Coppersmith's Method for Bivariate Modular Equations: Optimal Small-Root Bounds

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. In 1996, Coppersmith established the classical small-root bound for univariate modular equations, which was subsequently proved to be asymptotically optimal within the standard Coppersmith framework. In contrast, a general theoretical characterization of optimal small-root bounds for bivariate modular equations, together with a method for computing them, has remained elusive. The main difficulty is that the additional degrees of freedom introduced by a second variable turn shift selection into a complex combinatorial optimization problem. We prove that, for standard triangular shift constructions in Coppersmith's method, optimal small-root bounds for bivariate modular equations admit a unified characterization and can be computed effectively. To this end, we introduce a class of geometric objects called quasi-Newton polygons, which provide an exact geometric characterization of these optimal bounds and yield a method for computing them. This result extends the theory of optimal small-root bounds for Coppersmith's method to the general bivariate setting, marking its first breakthrough beyond the univariate case in three decades.

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

1 Introduction

In 1996, Coppersmith introduced methods based on lattice reduction for finding small roots of univariate modular equations [16] and bivariate polynomial equations over the integers [15]. Over the past three decades, Coppersmith's method has been extensively developed and has found widespread applications in cryptanalysis [12,11,19,13,5,6,31,27]. It has also been used in security proofs for cryptographic schemes [29,10,25,4,35] and has become one of the central tools in algebraic cryptanalysis.

Determining the range of small roots that can be recovered is a fundamental problem in the study of Coppersmith's method. In 1996, Coppersmith [16]

established the classical bound $X < N^{1/d}$ for small roots of a degree-d univariate polynomial congruence modulo N. Subsequently, Aono et al. [2] proved that no standard set of shifts can yield a better asymptotic bound. Using capacity theory, Chinburg et al. [14] further established a stronger nonexistence result: within Coppersmith's standard auxiliary-polynomial framework, for every $\varepsilon > 0$, there exists no auxiliary polynomial that extends the asymptotic bound to $N^{1/d+\varepsilon}$. Since this result is independent of the particular choice of shifts, the lattice reduction algorithm, and the short-vector estimate, it establishes the asymptotic optimality of the classical exponent $1/d$ within the framework of Coppersmith's method.

Unlike the univariate case, determining small-root bounds for bivariate and multivariate Coppersmith methods is considerably more challenging, as these bounds generally cannot be characterized solely by the degree of the polynomial. For instance, consider a bivariate polynomial equation $f(x, y) = 0$ over the integers, where the root (x_0, y_0) to be recovered satisfies $|x_0| \leq X$ and $|y_0| \leq Y$. Let $W = \|f(Xx, Yy)\|_\infty$. For polynomials of total degree at most δ, Coppersmith [15] established the classical asymptotic bound $XY < W^{2/(3\delta)}$ for recovering such roots. Subsequently, Blömer and May [9] investigated this problem from the perspective of Newton polytopes, showing that small-root bounds depend not only on the degree of the polynomial but also on its monomial structure. For example, for polynomials of total degree δ with a lower-triangular Newton polygon, they obtained the stronger asymptotic bound $XY < W^{1/\delta}$.

For multivariate modular equations, traditional applications of Coppersmith's method typically require manually selecting shift sets tailored to specific polynomials and deriving the corresponding asymptotic small-root bounds. When the underlying polynomial changes, the lattice construction and its analysis often need to be carried out again. This problem-specific approach is widely encountered in cryptanalysis involving partial exposure of RSA private keys [12,8,1,30], small private and CRT exponents [11,23,26,7,31], hidden number problems [13,32,34], and pseudorandom number generators [22,5,33,21].

To reduce reliance on problem-specific manual design, Jochemsz and May proposed a general strategy for constructing shift polynomials [24]. More recently, Meers and Nowakowski introduced Automated Coppersmith, which automates the selection of shift polynomials for systems of multivariate modular equations [27]. Ryan studied shift construction from the perspective of polynomial ideals and Gröbner bases [28]; Feng et al. used sumset theory and Newton polytopes to automate the computation of asymptotic bounds for Jochemsz–May-type constructions [20]; and Ding et al. reduced the computation of asymptotic bounds for Automated Coppersmith to a linear programming problem [18].

However, these methods typically assume that the monomial set S is fixed in advance, and address only how to construct shifts over a given S and compute the resulting small-root bound. Existing work still lacks a general method for finding a globally optimal S among all admissible monomial sets and for proving the optimality of the corresponding small-root bound. The fundamental difficulty is that, when passing from the univariate to the bivariate setting, candidate

monomial sets are no longer one-dimensional intervals but two-dimensional lattice-point regions, so determining the optimal bound becomes a substantially more intricate combinatorial and geometric optimization problem.

This naturally raises a fundamental question: Can we develop a unified and provable method for computing the optimal small-root bounds for arbitrary polynomials in the bivariate and multivariate settings, and thereby determine whether the bounds established in previous works are optimal? This paper addresses precisely this question.

Our Contributions. Thirty years ago, Coppersmith [16] established the classical small-root bound for univariate modular equations, which was subsequently proved to be asymptotically optimal within the standard Coppersmith framework. Three decades later, we revisit the problem of finding optimal small-root bounds for bivariate modular equations. We prove that, within the framework of standard triangular shifts, which encompasses nearly all lattice constructions considered in the existing literature, the optimal small-root bound for the bivariate case can likewise be characterized in a unified manner and computed effectively. This optimal bound is precisely characterized by a class of geometric objects that we introduce, called "quasi-Newton polygons". More specifically, our proof proceeds as follows:

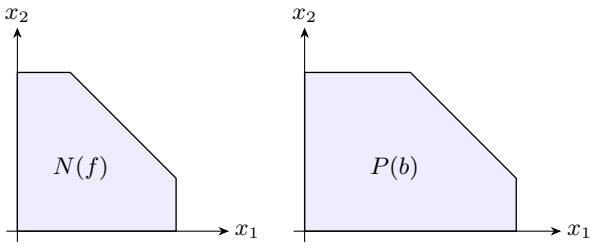
1. Problem Reduction. For any triangular lattice construction generated by the standard shifts $x^{i_1}y^{i_2}f^jN^{m-j}$, with small-root bounds $X = N^\alpha$ and $Y = N^\beta$, we prove that finding the optimal small-root bound reduces to determining

$$\sup_{S\subseteq\mathbb{Z}^2_{\geq 0}} \frac{\#\{(u,j)\in\mathbb{Z}^2_{\geq 0}\times\mathbb{Z}_{\geq 1} : u + jA(f)\subseteq S\}}{\sum\limits_{(s_1,s_2)\in S}(\alpha s_1+\beta s_2)}.$$

Here, $A(f)$ denotes the set of exponent vectors of the monomials appearing in f.

2. Weighted Counting on Interval Regions. To handle arbitrary monomial sets S in a unified manner, we introduce interval regions and prove that the number of integer intervals fully contained in a finite integer set T can be bounded by a weighted sum over T. This allows us to reduce the counting of $u + jA(f)$ to one-dimensional integer interval counting, providing a key tool for establishing a uniform upper bound valid for arbitrary S.

3. Balanced Quasi-Newton Polygons. We first define quasi-Newton polygons $P(b)$.



And construct an interval region on each non-coordinate edge to describe the sets $u + jN(f)$ classified by their contact edges. We then introduce a balance condition requiring that the ratio of the area of each interval region to the weighted integral along the corresponding edge equals a common constant c. We prove that this condition yields a uniform upper bound

$$\#\big\{(u,j) \in \mathbb{Z}_{\geq 0}^2 \times \mathbb{Z}_{\geq 1} : u + jA(f) \subseteq S\big\} \leq c \sum_{(s_1,s_2)\in S} (\alpha s_1 + \beta s_2)$$

for every finite monomial set S. Moreover, when $S = mP \cap \mathbb{Z}_{\geq 0}^2$, the corresponding ratio asymptotically approaches c. Thus, balanced quasi-Newton polygons precisely characterize the optimal small-root bound.

4. Existence and Computation. We further prove that, for any $\alpha, \beta > 0$, there exists a quasi-Newton polygon satisfying the balance condition. Moreover, the balance condition can be reduced to a finite system of algebraic equations. By solving these equations, we can determine the balanced polygon and its balance constant, thereby computing the optimal small-root bound.

We thus establish a complete theoretical framework for computing the optimal small-root bounds of the bivariate Coppersmith method under standard triangular shifts. This overcomes the longstanding lack of a unified optimality theory in this area and marks the first breakthrough in extending Coppersmith's optimal small-root bound theory from the univariate to the multivariate setting.

Notation. We use the following notation and conventions throughout the paper. For a nonzero polynomial $f(x,y) = \sum_{(a,b)\in\mathbb{Z}_{\geq 0}^2} c_{a,b}x^a y^b \in \mathbb{Z}[x,y]$, its support is defined as $A(f) = \{(a,b) \in \mathbb{Z}_{\geq 0}^2 : c_{a,b} \neq 0\}$, and its Newton polygon is $N(f) = \mathrm{conv}(A(f))$, where conv denotes the convex hull. The exponent vector (a,b) corresponds to the monomial $x^a y^b$.

For subsets $A, B \subseteq \mathbb{R}^2$, their Minkowski sum is defined by $A + B = \{a + b : a \in A,\ b \in B\}$. For an integer $j \geq 1$, we write $jA = \underbrace{A + \cdots + A}_{j \text{ times}}$, with $0A = \{0\}$.

For a finite set S, we use $\#S$ or $|S|$ to denote its cardinality. For vectors u, v, the notation $\langle u, v\rangle$ and $\|u\|$ denotes the standard inner product and Euclidean norm, respectively. Vector inequalities $u \leq v$ are understood componentwise. We use e_i to denote the i-th standard basis vector in the corresponding Euclidean space. For $u = (u_1, u_2)$ and $v = (v_1, v_2)$, we define $\det(u, v) = u_1 v_2 - u_2 v_1$. Finally, $\mathrm{vol}_d(C)$ denotes the usual d-dimensional Euclidean volume of a set C. In particular, vol_2 and vol_3 denote area and three-dimensional volume, respectively.

2 Bivariate Coppersmith method.

2.1 Basic Framework

Definition 1 (Bivariate Coppersmith Problem). *Given an integer $N \geq 2$, a bivariate polynomial $f(x,y) \in \mathbb{Z}[x,y]$, and bounds $X, Y > 0$, the bivariate Coppersmith problem is to find all integer pairs (x_0, y_0) satisfying*

$$f(x_0, y_0) \equiv 0 \pmod{N}, \qquad |x_0| < X, \qquad |y_0| < Y.$$

Throughout this paper, we assume that $f(0,0) \neq 0$, that $N(f)$ is two-dimensional, and that $A(f^j) = jA(f)$ for every integer $j \geq 1$.

We briefly review the basic framework of Coppersmith's method for finding small roots of bivariate modular equations. Further details can be found in [17,11,24].

Fix a positive integer m, and consider the standard shift polynomials widely used in the literature:

$$g(x,y) = x^{i_1} y^{i_2} f(x,y)^j N^{m-j}, \quad i_1, i_2 \geq 0, \quad 0 \leq j \leq m. \tag{1}$$

Each such shift satisfies $g(x_0, y_0) \equiv 0 \pmod{N^m}$. We select finitely many of these shifts and construct a lattice L generated by the coefficient vectors of the corresponding polynomials $g(xX, yY)$. In this paper, we consider lattice constructions satisfying the following triangularity condition.

Definition 2 (Standard triangular shift construction). *If the coefficient matrix formed from a collection of shifts of the form* (1) *can be transformed, by permuting its rows and columns, into a triangular square matrix with nonzero diagonal entries, then the resulting construction is called a* standard triangular shift construction.

For example, let $f = x^2 + xy + y^2 + 1$, and take the shifts N, xN, x^2N, y^2N, f. Ordering the columns by $1, x, x^2, y^2, xy$ gives the lattice basis matrix

$$\begin{pmatrix} N & 0 & 0 & 0 & 0 \\ 0 & NX & 0 & 0 & 0 \\ 0 & 0 & NX^2 & 0 & 0 \\ 0 & 0 & 0 & NY^2 & 0 \\ 1 & 0 & X^2 & Y^2 & XY \end{pmatrix}.$$

Note that the ordering of the monomials in the above example is not induced by any monomial order. Therefore, unlike Meers and Nowakowski [27], who construct a triangular square matrix with respect to a prescribed monomial order, we require only the existence of some permutation under which the resulting coefficient matrix is triangular.

We then apply LLL reduction to the lattice L. In the asymptotic analysis, if the determinant condition

$$\big(\det L\big)^{1/\dim L} < N^m \tag{2}$$

is satisfied, we can obtain the auxiliary polynomials $h_1(x,y), \ldots, h_s(x,y)$ such that (x_0, y_0) is a common root not only modulo N^m but also over the integers.

However, these auxiliary polynomials do not necessarily suffice to recover (x_0, y_0) directly. One typically considers the ideal they generate:

$$\mathfrak{a} = \langle h_1, \ldots, h_s \rangle \subseteq \mathbb{Q}[x,y].$$

If $\mathfrak{a}$ is zero-dimensional, the common zeros of these polynomials can be computed using Gröbner bases or resultants. Since this property cannot currently be

guaranteed in general, the multivariate Coppersmith method typically relies on the following heuristic assumption.

Assumption 1. *The auxiliary polynomials obtained by the multivariate Coppersmith method generate a zero-dimensional ideal in* $\mathbb{Q}[x, y]$.

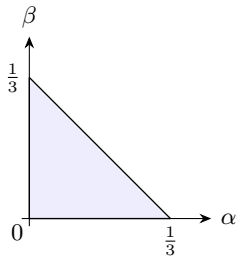
Definition 3 (Optimal Small-Root Bound Problem). *Let* $X = N^\alpha$ *and* $Y = N^\beta$, *where* $\alpha, \beta > 0$. *If there exists a standard triangular shift construction such that condition 2 is satisfied, then the exponent pair* (α, β) *is said to be feasible. The optimal small-root bound problem is to determine the feasible region* $\mathsf{FR}(f)$ *consisting of all feasible exponent pairs.*

For example, when the RSA public exponent is 3 and the plaintext contains two unknown blocks, recovering the plaintext can be reduced to finding small roots of the bivariate congruence

$$f(x, y) = (B + 2^s x + y)^3 - c \equiv 0 \pmod{N}.$$

Coppersmith [17] established the asymptotic small-root bound $XY < N^{1/3}$. A natural question is whether this bound is optimal; equivalently, whether

$$\mathrm{FR} \stackrel{?}{=} \left\{(\alpha, \beta) \in \mathbb{R}^2_{>0} : \alpha + \beta < \tfrac{1}{3}\right\}.$$



In this paper, we determine the feasible regions for this class of bivariate Coppersmith problems, thereby establishing whether the previously known small-root bounds are optimal.

