Heliyon 10 (2024) e24766

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

The formal solutions of Diophantine equation $ag^y = bx + c$ ☆

Xiazhou Yang

ARTICLE INFO

Keywords:
Number theory
Diophantine equation
Congruence
Discrete logarithm

ABSTRACT

We develop a novel method to completely solve the 3-term partial exponential Diophantine equation that represents a generalization of the standard discrete logarithm problem. Our method not only reveals the internal structure of the equation's solution and yields a numerical algorithm to solve it systematically, but also provides an alternative approach to the discrete logarithm problem.

1. Introduction

The study of 3-term Diophantine equations is an important topic in the number theory. The research in this area has been resulted in many advanced ideas and techniques. Among the most famous examples are Fermat's Last Theorem [4] and Pell-Type Equations [1].

In this paper, we develop a novel method to completely solve the 3-term Diophantine equation with two unknowns, y and x, such that

$$ag^y = bx + c, \tag{1.1}$$

where all the parameters and the unknowns are non-negative integers. This equation also represents a generalization of the standard discrete logarithm problem, while the difficulty of resolving such problem is the base of the security of certain popular cryptography systems [2]. As will be shown in the following sections, the novel method developed here transforms the problem of solving above partial exponential equation to a finite set of congruence calculations. It not only reveals the internal structure of the equation's solutions and yields a numerical algorithm to solve it systematically, but also provides an alternative approach to the discrete logarithm problem.

In order to exclude certain trivial or special cases, in Section 2, we impose some restrictions on the parameters of Equation (1.1), so that we can focus on the essential issues of solving the equation. We call those restrictions "the normalization restrictions", and an equation that satisfies such restrictions "a normalized equation". Under these restrictions, in Section 3, a subset of the least residue system modulo b is defined as "the associated residue set" to the equation, and its properties are analyzed. In Section 4, based on the properties of the associated residue set, the necessary and sufficient condition for a normalized equation to be solvable is obtained, and if it's solvable, the formal solutions are constructed explicitly. In Section 5, we show that, besides a few trivial cases, an equation that doesn't satisfy the normalization restrictions can be transformed to a normalized equation. The necessary and sufficient condition for the original equation to be solvable is that its corresponding normalized equation is solvable, and if it's solvable, the solutions are given by a subset of the solutions of its corresponding normalized equation. In Section 6, we discuss briefly the applications of

☆ This paper is in memory of my father Yun Ding, and my mother Ying Yang.
E-mail address: yangxiazhou@gmail.com.

https://doi.org/10.1016/j.heliyon.2024.e24766
Received 6 December 2022; Received in revised form 29 December 2023; Accepted 14 January 2024

Available online 18 January 2024

our method to the discrete logarithm problem and the 3-term Diophantine equation with two exponential unknowns. A summary is given in Section 7.

2. Normalization restrictions

To exclude certain trivial or special cases, so that we can focus on the essential issues of solving Equation (1.1), we impose three groups of restrictions on the parameters of the equation:

$$\begin{cases} 0 < a, \\ 0 < c, \\ 1 < g, \\ 1 < b, \end{cases} \tag{2.1}$$

$$\begin{cases} gcd(a,b) = 1, \\ gcd(a,c) = 1, \end{cases} \tag{2.2}$$

$$\begin{cases} a \not\equiv 0 \pmod{g}, \\ c < b, \\ gcd(g,b) = 1, \end{cases} \tag{2.3}$$

where $gcd(m_1, m_2)$ is the greatest common divisor of m_1 and m_2. We call these restrictions "the normalization restrictions", and an equation that satisfies such restrictions "a normalized equation".

2.1. On the restriction delineated by Equation (2.1)

Because all the parameters and the unknowns being non-negative integers is the prerequisite for the equation, there are only a few trivial cases that don't satisfy the restriction (2.1).

For example, the cases that don't satisfy the restriction $1 < b$ are $b = 0$ and $b = 1$. The equation with $b = 0$ is a trivial 2-term equation with one unknown, so it is not to be processed any further here. In the case of $b = 1$, Equation (1.1) becomes

$$ag^y = x + c.$$

Let n_0 be the smallest non-negative integer such that $ag^{n_0} - c \geq 0$, the equation has solutions

$$(y, x) = (n, ag^n - c), \quad n \geq n_0,$$

where $(y, x) = (m_1, m_2)$ represents $y = m_1$ and $x = m_2$ simultaneously.

