CONVEX NETWORKS REMAIN HARD TO CERTIFY: DIMENSION-ACCURACY BARRIERS FOR LIPSCHITZ CONSTANTS

Pahan Dewasurendra
Johns Hopkins University

Subhashini Jayawardhana
San Diego Miramar College

ABSTRACT

Input-convex neural networks permit globally tractable minimization over their inputs, so one might expect their global regularity to be tractable in low input dimension. We prove exact and accuracy-sensitive barriers to this expectation. Given a bias-free one-hidden-layer ReLU network

$$f(x) = \sum_{r=1}^{n} \mathrm{ReLU}(a_r^\top x)$$

with unit positive output weights, deciding whether its global Euclidean Lipschitz constant is at least a rational threshold is NP-complete and W[1]-hard when parameterized by the input dimension d. The same holds on the unit ball and with integral first-layer weights having at most nine nonzeros. More sharply, no deterministic multiplicative approximation scheme runs in $g(d)\,\mathrm{poly}(\mathcal{B}, 1/\varepsilon)$ time unless FPT equals W[1]. Under the Exponential Time Hypothesis, no such algorithm runs in $g(d)(\mathcal{B} + 1/\varepsilon)^{o(d/\log d)}$ time. Thus accuracy cannot have a polynomial dependence separated from dimension.

The exact result resolves the Euclidean case of an open problem posed at COLT 2025 and left open by the ICLR 2026 parameterized hardness theory for general two-layer networks. The approximation barrier is specific to generator-presented zonotopes, complementing known $(1/\varepsilon)^{O(d)}$-time schemes and an analogous barrier for halfspace-presented polytopes. Our lifted-selector reduction has an inverse-polynomial radial gap, proved through a quantitative theorem for rational cyclic zonogons. Equivalently, the results apply to Euclidean zonotope radius and positive-semidefinite binary quadratic maximization parameterized by rank. Convexity makes minimization easy, but it does not make global sensitivity fixed-parameter tractable or permit a dimension-separated fully polynomial accuracy guarantee.

1 INTRODUCTION

The global Lipschitz constant measures the worst amplification of an input perturbation. It is used to study robustness, generalization, fairness, and stability of neural networks (Virmaux & Scaman, 2018; Jordan & Dimakis, 2020). Exact computation is difficult for general ReLU networks even with one hidden layer (Virmaux & Scaman, 2018). This motivates both certified upper bounds and exact exponential-time methods (Jordan & Dimakis, 2020).

Low input dimension offers a natural route around worst-case hardness. A bias-free two-layer ReLU network induces a central hyperplane arrangement in $\mathbb{R}^d$. Enumerating its cells takes $n^{O(d)}\,\mathrm{poly}(\mathcal{B})$ time, so exact verification and Lipschitz computation are polynomial-time solvable for every fixed d. Approximation improves this dependence: known schemes obtain $(1/\varepsilon)^{O(d)}\,\mathrm{poly}(n, \mathcal{B})$ time for multiplicative accuracy ε (Shenmaier, 2018; Froese et al., 2026). Two parameterized questions follow. Can exact computation have running time $g(d)\mathcal{B}^{O(1)}$? Can approximation have running time $g(d)\,\mathrm{poly}(\mathcal{B}, 1/\varepsilon)$, separating dimension from a fully polynomial accuracy dependence?

This question is especially compelling for input-convex neural networks (ICNNs) (Amos et al., 2017). Consider the minimal architecture

$$f_A(x) = \sum_{r=1}^{n} \mathrm{ReLU}(a_r^\top x), \qquad a_r \in \mathbb{Q}^d. \tag{1}$$

All output weights are positive, hence f_A is convex. It is also the support function of the zonotope

$$Z(A) = \sum_{r=1}^{n} [0, a_r] = \left\{ \sum_{r=1}^{n} \lambda_r a_r : \lambda \in [0,1]^n \right\}. \tag{2}$$

Consequently,

$$\mathrm{Lip}_2(f_A) = \max_{z \in Z(A)} \|z\|_2 = \max_{s \in \{0,1\}^n} \|As\|_2. \tag{3}$$

The last expression is positive-semidefinite binary quadratic maximization with a matrix of rank at most d.

Froese et al. (2025) highlighted the fixed-parameter complexity of these equivalent problems as an open question. The general two-layer network permits negative output weights and corresponds to a difference of two zonotope support functions. Recent work proved W[1]-hardness in that general setting and resolved the companion zonotope-containment question (Froese et al., 2026). It explicitly left norm maximization over a single zonotope, equivalently the positive-output ICNN case, as its most prominent open question. Existing general-network reductions fundamentally use cancellation between positive and negative output weights. They therefore do not apply to Equation (1).

We answer both questions negatively in the Euclidean case.

Theorem 1 (Main result, informal)**.** *Exact global ℓ_2-Lipschitz certification for the networks in Equation (1) is W[1]-hard with respect to input dimension. Moreover, unless FPT equals W[1], no deterministic $(1+\varepsilon)$-approximation has running time $g(d)\,\mathrm{poly}(\mathcal{B}, 1/\varepsilon)$. Both statements hold on the unit ball, with zero biases, integral first-layer weights of support size at most nine, and every output weight equal to one.*

The same theorem settles two equivalent formulations. It proves that the Longest Vector Sum problem is W[1]-hard in the ambient dimension, answering a question of Shenmaier (2020), and gives the first variable-accuracy parameterized barrier for its Euclidean generator representation. It also proves W[1]-hardness and the same approximation barrier for fixed-rank convex binary quadratic maximization. The classical $O(n^{d-1})$ arrangement algorithms remain polynomial for constant rank (Ferrez et al., 2005; Shenmaier, 2020). Quantitatively, both our exact and accuracy-sensitive lower bounds leave only a logarithmic gap in the exponent.

Why the reduction is not immediate. A zonotope is a Minkowski sum. Its generators are independently selected, while a partitioned-subgraph reduction needs an exclusive choice of one host vertex for each pattern vertex and one host edge for each pattern edge. Standard selector gadgets made from simplices are unavailable because a simplex with many vertices is not a two-dimensional zonotope. Further, every additive penalty is squared inside the Euclidean norm, so cancellation can create unintended large points.