2.2 Reduction to a Combinatorial Optimization Problem

Consider an arbitrary standard triangular shift construction $G = \{g_1, \ldots, g_\omega\}$ from Definition 2, where $g_i = x^{u_i} y^{v_i} f^{j_i} N^{m-j_i}$. Let $S \subseteq \mathbb{Z}^2_{\geq 0}$ be the set of exponent vectors of the monomials corresponding to the columns of the lattice basis matrix, where $s = (s_1, s_2) \in S$ corresponds to the monomial $x^{s_1} y^{s_2}$. Since the lattice basis matrix is triangular and square, $\omega = |S| = \dim L$.

Let $J(G) = \sum_{i=1}^{\omega} j_i$, and define the weight $w_{\alpha,\beta}(s) = \alpha s_1 + \beta s_2$. Under the convention that all diagonal coefficients are normalized, the diagonal entries of the triangular lattice basis matrix involve only powers of N, X, and Y. Hence,

$$\det L = N^{m\omega - J(G)} \prod_{s=(s_1,s_2)\in S} X^{s_1} Y^{s_2}.$$

Substituting $X = N^\alpha$ and $Y = N^\beta$, the determinant condition (2) is equivalent to

$$\sum_{s \in S} w_{\alpha,\beta}(s) < J(G). \tag{3}$$

For a fixed S, the left-hand side of (3) is independent of the specific choice of shifts. Therefore, the problem reduces to maximizing $J(G)$ over all standard triangular shift constructions whose column exponent set is S. To characterize this maximum, we introduce the following definition.

$$\mathcal{H}(S) = \big\{(u,j) \in \mathbb{Z}_{\geq 0}^2 \times \mathbb{Z}_{\geq 1} : \ u + jA(f) \subseteq S\big\}, \qquad H(S) = \#\mathcal{H}(S).$$

Lemma 1. *Fix a finite nonempty set $S \subseteq \mathbb{Z}_{\geq 0}^2$. For sufficiently large m, we have $\max_G J(G) = H(S)$, where the maximum is taken over all standard triangular shift constructions with column exponent set S.*

Proof. We first prove the upper bound. Fix an arbitrary construction G and an integer $j \geq 1$. For each shift g_i satisfying $j_i \geq j$, divide out the factors $N^{m-j_i} f^j$ to obtain $q_i = x^{u_i} y^{v_i} f^{j_i - j}$. Since the original shifts are linearly independent over $\mathbb{Q}$, so are these polynomials q_i.

If a is the exponent vector of a monomial appearing in some q_i, then $a \in (u_i, v_i) + (j_i - j)A(f)$, and hence

$$a + jA(f) \subseteq (u_i, v_i) + j_i A(f) \subseteq S.$$

Here, the second inclusion follows from $A(f^{j_i}) = j_i A(f)$ and the fact that all monomial exponents of g_i belong to S. Therefore, these polynomials q_i all lie in the vector space spanned by the monomials whose exponent vectors satisfy $a + jA(f) \subseteq S$. By linear independence, $\#\{i : j_i \geq j\}$ is at most the dimension of this space, namely, the number of exponent vectors a satisfying this condition. Summing over $j \geq 1$ gives

$$\begin{aligned} J(G) &= \sum_{j \geq 1} \#\{i : j_i \geq j\} \\ &\leq \sum_{j \geq 1} \#\big\{a \in \mathbb{Z}_{\geq 0}^2 : a + jA(f) \subseteq S\big\} = H(S). \end{aligned}$$

We next construct a shift set attaining this upper bound. Fix a monomial order, and let $\gamma = (\gamma_1, \gamma_2) \neq 0$ be the exponent vector of the leading monomial of f. For each $s \in S$, we choose a feasible shift with leading exponent s and with the largest possible power of f. More precisely, define

$$r(s) = \max\big\{j \in \mathbb{Z}_{\geq 0} : s - j\gamma + jA(f) \subseteq S\big\}.$$

The value $j = 0$ is always feasible. Moreover, since $0 \in A(f)$, every feasible j satisfies $s - j\gamma \in S$. Because $\gamma \neq 0$ and S is finite, $r(s)$ is well defined.

Choose $m \geq \max_{s \in S} r(s)$. For each $s = (s_1, s_2) \in S$, define

$$g_s^* = x^{s_1 - r(s)\gamma_1} y^{s_2 - r(s)\gamma_2} f^{r(s)} N^{m - r(s)}.$$

All monomial exponents of this shift belong to S, and the exponent of its leading monomial is exactly s. Therefore, after ordering the rows and columns according to the chosen monomial order, the coefficient matrix of $G^* = \{g_s^* : s \in S\}$ is triangular with nonzero diagonal entries, and its column exponent set is exactly S. Thus, G^* is a standard triangular shift construction.

It remains to verify that $J(G^*) = H(S)$. Since $\gamma \in A(f)$, for every $j \geq 1$,

$$s - (j-1)\gamma + (j-1)A(f) \subseteq s - j\gamma + jA(f).$$

Thus, for each fixed s, the feasible powers are precisely $0, 1, \ldots, r(s)$. Consequently, for every $k \geq 1$, $r(s) \geq k$ if and only if $s - k\gamma + kA(f) \subseteq S$. Since $0, k\gamma \in kA(f)$, the translation $s \mapsto s - k\gamma$ induces a bijection

$$\{s \in S : r(s) \geq k\} \longleftrightarrow \big\{u \in \mathbb{Z}^2_{\geq 0} : u + kA(f) \subseteq S\big\},$$

whose inverse is $u \mapsto u + k\gamma$. Therefore,

$$J(G^*) = \sum_{s \in S} r(s) = \sum_{k \geq 1} \#\{s \in S : r(s) \geq k\} = H(S).$$

Combining this equality with the upper bound yields $\max_G J(G) = H(S)$. □

Therefore, for a fixed S, the shift-selection problem is completely resolved: computing $H(S)$ gives the maximum possible value of $J(G)$ and hence the optimal determinant condition. Although Definition 2 allows more general row and column permutations, it suffices to consider triangular constructions induced by a monomial order in order to attain this optimum. By Lemma 1,

$$\mathrm{FR}(f) = \bigcup_S \left\{ (\alpha, \beta) \in \mathbb{R}^2_{>0} : \sum_{s \in S} w_{\alpha,\beta}(s) < H(S) \right\}.$$

Lemma 2. *For any $\alpha, \beta > 0$, define*

$$\delta(\alpha, \beta) = \sup_S \frac{H(S)}{\sum_{s \in S} w_{\alpha,\beta}(s)},$$

where S ranges over all finite subsets of $\mathbb{Z}^2_{\geq 0}$ containing at least one nonzero vector. Then $0 < \delta(\alpha, \beta) < \infty$, and

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

Proof. Choose $\gamma \in A(f) \setminus \{0\}$. Associate each pair $(u, j) \in \mathbb{Z}^2_{\geq 0} \times \mathbb{Z}_{\geq 1}$ satisfying $u + jA(f) \subseteq S$ with the point $s = u + j\gamma \in S$. For a fixed s, each j uniquely determines $u = s - j\gamma$. Since the coordinates of u are nonnegative, we have $j\, w_{\alpha,\beta}(\gamma) \leq w_{\alpha,\beta}(s)$. Therefore, counting these pairs according to their associated points s gives

$$H(S) \leq \sum_{s \in S} \left\lfloor \frac{w_{\alpha,\beta}(s)}{w_{\alpha,\beta}(\gamma)} \right\rfloor \leq \frac{\sum_{s \in S} w_{\alpha,\beta}(s)}{w_{\alpha,\beta}(\gamma)}.$$

Since $w_{\alpha,\beta}(\gamma) > 0$, it follows that $\delta(\alpha,\beta) \le 1/w_{\alpha,\beta}(\gamma) < \infty$. On the other hand, taking $S = A(f)$ gives $H(S) \ge 1$, and the denominator is positive. Hence $\delta(\alpha,\beta) > 0$.

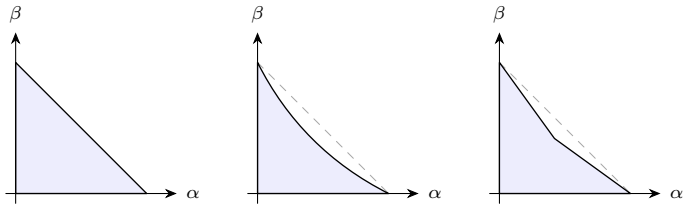
By the preceding characterization of the feasible region, $(\alpha,\beta) \in \mathsf{FR}$ if and only if there exists such a set S for which the corresponding ratio is strictly greater than 1. By the definition of the supremum, this is equivalent to $\delta(\alpha,\beta) > 1$. □

Thus, the problem of determining the optimal small-root bounds for the bivariate Coppersmith method reduces to determining the function $\delta(\alpha,\beta)$. □

Lemma 3. *The function $\delta : \mathbb{R}^2_{>0} \to \mathbb{R}_{>0}$ is homogeneous, monotone, 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 $\mathbb{R}^2_{>0} \setminus \mathrm{FR}$.*

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



3 Counting Tools for Interval Regions

In the proofs that follow, we will encounter a special class of two-dimensional regions, called interval regions, and will need to estimate the number of integer points they contain. This section studies lattice-point counting in these regions and establishes the estimates needed in the subsequent proofs.

3.1 Bounding Local Area by Total Area

Definition 4 (Interval region). *Let $a < b$. If a compact convex set $Q \subseteq \mathbb{R}^2$ satisfies*

$$\{(x,x) : a \le x \le b\} \subseteq Q \subseteq \{(x,y) : a \le x \le y \le b\},$$

then Q is called an interval region on $[a,b]$.

The term "interval region" reflects the fact that each point $(x,y) \in Q$ represents an interval $[x,y]$. The defining properties of Q ensure that, for every $(x,y) \in Q$, if $[x',y'] \subseteq [x,y]$, then $(x',y') \in Q$.

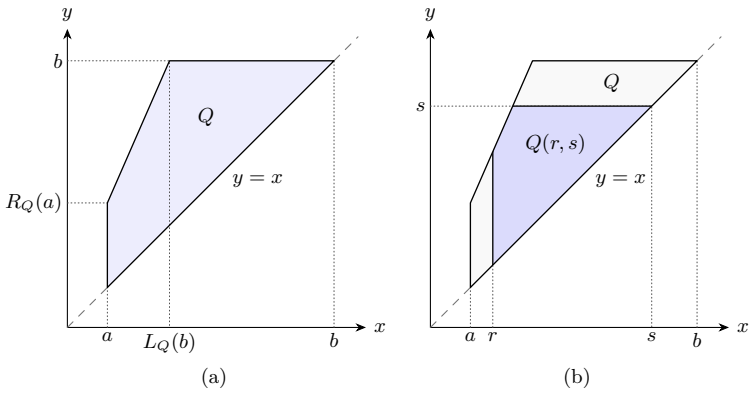
Define the two boundary functions of Q by

$$R_Q(x) = \max\{y : (x,y) \in Q\}, \qquad L_Q(y) = \min\{x : (x,y) \in Q\}.$$

It is straightforward to verify that R_Q and L_Q are both continuous and nondecreasing on $[a, b]$, with R_Q concave and L_Q convex. Moreover, for every $x \in [a, b]$, we have $x \le R_Q(x) \le b$ and $a \le L_Q(x) \le x$. In particular, $R_Q(b) = b$ and $L_Q(a) = a$.

For any $a \le r \le s \le b$, define the local interval region of Q on $[r, s]$ by $Q(r, s) = Q \cap \{(x, y) : r \le x \le y \le s\}$. Integrating the lengths of the vertical and horizontal cross sections, respectively, gives

$$\begin{aligned} \mathrm{vol}_2(Q(r,s)) &= \int_r^s \big(\min\{R_Q(x), s\} - x\big)\, dx \\ &= \int_r^s \big(y - \max\{L_Q(y), r\}\big)\, dy. \end{aligned} \tag{4}$$



Lemma 4 (Local area bound). *Let Q be an interval region on $[a, b]$, let $c > 0$, and let ω be an affine function that is strictly positive on $[a, b]$. If*

$$\begin{aligned} \mathrm{vol}_2(Q) &\le c \int_a^b \omega(t)\, dt, \\ R_Q(a) - a \ge c\omega(a), &\qquad b - L_Q(b) \ge c\omega(b), \end{aligned} \tag{5}$$

then, for any $a \le r \le s \le b$,

$$\mathrm{vol}_2(Q(r,s)) \le c \int_r^s \omega(t)\, dt.$$

Proof. Consider the function $J(r, s) = \mathrm{vol}_2(Q(r, s)) - c\int_r^s \omega(t)\, dt$. We first fix the left endpoint at a and prove that $J(a, s) \le 0$. By (4),

$$\begin{aligned} \frac{d}{ds} J(a,s) &= \frac{d}{ds} \int_a^s \big(y - \max\{L_Q(y), a\} - c\omega(y)\big)\, dy \\ &= s - \max\{L_Q(s), a\} - c\omega(s) \\ &= s - L_Q(s) - c\omega(s). \end{aligned}$$

This derivative is a concave function of s. Its value at $s = a$ is $-c\omega(a) < 0$, and its value at $s = b$ is nonnegative by (5). Thus, $J(a, s)$ is first nonincreasing and then nondecreasing, so its maximum is attained at an endpoint. Together with $J(a, a) = 0$ and $J(a, b) \le 0$, this gives $J(a, s) \le \max\{J(a, a), J(a, b)\} \le 0$.