2.2. On the restriction delineated by Equation (2.2)

An equation that doesn't satisfy the restriction (2.2) can be transformed into a corresponding normalized equation, and their solutions are identical.

For example, in the case of $gcd(a, b) = d > 1$, if d doesn't divide c, Equation (1.1) has no solution. Suppose that d divides c. Dividing the both sides of the equation by d, we have

$$\bar{a}g^y = \bar{b}x + \bar{c},$$

where $\bar{a} = a/d$, $\bar{b} = b/d$, and $\bar{c} = c/d$. For the reason that $gcd(\bar{a}, \bar{b}) = 1$, the equation has been normalized. Any solution of the original equation is the solution of the corresponding normalized equation, and vice versa. By solving the corresponding normalized equation, we get all the solutions of the original equation.

2.3. On the restriction delineated by Equation (2.3)

An equation that doesn't satisfy the restriction (2.3) can also be transformed into a corresponding normalized equation. However, in this case, although any solution of the original equation is the solution of the corresponding normalized equation, a solution of the corresponding normalized equation is not necessarily the solution of the original equation. Therefore, there is no guarantee yet that the solutions of the corresponding normalized equation contain the solutions of the original equation. Such guarantee is secured after we have completely solved the normalized equation. Details are discussed in Section 5.

3. Associated residue set

The core development of our novel approach to solve Equation (1.1) is to define a data set associated to the equation. We call this data set "the associated residue set", denoted by $< C >$. The parameters g, b, and c are deeply involved in its definition, and its relationship with the parameter a determines whether the equation is solvable, and if it's solvable, how the solutions are constructed.

Let γ be the least non-negative residue of b modulo g, i.e.,

$$b \equiv \gamma \pmod{g}, \quad 0 < \gamma < g, \tag{3.1}$$

where $\gamma \neq 0$ is the result of the restriction (2.3), namely $gcd(g, b) = 1$.

The associated residue set $< C >$ is defined by the following recursive formulas:

$$\begin{cases} c_0 = c, \\ c_{i+1} = \frac{b\mu_i + c_i}{g}, \end{cases} \tag{3.2}$$

where μ_i is solved from the linear congruence equation,

$$\gamma\mu_i + c_i \equiv 0 \pmod{g}, \quad 0 \leq \mu_i < g. \tag{3.3}$$

Lemma 3.1. *If c_i is an integer, so is c_{i+1}.*

Proof. If c_i is an integer, from Equations (3.1) and (3.3), we have

$$b\mu_i + c_i \equiv \gamma\mu_i + c_i \equiv 0 \pmod{g}.$$

Hence g divides $b\mu_i + c_i$ and c_{i+1} is an integer. □

Lemma 3.2. *If $c_i > 0$, so is c_{i+1}.*

Proof. Both μ_i and b are non-negative integers, thus, if $c_i > 0$, $b\mu_i + c_i > 0$ which leads immediately to $c_{i+1} > 0$. □

Lemma 3.3. *If $c_i < b$, so is c_{i+1}.*

Proof. Replacing c_i by b in Equation (3.2), and using the condition $c_i < b$, we have

$$c_{i+1} < b\frac{\mu_i + 1}{g}.$$

Equation (3.3) requires that $\mu_i < g$, which implies that $\mu_i + 1 \leq g$. As a result, the above equation becomes

$$c_{i+1} < b\frac{\mu_i + 1}{g} \leq b\,\frac{g}{g} = b. \quad \square$$

To avoid unnecessary tedious discussions, in the rest of this section and Section 4, we suppose that Equation (1.1) satisfies all the normalization restrictions described in Section 2.

Theorem 3.1. *All the elements of $< C >$ are integers, and their values are bounded within the range between 0 and b, i.e.,*

$$0 < c_i < b, \quad i \geq 0. \tag{3.4}$$

As a consequence, $< C >$ consists of a subset of the least residue system modulo b.

Proof. Combined with the initial condition $c_0 = c$, and the normalization restrictions $0 < c < b$, this theorem is concluded directly from Lemmas 3.1 to 3.3. □

The fact that all c_i are integers proves that $< C >$ is well-defined.