Our main technical device is a *lifted selector*. Start with an inscribed centrally symmetric polygon. Its first half of edge vectors generates the polygon as a zonotope. We lift every polygon vertex into extra payload coordinates before taking edge differences. The planar projection is unchanged, but each maximum-radius polygon vertex now carries an arbitrary prescribed payload. Independent selector blocks enforce one candidate per factor, while shared payload blocks compare labels. The new quantitative step proves that every nonvertex endpoint loses at least $1/[6(m^2+1)^2]$ in normalized squared radius for m rational-circle candidates. This inverse-polynomial loss replaces the denominator-based exponentially small bound and is what makes the approximation lower bound possible. Reducing from a cubic pattern uses only $11k+1$ coordinates for a k-vertex source instance.

Contributions. Our results are:

- W[1]-hardness and NP-completeness of Euclidean norm maximization over zonotopes given by generators, parameterized by ambient dimension.
- The same hardness for exact global and unit-ball Euclidean Lipschitz constants of bias-free one-hidden-layer ICNNs with integral first-layer weights, at most nine nonzeros per hidden unit, and unit positive output weights.
- The same hardness for maximizing a positive-semidefinite binary quadratic form, parameterized by its rank, even with an integral factor and threshold.
- Under the Exponential Time Hypothesis (ETH), a $g(d)\mathcal{B}^{o(d/\log d)}$ running-time lower bound in the total input encoding length $\mathcal{B}$, nearly matching the XP upper bound in its exponent.
- No deterministic $g(d)\operatorname{poly}(\mathcal{B}, 1/\varepsilon)$-time multiplicative approximation scheme for these three problems unless FPT equals W[1]. Under ETH, the joint lower bound is $g(d)(\mathcal{B} + 1/\varepsilon)^{o(d/\log d)}$.
- A quantitative lifted-selector lemma that embeds arbitrary labels into a rational zonogon and gives every noncanonical endpoint an inverse-polynomial radial deficit.

The approximation statement is deliberately about variable accuracy. For every fixed ε, deterministic polyhedral approximation and randomized sampling give FPT algorithms (Knauer et al., 2015; Shenmaier, 2018; Froese et al., 2026). Our result shows that the dependence on $1/\varepsilon$ cannot be polynomial with an exponent independent of d. Knauer et al. (2015) proved the analogous phenomenon for norm maximization over halfspace-presented polytopes. Generator-presented zonotopes require a different reduction because they are compact Minkowski sums and support fast specialized enumeration and sampling. Recent work also gives polynomial-time dimension-dependent approximations for broader containment and longest-vector-sum objectives (Eisenbrand et al., 2026).

2 Preliminaries and equivalent formulations

For a matrix $A = (a_1, \dots, a_n) \in \mathbb{Q}^{d\times n}$, define the zonotope $Z(A)$ by Equation (2). We study the following decision problem.

Definition 2 (Euclidean zonotope radius)**.** Given $A \in \mathbb{Q}^{d\times n}$ and $L \in \mathbb{Q}_{>0}$, decide whether

$$\max_{z\in Z(A)} \|z\|_2 \geq L.$$

The parameter is d.

The convexity of the squared norm implies that a maximum over $Z(A)$ is attained at an image of a cube vertex. Therefore

$$\max_{z\in Z(A)} \|z\|_2^2 = \max_{s\in\{0,1\}^n} s^\top A^\top A s. \tag{4}$$

This is often called fixed-rank convex 0–1 quadratic maximization (Ferrez et al., 2005). The rank of $A^\top A$ is at most d.

The neural-network equivalence is also exact and does not require a generic-position assumption.

Proposition 3 (Support-function equivalence)**.** *For f_A in Equation (1),*

$$f_A(x) = \max_{z\in Z(A)} z^\top x \quad \textit{and} \quad \operatorname{Lip}_2(f_A) = \max_{z\in Z(A)} \|z\|_2.$$

Proof. For every r, $\operatorname{ReLU}(a_r^\top x) = \max_{\lambda_r\in[0,1]} \lambda_r a_r^\top x$. Summing the independently optimized terms gives the support function of $Z(A)$. If $R = \max_{z\in Z(A)} \|z\|_2$, then

$$|f_A(x) - f_A(y)| \leq \max_{z\in Z(A)} |z^\top (x-y)| \leq R\|x-y\|_2.$$

Conversely, let $z_\star$ have norm R and set $u = z_\star/R$ when $R > 0$. Positive homogeneity and $f_A(0) = 0$ give $f_A(u) \geq z_\star^\top u = R$. Hence $\operatorname{Lip}_2(f_A) \geq R$. $\square$

The best general exact algorithms enumerate the cells of the central arrangement induced by the generators, or equivalently the vertices of the zonotope. A d-zonotope with n generators has at most $2\sum_{j=0}^{d-1}\binom{n-1}{j}$ vertices, yielding $n^{O(d)}$ time (Ferrez et al., 2005). This places the problem in XP with respect to d. Our result separates this guarantee from FPT under a standard assumption.

3 A ZONOTOPAL SELECTOR WITH ARBITRARY PAYLOADS

We first isolate the geometric gadget. Define rational unit-circle points

$$u(t) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right), \qquad t \in \mathbb{N}_{>0}. \tag{5}$$

They are distinct, lie in a common open semicircle, and occur in circular order as t increases.

