

Graded Neural Networks

Tony Shaska

SHASKA@OAKLAND.EDU

Department of Mathematics and Statistics, Oakland University, Rochester, MI, 48309.

Abstract

This paper presents a novel framework for graded neural networks (GNNs) built over graded vector spaces $\mathcal{V}_{\mathbf{q}}^n$, extending classical neural architectures by incorporating algebraic grading. Leveraging a coordinate-wise grading structure with scalar action $\lambda \star \mathbf{x} = (\lambda^{q_i} x_i)$, defined by a tuple $\mathbf{q} = (q_0, \dots, q_{n-1})$, we introduce graded neurons, layers, activation functions, and loss functions that adapt to feature significance. Theoretical properties of graded spaces are established, followed by a comprehensive GNN design, addressing computational challenges like numerical stability and gradient scaling. Potential applications span machine learning and photonic systems, exemplified by high-speed laser-based implementations. This work offers a foundational step toward graded computation, unifying mathematical rigor with practical potential, with avenues for future empirical and hardware exploration.

Keywords: Graded Neural Networks (GNN), Graded Vector Spaces

1. Introduction

Artificial neural networks are pivotal in artificial intelligence, tackling diverse problems from applied mathematics to pattern recognition. Typically, these networks model functions $f : k^n \rightarrow k^m$ (often $k = \mathbb{R}$), with coordinates of $\mathbf{v} \in k^n$ as *input features* and $f(\mathbf{v})$ as *output features*. While standard architectures treat features uniformly, many scenarios—such as document analysis or algebraic geometry—reveal inputs with varying significance, suggesting a graded approach where each feature x_i in $\mathbf{v} = (x_0, \dots, x_{n-1})$ carries a grade $\text{gr}(x_i) \in I$. Such structures align with graded vector spaces, where coordinates are assigned values from a set I , a concept we explore to extend neural computation.

Our motivation stems from the moduli space of genus two curves, isomorphic to a subspace of the weighted projective space $\mathbb{W}\mathbb{P}_{(2,4,6,10)}$ with grades $\mathbf{q} = (2, 4, 6, 10)$. In [Shaska and Shaska \(2024\)](#), neural networks predicting automorphism groups or (n, n) -split Jacobians of these curves achieved 40% accuracy with raw coefficients, but soared to 99% using graded invariants as inputs. This leap, while expected mathematically (as $\mathbb{W}\mathbb{P}_{(2,4,6,10)}$ captures isomorphism classes), prompts a deeper question: does grading inherently enhance performance? Parallel insights from [Beshaj et al. \(2020\)](#); [Salami and Shaska \(2023b, 2024\)](#) show weighted heights in graded projective spaces outpace standard height computations, hinting that graded norms and operations may simplify complex tasks—a hypothesis we test by designing neural networks over graded spaces.

Graded vector spaces, detailed in Section 2, generalize \mathbb{R}^n by assigning grades via $\text{gr}(x_i) = q_i$, with scalar action $\lambda \star \mathbf{x} = (\lambda^{q_i} x_i)$. In Section 3, we construct graded neural networks (GNNs), adapting neurons ($\sum w_i^{q_i} x_i + b$), activations ($\text{ReLU}_i(x_i) = \max\{0, |x_i|^{1/q_i}\}$), and losses (e.g., $\sum q_i |y_i - \hat{y}_i|^2$) to this structure. Unlike classical models, GNNs naturally handle graded inputs, reverting to standard networks when $q_i = 1$. Section 4 examines theoretical challenges—numerical stability, computational scaling—and potential applications, from machine learning to photonic systems like laser graded neurons [Nie et al. \(2024\)](#). We conclude in Section 5 with the framework’s implications and future directions.

This work investigates whether GNNs over graded vector spaces offer mathematical and practical advantages, building a foundation for graded computation in artificial intelligence.

2. Graded Vector Spaces

Here we give the bare minimum background on graded vector spaces. The interested reader can check details at [Bourbaki \(1974\)](#), [Roman \(2008\)](#), [Koszul \(1983\)](#), among other places. We use "grades" to denote the indices of grading (e.g., q_i), distinguishing them from "weights" used for neural network coefficients in Section 3.

A graded vector space is a vector space with an extra grading structure, typically a decomposition into a direct sum of subspaces indexed by integers. While we present the traditional decomposition $V = \bigoplus_{n \in \mathbb{N}} V_n$ and the coordinate-wise form $\mathcal{V}_{\mathbf{q}}^n(k) = k^n$ with scalar action $\lambda \star \mathbf{x} = (\lambda^{q_i} x_i)$, these definitions are equivalent, as the latter can represent the former via a basis choice, a perspective we adopt for neural networks in Section 3. For this paper, we focus on graded vector spaces indexed by integers, though we also define them for a general index set I below.

2.1. Integer Gradation

Let \mathbb{N} be the set of non-negative integers. An \mathbb{N} -graded vector space, often simply a **graded vector space** without the prefix \mathbb{N} , is a vector space V with a decomposition:

$$V = \bigoplus_{n \in \mathbb{N}} V_n,$$

where each V_n is a vector space. Elements of V_n are called homogeneous elements of degree n .

Graded vector spaces are common. For example, the set of all polynomials in one or several variables forms a graded vector space, where homogeneous elements of degree n are linear combinations of monomials of degree n .

Example 1 *Let k be a field and consider $\mathcal{V}_{(2,3)}$, the space of homogeneous polynomials of degrees 2 and 3 in $k[x, y]$. It decomposes as $\mathcal{V}_{(2,3)} = V_2 \oplus V_3$, where V_2 is the space of binary quadratics and V_3 the space of binary cubics. For $\mathbf{u} = [f, g] \in V_2 \oplus V_3$, scalar multiplication is:*

$$\lambda \star \mathbf{u} = \lambda \star [f, g] = [\lambda^2 f, \lambda^3 g],$$

reflecting grades 2 and 3. We use this example repeatedly throughout the paper.

The subspaces of a graded vector space need not be indexed by natural numbers and may use any set I . An I -graded vector space V is a vector space with a decomposition:

$$V = \bigoplus_{i \in I} V_i,$$

where $i \in I$ are the grades. The case $I = \mathbb{Z}/2\mathbb{Z}$ (elements 0 and 1) is notable in physics, termed a **supervector space**.