Lemma 3.4. *There exists an unique inverse of γ, denoted by γ^{-1}, such that*

$$\gamma^{-1}\gamma \equiv 1 \pmod{g}. \tag{3.5}$$

Hence Equation (3.3) has a single solution

$$\mu_i \equiv -\gamma^{-1}c_i \pmod{g}. \tag{3.6}$$

Proof. Equation (3.1) implies that there is an integer m such that $b = mg + \gamma$. Therefore, the normalization restriction $gcd(g, b) = 1$ indicates that $gcd(g, \gamma) = 1$ which guarantees the existence of the unique inverse γ^{-1} defined by Equation (3.5) (See Theorem 3.13 of [3], Page 72). □

Lemma 3.5. *If $c_i \equiv 0 \pmod g$, then $\mu_i = 0$, and vice versa.*

Proof. This is the result of Equations (3.6) and (3.3). □

Theorem 3.2. *Not all μ_i are zeros.*

Proof. Lemma 3.5 indicates that $c_i \not\equiv 0 \pmod g$ causes $\mu_i \neq 0$, hence we only need to show that there is at least one c_i such that $c_i \not\equiv 0 \pmod g$. If $c_0 \not\equiv 0 \pmod g$, the theorem is correct. Suppose $c_0 \equiv 0 \pmod g$. Due to the normalization restriction $c_0 = c > 0$, there exist two integers, m and $\bar{c}$, such that

$$c_0 = g^m \bar{c},$$

where $m > 0$ and $\bar{c} \not\equiv 0 \pmod g$. According to Lemma 3.5, $c_0 \equiv 0 \pmod g$ leads to $\mu_0 = 0$. Replacing μ_0 by 0 in Equation (3.2) results in $c_1 = g^{m-1}\bar{c}$, and $c_1 \equiv 0 \pmod g$ if $m > 1$. Repeating this procedure m times, we have

$$c_m = \bar{c} \not\equiv 0 \pmod g,$$

which proves the theorem. □

Lemma 3.6. *c_{i+1} is uniquely determined by c_i, thus $< C >$ has no forks in the forward direction.*

Proof. Due to Equation (3.6), μ_i is uniquely decided by c_i, so is the right side of Equation (3.2) which proves the lemma. □

Lemma 3.7. *μ_i is the smallest integer such that $gc_{i+1} - b\mu_i < b$.*

Proof. Combining Equation (3.2) and Equation (3.4), we have $gc_{i+1} - b\mu_i = c_i < b$. So we just need to prove that $gc_{i+1} - b(\mu_i - 1) > b$:

$$gc_{i+1} - b(\mu_i - 1) = gc_{i+1} - b\mu_i + b \quad = c_i + b > b,$$

where $c_i > 0$ is used. □

Lemma 3.7 provides a formula to uniquely derive c_i from c_{i+1}:

Lemma 3.8. *For a given c_{i+1}, let m be the smallest integer to make $gc_{i+1} - bm < b$, then*

$$\begin{cases} \mu_i & = m, \\ c_i & = gc_{i+1} - b\mu_i. \end{cases}$$

Hence $< C >$ has no forks in the backward direction.

Theorem 3.3. *$< C >$ is a pure cyclic sequence.*

Proof. Since $< C >$ is bounded from both below and above (Theorem 3.1), with the increase of the index i it will inevitably repeat its previous elements. Furthermore it is a sequence without any fork (Lemmas 3.6 and 3.8). Consequently, $< C >$ must be a pure cyclic sequence. □

Let L denote the length of the cycle, i.e., the smallest non-negative integer such that

$$c_L = c_0,$$

we have

$$\begin{cases} c_{nL+i} & = c_i, \\ \mu_{nL+i} & = \mu_i, \end{cases} \tag{3.7}$$

where n is any non-negative integer.

It is useful to have the compact form for the elements of $< C >$. Recursively using Equation (3.2) k times, we have

$$c_{i+1} = \frac{b \sum_{j=0}^{k-1} \mu_{i+1-k+j} g^j + c_{i+1-k}}{g^k}. \tag{3.8}$$