Lemma 4 (Lifted selector)**.** *Let $b_1, \ldots, b_m \in \mathbb{Q}^q$ be arbitrary payloads and let $M \in \mathbb{N}_{>0}$. Define*

$$c_j = (Mu(j), b_j) \in \mathbb{Q}^{2+q}.$$

For $m = 1$, set $g_1 = -2c_1$. For $m \geq 2$, set

$$g_j = c_{j+1} - c_j \quad (j < m), \qquad g_m = -c_1 - c_m.$$

Let $S = c_1 + \sum_{j=1}^{m} [0, g_j]$, and let π project onto the first two coordinates. Then:

1. *$\|\pi(x)\|_2 \leq M$ for every $x \in S$.*
2. *Equality holds if and only if $x \in \{\pm c_1, \ldots, \pm c_m\}$.*

Every endpoint sum $x = c_1 + \sum_j \varepsilon_j g_j$ that is not one of these $2m$ points satisfies

$$\|\pi(x)\|_2^2 \leq M^2 \left(1 - \frac{1}{6(m^2+1)^2}\right). \tag{6}$$

Proof. For $m = 1$, the set $S = [-c_1, c_1]$ and all claims are immediate. Assume $m \geq 2$. The sequence

$$Mu(1), \ldots, Mu(m), -Mu(1), \ldots, -Mu(m)$$

is the boundary order of a centrally symmetric polygon P inscribed in the radius-M circle. The vectors $\pi(g_1), \ldots, \pi(g_m)$ are one representative from each opposite pair of boundary edges. Hence

$$\pi(S) = Mu(1) + \sum_{j=1}^{m} [0, \pi(g_j)] = P.$$

Strict convexity of the Euclidean disk gives $\|y\|_2 \leq M$ on P, with equality only at the displayed polygon vertices.

Fix one polygon vertex and choose an exposing direction in the relative interior of its normal cone that is not orthogonal to any $\pi(g_j)$. Such a direction exists because only finitely many lines must be avoided. Maximizing this linear functional over the generating segments therefore selects a unique endpoint of every segment. The full lifted point above each polygon vertex is consequently unique. Telescoping gives

$$c_1 + \sum_{r<j} g_r = c_j, \qquad c_1 + \sum_{r=1}^{m} g_r - \sum_{r<j} g_r = -c_j.$$

This proves the first two claims. For Equation (6), consecutive rational-circle angles are at least $(m^2+1)^{-1}$. A support-loss argument using the cyclic order of the generator lines gives normalized squared-radius deficit at least $2/[\pi^2(m^2+1)^2] > 1/[6(m^2+1)^2]$. Appendix A.2 handles segment ties and the one-mismatch case in full. □

The important feature is that q, the payload dimension, does not affect the selector geometry. The payloads may overlap across gadgets, which lets Euclidean norm compare labels chosen by different factors.

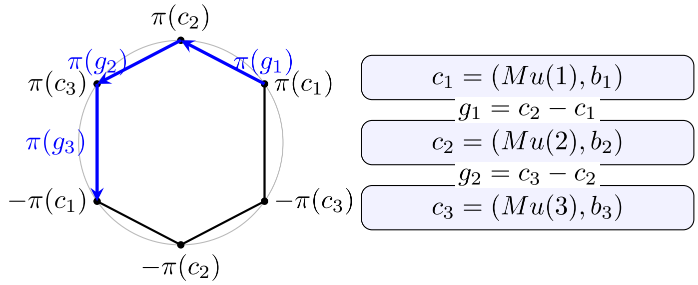

Figure 1: A schematic lifted selector for three candidates. Blue vectors are the first half of the boundary edges and generate the polygon as a translated zonotope. Lifting before taking differences preserves the planar selector and attaches the prescribed payloads.

4 Reduction from Cubic Partitioned Subgraph Isomorphism

An instance of Partitioned Subgraph Isomorphism consists of a pattern graph $P = (U, E_P)$, a host graph $G = (V, E_G)$, and a partition $V = \dot{\bigcup}_{u \in U} V_u$. It asks for a map $\phi : U \to V$ such that $\phi(u) \in V_u$ and $\phi(u)\phi(v) \in E_G$ whenever $uv \in E_P$. In the cubic restriction, every pattern vertex has degree three. This restriction is W[1]-complete with respect to $k = |U|$ (Eppstein & Lokshtanov, 2019). Under ETH, it has no $\rho(k)n^{o(k/\log k)}$-time algorithm for any computable ρ, where n is the host-instance size (Marx, 2010; Eppstein & Lokshtanov, 2019).

Write $m = |E_P| = 3k/2$. We create a *factor* for every pattern vertex and every pattern edge:

$$\mathcal{F} = U \cup E_P, \qquad F = |\mathcal{F}| = k + m = 5k/2.$$

The candidates of vertex factor u are the vertices in V_u. For a pattern edge $e = uv$, the candidates of edge factor e are the host edges

$$E_e = \{xy \in E_G : x \in V_u,\ y \in V_v\}.$$

Let $N = |V| + \sum_{e \in E_P} |E_e|$ be the total number of candidates. Each factor receives a private two-dimensional selector block. If a factor has no candidate, we output a fixed no-instance.

We also create a two-dimensional consistency block for every pattern incidence:

$$\mathcal{I} = \{(u, e) : u \in e \in E_P\}, \qquad I = |\mathcal{I}| = 2m = 3k.$$

Assign every host vertex x a distinct identity code $p_x = u(r_x)$, where $r_x \in [|V|]$. No two identity codes are equal or antipodal.

4.1 Candidate vectors

All candidate vectors initially live in

$$\mathbb{R}^{2F+2I} = \left(\bigoplus_{f \in \mathcal{F}} \mathbb{R}^2_{\mathrm{sel},f} \right) \oplus \left(\bigoplus_{\iota \in \mathcal{I}} \mathbb{R}^2_{\mathrm{con},\iota} \right).$$

Order the candidates of each factor arbitrarily. If $x \in V_u$ is the r-th candidate of vertex factor u, define $c_{u,x}$ by placing $Mu(r)$ in selector block u, placing p_x in every consistency block (u, e) incident to u, and placing zero elsewhere. If $xy \in E_e$, where $e = uv$, is the r-th candidate of edge factor e, define $c_{e,xy}$ by placing $Mu(r)$ in selector block e, placing p_x in block (u, e), placing p_y in block (v, e), and placing zero elsewhere.

Apply Lemma 4 to the candidate list of every factor. This produces one generator per candidate. Embed these generators into their selector block and the shared consistency blocks. Write $c_{f,1}$ for the first candidate vector of factor f.

The lemma describes a translated zonotope. We implement each factor translation with a sparse ordinary generator. Add one fresh anchor coordinate and, for every factor f, define

$$a_f = He_0 + c_{f,1}. \tag{7}$$

The output zonotope is generated by the F anchors a_f and all lifted-selector generators. Its ambient dimension is

$$d = 1 + 2F + 2I = 1 + 2k + 6m = 11k + 1. \tag{8}$$

If an endpoint sum selects anchor a_f, the translated contribution of factor f lies in the set from Lemma 4. If it omits a_f, that factor contributes an untranslated selector sum. We next choose M and H so every radius maximizer selects every anchor and every factor is canonical.