Now fix $s \in (a, b]$ and vary the left endpoint r. Again, by (4),

$$\begin{aligned}\frac{\partial J}{\partial r}(r, s) &= \frac{d}{dr}\int_r^s \big(\min\{R_Q(x), s\} - x - c\omega(x)\big)\, dx \\ &= r + c\omega(r) - \min\{R_Q(r), s\} \\ &= \max\{r + c\omega(r) - R_Q(r),\ r + c\omega(r) - s\}.\end{aligned}$$

The first expression in the maximum is convex in r and, by (5), is nonpositive at $r = a$. Hence, the set on which it is nonpositive in $[a, s]$ is an interval containing a: if its values at two points are nonpositive, convexity ensures that it is nonpositive between them. The second expression is affine and strictly positive at $r = s$, so the set on which it is nonpositive in $[a, s]$ is either empty or also an interval starting at a. The derivative above is nonpositive if and only if both expressions are nonpositive. It follows that, as a function of r, $J(r, s)$ is either first nonincreasing and then nondecreasing, or increasing throughout. Therefore,

$$J(r, s) \le \max\{J(a, s), J(s, s)\} = 0.$$

This completes the proof. □

3.2 From Area Bounds to Integer Interval Counts

Lemma 5 (Lattice points and area). *Let n be a positive integer, and let Q be an interval region on $[0, n]$. Then*

$$\#\{(x, y) \in Q \cap \mathbb{Z}^2 : x < y\} \le \frac{n+1}{n}\,\mathrm{vol}_2(Q).$$

Proof. The vertical cross-section length $h(x) = R_Q(x) - x$ is concave, and $h(n) = 0$. Let $I = \int_0^n h(x)\, dx = \mathrm{vol}_2(Q)$. By concavity, the graph of h lies above its chords. Comparing I with the sum of the trapezoidal areas over the unit intervals and with the triangular area under the chord over $[0, n]$, respectively, gives

$$\sum_{k=0}^{n} h(k) - \frac{h(0)}{2} \le I, \qquad \frac{nh(0)}{2} \le I.$$

For each $k \in \{0, \dots, n\}$, Q contains exactly $\lfloor h(k) \rfloor$ integer points with $x = k$ and $y > k$. Therefore,

$$\#\{(x, y) \in Q \cap \mathbb{Z}^2 : x < y\} = \sum_{k=0}^{n} \lfloor h(k) \rfloor \le \sum_{k=0}^{n} h(k) \le I + \frac{h(0)}{2} \le \frac{n+1}{n} I.$$

This completes the proof. □

In what follows, we use $[x, y]_{\mathbb{Z}} = [x, y] \cap \mathbb{Z}$ to denote an integer interval.

Theorem 1 (Weighted counting on finite sets of integers). *Suppose that Q, c, ω satisfy the conditions of Lemma 4. Then, for every finite set $T \subseteq [a, b] \cap \mathbb{Z}$,*

$$\#\{(x, y) \in Q \cap \mathbb{Z}^2 : x < y,\ [x, y]_{\mathbb{Z}} \subseteq T\} \leq c \sum_{k \in T} \omega(k). \tag{6}$$

Proof. First consider $T = [r, s]_{\mathbb{Z}}$, where r, s are integers. The case $r = s$ is immediate. Assume henceforth that $r < s$, and set $n = s - r$. Translating $Q(r, s)$ by $-r$ in both coordinates gives an interval region on $[0, n]$; this integer translation preserves both area and the number of integer points. By Lemmas 4 and 5,

$$\begin{aligned}\#\{(x, y) \in Q(r, s) \cap \mathbb{Z}^2 : x < y\} &\leq \frac{n+1}{n} \operatorname{vol}_2(Q(r, s)) \\ &\leq c\, \frac{n+1}{n} \int_r^s \omega(t)\, dt = c \sum_{k=r}^{s} \omega(k).\end{aligned}$$

The last equality uses the fact that ω is affine: the integral equals $n(\omega(r)+\omega(s))/2$, while $\omega(r), \ldots, \omega(s)$ form an arithmetic progression whose sum is $(n+1)(\omega(r) + \omega(s))/2$.

For a general nonempty finite set T, write it as the disjoint union of its maximal blocks of consecutive integers, $T = \biguplus_{\ell=1}^{m} [r_\ell, s_\ell]_{\mathbb{Z}}$, where $s_\ell + 1 < r_{\ell+1}$. Every integer interval $[x, y]_{\mathbb{Z}} \subseteq T$ is contained in exactly one of these blocks. Counting separately within each block and applying the preceding bound to each block gives

$$\begin{aligned}\#\{(x, y) \in Q \cap \mathbb{Z}^2 : x < y,\ [x, y]_{\mathbb{Z}} \subseteq T\} &= \sum_{\ell=1}^{m} \#\{(x, y) \in Q(r_\ell, s_\ell) \cap \mathbb{Z}^2 : x < y\} \\ &\leq c \sum_{\ell=1}^{m} \sum_{k=r_\ell}^{s_\ell} \omega(k) = c \sum_{k \in T} \omega(k).\end{aligned}$$

When $T = \varnothing$, both sides are zero. □

4 The Bridge: Balanced Quasi-Newton Polytopes

4.1 Quasi-Newton Polygons and Interval Regions on Their Edges

We call an edge of $N(f)$ a noncoordinate edge if it does not lie on either coordinate axis. Following these edges counterclockwise from the rightmost point on the horizontal axis to the highest point on the vertical axis, label the vertices $v_0, \ldots, v_r$. If $N(f)$ meets one of the coordinate axes only at the origin, take the corresponding endpoint to be the origin. For each directed edge $[v_{i-1}, v_i]$, set

$$D_i = v_i - v_{i-1}, \qquad \nu_i = ((D_i)_2, -(D_i)_1), \qquad k_i = \langle \nu_i, v_{i-1} \rangle = \langle \nu_i, v_i \rangle. \tag{7}$$

Here ν_i is the outward normal vector obtained by rotating D_i clockwise through $90°$, so $\|\nu_i\| = \|D_i\|$. Since the origin belongs to $N(f)$, we have $k_i \geq 0$, and $k_i = \max_{v \in N(f)} \langle \nu_i, v \rangle$.

Keeping the normal vectors fixed, translate the ith supporting line $\langle \nu_i, x \rangle = k_i$ to $\langle \nu_i, x \rangle = b_i$, and define

$$P(b) = \{x \in \mathbb{R}^2_{\geq 0} : \langle \nu_i, x \rangle \leq b_i,\ 1 \leq i \leq r\}. \tag{8}$$

Let z_0 be the intersection of the first supporting line with the horizontal axis, and let z_r be the intersection of the last supporting line with the vertical axis. For $1 \leq i < r$, let z_i be the intersection of the supporting lines with indices i and $i+1$. Consecutive lines are not parallel, so these intersections are uniquely determined and depend linearly on b. The points z_{i-1} and z_i lie on the same line parallel to D_i. Thus, there is a linear function $\ell_i(b)$ of b such that

$$z_i - z_{i-1} = \ell_i(b) D_i. \tag{9}$$

Here ℓ_i is the scale factor relative to the original edge vector; when $\ell_i > 0$, the Euclidean edge length is $\ell_i \|D_i\|$.

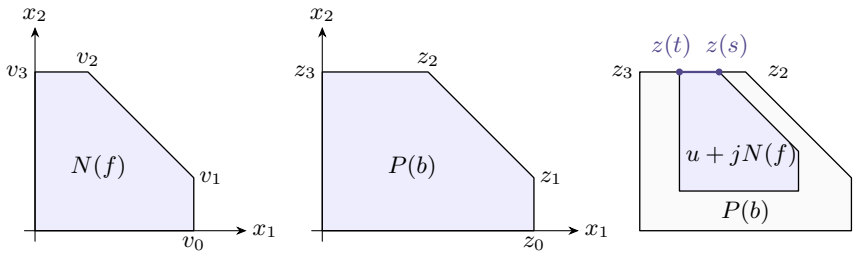
Definition 5 (Quasi-Newton polygon). *Let*

$$B_f = \{b \in \mathbb{R}^r_{>0} : \ell_i(b) > 0 \text{ for all } 1 \leq i \leq r\}.$$

For $b \in B_f$, we call $P(b)$ a quasi-Newton polygon of f.

Lemma 6. *The set B_f is a nonempty open cone. For every $b \in B_f$, the points $0, z_0, \ldots, z_r$ are the vertices of $P(b)$ in counterclockwise order. Its noncoordinate edges are $E_i = [z_{i-1}, z_i]$, and both coordinate edges have positive length. In particular, $P(b)$ is bounded and has nonempty interior.*

Proof. The first noncoordinate edge leaves the horizontal axis upwards, and the last approaches the vertical axis from the right. Hence $(\nu_1)_1 > 0$ and $(\nu_r)_2 > 0$. It follows that $z_0 = (b_1/(\nu_1)_1, 0)$ and $z_r = (0, b_r/(\nu_r)_2)$ lie on the positive horizontal and vertical axes, respectively. Equation (9), together with $\ell_i > 0$, ensures that each noncoordinate edge retains its original orientation. Together with the two coordinate edges, these directions make one full turn, with every turning angle between consecutive edges lying in $(0, \pi)$. Thus, the closed polygonal chain $0, z_0, \ldots, z_r, 0$ bounds a convex polygon. The half-planes on the interior side of its edges are precisely those defining $P(b)$ in (8), so this polygon is $P(b)$. Since each ℓ_i is linear, B_f is defined by strict linear inequalities and is therefore an open cone. At $b = (k_1, \ldots, k_r)$, we have $z_i = v_i$ and $\ell_i = 1$. Increasing all the k_i by the same sufficiently small positive amount makes every b_i positive while keeping every ℓ_i positive. Hence $B_f \neq \varnothing$. □



(a) (b) (c)

We next explain how quasi-Newton polytopes allow us to use interval regions to describe translated and scaled copies $u + jN(f)$ of $N(f)$.

Fix $P = P(b)$ and parameterize its ith edge by $z(s) = z_{i-1} + sD_i$, where $0 \le s \le \ell_i$. Two parameters $s \le t$ determine the copy $z(s) + (t-s)(N(f) - v_{i-1})$. The corresponding transformation maps v_{i-1} and v_i to $z(s)$ and $z(t)$, respectively, so the ith edge of this copy is exactly $[z(s), z(t)]$. Requiring the entire copy to lie in P gives

$$Q_i = \left\{ (s,t) : \begin{array}{l} 0 \le s \le t \le \ell_i, \\ z_{i-1} + sD_i + (t-s)(N(f) - v_{i-1}) \subseteq P \end{array} \right\}. \tag{10}$$

Lemma 7. *The set Q_i is an interval region on $[0, \ell_i]$.*

Proof. For any $s \in [0, \ell_i]$, when $t = s$, the set $z(s) + (t-s)(N(f) - v_{i-1})$ reduces to the single point $z(s)$. Since $z(s)$ lies on the ith edge of P, we have $(s,s) \in Q_i$. Thus,

$$\{(s,s) : 0 \le s \le \ell_i\} \subseteq Q_i \subseteq \{(s,t) : 0 \le s \le t \le \ell_i\}.$$

It remains to show that Q_i is compact and convex. By the convexity of P, the containment condition defining Q_i is equivalent to requiring $z_{i-1} + sD_i + (t-s)(v - v_{i-1}) \in P$ for every vertex v of $N(f)$. This point is affine in (s,t), and P is defined by finitely many non-strict linear inequalities. Hence Q_i is the intersection of the triangle $0 \le s \le t \le \ell_i$ with finitely many closed half-planes, and is therefore compact and convex. □

4.2 The Balance Condition and Vertex Expansion

Applying Theorem 1 requires both the area inequality and the two endpoint conditions. We show below that if the corresponding area inequality holds with equality for every noncoordinate edge of a quasi-Newton polygon, then both endpoint inequalities hold strictly on every such edge. It therefore suffices to find a quasi-Newton polygon satisfying this system of area equalities.

For the ith noncoordinate edge $E_i = [z_{i-1}, z_i]$ of a quasi-Newton polygon $P = P(b)$, retain the parameterization $z(s) = z_{i-1} + sD_i$, where $0 \le s \le \ell_i$, and define

$$\psi_i(b) = \mathrm{vol}_2(Q_i), \qquad \phi_i(b) = \int_0^{\ell_i} w_{\alpha,\beta}\big(z(s)\big)\, ds.$$

Since no noncoordinate edge contains the origin and $\ell_i > 0$, we have $\phi_i(b) > 0$.

Definition 6 (Balanced quasi-Newton polygon). *If there exists $c > 0$ such that $\psi_i(b) = c\phi_i(b)$ for every $1 \le i \le r$, then P is called a balanced quasi-Newton polygon, and c is called its balance constant.*

Lemma 8. *Let P be a balanced quasi-Newton polygon with balance constant c. Then $c\, w_{\alpha,\beta}(v) < 1$ for every vertex v of $N(f)$.*

Proof. The claim is immediate when $v = 0$. If $v = v_{i-1} \neq 0$, consider the vertical cross-section length $h(s) = R_{Q_i}(s) - s$ of Q_i. By the definition and compactness of Q_i, we have $z(s) + h(s)(N(f) - v) \subseteq P$. In particular, substituting $0 \in N(f)$ gives $z(s) - h(s)v \in P \subseteq \mathbb{R}^2_{\ge 0}$, and hence

$$h(s) w_{\alpha,\beta}(v) \le w_{\alpha,\beta}\big(z(s)\big).$$

Since $\alpha, \beta > 0$, equality is possible only when $z(s) = h(s)v$. The supporting line of E_i satisfies $\langle \nu_i, x\rangle = b_i > 0$, so it meets the ray $\mathbb{R}_{\ge 0}v$ in at most one point. Thus, the inequality above is strict for all but at most one value of s. Integrating over $0 \le s \le \ell_i$ and computing the area by vertical cross sections gives

$$\psi_i(b) w_{\alpha,\beta}(v) < \phi_i(b).$$

Substituting $\psi_i(b) = c\phi_i(b)$ and using $\phi_i(b) > 0$ yields $c\, w_{\alpha,\beta}(v) < 1$.

