Efficient arithmetic on elliptic curves using a mixed Edwards–Montgomery representation

Wouter Castryck[1], Steven Galbraith[2], and Reza Rezaeian Farashahi[3]

[1] Department of Electrical Engineering, University of Leuven,
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
wouter.castryck@esat.kuleuven.be
[2] Department of Mathematics, Royal Holloway University of London,
Egham Hill, Egham, Surrey TW20 0EX, United Kingdom
steven.galbraith@rhul.ac.uk
[3] Department of Mathematics and Computer Science,
Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, Netherlands
r.rezaeian@tue.nl

Abstract. From the viewpoint of x-coordinate-only arithmetic on elliptic curves, switching between the Edwards model and the Montgomery model is quasi cost-free. We use this observation to speed up Montgomery's algorithm, reducing the complexity of a doubling step from $2\mathbf{M} + 2\mathbf{S}$ to $1\mathbf{M} + 3\mathbf{S}$ for suitably chosen curve parameters.

1 Montgomery's algorithm

Aiming for an improved performance of Lenstra's elliptic curve factorization method [6], Montgomery developed a very efficient algorithm to compute in the group associated to an elliptic curve over a non-binary finite field $\mathbb{F}_q$, in which only x-coordinates are involved [8].

The algorithm also proves useful for point compression in elliptic curve cryptography. More precisely, instead of sending a point as part of some cryptographic protocol, one can reduce the communication cost by sending just its x-coordinate. From this, the receiver can compute the x-coordinate of any scalar multiple using Montgomery's method. This idea was first mentioned in [7].

The type of curves Montgomery considered are of the following non-standard Weierstrass type

$$M_{A,B} : \; By^2 = x^3 + Ax^2 + x, \quad A \in \mathbb{F}_q \setminus \{\pm 2\}, \, B \in \mathbb{F}_q \setminus \{0\},$$

which is now generally referred to as a *Montgomery form*. His method works as follows. Let $P = (x_1, y_1, z_1)$ be a point on $\overline{M}_{A,B}$, the projective closure of $M_{A,B}$, and for any $n \in \mathbb{N}$ write $n \cdot P = (x_n, y_n, z_n)$, where the multiple is taken in the algebraic group $\overline{M}_{A,B}, \oplus$ with neutral element $O = (0, 1, 0)$. Then the following recursive relations hold: for any $m, n \in \mathbb{N}$ such that $m \neq n$ we have

$$\begin{aligned} x_{m+n} &= z_{m-n} \left((x_m - z_m)(x_n + z_n) + (x_m + z_m)(x_n - z_n) \right)^2, \\ z_{m+n} &= x_{m-n} \left((x_m - z_m)(x_n + z_n) - (x_m + z_m)(x_n - z_n) \right)^2. \end{aligned} \qquad \text{(ADD)}$$

and

$$\begin{aligned} 4x_n z_n &= (x_n + z_n)^2 - (x_n - z_n)^2, \\ x_{2n} &= (x_n + z_n)^2 (x_n - z_n)^2, \\ z_{2n} &= 4x_n z_n \left((x_n - z_n)^2 + ((A+2)/4)\,(4x_n z_n)\right) \end{aligned} \qquad \text{(DOUBLE)}$$

(see also [3]). One can then compute $((x_n, z_n), (x_{n+1}, z_{n+1}))$ from

$$\left((x_{(n \text{ div } 2)}, z_{(n \text{ div } 2)}), (x_{(n \text{ div } 2)+1}, z_{(n \text{ div } 2)+1})\right)$$

by one application of (ADD) and one application of (DOUBLE), the input of the latter depending on n mod 2. Thus approximately $\log_2 n$ applications of (ADD) and (DOUBLE) suffice to recover (x_n, z_n).

Every application of (ADD) has a rough time-cost of $3\mathbf{M} + 2\mathbf{S}$, where $\mathbf{M}$ is the time needed to multiply two general elements of $\mathbb{F}_q$, and $\mathbf{S}$ is the time needed to square a general element (which is typically faster). Here we used that $z_1 = 1$ in practice. Every application of (DOUBLE) needs $2\mathbf{M} + 2\mathbf{S} + 1\mathbf{C}$, where $\mathbf{C}$ is the cost of multiplication of a general element of $\mathbb{F}_q$ with a curve constant. In this case, the constant is $(A + 2)/4$ (hence, if A is chosen carefully then $\mathbf{C}$ may be much less than $\mathbf{M}$).

