Computing Isogenies between Montgomery Curves Using the Action of $(0, 0)$

Joost Renes*

Digital Security Group, Radboud University, Nijmegen, The Netherlands
j.renes@cs.ru.nl

Abstract. A recent paper by Costello and Hisil at Asiacrypt'17 presents efficient formulas for computing isogenies with odd-degree cyclic kernels on Montgomery curves. We provide a constructive proof of a generalization of this theorem which shows the connection between the shape of the isogeny and the simple action of the point $(0, 0)$. This generalization removes the restriction of a cyclic kernel and allows for any separable isogeny whose kernel does not contain $(0, 0)$. As a particular case, we provide efficient formulas for 2-isogenies between Montgomery curves and show that these formulas can be used in isogeny-based cryptosystems without expensive square root computations and without knowledge of a special point of order 8. We also consider elliptic curves in triangular form containing an explicit point of order 3.

Keywords: Vélu's formulas, Montgomery form, 2-isogenies, SIDH, Post-quantum cryptography.

1 Introduction

Ever since their introduction to public-key cryptography by Miller [Mil86] and Koblitz [Kob87], elliptic curves have been of interest to the cryptographic community. By using the group of points on an appropriately chosen elliptic curve where the discrete logarithm problem is assumed to be hard, many standard protocols can be instantiated. Notably, the Diffie–Hellman key exchange [DH76] and the Schnorr signature scheme [Sch89] and its variants [Acc99, BDL+12] allow for efficient implementations with high security and small keys. The efficiency of these curve-based algorithms is largely determined by the scalar multiplication routine, and as a result a lot of research has gone into optimizing this operation.

However, the threat of large-scale quantum computers has initiated the search for alternative algorithms that also resist quantum adversaries (which the classical curve-based systems do not [Sho94]). Building on the work of Couveignes [Cou06] and Rostovsev and Stolbunov [RS06], in 2011 Jao and De Feo [JF11] proposed supersingular isogeny Diffie–Hellman (SIDH) as a key exchange protocol offering post-quantum security. Being based on the theory of elliptic curves, SIDH inherits several operations from traditional curve-based cryptography. As such, it has immediately benefited from decades of prior research into optimizing their operations. In particular, the Montgomery form of an elliptic curve has resulted in great performance. Initially proposed by Montgomery to speed up factoring using ECM [Mon87, Len87] and having been used for very efficient Diffie–Hellman key exchange (eg. Bernstein's Curve25519 [Ber06]), the current fastest instantiations of SIDH also employ Montgomery curves [CLN16b, KAK16]. But, although the optimizations for scalar multiplication immediately carry over, the work on computing explicit isogenies on Montgomery curves is more limited.

* This work has been supported by the Technology Foundation STW (project 13499 – TYPHOON & ASPASIA), from the Dutch government.

For isogeny computations one commonly uses Vélu's formulas [Vél71]. However, if the elliptic curve has a form which is less general than (or different from) Weierstrass form, the formulas from Vélu are not guaranteed to preserve this. As isogenies are only well-defined up to isomorphism, one can post-compose with an appropriate isomorphism to return to the required form, but it may not be obvious with which isomorphism, or the isomorphism may be expensive to compute. A more elegant approach is to observe some extra structure on the curve model and require the isogenies to preserve this. For example, Moody and Shumow [MS16] apply this idea to Edwards and Huff curves by fixing certain points. Moreover, since the isogeny is invariant under addition by kernel points, there is a close connection between the isogeny and the action (by translation) of some chosen point. We make this more explicit in Theorem 1 for curves in Weierstrass form.

So far the approaches for obtaining formulas for isogenies on Montgomery curves have been rather ad hoc. In [FJP14], De Feo, Jao and Plût apply Vélu's formulas and compose with the appropriate isomorphisms to return to Montgomery form. As noted in [FJP14, §4.3.2], this approach fails to produce efficient results for 2-isogenies. That is, either one has to compute expensive square roots in a finite field (see eg. [CJL+17, §3.1]), or one relies on having an appropriate point of order 8. However, this point of order 8 is not readily available for the final two 2-isogenies. As one suggested workaround in [FJP14] they derive formulas for 4-isogenies between two curves in Montgomery form and propose to compute 2^e-isogenies as a chain of 4-isogenies. As a result, optimized SIDH implementations [CLN16a, KAK16] have employed curves where e is even so that 2^e-isogenies can be comprised entirely of 4-isogenies. In [CH17], Costello and Hisil present elegant formulas for isogenies between Montgomery curves, but their theorem covers only the case of odd cyclic kernels and subsequently also does not address the case of 2-isogenies. Moreover, there is no justification for the derivation of these isogenies (except for showing that they work).

We bridge this gap by providing a more thorough analysis on isogenies between Montgomery curves. We show that the isogenies arising in [CH17] are exactly those fixing $(0, 0)$. Since we enforce the isogeny to fix $(0, 0)$, this point cannot be in the kernel. We show in Proposition 1 that this is the only restriction, and as a result present a generalization of [CH17, Theorem 1]. As a special case, we obtain formulas for 2-isogenies for 2-torsion points other than $(0, 0)$. We then show that this point can be naturally avoided in well-designed isogeny-based cryptosystems (see §4.3), and discuss the application of the 2-isogeny formulas to isogeny-based cryptosystems.

Finally, although currently it does not give rise to faster isogeny formulas, we consider it worthwhile to point out that the same techniques immediately apply to other models. In particular, models derived from the *Tate Normal Form* [Hus04, §4.4], where one could expect to get simple ℓ-isogenies for $\ell \geq 3$. We work out the case $\ell = 3$, also known as the *triangular form* [BCKL15], and derive isogenies by again fixing the action of the special point $(0, 0)$.

Organization. We begin by recalling some background on elliptic curves, isogenies and SIDH in §2. We state a theorem in §3 that allows to describe an isogeny in terms of the abscissas of its kernel points and their translations by a chosen point Q. We apply this to Montgomery curves in §4 and to curves in triangular form in §5, in both cases using $Q = (0, 0)$.

2 Preliminaries

An elliptic curve E defined over a field K is by definition [Gal12, Sil09] the curve

$$E/K : Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3 \ , \tag{1}$$

where $a_1, a_2, a_3, a_4, a_6 \in K$ such that E is non-singular. It is embedded into $\mathbb{P}^2$ with a single point $\mathcal{O}_E = (0 : 1 : 0)$ on the line $Z = 0$. This form is commonly referred to as *Weierstrass form* and the specified base point (implicitly) is $\mathcal{O}_E$. On the open patch defined by $Z \neq 0$ we can set $x = X/Z$ and $y = Y/Z$ and work on the corresponding affine curve inside $\mathbb{A}^2$ given by

$$E/K : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \ .$$

We can move back to the projective curve by mapping $(x, y) \mapsto (x : y : 1)$. Therefore, although many equations are given in affine coordinates, they can easily be transformed into projective ones. For any extension L/K, the set of L-rational points $E(L)$ forms a group with identity $\mathcal{O}_E$ [Sil09, Prop. 2.2(f)]. A subgroup $G \subset E(\bar{K})$ is said to be defined over K if $\sigma(P) \in G$ for all σ in the Galois group $\mathrm{Gal}(\bar{K}/K)$.

Isogenies. Let E and $\widetilde{E}$ be elliptic curves. An isogeny ϕ from E to $\widetilde{E}$ is a surjective morphism such that $\phi(\mathcal{O}_E) = \mathcal{O}_{\widetilde{E}}$ [Sil09, §III.4]. The (finite) degree d of an isogeny is its degree as a morphism, and we say an isogeny is separable if $\#\ker(\phi) = d$. In this paper every isogeny that appears is assumed to be separable. Given a finite subgroup $G \subset E(\bar{K})$ defined over K, there exists a curve $\widetilde{E}$ and an isogeny $\phi : E \to \widetilde{E}$ such that $\ker(\phi) = G$ [Gal12, Theorem 9.6.19]. The curve $\widetilde{E}$ is unique up to isomorphism (over $\bar{K}$) and the isogeny ϕ is unique up to post-composition with an isomorphism. The isogeny ϕ can be made explicit by using Vélu's formulas [Vél71] (for some fixed choice for the isogeny).

