FINSLER METRIC CLUSTERING IN WEIGHTED PROJECTIVE SPACES

T. SHASKA

Abstract. This paper develops a hierarchical clustering algorithm for weighted projective spaces $\mathbb{P}_{\mathbf{q}}$, utilizing a Finsler metric $d_F([z],[w])$ and its rational analogue $d_{F,\mathbb{Q}}([z],[w])$ to define distances that preserve the non-Euclidean geometry of these quotient manifolds. Defined via geodesic integrals of a scaling-invariant Finsler norm weighted by the grades $\mathbf{q} = (q_0, q_1, \ldots, q_n)$, these metrics satisfy true metric properties—including the triangle inequality —overcoming the limitations of the non-metric dissimilarity measure from prior work.

1. Introduction

The study of data clustering in non-Euclidean manifolds presents significant challenges and opportunities, particularly in spaces endowed with intricate geometric structures, such as weighted projective spaces. These spaces, defined as quotients of complex vector spaces under weighted scaling actions, naturally arise in diverse fields, including arithmetic geometry, dynamical systems, and data analysis, where projective symmetries govern the underlying data. Traditional clustering methods, often reliant on Euclidean metrics, fail to capture the intrinsic geometry of such spaces, leading to distorted groupings that obscure meaningful patterns. Motivated by the need to address these limitations, this paper introduces a novel hierarchical clustering algorithm tailored for weighted projective spaces, employing a Finsler metric to define distances that respect the manifold's weighted structure. The theoretical framework developed herein builds on our prior work [1], which successfully applied a non-metric dissimilarity measure to cluster rational points in moduli spaces, and extends it through a rigorous metric-based approach, offering a robust foundation for geometric and arithmetic applications.

This work is inspired by advances in graded computational frameworks within our broader program to develop machine learning techniques for graded spaces, as detailed in [2,3]. In [2], neural networks operate on graded vector spaces where coordinates are assigned grades, analogous to the weights $\mathbf{q} = (q_0, q_1, \ldots, q_n)$ defining $\mathbb{P}_{\mathbf{q}}$. This grading is extended in [3] to graded neural networks (GNNs), which weight features by grades, as exemplified by the moduli space $\mathbb{P}_{(2,4,6,10)}$. These frameworks suggest that GNNs could preprocess points in $\mathbb{P}_{\mathbf{q}}$, learning graded representations that enhance our clustering algorithm's geometric fidelity.

Weighted projective spaces, denoted $\mathbb{P}_{\mathbf{q}}$ for weights $\mathbf{q} = (q_0, q_1, \ldots, q_n)$, are quotients of $\mathbb{C}^{n+1} \setminus \{0\}$ under the equivalence relation

$$(z_0, z_1, \ldots, z_n) \sim (\lambda^{q_0} z_0, \lambda^{q_1} z_1, \ldots, \lambda^{q_n} z_n)$$

Key words and phrases. Geometric clustering and Weighted projective spaces and Finsler metrics and Non-Euclidean manifolds.

for $\lambda \in \mathbb{C}^*$. These spaces generalize standard projective spaces, incorporating weights that reflect varying degrees of coordinates, as seen in the moduli space of genus two curves $\mathbb{P}_{(2,4,6,10)}$, where Igusa invariants have degrees 2, 4, 6, and 10 [4]. The geometric complexity of $\mathbb{P}_{\mathbf{q}}$, characterized by its quotient structure, necessitates distance measures that preserve scaling symmetries (see Section 2 for preliminaries). While our earlier work [1] utilized a dissimilarity measure $d([z],[w]) = \inf_{\lambda,\mu\in\mathbb{C}^*} \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\|$, which effectively clustered points but lacked the triangle inequality, this paper introduces a Finsler metric $d_F([z],[w])$ and its rational counterpart $d_{F,\mathbb{Q}}([z],[w])$, defined via geodesic integrals of a Finsler norm (Section 3), to provide true metric properties—including the triangle inequality, rigorously proven in Section 3.1—essential for robust clustering algorithms. Preprocessing steps, detailed in Section 5.2, normalize points using the weighted norm $\sum_{k=0}^{n} q_k |z_k|^2 = 1$ for geometric applications and wgcd $= 1$ for arithmetic ones, ensuring consistency across contexts.

The Finsler norm, defined for a point $[z] \in \mathbb{P}_{\mathbf{q}}$ with representative $z \in \mathbb{C}^{n+1} \setminus \{0\}$ and tangent vector $v \in \mathbb{C}^{n+1}$, is given in Eq. (35) and induces the Finsler distance in Eq. (36). This construction, inspired by Finsler geometry principles [5], ensures non-negativity, symmetry, zero distance implies equality, and the triangle inequality, making $d_F([z],[w])$ a true metric. For rational points in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, a similar norm $F_{\mathbb{Q}}([z], v)$ defines $d_{F,\mathbb{Q}}([z],[w])$, enabling arithmetic applications. The development of these metrics, detailed through the lifting of curves to $\mathbb{C}^{n+1} \setminus \{0\}$ and the characterization of Finsler geodesics, provides a geometrically faithful framework for clustering, theoretically extending the capabilities of our prior non-metric approach [6].

The hierarchical clustering algorithm, designed to operate directly in $\mathbb{P}_{\mathbf{q}}$, constructs a dendrogram by iteratively merging clusters based on the Finsler metric, using linkage criteria such as single or average linkage. The algorithm's correctness and stability, proven through rigorous mathematical statements, ensure reliable partitioning of datasets, preserving the weighted projective geometry. Preprocessing steps, including normalization and dimensionality reduction via weighted principal component analysis, enhance computational efficiency while maintaining geometric fidelity. The computational challenge of geodesic optimization, addressed through variational methods and discrete approximations, is theoretically manageable, though practical scalability awaits empirical validation [7].

Our primary applications lie in arithmetic geometry and dynamical systems. In the moduli space of genus two curves, represented as $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$, the algorithm clusters rational points by their Igusa invariants, identifying curves with (n,n)-split Jacobians and supporting isogeny-based cryptographic studies. Building on prior results [1], the Finsler metric theoretically refines these groupings, preserving arithmetic patterns such as weighted height distributions. In Arithmetic Dynamics, this method applies the algorithm to rational functions on the projective line $\mathbb{P}^1$, clustering points in weighted projective spaces to analyze dynamical invariants like periodic point structures [8]. Moreover, the graded neural networks from [3] offer promising applications, potentially enhancing clustering through graded feature representations.

Theoretically, the Finsler metric's true metric properties enable compatibility with a wide range of machine learning algorithms, from clustering to classification, directly in $\mathbb{P}_{\mathbf{q}}$. Future directions include developing efficient geodesic computation

methods, exploring alternative clustering algorithms like spectral clustering, and extending the framework to fields such as robotics and quantum computing, where non-Euclidean geometries are prevalent. A significant avenue is the development of graded neural networks, assigning weights to features to model projective data, potentially revolutionizing non-Euclidean data analysis [3]. While the practical efficacy of the Finsler metric awaits experimental validation, contrasting with the established performance of $d([z],[w])$, this paper establishes a robust theoretical foundation, advancing the study of clustering in weighted projective spaces with profound implications for arithmetic geometry, dynamical systems, and beyond.

2. Preliminaries

2.1. **Weighted projective spaces (WPS).** Let $\mathbb{F}$ be a field and $q_0, q_1, \ldots, q_n$ be positive integers called *weights*. The tuple of weights is denoted by $\mathbf{q} := (q_0, q_1, \ldots, q_n)$. The *weighted projective space* $\mathbb{P}_{\mathbf{q}}$ is defined as the quotient space of $\mathbb{F}^{n+1} \setminus \{0\}$ under the equivalence relation

$$(z_0, z_1, \ldots, z_n) \sim (\lambda^{q_0} z_0, \lambda^{q_1} z_1, \ldots, \lambda^{q_n} z_n) \tag{1}$$

for all $\lambda \in \mathbb{F}^*$, where $\mathbb{F}^*$ represents the multiplicative group of non-zero elements in $\mathbb{F}$. A point in $\mathbb{P}_{\mathbf{q}}$ is an equivalence class $[z] = [z_0 : z_1 : \cdots : z_n]$, with the weights q_i governing the scaling of each coordinate. This construction extends the standard projective space, recovered when $q_0 = q_1 = \cdots = q_n = 1$. For the purposes of this paper, we primarily consider $\mathbb{F} = \mathbb{C}$ for geometric clustering applications and $\mathbb{F} = \mathbb{Q}$ for arithmetic geometry contexts, addressing both the geometric and Diophantine aspects of weighted projective spaces. The weights $\mathbf{q}$ define a grading structure analogous to the graded vector spaces in [2,3], positioning $\mathbb{P}_{\mathbf{q}}$ as a natural framework for advancing machine learning techniques within our program for graded spaces.

Remark 1. *The quotient structure of $\mathbb{P}_{\mathbf{q}}$ under weighted scaling informs the construction of the Finsler metric in Section 3, ensuring that distances respect the manifold's weighted geometry.*

2.2. **Heights on WPS.** To study the arithmetic properties of points in weighted projective spaces, we focus on rational points in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, where coordinates $z_i \in \mathbb{Q}$ and the equivalence relation employs scalings $\lambda \in \mathbb{Q}^*$. Drawing on the framework established in [9], we normalize representatives of rational points to define a weighted height function that quantifies their arithmetic complexity. This graded structure supports our program's goal, as outlined in [2], to develop machine learning methods for arithmetic data analysis, potentially leveraging graded neural networks from [3]. For a point $[z] = [z_0 : z_1 : \cdots : z_n] \in \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, a normalized representative is chosen as $(x_0, x_1, \ldots, x_n) \in \mathbb{Z}^{n+1} \setminus \{0\}$ such that the weighted greatest common divisor, denoted $\mathrm{wgcd}(x_0, x_1, \ldots, x_n)$, equals 1. The wgcd is the largest positive integer d for which there exists a $\lambda \in \mathbb{Q}^*$ satisfying $\lambda^{q_i} x_i / d \in \mathbb{Z}$ for all $i = 0, 1, \ldots, n$. This normalization ensures that the representative is unique up to scaling by roots of unity in $\mathbb{Q}^*$, providing a canonical form for arithmetic analysis.

The *weighted height* of a point $[z] \in \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, using its normalized representative $(x_0, x_1, \ldots, x_n)$, is defined as

$$h_w([z]) = \max_{i=0,\ldots,n} \left(|x_i|^{1/q_i} \right). \tag{2}$$

This height function is invariant under the weighted scaling action, since scaling the representative $(x_0, \ldots, x_n)$ by $\lambda \in \mathbb{Q}^*$ transforms each coordinate x_i to $\lambda^{q_i} x_i$, and the term $|\lambda^{q_i} x_i|^{1/q_i} = |\lambda|\,|x_i|^{1/q_i}$ preserves the maximum up to a constant factor that cancels in the equivalence class. The weighted height serves as a measure of arithmetic complexity, enabling the ordering of rational points in databases or the analysis of Diophantine properties. For instance, in the moduli space $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$, the weighted height quantifies the complexity of Igusa invariants, facilitating applications in arithmetic geometry as explored in [1].

Remark 2. *For rational points in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, the normalization using* $\gcd = 1$ *is crucial for arithmetic applications, such as clustering with the rational Finsler distance $d_{F,\mathbb{Q}}([z],[w])$ in Section 5, distinct from the geometric normalization $\sum_{k=0}^{n} q_k |z_k|^2 = 1$ used in Section 5.2.*

2.3. **Distances on WPS.** For clustering applications, we assume $\mathbb{F} = \mathbb{C}$. The *distance* between two points $[z], [w] \in \mathbb{P}_{\mathbf{q}}$ is defined as

$$d([z],[w]) = \inf_{\lambda,\mu \in \mathbb{C}^*} \|(\lambda^{q_0} z_0, \lambda^{q_1} z_1, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \mu^{q_1} w_1, \ldots, \mu^{q_n} w_n)\|, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{C}^{n+1}$. This distance quantifies the minimal Euclidean separation between representatives of $[z]$ and $[w]$ under weighted scalings, respecting the quotient structure of $\mathbb{P}_{\mathbf{q}}$.

We first establish that this distance is well-defined and finite.

Lemma 1. *For any $[z], [w] \in \mathbb{P}_{\mathbf{q}}$, the distance function*

$$d([z],[w]) = \inf_{\lambda,\mu \in \mathbb{C}^*} \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\| \tag{4}$$

is well-defined and finite.