When $v = v_i$, instead use the horizontal cross-section length $h(t) = t - L_{Q_i}(t)$. Again, the definition and compactness of Q_i give

$$\begin{aligned} &z(L_{Q_i}(t)) + h(t)\big(N(f) - v_{i-1}\big) \\ &\quad = \underbrace{z(L_{Q_i}(t)) + h(t)(v_i - v_{i-1})}_{= z(t)} + h(t)\big(N(f) - v_i\big) \\ &\quad = z(t) + h(t)\big(N(f) - v_i\big) \subseteq P. \end{aligned}$$

The same argument now applies, proving the claim for all vertices. □

Lemma 9 (Vertex expansion). *Let P be a balanced quasi-Newton polygon with balance constant c. For every pair of corresponding vertices z_h and v_h, where $0 \le h \le r$, there exists $t > c\, w_{\alpha,\beta}(z_h)$ such that*

$$z_h + t\big(N(f) - v_h\big) \subseteq P. \tag{11}$$

Proof. Fix a pair of corresponding vertices and write $z^* = z_h$ and $v = v_h$. Write the two coordinate constraints in the same form $\langle \nu_j, x\rangle \le b_j$, taking the corresponding triples (ν_j, b_j, k_j) to be $(-e_1, 0, 0)$ and $(-e_2, 0, 0)$, respectively. Throughout this proof, j ranges over all constraints, and we always have $k_j = \max_{q \in N(f)} \langle \nu_j, q\rangle$. Then, for every $t \ge 0$, condition (11) is equivalent to

$$\langle \nu_j, z^*\rangle + t\big(k_j - \langle \nu_j, v\rangle\big) \le b_j \qquad \text{for all } j. \tag{12}$$

Suppose, for a contradiction, that no $t > c\, w_{\alpha,\beta}(z^*)$ satisfies all these inequalities. For every j, we have $b_j - \langle \nu_j, z^*\rangle \ge 0$ and $k_j - \langle \nu_j, v\rangle \ge 0$. If the latter

quantity is zero, the constraint holds for all $t \geq 0$; if it is positive, the constraint gives an upper bound on t. Since there are only finitely many constraints, there must be some j such that

$$\Delta = k_j - \langle \nu_j, v \rangle > 0, \qquad \frac{b_j - \langle \nu_j, z^* \rangle}{\Delta} \leq c\, w_{\alpha,\beta}(z^*).$$

Define the affine function

$$g(x) = \frac{b_j - \langle \nu_j, x \rangle}{\Delta} - c\, w_{\alpha,\beta}(x),$$

so that $g(z^*) \leq 0$.

Choose $q \in N(f)$ such that $\langle \nu_j, q \rangle = k_j$, and let $z^{(1)}, z^{(2)}$ be the two vertices of P adjacent to z^*. Let μ_1, μ_2 be the outward normal vectors of the two edges incident to z^*. The corresponding vertex v satisfies $\langle \mu_a, v \rangle = \max_{p \in N(f)} \langle \mu_a, p \rangle$, so $\langle \mu_a, q - v \rangle \leq 0$ for $a = 1, 2$. In other words, $q - v$ lies in the cone generated by $z^{(1)} - z^*$ and $z^{(2)} - z^*$. Hence there exist $\lambda_1, \lambda_2 \geq 0$ such that

$$q - v = \lambda_1 \big(z^{(1)} - z^*\big) + \lambda_2 \big(z^{(2)} - z^*\big).$$

Since g is affine,

$$\begin{aligned} \sum_{a=1}^{2} \lambda_a \big(g(z^{(a)}) - g(z^*)\big) &= -\frac{\langle \nu_j, q - v \rangle}{\Delta} - c\, w_{\alpha,\beta}(q - v) \\ &= -1 - c\, w_{\alpha,\beta}(q) + c\, w_{\alpha,\beta}(v) < 0. \end{aligned}$$

The last step uses $w_{\alpha,\beta}(q) \geq 0$ and Lemma 8. It follows that at least one a satisfies $g(z^{(a)}) < g(z^*) \leq 0$. Since g is affine, it is strictly negative on the edge $E = [z^*, z^{(a)}]$, except possibly at the endpoint z^*.

This edge cannot be a coordinate edge. Indeed, $z^* \neq 0$; if E were a coordinate edge, its other endpoint would be $z^{(a)} = 0$. But $g(0) = b_j/\Delta \geq 0$, contradicting $g(z^{(a)}) < 0$. Thus, $E = E_i$ is a noncoordinate edge.

Parameterize this edge by $z(s) = z_{i-1} + sD_i$, where $0 \leq s \leq \ell_i$. If $z^* = z_{i-1}$, then $v = v_{i-1}$. For any $(s, y) \in Q_i$, the definition of Q_i and the jth defining inequality of P give $\langle \nu_j, z(s) \rangle + (y - s)\Delta \leq b_j$. Taking y to be the upper endpoint of the vertical cross section yields

$$R_{Q_i}(s) - s \leq \frac{b_j - \langle \nu_j, z(s) \rangle}{\Delta}.$$

If $z^* = z_i$, then $v = v_i$. For any $(x, s) \in Q_i$, using $D_i = v_i - v_{i-1}$, we can rewrite the corresponding copy as

$$z(x) + (s - x)(N(f) - v_{i-1}) = z(s) + (s - x)(N(f) - v_i).$$

Applying the same defining inequality and taking the left endpoint of the horizontal cross section gives

$$s - L_{Q_i}(s) \leq \frac{b_j - \langle \nu_j, z(s) \rangle}{\Delta}.$$

Integrating along vertical or horizontal cross sections, respectively, yields in either case

$$\begin{aligned}\psi_i(b) &\le \int_0^{\ell_i} \frac{b_j - \langle \nu_j, z(s)\rangle}{\Delta}\, ds \\ &= c\phi_i(b) + \int_0^{\ell_i} g\big(z(s)\big)\, ds < c\phi_i(b).\end{aligned}$$

The strict inequality follows from $\ell_i > 0$ and the strict negativity of g on $E_i \setminus \{z^*\}$. This contradicts the balance equality $\psi_i(b) = c\phi_i(b)$, so the required t exists. □

Corollary 2 (Endpoint conditions). *Let P be a balanced quasi-Newton polygon with balance constant c. Then every noncoordinate edge E_i satisfies*

$$R_{Q_i}(0) > c\, w_{\alpha,\beta}(z_{i-1}), \qquad \ell_i - L_{Q_i}(\ell_i) > c\, w_{\alpha,\beta}(z_i).$$

Proof. Applying Lemma 9 at z_{i-1}, choose $t_- > c\, w_{\alpha,\beta}(z_{i-1})$ such that $z_{i-1} + t_-(N(f) - v_{i-1}) \subseteq P$. Since $v_i \in N(f)$, the point $z_{i-1} + t_- D_i = z(t_-)$ belongs to P and lies on the supporting line of E_i, so it belongs to E_i. Thus, $t_- \le \ell_i$ and $(0, t_-) \in Q_i$, giving $R_{Q_i}(0) \ge t_- > c\, w_{\alpha,\beta}(z_{i-1})$.

Similarly, at z_i, choose $t_+ > c\, w_{\alpha,\beta}(z_i)$ such that $z_i + t_+(N(f) - v_i) \subseteq P$. Since $v_{i-1} \in N(f)$, we have $z_i - t_+ D_i = z(\ell_i - t_+) \in E_i$, and hence $t_+ \le \ell_i$. Moreover,

$$z(\ell_i - t_+) + t_+(N(f) - v_{i-1}) = z_i + t_+(N(f) - v_i) \subseteq P$$

implies $(\ell_i - t_+, \ell_i) \in Q_i$. Therefore, $\ell_i - L_{Q_i}(\ell_i) \ge t_+ > c\, w_{\alpha,\beta}(z_i)$. □

4.3 An Upper Bound for Arbitrary Finite Sets

Theorem 2 (Uniform upper bound from balanced polygons). *Let $P = P(b)$ be a balanced quasi-Newton polygon with balance constant c. Then, for every finite set $S \subseteq \mathbb{Z}^2_{\ge 0}$,*

$$H(S) \le c \sum_{s \in S} w_{\alpha,\beta}(s). \tag{13}$$

Consequently, $\delta(\alpha, \beta) \le c$.

Proof. If $S \subseteq \{0\}$, then $H(S) = 0$, and the claim holds. Henceforth, fix any other finite set S.

Preparation For each nonzero point $x \in \mathbb{R}^2_{\ge 0}$, set $q(x) = \max_{1 \le i \le r}\big(\langle \nu_i, x\rangle / b_i\big)$, and let $I(x)$ be the smallest index attaining the maximum. It is straightforward to verify that $q(x)$ is the smallest positive t for which $x \in tP$, and that x lies on the $I(x)$th noncoordinate edge of $q(x)P$.

Similarly, for each $(u, j) \in \mathcal{H}(S)$, set $q(u, j) = \max_{1 \le i \le r}\big((\langle \nu_i, u\rangle + jk_i)/b_i\big)$, and let $I(u, j)$ be the smallest index attaining the maximum. Write $q = q(u, j)$ and $i = I(u, j)$. It is straightforward to verify that q is the smallest positive t

for which $u + jA(f) \subseteq tP$. At this scale, the copy $u + jN(f)$ is contained in qP, and its edge $[u + jv_{i-1},\, u + jv_i]$ lies on qE_i, the ith noncoordinate edge of qP. Among all noncoordinate edges of qP that meet this copy, qE_i has the smallest index; we call it the contact edge of the copy.

The points $x_k = u + (j-k)v_{i-1} + kv_i$ $(k = 0, \ldots, j)$ on the contact edge all belong to $u + jA(f) \subseteq S$ and satisfy $q(x_k) = q$ and $I(x_k) = i$.

Step 1: Partitioning the point set. Partition $S \setminus \{0\}$ according to the following equivalence relation:

$$x \sim y \iff q(x) = q(y), \quad I(x) = I(y), \quad x - y \in \mathbb{Z}D_{I(x)}.$$

Each $\mathbb{Z}D_i$ is an additive subgroup, so this is indeed an equivalence relation. Let $\mathcal{C}(S)$ be the set of all equivalence classes. For each class C, let q_C, i_C be its common scaling factor and edge index, choose $z_C^0 \in C$, and define

$$z_C(t) = z_C^0 + tD_{i_C}, \qquad T_C = \{k \in \mathbb{Z} : z_C(k) \in C\}.$$

By the definition of the classes and the fact that $D_{i_C} \neq 0$, the map $z_C : T_C \to C$ is a bijection. Hence

$$\sum_{k \in T_C} w_{\alpha,\beta}(z_C(k)) = \sum_{x \in C} w_{\alpha,\beta}(x). \tag{14}$$

Each class C lies on the i_Cth edge of $q_C P$. Write the entire edge as $z_C([a_C, b_C])$. Then $T_C \subseteq [a_C, b_C] \cap \mathbb{Z}$. Define the corresponding interval region by

$$Q_C = \{(s,t) : a_C \le s \le t \le b_C, \\ z_C(s) + (t-s)(N(f) - v_{i_C - 1}) \subseteq q_C P\}.$$

Step 2: Mapping each $u + jA(f)$ to a distinct labeled integer interval. Take $(u, j) \in \mathcal{H}(S)$, set $i = I(u, j)$, and let C be the class containing $u + jv_{i-1}$. Then $q_C = q(u, j)$ and $i_C = i$, and all points in the chain belong to C. Since $z_C : T_C \to C$ is a bijection, there is a unique $x \in T_C$ such that $z_C(x) = u + jv_{i-1}$. Set $y = x + j$, and map (u, j) to the triple (C, x, y).

We first check that the image represents an admissible integer interval. For $k = 0, \ldots, j$, we have

$$z_C(x + k) = u + (j-k)v_{i-1} + kv_i \in C,$$

so $[x, y]_{\mathbb{Z}} \subseteq T_C$ and $a_C \le x < y \le b_C$. Moreover,

$$z_C(x) + (y - x)(N(f) - v_{i-1}) = u + jN(f) \subseteq q_C P,$$

so $(x, y) \in Q_C$. Thus, the map is well defined. From the labeled image (C, x, y), we can uniquely recover

$$j = y - x, \qquad u = z_C(x) - (y - x)v_{i_C - 1},$$

so the map is injective. The label C distinguishes intervals with the same parameter values in different classes. Consequently,

$$H(S) \le \sum_{C \in \mathcal{C}(S)} \#\{(x, y) \in Q_C \cap \mathbb{Z}^2 : x < y,\ [x, y]_{\mathbb{Z}} \subseteq T_C\}. \tag{15}$$

Step 3: Summing the interval counts over the classes. Fix a class C, and write $i = i_C$ and $q = q_C$ for brevity. Since $z_C([a_C, b_C]) = qE_i$ and z_C has direction D_i, we have

$$z_C(a_C) = qz_{i-1}, \qquad z_C(b_C) = qz_i, \qquad b_C - a_C = q\ell_i.$$

Furthermore, $z_C(a_C + qs) = q(z_{i-1} + sD_i)$, and hence

$$Q_C = \{(a_C + qs, a_C + qt) : (s,t) \in Q_i\}.$$

This shows that Q_C is an interval region on $[a_C, b_C]$. Changing variables in the area and weighted integrals gives

$$\mathrm{vol}_2(Q_C) = q^2\psi_i(b),$$
$$\int_{a_C}^{b_C} w_{\alpha,\beta}(z_C(s))\,ds = q^2\phi_i(b).$$

Thus, the balance equality $\psi_i(b) = c\phi_i(b)$ continues to hold for Q_C with the same constant c. The endpoint cross sections satisfy

$$R_{Q_C}(a_C) - a_C = qR_{Q_i}(0),$$
$$b_C - L_{Q_C}(b_C) = q\big(\ell_i - L_{Q_i}(\ell_i)\big),$$

while the corresponding endpoint weights are $qw_{\alpha,\beta}(z_{i-1})$ and $qw_{\alpha,\beta}(z_i)$, respectively. By Corollary 2, the two endpoint conditions required for interval counting also remain valid. The function $s \mapsto w_{\alpha,\beta}(z_C(s))$ is strictly positive and affine along the entire edge, so we may apply Theorem 1 to Q_C and T_C.

Substituting the resulting bound into (15) and then using (14), we obtain

$$\begin{aligned} H(S) &\le c \sum_{C\in\mathcal{C}(S)} \sum_{k\in T_C} w_{\alpha,\beta}(z_C(k)) \\ &= c \sum_{C\in\mathcal{C}(S)} \sum_{x\in C} w_{\alpha,\beta}(x) = c\sum_{x\in S} w_{\alpha,\beta}(x). \end{aligned}$$

The last equality uses the fact that the classes form a disjoint partition of $S \setminus \{0\}$ and that $w_{\alpha,\beta}(0) = 0$. This proves (13). Taking the supremum over all finite sets for which the denominator is positive yields $\delta(\alpha,\beta) \le c$. □

The upper bound above imposes no restriction on the shape of S. In the next section, we use the balanced polygon itself to construct a sequence of lattice-point sets that asymptotically attains this bound.

