EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Algorithms on Elliptic Curves

Aleš Drápal
Charles University, Prague
School of Mathematics and Physics
2022

C. Curves and function fields

An elliptic curve is often regarded as a synonymous for a smooth Weierstraß curve. But in fact an elliptic curve is a much broader concept the essence of which can be expressed algebraically in the language of algebraic function fields.

This text does not aspire to provide a formal introduction into that theory. Nevertheless this introductory section presents several of its concepts and notions. The aim is to sketch what is the connection between the geometry of curves and the algebra of function fields.

Let us start by making some notational conventions and introductory definitions.

Let K be a field, and let $\bar{K}$ be an algebraic closure of K. Both K and $\bar{K}$ are regarded as fixed.

The *n-dimensional affine space* $\bar{K}^n$ is denoted by $\mathbb{A}^n$, and the set K^n by $\mathbb{A}^n(K)$. The elements of $\mathbb{A}^n(K)$ are called *K-rational points.*

If $f_1, \dots, f_k \in K[x_1, \dots, x_n]$ are polynomials, then $V_{f_1,\dots,f_k}$ denotes the set of all $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{A}^n$ such that $f_i(\alpha) = f_i(\alpha_1, \dots, \alpha_n) = 0$ for all $i \in \{1, \dots, k\}$.

A *planar affine curve* over K is any subset of $\mathbb{A}^2$ than can be expressed as V_f, where $f \in K[x_1, x_2]$, $\deg(f) \geq 0$.

In other words, planar affine curves are the zero points of nonzero polynomials in two variables. Since $K[x_1, x_2]$ is a UFD (unique factorization domain) each polynomial $f \in K[x_1, x_2]$, $\deg(f) \geq 1$, may be expressed uniquely, up to scalar multiples, as $f_1 \cdots f_k$, where each f_i is an irreducible polynomial.

If $f = f_1 \cdots f_k$, $k \geq 1$, then $V_f = V_{f_1} \cup \cdots \cup V_{f_k}$. For example, if $f = x_1 x_2$, then V_f is the union of the coordinate lines.

The case of $k > 1$ will be discussed only briefly. The main focus is upon the case $k = 1$.

C.1. **Coordinate rings and function fields.** Suppose that C is an affine planar curve. There are many $g \in K[x_1, x_2]$ such that $g(\alpha) = 0$ for every $\alpha \in C$. The set of all such g is closed under addition, and also under multiplication by another element of $K[x_1, x_2]$. This set is thus an ideal of the ring $K[x_1, x_2]$. It may be proved that this ideal is the principal ideal of a polynomial $f = f_1 \cdots f_k$, where $k \geq 1$, where each f_i is irreducible, $1 \leq i \leq k$, and where $(f_i) \neq (f_j)$ if $1 \leq i < j \leq k$. The latter condition says that the principal ideals of f_i and f_j are different. Since both f_i and f_j are irreducible, this means, in fact, that f_i is not a scalar multiple of f_j. (A scalar multiple always refers to a multiplication by a nonzero element of K. The group of nonzero elements of K is denoted by K^*).

For polynomials $g, h \in K[x_1, x_2]$ write $g \sim h$ if $g(\alpha) = h(\alpha)$ for each $\alpha \in C$. This is clearly an equivalence upon $K[x_1, x_2]$ such that $g \sim h$ if and only if $g - h$ vanishes on all points of C. In other words $g \sim h$ if and only if $g - h \in (f)$. Classes of $\sim$ thus coincide with cosets of the ideal (f).

Where there is an ideal, there is also a factor ring. The ring $K[x_1, x_2]/(f)$ is determined only by the curve C since f is determined by C uniquely, up to a scalar multiple. Hence it is correct to put

$$K[C] = K[x_1, x_2]/(f).$$

The ring $K[C]$ is called the *coordinate ring* of the curve C. Elements of $K[C]$ are cosets $a + (f)$, where a runs through $K[x_1, x_2]$.

Such a description of $K[C]$ is complete, correct and exhaustive. Nevertheless it is somewhat formal. Such a description will be called *algebraic.* Another term for it might be *syntactic.*

To move from syntax to semantics let us return to the above definition of $\sim$. By this definition, $g \sim h$ if and only if polynomials g and h *behave* identically upon C. The ring $K[C]$ may be understood as a collection of all possible polynomial behaviours on C. Note that in this way $K[C]$ could be defined without introducing any ideal since polynomials upon C may be both added and multiplied in a natural way. Such an approach to $K[C]$ will be termed *geometric* or, perhaps more exactly, *functional.*

Note that elements of $K[C]$ are defined with respect to all points of C. This is important to realize especially when working with finite fields. Points of C that are not K-rational always have to be taken into account. At first glance this may be regarded as superfluous since the group of an elliptic curve over K is defined only upon the K-rational points. However, there exist important and efficient algorithms that determine properties of such a group (like the order) that work with points that are not K-rational.

Note also that if K is finite then two elements of $K[C]$ may agree upon all K-rational points and yet be different.

Let $f \in K[x_1, x_2]$ be a polynomial of degree at least one and let $K[C] = K[x_1, x_2]/(f)$, $C = V_f$. The ring $K[C]$ is a domain if and only if f is irreducible. This is exactly when the curve C is called *irreducible.*

Recall that if R is a domain, then it is possible to construct the *fraction field* F, where $a/b = c/d$ if and only if $ad - bc = 0$.

Suppose that C is an irreducible planar affine curve. The fraction field of $K[C]$ will be denoted by $K(C)$ and called the *function field* of C.

The functions to which the name "function field" refers are the rational functions $a/b \in K(x_1, x_2)$. Note that $K(x_1, x_2)$ may be defined as the fraction field of the domain $K[x_1, x_2]$.

The algebraic approach to $K(C)$ stresses the formal description of its elements. Each element of $K(C)$ is equal to some $(a + (f))/(b + (f))$, where $C = V_f$, $f \in K[x_1, x_2]$ irreducible. Elements a, b run through $K[x_1, x_2]$, with $b \notin (f)$. The latter condition is equivalent to $b + (f) \neq 0_{K(C)}$. By the definition of fraction fields

$$\frac{a + (f)}{b + (f)} = \frac{c + (f)}{d + (f)} \;\Leftrightarrow\; ad - bc \in (f). \tag{C.1}$$

The functional interpretation of $K(C)$ is similar to that of $K[C]$. However, there is a technical difficulty which has to be cosidered. For $\sigma \in K(C)$ there are many $a/b \in K[x_1, x_2]$ such that $\sigma = (a + (f))/(b + (f))$. Each such a/b is said to be a *representative* of σ. Since $b \notin (f)$ there are only finitely many $\alpha \in C$ such that $b(\alpha) = 0$ (this is not a completely obvious fact, but the proof is relatively easy). Hence each representative of σ yields a mapping $C \to \bar{K}$ that is defined *nearly everywhere*, that is up to finitely many points of C. The technical difficulty mentioned above rests in the fact that if c/d is another representative of σ, then the points $\alpha \in C$ where $b(\alpha) = 0$ may be different from those where $d(\alpha) = 0$. However, (C.1) implies that if $b(\alpha) \neq 0$ and $d(\alpha) \neq 0$, then $a(\alpha)/b(\alpha) = c(\alpha)/d(\alpha)$.

With each $\sigma \in K(C)$ there thus may be associated a function $C \to \bar{K}$ that is defined for every $\alpha \in C$ for which there exists a representative a/b of σ such that $b(\alpha) \neq 0$. If a/b is such a representative, then $\sigma(\alpha) = a(\alpha)/b(\alpha)$. This definition is correct, as follows from (C.1).

The functional field $K(C)$ may be regarded as a collection of all partial mappings $C \to \bar{K}$ that may be obtained in such a way.

C.2. **Discrete valuations.** Let C be an irreducible planar affine curve. It turns out that many important properties of C depend only upon the algebraic structure of the function field $K(C)$.

The key notion in the algebraic analysis of $K(C)$ is the notion of *discrete valuation.* This is something quite natural that arose from the most basic properties of primes as they occur in every UFD.

Let R be a UFD and let F be the fraction field of R. For each irreducible $p \in R$ define $v_p(r)$, r a nonzero element of R, as the largest $k \geq 0$ such that $p^k \mid r$. By definition, $v_p(0) = \infty$.

Extend the definition of $v_p(r)$ from R to F by setting $v_p(r/s) = v_p(r) - v_p(s)$.

Put $\nu = v_p$. The following properties are true for all $a, b \in F$:

$$\nu(ab) = \nu(a) + \nu(b); \tag{DV1}$$

$$\nu(a+b) \geq \min\{\nu(a), \nu(b)\}, \tag{DV2}$$

$$\nu(a) = \infty \iff a = 0; \text{ and} \tag{DV3}$$

$$\exists a \in F,\ \nu(a) = 1. \tag{DV4}$$

Let now F be a field (no assumption is now being made about F being a fraction field of a domain R). A mapping $\nu\colon F \to \mathbb{Z} \cup \{\infty\}$ is called a *discrete valuation* if it fulfils (DV1)–(DV3). Discrete valuations that also fulfil (DV4) are called *normalized.*

Suppose that $F = K(C)$. The discrete valuations of F that are considered when investigating the curve C are those that fulfil this additional condition:

$$\nu(a) = 0 \text{ for every } a \in K^*.$$

They will be called valuations *over* K.

Let us pay attention to the way how the field K is embedded into $K(C)$. Both $K[C]$ and $K(C)$ are vector spaces over K. The unit in both of them is equal to $1 + (f)$, where $C = V_f$, $f \in K[x_1, x_2]$ irreducible. Consider $\lambda \in K$. The element λ is identified in both $K[C]$ and $K(C)$ with $\lambda \cdot 1_{K[C]} = \lambda \cdot 1_{K(C)} = \lambda + (f)$. The functional interpretation of $\lambda + (f)$ is clear: each $\alpha \in C$ is mapped upon λ.

For unique factorization domains (UFD) the notion of discrete valuation does not seem to bring much new. That is not completely true as shown by the ensuing analysis of $K(x)$. Furthermore, the coordinate ring $K[C]$ is rarely a UFD, and yet $K(C)$ contains many (in fact, infinitely many) normalized discrete valuations over K.

If $F = K(x)$ (the ring of rational functions in one variable), then each irreducible polynomial $p \in K[x]$ yields a normalized discrete valuation v_p. Besides them there exists exactly one normalized discrete valuation over K. This valuation is denoted by v_∞ and defined by $v_\infty(a/b) = \deg(b) - \deg(a)$.

Suppose now, for a while, that $\bar{K} = K$. In such a case the monic irreducible polynomials are the polynomials $x - \lambda$. With the exception of v_∞ each normalized discrete valuation over K thus may be identified with a unique point of the affine line $\mathbb{A}^1$. To give a geometric meaning to v_∞ extend the affine line $\mathbb{A}^1$ to the projective line $\mathbb{P}^1$. This means to add just one point. This point is called the *point at infinity.*

The connection "one point—one discrete valuation" is not limited to the projective line. The connection is valid for all irreducible projective planar curves over $\bar{K}$ that satisfy a certain additional condition. This will be precised later.

C.3. **Planar projective curves.** The formal definition of the n-dimensional *projective space* $\mathbb{P}^n$ states that $\mathbb{P}^n$ is equal to the set of all 1-dimensional subspaces of $\mathbb{A}^{n+1}$. However, a projective point (i.e., an element of $\mathbb{P}^n$) is usually treated by considering its homogeneous coordinates $(\alpha_1 : \alpha_2 : \cdots : \alpha_{n+1})$. The connection to the formal definition is made by considering these coordinates as representatives of the space of all $(\lambda\alpha_1, \lambda\alpha_2, \ldots, \lambda\alpha_{n+1})$, where λ runs through $\bar{K}$. This means that homogeneous coordinates represent the same point if and only if in all positions they differ by the same scalar multiple and that at least one position has to carry a nonzero entry.

A projective point is K*-rational* if it may be expressed as $(\alpha_1 : \dots : \alpha_{n+1})$, where $\alpha_i \in K$ for each $i \in \{1, \dots, n+1\}$.

It is usual to identify an affine point $(\alpha_1, \dots, \alpha_n) \in \mathbb{A}^n$ with the projective point $(\alpha_1 : \dots : \alpha_n : 1) \in \mathbb{P}^n$. The projective points that cannot be obtained in this way are called *points at infinity.* In $\mathbb{P}^1$ there is only one point at infinity and this point is equal to $(1 : 0)$.

Let $a = \sum a_{i_1,\dots,i_k} x_1^{i_1} \dots x_k^{i_k}$ be a polynomial over K. This polynomial is called *homogeneous* if its coefficients fulfil the implication

$$a_{i_1,\dots,i_k} \neq 0 \text{ and } a_{j_1,\dots,j_k} \neq 0 \;\Rightarrow\; i_1 + \dots + i_k = j_1 + \dots + j_k.$$

If $a \neq 0$, then this means that the degree of a coincides with the degree of each nonzero term. As a convention, the unknowns of a homogeneous polyomial are written in capital letters.

If $F \in K[X_1, \dots, X_{n+1}]$ is a homogeneous polynomial and $(\alpha_1 : \dots : \alpha_{n+1}) \in \mathbb{P}^n$, then $F(\lambda\alpha_1, \dots, \lambda\alpha_{n+1}) = \lambda^d F(\alpha_1, \dots, \alpha_{n+1})$, where $d = \deg(F)$. The equation $F(\alpha_1, \dots, \alpha_{n+1}) = 0$ thus may be interpreted by saying that the projective point $(\alpha_1 : \dots : \alpha_{n+1})$ is a *zero* of F. The set of all projectives zeros is denoted by V_F, similarly to the affine case.

Say that $C \subseteq \mathbb{P}^2$ is a *planar projective curve* if there exists a (homogeneous) $F \in K[X_1, X_2, X_3]$, $\deg(F) \geq 1$, such that $C = V_F$.

A projective curve may be connected to an affine curve by the process of *homogenization.* The homogenization of a polynomial $f = \sum a_{ij} x_1^i x_2^j \in K[x_1, x_2]$, $d = \deg(f) \geq 0$, is the polynomial $F = \sum a_{ij} X_1^i X_2^j X_3^{d-i-j}$. Now, $(\alpha_1 : \alpha_2 : 1) \in \mathbb{P}^2$ belongs to V_F if and only if $(\alpha_1, \alpha_2) \in V_f$. Hence V_F may differ from V_f only in points at infinity. These are the points $(\alpha_1 : \alpha_2 : 0)$ such that $\sum_{i+j=d} a_{ij} \alpha_1^i \alpha_2^j = 0$.

If F is a homogenization of f, then f is irreducible if and only if F is irreducible (this is not difficult to prove). A planar projective curve C is said to be *irreducible* if it may be expressed as V_F, where $F \in K[X_1, X_2, X_3]$ is an irreducible homogeneous polynomial. There is only one irreducible planar projective curve that may not be obtained by a homogenization of an affine (irreducible) curve, and that is the line $X_3 = 0$. This is because an irreducible homogeneous polynomial F that is divisible by X_3 has to be a scalar multiple of X_3.

Let C be a planar projective curve. Then there is no reasonable way how to define the coordinate ring of C. This is because we have to consider only those mappings $C \to K$ that give the same value for each expression of a point $\alpha \in C$ by homogeneous coordinates. Such a behaviour cannot be achieved by using polynomials only. However, if $A, B \in K[X_1, X_2, X_3]$ are homogeneous of the same degree, then $A(\alpha)/B(\alpha)$ is independent of the choice of homogeneous coordinates of $\alpha = (\alpha_1 : \alpha_2 : \alpha_3) \in \mathbb{P}^3$. This is utilized to define the *function field* $K(C)$, provided $C = V_F$, $F \in K[X_1, X_2, X_3]$ irreducible. Nonzero elements of $K(C)$ are $(A + (F))/(B + (F))$, where A and B are as above.

If F is a homogenization of $f \in K[x_1, x_2]$, then, as may be proved, $K(V_F) \cong K(V_f)$. This means that the algebraic structure of an irreducible planar affine curve is not influenced by homogenization.

C.4. **Smoothness.** Consider a polynomial $f \in K[x_1, \dots, x_n]$ and let $\alpha \in \mathbb{A}^n$ be such that $f(\alpha) = 0$, i.e., $\alpha \in V_f$. Say that f is *smooth* or (equivalently) *nonsingular* at α if $(\partial f/\partial x_i)(\alpha) \neq 0$ for at least one $i \in \{1, \dots, n\}$.

Let C be a planar affine curve, $K[C] = K[x_1, x_2]/(f)$. A point $\alpha \in C$ is said to be *smooth* (or *nonsingular*) if f is smooth at α. The remaining points of C are *singular.* If $\alpha \in C$ is a singular point, then it is also said that C has a *singularity* at α. An affine curve with no singularity is called *smooth (nonsingular).*

Similarly, if $F \in K[X_1, X_2, X_3]$ is homogeneous and $\alpha \in \mathbb{P}^2$ is such that $F(\alpha) = 0$, then F is said to be smooth (or nonsingular) at α if $(\partial F/\partial X_i)(\alpha) \neq 0$ for at least one $i \in \{1, 2, 3\}$. Notions of smoothness and singularity are being transferred to planar curves like in the affine case.

Suppose that the homogeneous polynomial F is not equal to 0. Then

$$X_1 \frac{\partial F}{\partial X_1} + X_2 \frac{\partial F}{\partial X_2} + X_3 \frac{\partial F}{\partial X_3} = dF, \text{ where } d = \deg(F).$$

This can be used to prove that if F is a homogenization of f and f is smooth at $\alpha = (\alpha_1, \alpha_2)$, then F is smooth at $(\alpha_1 : \alpha_2 : 1)$. This means that the smoothness of a point of an affine curve is not influenced by homogenization.

C.5. **Places.** Let $F \in K[X_1, X_2, X_3]$ be an irreducible homogeneous polynomial. Suppose that the projective curve $C = V_F$ is smooth. If $K = \bar{K}$, then each point of C determines in $K(C)$ exactly one normalized discrete valuation over K. The exact nature of this correspondence and its proof is beyond the extent of this overview. However, the structure of discrete valuations in $K(x)$ suggets how this correspondence may look like. Very briefly: in the affine case when $C = V_f$, $f \in K[x_1, x_2]$ irreducible, the valuation ν associated with $\alpha \in C$ treats those $\sigma \in K(C)$ that may be represented by a polynomial $g \in K[x_1, x_2]$ in such a way that $\nu(\sigma)$ indicates the degree of smoothness coincidence between g and f. Thus $\nu(\sigma) = 0$ if $g(\alpha) \neq 0$. If $g(\alpha) = 0$ and g and f have different tangents, then $\nu(\sigma) = 1$. If $g(\alpha) = 0$ and the tangents coincide, then $\nu(\sigma) \geq 2$.

The correspondence described above is partly valid also for curves over $\bar{K}$ that are not smooth everywhere. What remains true is that each smooth point uniquely determines a normalized discrete valuation over $\bar{K}$. However, a singularity may determine more discrete valuations.

In context of function fields it is usual to speak about places rather than discrete valuations. A *place* is every subset of $K(C)$ that may be expressed as $\{a \in K(C);\ \nu(a) \geq 1\}$, where ν is a normalized discrete valuation of $K(C)$ over K. If C is a projective curve that is smooth everywhere, then there is a natural bijection between points of C and places of $\bar{K}(C)$.

The situation is more complicated if $K \neq \bar{K}$. Consider again the case of $K(x)$. Valuations of $K(x)$ over K are equal to v_p or v_∞, where $p \in K[x]$ is irreducible. With each valuation (and thus with each place) there may be associated a positive integer that is called the *degree* of the valuation (and also of the place associated with the valuation). It turns out that $\deg(v_p) = \deg(p)$ and $\deg(v_\infty) = 1$. Note that $\deg(v_p) = 1$ if and only if $p = x - \lambda$ for some $\lambda \in K$. In $K(x)$ the places of degree one thus correspond to K-rational points of $\mathbb{P}^1$.

The degree may be defined for each place of a function field $K(C)$. Each smooth K-rational point of C determines a place of degree one. If C is an irreducible projective planar curve that is smooth at every K-rational point, then there is a natural bijection between K-rational points of C and places of degree one.

To get a feeling what are places of degree > 1 consider first $K(x)$ again. If $\deg(p) > 1$ and $p \in K[x]$ is irreducible, then the place of p is naturally associated with all roots of the polynomial p. There are thus more points of $\mathbb{P}^1$ that correspond to the place of p.

Something similar is true for places of a smooth curve over K. For simplicity let us formulate this just for affine points. Suppose that $C = V_f$ is a smooth affine curve, $f \in K[x_1, x_2]$ irreducible. Then $(\alpha_1, \alpha_2) \in C$ and $(\beta_1, \beta_2) \in C$ correspond to the same place if and only if there exists a field L, $K \leq L \leq \bar{K}$, and a K-automorphism ψ of L such that $\psi(\alpha_i) = \beta_i$ for both $i \in \{1, 2\}$. Recall that K-automorphisms are those automorphisms which fix each element of K.

With a little knowledge of field theory it is apparent that $L = \bar{K}$ may be always assumed. However, it is also clear that assuming $[L : K] < \infty$ is always possible too.

When working with curves it is usual to assume that the field K is *perfect* (either $\mathrm{char}(K) = 0$, or $\mathrm{char}(K) = p > 0$ and the mapping $\lambda \mapsto \lambda^p$ is an automorphism of K). The connection between places and K-automorphisms, as described above, assumes that K is perfect.

If $K = \mathbb{R}$ and $\bar{K} = \mathbb{C}$, then each place is either of degree 1 or degree 2. In the latter case (α_1, α_2) forms a pair with $(\bar{\alpha}_1, \bar{\alpha}_2)$, where $\overline{a + b\mathrm{i}} = a - b\mathrm{i}$.

W. What is an elliptic curve

W.1. **The genus.** By definition, an *elliptic curve* over K is a projective planar irreducible curve C over K that contains at least one K-rational point and is of genus 1. What is the *genus*? Unfortunately, that is not so easy to explain. A complete formal definition goes beyond the scope of this text. Neverthelees, the ensuing comments might give an idea what the genus means.

If C is an irreducible curve over K, then the genus of C may be derived from the structure of $K(C)$. It somehow reflects properties of principal divisors. (If σ is a nonzero element of $K(C)$, then there are only finitely many places P with $v_P(\sigma) \neq 0$. The formal sum $\sum v_P(\sigma)P$ is called the principal divisor of σ.) Genus is always a nonnegative number and is usually denoted by g.

Since complex numbers may be identified with the euclidean plane, a planar curve over $K = \mathbb{C}$ may be regarded as a 2-dimensional object. Let us first ponder what kind of a 2-dimensional object the projective line $\mathbb{P}^1(\mathbb{C})$ should be associated with. The affine line $\mathbb{A}^1(\mathbb{C})$ coincides with the euclidean plane. The existence of the point at infinity changes, however, the picture completely. The proper 2-dimensional object to identify $\mathbb{P}^1(\mathbb{C})$ with is the sphere. This may be envisioned by considering a stereographic projection of a sphere to the euclidean plane, with the point at infinity being represented by the north pole of the sphere.

The sphere is an example of a closed 2-dimensional surface in the 3-dimensional real space. In this context the exact shape of the surface is not important. What is important are topological properties of the surface. It turns out that two such surfaces may be identified by continuous deformations if and only if they possess the same number of holes. A sphere has no hole. A toroid has one hole. The surface of a pretzel has two holes (going from doughnut to pretzel adds one hole). The number of holes is thus a topological invariant and this invariant is called the *genus* of the surface. This is how the notion of the genus of a curve arose. If C is a smooth irreducible projective planar curve over $\mathbb{C}$, then C forms a 2-dimensional structure that may be embedded into the 3-dimensional real space as a closed surface. The curve is of genus one if the surface to which it may be embedded has the shape of torus.

The existence of such an embedding has to be proved. That is done in topology and goes far beyond the scope of this text. Note however that such an embedding cannot be "seen" since the graph of the curve is a subset of $\mathbb{C} \times \mathbb{C}$, and thus it lives in a 4-dimensional real space. However, the fact that the surface of a complex elliptic curve forms a torus has certain consequences for elliptic curves over real numbers. When cutting a torus there appears either one ellipse or two of them. Because of that it may be expected that an elliptic curve over reals will have one or two closed branches.

Many authors define an elliptic curve as a projective irreducible curve of genus 1 that is smooth everywhere. This is a traditional approach that may be justified by the prominent role of smooth Weierstraß curves. These curves present a universal model of elliptic curves in the sense that whenever C is a curve of genus one that contains at least one K-rational point, then there exists a smooth Weierstraß curve E such that the function fields $K(C)$ and $K(E)$ are K-isomorphic (that is there exists an isomorphism that fixes each $\lambda \in K$.)

W.2. **Weierstraß curves.** An *affine Weierstraß curve* is the set C of all points $(\alpha_1, \alpha_2) \in \mathbb{A}^2$ that fulfil a *Weierstrass equation* $x_2^2 + x_2 g(x_1) = f(x_1)$, that is an equation in which $f, g \in K[x_1]$ are polynomials such that $\deg(g) \leq 1$, $\deg(f) = 3$, f is monic. It may be proved that the polynomial $x_2^2 + x_2 g(x_1) - f(x_1)$ is always irreducible. Each Weierstraß curve is thus an irreducible planar curve.

By convention, the coefficients of g are denoted by a_1 and a_3, and the coefficients of f by a_2, a_4 and a_6. The Weierstraß equation thus often appears in the *standard form*

$$x_2^2 + a_1x_1x_2 + a_3x_2 = x_1^3 + a_2x_1^2 + a_4x_1 + a_6.$$

Set $b_2 = 4a_2 + a_1^2$, $b_4 = 2a_4 + a_1a_3$, $b_6 = 4a_6 + a_3^2$ and

$$b_8 = 4a_2a_6 + a_2a_3^2 + a_1^2a_6 - a_4^2 - a_1a_3a_4.$$

It may be established that the curve C is smooth if and only if the *discriminant*

$$\Delta(C) = -8b_4^3 + 9b_2b_4b_6 - 27b_6^2 - b_2^2b_8$$

is different from 0.

Applications of Weierstraß curves usually assume that $\mathrm{char}(K) \notin \{2,3\}$ and $a_1 = a_3 = 0$. Often it is also assumed that $a_2 = 0$. In those cases the smoothness of C correlates with the nonexistence of a multiple root of f. This will be now verified.

Suppose that $\mathrm{char}(K) \neq 2$ and that $C = V_w$, where $w(x_1,x_2) = x_2^2 - f(x_1)$. Then

$$\frac{\partial w}{\partial x_1} = -f'(x_1) \text{ and } \frac{\partial w}{\partial x_2} = 2x_2.$$

A point $(\alpha_1, 0) \in \mathbb{A}^2$ belongs to C if and only if $f(\alpha_1) = 0$. All of this means that (α_1,α_2) presents a singularity of C if and only if $\alpha_2 = 0$ and α_1 is a root of both f and f'. We have proved:

Theorem W.1. *Let C be the Weierstraß curve over K, $\mathrm{char}(K) \neq 2$, determined by $x_2^2 = f(x_1)$, $f \in K[x_1]$ cubic and monic. Then C is smooth if and only if f is separable (i.e., possesses no multiple root).*

If $a_1 = a_2 = a_3 = 0$, then the Weierstraß equation will often be written as $x_2^2 = x_1^3 + ax_1 + b$ or $y^2 = x^3 + ax + b$. The polynomial $x^3 + ax + b$ has multiple roots if and only $4a^3 + 27b^2 = 0$. The curve determined by $x_2^2 = x_1^3 + ax_1 + b$, $\mathrm{char}(K) \neq 2$, is thus smooth if and only if $4a^3 + 27b^2 \neq 0$.

This is the same condition as $4a_4^3 + 27a_6^2 \neq 0$. A mnemotechnical remark: Both terms of the sum may be expressed as $(i/2)^{i/2}a_i^{j/2}$, where $\{i,j\} = \{4,6\}$.

Projective Weierstraß curves are obtained by homogenization. They are thus determined by equation

$$X_2^2X_3 + X_2G(X_1,X_3) = F(X_1,X_3), \text{ where } G(X_1,X_3) = a_1X_1X_3 + a_3X_3^2$$
$$\text{and } F(X_1,X_3) = X_1^3 + a_2X_1^2X_3 + a_4X_1X_3^2 + a_6X_3^3.$$

A point at infinity $(\alpha_1 : \alpha_2 : 0)$ belongs to the curve if and only if $0 = \alpha_1^3$. There is thus only one such point, and this point is equal to $(0 : 1 : 0)$.