2 Switching to Edwards curves and back

Following recent work of Edwards [4], Bernstein and Lange [2] proved that the elliptic curves

$$E_d : X^2 + Y^2 = 1 + dX^2Y^2 \quad d \in \mathbb{F}_q \setminus \{0, 1\}$$

allow a very esthetic description of the algebraic group law on $\overline{E}_d$, the (desingularized) projective closure of E_d, with $O = (0, 1) \in E_d \subset \overline{E}_d$ as neutral element. Namely, the formula

$$(X_1, Y_1) \oplus (X_2, Y_2) = \left(\frac{X_1X_2 + Y_1Y_2}{1 + dX_1X_2Y_1Y_2}, \frac{Y_1Y_2 - X_1X_2}{1 - dX_1X_2Y_1Y_2} \right)$$

holds at all affine point pairs for which the above denominators are nonzero. The curve E_d is said to be in *Edwards form.* In [1, Theorem 3.2.] it is proven that every Edwards form is birationally equivalent to a Montgomery form via

$$\begin{aligned} \varphi &: M_{\frac{2(1+d)}{1-d}, \frac{4}{1-d}} \dashrightarrow E_d : (x, y) \mapsto \left(\tfrac{x}{y}, \tfrac{x-1}{x+1}\right), \\ \psi &: E_d \dashrightarrow M_{\frac{2(1+d)}{1-d}, \frac{4}{1-d}} : (X, Y) \mapsto \left(\tfrac{1+Y}{1-Y}, X\tfrac{1+Y}{1-Y}\right). \end{aligned}$$

The dashed arrows indicate that the maps are not defined everywhere. However, the maps can be extended to give an everywhere-defined isomorphism between the respective (desingularized) projective models

$$\overline{M}_{\frac{2(1+d)}{1-d}, \frac{4}{1-d}} \longrightarrow \overline{E}_d$$

that maps the neutral elements O to each other. In particular, wherever φ and ψ are defined, they commute with the group structures on $\overline{M}_{\frac{2(1+d)}{1-d}, \frac{4}{1-d}}$ and $\overline{E}_d$.

Now the Y-coordinate of $\varphi(x, y)$ only depends on x, and conversely the x-coordinate of $\psi(X, Y)$ only depends on Y. In projective coordinates this correspondence becomes remarkably simple:

$$\varphi : (x, z) \mapsto (x - z, x + z) \quad \text{and} \quad \psi : (Y, Z) \mapsto (Z + Y, Z - Y).$$

Therefore, from the x/Y-coordinate-only viewpoint, switching between Edwards curves and Montgomery curves is quasi cost-free. As a consequence, one is free to pick the best from either world. In the next section we show that it is worth considering the (DOUBLE) step in the Edwards setting.

3 Y-coordinate-only doubling on Edwards curves

A general affine point (X, Y) on E_d doubles to a point whose second coordinate equals

$$\frac{Y^2 - X^2}{1 - dX^2Y^2} = \frac{Y^2(1 - dY^2) - (1 - Y^2)}{(1 - dY^2) - dY^2(1 - Y^2)} = \frac{-1 + 2Y^2 - dY^4}{1 - 2dY^2 + dY^4}.$$

Here we used the curve equation $X^2+Y^2 = 1+dX^2Y^2$. Therefore the (DOUBLE) analog becomes

$$\begin{aligned} Y_{2n} &= -Z_n^4 + 2Y_n^2 Z_n^2 - dY_n^4 = -(Z_n^4 + dY_n^4) + 2Y_n^2 Z_n^2, \\ Z_{2n} &= Z_n^4 - 2dY_n^2 Z_n^2 + dY_n^4 \;= (Z_n^4 + dY_n^4) - 2dY_n^2 Z_n^2. \end{aligned}$$

Suppose that d has a square root $\sqrt{d}$ in $\mathbb{F}_q$. Then the above step can be done using $1\mathbf{M} + 3\mathbf{S} + 3\mathbf{C}$ by computing

$$Y_n^2, \;\; Z_n^2, \;\; Y_n^2 Z_n^2, \;\; \sqrt{d}Y_n^2, \;\; \sqrt{d}Y_n^2 Z_n^2, \;\; dY_n^2 Z_n^2, \;\; (Z_n^2 + \sqrt{d}Y_n^2)^2$$

and then recovering $Z_n^4 + dY_n^4$ as $(Z_n^2 + \sqrt{d}Y_n^2)^2 - 2\sqrt{d}Y_n^2 Z_n^2$. If d is nonsquare, one easily verifies that a time cost of $5\mathbf{S} + 2\mathbf{C}$ can be achieved.