Montgomery Form. Setting $a_1 = a_3 = a_6 = 0$ and $a_4 = 1$ gives a curve in the form $E : y^2 = x^3 + ax^2 + x$. We also consider curves in the form $by^2 = x^3 + ax^2 + x$, better known as *Montgomery form*. Over $\bar{K}$ these two curve forms are isomorphic via $(x, y) \mapsto (x, y\sqrt{b})$, but this isomorphism is only defined over K if $\sqrt{b} \in K$. In particular, if $K = \mathbb{F}_q$ and $\sqrt{b} \notin K$ then we call this curve a (non-trivial) quadratic twist. An easy check shows that $Q = (0, 0)$ is a K-rational point of order 2, while for any $Q_4 \in E(\bar{K})$ we have that $[2]Q_4 = Q$ if and only if

$$Q_4 \in \{(1, \pm\sqrt{(a+2)/b}), (-1, \pm\sqrt{(a-2)/b})\}$$

If P is any point of order 2 other than Q, then $x_P^2 + ax_P + 1 = 0$.

Tate Normal Form. Suppose we are given a curve E/K containing a point P of prime order $\ell \geq 3$. We can move P to $(0, 0)$ and its tangent line to the line $y = 0$. This transformation is completely K-rational and puts the curve in *Tate Normal Form* [Hus04, §4.4]

$$y^2 + axy + by = x^3 + cx^2 \,, \quad a, b, c \in K \ .$$

In §5 we focus on the case where $\ell = 3$, in which case $c = 0$ and $b \neq 0$. Moreover, if $b = \beta^3$, then the transformation $(x, y) \mapsto (x/\beta^2, y/\beta^3)$ lets us assume that $b = 1$ and thus gives the form

$$E/K : y^2 + axy + y = x^3 \ .$$

Note that β is not necessarily defined over K. However, Proposition 4 shows that once we start on such a curve, the 3-isogenies will preserve this form. It has discriminant $\Delta(E) = a^3 - 27$ and has a subgroup $\{\mathcal{O}_E, (0,0), (0,-1)\}$ of order 3. The point $(0,0)$ acts on points outside this subgroup by

$$(x,y) + (0,0) = \left(\frac{-y}{x^2}, \frac{-y}{x^3}\right) \ .$$

This curve is known as a *triangular* curve [BCKL15] and is isomorphic to the twisted Hessian curve [BCKL15, Theorem 5.3]

$$(a^3 - 27)x^3 + y^3 + 1 = 3axy \ .$$

SIDH. Let $e_A, e_B, f \in \mathbb{Z}_{\geq 0}$ such that $p = \ell_A^{e_A}\ell_B^{e_B} f - 1$ is prime. For $K = \mathbb{F}_{p^2}$ we can then find a supersingular curve E over K [Brö09] such that

$$\begin{aligned} \#E(K) &= (p+1)^2 \ , \\ E(K)[\ell_A^{e_A}] &= \mathbb{Z}/\ell_A^{e_A}\mathbb{Z} \times \mathbb{Z}/\ell_A^{e_A}\mathbb{Z} \ , \\ E(K)[\ell_B^{e_B}] &= \mathbb{Z}/\ell_B^{e_B}\mathbb{Z} \times \mathbb{Z}/\ell_B^{e_B}\mathbb{Z} \ . \end{aligned}$$

By having the two parties compute isogenies of degree $\ell_A^{e_A}$ resp. $\ell_B^{e_B}$ and composing we can define a key exchange algorithm [FJP14, §3.2] similar to Diffie–Hellman. Since these degrees are exponentially large, they cannot be computed directly by polynomial evaluation. Instead, we decompose an $\ell_A^{e_A}$-isogeny as a sequence of e_A isogenies of degree ℓ_A, which are efficiently computable for small ℓ_A [FJP14, §4] (typically $\ell_A \in \{2,3\}$). Focusing on one of the sides, the secret key is a tuple $(\gamma, \delta) \in \mathbb{Z}/\ell_A^{e_A}\mathbb{Z} \times \mathbb{Z}/\ell_A^{e_A}\mathbb{Z}$ where not both γ and δ are divisible by ℓ_A. Fixing a basis $E(K)[\ell_A^{e_A}] = \langle P, Q\rangle$, this corresponds to an isogeny with kernel $\langle [\gamma]P + [\delta]Q\rangle$. As the kernel is determined by its generator up to some invertible scalar multiple, and since at least one of the two scalars must be invertible, all keys can either be put in the form $(1, \delta)$ or $(\gamma, 1)$.

3 Isogenies on Weierstrass Curves

We begin by stating a straightforward, but rather useful theorem. By assuming to have knowledge on the action of an isogeny on a single point Q, we can translate this point by elements of the kernel to obtain a simple description of the isogeny. Many curve models have a natural choice for this point (eg. $Q = (0,0)$ in Montgomery form, see §4).

Theorem 1. *Let K be a field and E/K an elliptic curve in Weierstrass form. Let $G \subset E(\bar{K})$ be a finite subgroup defined over K and*

$$\phi : (x,y) \mapsto \left(f(x), c_0 y f'(x) + g(x)\right), \quad c_0 \in \bar{K}^* \ , \tag{2}$$

a separable isogeny such that $\ker(\phi) = G$. Let $Q \in E(\bar{K})$ such that $Q \notin G$. Then

$$f(x) = c_1(x - x_Q) \prod_{T \in G \setminus \{\mathcal{O}_E\}} \frac{(x - x_{Q+T})}{(x - x_T)} + f(x_Q), \quad \text{for } c_1 \in \bar{K}^* \ .$$

Proof. First note that the existence of ϕ follows from Vélu's formulas [Vél71], while a standard result [Gal12, Theorem 9.7.5] shows that it can be written in the form of (2) (where $f'(x)$ is the formal derivative df/dx of $f(x)$). More explicitly, following the notation of [Gal12, Theorem 25.1.6], there exist functions $u, t : G \setminus \{\mathcal{O}_E\} \to \bar{K}$ such that

$$f(x) = x + \sum_{T \in G_1 \cup G_2} \left(\frac{t(T)}{x - x_T} + \frac{u(T)}{(x - x_T)^2} \right) \ ,$$

where $G_2 \subset G$ is the set of points of order 2 and $G_1 \subset E(\bar{K})$ is such that

$$G = \{\mathcal{O}_E\} \cup G_2 \cup G_1 \cup \{-T : T \in G_1\} \ .$$

Moreover, $u(T) = 0$ if and only if T has order 2. Collecting denominators, it is then immediate that there exists a function $w \in \bar{K}[x]$ such that $\deg(w) = |G|$ and

$$f(x) = \frac{w(x)}{v(x)} \ , \quad \text{where } v(x) = \prod_{T \in G \setminus \{\mathcal{O}_E\}} (x - x_T) \ .$$

Now define

$$h(x) = w(x)v(x_Q) - w(x_Q)v(x) \ .$$

Note that clearly $h(x_Q) = 0$. Since the value of f is invariant under the action of points in G, we in fact have that $h(x_{Q+T}) = 0$ for all $T \in G$. Therefore it follows that $(x - x_{Q+T}) \mid h(x)$ for all $T \in G$. If for all $T_1, T_2 \in G$ such that $T_1 \neq T_2$ we have that $x_{Q+T_1} \neq x_{Q+T_2}$, then it immediately follows that

$$\prod_{T \in G} (x - x_{Q+T}) \mid h(x) \ .$$

Otherwise[1], assume we have $T_1, T_2 \in G$ such that $T_1 \neq T_2$ and $x_{Q+T_1} = x_{Q+T_2}$. Since any x-coordinate corresponds to at most 2 points on E, it follows that $Q + T_2 = \pm (Q + T_1)$. However, $Q + T_2 = Q + T_1$ implies that $T_1 = T_2$, which contradicts our assumption. Therefore $Q + T_2 = -(Q + T_1)$ and

$$\begin{aligned} [2]\phi(Q + T_1) &= \phi(Q + T_1) + \phi(Q + T_1) \\ &= \phi(Q + T_1) + \phi(Q + T_2) \\ &= \phi(Q + T_1) - \phi(Q + T_1) \\ &= \mathcal{O}_{\widetilde{E}} \ . \end{aligned}$$

