Finding Optimal Small-Root Bounds for the Bivariate Coppersmith Method over the Integers

Zhaopeng Ding[1], Yanshuo Zhang[1], Zhaopeng Dai[2], Ziyang Yan[1], and Ying Chen[1]

[1] Department of Cryptographic Science and Technology, Beijing Electronic Science and Technology Institute, Beijing, China
16678784491@163.com, zhang_yanshuo@163.com, yanziyang9853@outlook.com, ychen@besti.edu.cn
[2] School of Mathematics and Statistics, Qingdao University, Qingdao, China
dzpeng@amss.ac.cn

Abstract. Since Coppersmith introduced his lattice-based method for finding small roots of bivariate polynomial equations over the integers in 1996, a general theoretical characterization of the optimal small-root bounds achievable by this method has remained elusive. In this paper, we resolve this long-standing open problem by providing rigorous proofs of optimality and an explicit method for computing these bounds. Specifically, using our newly introduced notion of quasi-Newton polygons, we rigorously determine the optimal small-root bounds achievable by Coron's lattice construction (EUROCRYPT 2004). We further prove that computing these bounds reduces to solving a univariate equation uniquely determined by the vertices of the Newton polygon. Applying our results to cryptanalysis, we systematically evaluate the small-root bounds reported in the literature for more than twenty cryptographic settings and prove that the vast majority of these bounds are optimal. We believe that our approach will help resolve further optimality questions in cryptanalysis and provide a theoretical basis for assessing the security of the relevant cryptographic schemes.

Keywords: Coppersmith's method · Bivariate integer equations · Optimal small-root bounds · Newton polygons

1 Introduction

In 1996, Coppersmith [8,10] introduced a method based on lattice reduction for finding small roots of bivariate polynomial equations over the integers. Coron [11] subsequently simplified the original method by using monomial shifts to construct a full-rank lattice with a triangular basis matrix, making the lattice determinant and the corresponding small-root bounds more straightforward to analyze. He further reduced the lattice dimension in subsequent work [12]. Building on Coron's lattice construction, Blömer and May [4] and Jochemsz and May [18] further developed techniques that exploit polynomial structure to construct lattices and analyze small-root bounds. The bivariate Coppersmith method over the integers

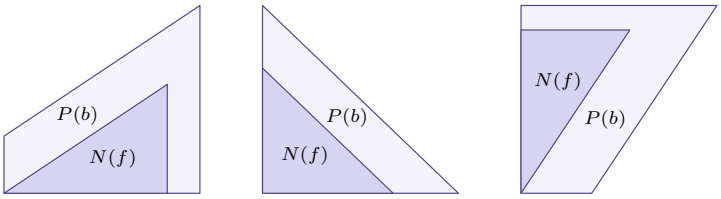
has been applied to a variety of problems in cryptanalysis and computational number theory. These include factoring integers when partial information about a prime factor is known or when the integers have a special multiplicative structure [10,5,4,23]; factoring RSA moduli using complete or partial private-key information [20,17]; recovering representations of integers as sums of squares from approximate information [16]; and analyzing the RSA-polynomial problem [22,24]. More recently, algorithms for computing supersingular isogenies of a prescribed degree have also combined norm equations over the integers with bivariate or multivariate Coppersmith methods [3,1], further extending this line of research to isogeny-based cryptography.

Determining the range of small roots that can be recovered is a fundamental problem in the study of Coppersmith's method. Let $f \in \mathbb{Z}[x, y]$ be an irreducible polynomial of degree d, and let $W = \|f(Xx, Yy)\|_\infty$. The classical bivariate Coppersmith theorem over the integers [10,13] states that integer roots within the prescribed bounds X and Y can be recovered whenever $XY \leq W^{2/(3d)}$. Sharper small-root bounds can be obtained by exploiting the geometry of the polynomial's support. Blömer and May [4] formulated the selection of monomial shifts as a combinatorial optimization problem over arbitrary monomial sets S and analyzed several families of shift sets suited to different Newton polygons. Subsequently, Jochemsz and May [18] introduced a basic strategy and an extended strategy for organizing such constructions. This approach has continued to yield improved small-root bounds for integer factorization and the RSA-polynomial problem [23,22,24]. Recently, Aardal et al. [1] used the volumes and first moments of Newton polytopes to derive a unified asymptotic small-root bound for Jochemsz and May's basic strategy.

Blömer and May [4] explicitly posed the problem of finding an optimal set S to determine the optimal small-root bounds achievable by Coron's lattice construction. However, this problem has remained unresolved to date. Instead, subsequent work has continued to expand construction strategies and develop more convenient methods for computing asymptotic bounds to address specific problems on a case-by-case basis.

Our Contributions. In this paper, we rigorously determine the optimal small-root bounds achievable by Coron's lattice construction for the bivariate Coppersmith method over the integers and provide an explicit method for computing these bounds. Our contributions are as follows.

Rigorous proof of the optimal small-root bounds. Following the approach of Blömer and May [4], we first reduce the problem of determining the optimal small-root bounds to a combinatorial optimization problem over arbitrary finite monomial sets S. We introduce a quasi-Newton polygon $P(b)$, illustrated below, as the central geometric construction for characterizing optimality.



By constructing a quasi-Newton polygon satisfying a balance condition, we obtain matching upper and lower bounds. On the one hand, the balance condition yields a uniform estimate valid for every finite nonempty monomial set S, giving an upper bound on the optimal value. On the other hand, by dilating the polygon to $mP(b)$ and taking the lattice points it contains, we construct a family of monomial sets that asymptotically attains this upper bound. We thus rigorously determine the optimal small-root bounds achievable over all admissible monomial sets.

Explicit computation of the optimal small-root bounds. Let $X = W^{\alpha}$ and $Y = W^{\beta}$, where $\alpha, \beta > 0$. We prove that computing the optimal small-root bounds reduces to solving a univariate equation determined directly by the vertices of the Newton polygon $N(f)$:

$$\sum_{i=1}^{r} \frac{\alpha\beta \det(a_{i-1}, a_i)}{\big(\lambda - w_{\alpha,\beta}(a_{i-1})\big)\big(\lambda - w_{\alpha,\beta}(a_i)\big)} = 1, \qquad \lambda > d.$$

Here, $a_0, \ldots, a_r$ are vertices of $N(f)$, $w_{\alpha,\beta}(x) = \alpha x_1 + \beta x_2$, and $d = \max_{0 \le i \le r} w_{\alpha,\beta}(a_i)$. This equation has a unique real root $\lambda_*(\alpha, \beta)$ in the interval (d, ∞), and the optimal small-root bound is given by

$$\lambda_*(\alpha, \beta) < 1.$$

Thus, computing this real root suffices to determine the optimal small-root bound.

Analysis of cryptographic instances. We explicitly compute the optimal small-root bounds for four representative classes of Newton polygons: rectangles, axis-intercept triangles, upper triangles, and nonrectangular trapezoids. We apply the resulting bounds to cryptographic problems involving partial information leakage in RSA, the factorization of integers of special forms, and the RSA-polynomial problem. For 22 specific cases from the literature, we compare the small-root bounds obtained from existing hand-crafted constructions with our optimal bounds on a case-by-case basis. We prove that the bounds in 20 of these cases are already optimal, while two existing constructions for the RSA-polynomial problem fall short of the optimal bounds, leaving room for further improvement.

Notation and preliminary conventions. $\mathbb{Z}_{\ge 0}$ denotes the set of nonnegative integers, while $\mathbb{R}_{\ge 0}$ and $\mathbb{R}_{>0}$ denote the sets of nonnegative and positive real numbers, respectively. The cardinality of a finite set S is denoted by $|S|$, and $\lfloor t \rfloor$ and $\lceil t \rceil$ denote the floor and ceiling of t, respectively.

For a nonzero polynomial $g(x,y) = \sum_{i,j\geq 0} c_{i,j}x^i y^j \in \mathbb{Z}[x,y]$, its support is $A(g) = \{(i,j) \in \mathbb{Z}^2_{\geq 0} : c_{i,j} \neq 0\}$, and its Newton polygon is $N(g) = \mathrm{conv}(A(g))$, where conv denotes the convex hull. Polynomial norms always refer to norms of the coefficient vector, namely $\|g\|_\infty = \max_{i,j} |c_{i,j}|$ and $\|g\|_2 = (\sum_{i,j} |c_{i,j}|^2)^{1/2}$.

For sets $A, B \subseteq \mathbb{R}^2$, their Minkowski sum is $A + B = \{a + b : a \in A, b \in B\}$, and their set difference is denoted by $A \setminus B$.

For real vectors u, v of the same dimension, $\langle u, v\rangle$ and $\|u\|$ denote the standard inner product and the Euclidean norm, respectively. Vector inequalities are interpreted coordinatewise. The jth coordinate of a vector u is written as u_j or $(u)_j$. For $u, v \in \mathbb{R}^2$, we define $\det(u,v) = u_1v_2 - u_2v_1$ and write $[u,v] = \mathrm{conv}\{u,v\}$ for the line segment joining u and v.

The Euclidean area of a planar region P is denoted by $\mathrm{area}(P)$, and $dx = dx_1\,dx_2$ and ds denote the area element and the arc length element, respectively.

Linear and affine functions have the forms $u \mapsto \langle c, u\rangle$ and $u \mapsto \langle c, u\rangle + c_0$, respectively. A function H is homogeneous of degree s if $H(tb) = t^s H(b)$ for every b in its domain and every $t > 0$.

All asymptotic notation is understood with respect to the limiting process specified by the context. For an eventually positive function G, $F = O(G)$ means that $|F|/G$ is eventually bounded, and $F = o(G)$ means that $F/G \to 0$. We write $F \asymp G$ if F/G eventually lies between two positive constants independent of the limiting variable.

2 The Bivariate Coppersmith Method over the Integers

Definition 1 (Bivariate Coppersmith problem over the integers). *Given an irreducible bivariate polynomial $f(x,y) \in \mathbb{Z}[x,y]$ and bounds $X, Y > 0$, the bivariate Coppersmith problem over the integers is to find all integer pairs (x_0, y_0) satisfying $f(x_0, y_0) = 0$, $|x_0| < X$, and $|y_0| < Y$.*

We briefly review the basic procedure of Coppersmith's method for finding small roots of bivariate equations over the integers. The two main classical approaches are Coppersmith's original construction [8] and Coron's simplified construction [11]. The latter has a simpler form and more closely resembles the lattice constructions commonly used to find small roots of modular equations. Both methods yield the same asymptotic bound, so we adopt Coron's construction to reduce the subsequent analysis of small-root bounds to a purely combinatorial optimization problem. For detailed studies of finding small roots of bivariate equations over the integers, see [8,11,4,18,12].

Let $W = \|f(xX, yY)\|_\infty$. Following Coron's preprocessing [11], it suffices to consider the case in which the constant term $a_0 = f(0,0)$ is nonzero. We additionally assume that $\dim N(f) = 2$. Choose arbitrary finite nonempty monomial sets S, M such that, for every $x^i y^j \in S$, all monomials appearing in $x^i y^j f(x,y)$ belong to M. Here, S specifies the monomial shifts used in the construction, and M indexes the columns of the coefficient matrix. Since f has a nonzero constant term, this condition also ensures that $S \subseteq M$.