4.4 Asymptotic Lower Bound

For a compact convex polygon $C \subseteq \mathbb{R}^2_{\ge 0}$ with nonempty interior, define

$$\widetilde{\mathcal{H}}(C) = \big\{(u,t) \in \mathbb{R}^2_{\ge 0} \times \mathbb{R}_{\ge 0} : u + tN(f) \subseteq C\big\},$$

$$\Psi(C) = \mathrm{vol}_3\big(\widetilde{\mathcal{H}}(C)\big), \qquad \Phi(C) = \int_C w_{\alpha,\beta}(x)\,dx.$$

It is straightforward to verify that $\widetilde{\mathcal{H}}(C)$ is a bounded convex polyhedron with a nonempty interior and that both $\Psi(C)$ and $\Phi(C)$ are finite and positive. For any $q > 0$, the identity $\widetilde{\mathcal{H}}(qC) = q\widetilde{\mathcal{H}}(C)$ and the fact that the weight function is homogeneous of degree one give

$$\Psi(qC) = q^3\Psi(C), \qquad \Phi(qC) = q^3\Phi(C). \tag{16}$$

Thus, the ratio $\Psi(C)/\Phi(C)$ is invariant under uniform scaling.

Lemma 10 (Asymptotic ratio for lattice-point sets in dilated polygons). *Set $S_n = nC \cap \mathbb{Z}^2$. As $n \to \infty$ through the positive integers,*

$$\frac{H(S_n)}{n^3} \longrightarrow \Psi(C), \qquad \frac{\sum_{s\in S_n} w_{\alpha,\beta}(s)}{n^3} \longrightarrow \Phi(C). \tag{17}$$

In particular, $\delta(\alpha,\beta) \geq \Psi(C)/\Phi(C)$.

Consequently, if P is a balanced quasi-Newton polygon with balance constant c and $\Psi(P)/\Phi(P) = c$, then $\delta(\alpha,\beta) = c$.

4.5 Boundary Derivatives and the Balance Condition

For $b \in B_f$, write $\Psi(b) = \Psi(P(b))$ and $\Phi(b) = \Phi(P(b))$. As before, the ith edge is $z_{i-1} + [0,\ell_i]D_i$, with associated interval region Q_i, and $\psi_i(b) = \mathrm{vol}_2(Q_i)$ and $\phi_i(b) = \int_0^{\ell_i} w_{\alpha,\beta}(z_{i-1} + sD_i)\,ds$.

Lemma 11 (Boundary derivatives). *The functions Ψ, Φ are continuously differentiable on B_f, and for $1 \leq i \leq r$,*

$$\frac{\partial\Psi}{\partial b_i}(b) = \psi_i(b), \qquad \frac{\partial\Phi}{\partial b_i}(b) = \phi_i(b). \tag{18}$$

Lemma 12 (Stationary points). *A quasi-Newton polygon $P(b)$ is balanced if and only if b is a stationary point of Ψ/Φ. If its balance constant is c, then $c = \Psi(b)/\Phi(b)$.*

Proof. Suppose first that b is a stationary point, and set $c = \Psi(b)/\Phi(b) > 0$. By Lemma 11,

$$0 = \partial_i(\Psi/\Phi)(b) = \frac{\psi_i(b) - c\phi_i(b)}{\Phi(b)}, \qquad 1 \leq i \leq r,$$

so all edges satisfy the balance condition with the same constant c.

Conversely, if $\psi_i(b) = c\phi_i(b)$ for every i, then $\partial_i\Psi(b) = c\partial_i\Phi(b)$. By (16), we have $\Psi(qb) = q^3\Psi(b)$ and $\Phi(qb) = q^3\Phi(b)$. Differentiating with respect to q at $q = 1$ gives

$$3\Psi(b) = \sum_{i=1}^{r} b_i\partial_i\Psi(b) = c\sum_{i=1}^{r} b_i\partial_i\Phi(b) = 3c\Phi(b).$$

Thus, $c = \Psi(b)/\Phi(b)$. Substituting this identity into the quotient rule yields $\nabla(\Psi/\Phi)(b) = 0$. □

Theorem 3. *If $P(b)$ is a balanced quasi-Newton polygon with balance constant c, then*

$$\delta(\alpha,\beta) = c = \frac{\Psi(P(b))}{\Phi(P(b))}. \tag{19}$$

Moreover, setting $S_n = nP(b) \cap \mathbb{Z}^2$, we have $H(S_n)/\sum_{s\in S_n} w_{\alpha,\beta}(s) \to c$.

Proof. By Theorem 2, every finite set S satisfies $H(S) \le c\sum_{s\in S} w_{\alpha,\beta}(s)$, so $\delta(\alpha,\beta) \le c$. On the other hand, Lemmas 10 and 12 give

$$\delta(\alpha,\beta) \ge \lim_{n\to\infty} \frac{H(S_n)}{\sum_{s\in S_n} w_{\alpha,\beta}(s)} = \frac{\Psi(P(b))}{\Phi(P(b))} = c.$$

The upper and lower bounds coincide, proving the claim. □

Thus, finding a balanced polygon yields both a uniform upper bound and a construction that asymptotically attains it. By Lemma 12, such a polygon can be obtained by finding a stationary point of Ψ/Φ in B_f.

Lemma 13 (Existence of a balanced quasi-Newton polygon). *The function $\Psi(P(b))/\Phi(P(b))$ attains its maximum on B_f. Consequently, there exists $b^* \in B_f$ such that $\nabla(\Psi/\Phi)(b^*) = 0$, and the corresponding $P(b^*)$ is a balanced quasi-Newton polygon.*

5 Computing the Feasible Region

Determining $\delta(\alpha,\beta)$ is equivalent to finding a balanced quasi-Newton polygon. This section explains how to formulate and solve the corresponding balance equations from $N(f)$, α, and β:

$$\psi_i(b) = c\phi_i(b) \quad (1 \le i \le r), \qquad b \in B_f, \quad c > 0. \tag{20}$$

Since ψ_i and ϕ_i are both homogeneous of degree two, uniform scaling leaves the balance constant unchanged, so we may fix $b_1 = 1$. Within B_f, each endpoint is always determined by the same pair of adjacent supporting lines, including the coordinate axes where applicable. These lines have fixed integer normal vectors, so each vertex z_i is a linear function of b with rational coefficients. Since $z_i - z_{i-1} = \ell_i D_i$ and $D_i \neq 0$, the same holds for ℓ_i. After normalization, the vertices and ℓ_i are affine functions of $b_2, \dots, b_r$ with rational coefficients, while the conditions $b_i > 0$ and $\ell_i > 0$ give the linear constraints defining the normalized parameter domain.

Computing ψ_i. By the definition of the interval region, $(\xi,\eta) \in Q_i$ if and only if $0 \le \xi \le \eta \le \ell_i$ and

$$z_{i-1} + \xi D_i + (\eta - \xi)\big(N(f) - v_{i-1}\big) \subseteq P(b).$$

Using $D_i = v_i - v_{i-1}$, we obtain, for every $v \in N(f)$,

$$\begin{aligned} z_{i-1} + \xi D_i + (\eta - \xi)(v - v_{i-1}) \\ = z_{i-1} + \xi v_i - \eta v_{i-1} + (\eta - \xi)v. \end{aligned} \tag{21}$$

Since $0 \in N(f) \subseteq \mathbb{R}^2_{\geq 0}$ and $\eta - \xi \geq 0$, both coordinates of the point above attain their minimum at $v = 0$. Thus, the copy lies in the nonnegative quadrant if and only if $z_{i-1}+\xi v_i-\eta v_{i-1} \geq 0$, where the vector inequality is interpreted coordinatewise. For the noncoordinate supporting inequalities, using $k_j = \max_{v \in N(f)} \langle \nu_j, v \rangle$ gives

$$\begin{aligned} \max_{v \in N(f)} \big\langle \nu_j, z_{i-1} + \xi v_i - \eta v_{i-1} + (\eta - \xi)v \big\rangle \\ = \langle \nu_j, z_{i-1} \rangle + (\langle \nu_j, v_i \rangle - k_j)\xi + (k_j - \langle \nu_j, v_{i-1} \rangle)\eta. \end{aligned}$$

When $j = i$, we have $\langle \nu_i, v_{i-1} \rangle = \langle \nu_i, v_i \rangle = k_i$ and $\langle \nu_i, z_{i-1} \rangle = b_i$. The corresponding constraint therefore holds identically as an equality and can be omitted. Hence

$$Q_i = \left\{ (\xi, \eta) \in \mathbb{R}^2 : \begin{array}{r} 0 \leq \xi \leq \eta \leq \ell_i, \\ \eta v_{i-1} - \xi v_i \leq z_{i-1}, \\ (\langle \nu_j, v_i \rangle - k_j)\xi + (k_j - \langle \nu_j, v_{i-1} \rangle)\eta \\ \leq b_j - \langle \nu_j, z_{i-1} \rangle \quad (1 \leq j \leq r,\ j \neq i) \end{array} \right\}. \tag{22}$$

These inequalities have fixed rational coefficients in (ξ, η), and their right-hand sides are affine functions of $b_2, \ldots, b_r$ with rational coefficients. By the piecewise polynomiality of parametric polytope volumes [3, Theorem 2.26(1)–(2)], the domain $B_f \cap \{b_1 = 1\}$ can be partitioned into finitely many regions defined by linear equalities and inequalities with rational coefficients, such that, on each region, $\psi_i(b) = \mathrm{vol}_2(Q_i)$ is a polynomial in $b_2, \ldots, b_r$ of total degree at most 2, with rational coefficients. The partition also includes the boundaries between these regions. More explicitly, within each region, each vertex of Q_i is determined by the same pair of boundary lines. List the vertices in counterclockwise order as $q_1, \ldots, q_s$, and set $q_{s+1} = q_1$. Then

$$\psi_i(b) = \frac{1}{2} \sum_{h=1}^{s} \det(q_h, q_{h+1}). \tag{23}$$

The vertex coordinates are affine functions of the parameters with rational coefficients, so this formula directly yields the corresponding area polynomial.

Computing ϕ_i. By the linearity of $w_{\alpha,\beta}$, direct integration gives

$$\begin{aligned} \phi_i(b) &= \int_0^{\ell_i} w_{\alpha,\beta}(z_{i-1} + sD_i)\, ds \\ &= \frac{\ell_i}{2} \big(w_{\alpha,\beta}(z_{i-1}) + w_{\alpha,\beta}(z_i) \big). \end{aligned} \tag{24}$$

Since the endpoints and ℓ_i depend affinely on the parameters, ϕ_i is represented throughout the normalized parameter domain by a single polynomial of total degree at most 2, whose coefficients are rational linear combinations of α and β.

On each region, solve the balance equations $\psi_i(b) = c\phi_i(b)$, subject to the linear constraints defining that region, together with $b \in B_f$, $b_1 = 1$, and $c > 0$. Lemma 13 guarantees the existence of at least one valid solution, while Theorem 3 ensures that the balance constant of every valid solution equals $\delta(\alpha, \beta)$. Therefore, $\delta(\alpha, \beta)$ can be determined by solving these finitely many systems of real polynomial equations.

6 Optimal Values for Representative Newton Polygons

In this section, we compute the optimal values for triangles with axis intercepts, triangles with one coordinate edge, rectangles, and nonrectangular trapezoids. Throughout, m, n, d are positive integers and $\alpha, \beta > 0$. Since uniform scaling leaves the balance constant unchanged, we normalize each quasi-Newton polygon P to have height n. We compute ψ_i, ϕ_i using (22) and (24), and then solve the balance equations. When there are two noncoordinate edges, it suffices to solve $\psi_1\phi_2 = \psi_2\phi_1$; the resulting common ratio $\psi_1/\phi_1 = \psi_2/\phi_2$ is $\delta(\alpha, \beta)$.

6.1 Triangle with Axis Intercepts

Let $N(f) = \mathrm{conv}\{0, (m, 0), (0, n)\}$. After normalization, the only possibility is $P = N(f)$, and the parameter interval of its unique noncoordinate edge has length $\ell_1 = 1$. Since $Q_1 = \{(\xi, \eta) : 0 \le \xi \le \eta \le 1\}$, we obtain

$$\psi_1 = \frac{1}{2}, \qquad \phi_1 = \frac{m\alpha + n\beta}{2}.$$

Therefore,

$$\delta(\alpha, \beta) = \frac{1}{m\alpha + n\beta}. \tag{25}$$

6.2 Upper Triangle

Let $N(f) = \mathrm{conv}\{0, (m, n), (0, n)\}$. Fixing the height of the quasi-Newton polygon at n gives

$$P_\rho = \mathrm{conv}\{0, (m\rho, 0), (m(1+\rho), n), (0, n)\}, \qquad \rho > 0.$$

Computing along the sloping edge and the top edge, respectively, gives

$$\psi_1 = \frac{1}{2}, \qquad \phi_1 = m\alpha\rho + \frac{m\alpha + n\beta}{2},$$
$$\psi_2 = \rho + \frac{1}{2}, \quad \phi_2 = \frac{m\alpha(1+\rho)^2}{2} + n\beta(1+\rho).$$

The balance equation $\psi_1\phi_2 = \psi_2\phi_1$ reduces to $3m\alpha\rho^2 + 2m\alpha\rho - n\beta = 0$. Substituting its unique positive root into ψ_1/ϕ_1 yields

$$\delta(\alpha,\beta) = \frac{3}{m\alpha + 3n\beta + 2\sqrt{m^2\alpha^2 + 3mn\alpha\beta}}. \tag{26}$$

6.3 Lower Triangle

Let $N(f) = \mathrm{conv}\{0,(m,0),(m,n)\}$. Again, fixing the height of the quasi-Newton polygon at n gives

$$P_\rho = \mathrm{conv}\{0,(m\rho,0),(m\rho,n),(0,n(1-\rho))\}, \qquad 0 < \rho < 1.$$

Computing along the right edge and the sloping edge, respectively, gives

$$\psi_1 = \rho - \frac{\rho^2}{2}, \quad \phi_1 = m\alpha\rho + \frac{n\beta}{2},$$
$$\psi_2 = \frac{\rho^2}{2}, \qquad \phi_2 = \frac{m\alpha\rho^2}{2} + n\beta\rho\left(1 - \frac{\rho}{2}\right).$$

The balance equation $\psi_1\phi_2 = \psi_2\phi_1$ reduces to $(m\alpha - n\beta)\rho^2 + 4n\beta\rho - 3n\beta = 0$. Substituting its unique root in $(0,1)$ into ψ_2/ϕ_2 yields

$$\delta(\alpha,\beta) = \frac{3}{3m\alpha + n\beta + 2\sqrt{n^2\beta^2 + 3mn\alpha\beta}}. \tag{27}$$