Moreover,

$$\begin{aligned} [2](Q + T_1) = \mathcal{O}_E &\iff Q + T_1 + Q + T_1 = \mathcal{O}_E \\ &\iff Q + T_1 - (Q + T_2) = \mathcal{O}_E \\ &\iff T_1 = T_2 \ , \end{aligned}$$

which contradicts the assumption that $T_1 \neq T_2$. Thus $\psi_2(Q + T_1) \neq 0$ and by Lemma 1 we can conclude that $f'(x_{Q+T_1}) = 0$. Since away from the zeros of v we have

$$h(x) = \big(f(x) - f(x_Q)\big)v(x)v(x_Q) \ ,$$

[1] This proof is quite elementary. An alternative method (which is perhaps more illuminating) is to consider the divisor of $x - f(x_Q)$ on E/G and to pull it back via ϕ. We can then use the fact that $\operatorname{div}(f(x) - f(x_Q)) = \phi^* \operatorname{div}(x - f(x_Q))$.

it follows from the fact that $f'(x_{Q+T_1}) = f(x_{Q+T_1}) - f(x_Q) = 0$ that $h'(x_{Q+T_1}) = 0$. That is, h has (at least) a double root at x_{Q+T_1}. In other words,

$$(x - x_{Q+T_1})(x - x_{Q+T_2}) \mid h(x) \ .$$

It is then clear that indeed

$$\prod_{T \in G} (x - x_{Q+T}) \mid h(x) \ .$$

As $\deg(h) \leq \max(\deg(w), \deg(v)) = |G|$, there exists a constant $c \in K^*$ such that

$$h(x) = c \prod_{T \in G} (x - x_{Q+T}) \ .$$

Thus,

$$f(x) = \frac{w(x)}{v(x)} = \frac{h(x)}{v(x)v(x_Q)} + f(x_Q) \ .$$

The result follows by setting $c_1 = c/v(x_Q)$. □

Lemma 1. *Let the setup be as in Theorem 1 and let $R \in E(\bar{K}) \setminus G$. Then*

$$[2]\phi(R) = \mathcal{O}_{\widetilde{E}} \iff \psi_2(R) f'(x_R) = 0 \ ,$$

where ψ_2 is the 2-division polynomial.

Proof. Firstly note that $R \notin G$ and thus $\phi(R) \neq \mathcal{O}_{\widetilde{E}}$. Therefore, by definition of the 2-division polynomial on $\widetilde{E} = E/G$ it follows that

$$[2]\phi(R) = \mathcal{O}_{\widetilde{E}} \iff 2y_{\phi(R)} + \widetilde{a}_1 x_{\phi(R)} + \widetilde{a}_3 = 0 \ ,$$

where $\widetilde{a}_1$ and $\widetilde{a}_3$ are Weierstrass constants of $\widetilde{E}$ conform (1). Using the definition of ϕ and by recalling that (see eg. [Gal12, Theorem 9.7.5])

$$2g(x_R) = c_0(a_1 x_R + a_3) f'(x_R) - \widetilde{a}_1 f(x_R) - \widetilde{a}_3 \ ,$$

a straightforward computation shows that

$$2y_{\phi(R)} + \widetilde{a}_1 x_{\phi(R)} + \widetilde{a}_3 = 0 \iff c_0 \left(2y_R + a_1 x_R + a_3\right) f'(x_R) = 0 \ .$$

Finally observe that $\psi_2(R) = 2y_R + a_1 x_R + a_3$ and $c_0 \neq 0$. □

Remark 1. Theorem 1 shows the connection between ϕ and the action of the point Q on abscissas of kernel elements, as ϕ is given by a product of functions

$$\frac{x - x_{Q+T}}{x - x_T} \ .$$

If this action is simple (eg. in Montgomery form where $x_{(0,0)+T} = 1/x_T$) then we can expect simple formulas for isogenies.

Remark 2. By relying on Theorem 1 we simplify the proof compared to earlier works [CH17, MS16]. Whereas those works present rational maps and prove them to be isogenies, we turn this argument around. We use the existence of the isogeny (by Vélu's formulas) and apply appropriate isomorphisms to enforce some structure to be maintained (eg. $(0,0) \mapsto (0,0)$ in Montgomery form). We can then apply Theorem 1 to get formulas for the isogeny up to some constants. Finally we also use the formal group law. However, as opposed to proving that the rational functions defining the isogeny satisfy the curve relation of the co-domain curve, we can assume them to vanish and therefore extract the constants and the coefficients of the co-domain curve. This significantly simplifies the proof compared to earlier works (eg. [MS16, Theorem 2] and [CH17, Theorem 1]).

4 Montgomery Form and 2-isogenies

In [CH17, Theorem 1] Costello and Hisil present rational maps which they prove to be isogenies between Montgomery curves. These isogenies are not unique, and are for example different from the formulas directly derived using Vélu's formulas. It is immediate that the isogenies in [CH17] have the property of fixing $(0,0)$. In §4.1 we show that this fact, together with the co-domain curve being in Montgomery form, characterizes their formulas (up to some sign choices). This generalizes the theorem by Costello and Hisil, by removing the restriction of kernels being cyclic and having odd order. In particular, in §4.2 we present formulas for 2-isogenies determined by points of order 2 other than $(0,0)$. Until now these had not appeared, and were considered to require the computation of a square root. In §4.3 we show how one could apply these formulas in an implementation. Although it requires only a modest change to the parameters, this does require care and can simplify the implementation. Finally in §4.4 we comment on a comparison to the state-of-the-art.

4.1 The General Formula

We begin by stating Proposition 1, which is the analogue of [CH17, Theorem 1].

Proposition 1. *Let K be a field with $\mathrm{char}(K) \neq 2$. Let $a \in K$ such that $a^2 \neq 4$ and $E/K : y^2 = x^3 + ax^2 + x$ is a Montgomery curve. Let $G \subset E(\bar{K})$ be a finite subgroup such that $(0,0) \notin G$ and let ϕ be a separable isogeny such that $\ker(\phi) = G$. Then there exists a curve $\widetilde{E}/K : y^2 = x^3 + Ax^2 + x$ such that, up to post-composition by an isomorphism,*

$$\begin{aligned} \phi : E &\to \widetilde{E} \\ (x,y) &\mapsto (f(x), c_0 y f'(x)) \end{aligned}$$

where

$$f(x) = x \prod_{T \in G \setminus \{\mathcal{O}_E\}} \frac{x x_T - 1}{x - x_T} \ .$$

Moreover, writing

$$\pi = \prod_{T \in G \setminus \{\mathcal{O}_E\}} x_T \,, \qquad \sigma = \sum_{T \in G \setminus \{\mathcal{O}_E\}} \left(x_T - \frac{1}{x_T} \right) \ ,$$

we have that $A = \pi(a - 3\sigma)$ and $c_0^2 = \pi$.