2.2. Graded Linear Maps

For an index set I , a linear map $f : V \rightarrow W$ between I -graded vector spaces is a **graded linear map** if it preserves the grading, $f(V_i) \subseteq W_i$, for all $i \in I$. Such maps are also called **homomorphisms (or morphisms) of graded vector spaces** or homogeneous linear maps. For a commutative monoid I (i.e., \mathbb{N}), maps homogeneous of degree $i \in I$ satisfy:

$$f(V_j) \subseteq W_{i+j}, \quad \text{for all } j \in I,$$

where $+$ is the monoid operation. If I embeds into an abelian group A (i.e., \mathbb{Z} for \mathbb{N}), maps of degree $i \in A$ follow the same property, with $+$ as the group operation. A map of degree $-i$ satisfies:

$$f(V_{i+j}) \subseteq W_j, \quad f(V_j) = 0 \text{ if } j - i \notin I.$$

Example 2 For $\mathcal{V}_{(2,3)} = V_2 \oplus V_3$, a linear map $L : \mathcal{V}_{(2,3)} \rightarrow \mathcal{V}_{(2,3)}$ satisfies:

$$\begin{aligned} L([\lambda \star \mathbf{u}]) &= L([\lambda^2 f, \lambda^3 g]) = [\lambda^2 L(f), \lambda^3 L(g)] = \lambda \star [L(f), L(g)] = \lambda \star L(\mathbf{u}), \\ L([f, g] \oplus [f', g']) &= L([f + f', g + g']) = [L(f) + L(f'), L(g) + L(g')] \\ &= [L(f), L(g)] \oplus [L(f'), L(g')] = L([f, g]) \oplus L([f', g']). \end{aligned}$$

Using the basis $\mathcal{B} = \{x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}$ (where $\mathcal{B}_1 = \{x^2, xy, y^2\}$ spans V_2 , $\mathcal{B}_2 = \{x^3, x^2y, xy^2, y^3\}$ spans V_3), the polynomial $F(x, y) = (x^2 + xy + y^2) + (x^3 + x^2y + xy^2 + y^3)$ has coordinates $\mathbf{u} = [1, 1, 1, 1, 1, 1, 1]^t$.

Further details are in [Bourbaki \(1974\)](#), [Balaba \(2005\)](#), [Bondarenko \(2003\)](#).

2.3. Operations over Graded Vector Spaces

Having established the structure of graded vector spaces in Section 2, we now define operations that extend their utility: the direct sum, tensor product, and dual space. These operations, rooted in the grading, are relevant to potential applications in graded neural networks, such as feature composition or optimization.

For I -graded spaces $V = \bigoplus_{i \in I} V_i$ and $W = \bigoplus_{i \in I} W_i$, the **direct sum** is $V \oplus W$ with gradation:

$$(V \oplus W)_i = V_i \oplus W_i.$$

Scalar multiplication is $\lambda(v_i, w_i) = (\lambda v_i, \lambda w_i)$. For differing grade sets I and J , index over $I \cup J$, with $(V \oplus W)_k = V_k \oplus W_k$ ($V_k = 0$ if $k \notin I$).

Consider two graded vector spaces $V = \bigoplus_{i \in I} V_i$ and $W = \bigoplus_{i \in I} W_i$, where I is a semigroup (e.g., \mathbb{N} with addition). The **tensor product** $V \otimes W$ is a graded vector space with components:

$$(V \otimes W)_i = \bigoplus_{(j,k):j+k=i} (V_j \otimes W_k),$$

where $v_j \in V_j$ and $w_k \in W_k$ form $v_j \otimes w_k$ of grade $j + k$, and scalar multiplication is given by $\lambda(v_j \otimes w_k) = (\lambda v_j) \otimes w_k$.

For example, take $V = \mathcal{V}_{(2,3)} = V_2 \oplus V_3$, with V_2 and V_3 as spaces of quadratic and cubic polynomials, respectively. The tensor product $V \otimes V$ is:

$$\mathcal{V}_{(2,3)} \otimes \mathcal{V}_{(2,3)} = (V_2 \otimes V_2)_4 \oplus (V_2 \otimes V_3)_5 \oplus (V_3 \otimes V_2)_5 \oplus (V_3 \otimes V_3)_6.$$

If $f = x^2 \in V_2$ (grade 2) and $g = x^3 \in V_3$ (grade 3), then $f \otimes g \in (V_2 \otimes V_3)_5$, since $2 + 3 = 5$.

For the coordinate-wise space $\mathcal{V}_{\mathbf{q}}^n(k) = k^n$ with $\text{gr}(x_i) = q_i$, the tensor product with $\mathcal{V}_{\mathbf{q}'}^m(k) = k^m$ (grades $\text{gr}(x'_j) = q'_j$) is:

$$\mathcal{V}_{\mathbf{q}}^n \otimes \mathcal{V}_{\mathbf{q}'}^m = \bigoplus_{i=0}^{n-1} \bigoplus_{j=0}^{m-1} k(e_i \otimes e'_j),$$

where $e_i \otimes e'_j$ has grade $q_i + q'_j$. This form accommodates varying grades across spaces, relevant to inputs of differing significance.

For three graded vector spaces U , V , and W over a semigroup I , the tensor product is **associative**: $(U \otimes V) \otimes W \cong U \otimes (V \otimes W)$. The graded components of $(U \otimes V) \otimes W$ are:

$$((U \otimes V) \otimes W)_i = \bigoplus_{(j,k,l):j+k+l=i} (U_j \otimes V_k) \otimes W_l,$$

where j , k , and l are grades in U , V , and W , respectively. This property ensures consistency in composing multiple tensor operations, analogous to stacking transformations in neural network layers.

Next, consider the dual space of $V = \bigoplus_{i \in I} V_i$, where I is a general index set (e.g., \mathbb{N} , \mathbb{Z}), not necessarily a semigroup. The **dual** $V^* = \text{Hom}_k(V, k)$ is graded as:

$$V^* = \bigoplus_{i \in I} V_{-i}^*,$$

with $V_{-i}^* = \{f : V \rightarrow k \mid f(V_i) \subseteq k, f(V_j) = 0 \text{ if } j \neq i\}$. The grade $-i$ arises because a functional on V_i (grade i) pairs to produce a scalar (grade 0), requiring $i + (-i) = 0$.