The compact form can be obtained by setting $k = i + 1$ in the above equation:

$$c_{i+1} = \frac{b \sum_{j=0}^{i} \mu_j g^j + c_0}{g^{i+1}}. \tag{3.9}$$

Let $i = L - 1$ in Equation (3.8), and notice that $c_L = c_0 = c$, we have

$$g^k c = b \sum_{j=0}^{k-1} \mu_{L-k+j} g^j + c_{L-k}. \tag{3.10}$$

Consequently,

$$g^k c \equiv c_{L-k} \pmod{b}, \ \ 0 \le k < L. \tag{3.11}$$

Denote g_k as the least non-negative residue of g^k modulo b, i.e.,

$$g^k \equiv g_k \pmod{b}, \ \ 0 < g_k < b, \tag{3.12}$$

The sequence of g_k is a multiplicative cyclic group generated by g, and L is the order of g modulo b. Since $g^L \equiv 1 \pmod{b}$, b divides $g^L - 1$. Replacing g^k by g_k in Equation (3.11), we have $g_k c \equiv c_{L-k} \pmod{b}$, or equivalently,

$$g_{L-k} c \equiv c_k \pmod{b}, \ \ 0 \le k < L. \tag{3.13}$$

This is another way to produce $< C >$. Notice that there is a significant difference between the two ways to produce $< C >$: Equation (3.13) is modulo b, while Equation (3.2) is modulo g.

Let $k = L$ in Equation (3.10), we have

$$\sum_{j=0}^{L-1} \mu_j g^j = c \frac{g^L - 1}{b} \tag{3.14}$$

which leads to

$$\sum_{j=0}^{nL+k-1} \mu_j g^j = c \frac{g^{nL} - 1}{b} + g^{nL} \sum_{j=0}^{k-1} \mu_j g^j. \tag{3.15}$$

The derivation is as follows: The summation on the left side of the equation is grouped by each cycle

$$\begin{aligned}
\sum_{j=0}^{nL+k-1} \mu_j g^j &= \sum_{j=0}^{L-1} \mu_j g^j + \sum_{j=L}^{2L-1} \mu_j g^j + \cdots + \sum_{j=(n-1)L}^{nL-1} \mu_j g^j + \sum_{j=nL}^{nL+k-1} \mu_j g^j \\
&= \sum_{j=0}^{L-1} \mu_j g^j + \sum_{j=0}^{L-1} \mu_{j+L} g^{j+L} + \cdots + \sum_{j=0}^{L-1} \mu_{j+(n-1)L} g^{j+(n-1)L} \\
&+ \sum_{j=0}^{k-1} \mu_{j+nL} g^{j+nL}.
\end{aligned}$$

Using the periodic property of μ_i given by Equations (3.7), above equation becomes:

$$\begin{aligned}
\sum_{j=0}^{nL+k-1} \mu_j g^j &= \sum_{j=0}^{L-1} \mu_j g^j + g^L \sum_{j=0}^{L-1} \mu_j g^j + \cdots + g^{(n-1)L} \sum_{j=0}^{L-1} \mu_j g^j + g^{nL} \sum_{j=0}^{k-1} \mu_j g^j \\
&= \sum_{i=0}^{n-1} g^{iL} \sum_{j=0}^{L-1} \mu_j g^j + g^{nL} \sum_{j=0}^{k-1} \mu_j g^j.
\end{aligned}$$

Notice that

$$\sum_{i=0}^{n-1} g^{iL} = \frac{g^{nL} - 1}{g^L - 1},$$

and using Equation (3.14) to replace $\sum_{j=0}^{L-1} \mu_j g^j$, we have Equation (3.15).

4. The formal solutions

To solve the normalized Equation (1.1), we expand x on the base of g:

$$x = \sum_{j=0} v_j g^j, \tag{4.1}$$

where the coefficients $0 \le v_j < g$ are to be determined by Equation (1.1). This expansion is general and any non-negative integer can be expanded in this way.

Lemma 4.1. *For any finite x, there always exists an index $k \ge 0$ such that*

$$\begin{cases} v_j = \mu_j, & j < k, \\ v_k \ne \mu_k, & j = k. \end{cases} \tag{4.2}$$

Proof. Since x is finite, there is an index $J \ge 0$ such that

$$v_j = 0, \quad j > J.$$

Therefore, in case that k defined by Equation (4.2) is not located in the range $0 \le j \le J$, it can be located in the range $j > J$, because the sequence of μ_j is cyclic and Theorem 3.2 guarantees the existence of nonzero μ_j in that range. □

Apparently k defined by Equation (4.2) is the smallest index that $v_k \ne \mu_k$. We rewrite Equation (4.1) by taking this into account:

$$x = g^k z + \sum_{j=0}^{k-1} \mu_j g^j, \tag{4.3}$$

where

$$z = \sum_{i=0} v_{i+k} g^i. \tag{4.4}$$

Equations (4.2) and (4.4) require respectively that z satisfies two restrictions:

$$\begin{cases} bz + c_k \equiv \gamma v_k + c_k \not\equiv 0 \pmod g, \\ z \ge 0. \end{cases} \tag{4.5}$$