Proof. Define the function

$$f(\lambda, \mu) = \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\|. \tag{5}$$

We must demonstrate that the infimum of $f(\lambda, \mu)$ over $\lambda, \mu \in \mathbb{C}^*$ is finite and independent of the choice of representatives for $[z]$ and $[w]$.

Since $\|\cdot\|$ is the Euclidean norm, $f(\lambda, \mu) \geq 0$. As $[z], [w] \in \mathbb{P}_{\mathbf{q}}$, the vectors $z = (z_0, \ldots, z_n)$, $w = (w_0, \ldots, w_n) \in \mathbb{C}^{n+1} \setminus \{0\}$. Normalize representatives using a weighted norm:

$$\|z\|_a = \left(\sum_{k=0}^{n} q_k \, |z_k|^2 \right)^{1/2}, \quad \|w\|_a = \left(\sum_{k=0}^{n} q_k \, |w_k|^2 \right)^{1/2}, \tag{6}$$

and set $z' = z/\|z\|_a$, $w' = w/\|w\|_a$, so $\sum_{k=0}^{n} q_k \, |z'_k|^2 = 1$, $\sum_{k=0}^{n} q_k \, |w'_k|^2 = 1$. The distance becomes:

$$d([z],[w]) = \inf_{\lambda,\mu \in \mathbb{C}^*} \|(\lambda^{q_0} z'_0, \ldots, \lambda^{q_n} z'_n) - (\mu^{q_0} w'_0, \ldots, \mu^{q_n} w'_n)\|. \tag{7}$$

Evaluate at $\lambda = 1$, $\mu = 1$:

$$f(1,1) = \|(z'_0 - w'_0, \ldots, z'_n - w'_n)\| = \left(\sum_{k=0}^{n} |z'_k - w'_k|^2 \right)^{1/2}. \tag{8}$$

Since $z'_k, w'_k \in \mathbb{C}$, each $|z'_k - w'_k|^2$ is finite, so $f(1,1) < \infty$, providing an upper bound for the infimum.

Consider the behavior as $|\lambda| \to \infty$, μ fixed:

$$f(\lambda,\mu)^2 \approx \sum_{k=0}^{n} |\lambda^{q_k} z'_k|^2 = \sum_{k=0}^{n} |\lambda|^{2q_k} |z'_k|^2 . \tag{9}$$

Since $z' \neq 0$, some $z'_k \neq 0$, and $|\lambda|^{2q_k} \to \infty$, so $f(\lambda,\mu) \to \infty$. Similarly, if $|\mu| \to \infty$, terms $|\mu|^{2q_k} |w'_k|^2$ dominate. If $|\lambda| \to 0$:

$$f(\lambda,\mu) \approx \|(\mu^{q_0} w'_0, \ldots, \mu^{q_n} w'_n)\| = |\mu| \left(\sum_{k=0}^{n} |w'_k|^2 \right)^{1/2}, \tag{10}$$

which is finite for fixed μ.

The function satisfies $f(t\lambda, t\mu) = |t|\, f(\lambda,\mu)$ for $t \in \mathbb{C}^*$. Set $\mu = e^{i\phi}$, so $|\mu| = 1$, and let $\lambda = re^{i\theta}$. Then:

$$f(r,\theta,\phi) = \left(\sum_{k=0}^{n} \left| r^{q_k} e^{iq_k\theta} z'_k - e^{iq_k\phi} w'_k \right|^2 \right)^{1/2} . \tag{11}$$

The infimum over $\theta, \phi \in [0, 2\pi)$ is over a compact set, and f is continuous. As $r \to \infty$, $f \to \infty$; as $r \to 0$, $f \to \left(\sum_{k=0}^{n} |w'_k|^2 \right)^{1/2}$. Thus, the infimum over r, θ, ϕ is finite.

The infimum is invariant under the choice of representatives, as scaling z by λ'^{q_k} adjusts λ to $\lambda\lambda'$. Hence, $d([z],[w])$ is well-defined and finite. □

Lemma 2. *The function $d([z],[w])$ on $\mathbb{P}_{\mathbf{q}}$, defined as*

$$d([z],[w]) = \inf_{\lambda,\mu \in \mathbb{C}^*} \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\|, \tag{12}$$

satisfies:

1. $d([z],[w]) \geq 0$,
2. $d([z],[w]) = d([w],[z])$,
3. $d([z],[w]) = 0$ *if and only if* $[z] = [w]$.

Proof. Define $f(\lambda,\mu) = \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\|$. Since $\|\cdot\|$ is the Euclidean norm, $f(\lambda,\mu) \geq 0$. Thus, $d([z],[w]) = \inf_{\lambda,\mu \in \mathbb{C}^*} f(\lambda,\mu) \geq 0$, establishing $d([z],[w]) \geq 0$.

To prove $d([z],[w]) = d([w],[z])$, compute

$$d([w],[z]) = \inf_{\lambda',\mu' \in \mathbb{C}^*} \|(\lambda'^{q_0} w_0, \ldots, \lambda'^{q_n} w_n) - (\mu'^{q_0} z_0, \ldots, \mu'^{q_n} z_n)\|. \tag{13}$$

Set $\lambda' = \mu$, $\mu' = \lambda$, yielding

$$\|(\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n) - (\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n)\| = \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\|, \tag{14}$$

as $\|v - u\| = \|u - v\|$. Hence, the infima are equal, so $d([w],[z]) = d([z],[w])$.

To show $d([z],[w]) = 0$ if and only if $[z] = [w]$, first assume $[z] = [w]$, so there exists $\alpha \in \mathbb{C}^*$ such that $w_k = \alpha^{q_k} z_k$ for all k. Choose $\lambda = \alpha$, $\mu = 1$, giving

$$\lambda^{q_k} z_k - \mu^{q_k} w_k = \alpha^{q_k} z_k - \alpha^{q_k} z_k = 0. \tag{15}$$

Thus, $f(\alpha, 1) = 0$, so $d([z],[w]) = 0$. Conversely, if $d([z],[w]) = 0$, then $\inf_{\lambda,\mu} f(\lambda,\mu) = 0$. There exist sequences $\lambda_n, \mu_n \in \mathbb{C}^*$ such that

$$\|(\lambda_n^{q_0} z_0 - \mu_n^{q_0} w_0, \ldots, \lambda_n^{q_n} z_n - \mu_n^{q_n} w_n)\| \to 0. \tag{16}$$

Hence, $\lambda_n^{q_k} z_k - \mu_n^{q_k} w_k \to 0$ for each k. For some k with $z_k \neq 0$,

$$w_k = \lim_{n\to\infty} \left(\frac{\lambda_n}{\mu_n}\right)^{q_k} z_k. \tag{17}$$

Set $\alpha_n = \lambda_n/\mu_n$. For all k, j with $z_k, z_j \neq 0$, we have $w_k/z_k = \alpha_n^{q_k}$, $w_j/z_j = \alpha_n^{q_j}$, implying $w_k/z_k = (w_j/z_j)^{q_k/q_j}$. Thus, $[z] = [w]$. If $z_k = 0$, then $w_k = 0$, consistent with the equivalence. □

This distance is particularly suitable for clustering in weighted projective spaces because it respects the equivalence relation defined by the weights. Specifically, it is invariant under the weighted scaling action:

$$d([z],[w]) = d([\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n], [\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n]) \tag{18}$$

for any $\lambda, \mu \in \mathbb{C}^*$. This ensures that the clustering is based on the intrinsic geometry of the space, rather than on specific choices of representatives for the points. In many applications, such as image analysis or genomic data, the data points naturally reside in a weighted projective space due to inherent symmetries or scaling properties. By employing a distance that accounts for these properties, our clustering algorithm can effectively group points that are similar in a geometrically meaningful way. This distance serves as a valid measure of dissimilarity for clustering purposes, enabling algorithms such as hierarchical clustering to partition the data effectively.

Remark 3. *Computing the exact value of $d([z],[w])$ involves solving an optimization problem over λ and μ, which can be computationally intensive. In practice, we approximate this infimum by sampling a finite set of λ and μ values or by employing numerical optimization methods to find sufficiently close approximations.*

Working directly in the weighted projective space allows us to leverage the inherent geometric structure of the data, which can lead to more efficient and accurate clustering compared to traditional methods that might require projecting the data into a different space. By preserving the weighted scaling equivalences, our approach can capture symmetries and invariances that are crucial in applications such as computer vision and genomic data analysis. Furthermore, as demonstrated in [1] and [8], this direct approach can offer computational advantages, particularly in high-dimensional or heterogeneous data settings.

2.4. **Rational points.** For arithmetic applications, we consider the subset $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ of points with rational coordinates $z_i \in \mathbb{Q}$, where the equivalence relation uses scalings $\lambda \in \mathbb{Q}^*$. Points in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ are normalized using the weighted greatest common divisor to facilitate arithmetic analysis. For a point $[z] = [z_0 : z_1 : \cdots : z_n] \in \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, we select a representative $[x_0 : x_1 : \cdots : x_n]$ with coordinates $x_i \in \mathbb{Z}$ such that the weighted greatest common divisor $\mathrm{wgcd}(x_0, x_1, \ldots, x_n) = 1$. The wgcd is defined as the largest positive integer d for which there exists a $\lambda \in \mathbb{Q}^*$ satisfying $\lambda^{q_i} x_i/d \in \mathbb{Z}$ for all $i = 0, 1, \ldots, n$. This normalization ensures a canonical representative, unique up to scaling by roots of unity in $\mathbb{Q}^*$.

The *weighted height* of a point $[z] \in \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, using its normalized representative $[x_0 : x_1 : \cdots : x_n]$, is defined as

$$h_w([z]) = \max_{i=0,\ldots,n} \left(|x_i|^{1/q_i}\right). \tag{19}$$

This height is invariant under the weighted scaling action, as scaling $(x_0, \ldots, x_n)$ by $\lambda \in \mathbb{Q}^*$ yields coordinates $\lambda^{q_i} x_i$, and $|\lambda^{q_i} x_i|^{1/q_i} = |\lambda|\,|x_i|^{1/q_i}$, preserving the maximum up to a factor that cancels in the equivalence class. The weighted height quantifies the arithmetic complexity of rational points, enabling their ordering in databases or the study of Diophantine properties, such as in moduli spaces of algebraic curves as explored in [1].

For rational points, we define a *rational distance* between $[z], [w] \in \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ as

$$d_{\mathbb{Q}}([z],[w]) = \inf_{\lambda,\mu\in\mathbb{Q}^*} \left(\sum_{i=0}^{n} |\lambda^{q_i} x_i - \mu^{q_i} y_i|^2 \right)^{1/2}, \tag{20}$$

where $[x_0 : x_1 : \cdots : x_n]$ and $[y_0 : y_1 : \cdots : y_n]$ are normalized representatives with $x_i, y_i \in \mathbb{Z}$ and $\mathrm{wgcd}(x_0, x_1, \ldots, x_n) = \mathrm{wgcd}(y_0, y_1, \ldots, y_n) = 1$. The use of 'infimum' reflects that, over the discrete set $\mathbb{Q}^*$, a minimum may not be attained due to the density of rational scalings in $\mathbb{R}^*$, but the infimum captures the limit of achievable distances, ensuring a well-defined measure. This distance extends the geometric clustering framework to rational points, respecting the weighted scaling action over $\mathbb{Q}^*$.

Lemma 3. *The function $d_{\mathbb{Q}}([z],[w])$ on $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, defined as*

$$d_{\mathbb{Q}}([z],[w]) = \inf_{\lambda,\mu\in\mathbb{Q}^*} \left(\sum_{i=0}^{n} |\lambda^{q_i} x_i - \mu^{q_i} y_i|^2 \right)^{1/2}, \tag{21}$$

where $[x_0 : x_1 : \cdots : x_n]$ and $[y_0 : y_1 : \cdots : y_n]$ are normalized representatives with $\mathrm{wgcd}(x_0, x_1, \ldots, x_n) = \mathrm{wgcd}(y_0, y_1, \ldots, y_n) = 1$, satisfies:

1. $d_{\mathbb{Q}}([z],[w]) \geq 0$,
2. $d_{\mathbb{Q}}([z],[w]) = d_{\mathbb{Q}}([w],[z])$,
3. $d_{\mathbb{Q}}([z],[w]) = 0$ *if and only if* $[z] = [w]$.