4.2 Polynomial scale separation

Let

$$D_p = 2I, \qquad P = D_p F^2 (2N+1)^2.$$

Every payload coordinate of a candidate has absolute value at most one. Every payload coordinate of a selector generator has absolute value at most two. For any endpoint sum, the magnitude of an aggregate payload coordinate is at most $F + 2N \le F(2N+1)$. Thus its total squared payload norm is at most P.

Let

$$M = 6(N^2+1)(P+1), \qquad H = 4FM + P + 1.$$

These scales have polynomial magnitude, not merely polynomial encoding length. Every factor contains at most N candidates, so Equation (6) gives a selector deficit of at least $M^2/[6(N^2+1)^2] = 6(P+1)^2$ whenever one factor is noncanonical.

Lemma 5 (Canonical maximizers)**.** *Every maximum-radius endpoint sum selects all F anchors. Conditional on selecting all anchors, every factor projection of a maximum-radius endpoint sum is one of its canonical points $\pm c_{f,r}$.*

Proof. If an endpoint sum selects only $q < F$ anchors, the anchor, selector, and payload bounds in Appendix A.5 put its squared norm strictly below $F^2H^2 + FM^2$, which every all-anchored canonical tuple attains or exceeds. A maximizer therefore selects every anchor.

Suppose an all-anchored endpoint sum is noncanonical in at least one factor. By Equation (6), the total selector contribution is at most

$$FM^2 - \frac{M^2}{6(N^2+1)^2} = FM^2 - 6(P+1)^2.$$

Even after adding the maximum payload contribution P, its squared norm is strictly below $F^2H^2 + FM^2$. Any canonical endpoint tuple has squared norm at least $F^2H^2 + FM^2$. Therefore no noncanonical endpoint sum maximizes the radius. □

Since squared Euclidean norm is convex, its maximum over a zonotope is attained at an endpoint sum. Lemma 5 therefore controls the full continuous zonotope.

4.3 Consistency detects a partitioned copy

Consider an all-anchored canonical endpoint tuple. A pattern-vertex factor chooses a signed host vertex $s_u x_u$, and a pattern-edge factor chooses a signed host edge $s_e x_e y_e$, where every sign belongs to $\{-1, 1\}$. For an incidence (u, e), let z be the endpoint of the selected host edge for e that lies in V_u. Consistency block (u, e) contains

$$s_u p_{x_u} + s_e p_z. \tag{9}$$

Its squared norm is at most four. Equality in the triangle inequality holds exactly when $s_u p_{x_u} = s_e p_z$. The identity codes are distinct and nonantipodal, so equality is equivalent to $x_u = z$ and $s_u = s_e$.

There are I consistency blocks. Every canonical tuple has selector squared norm FM^2, so its total squared norm is at most

$$T = F^2H^2 + FM^2 + 4I. \tag{10}$$

It reaches T exactly when every selected host edge has the selected host vertices as endpoints. Such choices are exactly the partition-respecting copies of P in G. We have proved the algebraic core of the reduction.

Lemma 6 (Explicit gap)**.** *The host graph contains a partition-respecting copy of the pattern if and only if the squared radius of the constructed zonotope equals T. If it contains no such copy, the squared radius is at most $T - \delta$, where $\delta > 0$ is rational and has polynomial encoding length.*

Proof. A partitioned copy, with all signs positive, gives equality in every consistency block and reaches T. Conversely, equality forces each selected host edge to agree with the selected host vertices at both endpoints by Equation (9).

Set

$$\delta = \frac{1}{3(|V|^2 + 1)^2}. \tag{11}$$

Appendix A.3 proves that every inconsistent signed identity-code pair has squared-norm deficit at least δ from four. Thus an all-anchored canonical but inconsistent endpoint has squared norm at most $T - \delta$.

An all-anchored noncanonical endpoint loses at least $6(P+1)^2$ in selector squared norm. Even after crediting the entire payload bound P, its deficit from T is greater than one. If an endpoint selects only $q < F$ anchors, the comparison in Appendix A.5 and the inequalities $H \geq 4M$ and $M \geq P+1$ give a deficit from the all-anchored canonical lower bound of at least

$$(F - q)(H^2 - 3M^2) - P \geq 13M^2 - P > 1.$$

Since $\delta < 1$, every no-instance endpoint has squared norm at most $T - \delta$. Convexity again extends the endpoint bound to the full zonotope. $\square$

The standard decision problem requires a rational radius rather than a rational squared radius. This causes no algebraic-number issue. Choose b with $2^{-b} < \delta/(2(T+1))$ and compute

$$L = 2^{-b} \left\lfloor 2^b \sqrt{T} \right\rfloor = 2^{-b} \left\lfloor \sqrt{2^{2b} T} \right\rfloor. \tag{12}$$

Integer square root computes L in polynomial time. Since

$$\sqrt{T} - \sqrt{T - \delta} = \frac{\delta}{\sqrt{T} + \sqrt{T - \delta}} \geq \frac{\delta}{2(T+1)} > 2^{-b},$$

we have $\sqrt{T - \delta} < L \leq \sqrt{T}$. Thus the zonotope radius is at least the rational threshold L exactly in yes-instances.

Theorem 7 (Euclidean zonotope radius)**.** *Euclidean zonotope radius is NP-complete and W[1]-hard parameterized by the ambient dimension d, even when every generator is integral and has at most nine nonzero coordinates. Unless FPT equals W[1], it has no $g(d)\mathcal{B}^{O(1)}$-time algorithm. Under ETH, it has no $g(d)\mathcal{B}^{o(d/\log d)}$-time algorithm for any computable g, where $\mathcal{B}$ denotes the total binary encoding length.*

Proof sketch. Lemma 6 gives a parameterized reduction with $N + F$ generators and $d = 11k + 1$. A binary endpoint is an NP witness. Clearing rational-circle denominators preserves support size nine, and the Cubic Partitioned Subgraph Isomorphism lower bound gives the ETH exponent. Appendix A.4 supplies the encoding and denominator details. $\square$

We now use the inverse-polynomial relative gap. Let $R(A) = \max_{z \in Z(A)} \|z\|_2$. A multiplicative estimate at accuracy $\varepsilon \in (0, 1)$ is a rational value $\widehat{R}$ satisfying

$$\frac{R(A)}{1 + \varepsilon} \leq \widehat{R} \leq (1 + \varepsilon) R(A). \tag{13}$$

This includes algorithms that return a feasible endpoint with value at least $R(A)/(1 + \varepsilon)$.