Substituting Equation (4.3) into Equation (1.1), we have

$$\begin{aligned} ag^y &= bg^k z + b \sum_{j=0}^{k-1} \mu_j g^j + c \\ &= bg^k z + g^k c_k, \end{aligned}$$

where Equation (3.9) is used. Dividing both sides of the equation by g^k, we get

$$ag^{y-k} = bz + c_k. \tag{4.6}$$

Lemma 4.2. *In order for Equation (4.6) to be valid, y must equal to k.*

Proof. If $y > k$, $bz + c_k = ag^{y-k} \equiv 0 \pmod g$ which conflicts the restriction (4.5). On the other hand, if $y < k$, the right side of Equation (4.6) is an integer, while due to the normalization restriction $a \not\equiv 0 \pmod g$, the left side is an irreducible fraction. Thus y must equal to k for the equation to be valid. □

As a consequence, Equation (4.6) is decomposed into two equations,

$$y = k \tag{4.7}$$

and

$$a = bz + c_k. \tag{4.8}$$

Considering the periodic property of the associated residue set $< C >$, it is convenient to express k by two parameters, n and $\bar{k}$, such that

$$k = nL + \bar{k},$$

where $0 \le n$ and $0 \le \bar{k} < L$. Substituting above equation into Equations (4.3) and (4.7), and replacing $\sum_{j=0}^{nL+\bar{k}-1} \mu_j g^j$ by Equation (3.15), we have

$$\begin{cases} x &= g^{nL+\bar{k}}z + c\frac{g^{nL}-1}{b} + g^{nL}\sum_{j=0}^{\bar{k}-1}\mu_j g^j, \\ y &= nL+\bar{k}. \end{cases} \tag{4.9}$$

Similarly, Equation (4.8) becomes

$$a = bz + c_{\bar{k}}, \tag{4.10}$$

where Equation (3.7) has been used.

Two unknowns, y and x, are now replaced by $\bar{k}$ and z, and solving Equation (1.1) is transformed to checking whether there exist $\bar{k}$ and z satisfying Equation (4.10).

Lemma 4.3. *If there exists an element $c_{\bar{k}}$ in the associated residue set $< C >$ such that*

$$a \equiv c_{\bar{k}} \pmod{b}, \tag{4.11}$$

then the integer z defined by

$$z = \frac{a - c_{\bar{k}}}{b}, \tag{4.12}$$

satisfies the restriction (4.5), and Equation (4.10) is valid. y and x constructed by Equation (4.9) are the solutions of Equation (1.1).

Proof. Equation (4.10) is a direct result of Equation (4.12) and its validity is inherited from (4.12). The normalization restrictions guarantee that Equation (4.5) is satisfied. In fact, Equation (4.10) gives rise to $z \geq 0$ due to that $c_{\bar{k}} < b$ and $a > 0$. It also gives rise to $zb + c_{\bar{k}} \not\equiv 0 \pmod{g}$ due to that $a \not\equiv 0 \pmod{g}$. With $\bar{k}$ and z available, y and x can be constructed by Equation (4.9). The validity of Equation (4.10) proves that y and x are the solutions of Equation (1.1). □

Lemma 4.4. *If Equation (1.1) has a solution (y, x), there must exist an element $c_{\bar{k}}$ in the associated residue set $< C >$ such that Equation (4.11) is valid.*

Proof. We begin the proof by using Equation (4.1) to expand x. The existence of k is guaranteed by Lemma 4.1. We follow the steps from Equation (4.3) to get Equation (4.10). Because (y, x) is a solution of Equation (1.1), Equation (4.10) is valid which immediately leads to Equation (4.11). □

We reach now the major theorem of this paper:

Theorem 4.1. *The normalized Equation (1.1) is solvable if and only if there exists an element $c_{\bar{k}}$ in the associated residue set $< C >$ that satisfies Equation (4.11). If it's solvable, z can be constructed from Equation (4.12), and the solutions are given by Equation (4.9) with n being any non-negative integer.*

Proof. Equation (4.11) as the sufficient condition is proved by Lemma 4.3, and it as the necessary condition is proved by Lemma 4.4. □

5. Non-normalized equations

We discuss now the equations that violate the restriction (2.3). We suppose that (2.1) and (2.2) are satisfied, because the equations that violate these two restrictions are trivial and have been briefly discussed in Section 2.