6.4 Rectangle

Let $N(f) = [0,m]\times[0,n]$. The normalized quasi-Newton polygon is $P_\rho = [0,m\rho]\times[0,n]$, where $\rho > 0$. The two noncoordinate edges have parameter intervals of lengths 1 and ρ, respectively. Direct computation gives

	ψ_1	ψ_2
$0 < \rho \le 1$	$\rho - \rho^2/2$	$\rho^2/2$
$\rho \ge 1$	$1/2$	$\rho - 1/2$

and

$$\phi_1 = m\alpha\rho + \frac{n\beta}{2}, \qquad \phi_2 = \frac{m\alpha\rho^2}{2} + n\beta\rho.$$

Substituting into the balance equation gives the following equations in the two cases:

$$m\alpha\rho^2 + 2n\beta\rho - 3n\beta = 0, \quad 0 < \rho \le 1,$$
$$3m\alpha\rho^2 - 2m\alpha\rho - n\beta = 0, \quad \rho \ge 1.$$

Solving within the indicated ranges and substituting into the common ratio yields

$$\delta(\alpha,\beta) = \begin{cases} \dfrac{3}{\left(\sqrt{n\beta} + \sqrt{n\beta + 3m\alpha}\right)^2}, & m\alpha \ge n\beta, \\ \dfrac{3}{\left(\sqrt{m\alpha} + \sqrt{m\alpha + 3n\beta}\right)^2}, & m\alpha < n\beta. \end{cases} \tag{28}$$

6.5 Nonrectangular Trapezoid

Let $N(f) = \operatorname{conv}\{0, (m, 0), (d, n), (0, n)\}$, where $m \neq d$. The normalized quasi-Newton polygon is

$$P_\rho = \operatorname{conv}\{0, (m - d + d\rho, 0), (d\rho, n), (0, n)\}, \qquad \rho > \max\{0, 1 - m/d\}.$$

Again, the two noncoordinate edges have parameter intervals of lengths 1 and ρ, respectively. Direct computation gives

	ψ_1	ψ_2
$\rho \le 1$	$\frac{1}{2} - \frac{d(1-\rho)^2}{2m}$	$\frac{\rho^2}{2}$
$\rho \ge 1$	$\frac{1}{2}$	$\rho - \frac{1}{2}$

and

$$\phi_1 = \alpha\left(d\rho + \frac{m-d}{2}\right) + \frac{n\beta}{2},$$
$$\phi_2 = \frac{d\alpha\rho^2}{2} + n\beta\rho.$$

For $\rho \le 1$, the balance equation reduces to

$$(d\alpha\rho + 2n\beta)(m - d + d\rho)^2 = mn\beta\big(3d\rho + 2(m-d)\big). \tag{29}$$

When $m\alpha \ge n\beta$, this equation has a unique root ρ^* in $(\max\{0, 1 - m/d\}, 1]$. For $\rho \ge 1$, the balance equation reduces to

$$3d\alpha\rho^2 + 2\alpha(m - 2d)\rho = \alpha(m-d) + n\beta.$$

Substituting the corresponding roots into the common ratio yields

$$\delta(\alpha, \beta) = \begin{cases} \dfrac{\rho^*}{d\alpha\rho^* + 2n\beta}, & m\alpha \ge n\beta, \\ \dfrac{3}{\alpha(m+d) + 3n\beta + 2\sqrt{\alpha^2(m^2 - md + d^2) + 3dn\alpha\beta}}, & m\alpha < n\beta. \end{cases} \tag{30}$$

7 Future Directions and Open Challenges

The results of this paper raise the following two questions for future research.

1. Optimal small-root bounds in three or more variables. A natural and important direction for future research is to characterize and compute the optimal small-root bounds for Coppersmith's method in three or more variables, particularly for systems of polynomial equations. We expect that extending the theory from two variables to three or more may present substantially greater structural difficulties

than extending it from one variable to two. These difficulties mainly arise from the geometric shapes of the optimal monomial regions. On the one hand, the optimal regions may exhibit new boundary directions that differ from the facet directions of the original Newton polytope, so that quasi-Newton polytopes may no longer suffice to describe the optimal constructions. On the other hand, the optimal regions may not even be polyhedral, and their boundaries may contain curved surfaces. Thus, the difficulty in higher dimensions may lie not only in determining additional boundary directions, but also in characterizing more general boundary shapes. Consequently, we believe that the elementary mathematical techniques employed in this paper may not be sufficient to provide a complete characterization of the optimal small-root bounds in three or more variables.

2. Small coefficients and unravelled linearization. It is well known that when the coefficients of a polynomial are small, Coppersmith's method can often achieve better small-root bounds than classical constructions by employing unravelled linearization, as illustrated by attacks on RSA with small private exponents. However, such improvements cannot be captured by the standard triangular shift framework considered in this paper. Therefore, extending the present theory to the small-coefficient setting and developing a unified characterization of the improvements obtained through unravelled linearization constitute another important direction for future research.

References

1. Aono, Y.: A new lattice construction for partial key exposure attack for RSA. In: Public Key Cryptography – PKC 2009. Lecture Notes in Computer Science, vol. 5443, pp. 34–53. Springer (2009). https://doi.org/10.1007/978-3-642-00468-1_3
2. Aono, Y., Agrawal, M., Satoh, T., Watanabe, O.: On the optimality of lattices for the Coppersmith technique. In: Susilo, W., Mu, Y., Seberry, J. (eds.) Information Security and Privacy – ACISP 2012. Lecture Notes in Computer Science, vol. 7372, pp. 376–389. Springer (2012). https://doi.org/10.1007/978-3-642-31448-3_28
3. Baldoni, V., Berline, N., Loera, J.A.D., Köppe, M., Vergne, M.: Three Ehrhart quasi-polynomials. Algebraic Combinatorics **2**(3), 379–416 (2019). https://doi.org/10.5802/alco.46
4. Barthe, G., Pointcheval, D., Zanella-Béguelin, S.: Verified security of redundancy-free encryption from rabin and RSA. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security. pp. 724–735. ACM (2012). https://doi.org/10.1145/2382196.2382272
5. Bauer, A., Vergnaud, D., Zapalowicz, J.C.: Inferring sequences produced by non-linear pseudorandom number generators using Coppersmith's methods. In: Public Key Cryptography – PKC 2012. Lecture Notes in Computer Science, vol. 7293, pp. 609–626. Springer (2012). https://doi.org/10.1007/978-3-642-30057-8_36
6. Bernstein, D.J., Chang, Y.A., Cheng, C.M., Chou, L.P., Heninger, N., Lange, T., van Someren, N.: Factoring RSA keys from certified smart cards: Coppersmith in the wild. In: Advances in Cryptology – ASIACRYPT 2013, Part II. Lecture Notes in Computer Science, vol. 8270, pp. 341–360. Springer (2013). https://doi.org/10.1007/978-3-642-42045-0_18

7. Bleichenbacher, D., May, A.: New attacks on RSA with small secret CRT-exponents. In: Public Key Cryptography – PKC 2006. Lecture Notes in Computer Science, vol. 3958, pp. 1–13. Springer (2006). https://doi.org/10.1007/11745853_1
8. Blömer, J., May, A.: New partial key exposure attacks on RSA. In: Advances in Cryptology – CRYPTO 2003. Lecture Notes in Computer Science, vol. 2729, pp. 27–43. Springer (2003). https://doi.org/10.1007/978-3-540-45146-4_2
9. 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
10. Boneh, D.: Simplified OAEP for the RSA and rabin functions. In: Advances in Cryptology – CRYPTO 2001. Lecture Notes in Computer Science, vol. 2139, pp. 275–291. Springer (2001). https://doi.org/10.1007/3-540-44647-8_17
11. Boneh, D., Durfee, G.: Cryptanalysis of RSA with private key d less than $n^{0.292}$. IEEE Transactions on Information Theory **46**(4), 1339–1349 (2000). https://doi.org/10.1109/18.850673
12. Boneh, D., Durfee, G., Frankel, Y.: An attack on RSA given a small fraction of the private key bits. In: Advances in Cryptology – ASIACRYPT 1998. Lecture Notes in Computer Science, vol. 1514, pp. 25–34. Springer (1998). https://doi.org/10.1007/3-540-49649-1_3
13. Boneh, D., Halevi, S., Howgrave-Graham, N.: The modular inversion hidden number problem. In: Advances in Cryptology – ASIACRYPT 2001. Lecture Notes in Computer Science, vol. 2248, pp. 36–51. Springer (2001). https://doi.org/10.1007/3-540-45682-1_3
14. Chinburg, T., Hemenway, B., Heninger, N., Scherr, Z.: Cryptographic applications of capacity theory: On the optimality of Coppersmith's method for univariate polynomials. In: Cheon, J.H., Takagi, T. (eds.) Advances in Cryptology – ASIACRYPT 2016, Part I. Lecture Notes in Computer Science, vol. 10031, pp. 759–788. Springer (2016). https://doi.org/10.1007/978-3-662-53887-6_28
15. 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
16. 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
17. 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
18. Ding, Z., Dai, Z., Wu, B., Zhang, Y., Zhang, K.: Computing asymptotic bounds for the automated Coppersmith method via linear programming. Cryptology ePrint Archive, Paper 2026/1027 (2026), https://eprint.iacr.org/2026/1027, to appear in Advances in Cryptology – ASIACRYPT 2026
19. Ernst, M., Jochemsz, E., May, A., de Weger, B.: Partial key exposure attacks on RSA up to full size exponents. In: Advances in Cryptology – EUROCRYPT 2005. Lecture Notes in Computer Science, vol. 3494, pp. 371–386. Springer (2005). https://doi.org/10.1007/11426639_22
20. Feng, Y., Luo, H., Chen, Q., Nitaj, A., Pan, Y.: Computing asymptotic bounds for small roots in Coppersmith's method via sumset theory. In: Advances in Cryptology

– CRYPTO 2025, Part I. Lecture Notes in Computer Science, vol. 16000, pp. 3–32. Springer (2025). https://doi.org/10.1007/978-3-032-01855-7_1
21. Feng, Y., Nitaj, A., Pan, Y., Zheng, M.: New asymptotic results on predicting polynomial congruential generators. In: Advances in Cryptology – ASIACRYPT 2026 (2026), https://eprint.iacr.org/2025/2044, to appear
22. Herrmann, M., May, A.: Attacking power generators using unravelled linearization: When do we output too much? In: Advances in Cryptology – ASIACRYPT 2009. Lecture Notes in Computer Science, vol. 5912, pp. 487–504. Springer (2009). https://doi.org/10.1007/978-3-642-10366-7_29
23. Herrmann, M., May, A.: Maximizing small root bounds by linearization and applications to small secret exponent RSA. In: Public Key Cryptography – PKC 2010. Lecture Notes in Computer Science, vol. 6056, pp. 53–69. Springer (2010). https://doi.org/10.1007/978-3-642-13013-7_4
24. Jochemsz, E., May, A.: A strategy for finding roots of multivariate polynomials with new applications in attacking RSA variants. In: Advances in Cryptology – ASIACRYPT 2006. Lecture Notes in Computer Science, vol. 4284, pp. 267–282. Springer (2006). https://doi.org/10.1007/11935230_18
25. Kurosawa, K., Takagi, T.: One-wayness equivalent to general factoring. IEEE Transactions on Information Theory **55**(9), 4249–4262 (2009). https://doi.org/10.1109/TIT.2009.2025532
26. May, A.: Cryptanalysis of unbalanced RSA with small CRT-exponent. In: Advances in Cryptology – CRYPTO 2002. Lecture Notes in Computer Science, vol. 2442, pp. 242–256. Springer (2002). https://doi.org/10.1007/3-540-45708-9_16
27. Meers, J., Nowakowski, J.: Solving the hidden number problem for CSIDH and CSURF via automated Coppersmith. In: Advances in Cryptology – ASIACRYPT 2023, Part IV. Lecture Notes in Computer Science, vol. 14441, pp. 39–71. Springer (2023). https://doi.org/10.1007/978-981-99-8730-6_2
28. Ryan, K.: Solving multivariate Coppersmith problems with known moduli. In: Advances in Cryptology – EUROCRYPT 2025, Part VI. Lecture Notes in Computer Science, vol. 15606, pp. 355–384. Springer (2025). https://doi.org/10.1007/978-3-031-91095-1_13
29. Shoup, V.: OAEP reconsidered. Journal of Cryptology **15**(4), 223–249 (2002). https://doi.org/10.1007/s00145-002-0133-9
30. Takayasu, A., Kunihiro, N.: Partial key exposure attacks on RSA: Achieving the Boneh–Durfee bound. Theoretical Computer Science **761**, 51–77 (2019). https://doi.org/10.1016/j.tcs.2018.08.021
31. Takayasu, A., Lu, Y., Peng, L.: Small CRT-exponent RSA revisited. Journal of Cryptology **32**(4), 1337–1382 (2019). https://doi.org/10.1007/s00145-018-9282-3
32. Xu, J., Sarkar, S., Hu, L., Wang, H., Pan, Y.: New results on modular inversion hidden number problem and inversive congruential generator. In: Advances in Cryptology – CRYPTO 2019, Part I. Lecture Notes in Computer Science, vol. 11692, pp. 297–321. Springer (2019). https://doi.org/10.1007/978-3-030-26948-7_11
33. Xu, J., Sarkar, S., Hu, L., Wang, H., Pan, Y.: Revisiting modular inversion hidden number problem and its applications. IEEE Transactions on Information Theory **69**(8), 5337–5356 (2023). https://doi.org/10.1109/TIT.2023.3263485
34. Xu, J., Sarkar, S., Wang, H., Hu, L.: New results on elliptic curve hidden number problem for ECDH key exchange. Journal of Cryptology **39** (2026). https://doi.org/10.1007/s00145-026-09577-6, article 20
35. Yamakawa, T., Yamada, S., Hanaoka, G., Kunihiro, N.: Adversary-dependent lossy trapdoor function from hardness of factoring semi-smooth RSA subgroup

moduli. In: Advances in Cryptology – CRYPTO 2016, Part II. Lecture Notes in Computer Science, vol. 9815, pp. 3–32. Springer (2016). https://doi.org/10.1007/978-3-662-53008-5_1

A Supplementary Proofs

A.1 Proof of Lemma 3

Proof. We first prove homogeneity. For any $t > 0$, the definition directly gives

$$\delta(t\alpha, t\beta) = t^{-1}\delta(\alpha, \beta).$$

Thus, δ is homogeneous of degree -1.