Proof. Define $g(\lambda,\mu) = \left(\sum_{i=0}^{n} |\lambda^{q_i} x_i - \mu^{q_i} y_i|^2\right)^{1/2}$. Since the sum of squared differences is non-negative, $g(\lambda,\mu) \geq 0$, so $d_{\mathbb{Q}}([z],[w]) = \inf_{\lambda,\mu\in\mathbb{Q}^*} g(\lambda,\mu) \geq 0$, establishing $d_{\mathbb{Q}}([z],[w]) \geq 0$.

To prove $d_{\mathbb{Q}}([z],[w]) = d_{\mathbb{Q}}([w],[z])$, compute

$$d_{\mathbb{Q}}([w],[z]) = \inf_{\lambda',\mu'\in\mathbb{Q}^*} \left(\sum_{i=0}^{n} |\lambda'^{q_i} y_i - \mu'^{q_i} x_i|^2 \right)^{1/2}. \tag{22}$$

Set $\lambda' = \mu$, $\mu' = \lambda$, yielding

$$\sum_{i=0}^{n} |\mu^{q_i} y_i - \lambda^{q_i} x_i|^2 = \sum_{i=0}^{n} |\lambda^{q_i} x_i - \mu^{q_i} y_i|^2, \tag{23}$$

as $|a-b|^2 = |b-a|^2$. Thus, the infima are equal, so $d_{\mathbb{Q}}([w],[z]) = d_{\mathbb{Q}}([z],[w])$.

To show $d_{\mathbb{Q}}([z],[w]) = 0$ if and only if $[z] = [w]$, first assume $[z] = [w]$, so there exists $\alpha \in \mathbb{Q}^*$ such that $y_i = \alpha^{q_i} x_i$ for all i. Choose $\lambda = \alpha$, $\mu = 1$, giving

$$\lambda^{q_i} x_i - \mu^{q_i} y_i = \alpha^{q_i} x_i - \alpha^{q_i} x_i = 0. \tag{24}$$

Thus, $g(\alpha, 1) = 0$, so $d_{\mathbb{Q}}([z],[w]) = 0$. Conversely, if $d_{\mathbb{Q}}([z],[w]) = 0$, then $\inf_{\lambda,\mu} g(\lambda,\mu) = 0$. There exists a sequence $(\lambda_n, \mu_n) \in (\mathbb{Q}^*)^2$ such that $g(\lambda_n, \mu_n) \to 0$, implying $\lambda_n^{q_i} x_i - \mu_n^{q_i} y_i \to 0$ for each i. For some i with $x_i \neq 0$, $y_i = \lim_{n\to\infty} (\lambda_n/\mu_n)^{q_i} x_i$.

Set $\alpha_n = \lambda_n/\mu_n$. For all i,j with $x_i, x_j \neq 0$, $y_i/x_i = \alpha_n^{q_i}$, $y_j/x_j = \alpha_n^{q_j}$, so $y_i/x_i = (y_j/x_j)^{q_i/q_j}$, implying $[z] = [w]$. If $x_i = 0$, then $y_i = 0$, consistent with the equivalence. □

2.5. **The triangle inequality.** The distance functions $d([z],[w])$ and $d_{\mathbb{Q}}([z],[w])$ are effective dissimilarity measures for clustering, but their behavior with respect to the triangle inequality merits careful consideration. For

$$d([z],[w]) = \inf_{\lambda,\mu\in\mathbb{C}^*} \|(\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n) - (\mu^{q_0} w_0, \ldots, \mu^{q_n} w_n)\|, \tag{25}$$

the infimum over complex scalings $\lambda, \mu \in \mathbb{C}^*$ introduces a continuum of choices, optimized independently for each pair of points. Similarly, for

$$d_{\mathbb{Q}}([z],[w]) = \inf_{\lambda,\mu\in\mathbb{Q}^*} \left(\sum_{i=0}^{n} |\lambda^{q_i} x_i - \mu^{q_i} y_i|^2 \right)^{1/2}, \tag{26}$$

the infimum over rational scalings $\lambda, \mu \in \mathbb{Q}^*$ is also pair-specific. This independent optimization can lead to configurations where the triangle inequality $d([u],[w]) \leq d([u],[v]) + d([v],[w])$ or $d_{\mathbb{Q}}([u],[w]) \leq d_{\mathbb{Q}}([u],[v]) + d_{\mathbb{Q}}([v],[w])$ does not hold, as the scalings that minimize distances for different pairs may not align additively.

To illustrate this for $d([z],[w])$, consider the weighted projective space $\mathbb{P}_{(2,1)}$ with weights $(2,1)$. Select points $[u] = [1:0]$, $[v] = [1:\epsilon]$, and $[w] = [0:1]$ for small $\epsilon > 0$. Compute the distances approximately:

For $d([u],[v])$, evaluate

$$d([u],[v]) = \inf_{\lambda,\mu\in\mathbb{C}^*} \sqrt{|\lambda^2 - \mu|^2 + |\mu\epsilon|^2}. \tag{27}$$

As $\epsilon \to 0$, the second term $|\mu\epsilon|^2$ dominates for small μ, suggesting $d([u],[v]) \approx \epsilon$ with appropriate λ, μ.

For $d([v],[w])$,

$$d([v],[w]) = \inf_{\lambda,\mu\in\mathbb{C}^*} \sqrt{|\lambda|^2 + |\lambda\epsilon - \mu|^2}. \tag{28}$$

Choosing $\lambda \approx \mu/\epsilon$, the second term minimizes to near zero, yielding $d([v],[w]) \approx 1 - \epsilon$.

For $d([u],[w])$,

$$d([u],[w]) = \inf_{\lambda,\mu\in\mathbb{C}^*} \sqrt{|\lambda^2|^2 + |\mu|^2} = \sqrt{1+1} = \sqrt{2}. \tag{29}$$

For small ϵ, we have $d([u],[v]) + d([v],[w]) \approx \epsilon + (1-\epsilon) = 1$, while $d([u],[w]) = \sqrt{2} \approx 1.414$. Thus, $\sqrt{2} > 1$, violating the triangle inequality. A similar construction applies to $d_{\mathbb{Q}}([z],[w])$, where rational scalings yield analogous violations due to independent optimization.

This violation arises because the optimal λ, μ for each pair are chosen independently, preventing the additive alignment required for the triangle inequality. Consequently, both $d([z],[w])$ and $d_{\mathbb{Q}}([z],[w])$ function as dissimilarity measures rather than true metrics, yet remain effective for clustering applications, such as hierarchical clustering, where the triangle inequality is not essential. The weighted height and distances provide a dual perspective: the height $h_w([z])$ orders points by arithmetic complexity, while the distances $d([z],[w])$ and $d_{\mathbb{Q}}([z],[w])$ group points

geometrically, enhancing data analysis in contexts like moduli spaces where both geometric and arithmetic structures are significant, as demonstrated in [1].

3. Finsler Metric on Weighted Projective Spaces

To define a distance on the weighted projective space $\mathbb{P}_{\mathbf{q}}$, we introduce a Finsler metric that induces a true metric, offering a theoretical framework for potential clustering applications. The weighted projective space $\mathbb{P}_{\mathbf{q}}$, as a quotient of $\mathbb{C}^{n+1} \setminus \{0\}$ under a weighted scaling action, requires careful consideration of curves and their tangent vectors to define a Finsler metric. We begin by detailing the process of lifting curves from $\mathbb{P}_{\mathbf{q}}$ to $\mathbb{C}^{n+1} \setminus \{0\}$, which leads to the definition of the tangent vector $\dot{\gamma}(t)$, essential for the Finsler distance.

3.1. **Curves and Lifting in Weighted Projective Spaces.** The weighted projective space $\mathbb{P}_{\mathbf{q}}$ is defined as the quotient of $\mathbb{C}^{n+1} \setminus \{0\}$ under the equivalence relation $(z_0, z_1, \ldots, z_n) \sim (\lambda^{q_0} z_0, \lambda^{q_1} z_1, \ldots, \lambda^{q_n} z_n)$ for $\lambda \in \mathbb{C}^*$, where $q_0, q_1, \ldots, q_n$ are positive integers called weights, denoted by $\mathbf{q} = (q_0, q_1, \ldots, q_n)$. A point $[z] \in \mathbb{P}_{\mathbf{q}}$ is an equivalence class $[z_0 : z_1 : \cdots : z_n]$, represented by a vector $z = (z_0, z_1, \ldots, z_n) \in \mathbb{C}^{n+1} \setminus \{0\}$. To define a Finsler metric, we consider smooth curves $\gamma : [0, 1] \to \mathbb{P}_{\mathbf{q}}$, which connect points $[z], [w] \in \mathbb{P}_{\mathbf{q}}$ and whose tangent vectors are used to measure distances in the Finsler geometry framework, as described in [5].

A curve $\gamma : [0, 1] \to \mathbb{P}_{\mathbf{q}}$ is smooth if, in local coordinates on $\mathbb{P}_{\mathbf{q}}$, its component functions are smooth (i.e., infinitely differentiable). Since $\mathbb{P}_{\mathbf{q}}$ is a complex manifold (or orbifold for non-coprime weights), smoothness implies that $\gamma(t)$ varies continuously and differentiably in the quotient space. However, $\mathbb{P}_{\mathbf{q}}$ is defined as a quotient, so to work with $\gamma(t)$, we must lift it to a curve in the covering space $\mathbb{C}^{n+1} \setminus \{0\}$, where differentiation is straightforward. The lifting process constructs a representative curve whose derivative defines the tangent vector $\dot{\gamma}(t)$.

Given a smooth curve $\gamma : [0, 1] \to \mathbb{P}_{\mathbf{q}}$ with $\gamma(0) = [z]$ and $\gamma(1) = [w]$, a **lift** of $\gamma(t)$ is a smooth curve

$$z(t) = (z_0(t), z_1(t), \ldots, z_n(t)) \in \mathbb{C}^{n+1} \setminus \{0\} \tag{30}$$

such that $[z(t)] = \gamma(t)$ for all $t \in [0, 1]$. That is, $z(t)$ is a representative of the equivalence class $\gamma(t)$, satisfying $z(t) \neq 0$ and mapping to $\gamma(t)$ under the quotient map $\pi : \mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{P}_{\mathbf{q}}$, defined by $\pi(z) = [z]$. The lift is not unique, as any scaled curve

$$z'(t) = (\lambda(t)^{q_0} z_0(t), \lambda(t)^{q_1} z_1(t), \ldots, \lambda(t)^{q_n} z_n(t)), \tag{31}$$

where $\lambda(t) \in \mathbb{C}^*$ is a smooth function, also satisfies $[z'(t)] = \gamma(t)$.

To construct a lift, consider a local coordinate chart on $\mathbb{P}_{\mathbf{q}}$. For a point $[z] \in \mathbb{P}_{\mathbf{q}}$, suppose $z_k \neq 0$ for some k. In the chart $U_k = \{[z_0 : \cdots : z_n] \in \mathbb{P}_{\mathbf{q}} \mid z_k \neq 0\}$, we can represent $[z]$ by normalizing the k-th coordinate to 1, yielding coordinates

$$(z_0/z_k^{q_0/q_k}, \ldots, z_{k-1}/z_k^{q_{k-1}/q_k}, 1, z_{k+1}/z_k^{q_{k+1}/q_k}, \ldots, z_n/z_k^{q_n/q_k}). \tag{32}$$

If $\gamma(t)$ lies in U_k, we can choose a representative $z(t) = (z_0(t), \ldots, z_n(t))$ with $z_k(t) = 1$, and smoothness of $\gamma(t)$ ensures the other coordinates $z_i(t)/z_k(t)^{q_i/q_k}$ are smooth functions of t. For a general curve $\gamma(t)$, which may exit one chart, we cover $[0, 1]$ with finitely many intervals where $\gamma(t)$ lies in charts U_{k_i}, and construct $z(t)$ piecewise, ensuring smoothness by adjusting scalings $\lambda(t) \in \mathbb{C}^*$ to glue the pieces

across chart transitions. Since $\mathbb{P}_{\mathbf{q}}$ is a smooth manifold (or orbifold), such a smooth lift exists, as the quotient map π is a submersion [4].