For $\mathcal{V}_{\mathbf{q}}^n(k) = k^n$ with $\text{gr}(x_i) = q_i$ and scalar action $\lambda \star \mathbf{x} = (\lambda^{q_i} x_i)$, the dual $(\mathcal{V}_{\mathbf{q}}^n)^* = k^n$ has basis functionals f_i of grade $\text{gr}(f_i) = -q_i$, with $\lambda \star f_i = \lambda^{-q_i} f_i$. This inverse scaling complements the original action, suggesting applications in defining graded loss functions or optimization procedures for neural networks.

2.4. Inner Graded Vector Spaces

Consider now the case when each V_i is a finite-dimensional inner space, and let $\langle \cdot, \cdot \rangle_i$ denote the corresponding inner product. Then we can define an inner product on $V = \bigoplus_{i \in I} V_i$ as follows. For $\mathbf{u} = u_1 + \dots + u_n$ and $\mathbf{v} = v_1 + \dots + v_n$, where $u_i, v_i \in V_i$, we define:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle u_1, v_1 \rangle_1 + \dots + \langle u_n, v_n \rangle_n,$$

which is the standard product across graded components. The Euclidean norm is then:

$$\|\mathbf{u}\| = \sqrt{u_1^2 + \dots + u_n^2},$$

where $\|u_i\|_i = \sqrt{\langle u_i, u_i \rangle_i}$ is the norm in V_i , and we assume an orthonormal basis for simplicity.

If such V_i are not necessarily finite-dimensional, then we have to assume that V_i is a Hilbert space (i.e., a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product). This case of Hilbert spaces is especially

important in machine learning and artificial intelligence due to their role in functional analysis and optimization.

Obviously, having a norm on a graded vector space is crucial for machine learning if we want to define a cost function of some type. The simpler case of Euclidean vector spaces and their norms was considered in [Moskowitz \(2014\)](#); [Sriwongsa and Wiboonton \(2020\)](#). However, graded structures allow for norms that reflect the grading, enhancing their utility in applications like neural networks over $\mathcal{V}_{\mathbf{q}}^n(k)$.

Example 3 *Let us continue with the space $\mathcal{V}_{(2,3)}$ from Example 1, with bases $\mathcal{B}_1 = \{x^2, xy, y^2\}$ for V_2 and $\mathcal{B}_2 = \{x^3, x^2y, xy^2, y^3\}$ for V_3 , as in Example 2. Hence, a basis for $\mathcal{V}_{(2,3)} = V_2 \oplus V_3$ is $\mathcal{B} = \{x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}$. Let $\mathbf{u}, \mathbf{v} \in \mathcal{V}_{(2,3)}$ be given by:*

$$\begin{aligned}\mathbf{u} &= \mathbf{a} + \mathbf{b} = (u_1x^2 + u_2xy + u_3y^2) + (u_4x^3 + u_5x^2y + u_6xy^2 + u_7y^3), \\ \mathbf{v} &= \mathbf{a}' + \mathbf{b}' = (v_1x^2 + v_2xy + v_3y^2) + (v_4x^3 + v_5x^2y + v_6xy^2 + v_7y^3).\end{aligned}$$

Then:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{a} + \mathbf{b}, \mathbf{a}' + \mathbf{b}' \rangle = \langle \mathbf{a}, \mathbf{a}' \rangle_2 + \langle \mathbf{b}, \mathbf{b}' \rangle_3 = u_1v_1 + u_2v_2 + u_3v_3 + u_4v_4 + u_5v_5 + u_6v_6 + u_7v_7,$$

and the Euclidean norm is $\|\mathbf{u}\| = \sqrt{u_1^2 + \dots + u_7^2}$, assuming \mathcal{B} is orthonormal. This treats all grades uniformly, which may not fully leverage the graded structure.

There are other ways to define a norm on graded spaces, particularly to emphasize the grading. Consider a Lie algebra \mathfrak{g} called **graded** if there is a finite family of subspaces V_1, \dots, V_r such that $\mathfrak{g} = V_1 \oplus \dots \oplus V_r$ and $[V_i, V_j] \subset V_{i+j}$, where $[V_i, V_j]$ is the Lie bracket. When \mathfrak{g} is graded, define a dilation for $t \in \mathbb{R}^\times$, $\alpha_t : \mathfrak{g} \rightarrow \mathfrak{g}$, by:

$$\alpha_t(v_1, \dots, v_r) = (tv_1, t^2v_2, \dots, t^r v_r).$$

We define a **homogeneous norm** on \mathfrak{g} as:

$$\|\mathbf{v}\| = \|(v_1, \dots, v_r)\| = (\|v_1\|_1^{2r} + \|v_2\|_2^{2r-2} + \dots + \|v_r\|_r^2)^{1/2r},$$

where $\|\cdot\|_i$ is the Euclidean norm on V_i , and $r = \max\{i\}$. This norm is homogeneous under α_t : $\|\alpha_t(\mathbf{v})\| = |t|\|\mathbf{v}\|$, reflecting the grading grades. It satisfies the triangle inequality, as shown in [Sriwongsa and Wiboonton \(2020\)](#), and is detailed in [Moskowitz \(2014, 2010\)](#). For $\mathcal{V}_{(2,3)}$ with $r = 3$, if $\mathbf{u} = (u_1, u_2) \in V_2 \oplus V_3$, then:

$$\|\mathbf{u}\| = (\|u_1\|_2^6 + \|u_2\|_3^2)^{1/6},$$

giving higher weight to lower-degree components. A more general approach is considered in [Salami and Shaska \(2023a\)](#), defining norms for line bundles and weighted heights on weighted projective varieties. For $\mathcal{V}_{\mathbf{q}}^n(k) = k^n$ with $\text{gr}(x_i) = q_i$, a **graded Euclidean norm** can be:

$$\|\mathbf{x}\|_{\mathbf{q}} = \left(\sum_{i=0}^{n-1} q_i |x_i|^2 \right)^{1/2},$$

weighting each coordinate by its grade q_i . Alternatively, a **max-graded norm** is:

$$\|\mathbf{x}\|_{\max} = \max_i \{q_i^{1/2} |x_i|\},$$

emphasizing the dominant graded component, akin to L_∞ norms but adjusted by q_i .

