

REAL MODULI SPACES FOR NETWORK QUIVERS

MARCO ARMENTA ARMENTA, JORGE ESQUIVEL ARAYA, AND ALEXANDER SCHMITT

ABSTRACT. Armenta and Jodoin introduced moduli spaces of quiver representations as a tool for the theoretical investigation of neural networks, and Armenta, Brüstle, Hassoun, and Reineke established many geometric properties of those moduli spaces in a generality which goes beyond the standard setting of neural networks. In this paper, we will first review the algebro-geometric background. We have tried to be elementary in order to make the techniques accessible to a broader audience. As originally suggested by Armenta and Jodoin, we derive a (generalization of a) result of Meng et al.. In addition, we prove some interesting results about the equations for the moduli spaces and their real points.

INTRODUCTION

A(n) (artificial) neural network is a structure which is inspired by brains. It features “neurons” and “synapses” which form a directed graph or quiver. It is subdivided into an input layer, an output layer, and hidden layers in between. One fixes an activation function for each hidden neuron (or vertex) and a weight for each synapse (or arrow). These data determine the network function, i.e., the rule according to which information is passed from the input layer to the output layer. The collection of weights determines a so-called thin representation of the network quiver. It was observed in [2] that the network function is invariant under certain rescalings and, therefore, can be factorized over a certain algebraic variety that is known in representation theory and algebraic geometry as the moduli space of (thin) representations of the (doubly) framed network quiver.¹ The idea of [2] is to investigate this moduli space in order to get insights into the workings of the network. For this reason, moduli spaces of - not necessarily thin - representations were investigated in [1]. That paper applied quite advanced techniques of representation theory and algebraic geometry.

In this note, we will focus only on the case of thin representations which seems to be, at the moment, the most relevant one for applications. We have provided an extensive introduction to the algebraic framework of group actions and quiver representations. We have tried to keep the discussion elementary, though this has not always been possible. Theorem 2.3.2 is a special case of a fundamental result about moduli spaces of framed quiver representations. We have provided an easy proof in the situation at hand. The generators of the invariant ring which appear in the statement of the theorem determine the moduli space within a certain ambient space $\mathbb{C}^s$, s being the number of generators. In fact, the equations cutting out the moduli space from that ambient space are obtained from relations among the generators. We recall an algorithm based on Gröbner bases for finding a set of relations which implies all other relations and is in a certain sense minimal with this property. Then, we provide an elementary way for determining a set of relations which implies all other relations, but which might be bigger than the one you find algorithmically.

2020 *Mathematics Subject Classification.* 16G20, 14D22, 14P05, 68T07.

¹The framings are determined by the input and the output layer where no rescalings are taking place.

The next important result is the slightly technical Proposition 2.4.2. Its proof nicely reflects the inductive structure of a network quiver. Stable, semistable, and polystable collections of weights are crucial notions for investigating the moduli spaces. They come from the general theory. Stable and semistable collections have already been investigated in [1] and [3] (see Theorem 3.1.1). We rephrase the result in Theorem 3.1.2 and show that collections consisting entirely of non-zero weights are stable (Corollary 3.1.3). The existence of stable collections and Proposition 2.4.2 lead to an algebraic proof and a generalization of a result in [16] (see Theorem 3.1.7). This is a first indication that the moduli space does indeed contain interesting information about the corresponding neural network. Another interesting finding is the characterization of polystable collections of weights in terms of their support (Proposition 3.2.1).

In the discussion, so far, it has been implicitly assumed that we are working over the complex numbers. This means that we are admitting complex weights, as well. However, this seems to be irrelevant for applications. To this end, we prove in Section 3.3 that the moduli space over the real numbers just consists of the real points of the complex moduli space (Theorem 3.3.4). This puts that object into a clear mathematical context, namely the realm of real algebraic geometry and should help in further investigations. The main ingredients of the proof of Theorem 3.3.4 are Propositions 2.4.2 and 3.2.1.

We again emphasize that the proofs use only elementary techniques and are of a more combinatorial flavor.