Put $W(X_1,X_2,X_3) = X_2^2X_3 + X_2G(X_1,X_3) - F(X_1,X_3)$. Then

$$\begin{aligned}
\frac{\partial W}{\partial X_1} &= X_2\frac{\partial G}{\partial X_1} - \frac{\partial F}{\partial X_1},\\
\frac{\partial W}{\partial X_2} &= 2X_2X_3 + G(X_1,X_3), \text{ and}\\
\frac{\partial W}{\partial X_3} &= X_2^2 + X_2(a_1X_1 + 2a_3X_3) - \frac{\partial F}{\partial X_3}.
\end{aligned}$$

Hence $(\partial W/\partial X_1)(0,1,0) = 0 = (\partial W/\partial X_2)(0,1,0)$ and $(\partial W/\partial X_3)(0,1,0) = 1$. Each projective Weierstraß curve is therefore smooth at the point at infinity. An affine Weierstraß curve is thus smooth if and only if the corresponding projective Weierstraß curve is smooth.

As examples of affine Weierstraß curves consider curves over real numbers given by equations $x_2^2 = x_1^3 - c^3$ and $x_2^2 = x_1^3 - c^2x_1$. The former curve has a single branch. In the central part it has a form of belly that is protruded to the point $(c, 0)$, with the body being to the right. If $c = 0$, then $(0, 0)$ is a singularity that is called *cusp.* Assume $c \neq 0$. Then in each case there are two inflexion points. If $c < 0$, then the curve passes through stationary inflexion points $(0, \pm c^{3/2})$. If $c > 0$, then the inflexion points are at $(2^{2/3}c, \pm 3^{1/2}c^{3/2}) \approx (1.6c, 1.7c^{3/2})$ and the slope of the inflexion line is equal to $2^{1/3}(3c)^{1/2} \approx 2.2c^{3/2}$.

If $x_2^2 = x_1^3 - c^2 x_1$, then it may be assumed that $c > 0$ since $x^3 - c^2 x = x(x - c)(x + c)$. In this case the curve has two affine branches. One has a form of an oval with the flat pole at $(-c, 0)$, with the other pole at $(0, 0)$ and with extreme points at $(-3^{-1/2}c, \pm 2^{1/2} \cdot 3^{-3/4}c^{3/2}) \approx (-0.58c, 0.62c^{3/2})$. The central part has again a belly-like form with the body right of $(c, 0)$. The inflexion points are at $\approx (1.5c, 1.3c^{3/2})$ and the slope of inflexion line is $\approx 2.1c^{1/2}$.

The shape of the unbounded branch in the above two examples does not seem to resemble a cut of a torus. However, the resemblence is topological, up to deformation. From the projective point of view the branch passes through the point at infinity, and that makes it closed.

To finish the classification of shapes of real Weierstrass curves consider the curve given by $x_2^2 = x_1(x_1 - 1)^2 = x_1^3 - 2x_1^2 + x_1$. The curve passes through points $(4, 6)$, $(1, 0)$, $(1/3, -4/\sqrt{27}) \approx (0.33, -0.77)$, $(0, 0)$, $(1/3, 4/\sqrt{27})$, $(1, 0)$, $(4, -6)$, forming thus a crossing point at $(1, 0)$. This type of singularity is called a *node.*

W.3. **The group of an elliptic curve.** The K-rational points of a projective smooth Weierstraß curve may be equipped with a group structure. This is well known and will be considered in detail later. The aim here is to give a certain idea what is the abstract background of such groups. It turns out that they may be defined only in terms of the function field $K(C)$, where C is an elliptic curve (thus each elliptic curve induces a group structure, not only smooth Weierstraß curves).

In this context the following metaphor may be of help. The genus of a surface measures, in some sense, what is missing. If $A \leq B$ are abelian groups, then what is missing to A may be expressed by factorization B/A.

Situations when B and A are infinite, but the factor may be finite and nontrivial, tend to be mathematically interesting. In our case B is a subgroup of free abelian group with the basis being equal to the set of all places of $K(C)$. Elements of that group are formal sums $\sum a_P P$, where P runs through all places and $a_P \in \mathbb{Z}$ is nonzero for only finitely many P. Elements with $\sum a_P \deg(P) = 0$ form the subgroup B, while the group A coincides with the set of all principal divisors. If Q is a fixed place of degree one, then it may be proved that each element of B/A (i.e., each coset modulo A) contains a unique element of B that is equal to $P - Q$, where P is a place of degree one.

If the curve C is smooth at each K-rational point, then each such P may be associated with a single K-rational point. Denote P by P_α if P is associated with a K-rational point α. Thus $Q = P_\omega$ for a K-rational point ω.

Facts above imply that adding the coset of $P_\alpha - P_\omega$ with the coset of $P_\beta - P_\omega$ yields a coset with some $P_\gamma - P_\omega$. Setting $\gamma = \alpha \oplus \beta$ equips the K-rational points of C with the structure of an abelian group, and ω is the neutral element of this group.

Formulae for computing $\oplus$ depend upon the definition of C. It occurs quite often that a choice has to be made between several formulae. The choice depends upon values of α and β, and upon their relationshiop. A situations when there exists a universal formula (called also a *closed* formula) which works for all α and β is of certain computational advantage.

W.4. **Applications of elliptic curves.** Some applications are standard and some are emerging. The *Elliptic curve cryptography* (ECC) usually refers to the bunch of applications that replace counting modulo a prime by computations in a subgroup of the group of an elliptic curve. If C is an elliptic curve over K, then $C(K)$ refers to the group operation $\oplus$ that is defined upon the K-rational points of C. The neutral element ω of the group is usually understood from the context. In applications K is a finite field. Thus $K = \mathbb{F}_q$, where q is a power of a prime. In present applications q is nearly always a large prime. Structure of $\mathbb{F}_q$ implies that for large q the group $C(\mathbb{F}_q)$ always contains a large cyclic subgroup. The order of this subgroup appears to be a random feature (while it has to occur in a certain interval). There are thus many situations when $C(\mathbb{F}_q)$ contains a large cyclic subgroup G that is of prime order. A generator of this subgroup, often denoted by P, usually constitutes, together with parameters of the curve C, the public key (or a part of it).

Note that making public the pair (P, C) does not imply knowledge of $|G|$ or $|C(K)|$. Classical protocols (Diffie-Hellman, Elgamal etc.) derive their security from the difficulty of the Discrete Logarithm Problem (DLP). Some of the attacks on the DLP require knowledge of the order of the group. The order of $C(K)$ is given by the number of K-rational points. There does not seem to exist any straightforward way how to determine this number from the parameters of the curve. In the context of ECC the point counting algorithms are thus of paramount importance.

The advantage of ECC over modular arithmetic rests in the fact that the DLP is more difficult, which allows for shorter keys. However, quantum computing makes all protocols based upon the DLP vulnerable. One of the promising alternatives for elliptic postquantum cryptography is based on isogenies of supersingular elliptic curves. That is presently beyond the scope of this text.

Classical applications of ECC need keys of considerable size (while much shorter than those needed for RSA). The speed of computation is hence a factor to be considered. The question is not only how to compute $\alpha \oplus \beta$, but also how to organize a computation of $[n]\alpha = \alpha \oplus \cdots \oplus \alpha$. In general, techniques used do not differ from those for other cyclic groups. In some cases (like Elliptic Curves Digital Signature Algorithm, the ECDSA) only the x-coordinate α of the point (α, β) is used. There are some speed-ups that take advantage of this fact.

Elliptic curves are also used for pseudorandom generators and in factorizing integers. Integers that are accessible by Lenstra elliptic-curve factorization are smaller than those accessible by the Number Field Sieve (NFS). However, the NFS uses many auxiliary factorizations of small integers, and for that the elliptic-curve factorization appears to be the most efficient.

B. Basic arithmetic

Multiplication of integers by computers used to be slower than addition by factor of 10 and more for many decades after the first computers have been constructed. Nowadays the speed of addition and multiplication does not much differ when performed in the length of computer word. This is not because the complexity of multiplication has diminished, but because this complexity has been transformed to hardware. In terms of tacts of the processor the multiplication needs two or three tacts while the addition only one tact. However, due to pipelining the actually observed behaviour may give an impression that the speed of multiplication is nearly the same as the speed of addition. The speed-up of division does not seem to have kept pace with the speed-up of multiplication. In microprocessors the divison takes 10 times more time than the multiplication, while the ratio in historical mainframes used to be around 3.

Computations with very long integers rely upon software packages of *multiple-precision arithmetic*. Since this arithmetic is realized by software and not by hardware, there is no decline in importance of replacing multiplication by addition whenever possible, and replacing division by multiplication whenever possible.

The division occurs naturally when working modulo p, p a large prime. The naive algorithm of computing x times y modulo p goes by performing first the multiplication of integers, and then taking the remainder of division by p.

The *Montgomery arithmetic* described below replaces the division by p by multiplications. The key concept is to replace each $x \bmod p$ by $xR \bmod p$, where R is an integer of special properties. Leaving implementation details aside, consider the situation when each $x \in \mathbb{Z}_p$ is represented in the memory of the computer by $X \equiv xR \bmod p$. To represent $z \equiv x + y \bmod p$ an algorithm is needed that derives $Z \equiv zR \bmod p$ from $X \equiv xR \bmod p$ and $Y \equiv yR \bmod p$. That is trivial since $Z \equiv X + Y \bmod p$ as $xR + yR = (x + y)R$.

What about $xy \bmod p$? Multiplying xR and yR modulo p yields ZR, where $Z \equiv (xy)R \bmod p$. Finding an efficient method that derives Z from ZR hence results into finding a way how to multiply efficiently modulo p, circumventing thus the division by p.

The goal hence is to devise an algorithm that transforms an integer, say x, that corresponds to ZR to an integer, say y, that corresponds to Z. The input restriction is $0 \le x < pR$. The output requires $0 \le y < p$ and $x \equiv yR \bmod p$, i.e. $y = xR^{-1}$ $(\text{mod } p)$. Such a transformation is known as *Montgomery reduction*.

Of course, an efficient Montgomery reduction is conceivable only under some external assumption. The assumption here comes from the reality of computers. The division by R requires much less resources if R is a power of two or, even better, if $R = b^t$, where b is the extent of the computer word ($b = 2^{32}$ or 2^{64} etc.).

It will be thus assumed that $R = b^t > p$ and that the division by b (and thus also by R) is 'cheap'. No other external assumption is being made. The integer b is considered as a *basis* and integers $< p$ are represented as $(a_{t-1}, \dots, a_1, a_0)_b = \sum a_i b^i$. Examples that rely on the pen and mental arithmetic may have $b = 10$ or $b = 100$ etc.

The idea of Montgomery reduction is as follows: The residue class modulo p does not change if x is replaced by $x + xpq$. Choose q so that $pq \equiv -1 \bmod R$. That makes $x + xpq$ divisible by R. Change now $x + xpq$ in such a way that xq is replaced by $u = xq$ $(\text{mod } R)$. This affects neither the residue class nor the divisibility by R. Hence $y = (x + up)/R$ is an integer, and $yR \equiv x \bmod p$. While there does not have to be $y < p$, there has to be $y < 2p$ if $x < pR$ is assumed. This is because $u < R$ and because $y < 2p$ may be expressed as $2pR > yR = x + up$.

The preceding observation will be recorded as a statement:

Lemma B.1. *Let $R > 1$ be an integer, and let $p, q \in \mathbb{Z}_R$ be such that $pq \equiv -1 \bmod R$, i.e. $q = -p^{-1} \pmod{R}$. Let x be an integer such that $0 \le x < pR$. Put $u = xq \pmod{R}$. Then $R \mid up{+}x$, and $y = (up{+}x)/R$ fulfils both $y < 2p$ and $yR \equiv x \bmod p$.*

Proof. Indeed, $up{+}x \equiv xpq + x \equiv 0 \bmod R$, and $yR = up{+}x \equiv x \bmod p$. Furthermore, $yR = pu + x < pR + pR = 2pR$. □

B.1. **Montgomery arithmetic.** Consider an algorithm that performs some task in the arithmetic modulo p, with inputs $a_1, \dots, a_m$ and $b_1, \dots, b_n$. To implement the algorithm by means of Montgomery arithmetic requires to determine $R = b^t > p$ and $q = -p^{-1} \pmod{R}$ in advance, and then, whenever the procedure is invoked, to convert the inputs a_i to $A_i \equiv a_i R \bmod p$, to perform all arithmetical operations of the procedure in this representation, and finally to convert each $B_j \equiv b_j R \bmod p$ to b_j at the time of output.

The Montgomery reduction $x \to y$, where $0 \le x < pR$, $0 \le y < p$ and $x \equiv yR \bmod p$, may be executed as suggested by Lemma B.1. That means to multiply x and q, and reduce it modulo R. The computations are exercised in the basis b. The reduction modulo R thus means to take the last t positions (i.e., the last t computer words) of the product. This is denoted by u. The output is equal to $x = (x + up)/R$ if $x < p$. If $x \ge p$, then the output is equal to $x - p$.

The disadvantage of this approach is that it requires two long multiplications (of x with q, and of u with p). A more efficient solution reduces this to a linear number of multiplications of a long integer with an integer in the size of the computer word (i.e., $< b$). It turns out that knowledge of q is not necessary. It suffices to know $q' = -p^{-1} \pmod{b}$.

Suppose that $x = \sum x_i b^i$, $0 \le x_i < b$. Let x be divisible by b^r, $r \ge 0$. Thus $b_0 = \cdots = b_{r-1} = 0$. Set $u = x_r q' \pmod{b}$ and $x' = x + upb^r$. Counting modulo b^{r+1} shows that $x' \equiv x_r b^r + x_r pq' b^r \equiv x_r b^r + x_r(-1)b^r \equiv 0$. Hence b^{r+1} divides x', and $x' - x < pb^{r+1}$ is a multiple of pb^r. Proceeding inductively from $r = 0$ increases x in k steps by an integer vp, where $0 \le v < b^k$. The Montgomery reduction can be thus performed as follows:

```
INPUT: $x = \sum_i x_i b^i \le pR$, $0 \le i \le 2t-1$, $0 \le x_i < b$.
OUTPUT: An integer $y$ with $0 \le y < p$ and $yR \equiv x \bmod R$.
PARAMETERS: $p$, $b$, $t$, $R$, where $R = b^t > p$,
            $q'$, where $q'p \equiv -1 \bmod b$ and $0 < q' < p$.
VARIABLES: $i$, $u$, $0 < u < p$.
i=0;
while $(i < t)$ do:
      $u = x_i q' \pmod{b}$;
      $x = x + pub^i$;
      $i = i + 1$;
$y = x/R$;
if $(y > p)$ then $y = y - p$;
return $y$.
```

Each multiplication in Montgomery arithmetics ends by the reduction. The efficiency may be raised by integrating both of these steps in an ensuing algorithm. The justification follows the description. Parameters and variable are the same as in the preceding algorithm.

```
INPUT: $x = \sum_i x_i b^i < p$, $y = \sum_i y_i b^i < p$.
OUTPUT: An integer $z = \sum_i z_i b^i < p$ such that $zR \equiv xy \bmod R$.
```

```
i=0;
z=0;
while $(i < t)$ do:
        $u = (z_0 + x_i y_0)q' \pmod b$;
        $z = (z + x_i y + pu)/b$;
        $i = i + 1$;
if $(z > p)$ then $z = z - p$;
return $z$.
```

To justify the division by b note that while counting modulo b

$$z + x_i y + pu \equiv z_0 + x_i y_0 + pu \equiv z_0 + x_i y_0 + pq'(z_0 + x_i y_0) \equiv 0$$

since $pq' \equiv -1 \bmod b$. To see that the procedure does what declared denote by $\bar{z}_i$ the value of z after the ith round. Thus $\bar{z}_t$ is equal to the output from the cycle. Put $\bar{x}_i = \sum_{j<i} x_j b^j$ and note that $\bar{x}_t = x$. The claim to verify is that there exists integer v_i such that

$$0 \leq \bar{z}_i b^i - \bar{x}_i y = pv_i \text{ and } v_i < b^i.$$

In the first step $u = \bar{x}_1 y_0 q' \pmod b$ since $\bar{x}_1 = x_0$ and $\bar{z}_1 b - \bar{x}_1 y = pu$. The condition thus holds for $i = 1$. For the induction step first observe that $\bar{z}_{i+1} b^{i+1} = \bar{z}_i b^i + x_i y b^i + pub^i$ and $\bar{x}_{i+1} y = x_i y b^i + \bar{x} + iy$. By the induction assumption

$$\bar{z}_{i+1} b^{i+1} - \bar{x}_{i+1} y = \bar{z}_i b^i - \bar{x}_i y + pub^i = p(v_i + ub^i).$$

Therefore $v_{i+1} = v_i + ub^i < b^{i+1}$. By the final step, $0 \leq zR - xy = pv_t < pR$. Thus $zR \equiv xy \bmod p$ and $zR < pR + p^2 < 2pR$.

Recall that in Montgomery arithmetic the procedure above is invoked with inputs $X \equiv xR \bmod p$ and $Y \equiv yR \bmod p$, and the output is equal to $Z \equiv xyR \bmod p$. In each step there are two multiplications of the form (long integer) × (computer word), and there is no multiplication of two long integers.

Note that whenever Montgomery arithmetic is applied, there is an initial cost of multiplying the inputs by R modulo p.

The final remark concerning the Montgomery arithmetic is about the computation of q'. The question thus is how to compute $p^{-1} \pmod b$ efficiently. In general the inverses may be computed by means of extended Euclidean algorithm. However, if $b = 2^w$, then there exists a more efficient procedure:

```
INPUT: An odd integer $x$, $0 < x < 2^w$.
OUTPUT: Integer $y$ such that $yx \equiv 1 \bmod 2^w$, $0 < y < 2^w$.
PARAMETER: Integer $w \geq 1$.
VARIABLES: Integers $i, j, u$.

$y = 1$;
$i = 1$;
while $(i < w)$ do:
      $j = i + 1$;
      $u = xy \pmod{2^j}$;
      if $(2^i < u)$ then $y = y + 2^i$;
      $i = j$;
return $y$.
```

To prove the correctness denote by y_i the value of y at the end of the ith round and set $y_0 = 1$. Thus $y = y_{w-1}$. The algorithm clearly implies that $y_i < 2^{i+1}$. It thus suffices to verify that $xy_i \equiv 1 \bmod 2^{i+1}$. For $i = 0$ this is true because x is odd.

Suppose that $i \geq 1$. By the induction assumption $xy_{i-1} \equiv 1 \bmod 2^i$. Hence $xy_{i-1} \pmod{2^{i+1}}$ is equal to 1 or to $1 + 2^i$. In the former case $y_i = y_{i-1}$. In the latter case $y_i = y_{i-1} + 2^i$ and $xy_i \equiv xy_{i-1} + 2^i \equiv 1 \bmod 2^{i+1}$.

A. Speeding up addition and doubling

Let K be a field and let C be a smooth Weierstraß curve over K given by

$$x_2^2 + a_1x_1x_2 + a_3x_2 = x_1^3 + a_2x_1^2 + a_4x_1 + a_6. \tag{A.1}$$

Then all K-rational points of C together with ∞, the point at infinity, can be interpreted as an abelian group. This group will be denoted by $C(K)$, the addition in this group by $\oplus$, the opposite elements by $\ominus$, and $[m]$ will be used when the addition is repeated m-times. The neutral element of $C(K)$ is the point at infinity ∞. Thus $\alpha \oplus \infty = \infty \oplus \alpha$ for all $\alpha \in C(K)$.

The group $C(K)$ may be also interpreted as a group on all projective K-rational points of C. Under this approach every affine K-rational point (α_1, α_2) is identified with $(\alpha_1 : \alpha_2 : 1)$, and ∞ with $(0 : 1 : 0)$.

Suppose that $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ are K-rational affine points of C. Then:

$$\ominus\, \alpha = (\alpha_1, -\alpha_2 - \alpha_1 a_1 - a_3). \tag{A.2}$$

If $\beta = \ominus\, \alpha$, then $\beta \oplus \alpha = \infty$. Suppose that $\beta \neq \ominus\, \alpha$. To define $\gamma = \alpha \oplus \beta$, $\gamma = (\gamma_1, \gamma_2)$, first set

$$\lambda = \frac{3\alpha_1^2 + 2a_2\alpha_1 - a_1\alpha_2 + a_4}{2\alpha_2 + a_1\alpha_1 + a_3} \text{ if } \alpha = \beta, \text{ and } \lambda = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} \text{ if } \alpha \neq \beta. \tag{A.3}$$

The value of γ_1 depends upon λ, α_1, β_1, a_1 and a_2, and γ_2 depends upon λ, γ_1, a_1 and a_3:

$$(\gamma_1, \gamma_2) = (-\alpha_1 - \beta_1 + \lambda^2 + a_1\lambda - a_2,\ \lambda(\alpha_1 - \gamma_1) - \alpha_2 - a_1\gamma_1 - a_3). \tag{A.4}$$

The formulas above describe what is known as the *chord and tangent process*. Let us recall its properties:

(CT1) For each $\alpha = (\alpha_1, \alpha_2) \in C(K)$ there is at most one $\beta = (\beta_1, \beta_2) \in C(K)$ such that $\alpha_1 = \beta_1$ and $\beta \neq \alpha$. If such a β exists, then $\beta = \ominus\, \alpha$. If no such β exists, then $[2]\alpha = \alpha \oplus \alpha = \infty$ and, thus, $\ominus\, \alpha = \alpha$. The latter happens if and only if $x_1 = \alpha$ yields the tangent line of C at α.

(CT2) Suppose that $\beta \neq \ominus\, \alpha$. The choice of λ in (A.3) is such that there exists a (unique) $\mu \in K$ for which $x_2 = \lambda x_1 + \mu$ describes a line that is (1) the tangent of C at α, provided $\alpha = \beta$, and (2) connects α and β, provided $\alpha \neq \beta$.

(CT3) Assume $\beta \neq \ominus\, \alpha$ and $\gamma = \alpha \oplus \beta = (\gamma_1, \gamma_2)$. We have $\alpha_2 = \lambda\alpha_1 + \mu$, $\mu = \alpha_2 - \lambda\alpha_1$ and $\gamma = (\gamma_1, -(\lambda\gamma_1 + \mu) - a_1\gamma_1 - a_3)$. Therefore $\ominus\, \gamma = (\gamma_1, \lambda\gamma_1 + \mu)$, by (A.2). All of the points α, β and $\ominus\, \gamma$ are incident to the line given by $x_2 = \lambda x_1 + \mu$. Denote this line by L. It is a fact that $L \cap C = \{\alpha, \beta, \ominus\, \gamma\}$.

(CT4) These possibilities can occur:
- The points α, β and $\ominus\, \gamma$ are pairwise distinct.
- $\alpha \neq \beta$ and $\beta = \ominus\, \gamma$. Then $\alpha \oplus [2]\beta = \infty$ and $\gamma = \ominus\, \beta$.
- $\alpha \neq \beta$ and $\alpha = \ominus\, \gamma$. Then $[2]\alpha \oplus \beta = \infty$ and $\gamma = \ominus\, \alpha$.
- $\alpha = \beta$ and $\alpha \neq \ominus\, \gamma$. Then $\gamma = [2]\alpha$.
- $\alpha = \beta = \ominus\, \gamma$. Then $[3]\alpha = \infty$ and $\gamma = \ominus\, \alpha = [2]\alpha$.

The natural question is how to perform efficiently both the addition $\alpha \oplus \beta$, and the doubling $[2]\alpha$. Note that the elliptic curve cryptography requires a computation of $[n]\alpha$ for very large n. The point α is usually denoted by P. It remains stable, while n varies. Standard algorithms, e.g. the sliding window, require many applications of doubling. The doubling hence deserves the same attention as the addition of distinct arguments.

For the rest of this section we shall assume that $\mathrm{char}(K) \neq 2$ and that C is given by $x_2^2 = x_1^3 + ax_1 + b$. Thus $a = a_4$, $b = a_6$, $a_1 = a_2 = a_3 = 0$ and $4a^3 + 27b^2 \neq 0$.

Then

$$\ominus(\alpha_1, \alpha_2) = (\alpha_1, -\alpha_2). \tag{A.5}$$

This means that opposite elements are symmetric along the axis x_1 (the line with $x_2 = 0$), and that (α_1, α_2) is of order two if and only if $\alpha_2 = 0$. An element of order two is sometimes called an *involution*.

If $\alpha \oplus \beta \neq \infty$, then there exists $\gamma = (\gamma_1, \gamma_2)$ such that $\gamma = \alpha \oplus \beta$ and

$$\gamma_1 = \lambda^2 - \alpha_1 - \beta_1, \quad \gamma_2 = \lambda(\alpha_1 - \gamma_1) - \alpha_2, \text{ where} \tag{A.6}$$

$$\lambda = \frac{\alpha_2 - \beta_2}{\alpha_1 - \beta_1} \text{ if } \alpha_1 \neq \beta_1, \quad \text{and} \quad \lambda = \frac{3\alpha_1^2 + a}{2\alpha_2} \text{ if } \alpha_1 = \beta_1. \tag{A.7}$$

Note that the parameter $b = a_6$ has no bearing upon any of the formulas above.

Let us now consider the time needed to perform $\alpha \oplus \beta$, $\alpha \neq \beta$, and to perform $[2]\alpha$. The time will be quantified in the number of needed arithmetical operations over the field K. Typically, K is equal to $\mathbb{F}_p$ for p a large prime. This implies that these operations are not built-in, but have to be algorithmically computed. If $\xi, \eta \in K$, then there exist algorithms which compute ξ^2 somewhat more quickly then $\xi\eta$. We shall use S for squaring ξ^2, M for multiplying $\xi\eta$, and I for inversion ξ^{-1}. An addition $\xi + \eta$ and/or a subtraction $\xi - \eta$ will be neglected since it is much more quicker than multiplication.

The cost of $\alpha \oplus \beta$ is I + 2M + S. Indeed, an inversion is needed to compute $(\alpha_1 - \beta_1)^{-1}$. If this is done, then a multiplication is needed to get λ. A squaring appears when computing γ_1, and one more multiplication appears in the formula that expresses γ_2. Small multiples can be replaced by additions. That makes the cost of doubling I + 2M + 2S.

To find an inversion modulo a prime means to employ the extended Euclidean algorithm. This includes many multiplications. Hence replacing I by kM, where k is fixed (and not too big) causes a significant speed-up. Such a speed-up is possible, but at a price. The price is that a point $\alpha = (\alpha_1, \alpha_2)$ may be addressed in several ways (using a triple or a quadruple instead of the pair (α_1, α_2)). That may pay off only if there are many intermediary stages at which the lack of uniqueness of point identification does not cause a difficulty. At the end an inversion usually cannot be avoided if the goal is to get a uniquely determined result. However, when computing $[n]P$, say in a cryptographic application, then the computation uses many additions and doublings that are of intermediary character. For such situations projective or Jacob or Chudonovski coordinates may be used.

A.1. **Projective coordinates.** The projective description of C is by the equation

$$X_2^2 X_3 = X_1^3 + aX_1X_3^2 + bX_3^3. \tag{A.8}$$

Let $\alpha = (\alpha_1 : \alpha_2 : \alpha_3) = (\alpha_1/\alpha_3 : \alpha_2/\alpha_3 : 1)$ and $\beta = (\beta_1 : \beta_2 : \beta_3) = (\beta_1/\beta_3, \beta_2/\beta_3 : 1)$ be two distinct points on C. Assume that $\alpha \oplus \beta \neq (0 : 1 : 0)$. Then $\alpha \oplus \beta = \gamma = (\gamma_1 : \gamma_2 : \gamma_3) = (\gamma_1/\gamma_3 : \gamma_2/\gamma_3 : 1)$. By (A.6)

$$\frac{\gamma_1}{\gamma_3} = \lambda^2 - \frac{\alpha_1}{\alpha_3} - \frac{\beta_1}{\beta_3} \text{ and } \frac{\gamma_2}{\gamma_3} = \lambda\left(\frac{\alpha_1}{\alpha_3} - \frac{\gamma_1}{\gamma_3}\right) - \frac{\alpha_2}{\alpha_3}, \tag{A.9}$$

where, by (A.7),

$$\lambda = \frac{\alpha_2/\alpha_3 - \beta_2/\beta_3}{\alpha_1/\alpha_3 - \beta_1/\beta_3} = \frac{\alpha_2\beta_3 - \beta_2\alpha_3}{\alpha_1\beta_3 - \beta_1\alpha_3}.$$

Put $U = \alpha_2\beta_3 - \beta_2\alpha_3$ and $V = \alpha_1\beta_3 - \beta_1\alpha_3$. The cost of computing U and V is 4M. The cost of computing

$$W = U^2\alpha_3\beta_3 - V^2(\alpha_1\beta_3 + \beta_1\alpha_3)$$

is 2S + 7M since $\alpha_1\beta_3$ and $\beta_1\alpha_3$ may be regarded as precomputed. Since $\alpha_1\beta_3 + \beta_1\alpha_3 = (\beta_1\alpha_3 - \alpha_1\beta_3) + 2\alpha_1\beta_3 = -V + 2\alpha_1\beta_3$ we also have

$$W = U^2\alpha_3\beta_3 + V^3 - 2\alpha_1\beta_3V^2. \tag{A.10}$$

If this formula is followed, the cost of W is 2S + 8M.