Proof. Over $\bar{K}$ we can always move E/G to Montgomery form. Let $P \in E(\bar{K})$ such that $x_P = 1$. Then $[2]P = (0,0)$, hence $[2]\phi(P) = \phi([2]P) \neq \mathcal{O}_{E/G}$ while $[4]\phi(P) = [2]\,(0,0) = \mathcal{O}_{E/G}$. Thus $\phi(P)$ is a point of exact order 4, and we apply an isomorphism such that $x_{\phi(P)} = (-1)^{|G|-1}$ (see eg. [FJP14, §4.3.2]). In particular this assures that $\phi : (0,0) \mapsto (0,0)$. We then twist the y-coordinate via another isomorphism to set the coefficient of y^2 to 1 and have

$$\widetilde{E} = E/G : y^2 = x^3 + Ax^2 + x \ .$$

Now apply Theorem 1 with $Q = (0,0)$. We obtain that

$$\begin{aligned} f(x) &= c_1(x - x_{(0,0)}) \prod_{T \in G \setminus \{\mathcal{O}_E\}} \frac{(x - x_{(0,0)+T})}{(x - x_T)} + f(x_{(0,0)}) \\ &= c_1 x \prod_{T \in G \setminus \{\mathcal{O}_E\}} \frac{\left(x - \frac{1}{x_T}\right)}{(x - x_T)} \ . \end{aligned}$$

As we set up $\widetilde{E}$ such that $f(1) = (-1)^{|G|-1}$, we find that

$$c_1 = \prod_{T \in G \setminus \{\mathcal{O}_E\}} x_T \ .$$

Feeding c_1 back into the equation for f puts it in the right form. At this point it only remains to find A and c_0 (observe that $g = 0$ in Montgomery form [Gal12, Theorem 9.7.5]). To this end we utilize the formal group law, similar to [CH17, MS16].

Let $t = x/y$ be a uniformizer at $\mathcal{O}_E$ and write $s = 1/y$. By observing that $s = t^3 + at^2 s + ts^2$ we can recursively substitute s into itself to get an expression $s(t) \in \mathbb{Z}[a][\![t]\!]$ as a power series[2]

$$s(t) = t^3 + at^5 + (a^2 + 1)t^7 + O(t^9)$$

This is well-defined, see for example [Sil09, §IV.1]. As a result we can write

$$\begin{aligned} 1/s(t) = y(t) &= t^{-3} - at^{-1} + O(t) \ , \\ ty(t) = x(t) &= t^{-2} - a + O(t^2) \ . \end{aligned}$$

Let $X(t) = f(x(t))$. Then

$$\begin{aligned} X(t) &= \pi t^{-2} + \pi(\sigma - a) + O(t^2) \ , \\ dX/dt &= -2\pi t^{-3} + O(t) \ , \\ dx/dt &= -2t^{-3} + O(t) \ , \\ (dx/dt)^{-1} &= -t^3/2 + O(t^7) \ . \end{aligned}$$

Now define

$$\begin{aligned} Y(t) &= c_0 y(t) \cdot (df/dx) \\ &= c_0 y(t) \cdot (dX/dt) \cdot (dx/dt)^{-1} \ , \end{aligned}$$

[2] We denote by $O(t^n)$ a series whose coefficients of t^m are zero for all $m < n$.

so that

$$Y(t) = c_0\pi t^{-3} - c_0 a\pi t^{-1} + O(t) \ .$$

Writing

$$F(t) = Y(t)^2 - \left(X(t)^3 + AX(t)^2 + X(t)\right)$$

it follows that

$$F(t) = F_{-6} \cdot t^{-6} + F_{-4} \cdot t^{-4} + O(t^{-2}) \ ,$$

with

$$F_{-6} = \pi^2(c_0^2 - \pi)\,, \quad F_{-4} = \pi^2\left(3\pi(a-\sigma) - 2ac_0^2 - A\right) \ .$$

Now since by assumption ϕ is an isogeny with co-domain curve $\widetilde{E}$, and since F is precisely the equation defining $\widetilde{E}$, we must have $F = 0$. Solving $F_{-6} = 0$ and $F_{-4} = 0$ simultaneously leads to the desired equations for c_0^2 and A. Note that this way we have only defined c_0 up to sign. However, the sign choice merely induces a composition with $[-1]$ and therefore does not affect ϕ up to isomorphism. □

Remark 3. It is perhaps not immediately obvious that Proposition 1 is a generalization of the result by Costello and Hisil [CH17, Theorem 1]. Our result assumes the domain curve E to be of the form $y^2 = x^3 + ax^2 + x$, while their theorem also accounts for curves $E_0 : by^2 = x^3 + ax^2 + x$. Moreover, the map itself looks slightly different. However, it is straightforward to check that if one pre-composes with the isomorphism

$$\begin{aligned} \psi_0 : E_0 &\to E \\ (x, y) &\mapsto (x, y\sqrt{b}) \end{aligned}$$

and post-composes with the isomorphism

$$\begin{aligned} \psi_1 : \widetilde{E} \to E_1 : By^2 + x^3 + Ax^2 + x \ , \\ (x, y) \mapsto \left(x, \frac{y}{\sqrt{\pi b}}\right) \end{aligned}$$

then one recovers the theorem from Costello and Hisil in the case of odd-degree cyclic kernels. Ignoring these twists in Proposition 1 simplifies the proof. For example, see Proposition 2.

Remark 4. If $K = \mathbb{F}_q$ is a (large-characteristic) finite field, then possibly π is a non-square in $\mathbb{F}_q$. As a result ϕ is not defined over $\mathbb{F}_q$. However, in that case the map

$$(x, y) \mapsto (f(x), yf'(x))$$

is defined over $\mathbb{F}_q$ with co-domain curve $\widetilde{E}^{(t)} : \pi y^2 = x^3 + Ax^2 + x$. This is the quadratic twist of $\widetilde{E}$. Since $\widetilde{E}$ and its twist have the same Kummer line, we eliminate this issue by projecting to $\mathbb{P}^1$ (ie. by using x-only arithmetic).

Remark 5. If we set up an SIDH instance with $\ell_A = 2$ and $e_A \geq 2$ then the x-coordinates of points of order 2 are in fact squares. This follows from [Hus04, Ch. 1, Thm 4.1] combined with the doubling formulas for Montgomery curves, as noted in [CJL+17, §3.2]. Since all x-coordinates of points with orders other than 2 appear twice in the equation for π, it follows that π is actually a square. Therefore ϕ is defined over $\mathbb{F}_{p^2}$, and in particular we always have $\#E(\mathbb{F}_{p^2}) = \#\widetilde{E}(\mathbb{F}_{p^2})$. This is (implicitly) used in formulas for public-key compression [CJL+17, ZJP+17].

4.2 2-isogenies

As an immediate consequence of Proposition 1 we obtain formulas for 2-isogenies for 2-torsion points other than $(0,0)$.

Proposition 2. *Let K be a field with $\mathrm{char}(K) \neq 2$. Let $a, b \in K$ such that $b \neq 0$ and $a^2 \neq 4$, and $E/K : by^2 = x^3 + ax^2 + x$ is a Montgomery curve. Let $P \in E(\bar{K})$ such that $P \neq (0,0)$ and $[2]P = \mathcal{O}_E$. Then*

$$\begin{aligned} \phi : E &\to \widetilde{E}/K : By^2 = x^3 + Ax^2 + x \\ (x, y) &\mapsto (f(x), yf'(x)) \ , \end{aligned}$$

with $B = x_P b$ and $A = 2(1 - 2x_P^2)$ is a 2-isogeny with $\ker(\phi) = \langle P \rangle$, where

$$f(x) = x \cdot \frac{xx_P - 1}{x - x_P} \ .$$

Proof. This is exactly the statement in Proposition 1 composed with the isomorphisms ψ_0 and ψ_1 from Remark 3. The result follows by using the identity $ax_P = -(x_P^2 + 1)$ to derive A. □

We also compute the kernel of the dual of ϕ, which will be helpful in §4.3 for larger degree isogenies.

Corollary 1. *Let the setup be as in Proposition 2. Then $\ker(\widehat{\phi}) = \langle (0,0) \rangle$.*

Proof. Let ψ be a separable isogeny with domain $\widetilde{E}$ and kernel $\langle (0,0) \rangle$. Then certainly $E[2] \subset \ker(\psi \circ \phi)$, and since $\deg(\psi \circ \phi) = 4$ we in fact have $E[2] = \ker(\psi \circ \phi)$. Thus $\psi = \widehat{\phi}$ up to isomorphism by uniqueness of the dual isogeny, and hence $\ker(\widehat{\phi}) = \ker(\psi)$. □

The statement and proof of Proposition 2 does not explain why we are able to compute 2-isogenies without explicit square roots, while earlier works [CH17, FJP14] could not. We provide a more direct computation in Remark 6 to show why this is the case.