The tangent vector $\dot{\gamma}(t)$ is defined via the lift $z(t)$. The derivative of the lifted curve is

$$\dot{z}(t) = \left(\frac{d}{dt} z_0(t), \frac{d}{dt} z_1(t), \ldots, \frac{d}{dt} z_n(t) \right) \in \mathbb{C}^{n+1}, \tag{33}$$

where each $\dot{z}_k(t) = \frac{d}{dt} z_k(t) \in \mathbb{C}$ is the derivative of the coordinate function $z_k(t)$. This vector $\dot{z}(t)$ lies in the tangent space

$$T_{z(t)}(\mathbb{C}^{n+1} \setminus \{0\}) \cong \mathbb{C}^{n+1}.$$

In the quotient space $\mathbb{P}_{\mathbf{q}}$, the tangent space $T_{[\gamma(t)]}\mathbb{P}_{\mathbf{q}}$ at $[\gamma(t)] = [z(t)]$ is the quotient of $T_{z(t)}(\mathbb{C}^{n+1} \setminus \{0\})$ by the tangent vectors of the scaling action's orbits. The Finsler norm $F([z], v)$, defined below, is invariant under this action, so we compute

$$F(\gamma(t), \dot{\gamma}(t)) = F([z(t)], \dot{z}(t)),$$

where $\dot{\gamma}(t)$ is represented by $\dot{z}(t)$ in the quotient tangent space.

To formalize $\dot{\gamma}(t)$, consider the differential of the quotient map

$$\pi : \mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{P}_{\mathbf{q}}.$$

For a point $z(t) \in \mathbb{C}^{n+1} \setminus \{0\}$, the tangent vector $\dot{z}(t)$ is mapped to $\dot{\gamma}(t) \in T_{[\gamma(t)]}\mathbb{P}_{\mathbf{q}}$ via

$$d\pi_{z(t)} : T_{z(t)}(\mathbb{C}^{n+1} \setminus \{0\}) \to T_{[\gamma(t)]}\mathbb{P}_{\mathbf{q}}.$$

The scaling action $z \mapsto (\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n)$ generates an orbit through $z(t)$, and vectors tangent to this orbit, such as $(\alpha^{q_0} z_0(t), \ldots, \alpha^{q_n} z_n(t))$, are quotiented out. If we choose a different lift

$$z'(t) = (\lambda(t)^{q_0} z_0(t), \ldots, \lambda(t)^{q_n} z_n(t)),$$

the derivative is

$$\dot{z}'(t) = \left(q_0 \lambda(t)^{q_0 - 1} \dot{\lambda}(t) z_0(t) + \lambda(t)^{q_0} \dot{z}_0(t), \ldots, q_n \lambda(t)^{q_n - 1} \dot{\lambda}(t) z_n(t) + \lambda(t)^{q_n} \dot{z}_n(t) \right). \tag{34}$$

The Finsler norm $F([z], v)$ is designed to be invariant under scaling, ensuring

$$F([z'(t)], \dot{z}'(t)) = F([z(t)], \dot{z}(t)),$$

so $\dot{\gamma}(t)$ is well-defined as the equivalence class of $\dot{z}(t)$ in $T_{[\gamma(t)]}\mathbb{P}_{\mathbf{q}}$. This lifting process, rooted in the quotient structure of $\mathbb{P}_{\mathbf{q}}$ as described in [4], allows us to define the Finsler distance using tangent vectors derived from lifted curves, as detailed in [5].

For a point $[z] \in \mathbb{P}_{\mathbf{q}}$ with representative

$$z = (z_0, z_1, \ldots, z_n) \in \mathbb{C}^{n+1} \setminus \{0\},$$

and a tangent vector $v = (v_0, v_1, \ldots, v_n) \in \mathbb{C}^{n+1}$, define the **Finsler norm**

$$F([z], v) = \frac{\left(\sum_{k=0}^{n} q_k \left| v_k \right|^2 \right)^{1/2}}{\left(\sum_{k=0}^{n} q_k \left| z_k \right|^2 \right)^{1/2}} \cdot \frac{\left| \sum_{k=0}^{n} q_k z_k \overline{v_k} \right|}{\left(\sum_{k=0}^{n} q_k \left| z_k \right|^2 \right)^{1/2}}. \tag{35}$$

This norm, weighted by the grading $\mathbf{q}$, aligns with the graded vector spaces in [2,3], enabling potential integration with graded neural networks for clustering in $\mathbb{P}_{\mathbf{q}}$.

The induced **Finsler distance** between points $[z], [w] \in \mathbb{P}_{\mathbf{q}}$ is defined as

$$d_F([z],[w]) = \inf_{\gamma} \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt, \tag{36}$$

where $\gamma : [0,1] \to \mathbb{P}_{\mathbf{q}}$ is a smooth curve satisfying $\gamma(0) = [z]$ and $\gamma(1) = [w]$. This construction, inspired by Finsler geometry principles as described in [5], adapts the weighted structure of $\mathbb{P}_{\mathbf{q}}$ as a quotient space, as studied in [4], to provide a true metric, enhancing geometric analysis in clustering contexts.

Lemma 4. *The function $F([z], v)$ defines a Finsler norm on $\mathbb{P}_{\mathbf{q}}$, and the induced distance $d_F([z],[w])$ is a well-defined, finite metric satisfying:*

1. $d_F([z],[w]) \geq 0$,
2. $d_F([z],[w]) = d_F([w],[z])$,
3. $d_F([z],[w]) = 0$ *if and only if* $[z] = [w]$,
4. $d_F([z],[v]) \leq d_F([z],[w]) + d_F([w],[v])$.

Proof. To confirm that $F([z], v)$ is a Finsler norm, we verify its defining properties as outlined in [5]. For positive homogeneity, consider $\lambda \in \mathbb{C}$:

$$\begin{aligned} F([z], \lambda v) &= \frac{\left(\sum_{k=0}^n q_k \left|\lambda v_k\right|^2\right)^{1/2}}{\left(\sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} \cdot \frac{\left|\sum_{k=0}^n q_k z_k \overline{\lambda v_k}\right|}{\left(\sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} \\ &= |\lambda| \, \frac{\left(\sum_{k=0}^n q_k \left|v_k\right|^2\right)^{1/2}}{\left(\sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} \cdot \frac{|\lambda| \left|\sum_{k=0}^n q_k z_k \overline{v_k}\right|}{|\lambda| \left(\sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} = |\lambda| \, F([z], v). \end{aligned} \tag{37}$$

For invariance under the scaling action defining $\mathbb{P}_{\mathbf{q}}$, let $z' = (\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n)$, $\lambda \in \mathbb{C}^*$. Then:

$$\sum_{k=0}^n q_k \left|z'_k\right|^2 = \sum_{k=0}^n q_k \left|\lambda^{q_k} z_k\right|^2 = |\lambda|^2 \sum_{k=0}^n q_k \left|z_k\right|^2, \quad \sum_{k=0}^n q_k z'_k \overline{v_k} = \sum_{k=0}^n q_k \lambda^{q_k} z_k \overline{v_k}, \tag{38}$$

so

$$\begin{aligned} F([z'], v) &= \frac{\left(\sum_{k=0}^n q_k \left|v_k\right|^2\right)^{1/2}}{\left(|\lambda|^2 \sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} \cdot \frac{\left|\sum_{k=0}^n q_k \lambda^{q_k} z_k \overline{v_k}\right|}{\left(|\lambda|^2 \sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} \\ &= \frac{\left(\sum_{k=0}^n q_k \left|v_k\right|^2\right)^{1/2}}{|\lambda| \left(\sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} \cdot \frac{|\lambda| \left|\sum_{k=0}^n q_k z_k \overline{v_k}\right|}{|\lambda| \left(\sum_{k=0}^n q_k \left|z_k\right|^2\right)^{1/2}} = F([z], v). \end{aligned} \tag{39}$$

Non-degeneracy requires $F([z], v) = 0$ only for trivial tangent vectors. If $F([z], v) = 0$, then either

$$\sum_{k=0}^n q_k \left|v_k\right|^2 = 0,$$

implying $v = 0$, or

$$\sum_{k=0}^n q_k z_k \overline{v_k} = 0,$$

corresponding to the orbit direction $v = (\alpha^{q_0} z_0, \ldots, \alpha^{q_n} z_n)$, which is quotiented out in the tangent space of $\mathbb{P}_{\mathbf{q}}$. Thus, $F([z], v) > 0$ for non-trivial tangent vectors in the quotient space. Smoothness of $F([z], v)$ follows from the continuity and differentiability of the norm and inner product terms on the tangent bundle of $\mathbb{P}_{\mathbf{q}}$ minus the zero section, ensuring $F([z], v)$ is a Finsler norm [6].

To show that $d_F([z], [w])$ is well-defined and finite, consider a smooth curve

$$\gamma : [0, 1] \to \mathbb{P}_{\mathbf{q}} \tag{40}$$

with $\gamma(0) = [z]$ and $\gamma(1) = [w]$. As described in the lifting process, we lift $\gamma(t)$ to a smooth curve $z(t) \in \mathbb{C}^{n+1} \setminus \{0\}$ such that $[z(t)] = \gamma(t)$, and the tangent vector $\dot{\gamma}(t)$ is represented by

$$\dot{z}(t) = (\dot{z}_0(t), \ldots, \dot{z}_n(t)).$$

The integrand $F(\gamma(t), \dot{\gamma}(t)) = F([z(t)], \dot{z}(t))$ is continuous, as $z(t)$ and $\dot{z}(t)$ are smooth and F is smooth on the tangent bundle. Since $[0, 1]$ is compact, the integral

$$\int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt \tag{41}$$

exists and is finite. The space $\mathbb{P}_{\mathbf{q}}$ is path-connected, as it is the quotient of $\mathbb{C}^{n+1}\setminus\{0\}$ under a continuous group action, ensuring such curves exist. The infimum over all smooth curves is finite, as $F([z], v) \geq 0$, and a geodesic path, which exists in Finsler manifolds [5], yields a finite length. Invariance of $F([z], v)$ under the scaling action ensures the integral depends only on the equivalence classes $[z]$ and $[w]$, making $d_F([z], [w])$ well-defined.

The metric properties of $d_F([z], [w])$ are established as follows: For non-negativity, since $F([z], v) \geq 0$, the integral

$$\int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt \geq 0,$$

so $d_F([z], [w]) \geq 0$. For symmetry, consider a curve $\gamma(t)$ from $[z]$ to $[w]$. The reversed curve $\gamma(1-t)$ from $[w]$ to $[z]$ has tangent vector $-\dot{\gamma}(1-t)$. Since $F([z], -v) = F([z], v)$ due to the absolute value in the norm, we have

$$F(\gamma(1-t), -\dot{\gamma}(1-t)) = F(\gamma(t), \dot{\gamma}(t)),$$

so the integral along $\gamma(1-t)$ equals that along $\gamma(t)$. Thus, $d_F([z], [w]) = d_F([w], [z])$. To show zero distance implies equality, if $[z] = [w]$, the trivial path $\gamma(t) = [z]$ has $\dot{\gamma}(t) = 0$, giving $F(\gamma(t), 0) = 0$, so $d_F([z], [w]) = 0$. Conversely, if $d_F([z], [w]) = 0$, the infimum of the integral is zero, implying the geodesic length is zero. Since geodesics in Finsler manifolds have positive length for distinct points [6], $[z] = [w]$ must hold. For the triangle inequality, consider geodesics $\gamma_1 : [0, 1] \to \mathbb{P}_{\mathbf{q}}$ from $[z]$ to $[w]$ and $\gamma_2 : [0, 1] \to \mathbb{P}_{\mathbf{q}}$ from $[w]$ to $[v]$. Concatenate them to form $\gamma : [0, 2] \to \mathbb{P}_{\mathbf{q}}$ with $\gamma(t) = \gamma_1(t)$ for $t \in [0, 1]$ and $\gamma(t) = \gamma_2(t - 1)$ for $t \in [1, 2]$. Then,

$$\begin{aligned} \int_0^2 F(\gamma(t), \dot{\gamma}(t))\, dt &= \int_0^1 F(\gamma_1(t), \dot{\gamma}_1(t))\, dt + \int_0^1 F(\gamma_2(t), \dot{\gamma}_2(t))\, dt \\ &\geq d_F([z], [w]) + d_F([w], [v]), \end{aligned} \tag{42}$$

since the infimum over all paths from $[z]$ to $[v]$ is less than or equal to this length. Thus, $d_F([z], [v]) \leq d_F([z], [w]) + d_F([w], [v])$, with scaling invariance ensuring consistency across the quotient structure. Hence, $d_F([z], [w])$ is a well-defined, finite metric. □

Remark 4. *The Finsler distance $d_F([z],[w])$ provides a true metric for theoretical clustering applications in $\mathbb{P}_{\mathbf{q}}$, satisfying all metric axioms. Computing $d_F([z],[w])$ requires numerical optimization to determine geodesics, achievable through variational methods or discrete path approximations, as discussed in* [5].