Acknowledgment. Jorge Esquivel Araya was funded by the Math+ project EF 1-16 *Quiver representations in big data and machine learning*.²

1. BACKGROUND FROM ALGEBRAIC GEOMETRY

In this section, we will collect some necessary concepts and results from algebraic geometry. These concern, in particular, group actions and quotients. We have tried to keep it as elementary as possible, so that a reader with a general mathematical background which includes polynomial rings in several variables will get a first impression. Eventually, it might be necessary to get immersed more deeply into the subject. For the basic set-up of algebraic geometry, Chapter I of [11] or [17] or Chapter AG of [6] will be sufficient. The latter book also contains information on group actions and quotients. The books [8] and [19] provide gentle introductions to geometric invariant theory, that is, the theory of group actions in algebraic geometry, their invariants, and quotients.

1.1. Regular functions. Let $n \geq 1$ and $\mathbb{C}^n$ the set of n -tuples of complex numbers. So, an element of $\mathbb{C}^n$ has the shape $(a_1, \dots, a_n)$ with $a_i \in \mathbb{C}$, $i = 1, \dots, n$. The functions that we admit on $\mathbb{C}^n$ in algebraic geometry are called *regular functions* and are given by the polynomials in the variables $x_1, \dots, x_n$. Here, a polynomial $f(x_1, \dots, x_n) \in \mathbb{C}[x_1, \dots, x_n]$ is identified with the function

$$\begin{aligned} f: \mathbb{C}^n &\longrightarrow \mathbb{C} \\ (a_1, \dots, a_n) &\longmapsto f(a_1, \dots, a_n). \end{aligned}$$

A polynomial $f \in \mathbb{C}[x_1, \dots, x_n]$ is said to be a *monomial*, if there exists a tuple $\underline{\mu} = (\mu_1, \dots, \mu_n) \in \mathbb{N}^{\times n}$, such that

$$f = \underline{x}^{\underline{\mu}} := x_1^{\mu_1} \cdots x_n^{\mu_n}.$$

²<https://mathplus.de/research-2/emerging-fields/ef1-extracting-dynamical-laws-from-complex-data/ef1-16/>

The number $d := \mu_1 + \cdots + \mu_n$ is the *degree* of the monomial. The monomials

$$\underline{x}^\mu, \quad \underline{\mu} \in \mathbb{N}^{\times n},$$

form a vector space basis for $\mathbb{C}[x_1, \dots, x_n]$. For $d \in \mathbb{N}$, let

$$\mathbb{C}[x_1, \dots, x_n]_d$$

be the linear subspace of $\mathbb{C}[x_1, \dots, x_n]$ that is spanned by the monomials of degree d . An element of $\mathbb{C}[x_1, \dots, x_n]_d$ is called a *homogeneous polynomial* of *degree* d . Note that the zero polynomial is homogeneous of any degree. Since the monomials of degree d form a basis of the polynomial ring, we have the direct sum decomposition

$$\mathbb{C}[x_1, \dots, x_n] = \bigoplus_{d=0}^{\infty} \mathbb{C}[x_1, \dots, x_n]_d.$$

Remark 1.1.1. There is a coordinate free description. Let V be a complex vector space of dimension n and $V^\vee := \text{Hom}_{\mathbb{C}}(V, \mathbb{C})$ its dual vector space. If we choose a basis $e_1, \dots, e_n$ for V , then the dual basis $e^1, \dots, e^n$ for $V^\vee$ is defined by the condition

$$e^i(e_j) = \delta_{ij}, \quad i, j = 1, \dots, n.$$

Clearly,

$$(1.1.1.1) \quad \begin{aligned} \mathbb{C}[x_1, \dots, x_n]_1 &\longrightarrow V^\vee \\ x_i &\longmapsto e^i, \quad i = 1, \dots, n, \end{aligned}$$

is an isomorphism of complex vector spaces. Now, for any $d \in \mathbb{N}$, one may associate with $V^\vee$ its d -fold symmetric power $\text{Sym}^d(V^\vee)$ (see [15], Chapter XVI, §8). For example, $\text{Sym}^0(V^\vee) = \mathbb{C}$ and $\text{Sym}^1(V^\vee) = V^\vee$. The isomorphism (1.1.1.1) induces an isomorphism

$$\mathbb{C}[x_1, \dots, x_n]_d \longrightarrow \text{Sym}^d(V^\vee), \quad d \geq 2.$$