Let ℓ_1, ℓ_2 be the largest exponents of x and y, respectively, among the monomials in S, and define the artificial modulus $R = WX^{\ell_1}Y^{\ell_2}$. Following Coron's preprocessing, we adjust X, Y, W as needed to ensure that $\gcd(a_0, R) = 1$, retaining the same notation for the adjusted parameters. To make the constant term equal to 1, we set $f'(x, y) = a_0^{-1} f(x, y) \bmod R$ and construct the shifted polynomials

$$\begin{aligned} g_{i,j}(x, y) &= x^i y^j f'(x, y) X^{\ell_1 - i} Y^{\ell_2 - j}, && x^i y^j \in S, \\ h_{i,j}(x, y) &= x^i y^j R, && x^i y^j \in M \setminus S. \end{aligned}$$

We refer to this as Coron's construction. Note that all the polynomials $g_{i,j}$ and $h_{i,j}$ vanish at (x_0, y_0) modulo R. We replace x, y by xX, yY, respectively, and arrange the coefficients according to the monomials in M. The resulting coefficient vectors form a basis of a lattice L of dimension $\omega = |M|$.

Ordering the monomials in increasing lexicographic order and arranging the rows accordingly yields an upper triangular basis matrix. Since f' has constant term 1, the diagonal entry in each row corresponding to $g_{i,j}$ is $X^{\ell_1}Y^{\ell_2}$, and the diagonal entry in each row corresponding to $h_{i,j}$ is RX^iY^j. Therefore,

$$\begin{aligned} \det L &= \left(X^{\ell_1}Y^{\ell_2}\right)^{|S|} \prod_{x^i y^j \in M \setminus S} RX^iY^j \\ &= \left(X^{\ell_1}Y^{\ell_2}\right)^{\omega} W^{\omega - |S|} X^{s_1} Y^{s_2}, \end{aligned}$$

where $s_1 = \sum_{x^i y^j \in M \setminus S} i$ and $s_2 = \sum_{x^i y^j \in M \setminus S} j$.

In the asymptotic analysis, we absorb factors depending only on the lattice dimension and the polynomial degree into the error term, so the required determinant condition can be written as $(\det L)^{1/\omega} < R$. Substituting the determinant expression above reduces this condition to

$$X^{s_1} Y^{s_2} < W^{|S|}. \tag{1}$$

More precisely, for fixed S, M and any fixed $\varepsilon > 0$, if $X^{s_1}Y^{s_2} \le W^{|S|-\varepsilon}$, then, for sufficiently large W, the LLL algorithm yields a nonzero auxiliary polynomial $q(x, y)$ satisfying $\|q(xX, yY)\|_2 < R/\sqrt{\omega}$. Since $q(x_0, y_0) \equiv 0 \pmod R$, Howgrave-Graham's lemma implies that $q(x_0, y_0) = 0$ over the integers.

Moreover, every coefficient of $q(xX, yY)$ is divisible by $X^{\ell_1}Y^{\ell_2}$. Combining this divisibility with Coron's lower bound on the norm of a multiple of f [11] shows that, under the asymptotic margin above, the resulting q cannot be a multiple of f. Since f is irreducible, f and q are coprime. The target integer roots can therefore be recovered by elimination using resultants, followed by univariate polynomial root finding.

Definition 2 (Optimal small-root bound problem). *Let $X = W^\alpha$ and $Y = W^\beta$, where $\alpha, \beta > 0$. An exponent pair (α, β) is called* feasible *if there exists a Coron shift construction for which condition* (1) *holds. The* optimal small-root bound problem *is to determine the feasible region $\mathsf{FR}(f)$ consisting of all feasible exponent pairs.*

Write $A = A(f)$. For ease of counting, we identify S, M with their corresponding sets of exponent vectors. Throughout what follows, S ranges over the finite nonempty subsets of $\mathbb{Z}^2_{\geq 0}$. For a fixed S, every admissible M contains $S + A$. Hence, choosing $M = S + A$ minimizes both exponent sums s_1, s_2 in the determinant condition (1), thereby maximizing both α and β. Let $w_{\alpha,\beta}(x) = \alpha x_1 + \beta x_2$. By the preceding discussion, for a fixed S, it suffices to consider the condition

$$\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) < |S|.$$

Thus, the feasible region defined by this condition is

$$\mathsf{FR} = \bigcup_S \left\{ (\alpha, \beta) \in \mathbb{R}^2_{>0} : \sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) < |S| \right\}.$$

We next describe this feasible region using an optimal-value function.

Lemma 1. *For every $\alpha, \beta > 0$, the quantity*

$$\delta(\alpha, \beta) = \sup_S \frac{|S|}{\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x)}$$

is well defined, and

$$\mathsf{FR} = \left\{ (\alpha, \beta) \in \mathbb{R}^2_{>0} : \delta(\alpha, \beta) > 1 \right\}.$$

Proof. Choose $\gamma \in A \setminus \{0\}$. Since $S \subseteq S + A$ and $S + \gamma \subseteq S + A$, the nonnegativity and linearity of $w_{\alpha,\beta}$ give

$$\begin{aligned} \sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) &= \sum_{x \in S+A} w_{\alpha,\beta}(x) - \sum_{x \in S} w_{\alpha,\beta}(x) \\ &\geq \sum_{x \in S+\gamma} w_{\alpha,\beta}(x) - \sum_{x \in S} w_{\alpha,\beta}(x) \\ &= |S| w_{\alpha,\beta}(\gamma). \end{aligned}$$

Because $w_{\alpha,\beta}(\gamma) > 0$, the denominator of each ratio is positive, and

$$\frac{|S|}{\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x)} \leq \frac{1}{w_{\alpha,\beta}(\gamma)}. \tag{2}$$

Thus, the supremum $\delta(\alpha, \beta)$ is finite. Taking $S = \{0\}$ also shows that $\delta(\alpha, \beta) > 0$.

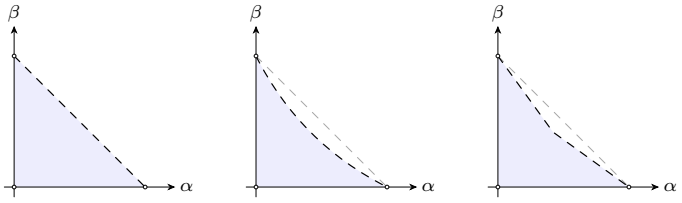
By the expression for the feasible region, $(\alpha, \beta) \in \mathsf{FR}$ if and only if there exists a finite nonempty set S for which the ratio above is strictly greater than 1. By the definition of the supremum, this is equivalent to $\delta(\alpha, \beta) > 1$. □

We have thus reduced the optimal small-root bound problem for the bivariate Coppersmith method over the integers to determining the function $\delta(\alpha, \beta)$.

Lemma 2. *The function $\delta : \mathbb{R}^2_{>0} \to \mathbb{R}_{>0}$ is homogeneous, nonincreasing, continuous, and convex.*

Corollary 1. *The line segment joining any two points on the curve $\{(\alpha, \beta) \in \mathbb{R}^2_{>0} : \delta(\alpha, \beta) = 1\}$ lies entirely in $\mathbb{R}^2_{>0} \setminus \mathsf{FR}$.*

Thus, the boundary of FR in the first quadrant is a continuous curve bending toward the origin, as illustrated below.



3 A Bridge from Combinatorial Optimization to Geometric Analysis

Lemma 3. *Let $A \subseteq \mathbb{Z}^2_{\geq 0}$ be a finite set containing the origin, and let $a_0, \ldots, a_r \in A$ be translation vectors, with repetitions allowed. If there exist decomposition functions $R_0, \ldots, R_r : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$ and a constant $\lambda \in \mathbb{R}$ such that, for every $x = (x_1, x_2) \in \mathbb{R}^2_{\geq 0}$,*

$$\sum_{i=0}^{r} R_i(x) = w_{\alpha,\beta}(x), \qquad \sum_{i=0}^{r} \big(R_i(x + a_i) - R_i(x)\big) \geq \lambda, \tag{3}$$

then, for every finite nonempty set $S \subseteq \mathbb{Z}^2_{\geq 0}$,

$$\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) \geq \lambda |S|.$$

Proof. Summing the inequality in (3) over S, interchanging the order of summation, and reindexing by the translation $x \mapsto x + a_i$, we obtain

$$\begin{aligned}
\lambda|S| &\leq \sum_{i=0}^{r} \left(\sum_{x \in S + a_i} R_i(x) - \sum_{x \in S} R_i(x) \right) \\
&\leq \sum_{i=0}^{r} \sum_{x \in (S+a_i)\setminus S} R_i(x) \\
&\leq \sum_{x \in (S+A)\setminus S} \sum_{i=0}^{r} R_i(x) \\
&= \sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x).
\end{aligned}$$

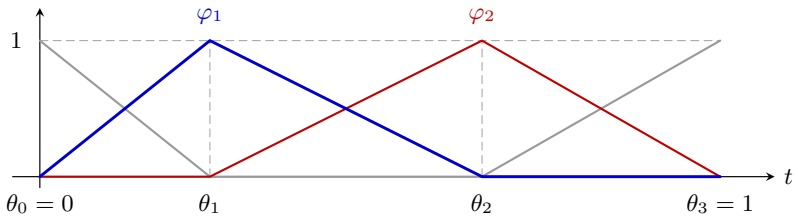
The second inequality follows by canceling the common terms in the sums over S and $S + a_i$ and then using the nonnegativity of R_i. For the third inequality, we use $S + a_i \subseteq S + A$ and the nonnegativity of R_i, and then interchange the order of summation. □

Remark 1. The simplest decomposition is obtained by taking $R_0(x) = \alpha x_1$, $R_1(x) = \beta x_2$, and $a_0 = a_1 = \gamma$, where $\gamma \in A \setminus \{0\}$. It is straightforward to verify that the conditions in (3) hold. By Lemma 3, for every finite nonempty set S, we have

$$\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) \geq w_{\alpha,\beta}(\gamma)|S|,$$

which exactly recovers the trivial bound (2). This construction uses only the coordinate decomposition and a single translation vector. We next consider finer decompositions and choices of translation vectors to obtain stronger lower bounds.

Decomposition functions. We use linear interpolation on an interval to decompose $w_{\alpha,\beta}(x)$ into a sum of nonnegative functions. Choose an integer $r \geq 1$ and nodes $0 = \theta_0 < \theta_1 < \cdots < \theta_r = 1$. For each $i = 0, \ldots, r$, define $\varphi_i : [0,1] \to \mathbb{R}$ to be the continuous piecewise linear function that equals 1 at θ_i, equals 0 at all other nodes, and is linear between consecutive nodes. These hat functions are nonnegative and satisfy $\sum_{i=0}^{r} \varphi_i(t) = 1$ for every $t \in [0,1]$. On each interval $[\theta_{i-1}, \theta_i]$, only φ_{i-1} and φ_i can be nonzero. The case $r = 3$ is illustrated below.



For $x \in \mathbb{R}^2_{\geq 0} \setminus \{0\}$, the ratio $\beta x_2 / w_{\alpha,\beta}(x) \in [0,1]$ is the fraction of the total weight $w_{\alpha,\beta}(x)$ contributed by the second coordinate. Substituting this ratio into the hat functions above, we define

$$R_i(x) = w_{\alpha,\beta}(x)\varphi_i\left(\frac{\beta x_2}{w_{\alpha,\beta}(x)}\right), \qquad R_i(0) = 0, \qquad i = 0, \ldots, r.$$

Then $R_i(x) \geq 0$ and $\sum_{i=0}^{r} R_i(x) = w_{\alpha,\beta}(x)$ for every $x \in \mathbb{R}^2_{\geq 0}$. When $r = 1$, we have $\varphi_0(t) = 1 - t$ and $\varphi_1(t) = t$, so $R_0(x) = \alpha x_1$ and $R_1(x) = \beta x_2$, recovering the trivial decomposition above.