Example 4 For $\mathbf{x} = (x_1, x_2) \in \mathcal{V}_{(2,3)}$ with coordinates in basis \mathcal{B} , let $x_1 = (1, 0, 1) \in V_2$, $x_2 = (1, -1, 0, 1) \in V_3$. The graded Euclidean norm is:

$$\|\mathbf{x}\|_{\mathbf{q}} = (2(1^2 + 0^2 + 1^2) + 3(1^2 + (-1)^2 + 0^2 + 1^2))^{1/2} = \sqrt{2 \cdot 2 + 3 \cdot 3} = \sqrt{13},$$

while the max-graded norm is:

$$\|\mathbf{x}\|_{\max} = \max\{2^{1/2} \cdot 1, 2^{1/2} \cdot 0, 2^{1/2} \cdot 1, 3^{1/2} \cdot 1, 3^{1/2} \cdot 1, 3^{1/2} \cdot 0, 3^{1/2} \cdot 1\} = 3^{1/2}.$$

These differ from the standard $\|\mathbf{x}\| = \sqrt{6}$, highlighting grading's impact.

Properties of Graded Norms. The graded Euclidean norm $\|\cdot\|_{\mathbf{q}}$ is a true norm: (i) $\|\mathbf{x}\|_{\mathbf{q}} \geq 0$, zero iff $\mathbf{x} = 0$; (ii) $\|\lambda\mathbf{x}\|_{\mathbf{q}} = |\lambda|\|\mathbf{x}\|_{\mathbf{q}}$; (iii) $\|\mathbf{x} + \mathbf{y}\|_{\mathbf{q}} \leq \|\mathbf{x}\|_{\mathbf{q}} + \|\mathbf{y}\|_{\mathbf{q}}$ (via Cauchy-Schwarz). The homogeneous norm $\|\cdot\|$ is also a norm, satisfying similar properties under the dilation α_t , and is differentiable except at zero [Sriwongsa and Wiboonton \(2020\)](#). The max-graded norm satisfies norm axioms but is less smooth. These norms extend to infinite I in Hilbert spaces with convergence conditions [Moskowitz \(2014\)](#).

Norm Convexity and Gradient Behavior. Further exploration of norm convexity and gradient behavior is warranted for graded vector spaces, as these properties illuminate their geometric structure. A norm $\|\cdot\|$ is convex if for all $\mathbf{x}, \mathbf{y} \in V$ and $t \in [0, 1]$, $\|t\mathbf{x} + (1-t)\mathbf{y}\| \leq t\|\mathbf{x}\| + (1-t)\|\mathbf{y}\|$. The Euclidean norm $\|\mathbf{x}\| = \sqrt{\sum x_i^2}$ is convex, as its square $\|\mathbf{x}\|^2$ is quadratic with Hessian $\nabla^2(\|\mathbf{x}\|^2) = 2I$, positive definite. For the graded Euclidean norm $\|\mathbf{x}\|_{\mathbf{q}} = (\sum q_i |x_i|^2)^{1/2}$ with $q_i > 0$, let $f(\mathbf{x}) = \|\mathbf{x}\|_{\mathbf{q}}^2 = \sum q_i |x_i|^2$; the Hessian is $\nabla^2 f = 2\text{diag}(q_0, \dots, q_{n-1})$, positive definite, so $\|\cdot\|_{\mathbf{q}}$ is convex.

The homogeneous norm $\|\mathbf{v}\| = \left(\sum \|v_i\|_i^{2r-2(i-1)}\right)^{1/2r}$ is less straightforward. For $\mathcal{V}_{(2,3)}$ ($r = 3$), $\|\mathbf{u}\| = (\|u_1\|_2^6 + \|u_2\|_3^2)^{1/6}$. Define $f(\mathbf{u}) = \|\mathbf{u}\|^6 = \|u_1\|_2^6 + \|u_2\|_3^2$; the Hessian includes $\partial^2 f / \partial u_{1j}^2 = 30u_{1j}^4$, positive for $\mathbf{u} \neq 0$, but near zero, high exponents (i.e., 6) disrupt convexity. However, $\|\mathbf{u}\|$ is quasiconvex, as sublevel sets $\{\mathbf{u} \mid \|\mathbf{u}\| \leq c\}$ are convex for $c > 0$ [Sriwongsa and Wiboonton \(2020\)](#), reflecting a weaker but useful property. The max-graded norm $\|\mathbf{x}\|_{\max} = \max\{q_i^{1/2} |x_i|\}$ is convex, as the maximum of convex functions $q_i^{1/2} |x_i|$, with sublevel sets being intersections of slabs $\{\mathbf{x} \mid q_i^{1/2} |x_i| \leq c\}$ [Boyd and Vandenberghe \(2004\)](#).

Gradient behavior is analyzed via the function $f(\mathbf{x}) = \|\mathbf{x}\|^2$. For the Euclidean norm, $f(\mathbf{x}) = \sum x_i^2$, $\nabla f = 2\mathbf{x}$, linear and isotropic. For $\|\cdot\|_{\mathbf{q}}$, $f(\mathbf{x}) = \sum q_i |x_i|^2$, $\nabla f = 2(q_0 x_0, \dots, q_{n-1} x_{n-1})$, scaling components by q_i , with magnitude $\|\nabla f\|_2 = 2\sqrt{\sum q_i^2 x_i^2}$. For the homogeneous norm on $\mathcal{V}_{(2,3)}$, $f(\mathbf{u}) = \|u_1\|_2^6 + \|u_2\|_3^2$, $\nabla f = (6\|u_1\|_2^4 u_1, 2u_2)$, nonlinear with steep growth in V_2 (exponent 4) versus V_3 (exponent 1). The max-graded norm's $f(\mathbf{x}) = (\max q_i^{1/2} |x_i|)^2$ has a subdifferential, i.e., $\partial f / \partial x_i = 2q_i^{1/2} \text{sgn}(x_i) \max\{q_j^{1/2} |x_j|\}$ if i achieves the max, zero otherwise, reflecting discontinuity [Boyd and Vandenberghe \(2004\)](#).

These properties—convexity and gradient variation—highlight how graded norms shape the geometry of $\mathcal{V}_{\mathbf{q}}^n$ and $\bigoplus V_i$, offering diverse tools for algebraic and analytic applications [Boyd and Vandenberghe \(2004\)](#); [Sriwongsa and Wiboonon \(2020\)](#).