Theorem 8 (Dimension-accuracy barrier)**.** *Unless FPT equals W[1], no deterministic algorithm produces an estimate satisfying Equation (13) in time $g(d)\,\mathrm{poly}(\mathcal{B}, 1/\varepsilon)$ for any computable g. Under ETH, no deterministic algorithm has running time*

$$g(d)(\mathcal{B} + 1/\varepsilon)^{o(d/\log d)}.$$

Both statements hold for integral generators with at most nine nonzero coordinates.

Proof sketch. Choose $\varepsilon = \delta/(16T)$. In a yes-instance every valid estimate satisfies $\widehat{R}^2 > T - \delta/2$, while in a no-instance it satisfies $\widehat{R}^2 < T - \delta/2$. Both T and $1/\delta$ have polynomial magnitude, so the required $1/\varepsilon$ is polynomial in the source size. The two claimed algorithms would therefore violate W[1]-hardness or the source ETH lower bound. Appendix A.4 gives the inequalities and the integral scaling argument. □

Theorem 8 is stated for deterministic algorithms. Its randomized analogue follows under the corresponding randomized parameterized assumption, or under randomized ETH for the quantitative claim. We keep the standard deterministic assumptions separate from the known randomized upper bounds.

5 Consequences for convex networks and low-rank QUBO

Proposition 3 transfers Theorem 7 without changing the dimension or adding any negative output weight.

Corollary 9 (Input-convex networks)**.** *Given rational $a_1, \dots, a_n \in \mathbb{Q}^d$, define the bias-free one-hidden-layer network*

$$f(x) = \sum_{r=1}^{n} \mathrm{ReLU}(a_r^\top x).$$

Deciding whether $\mathrm{Lip}_2(f) \geq L$ is NP-complete and W[1]-hard parameterized by d. The same statements hold for the Lipschitz constant restricted to the unit ball. The exact and dimension-accuracy lower bounds in Theorems 7 and 8 also hold. Every a_r may be integral with at most nine nonzero entries and every output weight is one, so f is nonnegative, positively homogeneous, and convex.

Positive homogeneity proves the unit-ball statement: the pair $0, u$ in Proposition 3 can be rescaled to $0, \rho u$ without changing its difference quotient.

This is a restricted case of the ICNN architecture of Amos et al. (2017). The hardness is not inherited from the known mixed-sign result for general two-layer networks (Froese et al., 2026). Mixed signs represent a difference of support functions and enable zonotope noncontainment. Corollary 9 instead uses one support function and resolves the restriction that remained open in that work.

Equation (4) gives the optimization consequence.

Corollary 10 (Low-rank convex binary quadratic maximization)**.** *Given an integral factor A and an integral threshold B, deciding whether*

$$\max_{s \in \{0,1\}^n} s^\top A^\top A s \geq B$$

is NP-complete and W[1]-hard parameterized by $r = \mathrm{rank}(A)$. Under ETH there is no $g(r)\mathcal{B}^{o(r/\log r)}$-time algorithm, where $\mathcal{B}$ is the factored input's encoding length. The dimension-accuracy lower bound in Theorem 8 also holds for multiplicative approximation of the optimum. The same conclusions hold in the explicit-matrix convention because $Q = A^\top A$ can be formed in polynomial time and $\mathrm{rank}(Q) = r$.

After the denominator-clearing step in Theorem 7, use the integer threshold $B = D^2 T$. Yes-instances attain B, while no-instances remain below it by the scaled gap $D^2\delta$. The construction may have $\mathrm{rank}(Q) < d$. This only strengthens the parameterized reduction because $\mathrm{rank}(Q) \leq 11k + 1$. No full-row-rank promise is needed.

The approximation frontier. A standard spherical net gives a deterministic feasible $(1+\varepsilon)$-approximation in $(1/\varepsilon)^{O(d)}\,\mathrm{poly}(n, \mathcal{B})$ time. Randomized schemes have comparable dependence (Knauer et al., 2015; Shenmaier, 2018; Froese et al., 2026). Theorem 8 rules out $g(d)\,\mathrm{poly}(\mathcal{B}, 1/\varepsilon)$ time and, under ETH, forces exponent $\Omega(d/\log d)$ on $\mathcal{B} + 1/\varepsilon$. Thus the upper and lower exponents match up to a logarithmic factor, while constant-factor approximation remains possible.

Why convexity does not help here. Input convexity makes $\min_x f(x)$ tractable, but Lipschitz certification maximizes a subgradient norm. For Equation (1), the subdifferential polytope is exactly $Z(A)$. The result therefore preserves the optimization benefit of ICNNs while separating worst-case sensitivity.

6 Related work

Neural Lipschitz certification. Exact Lipschitz computation is NP-hard for general two-layer ReLU networks, with exact formulations developed subsequently (Virmaux & Scaman, 2018; Jordan & Dimakis, 2020; Bhowmick et al., 2021; Couëllan et al., 2026; Splittgerber, 2026). Concurrent work studies reachability barriers and failures of semidefinite Lipschitz bounds for broader neural-network families (Kuang et al., 2026). Froese et al. (2025) asked whether low dimension makes the associated zonotope problems FPT. Froese et al. (2026) proved mixed-sign hardness but left positive-output networks open and gave randomized FPT approximation. We close the Euclidean open case and limit variable-accuracy dependence for its convex restriction. ICNN representation lower bounds address a different question (Hertrich & Loho, 2024).