5.1. *Case $a \equiv 0 \pmod{g}$*

Let m be the highest power of g occurring in a such that

$$a = g^m \bar{a}, \quad \bar{a} \not\equiv 0 \pmod{g}.$$

Replace a by $\bar{a}$ and y by $\bar{y} = y + m$, Equation (1.1) becomes a normalized equation

$$\bar{a}g^{\bar{y}} = bx + c. \tag{5.1}$$

Apparently, any solution (y, x) of the original equation (1.1) yields a solution $(\bar{y}, x)$ of its corresponding normalized equation (5.1) with

$$\bar{y} \geq m. \tag{5.2}$$

Hence the solvability of the corresponding normalized equation is necessary for the original equation to be solvable. On the other hand, if Equation (5.1) is solvable, its solutions $(\bar{y}, x)$ are given by Equation (4.9), thus there always exists an n_0 such that $n_0 L + \bar{k} \geq m$, so all the solutions with $n \geq n_0$ are the solutions of Equation (1.1). Therefore we have

Theorem 5.1. *Equation (1.1) with $a \equiv 0 \pmod{g}$ is solvable if and only if its normalized equation (5.1) is solvable, and its solutions are derived from the solutions of (5.1) by*

$$(y, x) = (\bar{y} - m, x),$$

where $\bar{y}$ satisfy the condition delineated by Equation (5.2).

5.2. Case $c > b$

By the division theorem (Theorem 1.1 of [3], Page 30), there is exactly one pair of positive integers m and $\bar{c}$ such that

$$c = mb + \bar{c}, \quad 0 < \bar{c} < b.$$

Notice that $\bar{c} = 0$ is the trivial case excluded by restriction (2.1) and we don't consider it here. Replacing c by $\bar{c}$ and x by $\bar{x} = x + m$, Equation (1.1) becomes a normalized equation

$$ag^y = b\bar{x} + \bar{c}. \tag{5.3}$$

Similarly, any solution (y, x) of the original equation (1.1) yields a solution $(y, \bar{x})$ of the corresponding normalized equation (5.3) with

$$\bar{x} \geq m. \tag{5.4}$$

Hence the solvability of the corresponding normalized equation is necessary for the original equation to be solvable. On the other hand, if Equation (5.3) is solvable, its solutions $(y, \bar{x})$ are given by Equation (4.9) and there always exists an n_0 such that $\bar{x} \geq m$, so all the solutions with $n \geq n_0$ are the solutions of Equation (1.1). Therefore we have

Theorem 5.2. *Equation (1.1) with $c > b$ is solvable if and only if its normalized equation (5.3) is solvable, and its solutions are derived from the solutions of (5.3) by*

$$(y, x) = (y, \bar{x} - m),$$

where $\bar{x}$ satisfy the condition delineated by Equation (5.4).

5.3. Case $gcd(g, b) > 1$

Suppose that the prime numbers q_i, $i = 1, 2, ..., s$, are the common divisors of g and b, and there is no other prime number as their common divisor. Let u_i, v_i, and w_i be the highest powers of q_i occurring in g, b, and c, respectively. Because q_i is a common divisor of g and b, $u_i > 0$ and $v_i > 0$, while w_i can be 0, i.e., $w_i \geq 0$. Let t be the count of q_i whose w_i is less than v_i. q_i can be sorted such that

$$\begin{cases} w_i < v_i, & 1 \leq i \leq t, \\ w_i \geq v_i, & t < i \leq s, \end{cases}$$

and g, b, and c are factorized as

$$\begin{cases} g &= q_1^{u_1} \dots q_t^{u_t} q_{t+1}^{u_{t+1}} \dots q_s^{u_s} \bar{g}, \\ b &= q_1^{v_1} \dots q_t^{v_t} q_{t+1}^{v_{t+1}} \dots q_s^{v_s} \bar{b}, \\ c &= q_1^{w_1} \dots q_t^{w_t} q_{t+1}^{w_{t+1}} \dots q_s^{w_s} \bar{c}. \end{cases}$$