3. Graded Neural Networks (GNN)

We define artificial neural networks over graded vector spaces, utilizing Section 2. Let k be a field, and for $n \geq 1$, denote \mathbb{A}_k^n (resp. \mathbb{P}_k^n) as the affine (resp. projective) space over k , omitting the subscript if k is algebraically closed. A tuple $\mathbf{q} = (q_0, \dots, q_{n-1}) \in \mathbb{N}^n$ defines the **grades**, with $\text{gr}(x_i) = q_i$. The graded vector space $\mathcal{V}_{\mathbf{q}}^n(k) = k^n$ has scalar multiplication:

$$\lambda \star \mathbf{x} = (\lambda^{q_0} x_0, \dots, \lambda^{q_{n-1}} x_{n-1}), \quad \mathbf{x} = (x_0, \dots, x_{n-1}) \in k^n, \lambda \in k,$$

as in Section 2, denoted $\mathcal{V}_{\mathbf{q}}$ when clear. This scalar action, denoted $\lambda \star \mathbf{x}$, mirrors the graded multiplication in Section 2, applicable to both the coordinate form here and the direct sum form (e.g., $\lambda \star [f, g]$) via basis representation. A **graded neuron** on $\mathcal{V}_{\mathbf{q}}$ is $\alpha_{\mathbf{q}} : \mathcal{V}_{\mathbf{q}}^n \rightarrow k$ such that

$$\alpha_{\mathbf{q}}(\mathbf{x}) = \sum_{i=0}^{n-1} w_i^{q_i} x_i + b,$$

where $w_i \in k$ are **neural weights**, and $b \in k$ is the **bias**. For $b = 0$,

$$\alpha_{\mathbf{q}}(\lambda \star \mathbf{x}) = \sum (\lambda w_i)^{q_i} x_i = \lambda \sum w'_i x_i$$

for ($w'_i = w_i^{q_i}$), approximating a graded linear map of degree 1 per Section 2. With $b \neq 0$, $\alpha_{\mathbf{q}}$ is affine, embedding grading via $w_i^{q_i}$. A **graded network layer** is:

$$\begin{aligned} \phi : \mathcal{V}_{\mathbf{q}}^n(k) &\rightarrow \mathcal{V}_{\mathbf{q}}^n(k) \\ \mathbf{x} &\rightarrow g(W\mathbf{x} + \mathbf{b}), \end{aligned}$$

where $W = [w_{j,i}^{q_i}] \in k^{n \times n}$, $\mathbf{b} = (b_0, \dots, b_{n-1}) \in k^n$, and ϕ preserves grading, with $\text{gr}(y_j) = q_j$. A **graded neural network** (GNN) is a composition of multiple layers given as

$$\hat{\mathbf{y}} = \phi_m \circ \dots \circ \phi_1(\mathbf{x}),$$

where each layer $\phi_l(\mathbf{x}) = g_l(W^l \mathbf{x} + \mathbf{b}^l)$ applies a transformation defined by the matrix of neural weights $W^l = [w_{j,i}^{q_i}]$, producing outputs $\hat{\mathbf{y}}$ and true values \mathbf{y} in $\mathcal{V}_{\mathbf{q}}^n$ with grades $\text{gr}(\hat{y}_i) = q_i$.

3.1. ReLU Activation

In classical neural networks, the rectified linear unit (ReLU) activation, defined as $\text{ReLU}(x) = \max\{0, x\}$, applies a simple thresholding to promote sparsity and efficiency. However, for graded neural networks over $\mathcal{V}_{\mathbf{q}}^n$, where $\mathbf{x} = (x_0, \dots, x_{n-1})$ has coordinates with grades $\text{gr}(x_i) = q_i$ and scalar action $\lambda \star \mathbf{x} = (\lambda^{q_0} x_0, \dots, \lambda^{q_{n-1}} x_{n-1})$, a direct application of this ReLU ignores the grading's intrinsic scaling. To adapt to this structure, we define a *graded ReLU* that adjusts nonlinearity by grade. For $\mathbf{x} \in \mathcal{V}_{\mathbf{q}}^n$, the graded ReLU is:

$$\text{ReLU}_i(x_i) = \max\{0, |x_i|^{1/q_i}\}, \quad \text{and} \quad \text{ReLU}(\mathbf{x}) = (\text{ReLU}_0(x_0), \dots, \text{ReLU}_{n-1}(x_{n-1})).$$

Unlike the classical $\max\{0, x_i\}$, which treats all coordinates uniformly, this version scales each x_i by $1/q_i$, reflecting the graded action. For $\lambda \star \mathbf{x} = (\lambda^{q_i} x_i)$, compute:

$$\text{ReLU}_i(\lambda^{q_i} x_i) = \max\{0, |\lambda^{q_i} x_i|^{1/q_i}\} = \max\{0, |\lambda| |x_i|^{1/q_i}\} = |\lambda| \max\{0, |x_i|^{1/q_i}\},$$

so $\text{ReLU}(\lambda \star \mathbf{x}) = |\lambda| \text{ReLU}(\mathbf{x})$ for $\lambda > 0$, aligning with $\mathcal{V}_{\mathbf{q}}^n$'s grading up to magnitude. This ensures the activation respects the differential scaling of coordinates (i.e., $q_i = 2$ vs. $q_i = 3$ in $\mathcal{V}_{(2,3)}$), unlike the classical ReLU, where $\text{ReLU}(\lambda x_i) = \lambda \text{ReLU}(x_i)$ for $\lambda > 0$ assumes homogeneity of degree 1.

This adaptation is motivated by the need to capture feature significance in graded spaces, as seen in applications like genus two curve invariants (J_2, J_4, J_6, J_{10} with grades 2, 4, 6, 10). A classical ReLU might underweight high-graded features (i.e., J_{10}) or overreact to low-graded ones (i.e., J_2), whereas the graded ReLU normalizes sensitivity via $1/q_i$, akin to the homogeneous norm's scaling in Section 2. It also mirrors weighted heights from [Salami and Shaska \(2023a, 2024\)](#), where exponents adjust to graded geometry.