Norm maximization and vector sums. Euclidean zonotope radius is Longest Vector Sum and fixed-rank PSD QUBO. NP-hardness and $O(n^{d-1})$ exact algorithms were known (Bodlaender et al., 1990; Ferrez et al., 2005; Shenmaier, 2020), as was $(1 + 2/\varepsilon)^d \operatorname{poly}(d, n)$-time randomized approximation (Shenmaier, 2018; Froese et al., 2026). A concurrent preprint unifies low-rank PSD optimization through projected-shadow geometry and recovers classical binary zonotope enumeration, without parameterized lower bounds (Soltanalian & Mousavi, 2026). For halfspace-presented polytopes, Knauer et al. (2015) ruled out $g(d) \operatorname{poly}(n, 1/\varepsilon)$ approximation. Their reduction does not apply to generator-presented zonotopes. Our lifted selector establishes the barrier for this compact representation, while broader recent containment approximations do not resolve it (Eisenbrand et al., 2026).

7 Conclusion

Exact global Euclidean Lipschitz certification remains parameterized intractable for the simplest bias-free ICNNs. The inverse-polynomial selector gap gives a sharper boundary: fixed accuracy is tractable in low dimension, but no deterministic scheme can separate dimension from polynomial accuracy dependence. Extensions to other smooth norms, removal of the logarithmic ETH gap, and randomized parameterized lower bounds remain open.

Reproducibility statement

The appendix gives full proofs and encoding details.

References

Brandon Amos, Lei Xu, and J. Zico Kolter. Input convex neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 146–155. PMLR, 2017.

Aritra Bhowmick, Meenakshi D'Souza, and G. Srinivasa Raghavan. LipBaB: Computing exact lipschitz constant of ReLU networks. In *Artificial Neural Networks and Machine Learning–ICANN 2021*, volume 12894 of *Lecture Notes in Computer Science*, pp. 151–162. Springer, 2021. doi: 10.1007/978-3-030-86380-7_13.

Hans L. Bodlaender, Peter Gritzmann, Victor Klee, and Jan van Leeuwen. Computational complexity of norm-maximization. *Combinatorica*, 10(2):203–225, 1990. doi: 10.1007/BF02123011.

Nicolas Couëllan, Sophie Jan, and Mohammed Sbihi. Mixed integer quadratically constrained quadratic programming for neural network lipschitz constant computation. *European Journal of Operational Research*, 331:916–924, 2026. doi: 10.1016/j.ejor.2025.12.019.

Friedrich Eisenbrand, Thomas Rothvoss, Matteo Russo, and Ruben Skorupinski. Nearly-tight bounds for zonotope containment and beyond. *arXiv preprint arXiv:2605.04183*, 2026. URL https://arxiv.org/abs/2605.04183.

David Eppstein and Daniel Lokshtanov. The parameterized complexity of finding point sets with hereditary properties. In *13th International Symposium on Parameterized and Exact Computation (IPEC 2018)*, volume 115 of *Leibniz International Proceedings in Informatics*, pp. 11:1–11:14. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2019. doi: 10.4230/LIPIcs.IPEC.2018.11.

Jean-Antoine Ferrez, Komei Fukuda, and Thomas M. Liebling. Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope construction algorithm. *European Journal of Operational Research*, 166(1):35–50, 2005. doi: 10.1016/j.ejor.2003.04.011.

Vincent Froese, Moritz Grillo, Christoph Hertrich, and Martin Skutella. Open problem: Fixed-parameter tractability of zonotope problems. In *Proceedings of the 38th Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 6210–6214. PMLR, 2025.

Vincent Froese, Moritz Leo Grillo, Christoph Hertrich, and Moritz Stargalla. Parameterized hardness of zonotope containment and neural network verification. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=y8N45EEW05.

Christoph Hertrich and Georg Loho. Neural networks and (virtual) extended formulations. *arXiv preprint arXiv:2411.03006*, 2024. URL https://arxiv.org/abs/2411.03006.

Matt Jordan and Alexandros G. Dimakis. Exactly computing the local lipschitz constant of ReLU networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7344–7353, 2020.

Christian Knauer, Stefan König, and Daniel Werner. Fixed-parameter complexity and approximability of norm maximization. *Discrete & Computational Geometry*, 53(2):276–295, 2015. doi: 10.1007/s00454-015-9667-0.

Simon Kuang, Yuezhu Xu, S. Sivaranjani, and Xinfan Lin. Demystifying lipschitz verification: Positive matrices, negative results. *arXiv preprint arXiv:2603.28113*, 2026. URL https://arxiv.org/abs/2603.28113.

Dániel Marx. Can you beat treewidth? *Theory of Computing*, 6:85–112, 2010. doi: 10.4086/toc.2010.v006a005.

Vladimir V. Shenmaier. Complexity and approximation of finding the longest vector sum. *Computational Mathematics and Mathematical Physics*, 58(6):850–857, 2018. doi: 10.1134/S0965542518060131.

Vladimir V. Shenmaier. Complexity and algorithms for finding a subset of vectors with the longest sum. *Theoretical Computer Science*, 818:60–73, 2020. doi: 10.1016/j.tcs.2018.04.018.

Mojtaba Soltanalian and Ahmad Mousavi. The rank-collapse principle for quadratic optimization. *arXiv preprint arXiv:2608.07828*, 2026. URL https://arxiv.org/abs/2608.07828.

Tom A. Splittgerber. ExLipBaB: Exact lipschitz constant computation for piecewise linear neural networks. *arXiv preprint arXiv:2602.15499*, 2026. URL https://arxiv.org/abs/2602.15499.

Aladin Virmaux and Kevin Scaman. Lipschitz regularity of deep neural networks: Analysis and efficient estimation. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

A Additional proof details

A.1 The planar zonogon identity

For completeness, consider a centrally symmetric polygon whose circularly ordered vertices are

$$v_1, \dots, v_m, -v_1, \dots, -v_m.$$

The first half of its boundary edge vectors is

$$e_j = v_{j+1} - v_j \quad (1 \le j < m), \qquad e_m = -v_1 - v_m.$$

The second half is $-e_1, \dots, -e_m$. The Minkowski sum $v_1 + \sum_{j=1}^{m} [0, e_j]$ has exactly these edge directions and lengths in circular order. Starting from v_1, its consecutive vertices include

$$v_1, \quad v_1 + e_1 = v_2, \quad \dots, \quad v_1 + \sum_{j=1}^{m-1} e_j = v_m, \quad v_1 + \sum_{j=1}^{m} e_j = -v_1.$$

Central symmetry supplies the remaining half. Therefore the Minkowski sum equals the polygon. The same telescoping identities hold before projection, which attaches the prescribed payload to each maximum-radius vertex.