Put $\gamma_3 = V^3\alpha_3\beta_3$. Note that $\lambda = U/V$. Then

$$\begin{aligned}\gamma_1 &= V(U^2\alpha_3\beta_3) - V^3(\alpha_1\beta_3 + \beta_1\alpha_3) = VW, \text{ and}\\ \gamma_2 &= (U/V)(V^3\alpha_1\beta_3 - VW) - V^3\alpha_2\beta_3 = U(\alpha_1\beta_3V^2 - W) - \alpha_2\beta_3V^3.\end{aligned}$$

Compute W by means of (A.10) and use precomputed values to get γ_3, γ_1 and γ_2. The cost is 1M, 1M and 2M, respectively. The overall cost of computing $\gamma = (\gamma_1, \gamma_2)$ thus amounts to 2S + 12M.

Formula (A.9) can be used for the doubling as well, with $\alpha = \beta$. However, in this case

$$\lambda = \frac{3(\alpha_1/\alpha_3)^2 + a}{2\alpha_2/\alpha_3} = \frac{3\alpha_1^2 + a\alpha_3^2}{2\alpha_2\alpha_3}.$$

The form of γ_2/γ_3 suggests to choose γ_3 as $8\alpha_2^3\alpha_3^3$. Then

$$\begin{aligned}\gamma_1 &= 2\alpha_2\alpha_3((3\alpha_1^2 + a\alpha_3^2)^2 - 8\alpha_1\alpha_2^2\alpha_3), \text{ and}\\ \gamma_2 &= (3\alpha_1^2 + a\alpha_3^2)(4\alpha_1\alpha_2^2\alpha_3 - \gamma_1/2\alpha_2\alpha_3) - 8\alpha_2^4\alpha_3^2\\ &= (3\alpha_1^2 + a\alpha_3^2)(4\alpha_1\alpha_2^2\alpha_3 - ((3\alpha_1^2 + a\alpha_3^2)^2 - 8\alpha_1\alpha_2^2\alpha_3)) - 8\alpha_2^4\alpha_3^2.\end{aligned}$$

To compute γ_i, $1 \leq i \leq 3$, it may be proceeded by computing (1) α_1^2, (2) α_3^2, (3) $U = 3\alpha_1^2 + a\alpha_3^2$, (4) U^2, (5) $V = 2\alpha_2\alpha_3$, (6) α_2V, (7) V^2, (8) $\gamma_3 = V^3$, (9) $W = U^2 - 4\alpha_1\alpha_2V$, (10) $\gamma_1 = VW$, (11) $(\alpha_2V)^2$ and (12) $\gamma_2 = U(2\alpha_1\alpha_2V - W) - 2(\alpha_2V)^2$. The cost of doubling hence is 5S + 7M. If a is small, then the cost of multiplying by a may be regarded as negligible. In such a case the cost of doubling is equal to 5S + 6M.

When computing $[n]P$ it often happens that the point P is being added to an intermediary result. If the intermediary result is denoted by α, and the point P as β, then $\beta_3 = 1$ since P is given as an affine point. By inspecting the above procedure for computing $\alpha \oplus \beta$ it may be observed that it includes exactly three instances of multiplying by β_3. The cost of computing $\alpha \oplus \beta$ is thus reduced to 2S + 9M if $\beta_3 = 1$.

M. Montgomery curves

Consider the Weierstraß equation in its general form (A.1). For simplicity let us write y in place of x_2 and x in place of x_1. Suppose that $\mathrm{char}(K) \neq 2$. Then $y^2+a_1xy+a_3y = (y+(a_1x+a_3)/2)^2-(a_1x+a_3)^2/4$. Two Weierstraß equations are called *K-equivalent* if one can be obtained from the other by a linear substitution over K. The equation $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ is thus K-equivalent to the equation $y^2 = x^3 + (a_2 + a_1^2/4)x^2 + (a_4 + a_1a_3/2)x + (a_6 + a_3^2/4)$, provided $\mathrm{char}(K) \neq 2$. Indeed, the latter equation can be turned into the equation (A.1) by $y \mapsto y + (a_1x + a_3)/2$ and $x \mapsto x$. (By a *linear substitution* over K we understand here any reversible substitution $x_i \mapsto \lambda_{1i}x_1 + \lambda_{2i}x_2 + \mu_i$, $i \in \{1, 2\}$, where $\lambda_{ij}, \mu_i \in K$. Such a substitution is reversible if and only if $\det(\lambda_{ij}) \neq 0$.)

If $\mathrm{char}(K) > 3$, then $x^3 + a_2x^2 + a_4x + a_6 = (x + a_2/3)^3 + (a_4 - a_2^2/3)(x + a_2/3) + (a_6 - a_4a_2/3 + 2a_2^3/27)$. Hence each Weierstraß equation is K-equivalent to a Weierstraß equation of the form $y^2 = x^3 + ax + b$, provided $\mathrm{char}(K) > 3$.

A linear substitution may turn an equation $u(x, y) = v(x, y)$ into an equation $\lambda\tilde{u}(x, y) = \lambda\tilde{v}(x, y)$, where $\lambda \in K^*$. The curve determined by the latter equation is the same as the curve determined by $\tilde{u}(x, y) = \tilde{v}(x, y)$. Hence we say that $\tilde{u}(x, y) = \tilde{v}(x, y)$ is K-equivalent to $u(x, y) = v(x, y)$ also in this case.

If $\lambda \in K^*$, then the curve defined by the equation $y^2 = x^3+ax+b$ coincides with the curve given by $(\lambda^3y)^2 = (\lambda^2x)^3+a\lambda^4(\lambda^2x)+b\lambda^6$. The equation $y^2 = x^3+ax+b$ is hence K-equivalent to the equation $y^2 = x^3 + \lambda^4ax + \lambda^6b$. This is the only way how Weierstraß equations $y^2 = x^3+ax+b$ and $y^2 = x^3+\tilde{a}x+\tilde{b}$ may be K-equivalent. They are K-equivalent if and only if

$$\text{there exists } \lambda \in K^* \text{ such that } \tilde{a} = \lambda^4a \text{ and } \tilde{b} = \lambda^6b. \tag{M.1}$$

Curves given by equations $By^2 = x^3+Ax^2+x$, $\mathrm{char}(K) \neq 2$, are also important. A curve of this form is called a *Montgomery curve*. We will also speak about a *Montgomery equation*. Elements A and B belong to K, and $B \neq 0$. Capital letters are used to avoid a confusion with a and b in the normal form of a Weierstraß equation.

When both sides of $By^2 = x^3 + Ax^2 + x$ are multiplied by B^3 we get

$$(B^2y)^2 = (Bx)^3 + AB(Bx)^2 + B^2(Bx).$$

A Montgomery equation is thus K-equivalent to a Weierstraß equation $y^2 = x^3 + ABx^2 + B^2x$. Weierstraß equations of the form $y^2 = f(x)$, $f \in K[x]$ cubic monic, $\mathrm{char}(K) \neq 2$, are smooth if and only if f is separable, i.e. it contains no multiple root. The polynomial $x(x^2 + ABx + B^2)$ has a multiple root if and only if $(AB)^2 - 4B^2 = B^2(A-2)(A+2)$ is equal to zero. Hence if $A \neq \pm 2$, then the curve given by $y^2 = x^3 + ABx^2 + B^2x$ is smooth—and this is also, not surprisingly, the condition for the Montgomery curve to be smooth.

Assume $A \neq \pm 2$. Denote the Montgomery curve by M and the Weierstraß curve of $y^2 = x^3 + ABx^2 + B^2x$ by C. Note that $\sigma\colon (\alpha_1, \alpha_2) \mapsto (B\alpha_1, B^2\alpha_2)$ is a bijection $M \to C$. Extend this bijection by $\infty \mapsto \infty$. The group structure of $C(K)$ may be transferred upon M in such a way that $\sigma(\alpha) \oplus \sigma(\beta) = \sigma(\alpha\tilde{\oplus}\beta)$ for all $\alpha, \beta \in M \cup \{\infty\}$. This may be also written as $\alpha\tilde{\oplus}\beta = \sigma^{-1}(\sigma(\alpha) \oplus \sigma(\beta))$ for all $\alpha, \beta \in M \cup \{\infty\}$.

Suppose that $\alpha = (\alpha_1, \alpha_2)$. Then

$$\tilde{\ominus}\alpha = \sigma^{-1}(\ominus(B\alpha_1, B^2\alpha_2)) = \sigma^{-1}(B\alpha_1, -B^2\alpha_2) = (\alpha_1, -\alpha_2).$$

The formula for opposite elements thus does not change, and so we can write $\ominus$ in place of $\tilde{\ominus}$ when the unary minus is being used.

The value of λ for $(B\alpha_1, B^2\alpha_2) \oplus (B\beta_1, B^2\beta_2)$ comes from (A.3) as

$$\frac{3B^2\alpha_1^2 + 2AB^2\alpha_1 + B^2}{2B^2\alpha_2} = \frac{3\alpha_1^2 + 2A\alpha_1 + 1}{2\alpha_2} \text{ if } \alpha = \beta, \text{ and } B\frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} \text{ if } \alpha \neq \beta.$$

Assume that $\alpha \neq \ominus\beta$, and set $\gamma = \alpha\tilde{\oplus}\beta$, $\gamma = (\gamma_1, \gamma_2)$. By (A.4),

$$(\gamma_1, \gamma_2) = (-\alpha_1 - \beta_1 + B^{-1}\lambda^2 - A, B^{-1}\lambda(\alpha_1 - \gamma_1) - \alpha_2).$$

Let us express the latter formula using $\tilde{\lambda} = B^{-1}\lambda$. Note that $\tilde{\lambda}$ expresses the slope of the line connecting α and β, if $\alpha \neq \beta$. Indeed,

$$\tilde{\lambda} = \frac{3\alpha_1^2 + 2A\alpha_1 + 1}{2B\alpha_2} \text{ if } \alpha = \beta, \quad \tilde{\lambda} = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} \text{ if } \alpha \neq \beta, \text{ and} \tag{M.2}$$

$$(\gamma_1, \gamma_2) = (-\alpha_1 - \beta_1 + B\tilde{\lambda}^2 - A,\ \tilde{\lambda}(\alpha_1 - \gamma_1) - \alpha_2). \tag{M.3}$$

Assume $\alpha_1 \neq \beta_1$ and use the fact that $\alpha\tilde{\ominus}\beta = \alpha\tilde{\oplus}(\beta_1, -\beta_2)$. Let $\alpha\tilde{\ominus}\beta = \delta = (\delta_1, \delta_2)$. By (M.3),

$$(\delta_1, \delta_2) = (-\alpha_1 - \beta_1 + B\tilde{\tilde{\lambda}}^2 - A, \tilde{\tilde{\lambda}}(\alpha_1 - \delta_1) - \alpha_2), \text{ where } \tilde{\tilde{\lambda}} = \frac{\alpha_2 + \beta_2}{\alpha_1 - \beta_1}. \tag{M.4}$$

Proposition M.1. *Let $\tilde{\oplus}$ be the group operation upon a Montgomery curve M given over K by $By^2 = x^3 + Ax^2 + x$. Let $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ be K-rational points of M, $\alpha_1 \neq \beta_1$. Put $\gamma = \alpha\tilde{\oplus}\beta = (\gamma_1, \gamma_2)$ and $\delta = \alpha\tilde{\ominus}\beta = (\delta_1, \delta_2)$. Then*

$$\gamma_1\delta_1(\alpha_1 - \beta_1)^2 = (\alpha_1\beta_1 - 1)^2. \tag{M.5}$$

Proof. Start with (M.3) and express $B\alpha_2^2$ and $B\beta_2^2$ by means of the Montgomery equation to get

$$\begin{aligned}\gamma_1(\alpha_1 - \beta_1)^2 &= B(\alpha_2 - \beta_2)^2 - (A + \alpha_1 + \beta_1)(\alpha_1 - \beta_1)^2 \\ &= -2B\alpha_2\beta_2 + (\alpha_1^3 + A\alpha_1^2 + \alpha_1) + (\beta_1^3 + A\beta_1^2 + \beta_1) \\ &\quad - \alpha_1^3 - \beta_1^3 + \alpha_1^2\beta_1 + \alpha_1\beta_1^2 - A\alpha_1^2 - A\beta_1^2 + 2A\alpha_1\beta_1 \\ &= -2B\alpha_2\beta_2 + \alpha_1\beta_1(\alpha_1 + \beta_1 + 2A) + \alpha_1 + \beta_1.\end{aligned}$$

Therefore

$$\begin{aligned}\gamma_1(\alpha_1 - \beta_1)^2\alpha_1\beta_1 &= -2B\alpha_2\beta_2\alpha_1\beta_1 + \beta_1^2(\alpha_1^3 + A\alpha_1^2 + \alpha_1) + \alpha_1^2(\beta_1^3 + A\beta_1^2 + \beta_1) \\ &= -2B\alpha_1\beta_1\alpha_2\beta_2 + B\beta_1^2\alpha_2^2 + B\alpha_1^2\beta_2^2 = B(\beta_1\alpha_2 - \beta_2\alpha_1)^2.\end{aligned}$$

The right hand side of (M.4) is obtained from the right hand side of (M.3) by replacing β_2 with $-\beta_2$. Hence we have

$$\begin{aligned}\gamma_1(\alpha_1 - \beta_1)^2\alpha_1\beta_1 = B(\beta_1\alpha_2 - \beta_2\alpha_1)^2 \text{ and} \\ \delta_1(\alpha_1 - \beta_1)^2\alpha_1\beta_1 = B(\beta_1\alpha_2 + \beta_2\alpha_1)^2.\end{aligned} \tag{M.6}$$

By multiplying, $\gamma_1\delta_1(\alpha_1 - \beta_1)^4\alpha_1^2\beta_1^2 = B^2(\beta_1^2\alpha_2^{2}` - \beta_2^2\alpha_1^2)^2$. Now,

$$\begin{aligned}B(\beta_1^2\alpha_2^2 - \beta_2^2\alpha_1^2) = \beta_1^2(\alpha_1^3 + A\alpha_1^2 + \alpha_1) - \alpha_1^2(\beta_1^2 + A\beta_1^2 + \beta_1) \\ = (\alpha_1\beta_1)^2(\alpha_1 - \beta_1) + \alpha_1\beta_1(\beta_1 - \alpha_1) = (\alpha_1 - \beta_1)\alpha_1\beta_1(\alpha_1\beta_1 - 1).\end{aligned}$$

Hence $\gamma_1\delta_1(\alpha - \beta_1)^4(\alpha_1\beta_1)^2 = (\alpha_1\beta_1)^2(\alpha - \beta_1)^2(\alpha_1\beta_1 - 1)^2$, and so

$$\gamma_1\delta_1(\alpha_1 - \beta_1)^2 = (\alpha_1\beta_1 - 1)^2.$$

This yields (M.5) if $\alpha_1\beta_1 \neq 0$. If $\beta_1 = 0$, then $\beta_2 = 0$, $\gamma_1 = \delta_1 = -\alpha_1 + B\alpha_2^2\alpha_1^{-2} - A$ and $\alpha_1^2\gamma_1 = -\alpha_1^3 + B\alpha_2^2 - A\alpha_1^2 = \alpha_1$. Thus $\alpha_1\gamma_1 = \alpha_1\delta_1 = 1$, and both sides of (M.5) are equal to 1.

If $\alpha_1 = 0$, then $\alpha_2 = 0$, $\gamma_1 = -\beta_1 + B\beta_2^2\beta_1^{-2} - A = \delta_1$ and $\gamma_1\beta_1^2 = -\beta_1^3 + B\beta_2^2 - A\beta_1^2 = \beta_1$. Hence $\gamma_1\beta_1 = \delta_1\beta_1 = 1$, and both sides of (M.5) are equal to 1 again. □

There exists a natural technique how to compute $[n]P$ by means of a sequence $1 = n_1, \dots, n_k$ of integers such that in the ith round both $[n_i]P$ and $[n_i{+}1]P$ are known. This is known as *Montgomery's ladder* and is discussed below. If $\beta = [n_i]P$ and $\alpha = [n_i + 1]P$, then $\alpha\tilde{\ominus}\beta = P$. Hence (M.5) may be used to obtain $\gamma = \alpha\tilde{\oplus}\beta = [2n_i + 1]P$. The practicality of such a procedure follows from the fact that we may work only in the first coordinate. For all $[n_i]P$ and $[n_i + 1]P$ only the first coordinate is being computed, and the second coordinate of $[n]P$ is retrieved from the last two elements of the sequence, cf. Lemma M.2.

Since Montgomery's ladder needs also doubling, we have to verify that doubling can be performed in the first coordinate only too:

Let $(\gamma_1, \gamma_2) = [2]\alpha$, where $\alpha = (\alpha_1, \alpha_2)$ and $\alpha_2 \neq 0$. By (M.2) and (M.3), γ_1 is equal to $-2\alpha_1 - A + B(3\alpha_1^2 + 2A\alpha_1 + 1)^2(2B\alpha_2)^{-2}$. Thus

$$\begin{aligned}
4B\gamma_1\alpha_2^2 &= -8\alpha_1(B\alpha_2^2) + (3\alpha_1^2 + 2A\alpha_1 + 1)^2 - 4A(B\alpha_2^2) \\
&= -(8\alpha_1 + 4A)(\alpha_1^3 + A\alpha_1^2 + \alpha_1) + 9\alpha_1^4 + 12A\alpha_1^3 + (6 + 4A^2)\alpha_1^2 + 4A\alpha_1 + 1 \\
&= \alpha_1^4 - 2\alpha_1^2 + 1.
\end{aligned}$$

Hence

$$\gamma_1 = \frac{(\alpha_1^2 - 1)^2}{4B\alpha_2^2} = \frac{(\alpha_1^2 - 1)^2}{4(\alpha_1^3 + A\alpha_1^2 + \alpha_1)}. \tag{M.7}$$

In the context of Montgomery's ladder the points occurring in the following statement have this meaning: $\gamma = [n + 1]P$, $\alpha = [n]P$ and $\beta = P \neq (0, 0)$. The goal is to determine α_2 from knowledge of α_1, γ_1, β_1 and β_2.

Lemma M.2. *Let $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$ and $\gamma = (\gamma_1, \gamma_2)$ be points of a Montgomery curve over K given by $By^2 = x^3 + Ax^2 + x$. Suppose that $\alpha_1 \neq \beta_1$, $\beta \neq (0, 0)$ and that $\gamma = \alpha\tilde{\oplus}\beta$, where $\tilde{\oplus}$ is the group operation upon $M \cup \{\infty\}$. Then*

$$\alpha_2 = \frac{\alpha_1\beta_1(\alpha_1 + \beta_1 + 2A) + \alpha_1 + \beta_1 - \gamma_1(\alpha_1 - \beta_1)^2}{2B\beta_2}$$

Proof. The first equation in the proof of Proposition M.1 is

$$\gamma_1(\alpha_1 - \beta_1)^2 = -2B\alpha_2\beta_2 + \alpha_1\beta_1(\alpha_1 + \beta_1 + 2A) + \alpha_1 + \beta_1.$$

It remains to express α_2 using this equation. □

M.1. **Montgomery's ladder.** Let us start by an example. The binary expansion of, say, $n = 49$ is 110001 since $49 = 32 + 16 + 1$. The decimal expression of binary integers 1, 11, 110, 1100, 11000 and 110001 is 1, 3, 6, 12, 24 and 49. Put $n_1 = 1$, $n_2 = 3$, $n_3 = 6$, $n_4 = 12$, $n_5 = 24$ and $n_6 = 49$, and set $n_i' = n_i + 1$, $1 \leq i \leq 6$. Note that $(3, 4) = (1 + 2, 2 + 2)$, $(6, 7) = (3 + 3, 3 + 4)$, $(12, 13) = (6 + 6, 6 + 7)$, $(24, 25) = (12 + 12, 12 + 13)$ and $(49, 50) = (24 + 25, 25 + 25)$. Obviously there are two patterns. Either $(n_{i+1}, n_{i+1}') = (2n_i, n_i + n_i')$, or $(n_{i+1}, n_{i+1}') = (n_i + n_i', 2n_i')$. The former equality holds if the rightmost bit of n_{i+1} is equal to 0, while the latter equality holds if the rightmost bit of n_{i+1} is equal to 1. This will be proved below.

Now suppose that our goal is to compute $[n]P$, $P \neq (0, 0)$ a point of a Montgomery curve M. Let $x_i, y_i, x_i', y_i' \in K$ be such that $[n_i]P = (x_i, y_i)$ and $[n_i']P = (x_i', y_i')$. The sequence $n_1, n_2, \dots, n_k$ is defined so that $n_k = n$. Thus $[n]P = (x_k, y_k)$.

The recommended procedure is to compute x_i and x_i' by means of (M.5) and (M.7), and then use Lemma M.2 to retrieve y_k from knowledge of x_k, x_k', and $P = (x_1, y_1)$.

Let us be more concrete. Suppose first that $(n_{i+1}, n_{i+1}') = (2n_i, n_i + n_i')$. Then

$$x_{i+1} = \frac{(x_i^2 - 1)^2}{4(x_i^3 + Ax_i^2 + x_i)}, \text{ and } x_{i+1}' = \frac{(x_i x_i' - 1)^2}{x_1(x_i' - x_i)^2}. \tag{M.8}$$

If $(n_{i+1}, n'_{i+1}) = (n_i + n'_i, 2n'_i)$, then

$$x_{i+1} = \frac{(x_i x'_i - 1)^2}{x_1(x'_i - x_i)^2}, \text{ and } x'_{i+1} \;\; = \frac{(x'^2_i - 1)^2}{4(x'^3_i + Ax'^2_i + x'_i)}. \tag{M.9}$$

Finally, by Lemma M.2,

$$y_k = \frac{x_1 x_k(x_1 + x_k + 2A) + x_1 + x_k - x'_k(x_k - x_1)^2}{2By_1}.$$

Of course, the scheme assumes that the order of P is greater than $n + 1$. Thus $[m]P \neq \infty$ for any m, $1 \leq m \leq n + 1$.

When implementing the arithmetic of Montgomery curves the effeciency may be enhanced by using projective coordinates.

Let us formalize observations deduced from the initial example. Note that if $n = \sum_{0 \leq i < k} a_i 2^i$ is a binary expansion of n (thus $a_i \in \{0, 1\}$ and $a_{k-1} = 1$), then the sequence $n_1, n_2, \ldots, n_k$ constructed above can be expressed as $n_1 = 1$, $n_2 = 2 + a_{k-2} = 2a_{k-1} + a_{k-2}$, $n_3 = 4a_{k-1} + 2a_{k-2} + a_{k-3}$, etc. Thus $n_j = \sum_{1 \leq i \leq j} a_{k-i} 2^{j-i}$.

Lemma M.3. *Let $n \geq 1$ be an integer, and let $\sum_{0 \leq i < k} a_i 2^i$ be its binary expansion, $a_{k-1} = 1$. For $j \in \{1, \ldots, k\}$ define n_j as $\sum_{1 \leq i \leq j} a_{k-i} 2^{j-i}$, and put $n'_j = n_j + 1$. Then $n_1 = 1$, $n_k = n$, and for every j, $1 \leq j < k$ the following holds:*

- *If $a_{k-j-1} = 0$, then $n_{j+1} = 2n_j$ and $n'_{j+1} = n_j + n'_j$.*
- *If $a_{k-j-1} = 1$, then $n_{j+1} = n_j + n'_j$ and $n'_{j+1} = 2n'_j$.*

Proof. Put $\varepsilon = a_{k-j-1}$. By the definition, $n_{j+1} = 2n_j + \varepsilon$. If $\varepsilon = 0$, then $n_{j+1} = 2n_j$. If $\varepsilon = 1$, then $n_{j+1} + 1 = 2(n_j + 1)$. □

M.2. **Turning Weierstraß into Montgomery.** Recall that by multiplying the equation $By^2 = x^3 + Ax^2 + x$ by B^3 we obtain a K-equivalent Weierstraß equation $y^2 = x^3 + ABx^2 + B^2x$. Hence we may immediately claim the following fact:

Lemma M.4. *A Weierstraß equation $y^2 = f(x)$, where $f(x) = x^3 + a_2x^2 + a_4x + a_6$, is K-equivalent to a Montgomery equation if and only if it is K-equivalent to a Weierstraß equation $y^2 = x^3 + \tilde{a}_2x^2 + \tilde{a}_4x$ in which $\tilde{a}_4$ is in K a nonzero square.*

Assume $\mathrm{char}(K) > 3$. Expressing $x^3 + ABx^2 + B^2x$ as a polynomial in $x + AB/3$ shows that $By^2 = x^3 + Ax^2 + x$ is K-equivalent to

$$y^2 = x^3 + B^2\left(1 - \frac{A^2}{3}\right)x - \frac{AB^3}{3} + \frac{2(AB)^3}{27}. \tag{M.10}$$

If $y^2 = x^3 + ax + b$, then it may not be easy to decide whether there exist A and B such that $a = B^2(1 - A^2/3)$ and $b = -(AB^3)/3 + 2(AB)^3/27$. The following structural description may be then useful.

Proposition M.5. *A Weierstraß equation $y^2 = f(x)$ is K-equivalent to a Montgomery equation if and only if there exists $\zeta \in K$ such that $f(\zeta) = 0$ and $f'(\zeta)$ is in K a nonzero square.*

Proof. If $f(x) = x^3 + ABx^2 + B^2x$, then $f'(x) = 3x^2 + 2ABx + B^2$, $f(0) = 0$ and $f'(0) = B^2$. For the converse direction suppose that $y^2 = f(x)$, $f'(\zeta) = B^2$ and $f(\zeta) = 0$. Put $\tilde{f}(x) = f(x + \zeta)$. Then $\tilde{f}(x) = x^3 + \tilde{a}_2x^2 + \tilde{a}_4x + \tilde{a}_6$, and $\tilde{a}_6 = \tilde{f}(0) = f(\zeta) = 0$. Furthermore, $\tilde{a}_4 = \tilde{f}'(0) = f'(\zeta)$ is assumed to be square. The equation $y^2 = \tilde{f}(x)$ is thus equivalent to a Montgomery equation, by Lemma M.4.

To finish the proof we have to show that if $y^2 = f(x)$ and $y^2 = \tilde{f}(x)$ are K-equivalent Weierstraß equations, then from the existence of ζ with $f(\zeta) = 0$ and $f'(\zeta) \in (K^*)^2$ there follows the existence of $\tilde{\zeta}$ with the same properties. (This part of the proof is necessary since without it there would remain open a possibility that a Montgomery equation is K-equivalent to a Weierstraß equation that does not have the required property.) If $\tilde{f}(x) = f(x+\mu)$, set $\tilde{\zeta} = \zeta - \mu$. If $\tilde{f}(x)$ is obtained from $f(\lambda_1 x)$, $\lambda_1 \in K^*$, then λ_1 must be a square (cf. the discussion before (M.1)). Suppose that $f(x) = x^3 + a_2x^2 + a_4x + a_6$. Then $\lambda^6y^2 = (\lambda^2x)^3 + a_2\lambda^2(\lambda^2x)^2 + a_4\lambda^4(\lambda^2x) + a_6\lambda^6$. Thus $\tilde{f}(x) = x^3 + a_2\lambda^2x^2a + a_4\lambda^4x + a_6\lambda^6$. Put $\tilde{\zeta} = \lambda^2\zeta$. Then $\tilde{f}(\tilde{\zeta}) = \lambda^6 f(\zeta) = 0$, and $\tilde{f}'(\tilde{\zeta}) = 3\tilde{\zeta}^2 + 2a_2\lambda^2\tilde{\zeta} + a_4\lambda^4 = \lambda^4(3\zeta^2 + 2a_2\zeta + a_4) = \lambda^4(f'(\zeta))$ is a square. □

Corollary M.6. *Let $p \equiv 1 \bmod 4$ be a prime, and let $f \in \mathbb{Z}_p[x]$ be a cubic monic separable polynomial that splits over $\mathbb{Z}_p$ (i.e. all roots of f are in $\mathbb{Z}_p$). If $f(0) \neq 0$, then the Weierstraß equation $y^2 = f(x)$ is K-equivalent to a Montgomery equation.*

Proof. By the assumptions, $f(x) = (x-\zeta_1)(x-\zeta_2)(x-\zeta_3)$, where $\zeta_i \in \mathbb{Z}_p$. We have

$$-\prod f'(\zeta_i) = \prod_{i<j}(\zeta_i - \zeta_j)^2.$$

This is because both sides of the equality express the discriminant of f. (This equality can be also verified directly, which is an option for those who are not familiar with discriminants.) Because -1 is modulo p a square, $\prod f'(\zeta_i)$ is also a square. Therefore at least one of $f'(\zeta_i)$ has to be a square too. □

It is not difficult to solve completely the question when two Montgomery equations are K-equivalent. Here we shall restrict our attention only to the fact that $By^2 = x^3 + Ax^2 + x$ holds if and only if $-By^2 = (-x)^3 - A(-x)^2 + (-x)$. A Montgomery equation with parameters (A, B) is hence K-equivalent to a Montgomery equation with parameters $(-A, -B)$.

