order 1 or q in the group $E(k)$.

By Theorem 4.3 again, $X(fQ) = x_f/z_f \neq \infty$ in k, so $fQ \neq \infty$. Thus fQ has order q.

Consequently $\#E(k) \geq q > (\lceil p^{1/4} \rceil + 1)^2 \geq (p^{1/4}+1)^2$. But $\#E(k) \leq (\sqrt{\#k}+1)^2$ by Hasse's theorem. Thus $(\sqrt{\#k}+1)^2 > (p^{1/4}+1)^2$; i.e., $\#k > p^{1/2}$; i.e., every prime divisor of p is larger than $p^{1/2}$. Consequently p is prime. $\square$

7 The elliptic-curve integer-factorization method (ECM)

Lenstra in [9] suggested finding small factors of an integer m by choosing n with many divisors, such as $n = \mathrm{lcm}\{1, 2, \ldots, 1000\} \approx 2^{1438}$, and computing the nth multiple of a random point on a random elliptic curve modulo m. Computing this multiple involves divisions modulo m, as in the Goldwasser-Kilian algorithm discussed in Section 6; one hopes that a division fails, revealing a factor of m. This is guaranteed to work if the multiple is ∞ modulo one factor of m (i.e., the original point modulo that factor has order dividing n) and not ∞ modulo another factor of m.

Montgomery in [12, Section 10.3.1] introduced his recurrences to speed up Lenstra's elliptic-curve factorization method. Montgomery's improved ECM is remarkably easy to state: choose a small $a \in \{6, 10, 14, \ldots\}$; choose $(x_1, z_1) = (2, 1)$; choose n; and compute $\gcd\{m, z_n\}$. The connection to elliptic curves is clear from Theorem 4.3: if $n(2, 1) = \infty$ on an elliptic curve $(4a + 10)y^2 = x^3 + ax^2 + x$ over a field k then $z_n = 0$ in k. Of course, ECM's success doesn't depend on this connection being perfectly reliable; what matters for ECM are the common cases analyzed by Montgomery, not the exceptional cases analyzed in this paper.

For small x_1, z_1 and large a, Montgomery's recurrences use 9 multiplications for each bit of n. For small x_1, z_1 and small a, Montgomery's recurrences use 8 multiplications for each bit of n. This improvement is stated in [12, page 261, bottom] but doesn't seem to be widely appreciated. The standard choice of a—see, e.g., [12, Section 10.3.2] and [15, Section 1, subsection "Suyama's parametrization"]—is large. There are slight advantages of the standard choice, but those advantages are outweighed by the extra multiplication when n is not very small.

There are many other ECM improvements due to Pollard, Montgomery, and others; for example, using many n's simultaneously. See [15] for a survey of the state of the art.

References

1. — (no editor), *Proceedings of the 18th annual ACM symposium on theory of computing*, Association for Computing Machinery, New York, 1986. ISBN 0–89791–193–8. See [5].
2. Daniel J. Bernstein, *Curve25519: new Diffie-Hellman speed records* (2006). URL: `http://cr.yp.to/papers.html#curve25519`. Citations in this document: §1, §5.
3. Daniel J. Bernstein, *Differential addition chains* (2006). URL: `http://cr.yp.to/papers.html#diffchain`. Citations in this document: §1.
4. Andrew M. Gleason (editor), *Proceedings of the International Congress of Mathematicians, volume 1*, American Mathematical Society, Providence, 1987. ISBN 0–8218–0110–4. MR 89c:00042. See [10].
5. Shafi Goldwasser, Joe Kilian, *Almost all primes can be quickly certified*, in [1] (1986), 316–329; see also newer version [6]. Citations in this document: §6.

6. Shafi Goldwasser, Joe Kilian, *Primality testing using elliptic curves*, Journal of the ACM **46** (1999), 450–472; see also older version [5]. ISSN 0004–5411. MR 2002e:11182. Citations in this document: §6.
7. Helmut Hasse, *Zur Theorie der abstrakten elliptischen Funktionenkörper I, II, III*, Journal für die Reine und Angewandte Mathematik (1936), 55–62, 69–88, 193–208. ISSN 0075–4102. Citations in this document: §6.
8. Neal Koblitz, *Elliptic curve cryptosystems*, Mathematics of Computation **48** (1987), 203–209. ISSN 0025–5718. MR 88b:94017. Citations in this document: §5.
9. Hendrik W. Lenstra, Jr., *Factoring integers with elliptic curves*, Annals of Mathematics **126** (1987), 649–673. ISSN 0003–486X. MR 89g:11125. URL: `http://links.jstor.org/sici?sici=0003-486X(198711)2:126:3<649:FIWEC>2.0.CO;2-V`. Citations in this document: §7.
10. Hendrik W. Lenstra, Jr., *Elliptic curves and number-theoretic algorithms*, in [4] (1987), 99–120. MR 89d:11114. URL: `http://cr.yp.to/bib/entries.html#1987/lenstra-ecnta`. Citations in this document: §6.
11. Victor S. Miller, *Use of elliptic curves in cryptography*, in [14] (1986), 417–426. MR 88b:68040. Citations in this document: §5, §5, §5.
12. Peter L. Montgomery, *Speeding the Pollard and elliptic curve methods of factorization*, Mathematics of Computation **48** (1987), 243–264. ISSN 0025–5718. MR 88e:11130. URL: `http://links.jstor.org/sici?sici=0025-5718(198701)48:177<243:STPAEC>2.0.CO;2-3`. Citations in this document: §1, §2, §3, §7, §7, §7.
13. François Morain, *Implementing the asymptotically fast version of the elliptic curve primality proving algorithm* (2005). URL: `http://www.lix.polytechnique.fr/~morain/Articles/fastecpp-final.pdf`. Citations in this document: §6, §6.
14. Hugh C. Williams (editor), *Advances in cryptology: CRYPTO '85*, Lecture Notes in Computer Science, 218, Springer, Berlin, 1986. ISBN 3–540–16463–4. See [11].
15. Paul Zimmermann, *20 years of ECM* (2006). URL: `http://www.loria.fr/~zimmerma/papers/`. Citations in this document: §7, §7.