Sectors. For each $i = 1, \ldots, r$, the interval $[\theta_{i-1}, \theta_i]$ corresponds to the sector in $\mathbb{R}^2_{\geq 0}$ defined by $\theta_{i-1} w_{\alpha,\beta}(x) \leq \beta x_2 \leq \theta_i w_{\alpha,\beta}(x)$. Within this sector, only R_{i-1} and R_i can be nonzero, and they are given by

$$R_{i-1}(x) = \frac{\theta_i w_{\alpha,\beta}(x) - \beta x_2}{\theta_i - \theta_{i-1}}, \qquad R_i(x) = \frac{\beta x_2 - \theta_{i-1} w_{\alpha,\beta}(x)}{\theta_i - \theta_{i-1}}.$$

Both expressions are linear in x, and the formulas for each R_i agree on the common boundaries of adjacent sectors. Thus, each R_i is a continuous piecewise linear function.

Local increments. Choose nonnegative vectors $a_0, \dots, a_r \in \mathbb{R}^2_{\geq 0}$ as the respective translation vectors for $R_0, \dots, R_r$ when computing the increments in condition (3). Since the R_i are piecewise linear, we first consider, for each $i = 1, \dots, r$, the case in which x, $x + a_{i-1}$, and $x + a_i$ all lie in the ith sector. In this case, the linear expressions for R_{i-1} and R_i on this sector give

$$\begin{aligned} &R_{i-1}(x + a_{i-1}) - R_{i-1}(x) + R_i(x + a_i) - R_i(x) \\ &\quad = \boxed{\frac{\theta_i w_{\alpha,\beta}(a_{i-1}) - \beta(a_{i-1})_2}{\theta_i - \theta_{i-1}} + \frac{\beta(a_i)_2 - \theta_{i-1} w_{\alpha,\beta}(a_i)}{\theta_i - \theta_{i-1}} =: \lambda_i}. \end{aligned} \tag{4}$$

The right-hand side is independent of the starting point x. We call this value the *local increment* of the ith sector.

The following lemma shows that, if $\lambda_1 = \cdots = \lambda_r$, their common value provides a global lower bound on the total increment.

Lemma 4. *If $\lambda_1 = \cdots = \lambda_r = \lambda$, then, for every $x \in \mathbb{R}^2_{\geq 0}$,*

$$\sum_{i=0}^{r} \big(R_i(x + a_i) - R_i(x)\big) \geq \lambda.$$

Proof. Write $u_+ = \max\{u, 0\}$ for the positive part of u. For $j = 0, \dots, r$, define the linear function $\ell_j(z) = \beta z_2 - \theta_j w_{\alpha,\beta}(z)$. The line $\ell_j(z) = 0$ corresponds to the sector boundary determined by the node θ_j. On $\mathbb{R}^2_{\geq 0}$, we have $\ell_0(z) = \beta z_2 \geq 0$ and $\ell_r(z) = -\alpha z_1 \leq 0$. By the definition of the hat functions, the piecewise linear functions R_i can be expressed uniformly as

$$\begin{aligned} R_0(z) &= \frac{\ell_1(z)_+ - \ell_1(z)}{\theta_1}, \\ R_i(z) &= \frac{\ell_{i-1}(z)_+ - \ell_i(z)_+}{\theta_i - \theta_{i-1}} + \frac{\ell_{i+1}(z)_+ - \ell_i(z)_+}{\theta_{i+1} - \theta_i}, \qquad 1 \leq i < r, \\ R_r(z) &= \frac{\ell_{r-1}(z)_+}{1 - \theta_{r-1}}. \end{aligned} \tag{5}$$

For each $j = 1, \dots, r-1$, the definition of the local increments and the equality $\lambda_j = \lambda_{j+1}$ give

$$0 = \lambda_{j+1} - \lambda_j = \frac{\ell_j(a_{j-1}) - \ell_j(a_j)}{\theta_j - \theta_{j-1}} + \frac{\ell_j(a_{j+1}) - \ell_j(a_j)}{\theta_{j+1} - \theta_j}.$$

Since ℓ_j is linear, this identity remains valid when every a_k is simultaneously replaced by $x + a_k$. It follows that $\ell_j(x + a_j)$ is a weighted average of $\ell_j(x + a_{j-1})$ and $\ell_j(x + a_{j+1})$:

$$\ell_j(x + a_j) = \frac{\theta_{j+1} - \theta_j}{\theta_{j+1} - \theta_{j-1}} \ell_j(x + a_{j-1}) + \frac{\theta_j - \theta_{j-1}}{\theta_{j+1} - \theta_{j-1}} \ell_j(x + a_{j+1}).$$

Using the convexity of the positive-part function and the fact that $\theta_i > \theta_{i-1}$, we obtain

$$\frac{\theta_{j+1}-\theta_j}{\theta_{j+1}-\theta_{j-1}}\,\ell_j(x+a_{j-1})_+ + \frac{\theta_j-\theta_{j-1}}{\theta_{j+1}-\theta_{j-1}}\,\ell_j(x+a_{j+1})_+ \geq \ell_j(x+a_j)_+.$$

Rearranging this inequality gives

$$\frac{\ell_j(x+a_{j-1})_+ - \ell_j(x+a_j)_+}{\theta_j-\theta_{j-1}} + \frac{\ell_j(x+a_{j+1})_+ - \ell_j(x+a_j)_+}{\theta_{j+1}-\theta_j} \geq 0.$$

We substitute $z = x + a_i$ into the expression for $R_i(z)$ in (5), sum over $i = 0, \ldots, r$, and group the terms involving $\ell_j(\cdot)_+$ for each j. Using $\ell_0(z)_+ = \ell_0(z)$, $\ell_r(z)_+ = 0$, and the inequality above, we obtain

$$\begin{aligned}
\sum_{i=0}^{r} R_i(x+a_i) &= \frac{\ell_0(x+a_1)-\ell_1(x+a_0)}{\theta_1} \\
&\quad + \sum_{j=1}^{r-1} \underbrace{\left(\frac{\ell_j(x+a_{j-1})_+ - \ell_j(x+a_j)_+}{\theta_j-\theta_{j-1}} + \frac{\ell_j(x+a_{j+1})_+ - \ell_j(x+a_j)_+}{\theta_{j+1}-\theta_j} \right)}_{\geq 0} \\
&\geq \frac{-\ell_1(x+a_0)+\ell_0(x+a_1)}{\theta_1} \\
&= w_{\alpha,\beta}(x) + w_{\alpha,\beta}(a_0) + \frac{\beta\big((a_1)_2-(a_0)_2\big)}{\theta_1} \\
&= w_{\alpha,\beta}(x) + \lambda_1.
\end{aligned}$$

Finally, the conclusion follows from $\sum_{i=0}^{r} R_i(x) = w_{\alpha,\beta}(x)$ and $\lambda_1 = \lambda$. □

Theorem 1. *Let $A \subseteq \mathbb{Z}^2_{\geq 0}$ be a finite set containing the origin. Suppose there exist an integer $r \geq 1$, vectors $a_0, \ldots, a_r \in A$, and nodes $0 = \theta_0 < \theta_1 < \cdots < \theta_r = 1$ such that the corresponding local increments satisfy $\lambda_1 = \cdots = \lambda_r = \lambda$. Then, for every finite nonempty set $S \subseteq \mathbb{Z}^2_{\geq 0}$,*

$$\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) \geq \lambda |S|.$$

Proof. Construct the functions $R_0, \ldots, R_r$ using the given nodes. Then each R_i is nonnegative, and $\sum_{i=0}^{r} R_i(x) = w_{\alpha,\beta}(x)$. By Lemma 4, for every $x \in \mathbb{R}^2_{\geq 0}$,

$$\sum_{i=0}^{r} \big(R_i(x+a_i) - R_i(x)\big) \geq \lambda.$$

Since $a_0, \ldots, a_r \in A$, the conclusion follows from Lemma 3. □

Thus, to obtain a uniform lower bound valid for every finite nonempty set S, it suffices to choose translation vectors in A and interval nodes so that the local increments in all sectors are equal.

4 Geometric Framework: Quasi-Newton Polygons

4.1 Quasi-Newton polygons

In what follows, we use several elementary facts from plane analytic geometry and the theory of convex polygons; relevant background can be found in [6]. Traverse the boundary of $N(f)$ counterclockwise from its rightmost point on the horizontal axis to its highest point on the vertical axis, and label the successive vertices $a_0, \dots, a_r$. If $N(f)$ meets either coordinate axis only at the origin, take the corresponding endpoint to be the origin.

For each oriented edge $[a_{i-1}, a_i]$, define the outward normal vector

$$\nu_i = \big((a_i)_2 - (a_{i-1})_2,\, (a_{i-1})_1 - (a_i)_1\big), \qquad 1 \le i \le r.$$

This vector is obtained by rotating the edge vector clockwise through 90°, and the line containing the edge is $\langle \nu_i, x\rangle = \det(a_{i-1}, a_i)$. Keeping the normal vectors ν_i fixed, we translate these lines to $\langle \nu_i, x\rangle = b_i$ and define

$$P(b) = \left\{x \in \mathbb{R}^2_{\ge 0} : \langle \nu_i, x\rangle \le b_i,\ 1 \le i \le r\right\}, \tag{6}$$

where $b = (b_1, \dots, b_r)$.

For $1 \le i < r$, let z_i be the intersection of the adjacent lines $\langle \nu_i, x\rangle = b_i$ and $\langle \nu_{i+1}, x\rangle = b_{i+1}$. Let z_0 be the intersection of the first line with the horizontal axis, and z_r the intersection of the last line with the vertical axis. These intersection points are uniquely determined and depend linearly on b. Since z_{i-1} and z_i lie on the same translated line, there is a linear function $\ell_i(b)$ of b such that

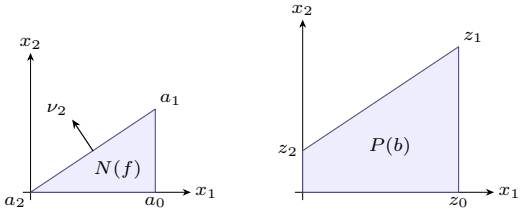
$$z_i - z_{i-1} = \ell_i(b)(a_i - a_{i-1}), \qquad 1 \le i \le r. \tag{7}$$

The condition $\ell_i(b) > 0$ means that this segment has the same orientation as the original edge and is $\ell_i(b)$ times as long.

To place the origin on the inner side of each translated line and ensure that every segment retains its original orientation and positive length, we take the parameter domain

$$\mathcal{B} = \left\{b \in \mathbb{R}^r_{>0} : \ell_i(b) > 0,\ 1 \le i \le r\right\}.$$

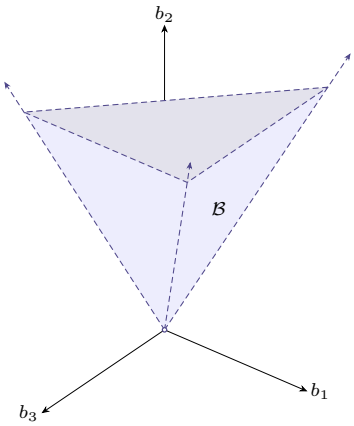
For $b \in \mathcal{B}$, we call $P(b)$ a *quasi-Newton polygon* of f. The following lemma shows that the intersection points defined above are indeed successive vertices of $P(b)$.