Example 5 Consider $\mathcal{V}_{(2,3)}$ from Example 1, with $\mathbf{q} = (2, 2, 2, 3, 3, 3, 3)$ and basis

$$\mathcal{B} = \{x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}.$$

Let $\mathbf{u} = (2, -3, 1, 1, -2, 1, 1)$. Here, $\mathbf{u} = (2, -3, 1, 1, -2, 1, 1)$ represents the coordinates of a polynomial $\mathbf{u} = [f, g] \in V_2 \oplus V_3$ in the basis $\mathcal{B} = \{x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}$ from Example 1, mapping $f = 2x^2 - 3xy + y^2$ and $g = x^3 - 2x^2y + xy^2 + y^3$ to k^7 :

$$\text{ReLU}(\mathbf{u}) = (\sqrt{2}, 3, 1, 1, \sqrt{2}, 1, 1),$$

e.g., $\text{ReLU}_0(2) = \sqrt{2}$ ($q_0 = 2$), $\text{ReLU}_1(-3) = 3$ ($q_1 = 2$), $\text{ReLU}_3(1) = 1$ ($q_3 = 3$). Compare to classical ReLU: $\text{ReLU}(-3) = 0$, $\text{ReLU}(2) = 2$, yielding $(2, 0, 1, 1, 0, 1, 1)$, which loses the graded nuance (e.g., $-3 \rightarrow 3$ vs. 0). The graded version preserves $\mathcal{V}_{\mathbf{q}}^n$ while adjusting output scale.

The graded ReLU thus balances nonlinearity with grading, potentially enhancing feature discrimination in $\mathcal{V}_{\mathbf{q}}^n$ compared to the uniform thresholding of classical ReLU. Its efficiency relative to other adaptations (i.e., $\max\{0, x_i/q_i\}$) remains to be explored, but its form leverages the algebraic structure established in Section 2.

3.2. Graded Loss Functions

In classical neural networks, loss functions like the mean squared error (MSE), $L = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2$, treat all coordinates equally, assuming a uniform vector space structure. However, on $\mathcal{V}_{\mathbf{q}}^n(k) = k^n$ with grading $\text{gr}(x_i) = q_i$ and scalar action $\lambda \star \mathbf{x} = (\lambda^{q_0} x_0, \dots, \lambda^{q_{n-1}} x_{n-1})$, this approach overlooks the differential significance of coordinates (i.e., $q_i = 2$ vs. $q_i = 10$ in genus two invariants). Graded loss functions adapt to this structure by weighting errors according to q_i , enhancing sensitivity to features of varying grades, as motivated by the improved accuracy in graded inputs observed in [Shaska and Shaska \(2024\)](#).

The **graded MSE** on $\mathcal{V}_{\mathbf{q}}^n$ is:

$$L_{\text{MSE}}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=0}^{n-1} q_i (y_i - \hat{y}_i)^2,$$

where $\mathbf{y}, \hat{\mathbf{y}} \in \mathcal{V}_{\mathbf{q}}^n$ are true and predicted values, and q_i amplifies errors for higher-graded coordinates. Unlike classical MSE, this scales with grading: for $\lambda \star (\mathbf{y} - \hat{\mathbf{y}}) = (\lambda^{q_i} (y_i - \hat{y}_i))$, $L_{\text{MSE}}(\lambda \star \mathbf{y}, \lambda \star \hat{\mathbf{y}}) = \frac{1}{n} \sum q_i \lambda^{2q_i} (y_i - \hat{y}_i)^2$, reflecting $\mathcal{V}_{\mathbf{q}}^n$'s geometry. Alternatively, using the graded Euclidean norm from Section 2:

$$L_{\text{norm}}(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|_{\mathbf{q}}^2 = \sum_{i=0}^{n-1} q_i |y_i - \hat{y}_i|^2,$$

omits the $1/n$ normalization, aligning directly with $\|\cdot\|_{\mathbf{q}}$'s definition.

For spaces like $\mathcal{V}_{(2,3)}$ with $\mathcal{V}_{\mathbf{q}}^n = k^7$, we partition coordinates as $\mathbf{y} = (\mathbf{y}_2, \mathbf{y}_3)$, where $\mathbf{y}_2 = (y_0, y_1, y_2) \in k^3$ corresponds to V_2 (grade 2) and $\mathbf{y}_3 = (y_3, y_4, y_5, y_6) \in k^4$ to V_3 (grade 3), matching the basis \mathcal{B} from Example 1. The **homogeneous loss** leverages the homogeneous norm from Section 2:

$$L_{\text{hom}}(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|^6 = \|(\mathbf{y} - \hat{\mathbf{y}})_2\|_2^6 + \|(\mathbf{y} - \hat{\mathbf{y}})_3\|_3^2,$$

where $r = 3$, emphasizing lower-graded errors (i.e., V_2 with exponent 6) over higher-graded ones (V_3 with 2).

Additional loss functions enrich this framework. A **max-graded loss** uses the max-graded norm:

$$L_{\text{max}}(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|_{\text{max}}^2 = \left(\max_i \{q_i^{1/2} |y_i - \hat{y}_i|\} \right)^2,$$

focusing on the largest grade-adjusted error, akin to L_{∞} but tuned to q_i . For classification in $\mathcal{V}_{\mathbf{q}}^n$, a **graded cross-entropy** could be:

$$L_{\text{CE}}(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=0}^{n-1} q_i y_i \log(\hat{y}_i),$$

assuming \hat{y}_i are probabilities (i.e., via a softmax on $\mathcal{V}_{\mathbf{q}}^n$), weighting log-losses by grade to prioritize high- q_i classes.

Example 6 For $\mathbf{y} = (1, 2, 0, 1, 0, 1, 1)$, $\hat{\mathbf{y}} = (0, 1, 1, 1, -1, 0, 1)$ in $\mathcal{V}_{(2,3)}$ ($\mathbf{q} = (2, 2, 2, 3, 3, 3, 3)$):

$$L_{\text{MSE}} = \frac{1}{7} [2 \cdot 1^2 + 2 \cdot 1^2 + 2 \cdot 1^2 + 3 \cdot 0^2 + 3 \cdot 1^2 + 3 \cdot 1^2 + 3 \cdot 0^2] = \frac{11}{7},$$

$$L_{\text{norm}} = 2 \cdot 3 + 3 \cdot 2 = 11,$$

$$L_{\text{hom}} = (3^3 + 2)^2 = 841, \quad \text{with } \|(\mathbf{y} - \hat{\mathbf{y}})_2\|_2^2 = 3, \quad \|(\mathbf{y} - \hat{\mathbf{y}})_3\|_3^2 = 2,$$

$$L_{\text{max}} = \left(\max\{2^{1/2} \cdot 1, 2^{1/2} \cdot 1, 2^{1/2} \cdot 1, 3^{1/2} \cdot 0, 3^{1/2} \cdot 1, 3^{1/2} \cdot 1, 3^{1/2} \cdot 0\} \right)^2 = 3.$$

Classical MSE gives $\frac{6}{7}$, underweighting V_3 errors (i.e., 1^2 vs. $3 \cdot 1^2$).