A.2 Quantitative selector gap

We prove the angular estimates used in Lemma 4. Normalize $M = 1$ and suppress payload coordinates. Write $u(j) = (\cos\theta_j, \sin\theta_j)$, where $\theta_j = 2\arctan j$. For $j < m$, the tangent subtraction identity gives

$$\alpha_j := \theta_{j+1} - \theta_j = 2\arctan\left(\frac{1}{1 + j(j+1)}\right) \ge \frac{1}{1 + j(j+1)} \ge \frac{1}{m^2 + 1}.$$

Here we used $\arctan t \ge t/2$ for $t \in [0, 1]$. Put $\theta_{m+1} = \theta_1 + \pi$. The closing gap satisfies $\alpha_m = \theta_1 + \pi - \theta_m > \pi/2$. Thus every α_j is at least $\Delta = (m^2 + 1)^{-1}$.

Let $e_j = u(j+1) - u(j)$, including $u(m+1) = -u(1)$. Its length obeys

$$\|e_j\|_2 = 2\sin(\alpha_j/2) \ge \frac{2\alpha_j}{\pi} \ge \frac{2\Delta}{\pi}. \tag{14}$$

The unoriented line containing e_j has projective angle

$$\lambda_j = \frac{\theta_j + \theta_{j+1}}{2} + \frac{\pi}{2} \pmod{\pi}.$$

After extending $\theta_{m+r} = \theta_r + \pi$, the cyclic separation of consecutive generator lines is

$$\lambda_{j+1} - \lambda_j = \frac{\alpha_j + \alpha_{j+1}}{2} \ge \Delta. \tag{15}$$

In particular, relative to any fixed projective line, at most one generator line can have distance less than $\Delta/2$.

Let $x = u(1) + \sum_j \varepsilon_j e_j$ be a noncanonical endpoint. If $x = 0$, the claim is immediate. Otherwise put $w = x/\|x\|_2$. Choose a bit vector η maximizing the linear functional $w^\top z$ over the generating segments. A generator orthogonal to w, if one exists, is unique, and we choose its tied bit to match ε_j. The resulting endpoint y is a polygon vertex. For $J = \{j : \varepsilon_j \ne \eta_j\}$, separability over the segments gives

$$h_P(w) - w^\top x = \sum_{j \in J} |w^\top e_j|. \tag{16}$$

If $|J| \ge 2$, Equation (15) implies that one mismatched generator line has distance at least $\Delta/2$ from $w^\perp$. Together with Equation (14),

$$h_P(w) - w^\top x \ge \frac{2\Delta}{\pi}\sin(\Delta/2) \ge \frac{2\Delta^2}{\pi^2}.$$

It remains to treat $|J| = 1$, with unique mismatch j. The two generator lines incident to y are consecutive in projective order and bound the normal cone of y after rotation by $\pi/2$. If e_j were incident to y, toggling it would move from y to the neighboring polygon vertex, making x canonical. Hence its line is nonincident and is separated from the rotated normal cone by at least one cyclic gap. Its distance from $w^\perp$ is therefore at least Δ, so

$$h_P(w) - w^\top x \geq \frac{2\Delta}{\pi} \sin \Delta \geq \frac{4\Delta^2}{\pi^2}.$$

Finally, P lies in the unit disk, so $h_P(w) \leq 1$ and Equation (16) gives

$$1 - \|x\|_2 \geq h_P(w) - w^\top x \geq \frac{2\Delta^2}{\pi^2}.$$

Thus $1 - \|x\|_2^2 \geq 2\Delta^2/\pi^2 > \Delta^2/6$, as claimed.

A.3 Consistency-code gap

For identity codes $p_r = u(r)$, $r \in [n]$, the preceding angular calculation gives $|\theta_r - \theta_s| \geq \Delta_n = (n^2+1)^{-1}$ whenever $r \neq s$. If two signs agree but the identities differ, then

$$4 - \|p_r + p_s\|_2^2 = 4 \sin^2\left(\frac{|\theta_r - \theta_s|}{2}\right) \geq \frac{4\Delta_n^2}{\pi^2} > \frac{1}{3(n^2+1)^2}.$$

If the signs disagree, all code angles lie in an interval of width less than $\pi/2$, so $\|p_r - p_s\|_2^2 \leq 2$ and the deficit from four is at least two. This proves the consistency deficit used in Equation (11).

A.4 Complexity and approximation proofs

We first complete Theorem 7. Membership in NP follows because a binary endpoint $s \in \{0,1\}^n$ certifies a yes-instance and its squared norm can be compared with L^2 in polynomial time. The construction has $N + F$ generators and dimension $d = 11k + 1$. Equations (5)–(12) use rational numbers with polynomial bit length and are computable in polynomial time. Define the common denominator

$$D = \prod_{r=1}^{N} (1 + r^2).$$

Multiplying every generator by D clears all coordinate denominators, so the generators may be required to be integral. We also multiply the rational radius threshold by D, which preserves the decision problem without requiring that threshold to become integral. The support accounting in Appendix A.5 gives at most nine nonzeros per generator. This is a parameterized reduction from Cubic Partitioned Subgraph Isomorphism. An algorithm with running time $g(d)\mathcal{B}^{o(d/\log d)}$ would solve the source problem in $\rho(k)n^{o(k/\log k)}$ time because $d = 11k+1$ and the output encoding length $\mathcal{B}$ is polynomial in the source size n, contradicting ETH.