Lemma 5. *The parameter domain $\mathcal{B}$ is a nonempty open cone. For every $b \in \mathcal{B}$, the points $0, z_0, \ldots, z_r$ are the vertices of $P(b)$ in counterclockwise order, and the edges $[z_{i-1}, z_i]$ satisfy $\det(z_{i-1}, z_i) = b_i \ell_i(b) > 0$. Moreover,*

$$\operatorname{area}(P(b)) = \frac{1}{2} \sum_{i=1}^{r} b_i \ell_i(b). \tag{8}$$

Proof. When $b_i = \det(a_{i-1}, a_i)$, each line contains the corresponding edge of $N(f)$, so $z_i = a_i$ and $\ell_i(b) = 1$. Since the origin belongs to $N(f)$, all these b_i are nonnegative. Increasing all the b_i by the same sufficiently small positive amount makes every b_i positive while preserving $\ell_i(b) > 0$. Thus, $\mathcal{B}$ is nonempty. Since $\mathcal{B}$ is defined by strict homogeneous linear inequalities, it is an open cone.



Now fix $b \in \mathcal{B}$. The first edge is directed upward from the horizontal axis, and the last edge is directed leftward toward the vertical axis, so $(\nu_1)_1 > 0$ and $(\nu_r)_2 > 0$. Thus,

$$z_0 = \left(\frac{b_1}{(\nu_1)_1}, 0 \right), \qquad z_r = \left(0, \frac{b_r}{(\nu_r)_2} \right)$$

lie on the positive horizontal and vertical axes, respectively. Since $\ell_i(b) > 0$, every edge of the closed polygonal chain $0, z_0, \ldots, z_r, 0$ has positive length. The edges are directed, in order, rightward along the horizontal axis, along the corresponding edges of $N(f)$, and downward along the vertical axis. These directions occur in counterclockwise order and complete one full turn, so the closed polygonal chain bounds a convex polygon. Indeed, fix an edge and traverse the remaining edges from its terminal vertex back to its initial vertex. Along this traversal, the scalar projections of the successive vertices onto the outward normal of the fixed edge are first nonincreasing and then nondecreasing, eventually returning to the projection value on that edge. Thus, all vertices lie in the closed half-plane on the inner side of the line containing that edge. The inequalities defining these half-planes are precisely $x_1 \geq 0$, $x_2 \geq 0$, and $\langle \nu_i, x \rangle \leq b_i$, so this convex polygon is exactly $P(b)$.

Finally, by (7) and the definition of the outward normal vectors,

$$\begin{aligned}\det(z_{i-1}, z_i) &= \det(z_{i-1}, z_i - z_{i-1}) \\ &= \ell_i(b)\det(z_{i-1}, a_i - a_{i-1}) \\ &= \ell_i(b)\langle \nu_i, z_{i-1}\rangle \\ &= b_i \ell_i(b) > 0.\end{aligned}$$

Connecting the origin to the vertices partitions $P(b)$ into triangles with bases $[z_{i-1}, z_i]$. Therefore,

$$\operatorname{area}(P(b)) = \frac{1}{2}\sum_{i=1}^{r} \det(z_{i-1}, z_i) = \frac{1}{2}\sum_{i=1}^{r} b_i \ell_i(b).$$

This completes the proof. □

4.2 Balanced Quasi-Newton polygons

We first use the Newton polygon to construct the interval nodes and translation vectors. Fix $b \in \mathcal{B}$. We take the rays from the origin through $z_0, \ldots, z_r$ as the sector boundaries and define the corresponding interval nodes by

$$\theta_i = \frac{\beta(z_i)_2}{w_{\alpha,\beta}(z_i)}, \qquad i = 0, \ldots, r.$$

Since z_0 and z_r lie on the positive horizontal and vertical axes, respectively, we have $\theta_0 = 0$ and $\theta_r = 1$. Moreover, $\det(z_{i-1}, z_i) > 0$ gives

$$\theta_i - \theta_{i-1} = \frac{\alpha\beta\det(z_{i-1}, z_i)}{w_{\alpha,\beta}(z_{i-1})w_{\alpha,\beta}(z_i)} > 0.$$

Thus, $0 = \theta_0 < \cdots < \theta_r = 1$, and these nodes can be used in the hat function construction above.

We use the vertices $a_0, \ldots, a_r$ of $N(f)$ as the translation vectors. Using the proportionality relation $z_i - z_{i-1} = \ell_i(b)(a_i - a_{i-1})$ between the corresponding edge vectors, we now compute the local increments associated with these nodes. By the definition of the nodes, $\beta(z_j)_2 = \theta_j w_{\alpha,\beta}(z_j)$, so, for each $i = 1, \ldots, r$, we have

$$\begin{aligned}\lambda_i &= \frac{\theta_i w_{\alpha,\beta}(a_{i-1}) - \beta(a_{i-1})_2}{\theta_i - \theta_{i-1}} + \frac{\beta(a_i)_2 - \theta_{i-1} w_{\alpha,\beta}(a_i)}{\theta_i - \theta_{i-1}} \\ &= \frac{(\theta_i - \theta_{i-1})w_{\alpha,\beta}(a_{i-1}) + \beta(a_i - a_{i-1})_2 - \theta_{i-1} w_{\alpha,\beta}(a_i - a_{i-1})}{\theta_i - \theta_{i-1}} \\ &= w_{\alpha,\beta}(a_{i-1}) + \frac{\beta(a_i - a_{i-1})_2 - \theta_{i-1} w_{\alpha,\beta}(a_i - a_{i-1})}{\theta_i - \theta_{i-1}} \\ &= w_{\alpha,\beta}(a_{i-1}) + \frac{\beta(z_i - z_{i-1})_2 - \theta_{i-1} w_{\alpha,\beta}(z_i - z_{i-1})}{\ell_i(b)(\theta_i - \theta_{i-1})}\end{aligned}$$

$$= w_{\alpha,\beta}(a_{i-1}) + \frac{\big[\beta(z_i)_2 - \theta_{i-1} w_{\alpha,\beta}(z_i)\big] - \big[\beta(z_{i-1})_2 - \theta_{i-1} w_{\alpha,\beta}(z_{i-1})\big]}{\ell_i(b)(\theta_i - \theta_{i-1})}$$
$$= w_{\alpha,\beta}(a_{i-1}) + \frac{w_{\alpha,\beta}(z_i)}{\ell_i(b)} = w_{\alpha,\beta}(a_i) + \frac{w_{\alpha,\beta}(z_{i-1})}{\ell_i(b)}. \tag{9}$$

Here, the penultimate equality uses $\beta(z_i)_2 - \theta_{i-1} w_{\alpha,\beta}(z_i) = (\theta_i - \theta_{i-1}) w_{\alpha,\beta}(z_i)$ and $\beta(z_{i-1})_2 - \theta_{i-1} w_{\alpha,\beta}(z_{i-1}) = 0$. The last equality follows from $w_{\alpha,\beta}(z_i) - w_{\alpha,\beta}(z_{i-1}) = \ell_i(b)\big(w_{\alpha,\beta}(a_i) - w_{\alpha,\beta}(a_{i-1})\big)$.

Balance. We call $P(b)$ a *balanced quasi-Newton polygon* if the local increments above satisfy $\lambda_1 = \cdots = \lambda_r = \lambda$.

By (9), the common value satisfies $\lambda > 0$. Since all the vertices a_i belong to A, Theorem 1 gives, for every finite nonempty set $S \subseteq \mathbb{Z}^2_{\geq 0}$,

$$\sum_{x \in (S+A)\setminus S} w_{\alpha,\beta}(x) \geq \lambda |S|.$$

The definition of δ then yields $\delta \leq 1/\lambda$.

4.3 Constructing Asymptotic Upper Bounds Using Quasi-Newton Polygons

We now use quasi-Newton polygons to construct sets S that asymptotically attain the upper bound. Fix $P = P(b)$, where $b \in \mathcal{B}$. Let $k_i = \det(a_{i-1}, a_i)$, so that $\max_{a \in N(f)} \langle \nu_i, a \rangle = k_i$.

For each positive integer n, take

$$\begin{aligned} M_n &= nP \cap \mathbb{Z}^2_{\geq 0}, \\ S_n &= \big\{ u \in \mathbb{Z}^2_{\geq 0} : u + N(f) \subseteq nP \big\} \\ &= \big\{ u \in \mathbb{Z}^2_{\geq 0} : \langle \nu_i, u \rangle \leq n b_i - k_i,\ 1 \leq i \leq r \big\}. \end{aligned} \tag{10}$$

Since $0 \in A \subseteq N(f)$, we have $S_n \subseteq S_n + A \subseteq M_n$, and hence

$$\sum_{u \in (S_n + A)\setminus S_n} w_{\alpha,\beta}(u) \leq \sum_{u \in M_n \setminus S_n} w_{\alpha,\beta}(u). \tag{11}$$

Here, $M_n \setminus S_n$ consists of the lattice points in the strips $nb_i - k_i < \langle \nu_i, u \rangle \leq nb_i$ within nP. We now compute the sum of their weights $\sum w_{\alpha,\beta}$ and show that the coefficient of its leading term is

$$B(P) = \frac{1}{2} \sum_{i=1}^{r} k_i \ell_i(b) \big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i) \big). \tag{12}$$

Since $\sum_{i=1}^{r} k_i = 2\,\mathrm{area}(N(f)) > 0$, we have $B(P) > 0$. Moreover, since z_i and $\ell_i(b)$ depend linearly on b, both $B(P)$ and $\mathrm{area}(P)$ are homogeneous quadratic polynomials in b.

Lemma 6. *As $n \to \infty$, we have*

$$\begin{aligned} |S_n| &= n^2 \operatorname{area}(P) + O(n), \\ |M_n \setminus S_n| &= O(n), \\ \sum_{u \in M_n \setminus S_n} w_{\alpha,\beta}(u) &= n^2 B(P) + O(n). \end{aligned}$$

Proof. Associate each lattice point $u \in M_n$ with the unit square $u + [0,1)^2$. The union of these squares differs from nP only within a neighborhood of fixed width around the boundary of nP. Since the perimeter of nP is $O(n)$, this neighborhood has area $O(n)$, giving $|M_n| = n^2 \operatorname{area}(P) + O(n)$.

We next count the lattice points in each strip within nP. Fix i, and let $g_i = \gcd\big(|(\nu_i)_1|, |(\nu_i)_2|\big)$. On the integer lattice, the values of $\langle \nu_i, u \rangle$ are integer multiples of g_i, each corresponding to a line parallel to the ith edge. Conversely, Bézout's identity guarantees that every line defined by setting $\langle \nu_i, x \rangle$ equal to an integer multiple of g_i contains a lattice point. Thus, the strip $nb_i - k_i < \langle \nu_i, u \rangle \leq nb_i$ contains exactly

$$\left\lfloor \frac{nb_i}{g_i} \right\rfloor - \left\lfloor \frac{nb_i - k_i}{g_i} \right\rfloor = \frac{k_i}{g_i}$$

parallel lines containing lattice points. Here, k_i/g_i is an integer because $k_i = \langle \nu_i, a_{i-1} \rangle$.