E. Edwards curves

An *elliptic curve* over a field K is a projective curve E such that the function field $K(E)$ is of genus 1, and E contains at least one K-rational point. The curve E is often considered in its affine version. This is particularly true if the curve is smooth and there is only one point at infinity.

For each elliptic curve there exists a smooth Weierstraß curve C such that $K(E) \cong K(C)$. For each elliptic curve it is possible to define a group operation $\oplus$. The group is then denoted by $E(K)$. If $K(E) \cong K(C)$, then $E(K) \cong C(K)$. Are there any reasons why there should be considered other elliptic curves but the smooth Weierstraß curves? One reason may be computational, and this is why Montgomery curves have been considered. Another reason may be structural. In cryptographic applications the fact that doubling and adding proceeds differently makes an implementation vulnerable to side channel attacks. We would like to have an elliptic curve with only one formula for both doubling of a point, and addition of two distinct points. Such a formula is sometimes known as a *closed formula* or a *uniform formula.*

Edwards curves fulfil such a requirement. Some of the Edwards curves have no K-rational point at infinity, and these are those for which a closed formula can be used indiscriminately. This is also true for the so called twisted Edwards curves, which is a somewhat more general notion. Twisted Edwards curves correspond to Montgomery curves. Assume $\operatorname{char}(K) \neq 2$. Then for each twisted Edwards curve E there exists a smooth Montgomery curve M such that $M(K) \cong E(K)$, and vice versa.

To define the group $E(K)$, E an elliptic curve, in full generality, the notion of a place is needed. Elements of $E(K)$ are places of degree one. If $\alpha \in E$ is a smooth point, then there is only one place at α. However, if α is a singular point, then there are either more places at α, or there is a place of degree > 1. That makes the connection between K-rational projective points of E and elements of $E(K)$ a bit more complicated. If E is a twisted Edwards curve, then all affine points are smooth and all points at infinity are singular. If all places induced by K-rational points at infinity are of degree > 1, then each element of $E(K)$ can be represented by exactly one affine point. Thus in this case $E(K)$ may be constructed directly upon the set of all affine K-rational points. In the general case the affine points may be also used, but their set has to be extended by two or four extra elements that correspond to “places at infinity”. There are several ways how to do that formally, and some of them have computational consequences. These are discussed below.

E.1. **Branches and the definition.** Consider a curve described by equation

$$y^2 + x^2 = 1 + dx^2y^2 \tag{E.1}$$

that is defined over real numbers. There are good reasons to expect that the corresponding projective curve will have one or two components of connectivity, each of them closed, similarly as in the case of Weierstraß curves. However, the number of connectivity components of an affine curve may be bigger than the number of components of its projective completion. (Think about a hyperbole which has only one component in projective coordinates, but two in affine coordinates.) For a while, for the sake of simplicity of expression, call an affine component of connectivity a *branch*. Denote the curve by E.

Suppose first that $d = s^2$ and $s > 1$. When (E.1) is written in the form $y^2 = (1 - x^2)/(1 - s^2x^2)$, then it is easy to deduce that in this case there exists exactly one branch which satisfies $y > 0$ and $x \in (-s^{-1}, s^{-1})$. This branch has a shape of the letter $\cup$, with $x = -s^{-1}$ and $x = s^{-1}$ being tangents at infinity, and $(0, 1)$

being the bottom point. By turning the branch bottom up, i.e. by reflecting it along the axis x (the line $y = 0$), we obtain another branch. This branch satisfies $y < 0$ and $x \in (-s^{-1}, s^{-1})$. Since the definition of the curve is $x \leftrightarrow y$ symmetric, the other two branches are obtained by right angle rotation of the branches that have been already described. So there are four branches, none of which is closed. Extreme points of these branches are $(0,1)$, $(0,-1)$, $(1,0)$ and $(-1,0)$. Note that these points belong to E for any field K and any element $d \in K$.

At this point the reader might wish to guess the number of points at infinity without actually computing them.

If $0 < s < 1$ and $d = s^2$, then there are five branches, and the central branch is closed.

If $d < 0$, then any point $(\alpha, \beta) \in E$ fulfils $|\alpha| \leq 1$ since $\beta^2 = (1-\alpha^2)/(1-d\alpha^2)$ and $1 - d\alpha^2 \geq 1$ for every α. Similarly, $|\beta| \leq 1$. In this case there is only one branch. The branch is closed and resembles a somewhat smoothed star from the logo of an Orion chocolate bar. Note that if $d = 0$, then the curve coincides with a circle. With decreasing d, the circle gets more and more pressed crosswise towards the centre (the pressure comes along the quadrangle axes).

Consider now the projective curve induced by (E.1), for any field K, $\mathrm{char}(K) \neq 2$. The equation is $Y^2Z^2 + X^2Z^2 = Z^4 + dX^2Y^2$. Assume $d \neq 0$. If $Z = 0$, then $dX^2Y^2 = 0$. There are thus two points at infinity, $(0:1:0)$ and $(1:0:0)$. Both of them are singular. If $K = \mathbb{R}$ and $d = s^2 > 1$, then the two affine branches with $x \in (-s^{-1}, s)$ make one projective branch in the shape of the digit 8. The point of crossing is equal to the projective point $(0:1:0)$. There are two distinct places of degree 1 at this point. Think about the point as if consisting of two “ideal” points. Separating them “resolves the singularity” and changes the shape of 8 into a (topological) circle. If $1 > d = s^2 > 0$, then the situation is similar but somewhat different since there is a central closed affine branch. The other four affine branches form a single projective branch the shape of which can be represented by two circles that intersect in two points. The points of crossing are $(1:0:0)$ and $(0:1:0)$. Singularities can be resolved in a similar manner, and that makes this projective branch a topological circle too.

If $d < 0$, then each place of degree one of $K(E)$ corresponds to a (unique) affine point. In fact, this is true for any field K, $\mathrm{char}(K) \neq 2$, when d is not a square. How does this relate to the fact that the projective curve contains K-rational points $(1:0:0)$ and $(0:1:0)$ in this case too? The answer is that at each of these points there sits a single place, and this place is not of degree one, but of degree two. These points thus do not influence the structure of the group $E(K)$. Of course, the situation changes if the same curve is considered over the field $K[\sqrt{d}]$.

An *Edwards curve* over K, $\mathrm{char}(K) \neq 2$, is any curve given by (E.1), with $d \notin \{0, 1\}$. A *twisted Edwards curve* over K, $\mathrm{char}(K) \neq 2$ is a curve given by

$$ax^2 + y^2 = 1 + dx^2y^2, \text{ where } a, d \in K^* \text{ and } a \neq d. \tag{E.2}$$

Usage of the adjective “twisted” indicates that the class of twisted Edwards curves extends the class of Edwards curves only modestly. To see this note that if $a = b^2$, then $(bx)^2 + y^2 = 1 + db^{-2}(bx)^2y^2$. A twisted Edwards curve with parameters (b^2, d) is K-equivalent to the Edwards curve given by $x^2 + y^2 = 1 + db^{-2}x^2y^2$. More generally, there is a K-equivalence between parameters (b^2c, d) and $(c, b^{-2}d)$. To cover the class of twisted Edwards curves over a finite field $\mathbb{F}_q$, q odd, it is thus enough to consider the Edwards curves, and the curves given by $\vartheta x^2 + y^2 = 1 + dx^2y^2$, where ϑ is a preselected nonsquare.

Let us now verify that the polynomial $ax_1^2 + x_2^2 - 1 - dx_1^2x_2^2 \in K[x_1, x_2]$ is absolutely irreducible if $a, d \in K^*$ and $a \neq d$. Up to now we have tacitly assumed that this is true. If it have not been true, we could not have had considered the function field $K(E)$ since this assumes that the polynomial defining E is irreducible.

Proposition E.1. *Let K be a field,* $\mathrm{char}(K) \neq 2$. *Assume that $a_1, a_2, d \in K^*$. The polynomial $f(x_1, x_2) = a_1x_1^2 + a_2x_2^2 - 1 - dx_1^2x_2^2$ is absolutely irreducible if and only if $d \neq a_1a_2$.*

Proof. If $d = a_1a_2$, then $f = (a_1x_1^2 - 1)(1 - a_2x_2^2)$. Let $f = g_1g_2$, where $g_1, g_2 \in \bar{K}[x_1, x_2]$. If $\deg_{x_1}(g_1) = \deg_{x_2}(g_1) = 2$, then $g_2 \in \bar{K}^*$. Assume that $g_1, g_2 \notin \bar{K}^*$. Suppose first that $g_i = \alpha_ix_i^2 + \beta_ix_i + \gamma_i \in \bar{K}[x_i]$, $i \in \{1, 2\}$. In the polynomial $f(x_1, x_2)$ the coefficients at both $x_1^2x_2$ and $x_1x_2^2$ vanish. We have $\alpha_1\alpha_2 = -d \neq 0$. Hence both β_1 and β_2 must vanish too. Now, $f = g_1g_2 = \alpha_1\alpha_2x_1^2x_2^2 + \alpha_1\gamma_2x_1^2 + \alpha_2\gamma_1x_2^2 + \gamma_1\gamma_2$. Therefore $\gamma_1\gamma_2 = -1$, $\alpha_1\gamma_2 = a_1$, $\alpha_2\gamma_1 = a_2$ and $-d = -\alpha_1\gamma_1\gamma_2\alpha_2 = -a_1a_2$.

Assume $d \neq a_1a_2$. We have shown that there cannot be $\deg_{x_i}(g_j) \in \{0, 2\}$ for all $i, j \in \{1, 2\}$. Hence there exists $i \in \{1, 2\}$ such that $\deg_{x_i}(g_1) = \deg_{x_i}(g_2) = 1$. Because of the $x_1 \leftrightarrow x_2$ symmetry it may be assumed that $i = 1$. This means that the polynomial f splits over the field of rational functions $\bar{K}(x_2)$ when regarded as a quadratic polynomial in one variable x_1. This can happen if and only if the discriminant $-4a_1(a_2x_2^2 - 1)(1 - da_1^{-1}x_2^2)$ is a square in $\bar{K}[x_2]$. If it is a square, then all of the roots have to have an even multiplicity. This is not possible since the polynomials $a_2x_2^2 - 1$ and $1 - da_1^{-1}x_2^2$ have no common root because $d \neq a_1a_2$ is assumed, and none of them has a double root because $\mathrm{char}(K) \neq 2$. □

Lemma E.2. *Let K be a field,* $\mathrm{char}(K) \neq 2$. *Assume that $a_1, a_2, d \in K^*$, $d \neq a_1a_2$ and $f(x_1, x_2) = a_1x_1^2 + a_2x_2^2 - 1 - dx_1^2x_2^2$. Let $\alpha_1, \alpha_2 \in \bar{K}$ be such that $f(\alpha_1, \alpha_2) = 0$. Then $(\partial f/\partial x_i)(\alpha_1, \alpha_2) \neq 0$ for at least one $i \in \{1, 2\}$.*

Proof. First note that $\partial f/\partial x_i = 2x_i(a_i - dx_j^2)$, where $i, j \in \{1, 2\}$ and $j \neq i$. If $\alpha_1 = 0$, then $\alpha_2^2 = a_2^{-1} \neq 0$ and $(\partial f/\partial x_2)(0, \alpha_2) = 2\alpha_2a_1 \neq 0$. Suppose that $\alpha_i \neq 0$ and $(\partial f/\partial x_i)(\alpha_1, \alpha_2) = 0$ for both $i \in \{1, 2\}$. Then $\alpha_1^2 = a_2d^{-1}$, $\alpha_2^2 = a_1d^{-1}$ and $f(\alpha_1, \alpha_2) = a_1a_2d^{-1} + a_1a_2d^{-1} - 1 - a_1a_2d^{-1} = d^{-1}(a_1a_2 - d) \neq 0$, a contradiction. □

Corollary E.3. *Let K be a field with* $\mathrm{char}(K) \neq 2$. *Any twisted Edwards curve over K is smooth at every of its affine points.*

If E is an elliptic curve over K, then any of its K-rational points may be chosen as the neutral element of $E(K)$. The choice is a matter of convention and is made so that the addition formula is as simple as possible. For twisted Edwards curves the neutral element has been chosen to be equal to $(0, 1)$. The closed formula for addition is

$$(\alpha_1, \alpha_2) \oplus (\beta_1, \beta_2) = \left(\frac{\alpha_1\beta_2 + \alpha_2\beta_1}{1 + d\alpha_1\alpha_2\beta_1\beta_2}, \frac{\alpha_2\beta_2 - a\alpha_1\beta_1}{1 - d\alpha_1\alpha_2\beta_1\beta_2} \right). \tag{E.3}$$

This formula works for any two affine points provided $d\alpha_1\alpha_2\beta_1\beta_2 \neq \pm 1$. If the latter condition is not satisfied, then the result is one of the places of infinity. Their number can be expressed as $2(\varepsilon_1 + \varepsilon_2)$, where $\varepsilon_1, \varepsilon_2 \in \{0, 1\}$ is defined so that $\varepsilon_1 = 1$ if d is a square, and $\varepsilon_2 = 1$ if ad^{-1} is square. Each of the places at infinity is an element of $E(K)$ that is either of order 2, or of order 4. Applications in cryptography are mainly concerned with points $P \in E(K)$ that are of large prime order. For these applications computation rules involving places at infinity are thus not needed. However, for other applications, like factorization algorithms, these formulas have to be established. This will be discussed later.

E.2. **Birational equivalence.** Let $C = V_f$, $f \in K[x_1, x_2]$ irreducible. Recall that $K(C)$ can be interpreted as a set of partial mappings $\rho\colon C \to K$ that can be represented by rational mappings $a(x_1, x_2)/b(x_1, x_2)$, $b \notin (f)$. If $\alpha \in C$ and $b(\alpha) \neq 0$, then $\rho(\alpha) = a(\alpha)/b(\alpha)$. Recall also that $a_1/b_1, a_2/b_2 \in K(x_1, x_2)$, $b_1, b_2 \notin (f)$, represent the same $\rho \in K(C)$ if and only if $a_1b_2 - a_2b_1 \in (f)$, i.e., if $(a_1 + (f))/(b_1 + (f))$ and $(a_2 + (f))/(b_2 + (f))$ denote the same element of $K(C)$. The partial mapping ρ is defined at $\alpha \in C$ whenever there exists a representative a/b such that $b(\alpha) \neq 0$. There are only finitely many $\alpha \in C$ at which $\rho(\alpha)$ is not defined. This is because if a/b represents ρ, $b \notin (f)$, then there are only finitely many $\alpha \in C$ such that $b(\alpha) = 0$. We say that ρ is defined *nearly everywhere.* This is meant as a synonym to *up to finitely many elements (or points).* Use $\mathrm{Dom}(\rho)$ to denote the *domain* of ρ, i.e. the set of elements where ρ is defined.

Let $C_1 = V_{f_1}$ and $C_2 = V_{f_2}$, where $f_1, f_2 \in K[x_1, x_2]$ are irreducible. A pair $\rho = (\rho_1, \rho_2) \in K(C)^2$ is said to be a *rational map* $C_1 \to C_2$ if $(\rho_1(\alpha), \rho_2(\alpha)) \in C_2$ whenever $\alpha \in \mathrm{Dom}(\rho) = \mathrm{Dom}(\rho_1) \cap \mathrm{Dom}(\rho_2)$. The curves C_1 and C_2 are *birationally equivalent* (over K) if there exist rational maps $\rho\colon C_1 \to C_2$ and $\sigma\colon C_2 \to C_1$ such that $\sigma\rho(\alpha) = \alpha$ for nearly all $\alpha \in C_1$ and $\rho\sigma(\beta) = \beta$ for nearly all $\beta \in C_2$ (an equivalent condition: $\sigma\rho(\alpha) = \alpha$ whenever $\alpha \in \mathrm{Dom}(\rho) \cap \rho^{-1}(\mathrm{Dom}(\sigma))$, similarly for β).

If $\rho\colon C_1 \to C_2$ and $\sigma\colon C_2 \to C_1$ yield a birational equivalence, then there exist mutually inverse K-isomorphisms $\sigma^*\colon K(C_1) \cong K(C_2)$ and $\rho^*\colon K(C_2) \cong K(C_1)$ such that $x_i + (f_1) \mapsto \sigma_i$ and $x_i + (f_2) \mapsto \rho_i$. In fact, $K(C_1)$ *and* $K(C_2)$ *are* K*-isomorphic if and only if* C_1 *and* C_2 *are birationally equivalent over* K.

To see that a birational equivalence induces mutually inverse isomorphisms of function fields is not too difficult. Nevertheless it is technically somewhat demanding. For a reader who would like to verify the statement the following comments may be useful. If $\tau \in K(C_1)$, then $\sigma^*(\tau) = \sigma^*(\tau(x_1 + (f_1), x_2 + (f_1))) = \tau(\sigma^*(x_1 + (f_1)), \sigma^*(x_2 + (f_1))) = \tau(\sigma_1, \sigma_2)$. Hence $\sigma^*(\tau)(\beta) = \tau(\sigma_1(\beta), \sigma_2(\beta)) = \tau\sigma(\beta)$ for every $\beta \in C_2$. Since $\sigma^*\rho^*(x_i + (f_2)) = \sigma^*(\rho_i)$ we get $\sigma^*\rho^*(x_1 + (f_2))(\beta) = \sigma^*(\rho_1)(\beta) = \rho_1\sigma(\beta)$. Now $\rho\sigma(\beta) = (\rho_1\sigma(\beta), \rho_2\sigma(\beta))$ is assumed to be equal to $\beta = (\beta_1, \beta_2)$ nearly everywhere. Hence $\rho_1\sigma(\beta) = \beta_1$ nearly everywhere, and therefore $\sigma^*\rho^*(x_1 + (f_2)) = \sigma^*(\rho_1) = x_1 + (f_2)$. Similarly, $\sigma^*\rho^*(x_2 + (f_2)) = x_2 + (f_2)$, and hence $\sigma^*\rho^* = \mathrm{id}_{K(C_2)}$. The equality $\rho^*\sigma^* = \mathrm{id}_{K(C_1)}$ follows in the same way.

If C_1 and C_2 are birationally equivalent elliptic curves, then $C_1(K) \cong C_2(K)$. This is because the structure of the abelian group $C_i(K)$ fully depends upon the structure of the function field $K(C_i)$. If the fields are isomorphic, then the groups are isomorphic too.

Recall that equations $f_1(x_1, x_2) = 0$ and $f_2(x_1, x_2) = 0$ are said to be K-equivalent if the polynomials can be obtained one from another by a linear substitution. Such substitutions induce a birational equivalence between C_1 and C_2 that is realized by affine mappings, i.e. by a linear change of coordinates, like in the case of Weierstraß and Montgomery curves. However, not every birational equivalence is affine. Below we shall observe that twisted Edwards curves are birationally equivalent to Montgomery curves. The advantage of invertible affine (or linear) mappings is that they are defined globally for all $\alpha \in \mathbb{A}^2 = \bar{K} \times \bar{K}$, and their inversions are affine (linear) too. The birational equivalence may be thus obtained by restricting a global mapping to curves.

A *linear fractional mapping* $x \mapsto (ax + c)/(bx + d)$, $ad - bc \neq 0$, nearly permutes an affine line (it may be extended to a permutation of the projective line by $\infty \mapsto a/b$ and $-d/b \mapsto \infty$). Linear fractional mappings thus may serve as a tool to define transformations of $\mathbb{A}^2$ that are very close to permutations. One of such transformations is used to associate Montgomery and Edwards curves:

Lemma E.4. *Assume* $\operatorname{char}(K) \neq 2$*. Then* $\vartheta : \beta \mapsto (\beta+1)/(\beta-1)$ *permutes the set* $K' = K \setminus \{0,1,-1\}$ *and* $\Psi\colon (\alpha,\beta) \mapsto (\vartheta(\beta), \vartheta(\beta)/\alpha)$ *is a bijection* $K^* \times K' \to K' \times K^*$*.*

Proof. If $\beta \neq 1$, then $\vartheta^2(\beta) = \beta$, $\vartheta(0) = -1$ and $\vartheta(-1) = 0$. Hence ϑ permutes K'. The mapping Ψ clearly sends $K^* \times K'$ to $K' \times K^*$ injectively. If $(\gamma,\delta) \in K' \times K^*$, then $(\gamma,\delta) = \Psi(\gamma/\delta, \vartheta^{-1}(\gamma))$. □

Lemma E.5. *Assume* $\operatorname{char}(K) \neq 2$*. The mappings*

$$(a,d) \mapsto \left(2\frac{a+d}{a-d}, 4\frac{1}{a-d}\right) \qquad (A,B) \mapsto \left(\frac{A+2}{B}, \frac{A-2}{B}\right) \tag{E.4}$$

are mutually inverse if $(a,d) \in K^* \times K^*$*,* $a \neq d$*, and* $(A,B) \in K \times K^*$*,* $A \neq \pm 2$*.*

Proof. Let $A = 2(a+d)/(a-d)$ and $B = 4/(a-d)$, where $a,d \in K$ and $a \neq d$. Then $B \neq 0$, $Aa - Ad = 2a + 2d$, $(A-2)a = (A+2)d$, $a = (4+Bd)/B$, $(A-2)(4+Bd) = ABd + 2Bd$, $-4 - Bd + 2A = Bd$, $d = (A-2)/B$, $4 + Bd = A + 2$ and $a = (A+2)/B$. This establishes a bijection between the set of all $(a,d) \in K \times K$, $a \neq d$, and the set $K \times K^*$. The rest is clear. □

Lemma E.6. *Let* $\operatorname{char}(K) \neq 2$ *and suppose that* $a,d \in K^*$ *are such that* $a \neq d$*. Set* $A = 2(a+d)/(a-d)$ *and* $B = 4/(a-d)$*, and assume that* $\alpha,\beta \in K$ *are such that* $\alpha \neq 0$ *and* $\beta \notin \{0,1,-1\}$*. Put* $u = (1+\beta)/(1-\beta)$ *and* $v = u/\alpha$*. Then*

$$a\alpha^2 + \beta^2 = 1 + d\alpha^2\beta^2 \iff Bv^2 = u^3 + Au^2 + u.$$

Proof. Multiplying the equality $Bv^2 = u^3 + Au^2 + u$ by $(1-\beta)^3$, dividing it by $1+\beta$, and using $(1+\beta)(1-\beta) = 1-\beta^2$ yields an equivalent equation

$$B(1-\beta^2)\alpha^{-2} = (1+\beta)^2 + A(1-\beta^2) + (1-\beta)^2 = A(1-\beta^2) + 2(1+\beta^2).$$

Hence $(1-\beta^2)(B\alpha^{-2} - A) = 2(1+\beta^2)$. Therefore

$$2(1-\beta^2)(2\alpha^{-2} - (a+d)) = 2(a-d)(1+\beta^2),$$

which is the same as $2\alpha^{-2} - (a+d) - 2\alpha^{-2}\beta^2 + \beta^2 d = a - d - d\beta^2$ and as $\alpha^{-2} - \alpha^{-2}\beta^2 + \beta^2 d = a$. The latter can be written as $1 + d\alpha^2\beta^2 = a\alpha^2 + \beta^2$. Nothing else is needed since none of the transformations changes the set of solutions because $a \neq 0$ and $\beta \notin \{-1,0,1\}$ has been assumed. □

Theorem E.7. *Let* K *be a field of characteristic* $\neq 2$*, and let* $a,d \in K^*$ *be such that* $a \neq d$*. Set* $A = 2(a+d)/(a-d)$ *and* $B = 4/(a-d)$*. The twisted Edwards curve* E *given by* $1 + dx_1^2x_2^2 = ax_1^2 + x_2^2$ *is birationally equivalent over* K *to the Montgomery curve* M *given by* $Bx_2^2 = x_1^3 + Ax_1^2 + x_1$*. The rational map* $E \to M$ *may be represented by* $((1+x_2)/(1-x_2), (1+x_2)/x_1(1-x_2))$*, and the inverse rational map* $M \to E$ *by* $(x_1/x_2, (x_1-1)/(x_1+1))$*.*

Proof. The described rational map $E \to M$ sends nearly all elements of E upon M by Lemma E.6. The mapping is injective and its image covers nearly all elements of M, by Lemma E.4. It is immediately clear that the described rational map $M \to E$ behaves as an inverse mapping at each point where it is possible to define composition of the both mappings. □

Corollary E.8. *Let* K *be a field of characteristic* $\neq 2$*. For each twisted Edwards curve* E *over* K *there exists a smooth Montgomery curve* M *that is birationally equivalent over* K *to* E*, and for each smooth Montgomery curve* M *over* K *there exists a twisted Edwards curve* E *that is birationally equivalent over* K *to* M*.*

Proof. This immediately follows from Theorem E.7 and Lemma E.5. □

E.3. **Completed curves and various formulas.** Formula (E.3) is not the only way how the addition upon a twisted Edwards curve may be expressed. The so called *dual addition law*

$$(\alpha_1, \alpha_2) \oplus (\beta_1, \beta_2) = \left(\frac{\alpha_1\alpha_2 + \beta_1\beta_2}{\alpha_2\beta_2 + a\alpha_1\beta_1}, \frac{\alpha_1\alpha_2 - \beta_1\beta_2}{\alpha_1\beta_2 - \alpha_2\beta_1} \right) \tag{E.5}$$

is an alternative. It gives the same result as (E.3) whenever the denominators in both (E.3) and (E.5) are nonzero. Obviously, (E.5) may never be used for doublings. However, it is important both theoretically and practically, since it is a source of various speed-ups. The speed-ups usually work differently for the doubling and for the addition of distinct points (which is often called a *generic addition*). They are used if the context does not require a closed formula that makes the computation resistant to side channel attacks.

Let us observe that the dual addition law really works. If the denominators are nonzero, then the equality

$$\frac{\alpha_1\alpha_2 + \beta_1\beta_2}{\alpha_2\beta_2 + a\alpha_1\beta_1} = \frac{\alpha_1\beta_2 + \alpha_2\beta_1}{1 + d\alpha_1\alpha_2\beta_1\beta_2}$$

holds if and only if

$$\begin{aligned}
&\alpha_1\alpha_2 + d\alpha_1^2\alpha_2^2\beta_1\beta_2 + \beta_1\beta_2 + d\alpha_1\alpha_2\beta_1^2\beta_2^2 \\
&\qquad = \alpha_1\alpha_2(1 + d\beta_1^2\beta_2^2) + \beta_1\beta_2(1 + d\alpha_1^2\alpha_2^2) \\
&\qquad\qquad = \alpha_1\alpha_2(a\beta_1^2 + \beta_2^2) + \beta_1\beta_2(a\alpha_1^2 + \alpha_2^2)
\end{aligned}$$

is equal to $(\alpha_2\beta_2 + a\alpha_1\beta_1)(\alpha_1\beta_2 + \alpha_2\beta_1)$. That is clearly true.

The proof for the second coordinate may be done similarly.

When the addition is computed upon **projective coordinates**, i.e. upon the zeros of $aX_1^2X_3^2 + X_2^2X_3^2 = X_3^4 + dX_1^2X_2^2$, then it is possible to order the operations in such a way that the addition of distinct point (the *generic addition*) costs 10M + 1S + 1a + 1d, where 1a + 1d refer to multiplications by a and d (which may be chosen small), while the cost of doubling is 3M + 4S + 1a.