Remark 6. In [FJP14, §4.3.2] the authors describe a 2-isogeny with kernel (0,0) as

$$\begin{aligned} \varphi : E &\to F : by^2 = x^3 + (a+6)x^2 + 4(2+a)x \\ (x, y) &\mapsto \left(\frac{(x-1)^2}{x}, y\left(1 - \frac{1}{x^2}\right) \right) \ . \end{aligned}$$

The coefficient of x can be removed by computing $2\sqrt{a+2}$ and composing with the isomorphism

$$(x, y) \mapsto \left(\frac{x}{2\sqrt{a+2}}, \frac{y}{2\sqrt{a+2}} \right) \ ,$$

putting F in the desired form. This requires computing a square root, which could be avoided by having knowledge of a point $P_8 = \left(2\sqrt{a+2}, —\right)$ of order 8 above $(0,0)$. Instead, we observe that we can compose with the isomorphism

$$\begin{aligned} \psi : F &\to G : \frac{b}{\sqrt{a^2-4}} y^2 = x^3 - \frac{2a}{\sqrt{a^2-4}} \cdot x^2 + x \\ (x, y) &\mapsto \left(\frac{x + a + 2}{\sqrt{a^2-4}}, \frac{y}{\sqrt{a^2-4}} \right) \ , \end{aligned}$$

which moves the kernel of $\widetilde{\varphi}$ to $(0,0)$. This requires computing $\sqrt{a^2-4}$ and therefore also relies on a square root. However, if $P_2 = (x_2, 0)$ is a point of order 2 on E with $x_2 \neq 0$, then $x_2^2 + ax_2 + 1 = 0$. Therefore it is immediate that

$$\sqrt{a^2-4} = 2x_2 + a \ ,$$

allowing us to compute the isomorphism efficiently. We have such a point by assumption in Proposition 2. We can now compute ϕ as $\psi \circ \varphi \circ \chi$, where χ is an isomorphism mapping P_2 to $(0,0)$ (eg. [FJP14, Equation (15)]).

To provide explicit operation counts[3] we move to projective space and project to $\mathbb{P}^1$. Let $P = (X_P : 0 : Z_P)$ be a point of order 2 on $E : bY^2Z = X^3 + aX^2Z + XZ^2$ such that $X_P \neq 0$. Then by Proposition 2

$$\begin{aligned} \phi : E &\to \widetilde{E} : BY^2Z = X^3 + AX^2Z + XZ^2 \\ (X : — : Z) &\mapsto (X(XX_P - ZZ_P) : — : Z(XZ_P - ZX_P)) \end{aligned}$$

is a 2-isogeny with kernel $\langle P \rangle$. We have

$$A = 2\,(Z_P^2 - 2X_P^2)/Z_P^2 \ ,$$

and to avoid inversions we represent it projectively as

$$(A : 1) = (2\,(Z_P^2 - 2X_P^2) : Z_P^2) \ .$$

However, the doubling formulas on Montgomery curves use $(A+2)/4$ instead of A, and we see that

$$(A + 2 : 4) = (Z_P^2 - X_P^2 : Z_P^2) \ .$$

This can be computed in $2\mathbf{S} + 1\mathbf{a}$. Moreover, we observe that

$$\begin{aligned} X(XX_P - ZZ_P) &= X\Big[(X - Z)(X_P + Z_P) - (X + Z)(Z_P - X_P)\Big] \ , \\ Z(XZ_P - ZX_P) &= Z\Big[(X - Z)(X_P + Z_P) + (X + Z)(Z_P - X_P)\Big] \ . \end{aligned}$$

This can be computed in $4\mathbf{M} + 6\mathbf{a}$ via the sequence of operations

$$\begin{gathered} T_0 = X_P + Z_P \,,\ T_1 = X_P - Z_P \,,\ T_2 = X + Z \,,\ T_3 = X - Z \,,\ T_4 = T_3 \cdot T_0 \,, \\ T_5 = T_2 \cdot T_1 \,,\ T_6 = T_4 - T_5 \,,\ T_7 = T_4 + T_5 \,,\ T_8 = X \cdot T_6 \,,\ T_9 = Z \cdot T_7 \ . \end{gathered}$$

If we assume $X_P + Z_P$ and $Z_P - X_P$ to be pre-computed, the cost reduces to $4\mathbf{M} + 4\mathbf{a}$. This would for example apply if we require multiple evaluations of the isogeny (eg. in SIDH). Also note that $Z_P^2 - X_P^2 = (X_P + Z_P)(Z_P - X_P)$ which allows us to compute the curve coefficient in $\mathbf{M} + \mathbf{S}$. This may or may not be worth it, depending on the underlying architecture.

[3] We denote by $\mathbf{M}$, $\mathbf{S}$ resp. $\mathbf{a}$ the cost of a field multiplication, squaring resp. addition or subtraction (which are assumed to have equal cost).

4.3 Application to Isogeny-based Cryptography

In the general setting it is not true that the kernels appearing in the computations cannot contain the point $(0,0)$, so it is not clear that the 2-isogenies can immediately be used. In a similar fashion, it is not true in general that kernels of 4-isogenies cannot contain $(1, \pm\sqrt{(a+2)/b})$ or $(-1, \pm\sqrt{(a-2)/b})$. In [CLN16a, §3] and [CH17] this assumption is used without justification (implicitly by replacing ψ_4 with $\widehat{\psi}_4$). This is dealt with by using a separate function `first_4_isog` for the first 4-isogeny, which is the only kernel that can contain such a point (a proof of which does not appear). However, Lemma 2 and Corollary 2 show that we can avoid these points with only a minor restriction on the keyspace. Applying this restriction to [CLN16a] makes the function `first_4_isog` redundant, simplifying the implementation.

Lemma 2. *Let $e, f \in \mathbb{Z}_{\geq 0}$ and let $p = 2^e 3^f - 1$ be prime. Let $E/\mathbb{F}_{p^2}$ be a supersingular elliptic curve in Montgomery form such that $\#E(\mathbb{F}_{p^2}) = (p+1)^2$. Let $P, Q \in E(\mathbb{F}_{p^2})$ such that $E[2^e] = \langle P, Q\rangle$ and $[2^{e-1}]Q = (0,0)$. Let $\alpha \in \mathbb{Z}_{2^e}$. Then $(0,0) \notin \langle P + [\alpha]Q\rangle$.*

Proof. It is clear that $\langle P + [\alpha]Q\rangle$ can only contain a single point of order 2, namely $[2^{e-1}](P + [\alpha]Q)$. But by assumption on Q we know that $[2^{e-1}](P + [\alpha]Q) \neq (0,0)$, hence the result follows.

By Lemma 2 we know that we can compute the 2^e-isogenies as defined in Proposition 1. However, as the degrees grow this will quickly be impractical. Instead, we do the computations as a sequence of 2-isogenies (ie. as in Proposition 2) [FJP14, §4]. Therefore we must show that none of these intermediate isogenies has a kernel generated by $(0,0)$.

Corollary 2. *Let the setup be as in Lemma 2 and write $R = P + [\alpha]Q$. Let ϕ be an isogeny such that $\ker(\phi) = \langle R\rangle$ and suppose that we compute*

$$\begin{aligned} \phi &= \phi_{e-1} \circ \cdots \circ \phi_0 \ , \\ \ker(\phi_0) &= \langle [2^{e-1}]R\rangle \ , \\ \ker(\phi_i) &= \left\langle \left[2^{e-i-1}\right] \phi_{i-1} \cdots \phi_0 \left(R\right)\right\rangle , \quad (\textit{for } 1 \leq i \leq e-1) \end{aligned}$$

as a sequence of 2-isogenies, each one computed as in Proposition 2. Then $(0,0) \notin \ker(\phi_i)$ for all $0 \leq i \leq e-1$.