Each of these lines is at distance at most $k_i/\|\nu_i\|$ from the ith edge of nP. Since the angles between adjacent edges are fixed, the two endpoints of the segment in which each line intersects nP are at distance $O(1)$ from nz_{i-1} and nz_i, respectively. Each such line has direction $a_i - a_{i-1}$, and $g_i = \gcd\big(|(\nu_i)_1|, |(\nu_i)_2|\big) = \gcd\big(|(a_i - a_{i-1})_1|, |(a_i - a_{i-1})_2|\big)$. Consequently, successive lattice points along the line differ by the vector $(a_i - a_{i-1})/g_i$, so the intersection segment contains $ng_i\ell_i(b) + O(1)$ lattice points. Their weights under $w_{\alpha,\beta}$ form an arithmetic progression, whose first and last terms are $nw_{\alpha,\beta}(z_{i-1}) + O(1)$ and $nw_{\alpha,\beta}(z_i) + O(1)$, respectively. By the formula for the sum of an arithmetic progression, the total weight of the lattice points on each line within nP is

$$\frac{n^2 g_i \ell_i(b)}{2} \big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)\big) + O(n).$$

Summing over the k_i/g_i lines in the strip gives

$$\sum_{\substack{u \in M_n \\ nb_i - k_i < \langle \nu_i, u \rangle \leq nb_i}} w_{\alpha,\beta}(u) = \frac{n^2 k_i \ell_i(b)}{2} \big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)\big) + O(n).$$

When summing over the strips, overlaps between distinct strips occur only in bounded neighborhoods of the vertices of nP. Thus, only $O(1)$ lattice points are counted more than once. Each such point has weight $O(n)$, so the overcount contributes only an $O(n)$ error. Together with the definition of $B(P)$, this yields

$$\sum_{u \in M_n \setminus S_n} w_{\alpha,\beta}(u) = n^2 B(P) + O(n).$$

The same counting argument also gives $|M_n \setminus S_n| = O(n)$. Finally, $|S_n| = |M_n| - |M_n \setminus S_n|$ yields $|S_n| = n^2 \operatorname{area}(P) + O(n)$. □

Corollary 2. *If there exists a balanced quasi-Newton polygon $P(b)$ whose common local increment satisfies $\lambda = B(P(b))/\operatorname{area}(P(b))$, then*

$$\delta(\alpha, \beta) = \frac{\operatorname{area}(P(b))}{B(P(b))} = \frac{1}{\lambda}.$$

Proof. Using (11), Lemma 6, and the definition of δ, together with the upper bound $\delta \le 1/\lambda$ obtained from the balance condition, we have

$$\frac{\operatorname{area}(P(b))}{B(P(b))} \ge \delta(\alpha, \beta) \ge \lim_{n\to\infty} \frac{|S_n|}{\sum_{u \in M_n \setminus S_n} w_{\alpha,\beta}(u)} = \frac{\operatorname{area}(P(b))}{B(P(b))}.$$

This completes the proof. □

5 Continuity Analysis

5.1 Stationarity Conditions

To find balanced quasi-Newton polygons, we study the stationary points of the ratio $\operatorname{area}(P(b))/B(P(b))$.

Lemma 7. *For every $b \in \mathcal{B}$, we have*

$$\frac{\partial \operatorname{area}(P(b))}{\partial b_i} = \ell_i(b), \qquad \frac{\partial B(P(b))}{\partial b_i} = \ell_i(b)\lambda_i, \qquad 1 \le i \le r. \tag{13}$$

Proof. Increasing only b_i by ε moves the line containing the corresponding edge of $P(b)$ a distance $\varepsilon/\|\nu_i\|$ in the outward normal direction. The edge has length $\|z_i - z_{i-1}\| = \ell_i(b)\|\nu_i\|$, so the area increases by $\ell_i(b)\varepsilon + O(\varepsilon^2)$, yielding the first partial derivative formula.

To compute the partial derivatives of $B(P(b))$, we first consider the integral of the weight function over the polygon. Since $w_{\alpha,\beta}$ varies linearly along the edge, its average over the edge is the average of its values at the two endpoints. Integrating over the newly added narrow strip gives

$$\begin{aligned}\frac{\partial}{\partial b_i}\int_{P(b)} w_{\alpha,\beta}(x)\,dx &= \frac{w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)}{2}\frac{\partial}{\partial b_i}\int_{P(b)} 1\,dx \\ &= \frac{\ell_i(b)}{2}\big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)\big).\end{aligned}$$

By the definition of $B(P(b))$,

$$B(P(b)) = \sum_{j=1}^{r} k_j \frac{\partial}{\partial b_j}\int_{P(b)} w_{\alpha,\beta}(x)\,dx. \tag{14}$$

We partition $P(b)$ into triangles $T_i = \mathrm{conv}\{0, z_{i-1}, z_i\}$, each having the origin as a vertex. The average of the linear function $w_{\alpha,\beta}$ over each triangle equals the arithmetic mean of its values at the three vertices. Since $w_{\alpha,\beta}(0) = 0$, we have

$$\begin{aligned}\int_{P(b)} w_{\alpha,\beta}(x)\,dx &= \sum_{i=1}^{r} \int_{T_i} w_{\alpha,\beta}(x)\,dx \\ &= \frac{1}{3}\sum_{i=1}^{r} \mathrm{area}(T_i)\big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)\big) \\ &= \frac{1}{6}\sum_{i=1}^{r} \det(z_{i-1}, z_i)\big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)\big).\end{aligned}$$

Since each vertex z_i depends linearly on b, the integral above is a cubic polynomial in b. We may therefore interchange the order of partial differentiation.

Now replace each b_j by $b_j + tk_j$ simultaneously. Since $\langle \nu_j, a_{j-1}\rangle = \langle \nu_j, a_j\rangle = k_j$, the new vertices of $P(b + tk)$ are $z_j(b + tk) = z_j + ta_j$, and the corresponding edge-length ratios are $\ell_j(b+tk) = \ell_j(b)+t$. Thus, differentiating (14) with respect to b_i and interchanging the order of partial differentiation, we obtain

$$\begin{aligned}\frac{\partial B(P(b))}{\partial b_i} &= \sum_{j=1}^{r} k_j \frac{\partial}{\partial b_j}\left[\frac{\ell_i(b)}{2}\big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)\big)\right] \\ &= \left.\frac{d}{dt}\right|_{t=0}\left[\frac{\ell_i(b)+t}{2}\big(w_{\alpha,\beta}(z_{i-1}+ta_{i-1}) + w_{\alpha,\beta}(z_i+ta_i)\big)\right] \\ &= \frac{w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i)}{2} + \frac{\ell_i(b)}{2}\big(w_{\alpha,\beta}(a_{i-1}) + w_{\alpha,\beta}(a_i)\big) \\ &= \ell_i(b)\lambda_i.\end{aligned}$$

The second equality uses the formula for the directional derivative along k, and the last equality uses the two expressions for the local increment in the final line of (9). □

Theorem 2. *A parameter vector b is a stationary point of the ratio* $\mathrm{area}(P(b))/B(P(b))$ *if and only if $P(b)$ is balanced. In this case, the common local increment satisfies* $\lambda = B(P(b))/\mathrm{area}(P(b))$*, and*

$$\delta(\alpha,\beta) = \frac{\mathrm{area}(P(b))}{B(P(b))} = \frac{1}{\lambda}.$$

Proof. By the partial derivative formulas in (13),

$$\frac{\partial}{\partial b_i}\frac{\mathrm{area}(P(b))}{B(P(b))} = \frac{\ell_i(b)}{B(P(b))^2}\left(B(P(b)) - \mathrm{area}(P(b))\lambda_i\right).$$

Since $\ell_i(b) > 0$, at any stationary point all the local increments equal $B(P(b))/\mathrm{area}(P(b))$. Hence, $P(b)$ is balanced.

Conversely, suppose that all the local increments equal λ. Since $B(P(b))$ is homogeneous of degree 2, Euler's identity for homogeneous functions gives

$$\begin{aligned} 2B(P(b)) &= \sum_{i=1}^{r} b_i \frac{\partial B(P(b))}{\partial b_i} \\ &= \lambda \sum_{i=1}^{r} b_i \ell_i(b) \\ &= 2\lambda \operatorname{area}(P(b)). \end{aligned}$$

The last equality uses (8). Thus, $\lambda = B(P(b))/\operatorname{area}(P(b))$. Substituting this into the partial derivative formula for the ratio shows that all partial derivatives vanish, so b is a stationary point.

Finally, the conclusion follows from Corollary 2. □

5.2 Construction and Uniqueness of Balanced Quasi-Newton Polygons

Theorem 2 shows that constructing a balanced quasi-Newton polygon suffices to determine $\delta(\alpha, \beta)$. We now prove that such a polygon exists and is unique after normalization, and give an explicit construction for computing $\delta(\alpha, \beta)$. To simplify the subsequent proof, we first carry out some preliminary quantitative analysis.

Let $P(b)$ be a balanced quasi-Newton polygon with common local increment λ. We show that, after normalization, its vertices are uniquely determined by λ.

Set $d = \max_{0 \le i \le r} w_{\alpha,\beta}(a_i)$. By (9), we have $\lambda > d$ and

$$\begin{aligned} \big(\lambda - w_{\alpha,\beta}(a_i)\big) w_{\alpha,\beta}(z_i) &= \big(\lambda - w_{\alpha,\beta}(a_{i-1})\big) w_{\alpha,\beta}(z_{i-1}) \\ &= \frac{w_{\alpha,\beta}(z_{i-1}) w_{\alpha,\beta}(z_i)}{\ell_i} > 0. \end{aligned}$$

Thus, $(\lambda - w_{\alpha,\beta}(a_i)) w_{\alpha,\beta}(z_i)$ takes the same positive value at all the vertices z_i. Under uniform scaling of $P(b)$, the vertex weights $w_{\alpha,\beta}(z_i)$ and the edge-length ratios ℓ_i are multiplied by the same factor, while the local increments remain unchanged. We may therefore rescale $P(b)$ to normalize this common value to 1, so that

$$w_{\alpha,\beta}(z_i) = \frac{1}{\lambda - w_{\alpha,\beta}(a_i)}, \qquad 0 \le i \le r.$$

Combining this with (7) gives

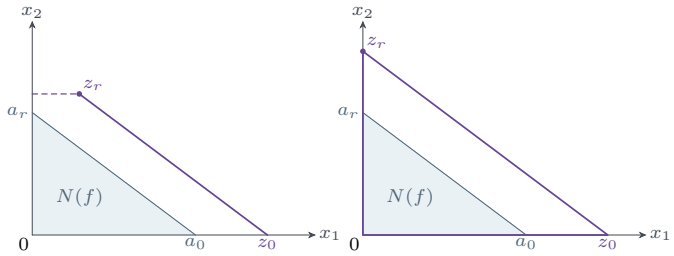
$$\begin{aligned} z_0 &= \left(\frac{1}{\alpha\big(\lambda - w_{\alpha,\beta}(a_0)\big)}, 0 \right), \\ \ell_i &= \frac{1}{\big(\lambda - w_{\alpha,\beta}(a_{i-1})\big)\big(\lambda - w_{\alpha,\beta}(a_i)\big)}, \\ z_i &= z_{i-1} + \ell_i (a_i - a_{i-1}), \qquad 1 \le i \le r. \end{aligned} \tag{15}$$

Hence, given λ, the vertices of the normalized balanced quasi-Newton polygon are uniquely determined by (15).

Conversely, for any $\lambda > d$, although (15) uniquely determines a sequence of points $z_0, \ldots, z_r$, it does not by itself guarantee that these points define a balanced quasi-Newton polygon. Surprisingly, the single endpoint condition

$$(z_r)_1 = 0$$

determines a unique normalized balanced quasi-Newton polygon.