There have been also used **inverted coordinates** which correspond to the equation $aX_1^{-2}X_3^{-2} + X_2^{-2}X_3^{-2} = X_3^{-4} + dX_1^{-2}X_2^{-2}$, and thus also to $aX_2^2X_3^2 + X_1^2X_3^2 = X_1^2X_2^2 + dX_3^4$. In these coordinates the cost of generic addition is 9M + 1S + 1a + 1d, and the doubling costs 3M + 4S + 1a + 1d.

We shall skip **extended coordinates** and turn directly to **completed coordinates**. They use projective coordinates, but not in $\mathbb{P}^2$ or $\mathbb{P}^3$, but in $\mathbb{P}^1 \times \mathbb{P}^1$. The curve, say U, is formed by all $((\alpha_1 : \alpha_2), (\beta_1 : \beta_2))$ for which the substitutions $(X_1, X_2) \mapsto (\alpha_1, \alpha_2)$ and $(Y_1, Y_2) \mapsto (\beta_1, \beta_2)$ fulfil

$$aX_1^2Y_2^2 + Y_1^2X_2^2 = X_2^2Y_2^2 + dX_1^2Y_1^2. \tag{E.6}$$

Note that $((\alpha_1 : \alpha_2), (\beta_1 : \beta_2)) = ((\mu\alpha_1 : \mu\alpha_2), (\nu\beta_1 : \nu\beta_2))$ for any $\mu, \nu \in \bar{K}^*$. The advantage of completed coordinates is that in this setting each K-rational point of U corresponds to exactly one place of degree one in the function field $K(E)$, where E is the curve given by $ax_1^2 + x_2^2 = 1 + dx_1^2x_2^2$. The points of $E(K)$ may hence be identified bijectively with the K-rational points of U. The affine points of E obviously embed into U by $(\alpha, \beta) \mapsto ((\alpha : 1), (\beta : 1))$. If d is a square in K, $d = s^2$, then $((1 : s), (1 : 0)) \in U$ and $((1 : -s), (1 : 0)) \in U$ express the two places at infinity that sit in the singular projective point $(0 : 1 : 0)$. If a/d is a square in K, $a/d = t^2$, then $((1 : 0), (t : 1))$ and $((1 : 0), (-t : 1))$ correspond to the places at infinity at $(1 : 0 : 0)$.

The computation of

$$((\alpha_1 : \alpha_2), (\beta_1 : \beta_2)) \oplus ((\gamma_1 : \gamma_2), (\delta_1 : \delta_2))$$

requires two formulas. One yields $((\mu_1 : \mu_2), (\nu_1 : \nu_2))$, and the other $((\mu_1' : \mu_2'), (\nu_1' : \nu_2'))$. Since $\mu_1\mu_2' = \mu_1'\mu_2$ and $\nu_1\nu_2' = \nu_1'\nu_2$, both formulas yield the same result if both of them belong to $\mathbb{P}^1 \times \mathbb{P}^1$. However, it may happen that $\mu_1 = \mu_2 = 0$ or $\nu_1 = \nu_2 = 0$. In such a case both (μ_1', μ_2') and (ν_1', ν_2') are distinct from $(0, 0)$, and $((\mu_1' : \mu_2'), (\nu_1' : \nu_2'))$ is the result of the addition. Similarly, if $\mu_1' = \mu_2' = 0$ or $\nu_1' = \nu_2' = 0$, then the result is $((\mu_1 : \mu_2), (\nu_1 : \nu_2))$. The formulas are as follows:

$$\begin{aligned}
\mu_1 &= \alpha_1\beta_2\gamma_2\delta_1 + \alpha_2\beta_1\gamma_1\delta_2, & \mu_1' &= \alpha_1\beta_1\gamma_2\delta_2 + \alpha_2\beta_2\gamma_1\delta_1,\\
\mu_2 &= \alpha_2\beta_2\gamma_2\delta_2 + d\alpha_1\beta_1\gamma_1\delta_1, & \mu_2' &= a\alpha_1\beta_2\gamma_1\delta_2 + \alpha_2\beta_1\gamma_2\delta_1,\\
\nu_1 &= \alpha_2\beta_1\gamma_2\delta_1 - a\alpha_1\beta_2\gamma_1\delta_2, & \nu_1' &= \alpha_1\beta_1\gamma_2\delta_2 - \alpha_2\beta_2\gamma_1\delta_1,\\
\nu_2 &= \alpha_2\beta_2\gamma_2\delta_2 - d\alpha_1\beta_1\gamma_1\delta_1, & \nu_2' &= \alpha_1\beta_2\gamma_2\delta_1 - \alpha_2\beta_1\gamma_1\delta_2.
\end{aligned} \tag{E.7}$$

Let us now observe how these formulas correspond to formulas (E.3) and (E.5). Let $\sigma = (\sigma_1, \sigma_2)$ and $\tau = (\tau_1, \tau_2)$. By (E.3) and (E.5) $\sigma \oplus \tau$ is equal to

$$\left(\frac{\sigma_1\tau_2 + \sigma_2\tau_1}{1 + d\sigma_1\sigma_2\tau_1\tau_2}, \frac{\sigma_2\tau_2 - a\sigma_1\tau_1}{1 - d\sigma_1\sigma_2\tau_1\tau_2}\right) \text{ and } \left(\frac{\sigma_1\sigma_2 + \tau_1\tau_2}{\sigma_2\tau_2 + a\sigma_1\tau_1}, \frac{\sigma_1\sigma_2 - \tau_1\tau_2}{\sigma_1\tau_2 - \sigma_2\tau_1}\right),$$

respectively.

Insert $\sigma, \tau \in \mathbb{A}^2$ into $\mathbb{P}^1 \times \mathbb{P}^1$ by

$$(\sigma_1, \sigma_2) \mapsto ((\sigma_1 : 1), (\sigma_2 : 1)) \text{ and } (\tau_1, \tau_2) \mapsto ((\tau_1 : 1), (\tau_2 : 1)).$$

Apply now (E.7) with $\alpha_1 = \sigma_1, \beta_1 = \sigma_2, \gamma_1 = \tau_1$, $\delta_1 = \tau_2$, and the other values being equal to 1. We obtain

$$\begin{aligned}
\mu_1 &= \sigma_1\tau_2 + \sigma_2\tau_1, & \mu_1' &= \sigma_1\sigma_2 + \tau_1\tau_2,\\
\mu_2 &= 1 + d\sigma_1\sigma_2\tau_1\tau_2, & \mu_2' &= a\sigma_1\tau_1 + \sigma_2\tau_2,\\
\nu_1 &= \sigma_2\tau_2 - a\sigma_1\tau_1, & \nu_1' &= \sigma_1\sigma_2 - \tau_1\tau_2,\\
\nu_2 &= 1 - d\sigma_1\sigma_2\tau_1\tau_2, & \nu_2' &= \sigma_1\tau_2 - \sigma_2\tau_1.
\end{aligned}$$

We see that rules (E.7) can be interpreted as a transformation of the main addition law (E.3) and the dual addition law (E.5) to projective points. However, in addition to that, rules (E.7) may be applied to points and places at infinity. For example consider $((1 : s), (1 : 0)) \oplus ((1 : -s), (1 : 0))$, where $s^2 = d$. Then $(\mu_1, \mu_2, \nu_1, \nu_2) = (0, d, -d, -d)$ and $(\mu_1', \mu_2', \nu_1', \nu_2') = (0, -d, 0, 0)$. Hence only the former quadruple may be used to compute the result of the addition. The result is

$$((0 : d), (-d : -d)) = ((0 : 1), (1 : 1)), \text{ i.e., the affine point } (0, 1).$$

Recall that $(0, 1)$ is the neutral element of the group. Points $((1 : s), (1 : 0))$ and $((1 : -s), (1 : 0))$ are thus opposite each to other.

G. Group structure and order, and examples over five elements

Let $q > 1$ be a prime power and E a (projective) elliptic curve over the finite field $\mathbb{F}_q$. Then

$$|q + 1 - |E(\mathbb{F}_q)|| \leq 2\sqrt{q}. \tag{G.1}$$

This is known as *Hasse's Theorem.* As a convention, the integer $q + 1 - |E(\mathbb{F}_q)|$ is often denoted by t. Now, $|E(\mathbb{F}_q)|$ is the order of the group $E(\mathbb{F}_q)$ that is constructed upon the set of K-rational points of E. When the number of points is the main focus, then the notation $\#E(\mathbb{F}_q)$ is often used. Hence

$$|E(\mathbb{F}_q)| = \#E(\mathbb{F}_q) = q + 1 - t \text{ and } |t| \leq 2\sqrt{q}. \tag{G.2}$$

Note that if E is an affine Weierstraß curve, then $t = q - r$, where r is the number of affine $\mathbb{F}_q$-rational points.

Let now K be any field, and E an elliptic curve over K. Let L be a subfield of $\bar{K}$ such that $L \supseteq K$. Each K-rational point of E is also L-rational. It follows that $E(K)$ is a subgroup of $E(L)$. In particular, $E(K) \leq E(\bar{K})$.

For each integer $m \geq 1$ put

$$E[m] = \{\alpha \in E; [m]\alpha = \mathcal{O}\}.$$

The symbol $\mathcal{O}$ is used to denote the neutral element of $E(K)$ and $E(\bar{K})$. That is a generic notation that makes especially sense when the form of E is not specified. (In Weierstraß curves the neutral element is denoted by ∞, while in Edwards curves $(0, 1)$ has been chosen. Some authors use $\mathcal{O}$ also in these situations, while some use 0—which may be confusing.)

Note that $E[m]$ is a subgroup of $E(\bar{K})$. This follows from $[m](\alpha \oplus \beta) = [m]\alpha \oplus [m]\beta$ and $[m](\ominus\alpha) = \ominus([m]\alpha)$.

Theorem G.1. *Let K be a field of characteristic p.*

- *If p does not divide $m \geq 2$, then $E[m] \cong \mathbb{Z}_m \times \mathbb{Z}_m$.*
- *If $m \geq 2$ is a power of $p > 0$, then either $E[m] \cong \mathbb{Z}_m$, or $E[m] = \mathcal{O}$.*

By basic properties of abelian groups,

$$E[m_1 m_2] \cong E[m_1] \times E[m_2] \text{ whenever } \gcd(m_1, m_2) = 1.$$

Hence $E[m]$ is known for any $m \geq 1$. Indeed, if $m = np^r$, $p \nmid n$, then $E[m] = E[n] \times E[p^r]$, and Theorem G.1 can be used.

If H is a finite subgroup of $E(\bar{K})$, then there exists $m \geq 1$ such that $H \leq E[m]$. (The choice of $m = |H|$ is always possible.) Hence each finite subgroup of $E(\bar{K})$ embeds into $\mathbb{Z}_m \times \mathbb{Z}_m$ for some $m \geq 1$.

Every subgroup of $\mathbb{Z}_m \times \mathbb{Z}_m$ is isomorphic to some $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$, where $m_1 \mid m_2$ and $m_2 \mid m$. We have:

Corollary G.2. *Let E be an elliptic curve over a field K, and let H be a finite subgroup of $E(K)$. Then there exist integers $m_2 \geq m_1 \geq 1$ such that $m_1 \mid m_2$ and $H \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$.*

If E is an elliptic curve over $\mathbb{F}_q$, then $E(\mathbb{F}_q)$ is finite. Hence Corollary G.2 applies to $E(\mathbb{F}_q)$. However, a somewhat stronger result is true:

Theorem G.3. *Let E be an elliptic curve over $\mathbb{F}_q$. Then there exist integers $m_2 \geq m_1 \geq 1$ such that $m_1 \mid m_2$, $m_1 \mid q - 1$, and $E(\mathbb{F}_q) \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$.*

There is a relationship between Theorem G.1 and the fact that complex elliptic curves take the shape of a torus. A rigorous description is not easy. Intuitively, think of the torus as being obtained from a rectangle by identifying the opposite sides. Suppose that the rectangle is a square of size m, and equip it with equidistant lines parallel to the axes so that a lattice of m^2 squares is formed. Think of the lattice points as of elements of $\mathbb{Z}_m \times \mathbb{Z}_m$.

Let us now turn to Hasse's theorem and its consequences. Suppose that $N = |E(\mathbb{F}_q)|$ is divisible by a prime $\ell > 4\sqrt{q}$. Then there exists a unique $c \geq 1$ such that $N = c\ell$ and this c can be easily established. This is because Hasse's Theorem stipulates that

$$q + 1 - 2\sqrt{q} \leq c\ell \leq q + 1 + 2\sqrt{q},$$

and there can be at most one multiple of ℓ in an interval of length $\leq 4\sqrt{q}$.

The existence of a big prime ℓ that divides $N = |E(\mathbb{F}_q)|$ is essential for elliptic curve cryptography. It is called a *factor* and $c = N/\ell$ is known as *cofactor.* There exist methods how to find E with a large factor and a small cofactor. They are based on what is known as *complex multiplication.* The first step is to choose (d, D), where d is square free, $D = d$ if $d \equiv 3 \bmod 4$ and $D = 4d$ otherwise, and to look for x and y such that $x^2 + dy^2 = \ell$. This method has much to do with algebraic number theory, and uses the fact that $-D$ is a discriminant of a primitive positive definite quadratic form. A recommendation is to choose d such that the class number of $\mathbb{Q}(\sqrt{-D})$ is small, but not too small.

Cryptosystems in public use are constructed in such a way that E is a fixed ingredient of the system (or possibly, there may be several options for the choice of E). The choice of E is a substantial part of devising the cryptosystem, and takes into account both the speed of computation and the resilience to possible attacks.

Another application of elliptic curves are factorization algorithms. An important fact that works in their favour is a uniform distribution of $\#E(\mathbb{F}_q)$ in the *Hasse interval*

$$[q + 1 - 2\sqrt{q}, q + 1 + 2\sqrt{q}].$$

If q is a prime, then not only for each N from the Hasse interval there exists a Weierstraß curve with N projective points, but the number of such curves seems to be, by experience, relatively independent of the choice of N.

In the remaining part of this section examples over $\mathbb{Z}_5$ are considered. We shall make an explicit list of all Weierstraß curves, up to $\mathbb{Z}_5$-equivalence, and associate them with Montgomery and twisted Edwards curves. We shall also list points incident to these curves, and describe their group structure.

G.1. **Weierstraß curves over $\mathbb{Z}_5$ and quadratic twists.** For the sake of brevity denote by $W_{a,b}$ a smooth Weierstraß curve over $\mathbb{Z}_5$ given by $y^2 = x^3 + ax + b$. Then condition for smoothness is $4a^3 + 27b^2 \neq 0$, i.e. either $a = b = 0$, or $2b^2a \not\equiv 1 \bmod 5$. The latter is the same as $b^2a \not\equiv 3 \bmod 5$. If $b^2 = 1$, then $a \neq 3$. If $b^2 = 4$, then $a \neq 2$. Hence we are considering (a, b) that **do not** belong to

$$\{(0, 0), (2, 2), (2, 3), (3, 1), (3, 4)\}.$$

Further on it is always assumed that (a, b) are not from such a set.

Let us now address the question when (a, b) and $(\tilde{a}, \tilde{b})$ yield K-equivalent curves. By (M.1) this happens if and only if $\tilde{a} = a$ and $\tilde{b}b$ is a nonzero square. Hence we may restrict our attention only to the case of $b \in \{0, 1, 2\}$.

Suppose for example that $(a, b) = (1, 1)$. By direct computation, the set of affine points is equal to

$$\{(0, 1), (0, 4), (2, 1), (2, 4), (3, 1), (3, 4), (4, 2), (4, 3)\}.$$

This will be recorded as $\{(0, \pm 1), (2, \pm 1), (3, \pm 1), (4, \pm 2)\}$. Note that if $\beta^2 = \alpha^3 + a\alpha + b$, then $(-\beta)^2 = \alpha^3 + a\alpha + b$ as well.

The following table enumerates the affine points of all smooth curves $W_{a,b}$, $b \in \{0,1,2\}$. It also gives the order $N = N_{a,b}$ of the group $W_{a,b}(\mathbb{Z}_5)$ and the parameter $t = 6 - N$. By Hasse's Theorem, $|t| \le [2\sqrt{5}] = 4$.

a	b	Affine points of $W_{a,b}$	N	t
0	1	$(0,\pm1),(2,\pm2),(4,0)$	6	0
0	2	$(2,0),(3,\pm2),(4,\pm1)$	6	0
1	0	$(0,0),(2,0),(3,0)$	4	$+2$
1	1	$(0,\pm1),(2,\pm1),(3,\pm1),(4,\pm2)$	9	-3
1	2	$(1,\pm2),(4,0)$	4	$+2$
2	0	$(0,0)$	2	$+4$
2	1	$(0,\pm1),(1,\pm2),(3,\pm2)$	7	-1
3	0	$(0,0),(1,\pm2),(2,\pm2),(3,\pm1),(4,\pm1)$	10	-4
3	2	$(1,\pm1),(2,\pm1)$	5	$+1$
4	0	$(0,0),(1,0),(2,\pm1),(3,\pm2),(4,0)$	8	-2
4	1	$(0,\pm1),(1,\pm1),(3,0),(4,\pm1)$	8	-2
4	2	$(3,\pm1)$	3	$+3$

Observations:

(1) The Hasse interval is equal to $[2,10]$. For each integer in the interval there exists at least one (a,b) with $N = N_{a,b}$.

(2) If $N_{a,b} \in \{2,3,5,6,7,10\}$, then $W_{a,b}(\mathbb{Z}_5)$ is cyclic since every abelian group of such an order is cyclic.

(3) Since $3 \nmid 4$, the group $W_{1,1}(\mathbb{Z}_5)$ is cyclic as well, by Theorem G.3.

(4) As will be explained, groups $W_{4,0}(\mathbb{Z}_5)$ and $W_{4,1}(\mathbb{Z}_5)$ are not isomorphic. The former group is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_4$, while the latter group to $\mathbb{Z}_8$. Similarly, $W_{1,0}(\mathbb{Z}_5) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ and $W_{1,2}(\mathbb{Z}_5) \cong \mathbb{Z}_4$.

(5) Pairs (a,b) for which $a \ne 0$ and $b \ne 0$ may be grouped by two into $\{(1,1),(4,2)\}$, $\{(2,1),(3,2)\}$, $\{(1,2),(4,1)\}$. The two pairs in each of these sets share the value of ab^2, and that is equal to 1, 2 and 4, respectively. If $\{(a,b),(\tilde{a},\tilde{b})\}$ is one of these sets, then $N_{a,b} + N_{\tilde{a},\tilde{b}} = 12 = 2(q+1)$. In other words, $t_{a,b} = -t_{\tilde{a},\tilde{b}}$.

The explanation of the last phenomenon needs the notion of *j-invariant.* If C is a smooth Weierstraß curve given by $y^2 = x^3 + ax + b$ and $\mathrm{char}(K) \ne 2,3$, then the j-invariant $j(C)$ is defined as $1728\jmath(C)$, where $\jmath(C) = 4a^3/(4a^3 + 27b^2)$.

Note that $\jmath(C) = 0 \Leftrightarrow a = 0$, and $\jmath(C) = 1 \Leftrightarrow b = 0$. Thus $\jmath(C) \in \{0,1\}$ if and only if $ab = 0$.

If $\jmath(C) \notin \{0,1\}$ and $\tilde{C}$ is given by $y^2 = x^3 + \tilde{a}x + \tilde{b}$, then $\jmath(C) = \jmath(\tilde{C})$ if and only if $b^2/a^3 = \tilde{b}^2/\tilde{a}^3$.

If this is true, and the equations $y^2 = x^3 + ax + b$ and $y^2 = x^3 + \tilde{a}x + \tilde{b}$ are **not** K-equivalent, then $\tilde{C}$ is said to be a (quadratic) *twist* of C.

Let K be equal to $\mathbb{R}$ or to $\mathbb{F}_q$, where q is not divisible by 2 or 3. If $\jmath(C) \ne 0,1$, then there exists $\tilde{C}$ that is a quadratic twist of C. If $\tilde{\tilde{C}}$ is another quadratic twist of C, then $\tilde{C}$ and $\tilde{\tilde{C}}$ are defined by K-equivalent Weierstraß equations.

If $K = \mathbb{F}_q$, $2 \nmid q$ and $3 \nmid q$, $\jmath(C) \ne 0,1$ and $\tilde{C}$ is a quadratic twist of C, then

$$|C(\mathbb{F}_q)| + |\tilde{C}(\mathbb{F}_q)| = 2(q+1). \tag{G.3}$$

This confirms the observations above since if $K = \mathbb{Z}_5$ and $ab \ne 0$, then $b^2/a^3 = b^2a$.

Suppose that $\jmath(C) = \jmath(\tilde{C}) \notin \{0,1\}$. To decide whether $\tilde{C}$ is a quadratic twist of C is easy. If $\tilde{b}b$ is a nonsquare, then $\tilde{C}$ is a quadratic twist. If $\tilde{b}b$ is a square, then $\tilde{C}$ and C are defined by K-equivalent Weierstraß equations.

To prove the latter from (M.1) is not difficult. Since $b^2/a^3 = \tilde{b}^2/\tilde{a}^3$ is assumed, we have $\alpha^3 = \beta^2$, where $\alpha = \tilde{a}/a$ and $\beta = \tilde{b}/b$. Put $\gamma = \beta/\alpha$. Then $\gamma^3 = (a^3\tilde{b}^3)/(\tilde{a}^3 b^3) = \tilde{b}/b$ and $\gamma^2 = (a^2\tilde{b}^2)/(\tilde{a}^2 b^2) = \tilde{a}/a$. If $\tilde{b}/b = \gamma^3$ is a square, then γ is also a square. If $\gamma = \lambda^2$, then $\tilde{b} = \lambda^6 b$ and $\tilde{a} = \lambda^4 a$, as required by (M.1).

The relationship between quadratic twists C and $\tilde{C}$ turns into a birational equivalence when the curves are considered over $\mathbb{F}_{q^2}$. Indeed, $\tilde{b}/b$ is always a square in $\mathbb{F}_{q^2}$. Therefore $\mathbb{F}_{q^2}(C) \cong \mathbb{F}_{q^2}(\tilde{C})$ and $C(\mathbb{F}_{q^2}) \cong \tilde{C}(\mathbb{F}_{q^2})$.

If $\tilde{\jmath}(C) \in \{0, 1\}$, then extensions of $\mathbb{F}_q$ offer even more symmetries. This is one of reasons why such curves are usually not considered to be safe for cryptographic purposes.

Let us give a proof of (G.3):

Proof. As explained above, it may be assumed that C is given by $y^2 = x^3 + ax + b$ and $\tilde{C}$ is given by $y^2 = x^3 + \gamma^2 ax + \gamma^3 b$, where $\gamma \in \mathbb{F}_q$ is a nonsquare.

For each $\alpha \in \mathbb{F}_q$ denote by $s(\alpha)$ the number of $\beta \in \mathbb{F}_q$ such that $(\alpha, \beta) \in C$, and by $\tilde{s}(\alpha)$ the number of $\beta \in \mathbb{F}_q$ such that $(\gamma\alpha, \beta) \in \tilde{C}$. Note that

$$|C(\mathbb{F}_q)| = 1 + \sum s(\alpha) \text{ and } |\tilde{C}(\mathbb{F}_q)| = 1 + \sum \tilde{s}(\alpha).$$

To finish it thus suffices to verify that $s(\alpha) + \tilde{s}(\alpha) = 2$ for each $\alpha \in \mathbb{F}_q$. Substituting $x = \gamma\alpha$ into $x^3 + \gamma^2 ax + \gamma^3 b$ yields $\gamma^3(\alpha^3 + b\alpha + c)$. This means that α is a root of $x^3 + ax + b$ if and only if $\gamma\alpha$ is a root of $x^3 + \gamma^2 ax + \gamma^3 b$. In such a case $(\alpha, 0) \in C$, $(\gamma\alpha, 0) \in \tilde{C}$ and $s(\alpha) = \tilde{s}(\alpha) = 1$. If α is not a root, then exactly one of $\alpha^3 + b\alpha + c$ and $\gamma^3(\alpha^3 + b\alpha + c)$ is a (nonzero) square in $\mathbb{F}_q$. This results into $s(\alpha) + \tilde{s}(\alpha) = 2 + 0 = 0 + 2 = 2$. □

G.2. **Tangents and cyclic subgroups.** Let C be a Weierstraß curve over K, and let P be an affine point of $C(K)$. Denote by t_P the (affine) tangent of C at P. To compute $[2]P$ is practically equivalent to finding an intersection of t_P with C. If t_P is parallel to the axis y, then $[2]P = \infty$. If this is not the case and t_P intersects C at no affine point, then $[3]P = \infty$ and $[2]P = \ominus P$. The other only remaining possibility is that t_P intersects C in $Q \neq P$. In such a case t intersects C in no other point, $[2]P = \ominus Q$, and $Q = [-2]P$.

Suppose we compute intersections of tangents with the curve for all K-rational points. Since the opposite element is easy to find, this gives us the value of $[2]P$ for every $P \in C(K)$. Hence we know $[2^k]P$ for each $k \geq 0$. If the set $\{[2^k]P;\ k \geq 1\}$ contains P, then P is of odd order, otherwise it is of even order. If P is of odd order, and $|C(K)| = 2^r\ell$, where ℓ is a prime $\equiv 3, 5 \bmod 8$, then the cyclic group generated by P coincides with the set $\{[2^k]P;\ k \geq 0\}$ since 2 is a primitive element of $\mathbb{Z}_\ell^*$.

Let us illustrate this by computing $[2]P$ for elements of of the curve C given by $y^2 = x^3 + 3x$. The tangent at $P = (\alpha, \beta)$ is given by the equation $\beta y + (\alpha^2 + 1)x + \mu = 0$, where $\mu = -\beta^2 - (\alpha^2 + 1)\alpha$. This is because $\partial(y^2 - x^3 + 2x)/\partial y = 2y$ and $\partial(y^2 - x^3 + 2x)/\partial x = 2(x^2 + 1)$.

If $\lambda y + \nu x + \mu$ gives t_P, then $-\lambda y + \nu x + \mu$ gives $t_{\ominus P}$. This is because $\ominus P = (\alpha, -\beta)$. It is thus needed to compute t_P only in four cases, as shown in the ensuing table. Recall that $P = (\alpha, \beta)$ is an involution if and only if $\beta = 0$. Thus $[2](0, 0) = \infty$, and $I = (0, 0)$ is the only involution of $C(K)$.

P	$\ominus P$	t_P and $t_{\ominus P}$	$[-2]P$	$[2]P$
$(1, 2)$	$(1, 3)$	$\pm y + x + 2$	$(4, 4)$	$(4, 1)$
$(2, 2)$	$(2, 3)$	$\pm y - 2$	$(1, 2)$	$(1, 3)$
$(3, 1)$	$(3, 4)$	$\pm y - 1$	$(4, 1)$	$(4, 4)$
$(4, 1)$	$(4, 4)$	$\pm y + 2x + 1$	$(1, 2)$	$(1, 3)$

Elements $[2]P$ form a subgroup of $W_{3,0}(\mathbb{Z}_5)$. The subgroup consists of ∞, $(1,2)$, $(1,3)$, $(4,1)$ and $(4,4)$. Set $Q = (1,2)$. Then $[2]Q = (4,1)$, $[4]Q = (1,3)$ and $[3]Q = (4,4)$.

Set now $P = (2,3)$. Then $Q = [2]P$, and we thus know each value of $[2m]P$, $m \in \mathbb{Z}$. To get $[2m+1]P$ let us consider an argument of general nature that can be used whenever I is the only involution of $C(K)$ and multiples of Q form a subgroup of $C(K)$ that is of index two and of odd order. If an affine point X is not a multiple of $Q = [2]P$, and $[2]X = [2m]Q$, then $[2](X \ominus [m]Q) = \infty$. This implies $X \ominus [m]Q = I$, and so $X = [m]Q \oplus I = [2m]P \oplus I$. Since multiples of Q form a subgroup of odd order, $2[X]$ can always be expressed as an even multiple of Q.

In our case $I = [5]P$. If, say, $X = (3,1)$, then $[2]X = (4,4) = [3]Q = [8]Q$, and so $(3,1) = [5+8]P = [3]P$. We have

$$P = (2,3),\ [2]P = (1,2),\ [3]P = (3,1),\ [4]P = (4,1),\ [5]P = (0,0),$$
$$[6]P = (4,4),\ [7]P = (3,4),\ [8]P = (1,3)\ \text{ and }\ [9]P = (2,2).$$

This describes the addition on $W_{3,0}$ completely, as $[n]P \oplus [m]P = [n+m]P$ for all $n, m \in \mathbb{Z}$.