Proof. We apply induction on i. The statement for $i = 0$ follows from Lemma 2. Let $i > 0$. Then $\ker(\widehat{\phi}_{i-1}) = \langle(0,0)\rangle$ by the inductive hypothesis and by Corollary 1. But since the walk determined by ϕ is non-backtracking, it follows that $\ker(\phi_i) \neq \langle(0,0)\rangle$. As $\#\ker(\phi_i) = 2$, we conclude that $(0,0) \notin \ker(\phi_i)$.

The keyspace is determined by tuples (γ, δ) which define kernels of the form $\langle [\gamma]P + [\delta]Q\rangle$, where not simultaneously $\gamma \equiv 0 \bmod 2$ and $\delta \equiv 0 \bmod 2$. We can divide the space into the three disjoint sets (of equal size)

$$\mathcal{K}_{(i,j)} = \{(\gamma, \delta) : \gamma \equiv i \bmod 2\,, \delta \equiv j \bmod 2\} \ ,$$

for $(i,j) \in \{(0,1),(1,0),(1,1)\}$. The restriction on the keyspace then corresponds exactly to disallowing $\mathcal{K}_{(0,1)}$, removing 1/3 of the keyspace. It is easy to see that these keys define the isogeny walks for which the first 2-isogeny has kernel $\langle(0,0)\rangle$. Note that this depends on the choice of 2^e-torsion basis $\{P, Q\}$, where we choose Q to lie above $(0,0)$. A similar argument applies to the use of 4-isogenies in [CLN16a].

Remark 7. The initial proposal to use curves in Montgomery form [CLN16a, §4] suggested taking P as an $\mathbb{F}_p$-rational point on the curve $E_0/\mathbb{F}_p : y^2 = x^3 + x$ with $j(E_0) = 1728$ and Q as the image of P under the distortion map $(x, y) \mapsto (-x, iy)$. This allows a compressed representation of $\{P, Q\}$. Although this does not work for the basis as chosen in Lemma 2, it only results in a small increase in the size of public parameters (which never need to be transferred).

4.4 Relating 2-isogenies and 4-isogenies

It is easy to see that the 4-isogenies from [CH17, Appendix A], which are currently the fastest formulas, can be derived by applying the 2-isogenies from §4.2 twice. That is, since they have equal kernel they are equal up to composition with an isomorphism. Both isogenies have a Montgomery curve as co-domain, of which there are at most six per isomorphism class (by looking at the formula for the j-invariant). Also, in both cases the dual is generated by a point $P \in \{(1, \pm\sqrt{(a+2)/b}), (-1, \pm\sqrt{(a-2)/b})\}$. Therefore we can transform one into the other by possibly composing with the simple isomorphisms $(x, y) \mapsto (x, -y)$ and $(x, y) \mapsto (-x, iy)$, where $i \in \bar{K}$ such that $i^2 = -1$. As a result, applying the 2-isogenies twice will not have more efficient formulas than the 4-isogenies. Indeed, if this were the case we could use the above transformation to obtain equally fast 4-isogenies. We summarize the costs in Table 1.

Table 1. Comparison of the costs of evaluating 2-isogenies and 4-isogenies.

Operation	**2-isogeny**	**2×2-isogeny**	**4-isogeny** [CH17]
Compute $(A + 2 : 4)$	$2\mathbf{S} + 1\mathbf{a}$	$4\mathbf{S} + 2\mathbf{a}$	$4\mathbf{S} + 5\mathbf{a}$[4]
First evaluation	$4\mathbf{M} + 6\mathbf{a}$	$8\mathbf{M} + 12\mathbf{a}$	$6\mathbf{M} + 2\mathbf{S} + 6\mathbf{a}$
Subsequent evaluations	$4\mathbf{M} + 4\mathbf{a}$	$8\mathbf{M} + 8\mathbf{a}$	$6\mathbf{M} + 2\mathbf{S} + 6\mathbf{a}$

Besides their theoretic value, there are some small upsides to using 2-isogenies in an implementation. Firstly, the computation leaks only a single bit as opposed to two [FJP14, §4.3.2]. Instead of leaking the dual of the final 4-isogeny, it would only leak the dual of the last 2-isogeny. Also, in some cases one may be able to select smaller parameters for a certain given security level. Primes of the form $2^e 3^f - 1$ where $e \approx \log_2(3^f)$ are somewhat sparse, and depending on one's requirements restricting e to be even could result in a (much) larger prime than hoped for. Alternatively, one could of course achieve this by doing a single 2-isogeny followed by a chain of 4-isogenies. However, this does come at the cost of having to implement more algorithms, increasing the size and complexity of an (already complex) implementation. Finally, having worked out formulas for isogenies of even degree and by showing how to avoid $(0, 0)$, we are able to straightforwardly write down formulas for 2^e-isogenies with $e \geq 3$. It remains to be seen if these can be made more efficient than repeated applications of 4-isogenies.

[4] Many of these additions are not needed to compute $(A + 2 : 4)$, but are used as pre-computation for the isogeny evaluation. We provide the counts as is to align with [CH17] since it does not affect our comparison of the costs of large degree isogeny evaluations.

5 Triangular Form and 3-isogenies

Given the generality of Theorem 1, an obvious question is whether there are other classes of curves which could possibly give rise to simple formulas for isogenies. In this section we analyze curves in triangular form $E/K : y^2 + axy + y = x^3$ containing a point $(0, 0)$ of order 3. Most of the ideas from earlier sections apply and in particular we get analogous statements for computing 3-isogenies (see §5.2). Although these allow to compute the co-domain curve very efficiently, the evaluation of the isogeny is not as efficient as its Montgomery counterpart. Moreover, since tripling formulas are currently slower, at this point Montgomery form still performs better with respect to 3-isogenies.

5.1 The General Formula

We start by presenting formulas for triangular curves that work for any separable isogeny whose kernel is an odd order subgroup. It is possible to include groups of even order, but this creates a case distinction which makes the proof more tedious. Since having (enough) rational points of even order would enable us to go to Montgomery form and reduce to §4, we discard that case here.

There are a couple of (minor) complications compared to the proof of Proposition 1. Firstly, we cannot assume that $g = 0$. If we work on $\mathbb{P}^1$ this will not affect the efficiency, but we will have to take it into account in the proof. Secondly, the action of $(0, 0)$ does not involve only x-coordinates. To eliminate the y-coordinates that arise, we group the kernel points into sets $\{T, -T\}$ (similar to [CH17, Theorem 1]).

Proposition 3. *Let K be a field with $\mathrm{char}(K) \neq 2$. Let $a \in K$ such that $a^3 \neq 27$ and $E/K : y^2 + axy + y = x^3$ in triangular form. Let $G \subset E(\bar{K})$ be a finite subgroup of odd order such that $(0, 0) \notin G$ and let ϕ be a separable isogeny such that $\ker(\phi) = G$. Let*

$$X = \left\{ x_P \;\middle|\; P \in G \setminus \{\mathcal{O}_E\} \right\} \ .$$

Then there exists a curve $\widetilde{E}/K : y^2 + Axy + y = x^3$ such that, up to post-composition by an isomorphism,

$$\begin{aligned} \phi : E &\to \widetilde{E} \\ (x, y) &\mapsto (f(x), c_0 y f'(x) + g(x)) \end{aligned}$$

where

$$f(x) = x \prod_{z \in X} \frac{x^2 z^2 - x(az + 1) - z}{(x - z)^2} \ .$$

Moreover, writing

$$\pi = \prod_{z \in X} z \,, \qquad \sigma = \sum_{z \in X} \left(\frac{1}{z^2} + \frac{a}{z} - 2z \right) \ ,$$

we have that $A^2 = \pi^2(a^2 + 12\sigma)$ and $c_0 = (-1)^{|X|}\pi$.