By the definitions of u_i, v_i, and w_i, we have

$$\begin{cases} gcd(\bar{g}, q_i) &= 1, \\ gcd(\bar{b}, q_i) &= 1, \\ gcd(\bar{c}, q_i) &= 1, \end{cases}$$

i.e., q_i doesn't divide $\bar{g}$, $\bar{b}$, and $\bar{c}$, where $1 \leq i \leq s$. Due to that g and b have no other common divisors,

$$gcd(g, \bar{b}) = 1.$$

Equation (1.1) becomes

$$\begin{aligned} aq_1^{yu_1}\dots q_t^{yu_t}q_{t+1}^{yu_{t+1}}\dots q_s^{yu_s}\bar{g}^y = q_1^{v_1}\dots q_t^{v_t}q_{t+1}^{v_{t+1}}\dots q_s^{v_s}\bar{b}x \\ + q_1^{w_1}\dots q_t^{w_t}q_{t+1}^{w_{t+1}}\dots q_s^{w_s}\bar{c}. \end{aligned} \tag{5.5}$$

Two subcases, $t > 0$ and $t = 0$, need to be treated differently.

5.3.1. Subcase $t > 0$

Divide both sides of Equation (5.5) by $q_1^{w_1}\dots q_t^{w_t}$, we have

$$\begin{aligned} &aq_1^{yu_1-w_1}\dots q_t^{yu_t-w_t}q_{t+1}^{yu_{t+1}}\dots q_s^{yu_s}\bar{g}^y = \\ &\quad q_1^{v_1-w_1}\dots q_t^{v_t-w_t}q_{t+1}^{v_{t+1}}\dots q_s^{v_s}\bar{b}x \\ &\quad + q_{t+1}^{w_{t+1}}\dots q_s^{w_s}\bar{c}. \end{aligned} \tag{5.6}$$

Since $v_i > w_i$, $1 \le i \le t$, the right side of the equation is an integer. In order for the left side of the equation to be an integer, we must have $yu_i - w_i \ge 0$, $1 \le i \le t$. However, if $yu_1 - w_1 > 0$, for example, dividing both sides of the equation by q_1 leads to

$$\begin{aligned} &aq_1^{yu_1-w_1-1}\dots q_t^{yu_t-w_t}q_{t+1}^{yu_{t+1}}\dots q_s^{yu_s}\bar{g}^y = \\ &\quad q_1^{v_1-w_1-1}\dots q_t^{v_t-w_t}q_{t+1}^{v_{t+1}}\dots q_s^{v_s}\bar{b}x \\ &\quad + q_{t+1}^{w_{t+1}}\dots q_s^{w_s}\bar{c}/q_1. \end{aligned}$$

The last term of the equation is an irreducible fraction while the rest of the equation are integers, thus for the equation to be valid we must have $yu_1 - w_1 = 0$. This argument is valid for all factor q_i, $1 \le i \le t$. We conclude

Theorem 5.3. *The necessary condition for Equation (5.6) to be solvable is that $w_1/u_1 = \dots = w_t/u_t = n_0$ is an integer, and $y = n_0$.*

When this condition is satisfied, Equation (5.6) becomes

$$\begin{aligned} a(q_{t+1}^{u_{t+1}}\dots q_s^{u_s}\bar{g})^{n_0} = q_1^{v_1-w_1}\dots q_t^{v_t-w_t}q_{t+1}^{v_{t+1}}\dots q_s^{v_s}\bar{b}x \\ + q_{t+1}^{w_{t+1}}\dots q_s^{w_s}\bar{c}. \end{aligned}$$

This is a trivial linear equation with one unknown x, so we are not to process it any further.

5.3.2. Subcase $t = 0$

We divide both sides of Equation (5.5) by $q_1^{v_1}\dots q_s^{v_s}$:

$$\begin{aligned} aq_1^{yu_1-v_1}\dots q_s^{yu_s-v_s}\bar{g}^y = \bar{b}x \\ + q_1^{w_1-v_1}\dots q_s^{w_s-v_s}\bar{c}. \end{aligned}$$

Because all $w_i \ge v_i$, the right side of the equation is an integer. In order for the left side of the equation to be an integer, it is needed that

$$y \ge \lceil max(v_1/u_1, \dots, v_s/u_s)\rceil, \tag{5.7}$$

where $\lceil z\rceil$ is the ceiling function and $max(z_1, \dots, z_m)$ is the largest value among z_i.

Replacing b by $\bar{b}$ and x by $\bar{x} = q_1^{v_1}\dots q_s^{v_s}x$, Equation (5.5) is normalized:

$$ag^y = \bar{b}\bar{x} + c, \tag{5.8}$$

where it is supposed that $c < \bar{b}$. If $c > \bar{b}$, then the equation can be further processed following the steps of Subsection 5.2.