Let us now turn to $W_{1,1}$.

P	$\ominus P$	t_P and $t_{\ominus P}$	$[-2]P$	$[2]P$
$(0,1)$	$(0,4)$	$\pm y + 2x - 1$	$(4,3)$	$(4,2)$
$(2,1)$	$(2,4)$	$\pm y + x + 2$	$(2,1)$	$(2,4)$
$(3,1)$	$(3,4)$	$\pm y + x + 1$	$(0,4)$	$(0,1)$
$(4,2)$	$(4,3)$	$\pm y - x + 2$	$(3,1)$	$(3,4)$

Put $P = (0,1)$. Then $[2]P = (4,2)$, $[4]P = (3,4)$, $[8]P = (0,4)$, $[7]P = (4,3)$ and $[5]P = (3,1)$. The value of $[3]P = (0,1) \oplus (4,2)$ has to be computed by means of (A.6) and (A.7).

We have $\lambda = 4/1 = -1$, $1 - 4 = 2$ and $[3]P = (2,1)$. Therefore $[6]P = (2,4)$. This completes the description of $W_{1,1}(\mathbb{Z}_5)$.

If C is a smooth Weierstraß curve given by $y^2 = f(x)$, then the involutions are all elements $(\alpha, 0) \in C$, and $(\alpha, 0) \in C$ if and only if $f(\alpha) = 0$. Both $W_{1,0}(\mathbb{Z}_5)$ and $W_{1,4}(\mathbb{Z}_5)$ contain three involutions. Hence they are isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathbb{Z}_4 \times \mathbb{Z}_2$, respectively. Groups $W_{1,2}(\mathbb{Z}_5)$ and $W_{4,1}(\mathbb{Z}_5)$ contain one involution each. They are cyclic.

When the order of a group is small enough to represent each of its element in computer memory, then computing with the group is easy since it suffices to choose one or two generators, and to express each of the group elements by means of these generators. For large orders this is not a viable way.

G.3. **Montgomery curves over $\mathbb{Z}_5$ and the parameter B.** By Proposition M.5, a Weierstraß equation C given by $y^2 = x^3 + ax + b$ is K-equivalent to a Montgomery curve if and only if there exists $\zeta \in K$ such that (1) $\zeta^3 + a\zeta + b = 0$, i.e. $(\zeta, 0) \in C(K)$, and (2) $f'(\zeta)$ is a nonzero square in K.

Suppose that $K = \mathbb{F}_q$, $\text{char}(K) \neq 2, 3$. If (1) holds, then $|C(K)|$ cannot be a prime > 2 since $|C(K)|$ is even. If both (1) and (2) hold, then $|C(K)|$ is divisible by four—a fact that is not completely obvious, but may be proved with a bit of effort. The ideal situation when the cofactor c is equal to 1 hence cannot occur. This is not the only situation when there is a tradeoff between the efficiency of computation and structural parameters.

Before turning to $\mathbb{Z}_5$ let us make a general observation. If 3 is a nonsquare in K, then (2) cannot hold if $a = 0$. Indeed, in that case $f'(\zeta) = 3\zeta^2$ is a nonsquare or a zero.

As explained above, up to $\mathbb{Z}_5$-equivalence there are 12 smooth Weierstraß curves over $\mathbb{Z}_5$. Eight of them are of even order. These are those for which condition (1) may be fulfilled. To fulfill (2) the curves $W_{a,b}$ with $a = 0$ may be put aside. This leaves us with (A) $(a, b) \in \{(1, 0), (2, 0), (3, 0), (4, 0)\}$ and (B) $(a, b) \in \{(1, 2), (4, 1)\}$. If $b = 0$, then $\zeta = 0$ is always a possibility. In that case $f'(\zeta) = a$ should be a square, and that restricts (A) to $(1, 0)$ and $(4, 0)$. In the former case $\zeta = \pm 2$ needs also be tested. However, $3 \cdot 4 + 1$ is not a square. In the latter case $\zeta = \pm 1$ does not supply a square as well. This means that (A) supplies two possibilities. Another two possibilities come from (B).

Now, $(x + \zeta)^3 + a(x + \zeta) + b = x^3 + 3\zeta x^2 + f'(\zeta)x$ holds over any field K, $\mathrm{char}(K) \neq 2, 3$, cf. the proof of Proposition M.5. That makes $y^2 = x^3 + ax + b$ an equation that is K-equivalent to $y^2 = x^3 + 3\zeta x^2 + B^2 x$, where $B^2 = f'(\zeta)$. Setting $A = 3\zeta/B$ gives the other parameter of the Montgomery curve.

The four cases over $\mathbb{Z}_5$ that were identified above yield the following parameters:

a	b	ζ	3ζ	$f'(\zeta)$	B	A
1	0	0	0	1	1	0
4	0	0	0	4	2	0
1	2	4	2	4	2	1
4	1	3	4	1	1	4

Up to $\mathbb{Z}_5$-equivalence there are thus four smooth Montgomery curves over $\mathbb{Z}_5$:

$$M_{1,0}\colon y^2 = x^3 + x, \qquad M_{4,0}\colon 2y^2 = x^3 + x,$$
$$M_{1,2}\colon 2y^2 = x^3 + x^2 + x, \qquad M_{4,1}\colon y^2 = x^3 + 4x^2 + x.$$

Curves $M_{1,0}$ and $W_{1,0}$ are the same. For the other three curves the affine points are listed below, using the rational mapping $C \to M$, $(\alpha_1, \alpha_2) \mapsto ((\alpha_1 - \zeta)/B, \alpha_2/B^2)$.

$$M_{4,0}\colon (0, 0), (1, \pm 1), (2, 0), (3, 0), (4, \pm 2);$$
$$M_{1,2}\colon (0, 0), (1, \pm 2);$$
$$M_{4,1}\colon (0, 0), (1, \pm 1), (2, \pm 1), (3, \pm 1).$$

The change $(A, B) \mapsto (-A, -B)$ gives $\mathbb{Z}_5$-equivalent equations $4y^2 = x^3 + x$, $3y^2 = x^3 + x$, $3y^2 = x^3 + 4x^2 + x$ and $4y^2 = x^3 + x^2 + x$. This still does not cover all possible parameters for Montgomery curves that may occur over $\mathbb{Z}_5$. We shall now explain how the remaining cases are $\mathbb{Z}_5$-equivalent to the already described cases.

Let us take a more general perspective. Let M be a Montgomery curve given by (A, B) over K. If $\tilde{M}$ is given by $(-A, -B)$, then $(\alpha, \beta) \mapsto (-\alpha, \beta)$ is a K-rational mapping $M \leftrightarrow \tilde{M}$.

If $\lambda \in K^*$ and $\tilde{M}$ is given by $(A, \lambda^2 B)$, then $(\alpha, \beta) \mapsto (\alpha, \lambda\beta)$ is a K-rational mapping $\tilde{M} \to M$.

The latter means that if $\vartheta \in K$ is a nonsquare such that each element of K is equal to λ^2 or $\vartheta\lambda^2$ for some $\lambda \in K$, then each Montgomery curve over K is K-equivalent to a Montgomery curve with parameters $(A, 1)$ or (A, ϑ). This means that if $K = \mathbb{F}_q$, then the following is true:

Proposition G.4. *Let $q > 1$ be a prime power not divisible by 2 and 3.*

- *If $q \equiv 3 \bmod 4$, then each Montgomery curve is $\mathbb{F}_q$-equivalent to a curve given by $y^2 = x^3 + Ax^2 + x$, $A \in \mathbb{F}_q$.*
- *If $q \equiv 1 \bmod 4$ and $\vartheta \in \mathbb{F}_q$ is a nonsquare, then each Montgomery curve is $\mathbb{F}_q$-equivalent to a curve given by $y^2 = x^3 + Ax^2 + x$, or by $\vartheta y^2 = x^3 + Ax^2 + x$, $A \in \mathbb{F}_q$.*

Proof. As explained above, each Montgomery curve is $\mathbb{F}_q$-equivalent to $y^2 = x^3 + Ax^2 + x$ or $\vartheta y^2 = x^3 + Ax^2 + x$. If $q \equiv 3 \bmod 4$, then ϑ may be chosen as

-1. If this is true, then the latter curve is $\mathbb{F}_q$-equivalent to the curve given by $y^2 = x^3 - Ax^2 + x$. □

G.4. **Edwards curves over $\mathbb{Z}_5$.** Consider an Edwards curve $y^2 + x^2 = 1 + dx^2y^2$. If d is a nonsquare, then the addition upon the curve may be described by a uniform (i.e. closed) formula.

There exists a simple criterion that decides whether a smooth Weierstraß curve C over K, $\mathrm{char}(K) \neq 2, 3$ is K-equivalent to an Edwards curve with d a nonsquare. This happens if and only if $C(K) \cong \mathbb{Z}_4 \times H$, where $|H|$ is odd.

This criterion is satisfied over $\mathbb{Z}_5$ if and only if $C = W_{1,2}$ or $C = W_{4,1}$. This matches the fact that 2 and 3 are the only nonsquares modulo 5.

By Theorem E.7, an Edwards curve E with parameter d is birationally equivalent to a Montgomery curve M with parameters $A = 2(1+d)/(1-d)$ and $B = 4/(1-d)$, and the birational mapping $M \to E$ sends (α, β) to $(\alpha/\beta, (\alpha-1)/(\alpha+1))$, assuming $\beta \neq 0$ and $\alpha \neq -1$.

Let E be defined over $\mathbb{Z}_5$, and let $d = 2$. Then $A = 4$ and $B = 1$. Thus $M = M_{4,1}$. This is the reason why E will be denoted by $E_{4,1}$. Similarly, if $d = 3$, then E is denoted by $E_{1,2}$ since it is birationally equivalent to $M_{1,2}$. Both $E_{1,2}$ and $E_{4,1}$ contain points $(0, \pm 1)$ and $(\pm 1, 0)$. Using this fact and the birational mapping described above we get:

$$d = 3 \quad E_{1,2}\colon (0, \pm 1), (\pm 1, 0);$$

$$d = 2 \quad E_{4,1}\colon (0, \pm 1), (\pm 1, 0), (\pm 2, 2), (\pm 2, 3);$$

Suppose now that $d = 4$. Then $A = 0$, $B = 2$. The Edwards curve is hence denoted by $E_{4,0}$. For this curve the completed coordinates have to be used if all $\mathbb{Z}_5$-rational points are to be described by coordinates. Since $((2:1),(1:0))$ fulfils (E.6), the $X \leftrightarrow Y$ symmetry yields the following list of points:

$$d = 4 \quad E_{4,0}\colon ((\pm 1 : 1), (0 : 1)), ((0 : 1), (\pm 1, 1)), ((\pm 2 : 1), (1 : 0)), ((1 : 0), (\pm 2 : 1)).$$

There remains to consider only one class of Montgomery curves over $\mathbb{Z}_5$, and that is the class represented my $M_{1,0} = W_{1,0}$. By Theorem E.7 and Lemma E.6 this curve is birationally equivalent to the twisted Edwards curve with $(a, d) = (2, 3)$. The curve is denoted by $E_{1,0}$. Since we know that $|E_{1,0}(\mathbb{Z}_5)| = 4$, it is easy to verify that it consists of the following points:

$$(a, d) = (2, 3) \quad E_{1,0}\colon ((0 : 1), (\pm 1, 1)), ((1 : 0), (\pm 2 : 1)).$$

Up to now four different twisted Edwards curves have been explicitly described. Of course, that does not exhaust all possible parameters (a, d). However, results of this section allow to find a birational equivalence over $\mathbb{Z}_5$ for each other possible choice. As an example consider the case $(a, d) = (2, 1)$. This is birationally equivalent, by Theorem E.7, to a Montgomery curve with parameters $(A, B) = (1, 4)$, and thus to a Montgomery curve with parameters $(A, B) = (1, 1)$. The equation for this curve is $y^2 = x^3 + x^2 + x$, which is $\mathbb{Z}_5$-equivalent to $y^2 = x^3 - x - 1$, and thus also to $y^2 = x^3 - x + 1$. That is $W_{4,1}$.

While $E_{4,1}$ consists of 8 affine points, the curve given by $2x^2 + y^2 = 1 + x^2y^2$ contains exactly 6 affine points. These are $(\pm 2, \pm 2)$ and $(0, \pm 1)$. The other two points need the completed coordinates. The two points at infinity are $((\pm 1 : 1), (1 : 0))$.

D. Division polynomials

Let us fix a field K of characteristic $p \neq 2, 3$, and let $a, b \in K$ be such that $4a^2 + 27b^2 \neq 0$. Use E to denote the smooth Weierstraß curve given by $y^2 = x^3 + ax + b$. Recall that $E[m]$ denotes the group of all $P \in E$ such that $[m]P = \infty$. This group is a subgroup of $E(\bar{K})$.

If $p \nmid m$, then $|E[m]| = m^2$, by Theorem G.1. There are thus $m^2 - 1$ affine points $P = (\alpha, \beta)$ for which $[m]P = \infty$.

Note that $(\alpha, \beta) \in E[m] \Leftrightarrow (\alpha, -\beta) \in E[m]$. This is because $(\alpha, -\beta) = \ominus P$. Hence, if m is odd and $p \nmid m$, then there are exactly $(m^2 - 1)/2$ different values of α that occur within the affine points $(\alpha, \beta) \in E$ that are of order that divides m.

If m is even, then we have to be a bit more cautious since in this case $E[m]$ contains involutions. There are three of them, and they are equal to $(\zeta_i, 0)$, where $x^3 + ax + b = \prod(x - \zeta_i)$, $1 \leq i \leq 3$. Hence in this case, provided $p \nmid m$, the number of α is exactly $((m^2 - 1) - 3)/2 + 3 = (m^2 + 2)/2$.

It is thus not surprising that there exist polynomials $\tilde{\psi}_m \in K[x]$ of respective orders $(m^2 - 1)/2$ and $(m^2 + 2)/2$ such that $(\alpha, \beta) \in E[m] \Leftrightarrow \tilde{\psi}_m(\alpha) = 0$.

Of course, if $m_1 \mid m_2$, then $E[m_1] \leq E[m_2]$ and $\tilde{\psi}_{m_1}$ divides $\tilde{\psi}_{m_2}$.

Therefore $\tilde{\psi}_2$ divides $\tilde{\psi}_m$ if m is even. A point $(\alpha, \beta) \in E$ is an involution if and only if $\alpha^3 + a\alpha + b = 0$. Hence $\tilde{\psi}_2 = x^3 + ax + b$.

Another criterion for (α, β) being an involution is that $\beta = 0$. This criterion is more easy to check. Because of that (and because of compatibility with the theory of Weierstraß equations in characteristics 2 and 3) it is usual to use polynomials ψ_m that are in defined in variables x and y, and not polynomials $\tilde{\psi}_m \in K[x]$ that are defined only in x. The difference is small. In our case of $y^2 = x^3 + ax + b$, $\mathrm{char}(K) \neq 2, 3$, the polynomial ψ_2 is defined as $2y$. Furthermore, $\psi_m = \tilde{\psi}_m$ if m is odd and $\psi_m = 2y\tilde{\psi}_m/(x^3 + ax + b)$ if m is even.

What is extremely important is the fact that the *division polynomials* ψ_m may be defined recursively, e.g. in the following way:

$$\begin{aligned}
\psi_0 &= 0,\\
\psi_1 &= 1,\\
\psi_2 &= 2y,\\
\psi_3 &= 3x^4 + 6ax^2 + 12bx - a^2,\\
\psi_4 &= 4y(x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3),\\
\psi_{2m+1} &= \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3, \text{ where } m \geq 2, \text{ and}\\
\psi_{2m} &= (\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2)\psi_m/2y, \text{ where } m \geq 3.
\end{aligned} \tag{D.1}$$

However, the definition of ψ_{2m+1} and ψ_{2m} as given above is <u>not</u> correct without a further adjustment. The formula upon the right always yields a polynomial in x and y. In this polynomial there may be occurences of y^i with $i \geq 2$. If this happens then y^i is replaced by $y^{i-2}(x^3 + ax + b)$ until the polynomial contains y in power at most 1. The final polynomial is equal to some $a(x)$ in the case of $2m + 1$, and to $ya(x)$ in the case of $2m$.

Every $P = (\alpha, \beta) \in E$ satisfies

$$[m]P = \infty \iff \psi_m(\alpha, \beta) = 0. \tag{D.2}$$

This is true for all $m \geq 1$, even for those with $p \mid m$. In addition to that the division polynomials can be used to express $[m]P$ for those $P = (\alpha, \beta) \in E$ that do not belong to $E[m]$. If $P \notin E[m]$, $m \geq 2$ and $P \notin E[2]$, then

$$[m]P = \left(\alpha - \frac{\psi_{m-1}\psi_{m+1}}{\psi_m^2}, \ \frac{\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2}{4\beta\psi_m^3}\right). \tag{D.3}$$

The above formula is written compactly, for the sake of clarity. For example the numerator in the former fraction should be read as $\psi_{m-1}(\alpha,\beta)\psi_{m+1}(\alpha,\beta)$.

None of (D.1) and (D.3) is easy to prove. Below we shall verify (D.1) for $m \in \{3,4,5\}$, and (D.3) for $m = 2$.

Instead of polynomials $\tilde{\psi}_m$ it is usual to work with polynomials $\bar{f}_m \in K[x]$. The meaning is nearly the same. The difference is that polynomials $\bar{f}_m$ ignore the involutions. They are defined so that if $P = (\alpha,\beta) \in E$, then

$$P \in E[m] \setminus E[2] \iff \bar{f}_m(\alpha) = 0. \tag{D.4}$$

The connection between $\bar{f}_m$ and ψ_m is such that

$$\bar{f}_m = \begin{cases} \psi_m & \text{if } m \text{ is odd, and} \\ \psi_m/2y & \text{if } m \text{ is even.} \end{cases} \tag{D.5}$$

Thus $\bar{f}_0 = 0$, $\bar{f}_1 = 1$, $\bar{f}_2 = 1$, $\bar{f}_3 = 3x^4 + 6ax^2 + 12bx - a^2$ and $\bar{f}_4 = 2(x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3)$.

For $m \ge 5$ the polynomials $\bar{f}_m$ may be defined recursively. While the formula is straightforwardly derived from (D.1), it looks slightly more complicated. This is because only the variable x is involved.

$$\bar{f}_{2m+1} = \begin{cases} \bar{f}_{m+2}\bar{f}_m^3 - 16(x^3+ax+b)^2\bar{f}_{m-1}\bar{f}_{m+1}^3 & \text{if } m \ge 3 \text{ is odd,} \\ 16(x^3+ax+b)^2\bar{f}_{m+2}\bar{f}_m^3 - \bar{f}_{m-1}\bar{f}_{m+1}^3 & \text{if } m \ge 2 \text{ is even, and} \end{cases}$$
$$\bar{f}_{2m} = \bar{f}_m(\bar{f}_{m+2}\bar{f}_{m-1}^2 - \bar{f}_{m-2}\bar{f}_{m+1}^2) \quad \text{for any } m \ge 3. \tag{D.6}$$

As may be guessed from the formulas above, division polynomials contain many nonzero coefficients of large values. Hence for large q it is not possible to represent them in computer memory if m is very big. Because of that the division polynomials cannot be used, say, to directly verify the order of $E(\mathbb{F}_q)$. Nevertheless this order can be determined by considering the behaviour of polynomials $\bar{f}_m$ where m runs through a set of not too large primes. This is how Schoof's algorithm works.

Note that polynomials $\bar{f}_m$ are not monic. In fact the leading coefficient of $\bar{f}_m$ is equal to m when m is odd, and to $m/2$ when m is even. This is important since when $m = p$ is the characteristic of the field, then $\deg(\bar{f}_m) < (m^2-1)/2$.

D.1. **The division polynomial for order 3.** Let $P = (\alpha,\beta)$ be a point upon E, $\beta \ne 0$. The tangent of E at P can be expressed by the equation $y = \lambda x + \mu$ in which $\lambda = (3\alpha^2 + a)/2\beta$ and $\mu = \beta - \lambda\alpha$. The chord and tangent process, as described in Section A, considers the intersections of the tangent and the curve E.

The first coordinate of such an intersection is a solution to the equation

$$(\lambda x + \mu)^2 = x^3 + ax + b. \tag{D.7}$$

From the logic of the chord and tangent process it follows that α is always a double root of the polynomial

$$x^3 + ax + b - (\lambda x + \mu)^2 = x^3 - \lambda^2 x^2 + (a - 2\lambda\mu)x + b - \mu^2. \tag{D.8}$$

This may also be seen immediately if we write (D.7) in the form

$$(\lambda x + \mu - \beta)^2 = x^3 + ax + b - 2\beta(\lambda x + \mu) + \beta^2$$

and observe that α is a root not only of the polynomials on both sides of this equation, but also of their derivatives.

The point P is of order 3 if and only if the tangent intersects E in no other point of E. This happens if and only if α is the triple root of the polynomial in (D.8). We already know that the multiplicity of α is at least two. The multiplicity is hence equal to three if and only if $\lambda^2 = 3\alpha$. Substituting $\alpha^3 + a\alpha + b$ for β^2 in the denominator of λ^2 turns the equation $\lambda^2 = 3\alpha$ into

$$\begin{aligned}(3\alpha^2 + a)^2 &= 12\alpha(\alpha^3 + a\alpha + b),\\ 9\alpha^4 + 6a\alpha^2 + a^2 &= 12\alpha^4 + 12a\alpha^2 + 12b\alpha \text{ and}\\ 3\alpha^4 + 6a\alpha^2 + 12b\alpha - a^2 &= 0.\end{aligned} \tag{D.9}$$

We have verified the formula for $\psi_3 = \bar{f}_3$. A point $(\alpha, \beta) \in E$ is of order 3 if and only if α is a root of $3x^4 + 6ax^2 + 12bx - a^2$.

Note that in this way we obtain all elements of $E[3]$. Only some of them are K-rational. To get a K-rational point of $E[3]$ the root α has to be from K and $\alpha^3 + a\alpha + b$ has to be a square in K.

D.2. **The division polynomial for order 4.** Suppose that $P = (\alpha, \beta) \in E$ is not an involution. This means that $\beta \neq 0$. In such a case $[4]P = \infty$ if and only if $[2]P = (\alpha', \beta')$ is an involution. This takes place if and only if $\beta' = 0$.

By (A.6) and (A.7), $\beta' = \lambda(\alpha - \alpha') - \beta$, $\alpha' = \lambda^2 - 2\alpha$ and $\lambda = (3\alpha^2 + a)/2\beta$. This gives the following expression of $\beta' = \lambda(\alpha - \alpha') - \beta$:

$$\lambda(3\alpha - \lambda^2) - \beta = (2\beta)^{-3}\left((3\alpha^2 + a)(12\alpha\beta^2 - (3\alpha^2 + a)^2) - 8\beta^4\right). \tag{D.10}$$

If $\beta \neq 0$, then $\beta' = 0$ if and only if $(2\beta)^3\beta' = 0$. In order to express $(2\beta)^3\beta'$ in terms of α, observe that

$$\begin{aligned}12x(x^3 + ax + b) - (3x^2 + a)^2 &= 3x^4 + 6ax^2 + 12bx - a^2,\\ (3x^2 + a)(3x^4 + 6ax^2 + 12bx - a^2) &= 9x^6 + 21ax^4 + 36bx^3 + 3a^2x^2 + 12abx - a^3,\\ \text{and } \; -8(x^3 + ax + b)^2 &= -8x^6 - 16ax^4 - 16bx^3 - 8a^2x^2 - 16abx - 8b^2.\end{aligned}$$

By summing up the latter two rows we obtain that

$$\begin{aligned}&(3x^2 + a)(12x(x^3 + ax + b) - (3x^2 + a)^2) - 8(x^3 + ax + b)^2\\ &\qquad = x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - a^3 - 8b^2 = \bar{f}_4(x)/2.\end{aligned}$$

This verifies that

$$(3\alpha^2 + a)(12\alpha\beta^2 - (3\alpha^2 + a)^2) - 8\beta^4 = \bar{f}_4(\alpha)/2 \text{ for all } (\alpha, \beta) \in E. \tag{D.11}$$

Hence if $(\alpha, \beta) \in E$ and $\beta \neq 0$, then $(2\beta)^3\beta' = 0$ if and only if $\bar{f}_4(\alpha) = 0$.

D.3. **Doubling.** Assume $m = 2$ and suppose that $P = (\alpha, \beta) \in E$ is not an involution. By (D.1), $\psi_{m-1}(\alpha, \beta) = 1$, $\psi_m^2(\alpha, \beta) = 4\beta^2$ and $\psi_{m+1}(\alpha, \beta) = 3\alpha^4 + 6\alpha^2 + 12b\alpha - a^2$.

By (D.9) the latter is equal to $12\alpha\beta^2 - (3\alpha^2 + a)^2$. Set $\lambda = (3\alpha^2 + a)/2\beta$. We have

$$\alpha - \left(\frac{\psi_1\psi_3}{\psi_2^2}\right)(\alpha, \beta) = \alpha - 12\alpha/4 + \lambda^2 = \lambda^2 - 2\alpha.$$

This verifies that if $m = 2$, then the first coordinate of (D.3) corresponds to the doubling formula (A.6) and (A.7).

By these formulas the second coordinate of $[2]P$ is equal to $\lambda(3\alpha - \lambda^2) - \beta$, and that can be expressed, by (D.10) and (D.11), as $(2\beta)^{-3}\bar{f}_4(\alpha)/2$. This agrees with formula (D.3) since for $m = 2$ the second coordinate at the right hand side of (D.3) is equal to

$$\psi_4(\alpha, \beta)/4\beta\psi_2^3(\alpha, \beta) = 2\beta\bar{f}_4(\alpha)/4\beta(2\beta)^3 = \bar{f}_4(\alpha)/16\beta^3.$$

D.4. **Order and characteristic 5.** As already mentioned, verifying formulas (D.1) and (D.3) in their generality is technically demanding. Here it will not be performed. However, we shall illustrate upon the case of $m = 5$ why ψ_m has much smaller number of roots when $\mathrm{char}(K)$ divides m.

What we shall do first is to use (D.6) to get the general formula for $\bar{f}_5$, and then we shall observe how dramatically $\bar{f}_5$ changes when it is considered in characteristic 5. By (D.6),

$$\bar{f}_5 = 16(x^3 + ax + b)^2 \bar{f}_4 \bar{f}_2^3 - \bar{f}_1 \bar{f}_3^3 = 16(x^3 + ax + b)^2 \bar{f}_4 - \bar{f}_3^3.$$

$$\text{Since } (x^3 + ax + b)^2 = x^6 + 2ax^4 + 2bx^3 + a^2x^2 + 2abx + b^2$$

$$\text{and } \bar{f}_4/2 = x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3$$

we may express $(x^3 + ax + b)^2 \bar{f}_4/2$ as

$$x^{12} + 7ax^{10} + 22bx^9 + 6a^2x^8 + 48abx^7 + (33b^2 - 6a^3)x^6 + 12a^2bx^5 + (21ab^2 - 7a^4)x^4 \\ + (4b^3 - 16a^3b)x^3 - (21b^2a^2 + a^5)x^2 - (20ab^3 + 2a^4b)x - 8b^4 - a^3b^2,$$

while $\bar{f}_3^3 = (3x^4 + 6ax^2 + 12bx - a^2)^3$ is equal to

$$27x^{12} + 162ax^{10} + 324bx^9 + 297a^2x^8 + 1296abx^7 + (108a^3 + 1296b^2)x^6 + 1080a^2bx^5 \\ + (2592ab^2 - 99a^4)x^4 + (1728b^3 - 432a^3b)x^3 - (432a^2b^2 - 18a^5)x^2 + 36a^4bx - a^6.$$

Therefore $\bar{f}_5 = 16(x^3 + ax + b)^2 \bar{f}_4 - \bar{f}_3^3$ is equal to

$$5x^{12} + 62ax^{10} + 380bx^9 - 105a^2x^8 + 240abx^7 - (240b^2 + 300a^3)x^6 \\ - 696a^2bx^5 - (1920ab^2 + 125a^4)x^4 - (1600b^3 + 80a^3b)x^3 - (240b^2a^2 + 50a^5)x^2 \\ - (640ab^3 + 100a^4b)x - (256b^4 + 32a^3b^2 - a^6).$$

Modulo 5 this yields $2ax^{10} - a^2bx^5 - b^4 - 2a^3b^2 + a^6$. Let $r, s, t \in \bar{K}$ be such that $r^5 = 2a$, $s^5 = -a^2b$ and $t^5 = -b^4 - 2a^3b^2 + a^6$. If K is assumed, as usual, to be a perfect field, then $r, s, t \in K$.