Proof. Let $P = (0,0)$. As $\phi(P) \neq \mathcal{O}_{E/G}$, while $[3]\phi(P) = \phi([3]P) = \mathcal{O}_{E/G}$, it follows that $\phi(P)$ must have exact order 3 on E/G. Therefore by moving $\phi(P)$ to the origin we can put E/G in triangular form and therefore assume that

$$\widetilde{E} = E/G : y^2 + Axy + y = x^3 \ .$$

Now apply Theorem 1 with $Q = P$. We find that

$$\begin{aligned} f(x) &= c_1(x - x_{(0,0)}) \prod_{T \in G \setminus \{\mathcal{O}_E\}} \frac{(x - x_{(0,0)+T})}{(x - x_T)} + f(x_{(0,0)}) \\ &= c_1 \prod_{T \in G \setminus \{\mathcal{O}_E\}} \frac{\left(x + \frac{y_T}{x_T^2}\right)}{(x - x_T)} \\ &= c_1 \prod_{x_T \in X} \frac{\left(x + \frac{y_T}{x_T^2}\right)\left(x + \frac{y_{-T}}{x_T^2}\right)}{(x - x_T)^2} \\ &= c_1 \prod_{x_T \in X} \frac{x^2 - \frac{x(ax_T+1)}{x_T^2} - \frac{1}{x_T}}{(x - x_T)^2} \ . \end{aligned}$$

Observe that we use the fact that there are no points of order 2, and that

$$y_T y_{-T} = -x_T^3 \,, \quad \text{and} \quad y_T + y_{-T} = -ax_T - 1 \ .$$

By [Gal12, Theorem 9.7.5] we can write

$$g(x) = \frac{1}{2}\left(-Af(x) - 1 + c_0 a x f'(x) + c_0 f'(x)\right) \ ,$$

so that $g(0) = \left(-1 + c_0 f'(0)\right)/2$. Now we use the fact that $\phi([2]P) = [2]\phi(P)$, ie. $\phi : (0,-1) \mapsto (0,-1)$. Therefore

$$\begin{aligned} & -1 = -c_0 f'(0) + g(0) \\ \iff & -1 = \left(-1 - c_0 f'(0)\right)/2 \\ \iff & c_0 = 1/f'(0) \\ \iff & c_0 = (-1)^{|X|}\pi^3/c_1 \ . \end{aligned} \tag{3}$$

It remains to find A and c_1 and for this we use the same strategy as earlier. Let $t = x/y$ be the uniformizer at $\mathcal{O}_E$ and write $s = 1/y$. Then as a power series

$$s(t) = t^3 - at^4 + a^2 t^5 + O(t^6) \ .$$

As $y = 1/s$ and $x = ty$ we find that

$$\begin{aligned} x(t) &= t^{-2} + at^{-1} + O(t) \ , \\ y(t) &= t^{-3} + at^{-2} + O(1) \ . \end{aligned}$$

Letting $X(t) = f(x(t))$ we get

$$\begin{aligned} X(t) &= c_1 t^{-2} + a c_1 t^{-1} - c_1 \sigma + O(t) \ , \\ dX/dt &= -2c_1 t^{-3} - a c_1 t^{-2} + O(1) \ , \\ dx/dt &= -2t^{-3} - a t^{-2} + O(t) \ , \\ (dx/dt)^{-1} &= -t^3/2 + a t^4/4 - a^2 t^5/8 + O(t^6) \ . \end{aligned}$$

It follows that

$$\begin{aligned} g(x(t)) &= \frac{1}{2}\left(-AX(t) - 1 + c_0 \cdot (dX/dt) \cdot (dx/dt)^{-1} \cdot (ax(t) + 1)\right) \\ &= \frac{1}{2}\left(a c_0 c_1 - A c_1\right) t^{-2} + \frac{1}{2} a \left(a c_0 c_1 - A c_1\right) t^{-1} + O(1) \ . \end{aligned}$$

Now define $Y(t) = c_0 y(t) \, (dX/dt) \cdot (dx/dt)^{-1} + g(x(t))$ and

$$F(t) = Y(t)^2 + AX(t)Y(t) + Y(t) - X(t)^3 \ .$$

We get that

$$F(t) = F_{-6} \cdot t^{-6} + F_{-5} \cdot t^{-5} + F_{-4} \cdot t^{-4} + O(t^{-2}) \ ,$$

where

$$\begin{aligned} F_{-6} &= c_1^2 \left(c_0^2 - c_1\right) \ , \\ F_{-5} &= 3 a c_1^2 \left(c_0^2 - c_1\right) \ , \\ F_{-4} &= c_1^2 \left(13 a^2 c_0^2/4 - A^2/4 + 3 c_1 \sigma - 3 a^2 c_1\right) \ . \end{aligned}$$

Again, as F is precisely the equation defining $\widetilde{E}$, we must have $F_{-6} = F_{-5} = F_{-4} = 0$. The first two identities lead to $c_1 = c_0^2$, which together with (3) gives $c_1^3 = \pi^6$. Therefore $c_1 = \zeta_3 \pi^2$ where $\zeta_3 \in \bar{K}$ is such that $\zeta_3^3 = 1$. Inserting this into F_{-4} and equating to zero we find that

$$A^2 = \pi^2 \left(a^2 + 12\sigma\right) / \zeta_3^2 \ .$$

Therefore, by composing with the isomorphism $(x, y) \mapsto (\zeta_3^2 x, y)$ we can assume that $\zeta_3 = 1$. From (3) we get that $c_0 = (-1)^{|X|}\pi$. The result is now clear. □

5.2 3-isogenies

We work out explicit formulas for 3-isogenies.

Proposition 4. *Let K be a field with $\mathrm{char}(K) \neq 2$. Let $a \in K$ such that $a^3 \neq 27$ and $E/K : y^2 + axy + y = x^3$ in triangular form. Let $P \in E(\bar{K})$ a point such that $[3]P = \mathcal{O}_E$ and $x_P \neq 0$. Then*

$$\begin{aligned} \phi : E &\to \widetilde{E}/K : y^2 + Axy + y = x^3 \\ (x, y) &\mapsto \left(f(x), -x_P y f'(x) + g(x)\right) \end{aligned}$$

with $A = -3\left(2 + a x_P\right)$ is a 3-isogeny such that $\ker(\phi) = \langle P \rangle$, where

$$f(x) = x \cdot \frac{x^2 x_P^2 - x(a x_P + 1) - x_P}{\left(x - x_P\right)^2} \ .$$

Proof. This is Proposition 3 with $X = \{x_P\}$. Using the division polynomial

$$\psi_3(x) = x\left(3x^3 + a^2x^2 + 3ax + 3\right)$$

it follows that $9\,(2 + ax_P)^2 = \pi^2\left(a^2 + 12\sigma\right)$. Hence $A = \pm 3\,(2 + ax_P)$ and the only remaining uncertainty is the choice of sign. However, setting $A = -3\,(2 + ax_P)$, a direct computation shows that

$$f'(x) = x_P^2 \cdot \frac{\left((x - x_P)^3 - (6x_P^2 + a^2x_Px_P + a)x + x_P^3 + 1\right)}{(x - x_P)^3} \ ,$$

while

$$g(x) = x^3 \cdot \frac{\left((3 + ax_P)x_P^2x + x_P^3 + 1\right)}{(x - x_P)^3} \ .$$

For $X = f(x)$ and $Y = -x_Pyf'(x) + g(x)$, a straightforward calculation shows that $Y^2 + AXY + Y = X^3$. It is then clear that ϕ is an isogeny and that $\ker(\phi) = \langle P \rangle$. □

Again, as a consequence of fixing $(0, 0)$ the dual will be generated by it.

Corollary 3. *Let the setup be as in Proposition 4. Then* $\ker(\widehat{\phi}) = \langle (0, 0) \rangle$.