We next prove monotonicity. If $\alpha' \geq \alpha$ and $\beta' \geq \beta$, then for every S we have $\alpha' W_1(S) + \beta' W_2(S) \geq \alpha W_1(S) + \beta W_2(S)$, and hence $\delta_S(\alpha', \beta') \leq \delta_S(\alpha, \beta)$. Taking the supremum over S gives $\delta(\alpha', \beta') \leq \delta(\alpha, \beta)$. Thus, δ is nonincreasing in each variable.

We now prove convexity. Take any $(\alpha_1, \beta_1), (\alpha_2, \beta_2) \in \mathbb{R}^2_{>0}$ and $\lambda \in [0, 1]$. Since the function $x \mapsto x^{-1}$ is convex on $\mathbb{R}_{>0}$, for every S we have

$$\begin{aligned}\delta_S\big(\lambda\alpha_1 + (1-\lambda)\alpha_2, \lambda\beta_1 + (1-\lambda)\beta_2\big) \\ \leq \lambda\delta_S(\alpha_1, \beta_1) + (1-\lambda)\delta_S(\alpha_2, \beta_2).\end{aligned}$$

Therefore,

$$\begin{aligned}&\delta\big(\lambda\alpha_1 + (1-\lambda)\alpha_2, \lambda\beta_1 + (1-\lambda)\beta_2\big) \\ &\quad\leq \sup_S\big(\lambda\delta_S(\alpha_1, \beta_1) + (1-\lambda)\delta_S(\alpha_2, \beta_2)\big) \\ &\quad\leq \lambda\delta(\alpha_1, \beta_1) + (1-\lambda)\delta(\alpha_2, \beta_2).\end{aligned}$$

Hence, δ is convex.

Finally, we prove continuity. Fix $(\alpha, \beta) \in \mathbb{R}^2_{>0}$ and let $(\alpha', \beta') \to (\alpha, \beta)$. Set

$$\varepsilon = \max\left\{\left|\frac{\alpha'}{\alpha} - 1\right|, \left|\frac{\beta'}{\beta} - 1\right|\right\}.$$

When (α', β') is sufficiently close to (α, β), we have $0 \leq \varepsilon < 1$, and the following inequalities hold coordinatewise:

$$(1-\varepsilon)(\alpha, \beta) \leq (\alpha', \beta') \leq (1+\varepsilon)(\alpha, \beta).$$

By the monotonicity and homogeneity of δ,

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

As $(\alpha', \beta') \to (\alpha, \beta)$, we have $\varepsilon \to 0$, so the squeeze theorem gives $\delta(\alpha', \beta') \to \delta(\alpha, \beta)$. Thus, δ is continuous on $\mathbb{R}^2_{>0}$.

A.2 Proof of corollary 1

Proof. Take any two points (α_1, β_1) and (α_2, β_2) on the curve $\{(\alpha, \beta) \in \mathbb{R}^2_{>0} : \delta(\alpha, \beta) = 1\}$. For any $\lambda \in [0, 1]$, let $(\alpha_\lambda, \beta_\lambda) = \lambda(\alpha_1, \beta_1) + (1 - \lambda)(\alpha_2, \beta_2)$. Since $\mathbb{R}^2_{>0}$ is convex, we have $(\alpha_\lambda, \beta_\lambda) \in \mathbb{R}^2_{>0}$. By the convexity of δ established in Lemma 2.11,

$$\delta(\alpha_\lambda, \beta_\lambda) \le \lambda\delta(\alpha_1, \beta_1) + (1 - \lambda)\delta(\alpha_2, \beta_2) = 1.$$

On the other hand, since $\mathrm{FR} = \{(\alpha, \beta) \in \mathbb{R}^2_{>0} : \delta(\alpha, \beta) > 1\}$, we have $(\alpha_\lambda, \beta_\lambda) \notin \mathrm{FR}$. Therefore, the entire chord joining (α_1, β_1) and (α_2, β_2) is contained in $\mathbb{R}^2_{>0} \setminus \mathrm{FR}$.

A.3 Proof of Lemma 10

Proof. For any $u \in \mathbb{Z}^2_{\ge 0}$ and $j \in \mathbb{Z}_{\ge 0}$, using $S_n = nC \cap \mathbb{Z}^2_{\ge 0}$, the convexity of C, and $\mathrm{conv}(jA(f)) = j\bar{N}(f)$, we obtain

$$u + jA(f) \subseteq S_n \quad \Longleftrightarrow \quad u + jN(f) \subseteq nC \quad \Longleftrightarrow \quad (u, j) \in n\widetilde{\mathcal{H}}(C).$$

The layer $j = 0$ consists precisely of the points $(u, 0)$ with $u \in S_n$. Since $H(S_n)$ counts only $j \ge 1$, we have

$$H(S_n) = \#\big(n\widetilde{\mathcal{H}}(C) \cap \mathbb{Z}^3\big) - \#S_n. \tag{31}$$

Since C and $\widetilde{\mathcal{H}}(C)$ are bounded polyhedra, the convergence of grid Riemann sums [**?**], together with $w_{\alpha,\beta}(s) = n\, w_{\alpha,\beta}(s/n)$, gives

$$\frac{\#\big(n\widetilde{\mathcal{H}}(C) \cap \mathbb{Z}^3\big)}{n^3} \longrightarrow \Psi(C),$$
$$\frac{\sum_{s \in S_n} w_{\alpha,\beta}(s)}{n^3} = \frac{1}{n^2} \sum_{s \in S_n} w_{\alpha,\beta}(s/n) \longrightarrow \int_C w_{\alpha,\beta}(x)\, dx = \Phi(C).$$

Since C is bounded, we also have $\#S_n = O(n^2)$. Together with (31), this yields $H(S_n)/n^3 \to \Psi(C)$.

The two limits above and the definition of $\delta(\alpha, \beta)$ therefore give

$$\frac{\Psi(C)}{\Phi(C)} = \lim_{n \to \infty} \frac{H(S_n)}{\sum_{s \in S_n} w_{\alpha,\beta}(s)} \le \delta(\alpha, \beta).$$

□

A.4 Proof of Lemma 11

Proof. Fix $b \in B_f$ and an edge index i. Parameterize the ith edge by $z(s) = z_{i-1} + sD_i$, and set $n_i = \nu_i / \|\nu_i\|^2$. Since $\nu_i = ((D_i)_2, -(D_i)_1)$, we have $\langle \nu_i, n_i \rangle = 1$

and $|\det(D_i, n_i)| = 1$. Moreover, since B_f is open, we have $b+\rho e_i \in B_f$ whenever $|\rho|$ is sufficiently small.

We first compute the derivative of Φ. Make the change of variables $x = z(s) + \rho n_i$. Then $\langle \nu_i, x\rangle = b_i + \rho$, and the area element is $dx = ds\, d\rho$. For each fixed ρ, let $I(\rho)$ be the interval of parameters s for which x satisfies the remaining supporting inequalities and the coordinate constraints. As s ranges over $I(\rho)$, the point x traces precisely the ith edge of $P(b + \rho e_i)$. In particular, $I(0) = [0, \ell_i]$.

Changing the support value b_i to $b_i + \varepsilon$ adds or removes only the region swept out by the ith supporting line during this motion. We therefore obtain

$$\Phi(b + \varepsilon e_i) - \Phi(b) = \int_0^{\varepsilon} \left(\int_{I(\rho)} w_{\alpha,\beta}(z(s) + \rho n_i)\, ds \right) d\rho. \tag{32}$$

Here and below, integrals with $\varepsilon < 0$ are interpreted as oriented integrals.

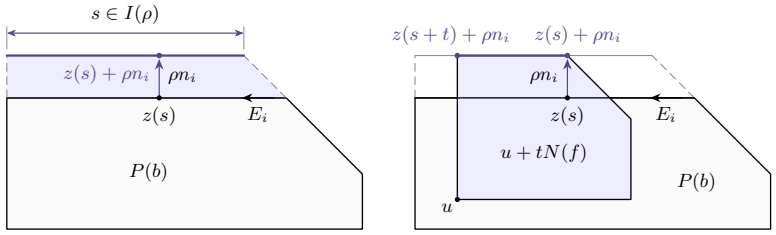
The endpoints of $I(\rho)$ correspond to the intersections of the shifted ith supporting line with the adjacent supporting lines or coordinate axes. The normal vectors of these lines are fixed, and adjacent lines are not parallel. Hence the intersection points, and therefore the endpoints of $I(\rho)$, depend affinely on ρ. Together with the continuity of $w_{\alpha,\beta}$, this gives

$$\int_{I(\rho)} w_{\alpha,\beta}(z(s) + \rho n_i)\, ds \longrightarrow \int_0^{\ell_i} w_{\alpha,\beta}(z(s))\, ds = \phi_i(b) \qquad (\rho \to 0).$$

Dividing (32) by ε and letting $\varepsilon \to 0$ yields $\partial_i \Phi(b) = \phi_i(b)$. Moreover, by the linearity of the weight function,

$$\phi_i(b) = \frac{\ell_i(b)}{2} \big(w_{\alpha,\beta}(z_{i-1}(b)) + w_{\alpha,\beta}(z_i(b)) \big).$$

Since the vertices and ℓ_i depend linearly on b, the function ϕ_i is continuous on B_f.



(a) $\partial_i \Phi(b) = \phi_i(b)$ (b) $\partial_i \Psi(b) = \psi_i(b)$

Next, we compute the derivative of Ψ. Since $k_h = \max_{v \in N(f)} \langle \nu_h, v\rangle$, we have

$$\widetilde{\mathcal{H}}(P(b)) = \big\{ (u,t) : u \geq 0,\ t \geq 0,\ \langle \nu_h, u\rangle + k_h t \leq b_h\ (1 \leq h \leq r) \big\}.$$

To make the ith supporting constraint depend on a single coordinate, use the change of variables

$$u = z(s) - t v_{i-1} + \rho n_i.$$

Using $\langle \nu_i, v_{i-1} \rangle = k_i$, we obtain $\langle \nu_i, u \rangle + k_i t = b_i + \rho$. Thus, the original ith constraint becomes $\rho \le 0$, whereas replacing b_i with $b_i + \varepsilon$ changes this constraint to $\rho \le \varepsilon$.

This three-dimensional transformation leaves the coordinate t unchanged. Its Jacobian determinant has absolute value $|\det(D_i, n_i)| = 1$, so the volume element is $du_1\, du_2\, dt = ds\, dt\, d\rho$. For each fixed ρ, let $T_b(\rho)$ be the region of parameter pairs (s,t) satisfying $t \ge 0$, $u \ge 0$, and the remaining supporting inequalities. Integrating the cross-sectional areas gives

$$\Psi(b + \varepsilon e_i) - \Psi(b) = \int_0^{\varepsilon} \mathrm{vol}_2(T_b(\rho))\, d\rho. \tag{33}$$

We next identify the cross section at $\rho = 0$ with Q_i. For $(s,t) \in T_b(\rho)$, the ith support value of the copy is exactly $b_i + \rho$, so $u + tN(f) \subseteq P(b + \rho e_i)$. The two points in this copy corresponding to v_{i-1} and v_i are

$$u + tv_{i-1} = z(s) + \rho n_i, \qquad u + tv_i = z(s+t) + \rho n_i.$$

Both points lie on the ith edge of $P(b + \rho e_i)$. In particular, when $\rho = 0$, we have $0 \le s \le s + t \le \ell_i$ and $z(s) + t(N(f) - v_{i-1}) \subseteq P(b)$. Hence $(s, s+t) \in Q_i$.

Conversely, suppose that $(s, s+t) \in Q_i$, and set $u = z(s) - tv_{i-1}$. By the definition of Q_i, we have $t \ge 0$ and $u + tN(f) \subseteq P(b)$. Since $0 \in N(f)$, we also have $u \in P(b) \subseteq \mathbb{R}^2_{\ge 0}$, so $(s,t) \in T_b(0)$. Thus, the map $(s,t) \mapsto (s, s+t)$ is a bijection from $T_b(0)$ to Q_i. Its determinant is 1, and therefore

$$\mathrm{vol}_2(T_b(0)) = \mathrm{vol}_2(Q_i) = \psi_i(b). \tag{34}$$

It remains to verify continuity of the cross-sectional area. To treat variations in ρ and b simultaneously, write out the constraints defining $T_b(\rho)$ as

$$\begin{aligned} t &\ge 0, \\ -sD_i + tv_{i-1} &\le z_{i-1}(b) + \rho n_i, \\ \langle \nu_h, D_i \rangle s + \big(k_h - \langle \nu_h, v_{i-1} \rangle\big) t &\le b_h - \langle \nu_h, z_{i-1}(b) \rangle - \rho \langle \nu_h, n_i \rangle \qquad (h \ne i). \end{aligned} \tag{35}$$

The second line is interpreted coordinatewise. The coefficients on the left-hand sides are fixed, while the right-hand sides are affine functions of (b, ρ).

For each scalar constraint, the coefficients of s and t are not both zero. Indeed, if both coefficients in a coordinate constraint were zero, then $[v_{i-1}, v_i]$ would lie on the corresponding coordinate axis, contradicting the fact that it is a noncoordinate edge. If both coefficients in a supporting constraint with $h \ne i$ were zero, then $\langle \nu_h, v_{i-1} \rangle = \langle \nu_h, v_i \rangle = k_h$. This would place the ith edge on the supporting line of the hth edge, which is also impossible.

On the other hand, if b' is sufficiently close to b and $|\rho|$ is sufficiently small, then $b' + \rho e_i \in B_f$. For any point in $T_{b'}(\rho)$, the two contact points described above lie on the ith edge of $P(b' + \rho e_i)$. Since the vertices depend linearly on the parameters, these contact points and $z_{i-1}(b')$ are uniformly bounded. Their expressions, together with $D_i \ne 0$, show that s and $s + t$ are uniformly bounded,

and hence so are s and t. Thus, all these cross sections lie in a fixed bounded rectangle K.