We see now that if $\mathrm{char}(K) = 5$, then $\bar{f}_5(x) = (rx^2 + sx + t)^5$. This implies $|E[5]| = 5$, provided $a \neq 0$. If $a = 0$, then $E[5]$ is a trivial group.

I. Ingredients of Schoof's algorithm and its main idea

Let E be a projective elliptic curve over $\mathbb{F}_q$. By Hasse's theorem, $|E(\mathbb{F}_q)| = q - t + 1$, where $|t| \le 2\sqrt{q}$. A related fact states that

$$\varphi^2 \ominus [t]\varphi \oplus [q] = \mathcal{O}, \tag{I.1}$$

where φ stands for the *Frobenius endomorphism* of E.

To explain the meaning of (I.1) let us start with the meaning of φ. If $P = (\alpha_1 : \alpha_2 : \alpha_3) \in E$, then $\varphi(P) = (\alpha_1^q : \alpha_2^q : \alpha_3^q) \in E$ too. To see this consider the equation, say $w(X_1, X_2, X_3) = 0$, that determines E. If $w(\alpha_1, \alpha_2, \alpha_3) = 0$, then $0 = \big(w(\alpha_1, \alpha_2, \alpha_3)\big)^q = w(\alpha_1^q, \alpha_2^q, \alpha_3^q)$. For example if E is given by a smooth Weierstraß curve $y^2 = x^3 + ax + b$ and $P = (\alpha, \beta) \in E$, then $\varphi(P) = (\alpha^q, \beta^q)$. Indeed $\beta^{2q} = (\beta^2)^q$ is equal to $\alpha^{3q} + a\alpha^q + b = (\alpha^3 + a\alpha + b)^q$, as $a^q = a$ and $b^q = b$.

The Frobenius endomorphism φ sends points of E upon the points of E. Equation (I.1) implicitly uses the fact that φ is also an endomorphism of the group $E(\bar{\mathbb{F}}_q)$, i.e. that $\varphi(P \oplus Q) = \varphi(P) \oplus \varphi(Q)$ for all $P, Q \in E$. This can be proved from the addition formulas. However, this is also a consequence of a more general fact that is explained below when introducing the notion of *isogeny*.

Equation (I.1) thus means that if three endomorphisms of $E(\bar{\mathbb{F}}_q)$, i.e., $P \mapsto \varphi^2(P)$, $P \mapsto [-t](\varphi(P))$ and $P \mapsto [q]P$, are summed up, then the result is the trivial endomorphism $P \mapsto \mathcal{O}$. This can also be expressed as

$$\varphi^2(P) \ominus [t]\varphi(P) \oplus [q]P = \mathcal{O} \text{ for every } P \in E. \tag{I.2}$$

In fact, the latter form occurs in literature more often than (I.1). However, it may be argued that the expression via (I.1) is more instructive since it conveys better the fact that we are dealing with a property of the group $E(\bar{\mathbb{F}}_q)$. This is important since the structure of the group does not change under birational equivalence.

It is usual to call $T^2 - tT + q$ the *characteristic polynomial of the Frobenius endomorphism* and t the *trace of the Frobenius endomorphism*. Here T stands for a variable and carries no specific meaning. Reasons for calling t a 'trace' will be explained at the end of this section.

If P is a $\mathbb{F}_q$-rational point of E, then $\varphi(P) = P$. In such a case (I.2) states that $[P] \ominus [t]P \oplus [q]P = [q-t+1]P$ is equal to $\mathcal{O}$. This is true because $P \in E(\mathbb{F}_q) \le E(\bar{\mathbb{F}}_q)$ and $|E(\mathbb{F}_q)| = q - t + 1$.

I.1. **Isogenies.** To understand Schoof's algorithm it is not completely necessary to absorb the content of this subsection. Its purpose is to set the endomorphisms occurring in (I.1) into a broader context. It explains the notion of morphism and the notion of isogeny, and states some of the basic properties that morphisms and isogenies fulfil. Morphisms and isogenies belong to central notions of elliptic curves theory, and are used in quite a few algorithms.

How to transfer the notion of a rational map to projective curves, say C and D? This question can be answered in several ways. Here we shall discuss, for the sake of simplicity, only the situation when both C and D are smooth. In that case every rational map from an affine part of C to an affine part of D may be extended to a *morphism* $C \to D$.

Suppose that $C = V_F$ and $D = V_G$. A morphism $\psi \colon C \to D$ is *represented* by $A = (A_1 : A_2 : A_3)$ if *the polynomials* $A_1, A_2, A_3 \in K[X_1, X_2, X_3]$ *are homogeneous and of the same degree and, with only finitely many exceptions, for each* $\alpha = (\alpha_1 : \alpha_2 : \alpha_3) \in C$ *at least one of* $A_1(\alpha)$, $A_2(\alpha)$ *and* $A_3(\alpha)$ *is nonzero, and* $(A_1(\alpha) : A_2(\alpha) : A_3(\alpha)) = \psi(\alpha) \in D$.

Triples $(A_1 : A_2 : A_3)$ and $(B_1 : B_2 : B_3)$ represent the same morphism if $A_iB_j - A_jB_i \in (F)$ whenever $1 \le i < j \le 3$. It can be proved that if $\alpha \in C$, and if $\psi \colon C \to D$ is a morphism, then there exists $(A_1 : A_2 : A_3)$ representing ψ such that at least one of $A_i(\alpha)$ is not zero. This means that a **morphism** $\psi\colon C \to D$ **is defined everywhere**. This is the main theoretical advantage of morphisms when compared to rational maps.

Any constant mapping $C \to D$ is a morphism. Because of that (and for other reasons too) it is useful, while not necessary, to allow in the definition of morphism that one or two of A_is are zero polynomials.

If C is an elliptic curve over K, then any K-rational point of C may be chosen as the zero element $\mathcal{O}$ of the group $C(K)$. In fact, $C(K)$ is completely determined by C and the choice of $\mathcal{O}$. This is why some authors define an elliptic curve as a pair $(C, \mathcal{O})$. Here it is assumed that $\mathcal{O}$ is known from the context. By context we understand, e.g., the convention that $\mathcal{O} = \infty$ for a Weierstraß curve, and $\mathcal{O} = (0,1)$ for a (twisted) Edwards curve. (Of course, choosing a different neutral element induces different addition formulas.)

Let C and D be smooth elliptic curves over K, and let $\mathcal{O}_C$ and $\mathcal{O}_D$ be the neutral elements. An *isogeny* $C \to D$ is any morphism $C \to D$ that sends $\mathcal{O}_C$ upon $\mathcal{O}_D$. It can be proved (and the proof is not completely easy) that **each isogeny is also a group homomorphism** $C(K) \to D(K)$. A related result states that **if** ψ_1 **and** ψ_2 **are isogenies** $C \to D$**, then** $\psi_1 \oplus \psi_2$ **is also an isogeny** $C \to D$. (The mapping $\psi_1 \oplus \psi_2$ sends a point $P \in C$ to $\psi_1(P) \oplus \psi_2(P) \in D$, the addition being performed in $D(\bar{K})$.) Note that if $n > 0$, then the mapping $P \mapsto [n]P$ can be expressed as $\mathrm{id}_C \oplus \cdots \oplus \mathrm{id}_C$, where id_C occurs n times. To prove that $P \mapsto [n]P$ is an isogeny thus does not require knowledge of formula (D.3).

An *endomorphism* of C is an isogeny $C \to C$. This is seemingly inconsistent with usual conventions since here an endomorphism of C is something different than a morphism $C \to C$. As an example of the latter take a point $Q \in C$. The *translation* $t_Q\colon P \mapsto P \oplus Q$ is a morphism $C \to C$, but not an endomorphism (unless $Q = \mathcal{O}$) since it maps $\mathcal{O}$ upon Q.

Without going into details let us justify the convention that an endomorphism of C has to be an isogeny by saying that endomorphisms of C are, in fact, assumed to be endomorphisms of $(C, \mathcal{O})$.

All endomorphisms of C form a ring. The ring is denoted by $\mathrm{End}(C)$. This ring contains a subring that is isomorphic to $\mathbb{Z}$ and consists of all mappings $[n]\colon P \to [n]P$. If $K = \mathbb{F}_q$, then $\mathrm{End}(C)$ also contains the Frobenius endomorphism φ.

As an example how to express a rational map (ρ_1, ρ_2) as a morphism represented by $(A_1 : A_2 : A_3)$ let us consider the doubling upon a smooth Weierstraß curve C given by $y^2 = x^3 + ax + b$. The strategy is always the same. Replace $r_i/s_i = r_i(x_1, x_2)/s_i(x_1, x_2)$ that represents ρ_i by $R_i(X_1, X_2, X_3)/S_i(X_1, X_2, X_3)$, where $\deg(R_i) = \deg(S_i)$, $\gcd(R_i, S_i) = 1$ and $R_i(X_1, X_2, 1)/S_i(X_1, X_2, 1) = r_i/s_i$, and then replace $(R_1/S_1 : R_2/S_2 : 1)$ by $(R_1S/S_1 : R_2S/S_2 : S) = (A_1 : A_2 : A_3)$, where $S = \mathrm{lcm}(S_1, S_2)$.

In our example we may proceed similarly as when expressing the doubling in projective coordinates, as done at the end of Section A. We have

$$\begin{aligned}
\frac{r_1(x_1,x_2)}{s_1(x_1,x_2)} &= \frac{(3x_1^2+a)^2 - 8x_1x_2^2}{4x_2^2}, \\
\frac{r_2(x_1,x_2)}{s_2(x_1,x_2)} &= \frac{(3x_1^2+a)(12x_1x_2^2 - (3x_1^2+a)^2) - 8x_2^4}{8x_2^3}, \\
R_1(X_1,X_2,X_3) &= (3X_1^2 + aX_3^2)^2 - 8X_1X_2^2X_3, \\
S_1(X_1,X_2,X_3) &= 4X_2^2X_3^2, \\
R_2(X_1,X_2,X_3) &= (3X_1^2 + aX_3^2)(12X_1X_2^2X_3 - (3X_1^2 + aX_3^2)^2) - 8X_2^4X_3^2, \text{ and} \\
S_2(X_1,X_2,X_3) &= 8X_2^3X_3^3 = S(X_1,X_2,X_3).
\end{aligned}$$

This shows that the morphism $P \mapsto [2]P$ may be represented by $(A_1 : A_2 : A_3) = (2X_2X_3R_1(X_1, X_2, X_3) : R_2(X_1, X_2, X_3) : 8X_2^3X_3^3)$. Unlike the rational maps, morphisms are defined everywhere. To illustrate this assume that $P = (\alpha, \beta) = (\alpha : \beta : 1)$ is an involution. This means that $\beta = 0$. In such a case $(A_1 : A_2 : A_3)$ sends P upon $(0 : -(3\alpha^2 + a)^3 : 0) = (0 : 1 : 0) = \infty$, as expected. (Recall that $3\alpha^2 + a \neq 0$ since α is a simple root of $x^3 + ax + b$.)

I.2. **The idea of Schoof's algorithm.** Schoof's algorithm counts the number of $\mathbb{F}_q$-rational points upon an elliptic curve E. It will be assumed that E is given by $y^2 = x^3 + ax + b$ and that q is divisible by neither 2 nor 3.

While we shall be concerned only with Weierstraß curves, the general framework of Schoof's algorithm is clearly applicable to other forms of elliptic curves. Nevertheless, details of the algorithm are tightly bounded with the specific properties of Weierstraß curves. The algorithm may be adapted to normal forms in characteristics 2 and 3. However, the case of $y^2 = x^3 + ax + b$ is technically the least complicated.

Recall that the order of $E(K)$ does not change under a birational equivalence. Hence there is always a possibility of finding a Weierstraß curve that is birationally equivalent to a given curve E.

The complexity of Schoof's algorithm is $O(\log^8 q)$ bit operations. This is an upper estimate that has been confirmed by practical experience. Theoretical complexity that uses different estimates for the complexity of multiplication is somewhat lower.

More advanced counting algorithms by Elkies and Atkins develop Schoof's ideas further on. A complete understanding of the Schoof-Elkies-Atkins algorithm (the SEA algorithm) requires knowledge of *modular polynomials.*

We shall now give an overall description of Schoof's algorithm.

Denote by t the trace of the Frobenius endomorphism. By Hasse's theorem, $|t| \leq 2\sqrt{q}$. If $\ell_1 < \cdots < \ell_r$ are primes such that $\prod \ell_i > 4\sqrt{q}$ and $t \bmod \ell_i$ is known for each $i \in \{1, \ldots, r\}$, then the Chinese Remainder Theorem determines t uniquely.

Primes $\ell_1, \ldots, \ell_r$ are taken to be the first r primes for which $\prod \ell_i$ is big enough. The main part of Schoof's algorithm thus is to determine $t_\ell = t \bmod \ell$, where ℓ is a prime that is significantly smaller than q.

If $\ell = 2$, then $t_\ell = 0$ when $E(K)$ contains an involution, and $t_\ell = 1$ otherwise. Thus $t_2 = 1$ if and only if the polynomial $x^3 + ax + b$ is irreducible in $K[x]$. Note that the latter happens if and only if $x^3 + ax + b$ is coprime to $x^q - x$.

For the rest we may thus assume that ℓ is an odd prime.

Let us denote by $E[\ell]^*$ the nonzero elements of $E[\ell]$. Hence each $P \in E[\ell]^*$ is of order ℓ. Each such P fulfils (I.2). Since $[\ell]P = \mathcal{O}$, we have, in fact,

$$\varphi^2(P) \oplus [q_\ell]P = [t_\ell]\varphi(P), \text{ where } q_\ell = q \bmod \ell. \tag{I.3}$$

This holds for every $P \in E[\ell]^*$. Hence if we find $\tau \in \{0, 1, \ldots, \ell - 1\}$ such that for *some* $P \in E[\ell]^*$

$$\varphi^2(P) \oplus [q_\ell]P = [\tau]\varphi(P),$$

then there must be $\tau = t_\ell$. The algorithm proceeds by taking values of $\tau = 0, 1, \ldots, (\ell-1)/2$ one after another. For each such τ the algorithm tests the existence of $P \in E[\ell]^*$ such that

$$\varphi^2(P) \oplus [q_\ell]P = [\pm\tau]\varphi(P) \tag{I.4}$$

until it succeeds.

Imagine for a while that all points $P = (\alpha, \beta) \in E$ fulfilling (I.4) were at our disposal. In such a case the obvious step to do would be to test whether some of them belongs to $E[\ell]$. Of course, $P \in E[\ell]$ if and only if $\psi_\ell(\alpha, \beta) = 0$, where ψ_ℓ is the ℓth division polynomial.

However, the algorithm does not run by finding all points P that fulfil (I.4). That would be difficult to achieve. What the algorithm does is to look for properties that such a point P has to fulfil, and to refute the incorrect values of τ when such a property is not fulfilled.

Suppose for a while that τ is fixed and that $\tau > 0$. Let us compare symbolically the first coordinate of $\varphi^2(P) \oplus [q_\ell]P$ (i.e., the x-coordinate) with the first coordinate of $[\tau]P$. It turns out that there exists a polynomial $h_X = h_{X,\tau} \in \mathbb{F}_q[x]$ such that **a point** $P = (\alpha, \beta) \in E$ **fulfils (I.4) if and only if** $h_X(\alpha) = 0$. To be exact, the "if and only if" relationship holds only for those P that do not belong to $E[q_\ell]^*$ or $E[\pm\tau]^*$. These exceptions cause no problem since the goal is to decide whether such a P can be found in $E[\ell]^*$. This is true if and only if $\gcd(\bar{f}_\ell, h_X) \neq 1$.

Suppose thus that $\bar{f}_\ell$ and h_X have a common root, say α. This means that $t_\ell \in \{-\tau, \tau\}$, and that there exists $\beta \in \bar{\mathbb{F}}_q$ such that $P = (\alpha, \beta)$ belongs to $E[\ell]^*$ and the point $\varphi^2(P) \oplus [q_\ell]P$ shares the first coordinate with $[\tau](\varphi(P))$. If these two points share also the second coordinate, then they are equal. In such a case $t_\ell = \tau$. If the points to not agree then $t_\ell = -\tau$. Hence the second coordinates either agree for all $P \in E[\ell]^*$, or for none $P \in E[\ell]^*$.

It turns out that if the second coordinates are compared, then the value of β may be cancelled out. Therefore there exists a polynomial h_Y such that $h_Y(\alpha) = 0$ if and only if the second coordinates agree, for any $P = (\alpha, \beta) \in E[\ell]^*$. If h_Y and $\bar{f}_\ell$ have a nontrivial common divisor, then $t_\ell = \tau$. Otherwise $t_\ell = -\tau$.

The construction of polynomials h_X and h_Y can be regarded as the computational core of Schoof's algorithm.

Because we are interested only in $\gcd(h_X, \bar{f}_\ell)$, the polynomial h_X may be actually computed modulo $\bar{f}_\ell$ all the time. This reduces the computational complexity. The degree of $\bar{f}_\ell$ is $\leq (\ell^2 - 1)/2$. The same reduction may be done for h_Y and other polynomials.

Polynomials h_X and h_Y are not computed when $t_\ell = 0$, and also in some other cases. What exactly are these exceptional cases and how they are handled is explained below.

While points $P = (\alpha, \beta) \in E[\ell]$ are considered throughout the description of the algorithm, neither α nor β is ever explicitly computed. All needed tests are turned into a polynomial form that involves α only, and we are asking if such a polynomial has a root in $E[\ell]^*$. Since any $P \in E \setminus E[2]$ belongs to $E[\ell]$ if and only if $\bar{f}_\ell(\alpha) = 0$, such a test may be performed by testing whether the polynomial and $\bar{f}_\ell$ possess a nontrivial common divisor.

I.3. **When the first coordinates coincide.** When starting to process an odd prime ℓ, the first step to be performed is to add $\varphi^2(P)$ and $[q_\ell]P$ under the assumption that $P \in E[\ell]^*$. But which formula to use? To decide that, the algorithm finds out whether there exists $P \in E[\ell]^*$ such that $\varphi^2(P) = [\pm q_\ell]P$. If $P = (\alpha, \beta) \in E \setminus E[q_\ell]$, then $[q_\ell](P)$ can be expressed by means of (D.3). Since $\varphi^2(P) = (\alpha^{q^2}, \beta^{q^2})$ it is easy to see that the first coordinates of both $\varphi^2(P)$ and $[q_\ell](P)$ depend only upon α. This yields a polynomial $\bar{s}_\ell \in \mathbb{F}_q[x]$ such that the first coordinates agree if and only if $\bar{s}_\ell(\alpha) = 0$. The existence of $P \in E[\ell]^*$ with $\varphi^2(P) = [\pm q_\ell]P$ is thus equivalent to $\gcd(\bar{s}_\ell, \bar{f}_\ell) \neq 1$. Let the latter be true.

Thus either $\varphi^2(P) = [q_\ell]P$ or $\varphi^2(P) = [-q_\ell]P$. In the latter case $t_\ell = 0$. To test whether $t_\ell = 0$ compare the second variables of $\varphi^2(P)$ and $[-q_\ell]P$. It turns out that by using $\beta^2 = \alpha^3 + a\alpha + b$ the value of β can be cancelled out from such an equation, and we get a polynomial in x. Now, $t_\ell = 0$ if and only if α is the root of this polynomial for each $(\alpha, \beta) \in E[\ell]^*$, and that takes place if and only if this polynomial is a multiple of $\bar{f}_\ell$.

If the polynomial is coprime to $\bar{f}_\ell$, then $\varphi^2(P) = [q_\ell]P$ for some (but necessarily for all) $P \in E[\ell]^*$. This is a special case which differs from the cases considered above. Historically it is important since this has been the departing point for Elkies improvements.

The equality $\varphi^2(P) = [q_\ell]P$ does not yield immediately the value of t_ℓ. Replacing $\varphi^2(P)$ with $[q_\ell]P$ in (I.3) gives $[2q_\ell]P = [t_\ell]\varphi(P)$. Thus $\varphi(P) = [2q_\ell/t_\ell]P$ (the fraction is evaluated modulo ℓ) and

$$[q_\ell]P = \varphi^2(P) = \varphi([2q_\ell/t_\ell]P) = [(2q_\ell/t_\ell)^2]P.$$

Therefore $[t_\ell^2]P = [4q_\ell]P$ and $t_\ell^2 \equiv 4q_\ell \bmod \ell$. This gives two possible values for t_ℓ. Denote one of them by τ. We are asking whether $[2q_\ell]P = [\tau]\varphi(P)$ for some $P \in E[\ell]^*$. This can be written as $\varphi(P) = [\gamma]P$, where $\gamma = 2q_\ell/\tau$. A test for that can be devised similarly as the tests described earlier. If no such P exists, then $t_\ell = -\tau$.

I.4. **Comments on the SEA algorithm.** In Schoof's algorithm, when there is computed the gcd of a polynomial and $\bar{f}_\ell$, the polynomial is in most cases either coprime to $\bar{f}_\ell$ or a multiple of $\bar{f}_\ell$. This is because the equation (I.3) either holds for all $P \in E[\ell]^*$, or for none $P \in E[\ell]^*$. However the equation $\varphi^2(P) = [q_\ell]P$ may hold only for some $P \in E[\ell]^*$, and not for all of them. What is behind this phenomenon?

We have $E[\ell] \cong \mathbb{Z}_\ell \times \mathbb{Z}_\ell$. This means that $(E[\ell], \oplus)$ can be regarded as a vector space of dimension 2 over $\mathbb{Z}_\ell$. The Frobenius endomorphism when restricted to this vector space is a linear automorphism, i.e., a linear transformation with trivial kernel. Denote this restriction by ψ. By Cayley-Hamilton Theorem, $\psi^2 - \mathrm{tr}(\psi)\psi + \det(\psi) = 0$. It is now clear why t is called the *trace of Frobenius endomorphism.*

The polynomial $T^2 - t_\ell T + q_\ell$ may have a root in $\mathbb{Z}_\ell$. If it does have a root, then ℓ is called an *Elkies prime.* If the polynomial is irreducible over $\mathbb{Z}_\ell$, then ℓ is called an *Atkin prime.*

Assume that ℓ is an Elkies prime. Then ψ possesses one or two eigenvalues. If λ is such an eigenvalue, then there exists $P \in E[\ell]^*$ such that $\varphi(P) = [\lambda]P$. We have encountered such a situation above, with $\lambda = 2q_\ell/t_\ell$. That is a special case. In the SEA algorithm an eigenvalue λ is determined for each Elkies prime ℓ.

Since we do not know t_ℓ in advance we also do not know in advance whether ℓ is an Elkies or Atkin prime. However, there exist methods using modular polynomials that allow to establish this without actually computing t_ℓ. Furthermore there exist methods involving modular polynomials and curves isogenous to E that allow, for each Elkies prime, to perform the testing for λ more efficiently. Once λ is known, we can use the existence of $P \in E[\ell]^*$ with $\varphi(P) = [\lambda]P$ to express (I.3) as $[\lambda^2](P) \oplus [q_\ell]P = [t_\ell\lambda]P$, which implies that $t_\ell = \lambda + q_\ell/\lambda$ (the fraction and the addition is evaluated modulo ℓ).

Another ingredient of the SEA algorithm is a method how to obtain, in case of an Atkin prime, a relatively small set T_ℓ such that t_ℓ has to belong to T_ℓ.

S. Schoof's algorithm

Let it be assumed that q is a prime power not divisible by 2 and 3, and that $a, b \in \mathbb{F}_q$ are such that $y^2 = x^3 + ax + b$ determines a smooth Weierstraß curve E. Polynomials h_X, h_Y and $\bar{s}_\ell$ are assumed to have the same meaning as in Section I. Here we shall explain how exactly they are computed.

Any polynomial in one variable that is computed in Schoof's algorithm may be immediately reduced modulo $\bar{f}_\ell$, where ℓ is the prime that is being processed. This fact is not being reflected in the ensuing description of Schoof's algorithm.

The description contains declarations of only those variables and procedures the meaning of which is not clear from the context. It skips declarations of procedures `equalx`, `nonequalx`, `tyzero` and `eigen` that are explained separately.

Procedure `equalx` is called when $\varphi^2(P)$ and $[q_\ell]P$ agree in the first variable for some $P \in E[\ell]^*$, while `noequalx` is used when $E[\ell]^*$ carries no such P.

```
Schoof's algorithm:
INPUT: $q$, $a$ and $b$ that determine a Weierstraß curve $E$.
OUTPUT: The order of $E(\mathbb{F}_q)$.

VARIABLES: $B$ is the product of primes.
           $M$ is the set of $(\ell, t_\ell)$.
           $r$ is the return value from nonequalx.

$B = 2$;
$\ell = 2$;
if $(\gcd(x^q - x, x^3 + ax + b) = 1)$ then $\tau = 1$ else $\tau = 0$;
$M = \{(2, \tau)\}$;
while $(B < 4\sqrt{q})$ do:
     $\ell =$ nextprime($\ell$);
     $B = B * \ell$;
     if (gcd($\bar{s}_\ell, \bar{f}_\ell$) $\neq 1$)
         then $\tau =$ equalx($\ell$)
         else do:
             $\tau = 0$;
             do:
                 $\tau = \tau + 1$;
                 $r =$ nonequalx($\ell, \tau$);
             until $(r \neq 0)$;
             if $(r = -1)$ then $\tau = -\tau$;
     $M = M \cup \{(\ell, \tau)\}$;
Recover $t$ using the set $M$ and the CRT.
Return $q+1-t$.
```

Suppose that $m \geq 2$ and that $P = (\alpha, \beta) \in E$. By (D.3) the first coordinate of $[m]P$ is equal to $\alpha - (\psi_{m-1}\psi_{m+1}\psi_m^{-2})(\alpha, \beta)$. Using the transformation of (D.5) this yields $\alpha - \bar{f}_{m-1}(\alpha)\bar{f}_{m+1}(\alpha)/4\beta^2\bar{f}_m^2(\alpha)$ if m is even, while for m odd we get $\alpha - \bar{f}_{m-1}(\alpha)\bar{f}_{m+1}(\alpha)4\beta^2/\bar{f}_m^2(\alpha)$. Therefore the first coordinate of $[m]P$, $m \geq 2$, is equal to

$$\begin{aligned} &\alpha - \frac{\bar{f}_{m-1}(\alpha)\bar{f}_{m+1}(\alpha)}{4(\alpha^3 + a\alpha + b)\bar{f}_m^2(\alpha)} && \text{if } m \text{ is even, and} \\ &\alpha - \frac{4(\alpha^3 + a\alpha + b)\bar{f}_{m-1}(\alpha)\bar{f}_{m+1}(\alpha)}{\bar{f}_m^2(\alpha)} && \text{if } m \text{ is odd.} \end{aligned} \tag{S.1}$$

Thus $\bar{s}_\ell(x) = x^{q^2} - x$ if $q_\ell = 1$,

$$\bar{s}_\ell(x) = 4(x^{q^2}-x)(x^3+ax+b)\bar{f}^2_{q_\ell}(x) + \bar{f}_{q_\ell-1}(x)\bar{f}_{q_\ell+1}(x) \text{ if } q_\ell \text{ is even, and}$$

$$\bar{s}_\ell(x) = (x^{q^2}-x)\bar{f}^2_{q_\ell}(x) + 4(x^3+ax+b)\bar{f}_{q_\ell-1}(x)\bar{f}_{q_\ell+1}(x) \text{ if } q_\ell > 1 \text{ is odd.}$$

From (D.3) there also may be derived a formula for the second coordinate of $[m]P$, $m \geq 2$:

$$\begin{aligned} &\beta\frac{\bar{f}_{m+2}(\alpha)\bar{f}^2_{m-1}(\alpha) - \bar{f}_{m-2}(\alpha)\bar{f}^2_{m+1}(\alpha)}{16(\alpha^3+a\alpha+b)^2\bar{f}^3_m(\alpha)} && \text{if } m \text{ is even, and} \\ &\beta\frac{\bar{f}_{m+2}(\alpha)\bar{f}^2_{m-1}(\alpha) - \bar{f}_{m-2}(\alpha)\bar{f}^2_{m+1}(\alpha)}{\bar{f}^3_m(\alpha)} && \text{if } m \text{ is odd.} \end{aligned} \tag{S.2}$$

The procedure `equalx` calls as a subprocedure the procedure $\texttt{tyzero}(\ell, m)$ with parameter m equal to q_ℓ. This procedure returns `TRUE` if there exists $P = (\alpha, \beta) \in E[\ell]^*$ such that $\varphi^2(P) = [-m]P$, under the assumption that there exists $P \in E[\ell]^*$ for which the first coordinates of $\varphi^2(P)$ and $[-m]P$ agree.