3.2. **Finsler Metric on Rational Points.** For arithmetic applications, we define a Finsler metric on the rational points $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$. For a point $[z] \in \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ with normalized representative $x = (x_0, x_1, \ldots, x_n) \in \mathbb{Z}^{n+1}$, where $\gcd(x_0, x_1, \ldots, x_n) = 1$, and a tangent vector $v = (v_0, v_1, \ldots, v_n) \in \mathbb{Q}^{n+1}$, define the **rational Finsler norm**

$$F_{\mathbb{Q}}([z], v) = \frac{\left(\sum_{k=0}^{n} q_k \left|v_k\right|^2\right)^{1/2}}{\left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \cdot \frac{\left|\sum_{k=0}^{n} q_k x_k v_k\right|}{\left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}}. \tag{43}$$

The induced **rational Finsler distance** is

$$d_{F,\mathbb{Q}}([z],[w]) = \inf_{\gamma} \int_0^1 F_{\mathbb{Q}}(\gamma(t), \dot{\gamma}(t))\, dt, \tag{44}$$

where $\gamma : [0,1] \to \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ is a piecewise smooth curve with rational coordinates, satisfying $\gamma(0) = [z]$ and $\gamma(1) = [w]$, aligning with the use of infimum in $d_{\mathbb{Q}}([z],[w])$ due to the discrete nature of $\mathbb{Q}^*$. This rational metric supports our program's aim, as outlined in [2], to develop machine learning techniques for graded arithmetic data, leveraging graded neural networks from [3]. This metric extends the Finsler framework to rational points, aligning with the arithmetic structure of $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ as studied in [4], potentially enabling clustering of Diophantine data.

Lemma 5. *The function $F_{\mathbb{Q}}([z], v)$ defines a Finsler norm on $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, and the induced distance $d_{F,\mathbb{Q}}([z],[w])$ is a well-defined, finite metric satisfying:*

1. $d_{F,\mathbb{Q}}([z],[w]) \geq 0$,
2. $d_{F,\mathbb{Q}}([z],[w]) = d_{F,\mathbb{Q}}([w],[z])$,
3. $d_{F,\mathbb{Q}}([z],[w]) = 0$ *if and only if* $[z] = [w]$,
4. $d_{F,\mathbb{Q}}([z],[v]) \leq d_{F,\mathbb{Q}}([z],[w]) + d_{F,\mathbb{Q}}([w],[v])$.

Proof. To verify that $F_{\mathbb{Q}}([z], v)$ is a Finsler norm, we check its properties, following the framework of [5]. For positive homogeneity, let $\lambda \in \mathbb{Q}$:

$$\begin{aligned} F_{\mathbb{Q}}([z], \lambda v) &= \frac{\left(\sum_{k=0}^{n} q_k \left|\lambda v_k\right|^2\right)^{1/2}}{\left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \cdot \frac{\left|\sum_{k=0}^{n} q_k x_k (\lambda v_k)\right|}{\left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \\ &= \frac{|\lambda| \left(\sum_{k=0}^{n} q_k \left|v_k\right|^2\right)^{1/2}}{\left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \cdot \frac{|\lambda| \left|\sum_{k=0}^{n} q_k x_k v_k\right|}{\left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} = |\lambda|\, F_{\mathbb{Q}}([z], v). \end{aligned} \tag{45}$$

For invariance, let $x' = (\lambda^{q_0} x_0, \ldots, \lambda^{q_n} x_n)$, $\lambda \in \mathbb{Q}^*$. Then:

$$\sum_{k=0}^{n} q_k \left|x'_k\right|^2 = \sum_{k=0}^{n} q_k \left|\lambda^{q_k} x_k\right|^2 = |\lambda|^2 \sum_{k=0}^{n} q_k \left|x_k\right|^2, \quad \sum_{k=0}^{n} q_k x'_k v_k = \sum_{k=0}^{n} q_k \lambda^{q_k} x_k v_k, \tag{46}$$

so

$$
\begin{aligned}
F_{\mathbb{Q}}([z'], v) &= \frac{\left(\sum_{k=0}^{n} q_k \left|v_k\right|^2\right)^{1/2}}{\left(\left|\lambda\right|^2 \sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \cdot \frac{\left|\sum_{k=0}^{n} q_k \lambda^{q_k} x_k v_k\right|}{\left(\left|\lambda\right|^2 \sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \\
&= \frac{\left(\sum_{k=0}^{n} q_k \left|v_k\right|^2\right)^{1/2}}{\left|\lambda\right| \left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} \cdot \frac{\left|\lambda\right| \left|\sum_{k=0}^{n} q_k x_k v_k\right|}{\left|\lambda\right| \left(\sum_{k=0}^{n} q_k \left|x_k\right|^2\right)^{1/2}} = F_{\mathbb{Q}}([z], v).
\end{aligned} \tag{47}
$$

Non-degeneracy holds, as $F_{\mathbb{Q}}([z], v) = 0$ implies either $\sum_{k=0}^{n} q_k \left|v_k\right|^2 = 0$, so $v = 0$, or $\sum_{k=0}^{n} q_k x_k v_k = 0$, corresponding to the quotiented orbit direction in the tangent space of $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$. The norm is well-defined and continuous on the rational tangent bundle, satisfying the smoothness requirement for a Finsler norm, adapted to the discrete rational coordinate structure.

To show that $d_{F,\mathbb{Q}}([z], [w])$ is well-defined and finite, consider a piecewise smooth curve $\gamma : [0, 1] \to \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ with $\gamma(0) = [z]$ and $\gamma(1) = [w]$. Lift $\gamma(t)$ to a piecewise smooth curve $x(t) \in \mathbb{Q}^{n+1} \setminus \{0\}$ such that $[x(t)] = \gamma(t)$. The integrand $F_{\mathbb{Q}}(\gamma(t), \dot{\gamma}(t)) = F_{\mathbb{Q}}([x(t)], \dot{x}(t))$ is piecewise continuous, as $x(t)$ and $\dot{x}(t)$ are piecewise smooth and $F_{\mathbb{Q}}$ is continuous on the rational tangent bundle. Since $[0, 1]$ is compact, the integral

$$
\int_0^1 F_{\mathbb{Q}}(\gamma(t), \dot{\gamma}(t))\, dt \tag{48}
$$

exists and is finite. Path-connectedness in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ is ensured by rational scalings, as points in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ can be connected via curves with rational coordinates, parameterized by rational functions. The infimum over all piecewise smooth curves is finite, as $F_{\mathbb{Q}}([z], v) \geq 0$, and a minimizing path (approximating a geodesic) yields a finite length. Invariance of $F_{\mathbb{Q}}([z], v)$ ensures the integral depends only on $[z]$ and $[w]$.

The metric properties are verified as follows. For non-negativity, since $F_{\mathbb{Q}}([z], v) \geq 0$, the integral $\int_0^1 F_{\mathbb{Q}}(\gamma(t), \dot{\gamma}(t))\, dt \geq 0$, so $d_{F,\mathbb{Q}}([z], [w]) \geq 0$. For symmetry, a curve $\gamma(t)$ from $[z]$ to $[w]$ has a reversed curve $\gamma(1 - t)$ from $[w]$ to $[z]$ with tangent vector $-\dot{\gamma}(1 - t)$. Since $F_{\mathbb{Q}}([z], -v) = F_{\mathbb{Q}}([z], v)$, the integral along $\gamma(1 - t)$ equals that along $\gamma(t)$, so $d_{F,\mathbb{Q}}([z], [w]) = d_{F,\mathbb{Q}}([w], [z])$. For zero distance implies equality, if $[z] = [w]$, the trivial path $\gamma(t) = [z]$ has $\dot{\gamma}(t) = 0$, giving $F_{\mathbb{Q}}(\gamma(t), 0) = 0$, so $d_{F,\mathbb{Q}}([z], [w]) = 0$. If $d_{F,\mathbb{Q}}([z], [w]) = 0$, the infimum is zero, implying the minimizing path has zero length, possible only if $[z] = [w]$, as non-trivial paths have positive length in Finsler geometry [6]. For the triangle inequality, concatenate piecewise smooth curves $\gamma_1 : [0, 1] \to \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ from $[z]$ to $[w]$ and $\gamma_2 : [0, 1] \to \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ from $[w]$ to $[v]$ to form a path $\gamma : [0, 2] \to \mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ by $\gamma(t) = \gamma_1(t)$ for $t \in [0, 1]$ and $\gamma(t) = \gamma_2(t - 1)$ for $t \in [1, 2]$. The integral is

$$
\begin{aligned}
\int_0^2 F_{\mathbb{Q}}(\gamma(t), \dot{\gamma}(t))\, dt &= \int_0^1 F_{\mathbb{Q}}(\gamma_1(t), \dot{\gamma}_1(t))\, dt + \int_0^1 F_{\mathbb{Q}}(\gamma_2(t), \dot{\gamma}_2(t))\, dt \\
&\geq d_{F,\mathbb{Q}}([z], [w]) + d_{F,\mathbb{Q}}([w], [v]),
\end{aligned} \tag{49}
$$

so $d_{F,\mathbb{Q}}([z], [v]) \leq d_{F,\mathbb{Q}}([z], [w]) + d_{F,\mathbb{Q}}([w], [v])$. Thus, $d_{F,\mathbb{Q}}([z], [w])$ is a well-defined, finite metric. □

Remark 5. *The rational Finsler distance $d_{F,\mathbb{Q}}([z], [w])$ offers a theoretical metric for clustering rational points in arithmetic applications, such as moduli spaces, by*

respecting the Diophantine structure of $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$. Its computation involves numerical approximation of geodesics over rational paths, feasible via methods described in [5], though constrained by the discrete nature of rational coordinates.

4. Finsler Geodesics in Weighted Projective Spaces

The Finsler distance $d_F([z],[w])$ on the weighted projective space $\mathbb{P}_{\mathbf{q}}$, defined as the infimum of the integral of the Finsler norm $F([z],v)$ over smooth curves, relies on the concept of Finsler geodesics to achieve its metric properties. Similarly, the rational Finsler distance $d_{F,\mathbb{Q}}([z],[w])$ on $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ depends on geodesics adapted to rational coordinates. This section elucidates Finsler geodesics, their definition, properties, and role in the Finsler geometry of $\mathbb{P}_{\mathbf{q}}$ and $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, building on the framework established for the Finsler metric and the lifting of curves to $\mathbb{C}^{n+1}\setminus\{0\}$, as described in [5] and [6].

A Finsler geodesic in a Finsler manifold equipped with a norm $F(x,v)$ on its tangent bundle is a curve that locally minimizes the length functional, defined by the integral of F along the curve. In the context of $\mathbb{P}_{\mathbf{q}}$, the Finsler norm is given in Eq. (35). The Finsler distance between points $[z],[w]\in\mathbb{P}_{\mathbf{q}}$ is given in Eq. (36). A Finsler geodesic is a curve $\gamma(t)$ that achieves this infimum or locally minimizes the integral, representing the shortest path in the Finsler geometry of $\mathbb{P}_{\mathbf{q}}$, as detailed in [5].

To formalize Finsler geodesics, consider a smooth curve $\gamma(t)$ in $\mathbb{P}_{\mathbf{q}}$. The length of $\gamma(t)$ is given by

$$L[\gamma]=\int_0^1 F(\gamma(t),\dot\gamma(t))\,dt, \tag{50}$$

where $\dot\gamma(t)$ is the tangent vector, represented by the derivative $\dot z(t)=(\dot z_0(t),\dots,\dot z_n(t))$ of a lifted curve $z(t)\in\mathbb{C}^{n+1}\setminus\{0\}$. These geodesics, governed by the graded norm $F([z],v)$, support our program's goal, as outlined in [2], to develop machine learning techniques for clustering in graded spaces, potentially enhanced by graded neural networks from [3]. A geodesic $\gamma(t)$ satisfies the Euler-Lagrange equations for the functional $L[\gamma]$, ensuring it is a critical point of the length. In local coordinates on $\mathbb{P}_{\mathbf{q}}$, say in a chart $U_k=\{[z]\in\mathbb{P}_{\mathbf{q}}\mid z_k\neq 0\}$ with coordinates

$$(x_1,\dots,x_{k-1},x_{k+1},\dots,x_n)=\left(\frac{z_0}{z_k^{q_0/q_k}},\dots,\frac{z_{k-1}}{z_k^{q_{k-1}/q_k}},\frac{z_{k+1}}{z_k^{q_{k+1}/q_k}},\dots,\frac{z_n}{z_k^{q_n/q_k}}\right),$$

the geodesic equation takes the form

$$\frac{d}{dt}\left(\frac{\partial F}{\partial\dot x^i}\right)-\frac{\partial F}{\partial x^i}=0,\quad i=1,\dots,n, \tag{51}$$

where $F(x,\dot x)=F(\gamma(t),\dot\gamma(t))$ is the Finsler norm evaluated along the curve [5]. However, the quotient structure of $\mathbb{P}_{\mathbf{q}}$, defined by the scaling action $(z_0,\dots,z_n)\sim(\lambda^{q_0}z_0,\dots,\lambda^{q_n}z_n)$, $\lambda\in\mathbb{C}^*$, complicates direct coordinate computations. The Finsler norm $F([z],v)$ is invariant under this action, allowing us to work with lifted curves in $\mathbb{C}^{n+1}\setminus\{0\}$, where the tangent vector $\dot z(t)$ is adjusted to the quotient tangent space $T_{[\gamma(t)]}\mathbb{P}_{\mathbf{q}}$, as described in the lifting process [4].