Proof. Since $(0, 0) \in E$ has order 3 and is not in $\ker(\phi)$, it follows from $\widehat{\phi} \circ \phi = [3]$ that $\phi\left((0, 0)\right) \neq \mathcal{O}_{\widetilde{E}}$, while $\left(\widehat{\phi} \circ \phi\right)\left((0, 0)\right) = \mathcal{O}_E$. Hence $\phi\left((0, 0)\right) \in \ker(\widehat{\phi})$, and since $\deg(\widehat{\phi}) = 3$ we have that $\ker(\widehat{\phi}) = \langle \phi\left((0, 0)\right) \rangle$. The result is now immediate by observing that $\phi\left((0, 0)\right) = (0, 0)$. □

5.3 Application to Isogeny-based Cryptography

By doing an analogous analysis as in §4.3 it is straightforward to see that it is theoretically possible to use the triangular form as above in isogeny-based systems. More specifically, by choosing a basis $E(\mathbb{F}_{p^2})[3^e] = \langle P, Q \rangle$ such that $[3^{e-1}]Q = (0, 0)$ and by only allowing secret kernels of the form $\langle P + [\alpha]Q \rangle$, we can always apply the isogeny from Proposition 4. However, to be seriously considered for implementations the efficiency must be at least on par with those coming from the Montgomery form. Although the computation of A can be done with only two multiplications, we have not been able to reduce the cost of the 3-isogeny evaluation far enough to be considered as efficient as its Montgomery counterpart. Moreover, the x-only tripling formulas (which can for example be obtained by using the 3-isogenies from [BCKL15, Theorem 5.4]) are significantly slower.

Acknowledgements. I would like to thank Craig Costello for valuable suggestions and feedback during the creation of this document, and Chloe Martindale for comments on a first version of the paper, in particular to improve the proof of Theorem 1. Thanks to Paulo Barreto for noticing an error in the operation counts of the 2-isogenies and to the anonymous reviewers of PQCrypto 2018 for their constructive comments.

References

[Acc99] Accredited Standards Committee X9. American National Standard X9.62-1999, Public key cryptography for the financial services industry: the elliptic curve digital signature algorithm (ECDSA). Technical report, ANSI, 1999. 1

[BCKL15] Daniel J. Bernstein, Chitchanok Chuengsatiansup, David Kohel, and Tanja Lange. Twisted Hessian Curves. In *Progress in Cryptology - LATINCRYPT 2015 - 4th International Conference on Cryptology and Information Security in Latin America, Guadalajara, Mexico, August 23-26, 2015, Proceedings*, pages 269–294, 2015. 2, 4, 17

[BDL+12] Daniel J. Bernstein, Niels Duif, Tanja Lange, Peter Schwabe, and Bo-Yin Yang. High-speed high-security signatures. *J. Cryptographic Engineering*, 2(2):77–89, 2012. 1

[Ber06] Daniel J. Bernstein. Curve25519: New Diffie-Hellman speed records. In M. Yung, Y. Dodis, A. Kiayias, and T. Malkin, editors, *Public Key Cryptography - PKC 2006, 9th International Conference on Theory and Practice of Public-Key Cryptography, New York, NY, USA, April 24-26, 2006, Proceedings*, volume 3958 of *Lecture Notes in Computer Science*, pages 207–228. Springer, 2006. 1

[Brö09] Reinier Bröker. Constructing Supersingular Elliptic Curves. *J. Comb. Number Theory*, 1(3):269–273, 2009. 4

[CH17] Craig Costello and Huseyin Hisil. A simple and compact algorithm for SIDH with arbitrary degree isogenies. Cryptology ePrint Archive, Report 2017/504, 2017. 2, 7, 8, 9, 10, 12, 13, 14

[CJL+17] Craig Costello, David Jao, Patrick Longa, Michael Naehrig, Joost Renes, and David Urbanik. Efficient compression of SIDH public keys. In *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part I*, pages 679–706, 2017. 2, 9

[CLN16a] Craig Costello, Patrick Longa, and Michael Naehrig. Efficient Algorithms for Supersingular Isogeny Diffie-Hellman. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, volume 9814 of *Lecture Notes in Computer Science*, pages 572–601. Springer, 2016. 2, 12, 13

[CLN16b] Craig Costello, Patrick Longa, and Michael Naehrig. SIDH Library, 2016. http://research.microsoft.com/en-us/downloads/bd5fd4cd-61b6-458a-bd94-b1f406a3f33f/. 1

[Cou06] Jean Marc Couveignes. Hard Homogeneous Spaces. *IACR Cryptology ePrint Archive*, 2006. 1

[DH76] Whitfield Diffie and Martin E. Hellman. New Directions in Cryptography. *IEEE Trans. Information Theory*, 22(6):644–654, 1976. 1

[FJP14] Luca De Feo, David Jao, and Jérôme Plût. Towards Quantum-Resistant Cryptosystems from Supersingular Elliptic Curve Isogenies. *J. Mathematical Cryptology*, 8(3):209–247, 2014. 2, 4, 8, 10, 11, 12, 13

[Gal12] Steven D. Galbraith. *Mathematics of Public Key Cryptography*. Cambridge University Press, 2012. 3, 4, 5, 6, 8, 15

[Hus04] Dale Husemöller. *Elliptic Curves*. Graduate Texts in Mathematics. Springer, 2004. 2, 3, 9

[JF11] David Jao and Luca De Feo. Towards Quantum-Resistant Cryptosystems from Supersingular Elliptic Curve Isogenies. In B. Yang, editor, *PQCrypto 2011*, volume 7071 of *LNCS*, pages 19–34. Springer, 2011. 1

[KAK16] Brian Koziel, Reza Azarderakhsh, and Mehran Mozaffari Kermani. Fast Hardware Architectures for Supersingular Isogeny Diffie-Hellman Key Exchange on FPGA. In *Progress in Cryptology - INDOCRYPT 2016 - 17th International Conference on Cryptology in India, Kolkata, India, December 11-14, 2016, Proceedings*, pages 191–206, 2016. 1, 2

[Kob87] N. Koblitz. Elliptic Curve Cryptosystems. *Mathematics of Computation*, 48:203–209, 1987. 1

[Len87] Hendrik W. Lenstra. Factoring Integers with Elliptic Curves. *The Annals of Mathematics*, 126:649–673, 1987. 1

[Mil86] Victor Miller. Use of Elliptic Curves in Cryptography. In *Advances in Cryptology - CRYPTO 85 Proceedings*, volume 218 of *Lecture Notes in Computer Science*, pages 417–426. Springer Berlin / Heidelberg, Berlin, Germany, 1986. 1

[Mon87] Peter L. Montgomery. Speeding the Pollard and elliptic curve methods of factorization. *Mathematics of computation*, 48(177):243–264, 1987. 1

[MS16] Dustin Moody and Daniel Shumow. Analogues of Vélu's formulas for isogenies on alternate models of elliptic curves. *Math. Comput.*, 85(300):1929–1951, 2016. 2, 7, 8

[RS06] Alexander Rostovtsev and Anton Stolbunov. Public-key cryptosystem based on isogenies. *IACR Cryptology ePrint Archive*, 2006:145, 2006. 1

[Sch89] Claus P. Schnorr. Efficient Identification and Signatures for Smart Cards. In Gilles Brassard, editor, *Advances in Cryptology - CRYPTO '89*, volume 435 of *LNCS*, pages 239–252. SV, 1989. 1

[Sho94] Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*, pages 124–134. IEEE, 1994. 1

[Sil09] Joseph H. Silverman. *The Arithmetic of Elliptic Curves, 2nd Edition*. Graduate Texts in Mathematics. Springer, 2009. 3, 8

[Vél71] Jacques Vélu. Isogénies entre courbes elliptiques. *Comptes Rendus de l'Académie des Sciences des Paris*, 273:238–241, 1971. 2, 3, 4

[ZJP+17] Gustavo H. M. Zanon, Marcos A. Simplicio Jr., Geovandro C. C. F. Pereira, Javad Doliskani, and Paulo S. L. M. Barreto. Faster isogeny-based compressed key agreement. Cryptology ePrint Archive, Report 2017/1143, 2017. https://eprint.iacr.org/2017/1143. 9