(a) $(z_r)_1 > 0$ (b) $(z_r)_1 = 0$

An equivalent equation. For any $\lambda > d$, first construct the sequence of points using (15). Since

$$\begin{aligned} w_{\alpha,\beta}(z_i) - w_{\alpha,\beta}(z_{i-1}) &= \ell_i\big(w_{\alpha,\beta}(a_i) - w_{\alpha,\beta}(a_{i-1})\big) \\ &= \frac{1}{\lambda - w_{\alpha,\beta}(a_i)} - \frac{1}{\lambda - w_{\alpha,\beta}(a_{i-1})}, \end{aligned}$$

these points still satisfy

$$w_{\alpha,\beta}(z_i) = 1/(\lambda - w_{\alpha,\beta}(a_i)). \tag{16}$$

Using the last recurrence in (15) and $k_i = \det(a_{i-1}, a_i)$, we can write the recurrence for the first coordinate as

$$\begin{aligned} \lambda\alpha\big((z_i)_1 - (z_{i-1})_1\big) &= \lambda\alpha\ell_i\big((a_i)_1 - (a_{i-1})_1\big) \\ &= \alpha\ell_i\Big[(a_i)_1\big(\lambda - w_{\alpha,\beta}(a_{i-1})\big) - (a_{i-1})_1\big(\lambda - w_{\alpha,\beta}(a_i)\big)\Big] \\ &\quad + \alpha\ell_i\Big[(a_i)_1 w_{\alpha,\beta}(a_{i-1}) - (a_{i-1})_1 w_{\alpha,\beta}(a_i)\Big] \\ &= \alpha(a_i)_1 w_{\alpha,\beta}(z_i) - \alpha(a_{i-1})_1 w_{\alpha,\beta}(z_{i-1}) \\ &\quad + \alpha\ell_i\Big[(a_i)_1 w_{\alpha,\beta}(a_{i-1}) - (a_{i-1})_1 w_{\alpha,\beta}(a_i)\Big] \\ &= \alpha(a_i)_1 w_{\alpha,\beta}(z_i) - \alpha(a_{i-1})_1 w_{\alpha,\beta}(z_{i-1}) \\ &\quad + \alpha\beta\ell_i\Big[(a_i)_1(a_{i-1})_2 - (a_{i-1})_1(a_i)_2\Big] \\ &= \alpha(a_i)_1 w_{\alpha,\beta}(z_i) - \alpha(a_{i-1})_1 w_{\alpha,\beta}(z_{i-1}) - \alpha\beta k_i\ell_i. \end{aligned}$$

When we sum over $i = 1, \dots, r$, the terms corresponding to the intermediate vertices cancel. Using $(a_0)_2 = (a_r)_1 = 0$ and the definition of z_0, we obtain

$$\lambda\alpha(z_r)_1 = 1 - \alpha\beta \sum_{i=1}^{r} k_i \ell_i.$$

Thus, the endpoint condition $(z_r)_1 = 0$ is equivalent to the univariate equation

$$\sum_{i=1}^{r} \frac{\alpha\beta k_i}{\big(\lambda - w_{\alpha,\beta}(a_{i-1})\big)\big(\lambda - w_{\alpha,\beta}(a_i)\big)} = 1, \qquad \lambda > d. \tag{17}$$

We next prove that this equation has a unique solution and that the corresponding sequence of points indeed defines a balanced quasi-Newton polygon.

Lemma 8 (Existence and uniqueness of balanced polygons). *Equation* (17) *has a unique solution* λ_**, satisfying* $d < \lambda_* \le 2d$*. Substituting* $\lambda = \lambda_*$ *into* (15) *yields points* $0, z_0, \dots, z_r$ *that are, in order, the vertices of a quasi-Newton polygon* $P(b)$ *for some* $b \in \mathcal{B}$*. This polygon is balanced, with common local increment* λ_**. Every balanced quasi-Newton polygon is a positive scalar multiple of this polygon.*

Proof. Since each $k_i \ge 0$ and $\sum_i k_i = 2\,\mathrm{area}(N(f)) > 0$, the left-hand side of (17) is continuous and strictly decreasing on (d, ∞), and tends to 0 as $\lambda \to \infty$.

Choose a vertex of $N(f)$ at which $w_{\alpha,\beta}$ attains its maximum d. This vertex is not the origin. The supporting lines of its two incident edges cannot both pass through the origin, since they would then coincide. Hence, at least one of these edges has $k_i > 0$. The denominator of the corresponding term in (17) contains the factor $\lambda - d$, so the left-hand side tends to $+\infty$ as $\lambda \downarrow d$. By continuity and strict monotonicity, the equation has a unique solution $\lambda_* > d$.

Furthermore,

$$N(f) \subseteq \{x \in \mathbb{R}^2_{\ge 0} : w_{\alpha,\beta}(x) \le d\},$$

and the triangle on the right has area $d^2/(2\alpha\beta)$. Hence,

$$\begin{aligned}\sum_{i=1}^{r} \frac{\alpha\beta k_i}{\big(2d - w_{\alpha,\beta}(a_{i-1})\big)\big(2d - w_{\alpha,\beta}(a_i)\big)} &\le \frac{\alpha\beta}{d^2} \sum_{i=1}^{r} k_i \\ &= \frac{2\alpha\beta\,\mathrm{area}(N(f))}{d^2} \le 1.\end{aligned}$$

Applying strict monotonicity again gives $\lambda_* \le 2d$.

Now set $\lambda = \lambda_*$. All ℓ_i are positive, and the closure equation (17) gives $(z_r)_1 = 0$. By (16), we also have

$$z_r = \left(0, \frac{1}{\beta\big(\lambda_* - w_{\alpha,\beta}(a_r)\big)}\right).$$

Thus, z_0 and z_r lie on the positive horizontal and vertical axes, respectively. Adding the two edges along the coordinate axes, from the origin to z_0 and from

z_r to the origin, gives the closed polygonal chain $0, z_0, \ldots, z_r, 0$. All its edges have positive length, the edge directions make one full counterclockwise turn, and the turning angles between successive edges lie in $(0, \pi)$. By the convexity argument in the proof of Lemma 5, this chain bounds a convex polygon with all the prescribed edges.

The two edges along the coordinate axes ensure that the polygon lies in the closed first quadrant. The origin is their common vertex and lies strictly on the inner side of the supporting lines of all the other edges. Therefore, $b_i = \langle \nu_i, z_{i-1} \rangle > 0$. Together with $\ell_i > 0$, this shows that $b \in \mathcal{B}$, and the resulting polygon is precisely $P(b)$. By (9), (16), and (15), we have

$$\lambda_i = w_{\alpha,\beta}(a_{i-1}) + \frac{w_{\alpha,\beta}(z_i)}{\ell_i} = w_{\alpha,\beta}(a_{i-1}) + \lambda_* - w_{\alpha,\beta}(a_{i-1}) = \lambda_*.$$

Thus, this polygon is a balanced quasi-Newton polygon.

Finally, the common local increment of any balanced quasi-Newton polygon must satisfy (17), and hence must equal λ_*. After normalization, its starting point and all its edge vectors are uniquely determined by (15). Hence, the balanced quasi-Newton polygon is unique up to positive scaling. □

Applying Theorem 2 to the balanced quasi-Newton polygon constructed above gives

$$\delta(\alpha, \beta) = \frac{\mathrm{area}(P)}{B(P)} = \frac{1}{\lambda_*}.$$

Thus, after substituting $k_i = \det(a_{i-1}, a_i)$, the feasible region is determined by the unique solution of the univariate equation

$$\boxed{\sum_{i=1}^{r} \frac{\alpha\beta \det(a_{i-1}, a_i)}{\big(\lambda - w_{\alpha,\beta}(a_{i-1})\big)\big(\lambda - w_{\alpha,\beta}(a_i)\big)} = 1, \qquad \lambda > d.}$$

This equation is uniquely determined by the vertices $a_0, \ldots, a_r$ of the Newton polygon.

The feasible region has the following two equivalent representations.

Representation via the unique real root. Computing the unique real root $\lambda_*(\alpha, \beta)$ of (17) in the interval (d, ∞) yields

$$\mathsf{FR}(f) = \left\{ (\alpha, \beta) \in \mathbb{R}^2_{>0} : \lambda_*(\alpha, \beta) < 1 \right\}.$$

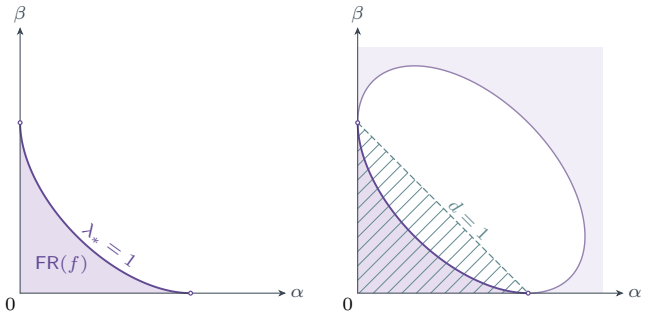
If this real root can be expressed in radicals, substituting its expression into the condition above gives an inequality in α, β involving radicals.

Representation via explicit inequalities. The feasible region can also be described directly by two explicit inequalities: $(\alpha, \beta) \in \mathsf{FR}(f)$ if and only if

$$d < 1 \qquad \text{and} \qquad \sum_{i=1}^{r} \frac{\alpha\beta k_i}{\big(1 - w_{\alpha,\beta}(a_{i-1})\big)\big(1 - w_{\alpha,\beta}(a_i)\big)} < 1.$$

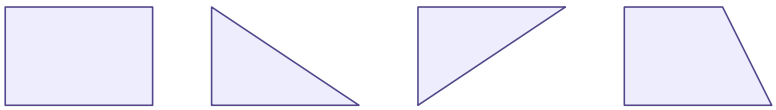
Indeed, the left-hand side of (17) is strictly decreasing on (d, ∞), and the unique real root of the equation satisfies $\lambda_* > d$. Consequently, $\lambda_* < 1$ holds if and only if $d < 1$ and the left-hand side evaluated at $\lambda = 1$ is less than 1. This establishes the equivalence of the two representations of the feasible region.

The figure below illustrates why these two representations are equivalent.



6 Explicit Feasible Regions for Representative Newton Polygons

In this section, we compute the feasible regions for four representative classes of Newton polygons and use them to assess the optimality of small-root bounds obtained in the literature for various cryptographic instances. Throughout this section, $\alpha, \beta > 0$. By the results of the previous section, it suffices to find the unique root λ_* of (17) with $\lambda > d$ and then use $\lambda_* < 1$ to determine the feasible region.



6.1 Rectangles

Let $N(f) = \mathrm{conv}\{0, (m, 0), (m, n), (0, n)\}$. Its vertex chain is $a_0 = (m, 0)$, $a_1 = (m, n)$, and $a_2 = (0, n)$, so $k_1 = k_2 = mn$ and $d = m\alpha + n\beta$. Substituting into (17) gives

$$\frac{mn\alpha\beta}{(\lambda - m\alpha)(\lambda - m\alpha - n\beta)} + \frac{mn\alpha\beta}{(\lambda - m\alpha - n\beta)(\lambda - n\beta)} = 1.$$

Clearing denominators and simplifying yields

$$\lambda\left[(\lambda - m\alpha - n\beta)^2 - mn\alpha\beta\right] = 0.$$

Since $\lambda > d = m\alpha + n\beta$, the required root is $\lambda_* = m\alpha + n\beta + \sqrt{mn\alpha\beta}$. Therefore, $(\alpha, \beta) \in \mathsf{FR}(f)$ if and only if

$$m\alpha + n\beta + \sqrt{mn\alpha\beta} < 1.$$

Cryptographic instances. We examine eight classes of cryptographic instances: RSA factoring with known most significant bits of a factor [8,10,11,12]; RSA factoring with known least significant bits of a factor [10]; factor approximations and approximations to both the product and its factors for $p^r q^s$ [23]; factoring with a known RSA private key [20,4]; factoring with a known CRT-RSA private key [19]; factor recovery for blinded RSA with small e [17]; and the RSA-polynomial problem after a change of variables [24]. As shown in Table 1, all the listed bounds from the literature match the optimal bounds established in this paper.