As previous cases, any solution (y, x) of the original equation (5.5) with $t = 0$ yields a solution $(y, \bar{x})$ of its corresponding normalized equation (5.8). Hence the solvability of the corresponding normalized equation is necessary for the original equation to be solvable. On the other hand, if Equation (5.8) is solvable, its solutions $(y, \bar{x})$ are given by Equation (4.9), thus there always exists an n_0 such that $n_0L + \bar{k} \ge \lceil max(v_1/u_1, \dots, v_s/u_s)\rceil$. The $\bar{x}$ of a solution with $n \ge n_0$ is divisible by $q_1^{v_1}\dots q_s^{v_s}$, because the other two terms of the equation (5.8) are divisible by the same factor. Therefore we have

Theorem 5.4. *Equation (5.5) with $t = 0$ is solvable if and only if its normalized equation (5.8) is solvable, and its solutions are derived from the solutions of (5.8) by*

$$(y, x) = (y, \frac{\bar{x}}{q_1^{v_1}\dots q_s^{v_s}})$$

where y satisfies the condition delineated by Equation (5.7).

Remark. In the case of $c > b$ or $gcd(g, b) > 1$, $< C >$ is still well-defined by Equation (3.2) and (3.3), but has more complicated structures than a pure cyclic sequence. The necessary and sufficient condition for Equation (1.1) to be solvable given by Theorem 4.1 is still valid, and the solutions are similar to Equation (4.9) with some exceptions that n has limited values. Those exceptions correspond to the solvable trivial cases of this section.

6. Two applications of the method

We discuss briefly two applications of our method in this section.

6.1. The discrete logarithm problem

Let $a = 1$ and c be a least non-negative residue g_s defined by Equation (3.12), i.e., $g^s \equiv g_s \pmod{b}$ where $0 \le s < L$, Equation (1.1) becomes the standard discrete logarithm problem

$$g^y = bx + g_s.$$

The goal is to resolve s from g_s with given g and b.

The necessary and sufficient condition for the equation to be solvable given by Equation (4.11) becomes:

$$c_{\bar{k}} = 1, \quad 0 \le \bar{k} < L.$$

There always exists a $\bar{k}$ satisfying this condition, because $c_0 = g_s$ and $g_L = 1$. Substituting the above result into Equation (3.13), we have

$$g_{L-\bar{k}} g_s \equiv g_{L-\bar{k}+s} \equiv c_{\bar{k}} \equiv 1 \pmod{b},$$

which leads to

$$s = \bar{k}.$$

As mentioned in Section 3, there are two ways to solve the problem: the traditional method of Equation (3.13) that is modulo b, and our novel method of Equation (3.2) that is modulo g. For certain values of b and g, our method can deliver better optimization algorithm than the traditional one.

6.2. Equation with two exponential unknowns

The 3-term Diophantine equation to be discussed is

$$ag^y = bh^x + c.$$

By letting $\bar{x} = h^x$, Equation (4.11) provides a necessary condition for the above equation to be solvable. Also a solution $\bar{x}$ should be in the form of Equation (4.3), hence we get another necessary condition for the solutions with $y > 0$

$$h^x \equiv \mu_0 \pmod{g}.$$

7. Summary

We develop a novel method to completely solve Equation (1.1). The foundation of the method is the associated residue set $< C >$ which exposes the different roles played by each parameters in solving the equation. By this method, solving the equation is transformed to a finite set of congruence calculations. Based on the properties of $< C >$, we obtain the necessary and sufficient condition for the equation to be solvable, and if it's solvable, the formal solutions are constructed explicitly. This method not only reveals the internal structure of the equation's solution and yields a numerical method to solve it systematically, but also provides an alternative approach to the discrete logarithm problem. Furthermore, the method furnishes inspiring insights on the general 3-term Diophantine equations.

CRediT authorship contribution statement

Xiazhou Yang: Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

References

[1] Titu Andreescu, Dorin Andrica, Ion Cucurezeanu, An Introduction to Diophantine Equations: A Problem-Based Approach, Springer Science and Business Media, 2010.

[2] James S. Kraft, Lawrence C. Washington, An Introduction to Number Theory with Cryptography, second edition, CRC Press, 2018.

[3] William J. LeVeque, Fundamental of Number Theory, Dover Publications, 1996.

[4] Takeshi Saito, Fermat's Last Theorem: The Proof, American Mathematical Society, 2014.