For Theorem 8, choose $\varepsilon = \delta/(16T)$. The quantities T and $1/\delta$ have polynomial magnitude in the source instance, so $1/\varepsilon$ is polynomial in its size. In a yes-instance, Equation (13) gives

$$\widehat{R}^2 \geq \frac{T}{(1+\varepsilon)^2} \geq T(1 - 2\varepsilon) > T - \frac{\delta}{2}.$$

In a no-instance, using $(1+\varepsilon)^2 \leq 1 + 3\varepsilon$,

$$\widehat{R}^2 \leq (1+\varepsilon)^2(T - \delta) \leq T - \delta + 3\varepsilon T < T - \frac{\delta}{2}.$$

Comparing $\widehat{R}^2$ with the rational value $T - \delta/2$ decides the source instance. Multiplication by D scales both squared radii, their additive gap, and the comparison threshold by D^2, while leaving ε unchanged. Thus a $g(d)\,\mathrm{poly}(\mathcal{B}, 1/\varepsilon)$-time scheme gives an FPT algorithm for the W[1]-hard source problem. The ETH claim follows because $d = 11k + 1$ and both $\mathcal{B}$ and $1/\varepsilon$ are polynomial in the source size.

For multiplicative approximation of the squared QUBO objective, take instead $\varepsilon_Q = \delta/(4T)$. A yes-instance estimate exceeds $T - \delta/2$, while a no-instance estimate is below $T - \delta/2$, so the same parameterized lower bounds follow.

For completeness, the matching elementary upper bound follows from a spherical net. Let U be a rational Euclidean ρ-net of $\mathbb{S}^{d-1}$, where $\rho^2 \leq 2\varepsilon/(1+\varepsilon)$. Such a net is constructible with $|U| = (1/\varepsilon)^{O(d)}$ and polynomial coordinate bit length. For each $u \in U$, maximize $u^\top z$ over the zonotope by selecting generator a_i exactly when $u^\top a_i > 0$, and return the resulting endpoint of largest norm. The case $R = 0$ is immediate. Otherwise, if $z_\star$ has norm R, choose $u \in U$ within ρ of $z_\star/R$. Then

$$\|z_u\|_2 \geq u^\top z_u \geq u^\top z_\star = R\left(1 - \frac{1}{2}\left\|u - \frac{z_\star}{R}\right\|_2^2\right) \geq \frac{R}{1+\varepsilon}.$$

This proves the deterministic $(1/\varepsilon)^{O(d)}\operatorname{poly}(n, \mathcal{B})$ upper bound quoted in the main text.

A.5 Endpoint, anchor, and sparsity bounds

The image of $[0,1]^n$ under a linear map is the convex hull of the images of its 2^n vertices. A convex function on a polytope attains a maximum at a polytope vertex, and every polytope vertex is the image of at least one cube vertex. It is therefore enough to bound endpoint sums of the generators.

Within a factor containing m_f candidates, each payload coordinate of $c_{f,1}$ has magnitude at most one and each generator coordinate has magnitude at most two. A contribution that selects its factor anchor has magnitude at most $1 + 2m_f$ in any payload coordinate. Summing over factors gives at most

$$\sum_f (1 + 2m_f) = F + 2N \leq F(2N+1).$$

A contribution that omits some factor anchors is no larger under the same bound. There are D_p payload coordinates. This proves the uniform squared bound $P = D_p F^2 (2N+1)^2$.

If a factor anchor is omitted, that factor's selector contribution is a point in the translated polygon $P_f - c_{f,1}$ and has norm at most $2M$. If it is selected, the selector contribution lies in P_f and has norm at most M. Hence an endpoint sum selecting exactly $q < F$ anchors has squared norm at most

$$q^2 H^2 + \big(q + 4(F-q)\big)M^2 + P.$$

Selecting every anchor and any canonical endpoint in every factor gives squared norm at least $F^2H^2 + FM^2$. The latter minus the former is

$$(F-q)\big((F+q)H^2 - 3M^2\big) - P \geq 13M^2 - P > 1,$$

where the inequalities follow from $H = 4FM + P + 1$, $M \geq P + 1$, and $F - q \geq 1$. This proves the missing-anchor comparison used in Lemma 5.

Finally, a vertex-factor candidate is supported on its two selector coordinates and three two-dimensional incidence blocks. An edge-factor candidate uses its selector block and two incidence blocks. Selector differences do not enlarge these block supports. A factor anchor uses the support of its first candidate plus the shared anchor coordinate. Thus vertex-factor anchors have at most nine nonzeros, vertex-factor selector generators have at most eight, and edge-factor generators have at most seven.

A.6 Polynomial bit complexity

Every coordinate of $u(r)$ has denominator dividing $1 + r^2$. The product $D = \prod_{r=1}^{N}(1+r^2)$ has bit length $O(N \log N)$ and clears every selector and identity-code denominator. The integers P, M, H, and T, as well as $1/\delta$, have polynomial magnitude in N and the pattern size. Candidate vectors and their differences are computable by rational arithmetic in polynomial time. Since a host edge is relevant to at most one pattern edge under the partition, $N \leq |V| + |E_G|$.

The threshold precision is also polynomial. One may take

$$b = 2 + \lceil \log_2(T+1) \rceil + \left\lceil \log_2\big(3(|V|^2+1)^2\big) \right\rceil.$$

Then $2^{-b} < \delta/(2(T+1))$. The numerator of L in Equation (12) is the integer square root of $2^{2b}T$, computable in time polynomial in its bit length.

A.7 Full-rank and zero-generator conventions

The reduction does not require its generators to span every displayed coordinate. Parameterized hardness uses the ambient input dimension as an upper bound, so a lower actual rank only decreases the parameter. If a problem definition forbids zero generators, none are introduced by the construction: selector projections of all lifted-selector generators are nonzero, and every anchor has a nonzero fresh coordinate.

B Further discussion

Extending the result to every fixed ℓ_p norm with $1 < p < \infty$ requires selector codes of equal norm with controlled arithmetic and a polynomial radial gap. The Euclidean circle supplies an especially clean rational family. The ETH lower bounds and the exact or approximate upper bounds also differ by a logarithmic factor in their dimension-dependent exponents. Removing it would require a stronger source lower bound or a different hypothesis. Finally, the best known fine-accuracy schemes are randomized. Deterministic FPT approximation and unconditional randomized parameterized lower bounds remain separate questions.

The lifted-selector principle may be useful beyond zonotopes. It converts an inscribed polygon into an exclusive-choice gadget while transporting arbitrary labels through the edge generators. More broadly, it shows that a compact Minkowski-sum representation can hide global consistency constraints even when the objective and represented function are convex.