6.2 Axis-Intercept Triangles

Let $N(f) = \text{conv}\{0, (m, 0), (0, n)\}$. Here, $a_0 = (m, 0)$ and $a_1 = (0, n)$, so $k_1 = mn$ and $d = \max\{m\alpha, n\beta\}$. Substituting into (17) gives

$$\frac{mn\alpha\beta}{(\lambda - m\alpha)(\lambda - n\beta)} = 1,$$

or equivalently, $\lambda(\lambda - m\alpha - n\beta) = 0$. The root satisfying $\lambda > d$ is $\lambda_* = m\alpha + n\beta$. Therefore, $(\alpha, \beta) \in \mathsf{FR}(f)$ if and only if

$$m\alpha + n\beta < 1.$$

Cryptographic instances. We examine four classes of cryptographic instances: formulations over the integers for RSA with partially known messages and for RSA with short random padding [9,10,4]; direct integer constructions for recovering representations as sums of squares [16,21]; and the integer approach to Fermat factorization [14]. As shown in Table 2, all the listed bounds from the literature match the optimal bounds established in this paper.

6.3 Upper Triangles

Let $N(f) = \text{conv}\{0, (m, n), (0, n)\}$. Its vertex chain begins at the origin: $a_0 = 0$, $a_1 = (m, n)$, and $a_2 = (0, n)$. Thus, $k_1 = 0$, $k_2 = mn$, and $d = m\alpha + n\beta$. Substitution into (17) leaves only one nonzero term, giving

$$\frac{mn\alpha\beta}{(\lambda - m\alpha - n\beta)(\lambda - n\beta)} = 1.$$

Rearranging yields the quadratic equation

$$\lambda^2 - (m\alpha + 2n\beta)\lambda + n^2\beta^2 = 0.$$

The root satisfying $\lambda > d$ is

$$\lambda_* = \frac{m\alpha + 2n\beta + \sqrt{m^2\alpha^2 + 4mn\alpha\beta}}{2}.$$

Therefore, $(\alpha, \beta) \in \mathsf{FR}(f)$ if and only if

$$m\alpha + 2n\beta + \sqrt{m^2\alpha^2 + 4mn\alpha\beta} < 2.$$

Cryptographic instances. We examine six classes of cryptographic instances: factoring $p^r q$ with known most significant bits of a factor [5,4]; RSA factoring with known least significant bits of a factor [10,4]; factor approximations and product approximations for $p^r q^s$ [23]; factoring with a known RSA private key [4]; and formulations over the integers for Φ-hiding/PIR [7,15,4]. As shown in Table 3, all the listed bounds from the literature match the optimal bounds established in this paper under the corresponding parameter relations.

6.4 Nonrectangular Trapezoids

Let $N(f) = \text{conv}\{0, (m,0), (c,n), (0,n)\}$, where c is an integer satisfying $0 < c < m$. Its vertex chain is $a_0 = (m,0)$, $a_1 = (c,n)$, and $a_2 = (0,n)$, so $k_1 = mn$, $k_2 = cn$, and $d = \max\{m\alpha, c\alpha + n\beta\}$. Substituting into (17) gives

$$\frac{mn\alpha\beta}{(\lambda - m\alpha)(\lambda - c\alpha - n\beta)} + \frac{cn\alpha\beta}{(\lambda - c\alpha - n\beta)(\lambda - n\beta)} = 1.$$

Clearing denominators and canceling the positive factor λ yields

$$\lambda^2 - \big((m+c)\alpha + 2n\beta\big)\lambda + mc\alpha^2 + mn\alpha\beta + n^2\beta^2 = 0.$$

The discriminant of this quadratic is $(m-c)^2\alpha^2 + 4cn\alpha\beta$. Its smaller root is less than $c\alpha + n\beta \le d$, so the root satisfying $\lambda > d$ is

$$\lambda_* = \frac{(m+c)\alpha + 2n\beta + \sqrt{(m-c)^2\alpha^2 + 4cn\alpha\beta}}{2}.$$

Therefore, $(\alpha, \beta) \in \mathsf{FR}(f)$ if and only if

$$(m+c)\alpha + 2n\beta + \sqrt{(m-c)^2\alpha^2 + 4cn\alpha\beta} < 2.$$

Cryptographic instances. We examine the original construction for the RSA-polynomial problem [2] and its subsequent improvements [22,24]. As shown in Table 4, the bounds in [22, Theorem 7] and [24, Theorem 4] match the optimal bounds established in this paper under the corresponding parameter relations. The other two listed bounds fall short of the corresponding optimal bounds.

6.5 Algebraic Complexity of the Feasible Region: A Toy Example

Let $N(f) = \text{conv}\{0, (m,c), (m,n), (0,n)\}$, where m, n, c are positive integers and $0 < c < n$. Its vertex chain is $a_0 = 0$, $a_1 = (m,c)$, $a_2 = (m,n)$, and $a_3 = (0,n)$, so $k_1 = 0$, $k_2 = m(n-c)$, $k_3 = mn$, and $d = m\alpha + n\beta$. Substituting into (17) gives

$$\frac{m(n-c)\alpha\beta}{(\lambda - m\alpha - c\beta)(\lambda - m\alpha - n\beta)} + \frac{mn\alpha\beta}{(\lambda - m\alpha - n\beta)(\lambda - n\beta)} = 1.$$

Clearing denominators and simplifying yields

$$(\lambda - c\beta)(\lambda - d)^2 = m\alpha\beta(n\lambda - cd),$$

or equivalently, the cubic equation

$$\lambda^3 - (2d + c\beta)\lambda^2 + \big(d^2 + 2c\beta d - mn\alpha\beta\big)\lambda - cn\beta^2 d = 0.$$

The signs of this cubic polynomial at $0, c\beta, d$ are negative, positive, and negative, respectively. Hence, its three real roots lie in $(0, c\beta)$, $(c\beta, d)$, and (d, ∞), respectively.

Define

$$\Delta_0 = (d - c\beta)^2 + 3mn\alpha\beta,$$
$$\Delta_1 = 2(d - c\beta)\big(9mn\alpha\beta - (d - c\beta)^2\big) - 27m^2c\alpha^2\beta.$$

By the cubic formula, the root satisfying $\lambda > d$ is

$$\lambda_* = \frac{2d + c\beta}{3} + \frac{1}{3}\left(\sqrt[3]{\frac{\Delta_1 + i\sqrt{4\Delta_0^3 - \Delta_1^2}}{2}} + \sqrt[3]{\frac{\Delta_1 - i\sqrt{4\Delta_0^3 - \Delta_1^2}}{2}}\right).$$

Here, both cube roots are taken to be their principal complex values, and the two values are complex conjugates. Therefore, $(\alpha, \beta) \in \mathsf{FR}(f)$ if and only if $\lambda_* < 1$, or equivalently,

$$2d + c\beta + 2\sqrt{\Delta_0}\,\cos\left(\frac{1}{3}\arccos\frac{\Delta_1}{2\Delta_0^{3/2}}\right) < 3.$$

This shows that the algebraic complexity of the optimal small-root bounds for the bivariate Coppersmith method over the integers may be higher than previously expected. Even when the Newton polygon has only four vertices, determining the bound may require solving an irreducible cubic equation, whereas the typical examples above lead to equations of degree at most two.

Table 1: Cryptographic instances and small-root bounds for rectangular Newton polygons

Cryptographic instance	**Integer polynomial** $f(x, y)$	**Published bound**	**Optimal?**
Known most significant bits of an RSA factor [8,10,11,12]	$(P + x)(Q + y) - N$	$\alpha = \beta$, $\alpha < 1/3$	Yes
Known least significant bits of an RSA factor [10]	$\dfrac{(Bx + p_0)(By + q_0) - N}{B}$	$\alpha = \beta$, $\alpha < 1/3$	Yes
Factor approximations for $p^r q^s$ [23, Theorem 4]	$(P + x)^r(Q + y)^s - N$	$\alpha = \beta$, $\alpha < \dfrac{1}{r + s + \sqrt{rs}}$	Yes

Table 1 (continued)

Cryptographic instance	**Integer polynomial** $f(x, y)$	**Published bound**	**Optimal?**
Product and factor approximations for $p^r q^s$, $r > s$ [23, Eq. (47)]	$(PQ + x)^s (P + y)^{r-s} - N$	$(r + s - 1)\alpha = (r + s - 2)\beta + 1,$ $s\alpha + (r - s)\beta + \sqrt{s(r - s)\alpha\beta} < 1$	Yes
Known RSA private key [20,4]	$(N - x)(\widetilde{k} + y) + 1 - ed$	$2\alpha + \beta = 1,$ $1/3 < \alpha < 1/2$	Yes
Known CRT-RSA private key [19, Theorem 1]	$\underline{(E + x)(F + y) - Nxy}$ g	$\alpha + \beta + \sqrt{\alpha\beta} < 1$	Yes
Factor recovery for blinded RSA (small e) [17, Theorem 1]	$\underline{(ex + p_e)(ey + q_e) - N}$ e	$\alpha = \beta,$ $\alpha < 1/3$	Yes
RSA-polynomial (after a change of variables) [24, Theorem 5]	$8xy - (e_0 + e_2)x + e_2 y + e_1$	$\alpha = \beta,$ $\alpha < 1/3$	Yes

Table 2: Cryptographic instances and small-root bounds for axis-intercept triangles

Cryptographic instance	**Integer polynomial** $f(x, y)$ $[\cdot]_N$: coefficientwise reduction modulo N	**Published bound**	**Optimal?**
RSA with a partially known message (integer formulation) [9,10,4]	$\left[(B + x)^e - c\right]_N - Ny$	$\beta = e\alpha,$ $\alpha < \frac{1}{2e}$	Yes
RSA with short random padding (integer formulation) [9,10,4]	$\left[\mathrm{Res}_z\left(z^e - c_1, (z + x)^e - c_2\right)\right]_N - Ny$	$\beta = e^2\alpha,$ $\alpha < \frac{1}{2e^2}$	Yes

Table 2 (continued)

Cryptographic instance	**Integer polynomial** $f(x,y)$ $[\cdot]_N$: coefficientwise reduction modulo N	**Published bound**	**Optimal?**
Recovery of sum-of-squares representations (direct integer construction) [16,21]	$r(A+x)^2 + s(B+y)^2 - N$ fixed $r,s>0,\ A,B \asymp N^{1/2}$	$\alpha=\beta,$ $\alpha<1/4$	Yes
Fermat factorization (integer approach) [14]	$(x+2R)^2 - y^2 - 4N,$ $R=\lceil\sqrt{N}\rceil, \quad N=pq$	$\lvert p-q\rvert < N^{1/4}$	Yes $(\alpha\to 0^+)$

Table 3: Cryptographic instances and small-root bounds for upper triangular Newton polygons