The existence of Finsler geodesics in $\mathbb{P}_{\mathbf{q}}$ is ensured by the completeness of the Finsler manifold, a property inherited from the completeness of $\mathbb{C}^{n+1}\setminus\{0\}$ under the quotient action. According to the Hopf-Rinow theorem for Finsler manifolds, any two points in a complete Finsler manifold can be joined by a minimizing geodesic,

whose length equals the Finsler distance [6]. For points $[z], [w] \in \mathbb{P}_{\mathbf{q}}$, there exists a geodesic $\gamma(t)$ such that

$$d_F([z],[w]) = \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt, \tag{52}$$

representing the shortest path in the Finsler geometry. The geodesic's tangent vectors $\dot{\gamma}(t)$ satisfy the geodesic equation, which, in the quotient space, accounts for the weighted scaling invariance of $F([z], v)$. The norm's structure, with terms like $\sum_{k=0}^n q_k |v_k|^2$ and $|\sum_{k=0}^n q_k z_k \overline{v_k}|$, weights the tangent vector's contribution by the geometry of $\mathbb{P}_{\mathbf{q}}$, ensuring geodesics reflect the manifold's weighted structure.

Lemma 6. *For any $[z], [w] \in \mathbb{P}_{\mathbf{q}}$, there exists a Finsler geodesic $\gamma : [0,1] \to \mathbb{P}_{\mathbf{q}}$ with $\gamma(0) = [z]$, $\gamma(1) = [w]$, such that*

$$d_F([z],[w]) = \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt. \tag{53}$$

Proof. The weighted projective space $\mathbb{P}_{\mathbf{q}}$ is a complete Finsler manifold, as it is the quotient of the complete manifold $\mathbb{C}^{n+1} \setminus \{0\}$ under the proper action of $\mathbb{C}^*$, defined by $(z_0, \ldots, z_n) \mapsto (\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n)$. The Finsler norm $F([z], v)$ is smooth on the tangent bundle minus the zero section, satisfying positive homogeneity, non-degeneracy, and invariance, as established previously. By the Hopf-Rinow theorem for Finsler manifolds, as detailed in [6], any two points in a complete Finsler manifold are connected by a minimizing geodesic, and the distance $d_F([z],[w]) = \inf_\gamma \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt$ is achieved by such a geodesic. For $[z], [w] \in \mathbb{P}_{\mathbf{q}}$, path-connectedness ensures the existence of smooth curves $\gamma : [0,1] \to \mathbb{P}_{\mathbf{q}}$ with $\gamma(0) = [z]$ and $\gamma(1) = [w]$. The infimum is finite, as $F([z], v) \geq 0$, and a geodesic $\gamma(t)$, whose lift $z(t) \in \mathbb{C}^{n+1} \setminus \{0\}$ satisfies the Euler-Lagrange equations adjusted for the quotient, achieves the minimum length, equaling $d_F([z],[w])$. □

For the rational Finsler distance $d_{F,\mathbb{Q}}([z],[w])$ on $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, geodesics are piecewise smooth curves with rational coordinates, minimizing the integral of the rational Finsler norm as defined in Eq. (44). The discrete nature of rational points requires piecewise smooth curves, as continuous rational paths may be constrained, but path-connectedness in $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ via rational scalings ensures the existence of such curves. A minimizing geodesic, approximated by rational paths, achieves the infimum $d_{F,\mathbb{Q}}([z],[w])$, satisfying a modified geodesic equation for the rational tangent bundle. The completeness of $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ in the Finsler sense, analogous to $\mathbb{P}_{\mathbf{q}}$, guarantees the existence of such paths [6]. Such rational geodesics, adapted to the graded structure of $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, advance our program's aim to apply machine learning to arithmetic data, as envisioned in [2,3].

Computing Finsler geodesics in $\mathbb{P}_{\mathbf{q}}$ or $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$ is complex due to the non-quadratic nature of $F([z], v)$ and $F_{\mathbb{Q}}([z], v)$, requiring numerical methods such as variational techniques or discrete approximations, as discussed in [5]. In $\mathbb{P}_{\mathbf{q}}$, geodesics determine the shortest paths for the metric $d_F([z],[w])$, enabling theoretical clustering by providing a true distance that respects the weighted geometry. In $\mathbb{P}_{\mathbf{q}}(\mathbb{Q})$, rational geodesics support arithmetic applications, aligning with the Diophantine structure of moduli spaces, as studied in [4]. Together, Finsler geodesics underpin the geometric and arithmetic framework of our Finsler metrics, offering a robust theoretical tool for distance-based analysis in weighted projective spaces.

5. Clustering in Weighted Projective Spaces

This section presents a hierarchical clustering algorithm tailored for the weighted projective space $\mathbb{P}_{\mathbf{q}}$, employing the Finsler metric $d_F([z],[w])$ to define distances between points. The algorithm exploits the intrinsic geometry of $\mathbb{P}_{\mathbf{q}}$, characterized by the weights $\mathbf{q} = (q_0, q_1, \ldots, q_n)$, to partition data into clusters, leveraging the true metric properties of $d_F([z],[w])$, as established previously. While our prior work utilized the dissimilarity measure $d([z],[w])$ [1], the use of $d_F([z],[w])$ offers a rigorous metric framework for clustering.

5.1. **Clustering Algorithm.** The hierarchical clustering algorithm constructs a dendrogram by iteratively merging clusters based on pairwise distances computed using the Finsler metric. Unlike the dissimilarity measure $d([z],[w])$, which does not satisfy the triangle inequality, the Finsler metric $d_F([z],[w])$ is a true metric, enabling compatibility with standard metric-based clustering techniques while respecting the non-Euclidean geometry of $\mathbb{P}_{\mathbf{q}}$, as defined in [4]. The algorithm operates on a dataset $S = \{[z_1],[z_2],\ldots,[z_N]\} \subset \mathbb{P}_{\mathbf{q}}$ of N points, producing a hierarchical structure of clusters.

Formally, let $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ be a partition of S into m clusters, initially $\mathcal{C} = \{\{[z_1]\},\{[z_2]\},\ldots,\{[z_N]\}\}$ with $m = N$. The algorithm iteratively merges pairs of clusters based on a linkage criterion, reducing m until a stopping condition is met (e.g., a fixed number of clusters or a distance threshold). This algorithm, utilizing the graded geometry of $\mathbb{P}_{\mathbf{q}}$, advances our program's objective, as outlined in [2], to develop machine learning techniques for clustering in graded spaces, potentially enhanced by graded neural networks from [3]. The Finsler distance between points $[z],[w] \in \mathbb{P}_{\mathbf{q}}$ is given in Eq. (36).

The linkage criterion defines the distance between clusters $C_i, C_j \in \mathcal{C}$. Common criteria include single linkage, minimizing the smallest distance between points in different clusters, complete linkage, minimizing the largest distance, and average linkage, minimizing the average distance, formally defined as

$$
\begin{aligned}
d_{\text{single}}(C_i, C_j) &= \min_{[z]\in C_i,[w]\in C_j} d_F([z],[w]),\\
d_{\text{complete}}(C_i, C_j) &= \max_{[z]\in C_i,[w]\in C_j} d_F([z],[w]),\\
d_{\text{average}}(C_i, C_j) &= \frac{1}{|C_i||C_j|} \sum_{[z]\in C_i,[w]\in C_j} d_F([z],[w]).
\end{aligned}
\tag{54}
$$

The algorithm proceeds by computing the pairwise distance matrix for S, merging clusters with the smallest linkage distance, updating the partition $\mathcal{C}$, and continuing until a desired number of clusters is reached or a threshold on the linkage distance is met.

Lemma 7. *The hierarchical clustering algorithm with the Finsler metric $d_F([z],[w])$ produces a valid dendrogram, correctly partitioning the dataset $S \subset \mathbb{P}_{\mathbf{q}}$ into a hierarchical structure of clusters.*

Proof. A dendrogram is a binary tree representing a sequence of cluster merges, where each merge combines two clusters into one, reducing the number of clusters from N to 1. Initially, set $\mathcal{C}_0 = \{\{[z_1]\},\{[z_2]\},\ldots,\{[z_N]\}\}$, with each point in its own cluster. At step k, the algorithm identifies clusters $C_i, C_j \in \mathcal{C}_{k-1}$ minimizing the linkage distance $d_{\text{link}}(C_i, C_j)$, where d_{link} is one of $d_{\text{single}}, d_{\text{complete}}$, or d_{average}.

Merge C_i and C_j into a new cluster $C_{ij} = C_i \cup C_j$, forming $\mathcal{C}_k = (\mathcal{C}_{k-1} \setminus \{C_i, C_j\}) \cup \{C_{ij}\}$. This process iterates for $N-1$ steps, resulting in $\mathcal{C}_{N-1} = \{S\}$.

The algorithm's correctness relies on the well-definedness of $d_F([z],[w])$ and the linkage criterion. Since $d_F([z],[w])$ is a metric, satisfying non-negativity, symmetry, zero distance implies equality, and the triangle inequality, the pairwise distance matrix is well-defined with $d_F([z_i],[z_j]) \geq 0$, $d_F([z_i],[z_j]) = d_F([z_j],[z_i])$, and $d_F([z_i],[z_j]) = 0$ if and only if $[z_i] = [z_j]$. Each linkage criterion produces a valid distance between clusters: single linkage ensures connectivity, complete linkage ensures compactness, and average linkage balances intra-cluster distances [7]. At each step, the minimum linkage distance exists, as $\mathcal{C}_{k-1}$ is finite, and merging reduces the number of clusters by one. The process terminates after $N-1$ merges, producing a dendrogram where each node represents a cluster merge, correctly encoding the hierarchical structure of S. □

Lemma 8. *The time complexity for computing the distance matrix is $O(N^2 \cdot T)$, where N is the number of points and T is the time to compute each $d_F([z],[w])$. The hierarchical clustering step has a time complexity of $O(N^2 \log N)$ with efficient implementations, making the overall complexity $O(N^2 \cdot T + N^2 \log N)$.*

Proof. The distance matrix requires computing $d_F([z_i],[z_j])$ for all $\binom{N}{2} = \frac{N(N-1)}{2}$ pairs $[z_i],[z_j] \in S$, which is $O(N^2)$ operations. Each computation of $d_F([z],[w]) = \inf_\gamma \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt$ involves optimizing a geodesic integral, requiring time T, dependent on the numerical method (e.g., variational optimization or discrete approximation). Thus, the total time for the distance matrix is $O(N^2 \cdot T)$.