These graded losses adapt classical metrics to $\mathcal{V}_{\mathbf{q}}^n$'s structure, offering flexibility— L_{MSE} and L_{norm} balance all errors, L_{hom} prioritizes grade hierarchy, L_{max} targets outliers, and L_{CE} suits classification—all leveraging q_i to reflect feature significance [Boyd and Vandenberghe \(2004\)](#).

3.3. Optimizers

Optimizers adjust weights $w_{j,i}$ and \mathbf{b}_j to minimize a loss function over $\mathcal{V}_{\mathbf{q}}^n$. Consider $L = L_{\text{norm}}(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|_{\mathbf{q}}^2$, using the graded Euclidean norm from Section 2, where $\|\mathbf{x}\|_{\mathbf{q}}^2 = \sum_{i=0}^{n-1} q_i |x_i|^2$. The gradient with respect to $\hat{\mathbf{y}}$, as derived in Section 2 (“Norm Convexity and Gradient Behavior”), is:

$$\nabla_{\hat{\mathbf{y}}} L = 2(q_0(\hat{y}_0 - y_0), \dots, q_{n-1}(\hat{y}_{n-1} - y_{n-1})),$$

reflecting the grading via q_i . This gradient scales components by their grades, emphasizing higher-graded coordinates (i.e., $q_i = 3$ in V_3 of $\mathcal{V}_{(2,3)}$).

Basic gradient descent updates parameters as:

$$w_{j,i}^{t+1} = w_{j,i}^t - \eta \frac{\partial L}{\partial w_{j,i}}, \quad \mathbf{b}_j^{t+1} = \mathbf{b}_j^t - \eta \frac{\partial L}{\partial \mathbf{b}_j},$$

where $\eta > 0$ is a step size, and partial derivatives are computed via the chain rule through ϕ_l , incorporating q_i from $w_{j,i}^{q_i}$ and W^l . For example, if $\hat{y}_j = \phi_l(x_j)$, $\partial L / \partial w_{j,i} = q_i w_{j,i}^{q_i-1} x_i \cdot \partial L / \partial \hat{y}_j$, adjusting for grading.

Other norms yield different gradients. For $L_{\text{hom}} = \|\mathbf{y} - \hat{\mathbf{y}}\|^{2r}$ (i.e., $r = 3$ for $\mathcal{V}_{(2,3)}$), the gradient from Section 2 is nonlinear:

$$\nabla_{\hat{\mathbf{y}}} L = 2r \|\mathbf{y} - \hat{\mathbf{y}}\|^{2r-6} (\|(\mathbf{y} - \hat{\mathbf{y}})_2\|_2^4 (\hat{\mathbf{y}}_2 - \mathbf{y}_2), (\hat{\mathbf{y}}_3 - \mathbf{y}_3)),$$

emphasizing magnitude disparities across grades. The max-graded norm $L = \|\mathbf{y} - \hat{\mathbf{y}}\|_{\max}^2$ has a subdifferential, i.e., $\partial L / \partial \hat{y}_i = 2q_i^{1/2} \text{sgn}(\hat{y}_i - y_i) \max\{q_j^{1/2} |\hat{y}_j - y_j|\}$ if i achieves the maximum [Boyd and Vandenberghe \(2004\)](#).

Alternative optimizers include momentum-based methods (i.e., adding a velocity term $v^{t+1} = \beta v^t - \eta \nabla L$), Adam (adaptive moment estimation), or RMSprop, which adjust η using gradient statistics. For $\|\cdot\|_{\mathbf{q}}^2$, these can use the same ∇L , but q_i -scaling may require grade-specific rates (i.e., $\eta_i \propto q_i^{-1}$) to balance updates. The homogeneous norm’s nonlinearity suggests cautious step sizes, while the max-graded norm’s sparsity suits subgradient methods [Boyd and Vandenberghe \(2004\)](#); [Goodfellow et al. \(2016\)](#).

4. Theoretical Implementation and Applications of Graded Neural Networks

Having defined GNNs over $\mathcal{V}_{\mathbf{q}}^n$ in Section 3, we now explore their computational implementation and potential applications. This section first examines theoretical challenges arising from the graded structure, then highlights how this framework could extend to practical domains, leveraging its algebraic properties established in Section 2 and Section 3.

4.1. Implementation Challenges

The graded scalar action $\lambda \star \mathbf{x} = (\lambda^{q_i} x_i)$ introduces numerical stability concerns, as large q_i amplify small λ , risking overflow or precision loss in finite arithmetic. For $\mathcal{V}_{\mathbf{q}}^n$ with $\mathbf{q} = (q_0, \dots, q_{n-1})$, inputs must be normalized to mitigate this, yet balancing scales across grades remains non-trivial.

Neuron computation $\alpha_{\mathbf{q}}(\mathbf{x}) = \sum w_i^{q_i} x_i + b$ and layers $\phi_l(\mathbf{x}) = g_l(W^l \mathbf{x} + \mathbf{b}^l)$ with $W^l = [w_{j,i}^{q_i}]$ face complexity from exponentiation. For large q_i , $w_i^{q_i}$ grows exponentially if $|w_i| > 1$, requiring careful weight initialization (e.g., $|w_i| < 1$) or precomputation, increasing memory demands. Sparse \mathbf{q} may reduce this, but dense grading scales poorly with n .

The graded ReLU $\text{ReLU}_i(x_i) = \max\{0, |x_i|^{1/q_i}\}$ is sensitive to q_i : small q_i (e.g., 2) yield smooth outputs, while large q_i (e.g., 10) flatten near zero, potentially reducing expressivity. This variability complicates uniform activation design across layers, unlike classical ReLU’s consistency.

Loss functions like $L_{\text{norm}} = \sum q_i |y_i - \hat{y}_i|^2$ amplify errors by q_i , skewing optimization toward high-graded coordinates, while L_{hom} requires partitioning (e.g., $k^7 \rightarrow V_2, V_3$), adding preprocessing overhead. Gradients $\nabla_{\hat{y}} L = 2(q_i(\hat{y}_i - y_i))$ scale with q_i , risking vanishing or exploding gradients for extreme q_i , necessitating adaptive step sizes (e.g., $\eta_i \propto q_i^{-1}$) or normalization, per Section 3.