Finally, the direct sum

$$\text{Sym}^\star(V^\vee) := \bigoplus_{d=0}^{\infty} \text{Sym}^d(V^\vee)$$

can be easily endowed with the structure of a $\mathbb{C}$ -algebra (loc. cit.). It is called the *symmetric algebra* of $V^\vee$. The isomorphism (1.1.1.1) yields an isomorphism

$$\mathbb{C}[x_1, \dots, x_n] \longrightarrow \text{Sym}^\star(V^\vee)$$

of $\mathbb{C}$ -algebras.

1.2. Tori. Let $n \geq 1$ and set

$$T^n := (\mathbb{C}^\star)^{\times n}.$$

We refer to T^n as the n -dimensional torus. Now, T^n may be viewed as the open subset

$$D(x_1 \cdots x_n) := \{ (a_1, \dots, a_n) \in \mathbb{C}^n \mid a_1 \cdots a_n \neq 0 \}.$$

So, the polynomial (function defined by) $x_1 \cdots x_n$ becomes invertible on T^n . But, then, we also have the functions

$$x_i^{-1} := \frac{1}{x_i} = \frac{x_1 \cdots x_{i-1} \cdot x_{i+1} \cdots x_n}{x_1 \cdots x_n}, \quad i = 1, \dots, n.$$

For this reason, we consider

$$\mathbb{C}[x_1, \dots, x_n, x_1^{-1}, \dots, x_n^{-1}]$$

to be the algebra of regular functions on T^n . An element $f \in \mathbb{C}[x_1, \dots, x_n, x_1^{-1}, \dots, x_n^{-1}]$ is a *monomial*, if there exists a tuple $\underline{\mu} = (\mu_1, \dots, \mu_n) \in \mathbb{Z}^n$, such that

$$f = \underline{x}^{\underline{\mu}} := x_1^{\mu_1} \cdot \dots \cdot x_n^{\mu_n}.$$

The monomials

$$\underline{x}^{\underline{\mu}}, \quad \underline{\mu} \in \mathbb{Z}^n,$$

form a vector space basis for $\mathbb{C}[x_1, \dots, x_n, x_1^{-1}, \dots, x_n^{-1}]$.

More generally, for $m \geq 1$, a map

$$\begin{aligned} F: T^n &\longrightarrow \mathbb{C}^m \\ (a_1, \dots, a_n) &\longmapsto (f_1(a_1, \dots, a_n), \dots, f_m(a_1, \dots, a_n)) \end{aligned}$$

is a *morphism*, if f_i is a regular function on T^n , $i = 1, \dots, m$. Similarly, we define a morphism

$$G: T^n \longrightarrow T^m.$$

Now, the multiplication

$$\begin{aligned} \cdot: T^n \times T^n &\longrightarrow T^n \\ ((a_1, \dots, a_n), (b_1, \dots, b_n)) &\longmapsto (a_1 \cdot b_1, \dots, a_n \cdot b_n) \end{aligned}$$

endows T^n with the structure of an abelian group. Using the identification $T^n \times T^n = T^{2n}$, we see that the multiplication is a morphism between tori. Likewise, the inversion

$$\begin{aligned} \cdot: T^n &\longrightarrow T^n \\ (a_1, \dots, a_n) &\longmapsto (a_1^{-1}, \dots, a_n^{-1}) \end{aligned}$$

is a morphism of tori. Since the group law and the inversion are morphisms, we say that T^n is an *algebraic group*. (We refer to [6], Chapter I, §1, for the general definition of an algebraic group and to [6], Chapter III, §1, for specific results on algebraic tori.)

For $m, n \geq 1$, a map $\alpha: T^n \longrightarrow T^m$ is a *homomorphism of tori*, if it is both a group homomorphism and a morphism between tori.