For the hierarchical clustering step, the algorithm performs $N-1$ merges. At step k, the partition $\mathcal{C}_{k-1}$ has $N-k+1$ clusters. Computing the linkage distance $d_{\text{link}}(C_i, C_j)$ for all pairs $C_i, C_j \in \mathcal{C}_{k-1}$ involves evaluating $d_F([z],[w])$ for points in C_i and C_j. For single linkage, this requires $O(|C_i||C_j|)$ evaluations, but distances are precomputed in the matrix. Using a priority queue to store pairwise linkage distances, initialized with $O(N^2)$ entries, finding the minimum distance at each step takes $O(\log(N-k+1)) = O(\log N)$. Updating the queue after merging C_i and C_j into C_{ij} involves computing $d_{\text{link}}(C_{ij}, C_l)$ for all other clusters $C_l \in \mathcal{C}_k$, taking $O(N-k)$ operations per merge. Over $N-1$ merges, the total clustering time is

$$\begin{aligned} \sum_{k=1}^{N-1} [O(\log N) + O(N-k)] &= O(N \log N) + O\left(\sum_{k=1}^{N-1} (N-k)\right) \\ &= O(N \log N) + O(N^2) = O(N^2 \log N), \end{aligned} \tag{55}$$

using efficient implementations [7]. The overall complexity is $O(N^2 \cdot T) + O(N^2 \log N)$, where T, typically $O(I \cdot n)$ for I iterations in n-dimensional space, dominates for large N. □

5.2. **Preprocessing Steps.** To ensure the consistency and efficiency of clustering in $\mathbb{P}_{\mathbf{q}}$, preprocessing steps are applied to the dataset $S = \{[z_1], \ldots, [z_N]\}$. Normalization mitigates the effects of arbitrary scaling in the quotient space. For geometric clustering using $d_F([z],[w])$, points are normalized such that $\sum_{k=0}^n q_k |z_k|^2 = 1$, ensuring consistency across the quotient action. For each point $[z_i] \in S$, select a representative $z_i = (z_{i,0}, \ldots, z_{i,n}) \in \mathbb{C}^{n+1} \setminus \{0\}$, and scale by $\alpha_i = \left(\sum_{k=0}^n q_k \, |z_{i,k}|^2\right)^{-1/2}$ to satisfy the condition. For arithmetic clustering using $d_{F,\mathbb{Q}}([z],[w])$, rational

points are normalized with $\text{wgcd}(x_0, x_1, \ldots, x_n) = 1$, as detailed in Section 2.2. Normalization is computed in $O(n)$ time per point, totaling $O(N \cdot n)$ for N points.

For high-dimensional data, dimensionality reduction preserves geometric structure while reducing computational cost. Weighted principal component analysis (PCA) constructs a weighted covariance matrix using the inner product $\langle z_i, z_j \rangle = \sum_{k=0}^{n} q_k z_{i,k} \overline{z_{j,k}}$, projecting points onto the top $k < n$ eigenvectors. The covariance matrix computation takes $O(N \cdot n^2)$, and eigenvalue decomposition requires $O(n^3)$, totaling $O(N \cdot n^2 + n^3)$. This reduces subsequent distance computations to $O(k)$ per pair, as points are embedded in a k-dimensional subspace. Alternatively, manifold learning methods, such as Isomap adapted to $d_F([z], [w])$, preserve geodesic distances, requiring $O(N^2 \cdot T)$ for distance matrix computation and additional processing, but are computationally intensive. These preprocessing steps, tailored to the graded structure of $\mathbb{P}_{\mathbf{q}}$, support our program's aim, as outlined in [2, 3], to apply machine learning to non-Euclidean geometric data.

Lemma 9. *Normalization by the weighted norm $\sum_{k=0}^{n} q_k |z_k|^2 = 1$ preserves the Finsler distance $d_F([z], [w])$, ensuring clustering consistency.*

Proof. Let $[z], [w] \in \mathbb{P}_{\mathbf{q}}$ with representatives $z, w \in \mathbb{C}^{n+1} \setminus \{0\}$. Normalize to $z' = z/\|z\|_a$, $w' = w/\|w\|_a$, where $\|z\|_a = \left(\sum_{k=0}^{n} q_k |z_k|^2\right)^{1/2}$, so $\sum_{k=0}^{n} q_k |z'_k|^2 = 1$, and similarly for w'. Since $[z'] = [z]$ and $[w'] = [w]$, we must show $d_F([z], [w]) = d_F([z'], [w'])$. The Finsler distance depends only on equivalence classes, as the Finsler norm $F([z], v)$ is invariant under scaling: for $z'' = (\lambda^{q_0} z_0, \ldots, \lambda^{q_n} z_n)$, $F([z''], v) = F([z], v)$, as proven previously. Thus, a curve $\gamma(t)$ from $[z]$ to $[w]$ with lift $z(t)$ has the same length as a curve with lift $z'(t) = z(t)/\|z(t)\|_a$, since $F([\gamma(t)], \dot{\gamma}(t)) = F([z(t)], \dot{z}(t)) = F([z'(t)], \dot{z}'(t))$ after adjusting for the quotient. The infimum over all curves yields $d_F([z], [w]) = d_F([z'], [w'])$, ensuring normalization does not affect clustering outcomes. □

5.3. **Computational Challenges.** Computing the Finsler distance $d_F([z], [w])$ involves optimizing the geodesic integral, a computationally intensive task due to the non-Euclidean geometry of $\mathbb{P}_{\mathbf{q}}$. The integral

$$d_F([z], [w]) = \inf_{\gamma} \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt \tag{56}$$

requires finding a geodesic $\gamma(t)$, typically via numerical methods, as the Finsler norm $F([z], v)$ is non-quadratic [5]. We address this challenge through discrete path approximation and variational optimization. In discrete path approximation, the curve $\gamma(t)$ is discretized into M segments, approximating the integral by numerical quadrature, such as the trapezoidal rule. For each segment, evaluating $F(\gamma(t_i), \dot{\gamma}(t_i))$ involves computing the norm in $O(n)$ time, totaling $O(M \cdot n)$ per distance, or $O(N^2 \cdot M \cdot n)$ for the distance matrix of N points. Variational optimization employs iterative methods, such as shooting methods or gradient-based solvers, to minimize the energy functional

$$E[\gamma] = \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt, \tag{57}$$

requiring $O(I \cdot n)$ time per distance for I iterations, totaling $O(N^2 \cdot I \cdot n)$. Parallelization distributes the $\binom{N}{2}$ distance calculations across P processors, reducing the time to $O\left(\frac{N^2 \cdot T}{P}\right)$, where $T = O(M \cdot n)$ or $O(I \cdot n)$ depending on the method.

Remark 6. *Computing geodesics for $d_F([z],[w])$ involves optimizing over curves in the quotient space $\mathbb{P}_{\mathbf{q}}$. Lifting curves to $\mathbb{C}^{n+1} \setminus \{0\}$, we minimize the energy functional $\int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt$ using numerical solvers, such as Euler-Lagrange equations or discrete optimization, with convergence ensured by the completeness of $\mathbb{P}_{\mathbf{q}}$ [6]. The non-Riemannian nature of $F([z], v)$ necessitates careful implementation to balance accuracy and efficiency.*

Theorem 1. *The hierarchical clustering algorithm with $d_F([z],[w])$ is stable under small perturbations of the input points in $\mathbb{P}_{\mathbf{q}}$, ensuring consistent dendrogram outputs for nearby datasets.*

Proof. Stability implies that small changes in the input dataset $S = \{[z_1], \ldots, [z_N]\} \subset \mathbb{P}_{\mathbf{q}}$ produce small changes in the dendrogram, measured by a metric on dendrograms, such as the Gromov-Hausdorff distance. Let $S' = \{[z'_1], \ldots, [z'_N]\}$ be a perturbed dataset with $d_F([z_i],[z'_i]) < \epsilon$ for all i. The distance matrix for S has entries $d_{ij} = d_F([z_i],[z_j])$, and for S', entries $d'_{ij} = d_F([z'_i],[z'_j])$. Since d_F is a metric, the triangle inequality gives

(58)
$$\left|d_{ij} - d'_{ij}\right| = \left|d_F([z_i],[z_j]) - d_F([z'_i],[z'_j])\right| \leq d_F([z_i],[z'_i]) + d_F([z_j],[z'_j]) < 2\epsilon.$$

Thus, the distance matrices are close in the sup-norm, with $\sup_{i,j} \left|d_{ij} - d'_{ij}\right| < 2\epsilon$. Hierarchical clustering with linkage criteria (single, complete, or average) is continuous with respect to the sup-norm on distance matrices, as small perturbations in distances result in small changes in merge decisions [7]. Each merge step depends on minimizing $d_{\text{link}}(C_i, C_j)$, and a perturbation of order 2ϵ alters the minimum by at most 2ϵ, preserving the dendrogram's structure up to small shifts in merge heights. Hence, the algorithm produces dendrograms for S and S' that are close, ensuring stability. □

This framework leverages the metric properties of $d_F([z],[w])$ to define a hierarchical structure, ensuring correctness and stability for clustering in $\mathbb{P}_{\mathbf{q}}$.

6. Applications and Examples

This section elucidates the theoretical utility of our hierarchical clustering algorithm using the Finsler metric $d_F([z],[w])$ and its rational counterpart $d_{F,\mathbb{Q}}([z],[w])$ in the weighted projective space $\mathbb{P}_{\mathbf{q}}$. The primary applications explored are the clustering of rational points in the moduli space of genus two curves, represented as $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$, and the analysis of rational functions on the projective line in the context of Arithmetic Dynamics, building on collaborative work with Eslam Badr and Elira Shaska [8]. Additionally, synthetic data experiments and comparisons with traditional methods demonstrate the algorithm's theoretical capabilities.

6.1. Clustering in the Moduli Space of Genus Two Curves.

The moduli space of genus two curves, represented as the weighted projective space $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ with coordinates $(x_0, x_1, x_2, x_3) = (J_2, J_4, J_6, J_{10})$ corresponding to Igusa invariants of degrees 2, 4, 6, and 10, provides a rich setting for applying our clustering algorithm. In prior work [1], a clustering approach using the dissimilarity measure

$d([z],[w])$ identified arithmetic patterns in this space, such as the distribution of fine moduli points and curves with (n,n)-split Jacobians. Here, we theoretically extend this analysis by employing the Finsler metric $d_{F,\mathbb{Q}}([z],[w])$, leveraging its true metric properties to cluster rational points and detect geometric structures.

Rational points $[z]=[x_0:x_1:x_2:x_3]\in\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ are normalized to satisfy $\mathrm{wgcd}(x_0,x_1,x_2,x_3)=1$, ensuring a canonical representative for arithmetic analysis. The weighted height $h_w([z])=\max_{i=0,\dots,3}\left(|x_i|^{1/q_i}\right)$, with weights $q_0=2,q_1=4,q_2=6,q_3=10$, quantifies the arithmetic complexity of these points. We focus on clustering points within the loci $\mathcal{L}_n\subset\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$, which are 2-dimensional hypersurfaces parameterizing genus two curves with (n,n)-split Jacobians for $n=2,3,5$. A dataset of 50,000 rational points per locus is generated using birational parametrizations, such as the (u,v)-parametrization for $\mathcal{L}_2$ described in [1], ensuring points lie on or near these loci. This clustering, leveraging the graded structure of $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$, supports our program's aim, as outlined in [2, 3], to develop machine learning for arithmetic geometry.

The hierarchical clustering algorithm, using single linkage defined as $d_{\text{single}}(C_i,C_j)=\min_{[z]\in C_i,[w]\in C_j}d_{F,\mathbb{Q}}([z],[w])$, groups points by their geometric proximity in $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$. The Finsler distance as in Eq. (44) is computed over piecewise smooth curves $\gamma:[0,1]\to\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$. The algorithm's correctness, established previously, ensures a valid dendrogram, grouping points into clusters that reflect the loci's geometric structure.

Lemma 10. *The hierarchical clustering algorithm with single linkage and $d_{F,\mathbb{Q}}([z],[w])$ identifies clusters in $\mathcal{L}_n\subset\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ corresponding to points with similar Igusa invariants, preserving arithmetic patterns.*

Proof. Consider a dataset $S=\{[z_1],\dots,[z_N]\}\subset\mathcal{L}_n\subset\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$, where $\mathcal{L}_n$ is the locus of curves with (n,n)-split Jacobians. The single linkage criterion merges clusters C_i,C_j minimizing $d_{\text{single}}(C_i,C_j)=\min_{[z]\in C_i,[w]\in C_j}d_{F,\mathbb{Q}}([z],[w])$. Since $d_{F,\mathbb{Q}}([z],[w])$ is a metric, it satisfies non-negativity, symmetry, zero distance implies equality, and the triangle inequality, ensuring well-defined distances. Points on $\mathcal{L}_n$ are parametrized by birational coordinates (e.g., (u,v) for $\mathcal{L}_2$), and $d_{F,\mathbb{Q}}([z],[w])$ measures their geometric proximity via geodesics in the rational Finsler geometry.