Cryptographic instance	**Integer polynomial** $f(x,y)$	**Published bound**	**Optimal?**
Known most significant bits of a factor of $p^r q$, $r \geq 1$ [5,4]	$(P+x)^r y - N$	$\beta = \dfrac{1}{r+1},$ $\alpha < \dfrac{r}{(r+1)^2}$	Yes
Known least significant bits of an RSA factor (integer formulation) [10,4]	$(Bx+p_0)y - N$ $B=2^k,\ p\equiv p_0 \pmod B$	$\beta = 1/2,$ $\alpha < 1/4$	Yes
Factor approximations for $p^r q^s$ $r\geq s,\ \gcd(r,s)=1$ [23, Eq. (54)]	$(P+x)^r y^s - N$	$\beta = \dfrac{1}{r+s},$ $\alpha < \dfrac{r}{(r+s)^2}$	Yes
Product approximations for $p^r q^s$ $s<r<3s,$ $\gcd(r,s)=1$ [23, Eq. (45)]	$(PQ+x)^s y^{r-s} - N$	$\beta = \dfrac{1}{r+s},$ $\alpha < \dfrac{4s}{(r+s)^2}$	Yes

Table 3 (continued)

Cryptographic instance	**Integer polynomial** $f(x, y)$	**Published bound**	**Optimal?**
Known RSA private key [4, Section 4.3]	$(N - x)y + 1 - ed$	$2\alpha + \beta = 1$, $1/4 < \alpha < 1/2$	Yes
Φ-hiding / PIR (integer formulation) [7,15,4]	$(\ell x + 1)y - N$ known ℓ with $\ell \mid p - 1$	$\beta = 1/2$, $\alpha < 1/4$	Yes

Table 4: Comparison of published bounds for the RSA-polynomial problem

$f(x, y) = 8x^2 - e_0x + e_1 + (8x + e_2)y$		
Reference and construction	**Published bound**	**Matches our optimal bound?**
Original construction [2,22]	$2\alpha + \beta < 1/2$	No
[22, Theorem 7]	$\alpha = \beta$, $\alpha < \frac{5 - \sqrt{5}}{10}$	Yes
[22, Theorem 8]	$\alpha = \beta$, $\alpha < 3/11$	No
Allowing unbalanced primes [24, Theorem 4]	$\beta \geq \alpha$, $3\alpha + 2\beta + \sqrt{\alpha^2 + 4\alpha\beta} < 2$	Yes

References

1. Aardal, M.A., Aranha, D.F., Feng, Y., Gao, Y., Pan, Y.: Better bounds for finding fixed-degree isogenies via Coppersmith's method. In: Daemen, J., Thomé, E. (eds.) Advances in Cryptology – EUROCRYPT 2026. Lecture Notes in Computer Science, vol. 16544, pp. 541–572. Springer (2026). https://doi.org/10.1007/978-3-032-25327-9_19
2. Bagherpour, B.: A bivariate polynomial-based cryptographic hard problem and its applications. Designs, Codes and Cryptography **91**(8), 2723–2735 (2023). https://doi.org/10.1007/s10623-023-01229-1
3. Benčina, B., Kutas, P., Merz, S.P., Petit, C., Stopar, M., Weitkämper, C.: Improved algorithms for finding fixed-degree isogenies between supersingular elliptic curves. In: Reyzin, L., Stebila, D. (eds.) Advances in Cryptology – CRYPTO 2024. Lecture Notes in Computer Science, vol. 14924, pp. 183–217. Springer (2024). https://doi.org/10.1007/978-3-031-68388-6_8

4. Blömer, J., May, A.: A tool kit for finding small roots of bivariate polynomials over the integers. In: Cramer, R. (ed.) Advances in Cryptology – EUROCRYPT 2005. Lecture Notes in Computer Science, vol. 3494, pp. 251–267. Springer (2005). https://doi.org/10.1007/11426639_15
5. Boneh, D., Durfee, G., Howgrave-Graham, N.: Factoring $N = p^r q$ for large r. In: Advances in Cryptology – CRYPTO '99. Lecture Notes in Computer Science, vol. 1666, pp. 326–337. Springer (1999). https://doi.org/10.1007/3-540-48405-1_21
6. Brøndsted, A.: An Introduction to Convex Polytopes, Graduate Texts in Mathematics, vol. 90. Springer-Verlag, New York (1983). https://doi.org/10.1007/978-1-4612-1148-8
7. Cachin, C., Micali, S., Stadler, M.: Computationally private information retrieval with polylogarithmic communication. In: Advances in Cryptology – EUROCRYPT '99. Lecture Notes in Computer Science, vol. 1592, pp. 402–414. Springer (1999). https://doi.org/10.1007/3-540-48910-X_28
8. Coppersmith, D.: Finding a small root of a bivariate integer equation; factoring with high bits known. In: Maurer, U. (ed.) Advances in Cryptology – EUROCRYPT '96. Lecture Notes in Computer Science, vol. 1070, pp. 178–189. Springer (1996). https://doi.org/10.1007/3-540-68339-9_16
9. Coppersmith, D.: Finding a small root of a univariate modular equation. In: Maurer, U. (ed.) Advances in Cryptology – EUROCRYPT '96. Lecture Notes in Computer Science, vol. 1070, pp. 155–165. Springer (1996). https://doi.org/10.1007/3-540-68339-9_14
10. Coppersmith, D.: Small solutions to polynomial equations, and low exponent RSA vulnerabilities. Journal of Cryptology **10**(4), 233–260 (1997). https://doi.org/10.1007/s001459900030
11. Coron, J.S.: Finding small roots of bivariate integer polynomial equations revisited. In: Cachin, C., Camenisch, J.L. (eds.) Advances in Cryptology – EUROCRYPT 2004. Lecture Notes in Computer Science, vol. 3027, pp. 492–505. Springer (2004). https://doi.org/10.1007/978-3-540-24676-3_29
12. Coron, J.S.: Finding small roots of bivariate integer polynomial equations: A direct approach. In: Advances in Cryptology – CRYPTO 2007. Lecture Notes in Computer Science, vol. 4622, pp. 379–394. Springer (2007). https://doi.org/10.1007/978-3-540-74143-5_21
13. Coron, J.S., Kirichenko, A., Tibouchi, M.: A note on the bivariate Coppersmith theorem. Journal of Cryptology **26**(2), 246–250 (2013). https://doi.org/10.1007/s00145-012-9121-x
14. Erra, R., Grenier, C.: The Fermat factorization method revisited. Cryptology ePrint Archive, Paper 2009/318 (2009), https://eprint.iacr.org/2009/318, revised version, 24 July 2009
15. Gentry, C., Ramzan, Z.: Single-database private information retrieval with constant communication rate. In: Automata, Languages and Programming – ICALP 2005. Lecture Notes in Computer Science, vol. 3580, pp. 803–815. Springer (2005). https://doi.org/10.1007/11523468_65
16. Gutierrez, J., Ibeas, Á., Joux, A.: Recovering a sum of two squares decomposition. Journal of Symbolic Computation **64**, 16–21 (2014). https://doi.org/10.1016/j.jsc.2013.12.003
17. Jiang, Z., Zhou, Y., Liu, Y.: New partial key exposure attacks on RSA with additive exponent blinding. Cybersecurity **7**, 26 (2024). https://doi.org/10.1186/s42400-024-00214-y

18. Jochemsz, E., May, A.: A strategy for finding roots of multivariate polynomials with new applications in attacking RSA variants. In: Lai, X., Chen, K. (eds.) Advances in Cryptology – ASIACRYPT 2006. Lecture Notes in Computer Science, vol. 4284, pp. 267–282. Springer (2006). https://doi.org/10.1007/11935230_18
19. Maitra, S., Sarkar, S.: On deterministic polynomial-time equivalence of computing the CRT-RSA secret keys and factoring. Defence Science Journal **62**(2), 122–126 (2012). https://doi.org/10.14429/dsj.62.1716
20. May, A.: Computing the RSA secret key is deterministic polynomial time equivalent to factoring. In: Franklin, M. (ed.) Advances in Cryptology – CRYPTO 2004. Lecture Notes in Computer Science, vol. 3152, pp. 213–219. Springer (2004). https://doi.org/10.1007/978-3-540-28628-8_13
21. Zhang, X., Wang, L.P., Xu, J., Hu, L., Peng, L., Huang, Z., Liu, Z.: Recovering a sum of two squares decomposition revisited. In: Lin, D., Wang, X., Yung, M. (eds.) Information Security and Cryptology – Inscrypt 2015. Lecture Notes in Computer Science, vol. 9589, pp. 178–192. Springer (2016). https://doi.org/10.1007/978-3-319-38898-4_11
22. Zheng, M.: Revisiting RSA-polynomial problem and semiprime factorization. Theoretical Computer Science **1004**, 114634 (2024). https://doi.org/10.1016/j.tcs.2024.114634
23. Zheng, M., Chen, Z., Wu, Y.: Solving generalized bivariate integer equations and its application to factoring with known bits. IEEE Access **11**, 34674–34684 (2023). https://doi.org/10.1109/ACCESS.2023.3264590
24. Zheng, M., Nitaj, A.: A more complete cryptanalysis of the RSA-polynomial problem. Theoretical Computer Science **1070**, 115837 (2026). https://doi.org/10.1016/j.tcs.2026.115837

A Supplementary Proofs

A.1 Proof of Lemma 2

Proof. By definition, for every $t > 0$, we have $\delta(t\alpha, t\beta) = t^{-1}\delta(\alpha, \beta)$. Thus, δ is homogeneous of degree -1. If $\alpha' \geq \alpha$ and $\beta' \geq \beta$, then the denominator of each ratio in the definition of δ does not decrease, and hence $\delta(\alpha', \beta') \leq \delta(\alpha, \beta)$. Therefore, δ is nonincreasing in each variable.

Next, we prove continuity. Fix $(\alpha, \beta) \in \mathbb{R}^2_{>0}$. For any $0 < \varepsilon < 1$, whenever $|\alpha' - \alpha| < \varepsilon\alpha$ and $|\beta' - \beta| < \varepsilon\beta$, monotonicity and homogeneity give

$$\frac{\delta(\alpha, \beta)}{1 + \varepsilon} \leq \delta(\alpha', \beta') \leq \frac{\delta(\alpha, \beta)}{1 - \varepsilon}.$$

Since $\delta(\alpha, \beta)$ is finite, these bounds show that $\delta(\alpha', \beta') \to \delta(\alpha, \beta)$ as $(\alpha', \beta') \to (\alpha, \beta)$. Thus, δ is continuous.

Finally, we prove convexity. For a fixed S, let δ_S denote the corresponding ratio in the definition of δ. Its denominator is a linear function of (α, β) that is positive on $\mathbb{R}^2_{>0}$. Since $u \mapsto u^{-1}$ is convex on $\mathbb{R}_{>0}$, for any $p, q \in \mathbb{R}^2_{>0}$ and $\lambda \in [0, 1]$, we have

$$\begin{aligned}\delta_S\big(\lambda p + (1 - \lambda)q\big) &\leq \lambda\delta_S(p) + (1 - \lambda)\delta_S(q) \\ &\leq \lambda\delta(p) + (1 - \lambda)\delta(q).\end{aligned}$$

Taking the supremum over S yields

$$\delta\big(\lambda p + (1-\lambda)q\big) \le \lambda\delta(p) + (1-\lambda)\delta(q).$$

Therefore, δ is convex.

A.2 Proof of Corollary 1

Proof. Let p, q be any two points on the curve. For every $\lambda \in [0,1]$, we have $\lambda p + (1-\lambda)q \in \mathbb{R}^2_{>0}$. By the convexity of δ,

$$\delta\big(\lambda p + (1-\lambda)q\big) \le \lambda\delta(p) + (1-\lambda)\delta(q) = 1.$$

The conclusion follows from $\mathsf{FR} = \{(\alpha,\beta) \in \mathbb{R}^2_{>0} : \delta(\alpha,\beta) > 1\}$.