Let us now describe the content of `tyzero`. The procedure is concerned with the equality $-\beta^{q^2} = \beta r_m(\alpha)/s_m(\alpha)$, where $r_m, s_m \in \mathbb{F}_q[x]$ correspond to (S.2). Thus $r_m = \bar{f}_{m+2}\bar{f}^2_{m-1} - \bar{f}_{m-2}\bar{f}^2_{m+1}$ if m is even, etc. Since $\beta^2 = \alpha^3+a\alpha+b$ and $\beta \neq 0$, the equality takes the form $(\alpha^3+a\alpha+b)^{(q^2-1)/2} = -r_m(\alpha)/s_m(\alpha)$. If $t_\ell = 0$, then each $\alpha \in E[\ell]^*$ fulfils this equality. That takes place if and only if $\bar{f}_\ell$ divides $s_m(x)(x^3+ax+b)^{(q^2-1)/2} + r_m(x)$.

The other procedure called by `equalx` is called `eigen`. The parameters are ℓ and m. The procedure returns `TRUE` if there exists $P \in E[\ell]^*$ such that $\varphi(P) = [m]P$. The procedure has two parts, the first part tests the first coordinate and produces a polynomial $g_\ell \in \mathbb{F}_q[x]$ that can be regarded as an input for the second part which tests the second coordinate. In Schoof's algorithm the first part may be skipped if $\gcd(\bar{s}_\ell, \bar{f}_\ell)$ is remembered, since at this point of the algorithm that polynomial coincides with g_ℓ (the exact meaning of g_ℓ is described below).

The first part is similar to the derivation of $\bar{s}_\ell$. The only difference is that the term $x^{q^2} - x$ is replaced by $x^q - x$. Indeed, we are asking whether there exists $(\alpha, \beta) \in E[\ell]^*$ such that $\alpha^q = \alpha - (\psi_{m-1}\psi_{m+1}\psi_m^{-2})(\alpha, \beta)$, and derive a polynomial in variable x for which α has to be a root. To see if there exists a root of such a polynomial that really belongs to $E[\ell]^*$ we compute the gcd of this polynomial with $\bar{f}_\ell$, and denote the gcd by g_ℓ. If $g_\ell = 1$, then the procedure returns `FALSE`. Assume that g_ℓ is nontrivial. There are some special situations when $g_\ell = \bar{f}_\ell$ (e.g. if λ is a double root of the characteristic polynomial induced by the Frobenius endomorphism). In the other situations the polynomial g_ℓ is of degree $(\ell - 1)/2$. The points $(\alpha, \beta) \in E[\ell]^*$ that fulfil $g_\ell(\alpha) = 0$ form a subgroup of $E[\ell]^*$. For the second part of the test only these points are to be considered because these are the points from which the eigenspace, if it exists, is constructed.

We are thus asking whether β^q is equal to $\beta r_m(\alpha)/s_m(\alpha)$, where r_m and s_m are derived from (S.2) as in the procedure `tyzero`, and where $g_\ell(\alpha) = 0$. This is true if $g_\ell(x)$ divides $(x^3+ax+b)^{(q-1)/2}s_m(x) - r_m(x)$ (alternatively: if the latter two polynomials possess a nontrivial common divisor).

```
PROCEDURE equalx($\ell$)
INPUT: Prime $\ell$ for which there exists $P \in E[\ell]^*$ such that there agree $x$-coordinates of $\varphi^2(P)$ and $[q_\ell]P$.
OUTPUT: The value of $t_\ell$.
if (tyzero($\ell, q_\ell$) $=$ TRUE)
    return 0;
$\tau =$ sqrt($4q_\ell$) mod $\ell$;
```

```
$\gamma = 2q_\ell/\tau$ mod $\ell$;
if (eigen($\ell, \gamma$) = TRUE)
    return $\tau$
else return $-\tau$;
```

The description of procedure nonequalx is short too. In this case the computational content is delegated to the description of polynomials h_X and h_Y (and not to subroutines).

```
PROCEDURE nonequalx($\ell, \tau$)
INPUT: Prime $\ell$ such that the $x$-coordinates of $\varphi^2(P)$ and $[q_\ell]P$ differ
       for every $P \in E[\ell]^*$.
       Positive $\tau < \ell/2$ that is a candidate for $t_\ell$.
OUTPUT: 0 if $t_\ell \neq \pm\tau$, 1 if $t_\ell = \tau$, $-1$ if $t_\ell = -\tau$.
if (gcd($h_X, \bar{f}_\ell$) = 1) return 0;
if (gcd($h_Y, \bar{f}_\ell$) = 1) return $-1$;
return 1;
```

When nonequalx is invoked, then it is already known that the generic addition formula holds for $\varphi^2(P) \oplus [q_\ell]P$ whenever $P \in E[\ell]^*$. Put $m = q_\ell$ to spare some indices.

Write (S.1) and (S.2) in a compact form

$$[m](\alpha, \beta) = \left(\alpha - \frac{c_m(\alpha)}{d_m(\alpha)}, \beta\frac{r_m(\alpha)}{s_m(\alpha)}\right). \tag{S.3}$$

Note that this can be used even for $m = 1$ if we set $d_1(x) = r_1(x) = s_1(x) = 1$ and $c_1(x) = 0$. With this notation $\varphi^2(P) \oplus [m]P = (\alpha^{q^2}, \beta^{q^2}) \oplus [m](\alpha, \beta)$ is equal to

$$\left(\lambda^2 - \alpha^{q^2} - \alpha + \frac{c_m(\alpha)}{d_m(\alpha)}, \ \lambda\left(2\alpha^{q^2} - \lambda^2 + \alpha - \frac{c_m(\alpha)}{d_m(\alpha)}\right) - \beta^{q^2}\right), \text{ where}$$

$$\lambda = \frac{\beta^{q^2} - \beta r_m(\alpha)/s_m(\alpha)}{\alpha^{q^2} - \alpha + c_m(\alpha)/d_m(\alpha)} = \beta\frac{d_m(\alpha)}{s_m(\alpha)} \frac{(\alpha^3+a\alpha+b)^{(q^2-1)/2}s_m(\alpha) - r_m(\alpha)}{d_m(\alpha)(\alpha^{q^2} - \alpha) + c_m(\alpha)}.$$

Since in the first coordinate λ occurs only as a square, the occurrence of β may be completely eliminated from the expression of the first coordinate of $\varphi^2(P) \oplus [m]P$.

We have

$$[\tau]\varphi(P) = \varphi([\tau]P) = \left(\alpha^q - \frac{c_\tau(\alpha^q)}{d_\tau(\alpha^q)}, \beta^q\frac{r_\tau(\alpha^q)}{s_\tau(\alpha^q)}\right).$$

Therefore comparing the first coordinate of $\varphi^2(P) \oplus [m]P$ with $\alpha^q - c_\tau(\alpha^q)/d_\tau(\alpha^q)$ results into a polynomial condition on α. This is how polynomial h_X is derived. The first coordinates thus agree if and only if $h_X(\alpha) = 0$, assuming $d_m(\alpha) \neq 0$, $s_m(\alpha) \neq 0$ and $d_\tau(\alpha) \neq 0$. The latter assumptions cause no difficulty since an element of $E[q_\ell]^*$ or $E[\tau]^*$ is never an element of $E[\ell]^*$.

Since β may be eliminated from λ^2 and since $\beta^{q^2} = \beta(\alpha^3 + a\alpha + b)^{(q^2-1)/2}$ and $\beta^q = \beta(\alpha^3 + a\alpha + b)^{(q-1)/2}$ we see that when comparing the second coordinate of $[\tau]\varphi(P)$ with the second coordinate of $\varphi^2(P) \oplus [m]P$ the value of β may be cancelled out. Therefore the equality of the second coordinates may be expressed via a polynomial in α too. This is the polynomial h_Y.

J. Normal forms, discriminant and j-invariant

Applications of elliptic curves that rely upon the difficulty of the discrete logarithm problem use equations with as little parameters as possible for the sake of efficiency when computing the group operation. The related cryptosystems have shorter keys than similar cryptosystems that are based on the difficulty of number factorization because the latter problem problem is easier to solve than the former problem. The reason is the existence of the number field sieve. No similar algorithm is known for elliptic curves. If the parameters of an elliptic curve are well chosen, then there seem to be no other attacks on mathematical principles of the problem but those that correspond to general (black box) attacks on the DLP. Of course, quantum computers may change the landscape completely, and dramatically, in particular if they will be widely available.

In the world of postquantum cryptography various concepts arise, and some of them use elliptic curves in a completely different way. This requires concepts that are different than the DLP.

As an example how the focus may change let us mention that in the advent of elliptic curve cryptography curves in characteristic two were considered as an attractive alternative to curves over primes since at that time no efficient methods solving the DLP in small characteristics were known.

Normal forms, discriminant and j-invariant are used in such discussions freely. They are considered as something that is well known and does not need an explanation. The purpose of this section is to provide such an explanation for the case of Weierstraß equations.

J.1. **Normal forms.** Recall that a Weierstraß curve is given by an equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \tag{J.1}$$

We assume that the coefficients a_i belong to a field K, and say that two Weierstraß equations are *linearly K-equivalent*, if there exists an invertible linear substitution over K that turns one such equation into a multiple of the other equation. For the sake of brevity the term "linearly K-equivalent" is shortened to "K-equivalent."

Let us first consider a somewhat weaker notion, in which only those substitutions are considered that turn an equation into an equation in a way that never allows for a possibility of a nontrivial multiple. Such substitutions necessarily have the form of $y \mapsto y + sx + t$ and $x \mapsto x + r$. Indeed, the substitution for x may not include y with a nonzero coefficient since in a Weierstraß equation the unknown y does not occur in the third power. Note that the substitution is invertible for any choice of $r, s, t \in K$.

These substitutions turn (J.1) into

$$(y{+}sx{+}t)^2{+}a_1(x{+}r)(y{+}sx{+}t){+}a_3(y{+}sx{+}t) = (x{+}r)^3{+}a_2(x{+}r)^2{+}a_4(x{+}r){+}a_6,$$

and that can be expressed as

$$\begin{aligned} y^2 + (2s{+}a_1)xy + (2t{+}a_1r{+}a_3)y = \\ x^3 + (3r{+}a_2{-}s^2{-}a_1s)x^2 + (3r^2{+}2a_2r{+}a_4{-}2st{-}a_1rs{-}a_1t{-}a_3s)x \\ + (r^3{+}a_2r^2{+}a_4r{+}a_6{-}t^2{-}a_1rt{-}a_3t). \end{aligned} \tag{J.2}$$

If $\mathrm{char}(K) \neq 2, 3$, then there exists exactly one triple $(r, s, t) \in K^3$ such that

$$\begin{gathered} 2s + a_1 = 0. \\ 3r + a_2 - s^2 - a_1s = 0, \text{ and} \\ 2t + a_1r + a_3 = 0. \end{gathered} \tag{J.3}$$

In other words, there exists exactly one triple $(r,s,t) \in K^3$ that transforms (J.1) into an equation $y^2 = x^3 + ax + b$.

If $\mathrm{char}(K) = 3$, then (J.2) takes the form

$$\begin{aligned} y^2 + (a_1{-}s)xy + (a_1r{+}a_3{-}t)y = \\ x^3 + (a_2{-}s^2{-}a_1s)x^2 + (a_4{-}a_2r{+}st{-}a_1rs{-}a_1t{-}a_3s)x \\ + (r^3{+}a_2r^2{+}a_4r{+}a_6{-}t^2{-}a_1rt{-}a_3t). \end{aligned}$$

Setting $s = a_1$ yields $a_2 - s^2 - a_1s = a_2 + a_1^2$. The equations in which $a_2 + a_1^2 = 0$ are termed *supersingular*. By setting $s = a_1$ and $t = a_1r + a_3$ they are transformed into

$$y^2 = x^3 + (a_4 - a_3a_1)x + (r^3 + (a_4{-}a_3a_1)r + (a_3^2{+}a_6)).$$

A supersingular curve in characteristic three may thus attain the form $y^2 = x^3 + ax + b$, but with a much bigger degree of freedom in the choice of b. Obviously, if K is algebraically closed, then r may be chosen in such a way that b vanishes.

If $\mathrm{char}(K) = 3$ and $a_2 + a_1^2 \neq 0$, the choice of $s = a_1$ and $t = a_1r + a_3$ produces

$$\begin{aligned} y^2 = x^3 + (a_2{+}a_1^2)x^2 + (a_4{-}a_1a_3{-}(a_2{+}a_1^2)r)x \\ + (r^3{+}a_2r^2{+}a_4r{+}a_6{+}a_1^2r^2{+}a_3^2{-}a_1a_3r). \end{aligned}$$

In the nonsupersingular case there is thus only one choice of $(r,s,t) \in K^3$ that transforms (J.1) into a form $y^2 = x^3 + ax^2 + b$, and in that case $a = a_2 + a_1^2$.

Suppose now that $\mathrm{char}(K) = 2$. Then (J.1) attains the form

$$\begin{aligned} y^2 + a_1xy + (a_1r{+}a_3)y = x^3 + (r{+}a_2{+}s^2{+}a_1s)x^2 \\ + (r^2{+}a_4{+}a_1rs{+}a_1t{+}a_3s)x + (r^3{+}a_2r^2{+}a_4r{+}a_6{+}t^2{+}a_1rt{+}a_3t). \end{aligned}$$

A *supersingular* curve is obtained when $a_1 = 0$. In such a case the choice $r = a_2{+}s^2$ yields

$$y^2 + a_3y = x^3 + (s^4{+}a_3s{+}a_2^2{+}a_4)x + (s^6{+}(a_2^2{+}a_4)s^2{+}t^2{+}a_3t{+}a_4a_2{+}a_6).$$

If K is algebraically closed, then s and t may be chosen in such a way that the equation is K-equivalent to $y^2 + a_3y = x^3$.

If $a_1 \neq 0$, then t and r may be chosen in such a way that there exists a $c \in K$ such that the equation is K-equivalent to

$$y^2 + a_1xy = x^3 + (a_3a_1^{-1}{+}a_2{+}s^2{+}a_1s)x^2 + c.$$

If K is algebraically closed, then s may be chosen in such a way that the form $y^2 + a_1xy = x^3 + c$ is attained.

To sum up, for every Weierstraß equation there exist substitutions $x \mapsto x + r$ and $y \mapsto y + sx + t$ such that exactly one of the following forms is attained:

$$\begin{aligned} y^2 &= x^3 + a_4x + a_6, \text{ if } \mathrm{char}(K) \neq 2,3, \\ y^2 &= x^3 + a_4x + a_6, \text{ where } \mathrm{char}(K) = 3, \\ y^2 &= x^3 + a_2x^2 + a_6, \text{ where } \mathrm{char}(K) = 3 \text{ and } a_2 \neq 0, \\ y^2 + a_3y &= x^3 + a_4x + a_6, \text{ where } \mathrm{char}(K) = 2, \text{ and} \\ y^2 + a_1xy &= x^3 + a_2x^2 + a_6, \text{ where } \mathrm{char}(K) = 2 \text{ and } a \neq 0. \end{aligned}$$

If $\mathrm{char}(K) = 2$, then in the nonsupersingular case the coefficients a_2 and a_6 are not determined uniquely and can be expressed by polynomials in one parameter, while in the supersingular case the coefficients a_4 and a_6 polynomially depend on two parameters. Similarly, the coefficient a_6 is polynomially dependent in the supersingular case of characteristic three.

To determine if two Weierstraß equations are K-equivalent it is possible to proceed in two stages, firstly applying the substitutions $x \mapsto x + r$ and $y \mapsto y + sx + t$ described above, and then substitutions $x \mapsto u^{-2}x$ and $y \mapsto u^{-3}y$. If the first stage produces an equation that is determined uniquely (as if $\mathrm{char}(K) \neq 2, 3$), then the second stage can be used straightforwardly to decide if the equations are K-equivalent or not. However, if the first stage produces equations with coefficients that may be parameterized, then all possible values of these parameters have to be taken into account when deciding the K-equivalence.

If the substitutions $x \mapsto u^{-2}x$ and $y \mapsto u^{-3}y$ are applied to $y^2 = x^3 + a_4x + a_6$, then we get $u^{-6}y^2 = u^{-6}x^3 + a_4u^{-2}x + a_6$, and that is

$$y^2 = x^3 + u^4a_4x + u^6a_6. \tag{J.4}$$

In the remaining three cases we obtain

$$\begin{aligned} y^2 &= x^3 + u^2a_2x^2 + u^6a_6, \text{ where } \mathrm{char}(K) = 3, \\ y^2 + a_3u^3y &= x^3 + u^4a_4x + u^6a_6, \text{ where } \mathrm{char}(K) = 2, \text{ and} \\ y^2 + a_1uxy &= x^3 + u^2a_2x^2 + u^6a_6, \text{ where } \mathrm{char}(K) = 2 \text{ as well.} \end{aligned}$$

If $\mathrm{char}(K) = 2$, then u is chosen so that $u^4a_4 = 1$ and $u = a_1^{-1}$, respectively. The former choice is possible if K is perfect, which is usually assumed.

Standardly there are considered five normal forms:

(SH1) $y^2 = x^3 + a_4x + a_6$ and $\mathrm{char}(K) \notin \{2, 3\}$;
(SH2a) $y^2 + xy = x^3 + a_2x^2 + a_6$ and $\mathrm{char}(K) = 2$;
(SH2b) $y^2 + a_3y = x^3 + a_4x + a_6$ and $\mathrm{char}(K) = 2$;
(SH3a) $y^2 = x^3 + a_2x^2 + a_6$, $a_2 \neq 0$ and $\mathrm{char}(K) = 3$; and
(SH3b) $y^2 = x^3 + a_4x + a_6$ and $\mathrm{char}(K) = 3$.

Only (SH2a) uses a nontrivial application of u, setting $u = a_1^{-1}$. If K is algebraically closed, then there exists a choice of u and of the other parameters such that the equation is transformed to one of the following forms:

(SH1) $y^2 = x^3 + x + a_6$ or $y^2 = x^3 + 1$ or $y^2 = x^3$;
(SH2a) $y^2 + xy = x^3 + x^2 + a_6$;
(SH2b) $y^2 + y = x^3$ or $y^2 = x^3$;
(SH3a) $y^2 = x^3 + x^2 + a_6$; and
(SH3b) $y^2 = x^3 + x$ or $y^2 = x^3$.

The above equations are determined uniquely, with the exception of varying the sign of a_6 in (SH1), and replacing a_6 by ηa_6, $\eta^3 = 1$, in (SH3a). Note that to get one of these forms only a finite degree extension of K is necessary since what we need is a split field for one or two polynomials over K.

There is another way how to decide whether two Weierstraß equations are $\bar{K}$-equivalent. If they are nonsingular (smooth), then this is true if and only if the have the same j-invariant. To define j-invariant we first need to define the discriminant.

J.2. **Discriminant.** The discriminant $D(a)$ of a polynomial $= \sum a_ix^i \in K[x]$ is often used just for the purpose of deciding whether a has or does not have multiple roots. Indeed, $D(a) = 0$ if and only if a posseses a multiple root, as implied by the following well known result:

Proposition J.1. *Assume that $a = \sum a_ix^i \in K[x]$, $n = \deg(a) \geq 1$. Then $D(a) = 0$ if and only if a possesses a multiple root. If $\alpha_1, \ldots, \alpha_n$ are the roots of a, then*

$$D(a) = a_n^{2n-2} \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

However, this formula should not be considered as the definition of $D(a)$ since it refers to roots, not coefficients. The definition is based upon the more general notion of resultant, and can be presented in this way:

The *discriminant* $D(a)$ of a polynomial $a = \sum a_i x^i \in K[x]$, $n = \deg(a) \geq 1$, is equal

$$(-1)^{\binom{n}{2}} a_n^{-1} \det(R(a, a'), \text{ where}$$

$$R(a, a') = \begin{pmatrix} a_n & a_{n-1} & a_{n-2} & \cdots & 0 & 0 & 0 \\ 0 & a_n & a_{n-1} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_1 & a_0 & 0 \\ 0 & 0 & 0 & \cdots & a_2 & a_1 & a_0 \\ na_n & (n{-}1)a_{n-1} & (n{-}2)a_{n-2} & \cdots & 0 & 0 & 0 \\ 0 & na_n & (n{-}1)a_{n-1} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2a_2 & a_1 & 0 \\ 0 & 0 & 0 & \cdots & 3a_3 & 2a_2 & a_1 \end{pmatrix}$$

The discriminant of a cubic polynomial is thus given by

$$D(ax^3 + bx^2 + cx + d) = b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2 \tag{J.5}$$

since

$$\begin{aligned} - a^{-1} \begin{vmatrix} a & b & c & d & 0 \\ 0 & a & b & c & d \\ 3a & 2b & c & 0 & 0 \\ 0 & 3a & 2b & c & 0 \\ 0 & 0 & 3a & 2b & c \end{vmatrix} &= -a^{-1} \begin{vmatrix} a & b & c & d & 0 \\ 0 & a & b & c & d \\ 0 & b & 2c & 3d & 0 \\ 0 & 0 & b & 2c & 3d \\ 0 & 0 & 3a & 2b & c \end{vmatrix} \\ = -a^{-1} \begin{vmatrix} a & b & c & d \\ ab & 2ca & 3da & 0 \\ 0 & b & 2c & 3d \\ 0 & 3a & 2b & c \end{vmatrix} &= - \begin{vmatrix} 2ca - b^2 & 3da - cb & -db \\ b & 2c & 3d \\ 3a & 2b & c \end{vmatrix} \\ = - \begin{vmatrix} 2ca & 3da + cb & 2db \\ b & 2c & 3d \\ 3a & 2b & c \end{vmatrix} &= -4ac^3 - 27d^2a^2 - 9abcd - 4db^3 \\ &\quad +12abcd + 3abcd + c^2b^2 + 12abcd. \end{aligned}$$

Suppose now that the Weierstraß equation is given by $y^2 = f(x)$, where $f(x) = x^3 + a_2x^2 + a_4x + a_6$. By (J.5),

$$D(f) = a_2^2a_4^2 - 4a_4^3 - 4a_2^3a_6 + 18a_2a_4a_6 - 27a_6^2.$$

For reasons that will become apparent later, define b_2, b_4 and b_6 so that $f(x) = x^3 + \frac{b_2}{4}x^2 + \frac{b_4}{2}x + \frac{b_6}{4}$. Thus $b_2 = 4a_2$, $b_4 = 2a_4$ and $b_6 = 4a_6$.

Obviously,

$$16D(f) = -8b_4^3 + 9b_2b_4b_6 - 27b_6^2 + b_2^2(b_4^2 - b_2b_6)/4. \tag{J.6}$$

The transformation of (J.3) cannot be used when $\operatorname{char}(K) = 3$. However, the standard completion of a quadratic equation to square works for any characteristic different from two. This means to set $s = -a_1/2$, $t = -a_3/2$ and $r = 0$. With these values the equation (J.3) turns into

$$y^2 = x^3 + (a_2{+}a_1^2/4)x^2 + (a_4{+}a_1a_3/2)x + (a_6{+}a_3^2/4).$$

Define now b_2, b_4 and b_6 for any Weierstraß equation given by (J.1) in such a way that the above equation gets the form $y^2 = x^3 + \frac{b_2}{4}x^2 + \frac{b_4}{2}x + \frac{b_6}{4}$. Thus

$$b_2 = 4a_2 + a_1^2, \quad b_4 = 2a_4 + a_1a_3 \quad \text{and} \quad b_6 = 4a_6 + a_3^2. \tag{J.7}$$

Define one more quantity, and that is b_8, by

$$b_8 = 4a_2a_6 + a_2a_3^2 + a_1^2a_6 - a_4^2 - a_1a_3a_4. \tag{J.8}$$

If $\operatorname{char}(K) \neq 2$, then

$$\begin{aligned}\frac{b_2b_6 - b_4^2}{4} = \frac{(4a_2 + a_1^2)(4a_6 + a_3^2) - (2a_4 + a_1a_3)^2}{4}& \\ = 4a_2a_6 + a_2a_3^2 + a_1^2a_6 - a_4^2 - a_1a_3a_4 = b_8.&\end{aligned}$$

For a Weierstraß curve C given by (J.1) define the *discriminant* by

$$\Delta(C) = -8b_4^3 + 9b_2b_4b_6 - 27b_6^2 - b_2^2b_8.$$

Comparing this definition with (J.6) shows that $\Delta(C) = 16D(f)$ if $\operatorname{char}(K) \neq 2$ and C is given by $y^2 = f(x) = x^3 + a_2x^2 + a_4x + a_6$.

Theorem J.2. *Let C be a Weierstraß curve given by equation (J.1). Then $\Delta(C) = 0$ if and only if C is singular.*

If $\tilde{C}$ is given by an equation obtained via transformations $x \mapsto x + r$ and $y \mapsto y + sx + t$, then $\Delta(\tilde{C}) = \Delta(C)$.

If $\tilde{C}$ is given by an equation obtained via transformations $x \mapsto u^{-2}x$ and $y \mapsto u^{-3}y$, then $\Delta(\tilde{C}) = u^{12}\Delta(C)$.

This theorem may be proved by a direct verification. However, the polynomials that have to be compared are very long. There exists a short proof that relies upon the properties of the polynomial discriminants, upon the connection (J.6), and upon the fact that in both (J.7) and (J.8) there appears no fraction. The latter may be used for an argument that transfers the validity of the theorem in characteristic zero to a positive characteristic via factorization.

The definition of the discriminant together with (J.7) and (J.8) can be used to compute the discriminant value for the normal forms:

type	b_2	b_4	b_6	b_8	$\Delta(C)$
SH1	0	$2a_4$	$4a_6$	$-a_4^2$	$-64a_4^3 - 432a_6^2 = -8b_4^3 - 27b_6^2$
SH2a	1	0	0	a_6	$a_6 = b_8$
SH2b	0	0	a_3^2	a_4^2	$a_3^4 = b_6^2$
SH3a	a_2	0	a_6	a_2a_6	$-a_2^3a_6 = -b_2^2b_8$
SH3b	0	$-a_4$	a_6	$-a_4^2$	$-a_4^3 = -b_4^3$

J.3. **The j-invariant.** Substitutions $x \mapsto x + r$ and $y \mapsto y + sx + t$ do not change the value of b_{2i}, $1 \leq i \leq 4$. On the other hand the substitutions $x \mapsto u^{-2}x$ and $y \mapsto u^{-3}y$ change b_{2i} to $u^{2i}b_{2i}$. Because of that they also change c_4 to u^4c_4 and u_6 to u^6c_6 if

$$c_4 = b_2^2 - 24b_4 \text{ and } c_6 = -b_2^3 + 36b_2b_4 - 216b_6.$$

Let C be a Weierstraß curve given by (J.1). Suppose that C is nonsingular, i.e. that $\Delta(C) \neq 0$. The j-invariant of C is defined by

$$j(C) = \frac{c_4^3}{\Delta(C)}.$$

From Theorem J.2 it follows that it C and $\tilde{C}$ are K-equivalent, then $j(C) = j(\tilde{C})$. Furthermore,

$$j(C) = j(\tilde{C}) \quad \Longleftrightarrow \quad C \text{ and } \tilde{C} \text{ are } \bar{K}\text{-equivalent.}$$

The value of the j-invariant for the normal forms is as follows:

type	c_4	c_6	$j(C)$
SH1	$-48a_4$	$-864a_6$	$6912a_4^3/(4a_4^3+27a_6^2) = c_4^3/(c_4^3-c_6^2)$
SH2a	1	1	$1/a_6$
SH2b	0	0	0
SH3a	a_2^2	$-a_2^3$	$-a_2^3/a_6$
SH3b	0	0	0

Let the curve C be defined over a field of characteristic $p > 0$. The curve is said to be *supersingular* if $C[p] = \mathcal{O}$ (i.e., the group of C contains no element of order p). Note that supersingular curves are nonsingular, by definition. If $p \in \{2, 3\}$, then C is supersingular if and only if $j(C) = 0$.

Two smooth Weierstraß curves are birationally equivalent over K if and only if they are given by K-equivalent Weierstraß equations. Since any elliptic curve E is birationally equivalent to a Weierstraß curve, the j-invariant of E is well defined too. In fact, $j(E)$ is an invariant of the function field $\bar{K}(E)$.