The algorithm constructs a dendrogram by merging clusters with minimal $d_{F,\mathbb{Q}}$, grouping points $[z],[w]$ with small distances, corresponding to similar Igusa invariants (J_2,J_4,J_6,J_{10}). The triangle inequality ensures that clusters form connected components in the metric space, reflecting the 2-dimensional structure of $\mathcal{L}_n$. For points with weighted height $h_w([z])$, the metric $d_{F,\mathbb{Q}}([z],[w])$ respects the scaling action, preserving arithmetic patterns (e.g., points with $h_w([z])\le 3$). The dendrogram's correctness, proven previously, guarantees that clusters align with the loci's geometry, identifying sets of curves with analogous splitting properties, validated by the parametrization's coverage of $\mathcal{L}_n$ [1]. □

6.2. **Synthetic Data.** Synthetic data experiments in $\mathbb{P}_{(2,4,6,10)}$ test the theoretical clustering algorithm with $d_F([z],[w])$. Points are sampled with rational coordinates (x_0,x_1,x_2,x_3), weights 2, 4, 6, 10, normalized to $\mathrm{wgcd}(x_0,x_1,x_2,x_3)=1$, and constrained near the loci $\mathcal{L}_n$ using parametrizations from [1]. The hierarchical clustering algorithm, employing average linkage defined as $d_{\text{average}}(C_i,C_j)=\frac{1}{|C_i||C_j|}\sum_{[z]\in C_i,[w]\in C_j}d_F([z],[w])$, groups points based on their Finsler distances.

The resulting dendrogram is visualized by projecting points onto absolute invariants, such as $t_1 = J_2^5/J_{10}$, which normalize the weighted degrees. These experiments, utilizing the graded Finsler metric $d_F([z],[w])$, advance our program's objective, as outlined in [2,3], to develop machine learning for graded geometric data. The clusters theoretically correspond to sets of points with similar invariant structures, reflecting the automorphism groups of the underlying curves, as the metric $d_F([z],[w])$ captures the non-Euclidean geometry of $\mathbb{P}_{(2,4,6,10)}$.

Lemma 11. *The hierarchical clustering algorithm with average linkage and $d_F([z],[w])$ produces clusters in synthetic data that align with the geometric structure of $\mathcal{L}_n \subset \mathbb{P}_{(2,4,6,10)}$.*

Proof. Let $S = \{[z_1],\dots,[z_N]\} \subset \mathbb{P}_{(2,4,6,10)}$ be a synthetic dataset sampled near $\mathcal{L}_n$, with points normalized to $\text{wgcd}(x_0,x_1,x_2,x_3) = 1$. The average linkage criterion merges clusters minimizing $d_{\text{average}}(C_i,C_j)$, computed using $d_F([z],[w])$. Since $d_F([z],[w])$ is a metric, the distance matrix is symmetric, non-negative, and satisfies the triangle inequality, ensuring robust clustering. Points near $\mathcal{L}_n$ are generated via parametrizations, placing them on or close to the 2-dimensional hypersurface. The Finsler norm $F([z],v)$ weights distances by the projective geometry, so $d_F([z],[w])$ is small for points with similar Igusa invariants. The algorithm's correctness ensures a dendrogram where merges reflect geometric proximity, grouping points into clusters that align with $\mathcal{L}_n$'s structure, as the average linkage criterion balances intra-cluster distances [7]. Projection onto invariants like $t_1 = J_2^5/J_{10}$ preserves this structure, confirming theoretical cluster alignment. □

6.3. **Applications.** The clustering algorithm with $d_F([z],[w])$ and $d_{F,\mathbb{Q}}([z],[w])$ is theoretically applicable to several domains, with primary focus on arithmetic geometry. In the analysis of curve automorphisms, clustering rational points in $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ using $d_{F,\mathbb{Q}}([z],[w])$ identifies genus two curves with extra automorphisms, corresponding to singular points on loci like $\mathcal{L}_2$. These points, associated with distinct geometric properties, are grouped by their Finsler distances, aiding in the classification of curves by automorphism groups. In cryptographic curve enumeration, the algorithm theoretically groups moduli points to estimate the distribution of curves with (n,n)-split Jacobians over number fields or finite fields, informing the design of secure isogeny-based cryptosystems by quantifying vulnerable curve classes.

In Arithmetic Dynamics, the algorithm is applied to study rational functions on the projective line $\mathbb{P}^1$, building on collaborative work with Eslam Badr and Elira Shaska [8]. Rational functions of degree n, represented as points in a weighted projective space (e.g., $\mathbb{P}_{(1,1,\dots,1,2n)}(\mathbb{Q})$ for coefficients of numerator and denominator polynomials), are clustered using $d_{F,\mathbb{Q}}([z],[w])$. This groups functions with similar dynamical properties, such as periodic point structures or multiplier spectra, facilitating the analysis of arithmetic and geometric invariants in dynamical systems. The Finsler metric's ability to capture projective symmetries ensures clusters reflect intrinsic dynamical behaviors, supporting theoretical studies of iteration and conjugacy classes.

Theorem 2. *Clustering with $d_{F,\mathbb{Q}}([z],[w])$ in $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ preserves the geometric and arithmetic structure of the moduli space, grouping points by their Igusa invariants and dynamical properties in Arithmetic Dynamics.*

Proof. Consider a dataset $S \subset \mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ of rational points on or near $\mathcal{L}_n$. The Finsler metric $d_{F,\mathbb{Q}}([z],[w])$ is invariant under the weighted scaling action $(x_0,x_1,x_2,x_3) \sim (\lambda^2 x_0, \lambda^4 x_1, \lambda^6 x_2, \lambda^{10} x_3)$, $\lambda \in \mathbb{Q}^*$, ensuring distances depend only on equivalence classes. For points $[z],[w] \in \mathcal{L}_n$, the geodesic integral reflects their proximity in the 2-dimensional hypersurface, weighted by the Finsler norm $F_{\mathbb{Q}}([z],v)$. The hierarchical clustering algorithm, with linkage criteria like single or average, groups points minimizing $d_{F,\mathbb{Q}}([z],[w])$, producing clusters that align with the loci's geometry, as shown in the correctness lemma. For Igusa invariants, small $d_{F,\mathbb{Q}}([z],[w])$ implies similar (J_2,J_4,J_6,J_{10}), preserving arithmetic properties like weighted height. In Arithmetic Dynamics, rational functions on $\mathbb{P}^1$, represented in a weighted projective space, are clustered by dynamical invariants (e.g., multipliers), as $d_{F,\mathbb{Q}}([z],[w])$ captures projective similarities [8]. The metric's triangle inequality ensures stable, geometrically meaningful clusters, theoretically grouping points by both moduli and dynamical structures. □

6.4. **Comparison with Traditional Methods.** Traditional clustering methods, such as k-means applied to Euclidean embeddings of absolute invariants (e.g., $t_1 = J_2^5/J_{10}$), fail to capture the weighted scaling symmetries of $\mathbb{P}_{(2,4,6,10)}$. In [1], k-means on such invariants produced inaccurate groupings, as Euclidean distances distort the non-Euclidean geometry of the moduli space. Our algorithm, using $d_F([z],[w])$ and $d_{F,\mathbb{Q}}([z],[w])$, operates directly in $\mathbb{P}_{(2,4,6,10)}$, preserving the projective structure. The Finsler metric's geodesic distances ensure clusters reflect the true geometry, theoretically improving grouping accuracy for loci like $\mathcal{L}_n$, as supported by the stability theorem.

7. Conclusion and Future Work

This paper presents a novel hierarchical clustering algorithm for weighted projective spaces, employing the Finsler metric $d_F([z],[w])$ and its rational counterpart $d_{F,\mathbb{Q}}([z],[w])$. This work is a key component of a broader program to develop fully operational machine learning (ML) and artificial intelligence (AI) techniques for graded spaces, as outlined in [2]. By leveraging the grading structure of weighted projective spaces $\mathbb{P}_{\mathbf{q}}$, defined by weights $\mathbf{q} = (q_0, q_1, \ldots, q_n)$, our algorithm advances this vision, building on the graded vector spaces and neural networks introduced in [2,3] to enable robust data analysis in non-Euclidean manifolds.

The proposed metrics, defined rigorously to capture the weighted scaling symmetries of $\mathbb{P}_{\mathbf{q}}$, offer a true metric framework that satisfies non-negativity, symmetry, zero distance implies equality, and the triangle inequality, enabling robust clustering in non-Euclidean manifolds. The algorithm's theoretical correctness and stability ensures its applicability to datasets with projective structures, such as the moduli space of genus two curves and rational functions on the projective line in Arithmetic Dynamics. While our prior work with the dissimilarity measure $d([z],[w])$ demonstrated empirical success [1], the Finsler metric's use remains a theoretical exploration, pending experimental validation to confirm its practical efficacy.

The primary application in the moduli space $\mathbb{P}_{(2,4,6,10)}(\mathbb{Q})$ clusters rational points representing genus two curves, grouping them by Igusa invariants and detecting curves with (n,n)-split Jacobians, as theoretically validated for loci $\mathcal{L}_n$. This supports arithmetic geometry studies and isogeny-based cryptography by classifying curve properties. In Arithmetic Dynamics [8] applies the algorithm to rational functions on $\mathbb{P}^1$, clustering points in weighted projective spaces to analyze dynamical

invariants, enhancing understanding of iteration and conjugacy classes. Further applications in Galois Theory [10,11], in reduction theory of binary forms [12], and [13] illustrate the power of machine learning over weighted projective spaces. Synthetic data experiments further demonstrate the algorithm's ability to capture geometric structures, while comparisons with Euclidean methods highlight the Finsler metric's theoretical superiority in preserving projective symmetries.

Future work will advance this framework within our program to develop ML and AI techniques for graded spaces, exploring several theoretical and computational directions. Efficient geodesic computation methods for $d_F([z],[w])$ and $d_{F,\mathbb{Q}}([z],[w])$ are critical to reduce complexity, potentially through advanced variational techniques or approximate solvers, building on [5]. Alternative clustering algorithms, such as spectral clustering adapted to the Finsler metric, could leverage its geometric properties to enhance partitioning accuracy. Extending the approach to fields like robotics, where non-Euclidean geometries model configuration spaces, or quantum computing, where projective spaces represent quantum states, offers promising theoretical avenues. A priority direction is the development of graded neural networks, inspired by [2, 3], where features are assigned weights mirroring $\mathbf{q} = (q_0, q_1, \ldots, q_n)$. By defining activation functions, layer transformations, and loss functions using $d_F([z],[w])$, GNNs could preprocess or classify data in $\mathbb{P}_{\mathbf{q}}$, addressing challenges in convergence and quotient-space architectures to realize fully operational ML/AI systems for graded spaces. These efforts aim to establish a robust theoretical and practical foundation for non-Euclidean data analysis, with transformative impacts in arithmetic geometry, cryptography, and beyond.

References

[1] Elira Shaska and Tanush Shaska, *Machine learning for moduli space of genus two curves and an application to isogeny-based cryptography*, J. Algebraic Combin. **61** (2025), no. 2, Paper No. 23, 35. MR4870337

[2] T. Shaska, *Artificial neural networks on graded vector spaces* (2024), available at 2407.19031.

[3] ———, *Graded Neural Networks* (2025), available at 2502.17751.

[4] Igor Dolgachev, *Weighted projective varieties*, Group actions and vector fields, 1982, pp. 34–71.

[5] David Bao, Shiing-Shen Chern, and Zhongmin Shen, *An introduction to riemann-finsler geometry*, Graduate Texts in Mathematics, vol. 200, Springer, New York, 2000.

[6] Zhongmin Shen, *Lectures on finsler geometry*, World Scientific, Singapore, 2012.

[7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, 2nd ed., Springer, New York, 2009.

[8] Eslam Badr, Elira Shaska, and Tony Shaska, *Rational functions on the projective line from a computational viewpoint* (2025), available at 2503.10835.

[9] Sajad Salami and Tony Shaska, *Local and global heights on weighted projective varieties*, Houston J. Math. **49** (2023), no. 3, 603–636. MR4845203

[10] Elira Shaska and Tony Shaska, *Galois groups of polynomials and neurosymbolic networks* (2025), available at 2501.12978.

[11] ———, *Neuro-symbolic learning for galois groups: Unveiling probabilistic trends in polynomials* (2025), available at 2502.20844.

[12] Ilias Kotsireas and Tony Shaska, *A neurosymbolic framework for geometric reduction of binary forms* (2025), available at 2501.15404.

[13] Elira Shaska, Jorge Mello, Sajad Salami, and Tony Shaska, *Rational points and zeta functions of Humbert surfaces with square discriminant* (2025), available at 2504.19268.

Department of Mathematics and Statistics,, Oakland University, Rochester, MI, 48326

Email address: shaska@oakland.edu