4 Conclusion and additional remarks

To sum up, our proposal is to work with a Montgomery curve of the type $M_{\frac{2(1+d)}{1-d}, \frac{4}{1-d}}$, and to replace (DOUBLE) by

$$\begin{aligned} Y_n &= x_n - z_n \\ Z_n &= x_n + z_n \\ Y_{2n} &= -(Z_n^4 + dY_n^4) + 2Y_n^2 Z_n^2 \\ Z_{2n} &= (Z_n^4 + dY_n^4) - 2dY_n^2 Z_n^2 \\ x_{2n} &= Z_{2n} + Y_{2n} \\ z_{2n} &= Z_{2n} - Y_{2n}. \end{aligned}$$

These formulas are complete, in the sense that for *every* input (x_n, z_n) they give the correct output (x_{2n}, z_{2n}). This is in contrast with the switching maps φ and ψ

and with the Edwards doubling formulas. But under the above composition, the incompleteness disappears: this can be checked by directly expressing (x_{2n}, z_{2n}) in terms of (x_n, z_n) and verifying that – up to scalar multiplication by $-2d+2$ – it matches with classical Montgomery doubling.

If the curve constant d is a square such that multiplication by $\sqrt{d}$ is cheap, then the above method improves upon Montgomery doubling by roughly $\mathbf{M}-\mathbf{S}$, i.e. it replaces a multiplication by a squaring. Therefore, our simple ideas can serve in constructing slightly improved ECC protocols for devices with limited computational power and memory. We remark that an even better speed-up of $\mathbf{2M}-\mathbf{2S}$ has been independently[1] obtained by Gaudry and Lubicz [5], who work however on a Kummer line instead of directly on a Montgomery form.

Not every Montgomery form is birationally equivalent to an Edwards curve, but this is resolved by extending to the class of so-called *twisted* Edwards forms $aX^2+Y^2 = 1+dX^2Y^2$ $(a \neq d)$, as was pointed out in [1]. For this class, exactly the same ideas apply, resulting in a doubling algorithm using $1\mathbf{M}+3\mathbf{S}+6\mathbf{C}$ if ad is a square, and $5\mathbf{S}+4\mathbf{C}$ in general.

We end by recalling that the Edwards-Montgomery setting only covers non-binary fields. Over binary fields there is less need for arithmetic directly on compressed representations, since a received point can be typically decompressed by solving a quadratic equation, which is easy in characteristic two. The transmission of an extra bit then allows the decompressor to decide upon the correct solution.

References

1. D. Bernstein, P. Birkner, M. Joye, T. Lange, C. Peters, Twisted Edwards Curves, *AFRICACRYPT 2008, Springer Lecture Notes in Computer Science, Springer* **5023**, pp. 389–405 (2008)
2. D. Bernstein and T. Lange, Faster addition and doubling on elliptic curves, *Advances in Cryptology - ASIACRYPT 2007, Springer Lecture Notes in Computer Science* **4833**, pp. 29–50 (2007)
3. C. Doche and T. Lange, Arithmetic of Elliptic Curves, Chapter 13 in H. Cohen and G. Frey (Eds.), Handbook of elliptic and hyperelliptic curve cryptography, *Chapman & Hall/CRC Press* (2005)
4. H. Edwards, A normal form for elliptic curves. *Bulletin of the American Mathematical Society* **44**, pp. 393–422 (2007)
5. P. Gaudry and D. Lubicz, The arithmetic of characteristic 2 Kummer surfaces, preprint
6. H. Lenstra, Factoring integers with elliptic curves, *Annals of Mathematics* **126**, pp. 649-673 (1987)
7. V. Miller, Use of elliptic curves in cryptography, *CRYPTO '85, Springer Lecture Notes in Computer Science* **218**, pp. 417–426 (1986)
8. P. Montgomery, Speeding the Pollard and elliptic curve methods of factorization, *Mathematics of Computation* **48**, pp. 243-264 (1987)

[1] This is an euphemistic rephrasing of our ignorance about Gaudry and Lubicz' result, which is somewhat hidden in a different framework. Its existence was pointed out to us by Dan Bernstein and Tanja Lange.