4.2. Potential Applications

The graded structure of GNNs offers versatility across domains. In machine learning, assigning grades to features based on significance (e.g., genus two invariants with $\mathbf{q} = (2, 4, 6, 10)$) enhances sensitivity, as seen in [Shaska and Shaska \(2024\)](#), potentially improving tasks like regression or classification where features vary in importance. Temporal signal processing could leverage grading to prioritize recent data (e.g., $\mathbf{q} = (1, 2, 3, \dots)$), adapting loss functions like L_{norm} to time-weighted errors.

Beyond traditional computing, photonic implementations, such as laser-based systems, present intriguing possibilities. Recent advances emulate graded responses using quantum-dot lasers for high-speed reservoir computing [Nie et al. \(2024\)](#), achieving rates like 10 GBaud without feedback loops. GNNs’ graded neurons ($\sum w_i^{q_i} x_i$) and activations ($\text{ReLU}_i(x_i)$) could map to such hardware, where q_i tunes photonic dynamics, offering ultrafast processing for real-time applications. This suggests a synergy: the algebraic grading of $\mathcal{V}_{\mathbf{q}}^n$ might inform novel hardware designs, extending beyond conventional platforms while addressing challenges like computational scalability for large n or diverse \mathbf{q} .

5. Conclusions

This paper introduces a novel framework for graded neural networks (GNNs) over graded vector spaces $\mathcal{V}_{\mathbf{q}}^n$, unifying algebraic grading from Section 2 with neural computation in Section 3. By defining neurons, layers, activations, and loss functions with grade-sensitive operations (e.g., $\sum w_i^{q_i} x_i$, $\text{ReLU}_i(x_i) = \max\{0, |x_i|^{1/q_i}\}$), we extend classical neural networks to capture feature hierarchies, as motivated by applications like genus two curve classification [Shaska and Shaska \(2024\)](#). The flexibility of $\mathbf{q} = (q_0, \dots, q_{n-1})$ enables tailored grading, distinct from uniform vector spaces.

Section 4 underscores the theoretical challenges—numerical stability, computational complexity—and potential applications, from machine learning to photonic systems. While implementation demands careful handling of large q_i and gradient scaling, the framework’s generality suggests broad utility. Future work could explore empirical validation across diverse datasets, optimization refinements, and hardware realizations, such as laser-based graded neurons, to harness GNNs’ full potential. This foundation offers a stepping stone for advancing graded computation in both theory and practice.

References

- I. N. Balaba. Isomorphisms of graded rings of linear transformations of graded vector spaces. *Chebyshvskiu i Sb.*, 6(4(16)):7–24, 2005. ISSN 2226-8383,2587-7119.

- L. Beshaj, J. Gutierrez, and T. Shaska. Weighted greatest common divisors and weighted heights. *J. Number Theory*, 213:319–346, 2020. ISSN 0022-314X. doi: 10.1016/j.jnt.2019.12.012. URL <https://doi-org.huaryu.kl.oakland.edu/10.1016/j.jnt.2019.12.012>.
- Vitalij M. Bondarenko. Linear operators on S -graded vector spaces. volume 365, pages 45–90. 2003. doi: 10.1016/S0024-3795(02)00689-4. URL [https://doi.org/10.1016/S0024-3795\(02\)00689-4](https://doi.org/10.1016/S0024-3795(02)00689-4). Special issue on linear algebra methods in representation theory.
- N. Bourbaki. *Algebra I*. Springer, 1974. ISBN 978-3-540-64243-5. Chapter 3.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- J.-L. Koszul. Graded manifolds and graded Lie algebras. In *Proceedings of the international meeting on geometry and physics (Florence, 1982)*, pages 71–84. Pitagora, Bologna, 1983. ISBN 88-371-0280-1.
- Martin Moskowitz. An extension of Minkowski’s theorem to simply connected 2-step nilpotent groups. *Port. Math.*, 67(4):541–546, 2010. ISSN 0032-5155,1662-2758. doi: 10.4171/PM/1877. URL <https://doi.org/10.4171/PM/1877>.
- Martin Moskowitz. The triangle inequality for graded real vector spaces of length 3 and 4. *Math. Inequal. Appl.*, 17(3):1027–1030, 2014. ISSN 1331-4343,1848-9966. doi: 10.7153/mia-17-75. URL <https://doi.org/10.7153/mia-17-75>.
- Yikun Nie, Bo Yang, Dongliang Wang, Ting Wang, Jiawei Wang, Zihao Wang, and Chaoran Huang. Integrated laser graded neuron enabling high-speed reservoir computing without a feedback loop. *Optica*, 11(12):1690–1699, Dec 2024. doi: 10.1364/OPTICA.537231. URL <https://opg.optica.org/optica/abstract.cfm?URI=optica-11-12-1690>.
- Steven Roman. *Advanced linear algebra*, volume 135 of *Graduate Texts in Mathematics*. Springer, New York, third edition, 2008. ISBN 978-0-387-72828-5.
- S. Salami and T. Shaska. Local and global heights on weighted projective varieties. *Houston J. Math.*, 49(3):603–636, 2023a. ISSN 0362-1588. URL [www.math.uh.edu/~hjm/restricted/pdf49\(3\)/09shaska.pdf](http://www.math.uh.edu/~hjm/restricted/pdf49(3)/09shaska.pdf).
- S. Salami and T. Shaska. Vojta’s conjecture on weighted projective varieties. 2024. URL <https://arxiv.org/abs/2309.10300>.
- Sajad Salami and Tony Shaska. Local and global heights on weighted projective varieties. *Houston J. Math.*, 49(3):603–636, 2023b. ISSN 0362-1588. URL [www.math.uh.edu/~hjm/restricted/pdf49\(3\)/09shaska.pdf](http://www.math.uh.edu/~hjm/restricted/pdf49(3)/09shaska.pdf).
- E. Shaska and T. Shaska. Machine learning for moduli space of genus two curves. 2024. URL <https://arxiv.org/abs/2403.17250>.
- Songpon Sriwongsa and Keng Wiboonton. The triangle inequality for graded real vector spaces. *Math. Inequal. Appl.*, 23(1):351–355, 2020. ISSN 1331-4343,1848-9966. doi: 10.7153/mia-2020-23-27. URL <https://doi.org/10.7153/mia-2020-23-27>.