As $(b', \rho) \to (b, 0)$, the right-hand sides of the constraints in (35) converge to their values at $(b, 0)$. If a point belongs to exactly one of $T_{b'}(\rho)$ and $T_b(0)$, then it satisfies at least one constraint for one parameter choice but not the other. The point therefore lies between the boundary lines of that constraint for the two parameter choices. Since the coefficients on the left-hand side are fixed and not both zero, these lines are parallel, and the distance between them tends to zero as the difference between the right-hand sides tends to zero. Thus, the symmetric difference of the two cross sections is covered by finitely many strips whose widths tend to zero. The total area of the intersections of these strips with K tends to zero. Consequently,

$$\mathrm{vol}_2(T_{b'}(\rho)) \longrightarrow \mathrm{vol}_2(T_b(0)) \qquad \big((b', \rho) \to (b, 0)\big).$$

Setting $b' = b$ in this limit and using (33) and (34), we obtain

$$\frac{\partial \Psi}{\partial b_i}(b) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^{\varepsilon} \mathrm{vol}_2(T_b(\rho))\, d\rho = \mathrm{vol}_2(T_b(0)) = \psi_i(b).$$

Setting $\rho = 0$, the same continuity result gives $\psi_i(b') = \mathrm{vol}_2(T_{b'}(0)) \to \mathrm{vol}_2(T_b(0)) = \psi_i(b)$. Hence ψ_i is continuous on B_f.

Since $b \in B_f$ and the edge index i were arbitrary, all partial derivatives of Ψ, Φ exist and are continuous. Therefore, $\Psi, \Phi \in C^1(B_f)$. □

A.5 Proof of Lemma 13

Proof. **Obtaining a nondegenerate maximizer.** Consider the finite set of normal vectors $U = \{\nu_1, \ldots, \nu_r, -e_1, -e_2\}$. Let $\mathcal{P}$ be the family of all compact convex polygons in the nonnegative quadrant with nonempty interior whose outward edge normals are positive multiples of vectors in U. Some prescribed edge directions may be absent, and the supporting lines in the coordinate directions may be translated. Thus, $N(f)$ and all quasi-Newton polygons belong to $\mathcal{P}$.

Since Ψ and Φ are homogeneous of degree three, uniform scaling leaves $\Psi(C)/\Phi(C)$ unchanged. We may therefore normalize each $C \in \mathcal{P}$ so that $\max_{x \in C} w_{\alpha,\beta}(x) = 1$. All normalized polygons then lie in the fixed compact triangle $\Omega = \{x \in \mathbb{R}^2_{\geq 0} : w_{\alpha,\beta}(x) \leq 1\}$. Choose a vertex z satisfying $w_{\alpha,\beta}(z) = 1$, and triangulate C from z. The average value of an affine function over a triangle equals the average of its values at the three vertices. Since the weight function is nonnegative, we obtain $\Phi(C) \geq \mathrm{vol}_2(C)/3$.

If $(u, t) \in \widetilde{\mathcal{H}}(C)$, then $0 \in N(f)$ implies $u \in C$, while $u + tN(f) \subseteq C$ gives $t^2\, \mathrm{vol}_2(N(f)) \leq \mathrm{vol}_2(C)$. Hence

$$\Psi(C) \leq \mathrm{vol}_2(C) \sqrt{\frac{\mathrm{vol}_2(C)}{\mathrm{vol}_2(N(f))}}, \qquad \frac{\Psi(C)}{\Phi(C)} \leq 3 \sqrt{\frac{\mathrm{vol}_2(C)}{\mathrm{vol}_2(N(f))}}. \tag{36}$$

This shows both that the ratio is bounded and that, after normalization, it tends to zero whenever the area tends to zero.

Choose a normalized maximizing sequence (C_n) such that $\Psi(C_n)/\Phi(C_n) \to c := \sup_{C\in\mathcal{P}} \Psi(C)/\Phi(C)$. Since $N(f) \in \mathcal{P}$ and $\Psi(N(f)) > 0$, we have $c > 0$. Each C_n has at most $r + 2$ edges. There are only finitely many possible sets of edge directions. By first passing to a subsequence with a fixed set of edge directions and then to a further subsequence along which the vertices converge, we may label the vertices of each C_n as $p_{n,1}, \dots, p_{n,s}$ in a common counterclockwise order, with each $p_{n,j}$ converging to some $p_j \in \Omega$. Set $C = \operatorname{conv}\{p_1, \dots, p_s\}$. Convex combinations with the same coefficients differ by at most $\max_j \|p_{n,j} - p_j\|$. Thus, the distance from any point of either set to the other set is at most $\max_j \|p_{n,j} - p_j\|$, and these distances tend uniformly to zero. The shoelace formula also gives

$$\operatorname{vol}_2(C_n) = \frac{1}{2}\sum_{j=1}^{s} \det(p_{n,j}, p_{n,j+1}) \longrightarrow \frac{1}{2}\sum_{j=1}^{s} \det(p_j, p_{j+1}) = \operatorname{vol}_2(C),$$

where the indices are interpreted cyclically; repeated or collinear vertices in the limit do not affect this formula. Combining (36) with $c > 0$ yields $\operatorname{vol}_2(C) > 0$.

To verify that the limit still belongs to $\mathcal{P}$, write $h_D(\nu) = \max_{x\in D}\langle \nu, x\rangle$ for a compact convex set D. The convergence of the individual vertices gives $h_{C_n}(\nu) \to h_C(\nu)$, and

$$C = \bigcap_{\nu\in U} \{x : \langle \nu, x\rangle \le h_C(\nu)\}. \tag{37}$$

The inclusion from left to right is immediate. For the reverse inclusion, take a point x on the right-hand side and a point $p \in \operatorname{int}(C)$. For each $0 < \theta < 1$, the point $(1 - \theta)x + \theta p$ satisfies all the inequalities on the right strictly, and therefore belongs to C_n for all sufficiently large n. By the convergence of the sets established above, it also belongs to C. Letting $\theta \to 0$ gives $x \in C$. Equation (37) shows that the edge directions of C remain among the allowed directions, so $C \in \mathcal{P}$.

Next, we verify convergence of the integrals. Using the same half-plane representation, we have $(u, t) \in \widetilde{\mathcal{H}}(C_n)$ if and only if

$$u \ge 0, \qquad t \ge 0, \qquad \langle \nu, u\rangle + t h_{N(f)}(\nu) \le h_{C_n}(\nu) \quad (\nu \in U).$$

In all these regions, u lies in Ω, and t is bounded above by $\sqrt{\operatorname{vol}_2(\Omega)/\operatorname{vol}_2(N(f))}$. The left-hand sides of the constraints are fixed, and the right-hand sides converge. The linear coefficients in each constraint are not all zero, so we may apply the same thin-strip estimate used in the proof of the boundary derivative formula: as the changes in the right-hand sides tend to zero, the symmetric difference is contained in a union of finitely many slabs whose thicknesses tend to zero. Within the common bounded cylinder, the total volume of these slabs tends to zero. Hence $\Psi(C_n) \to \Psi(C)$. Applying the corresponding two-dimensional strip estimate to C_n, and using the boundedness of the weight function on Ω, gives $\Phi(C_n) \to \Phi(C) > 0$. Therefore, $\Psi(C)/\Phi(C) = c$. The polygon C maximizes the

ratio over the entire family $\mathcal{P}$, so the perturbations below need not preserve the normalization.

Establishing the two coordinate edges. Set $m = (\min_{x\in C} x_1, \min_{x\in C} x_2)$. Then $m \geq 0$ and $C - m \in \mathcal{P}$. Since every $(u,t) \in \widetilde{\mathcal{H}}(C)$ satisfies $u \in C$, we have $u - m \geq 0$. Thus, the map $(u,t) \mapsto (u-m,t)$ is a volume-preserving bijection between $\widetilde{\mathcal{H}}(C)$ and $\widetilde{\mathcal{H}}(C-m)$, and

$$\Psi(C-m) = \Psi(C), \qquad \Phi(C-m) = \Phi(C) - w_{\alpha,\beta}(m)\,\mathrm{vol}_2(C).$$

If $m \neq 0$, the denominator decreases strictly, contradicting maximality. Hence $m = 0$, so C meets both coordinate axes.

Suppose that $C \cap \{x_1 = 0\}$ consists of a single vertex. For sufficiently small $\varepsilon > 0$, set $C_\varepsilon = (C \cap \{x_1 \geq \varepsilon\}) - \varepsilon e_1$. The only new edge this operation can introduce has outward normal $-e_1$, so $C_\varepsilon \in \mathcal{P}$. The line $x_1 = \varepsilon$ meets the two edges incident to this vertex, and the removed region is a triangle of area $O(\varepsilon^2)$. Since $\min_{v\in N(f)} v_1 = 0$, a previously feasible copy $u + tN(f)$ is excluded by the new constraint if and only if $u_1 < \varepsilon$. For the retained copies, the map $(u,t) \mapsto (u - \varepsilon e_1, t)$ is a volume-preserving bijection between the parameter region after truncation and $\widetilde{\mathcal{H}}(C_\varepsilon)$. For the excluded copies, u lies in the removed triangle, and t has a fixed upper bound. Hence $0 \leq \Psi(C) - \Psi(C_\varepsilon) = O(\varepsilon^2)$. The weighted integral over the removed region is also $O(\varepsilon^2)$, while translating to the left decreases the weight of each retained point by $\alpha\varepsilon$. Therefore,

$$\Phi(C_\varepsilon) = \Phi(C) - \alpha\,\mathrm{vol}_2(C)\varepsilon + O(\varepsilon^2).$$

Using $\Psi(C) = c\Phi(C)$, we obtain

$$\Psi(C_\varepsilon) - c\Phi(C_\varepsilon) = c\alpha\,\mathrm{vol}_2(C)\varepsilon + O(\varepsilon^2) > 0$$

for sufficiently small ε, contradicting maximality. Thus, $C \cap \{x_1 = 0\}$ has positive length. Exchanging the two coordinates similarly shows that $C \cap \{x_2 = 0\}$ has positive length.

Now set $b_i = h_C(\nu_i)$. Each normal ν_i of a noncoordinate edge has at least one positive component. Otherwise, its support face on $N(f) \subseteq \mathbb{R}^2_{\geq 0}$ would lie on a coordinate axis or reduce to the origin. Since C contains nonzero points on both coordinate axes, we have $b_i > 0$. Together with (37), this gives $C = P(b)$. It remains to show that every prescribed noncoordinate edge has positive length.

Ruling out missing noncoordinate edges. Suppose, for a contradiction, that the supporting line with normal ν_i meets C only at a vertex z for some i. Since $\langle \nu_i, z\rangle = b_i > 0$, we have $z \neq 0$ and therefore $w_{\alpha,\beta}(z) > 0$. Let $\mu_1, \mu_2 \in U$ be outward normal vectors of the two edges incident to z. They are linearly independent, and there exist $\lambda_1, \lambda_2 > 0$ such that $\nu_i = \lambda_1\mu_1 + \lambda_2\mu_2$. Indeed, as we move away from z along the first incident edge, the inner product with μ_1 remains constant, whereas the inner products with μ_2 and ν_i decrease strictly. Hence $\lambda_2 > 0$. Moving along the other incident edge similarly gives $\lambda_1 > 0$.

Set

$$\Delta = \lambda_1 h_{N(f)}(\mu_1) + \lambda_2 h_{N(f)}(\mu_2) - k_i.$$

For any point v on the ith edge of $N(f)$, we have

$$\Delta = \sum_{j=1}^{2} \lambda_j \big(h_{N(f)}(\mu_j) - \langle \mu_j, v\rangle\big) \geq 0.$$

If $\Delta = 0$, every point on this edge would lie on both lines $\langle \mu_j, v\rangle = h_{N(f)}(\mu_j)$ for $j = 1, 2$. These lines are not parallel and intersect in at most one point, contradicting the positive length of the original edge. Therefore, $\Delta > 0$.

For sufficiently small $\varepsilon > 0$, set $C_\varepsilon = C \cap \{x : \langle \nu_i, x\rangle \leq b_i - \varepsilon\}$. This polygon still belongs to $\mathcal{P}$. The removed triangle scales linearly with ε along the two fixed incident edges. Its area is therefore $\kappa\varepsilon^2$, where $\kappa > 0$ is independent of ε. Every point in this triangle is at distance $O(\varepsilon)$ from z, so

$$\Phi(C) - \Phi(C_\varepsilon) = \kappa w_{\alpha,\beta}(z)\varepsilon^2 + O(\varepsilon^3). \tag{38}$$

To estimate the loss in Ψ, take any $(u, t) \in \widetilde{\mathcal{H}}(C) \setminus \widetilde{\mathcal{H}}(C_\varepsilon)$. The fact that the new constraint excludes this copy means that

$$0 \leq b_i - \langle \nu_i, u\rangle - k_i t < \varepsilon. \tag{39}$$

On the other hand, multiplying the constraints from the two incident edges by λ_1, λ_2 and adding them gives

$$\begin{aligned}\langle \nu_i, u\rangle + (k_i + \Delta)t &= \sum_{j=1}^{2} \lambda_j \big(\langle \mu_j, u\rangle + t h_{N(f)}(\mu_j)\big) \\ &\leq \sum_{j=1}^{2} \lambda_j h_C(\mu_j) = \langle \nu_i, z\rangle = b_i.\end{aligned}$$

Combining this with (39) yields $\Delta t < \varepsilon$. Since $k_i \geq 0$ and $u \in C$, we also have

$$0 \leq t < \frac{\varepsilon}{\Delta}, \qquad 0 \leq b_i - \langle \nu_i, u\rangle < \left(1 + \frac{k_i}{\Delta}\right)\varepsilon.$$

The latter condition confines u to a triangle of area $\kappa(1 + k_i/\Delta)^2\varepsilon^2$, while the former confines t to an interval of length ε/Δ. Hence

$$0 \leq \Psi(C) - \Psi(C_\varepsilon) \leq \frac{\kappa}{\Delta}\left(1 + \frac{k_i}{\Delta}\right)^2 \varepsilon^3. \tag{40}$$

Equations (38)–(40), together with $\Psi(C) = c\Phi(C)$, give

$$\Psi(C_\varepsilon) - c\Phi(C_\varepsilon) = c\kappa w_{\alpha,\beta}(z)\varepsilon^2 + O(\varepsilon^3) > 0$$

for sufficiently small ε, again contradicting maximality. Thus, all prescribed noncoordinate edges have positive length, so $b \in B_f$.

Set $b^* = b$. Then $P(b^*)$ attains the maximum ratio among all quasi-Newton polygons. Since B_f is open and Ψ/Φ is continuously differentiable there, we have $\nabla(\Psi/\Phi)(b^*) = 0$. By the boundary derivative formula, the corresponding edges satisfy the balance equalities with a common constant. This completes the proof. □