A homomorphism

$$\chi: T^n \longrightarrow T^1 = \mathbb{C}^\star$$

of tori is said to be a *character* of T^n . It is easy to see that a character is defined by a monomial, so that there is a unique tuple $\underline{\mu}(\chi) \in \mathbb{Z}^n$ with

$$\chi = \underline{x}^{\underline{\mu}(\chi)}.$$

Conversely, any monomial defines a character of T^n . Let $X(T^n)$ be the set of characters of T^n . Using the multiplication on $\mathbb{C}^\star$, we equip $X(T^n)$ with the structure of an abelian group. Now,

$$\begin{aligned} (1.2.0.1) \quad X(T^n) &\longrightarrow \mathbb{Z}^n \\ \chi &\longmapsto \underline{\mu}(\chi) \end{aligned}$$

is an isomorphism of groups. We call $M := X(T^n)$ the *character lattice* of T^n .

A homomorphism

$$\lambda: T^1 = \mathbb{C}^\star \longrightarrow T^n$$

is a *one parameter subgroup* or *cocharacter* of T^n . Let $\Lambda(T^n)$ be the set of all cocharacters of T^n . For each cocharacter $\lambda: \mathbb{C}^\star \rightarrow T^n$, there is a unique tuple $\underline{\mu}(\lambda) = (\mu_1, \dots, \mu_n) \in \mathbb{Z}^n$, such that

$$\forall z \in \mathbb{C}^\star : \quad \lambda(z) = (z^{\mu_1}, \dots, z^{\mu_n}).$$

The group structure on T^n induces on $\Lambda(T^n)$ the structure of an abelian group, and

$$\begin{aligned} \Lambda(T^n) &\longrightarrow \mathbb{Z}^n \\ \lambda &\longmapsto \underline{\mu}(\lambda) \end{aligned}$$

is an isomorphism of groups. We call $N := \Lambda(T^n)$ the *cocharacter lattice* of T^n .

Note that

$$\begin{aligned} \langle \cdot, \cdot \rangle : \Lambda(T^n) \times X(T^n) &\longrightarrow \mathbb{Z} \\ (\lambda, \chi) &\longmapsto \mu(\chi \circ \lambda) \end{aligned}$$

is a perfect pairing, i.e.,

$$\begin{aligned} \Lambda(T^n) &\longrightarrow X(T^n)^\vee = \text{Hom}_{\mathbb{Z}}(X(T^n), \mathbb{Z}) \\ \lambda &\longmapsto \langle \lambda, \cdot \rangle \end{aligned}$$

is an isomorphism of groups. Also,

$$\begin{aligned} (1.2.0.2) \quad \text{ev} : \Lambda(T^n) \otimes_{\mathbb{Z}} \mathbb{C}^\star &\longrightarrow T^n \\ \lambda \otimes z &\longmapsto \lambda(z) \end{aligned}$$

is an isomorphism.

Now, let $m, n \geq 1$ and

$$\alpha : T^n \longrightarrow T^m$$

a homomorphism of tori. Then,

$$\begin{aligned} X(\alpha) : X(T^m) &\longrightarrow X(T^n) \\ \chi &\longmapsto \chi \circ \alpha \end{aligned}$$

is a homomorphism of groups. Conversely, for a group homomorphism

$$\varphi : X(T^m) \longrightarrow X(T^n),$$

there is the dual homomorphism

$$\begin{aligned} \varphi^\vee : \Lambda(T^n) = X(T^n)^\vee &\longrightarrow \Lambda(T^m) = X(T^m)^\vee \\ (\beta : X(T^n) \longrightarrow \mathbb{Z}) &\longmapsto \beta \circ \varphi, \end{aligned}$$

and one checks that

$$\begin{aligned} \varphi^\vee \otimes \text{id}_{\mathbb{C}^\star} : \Lambda(T^n) \otimes_{\mathbb{Z}} \mathbb{C}^\star &\longrightarrow \Lambda(T^m) \otimes_{\mathbb{Z}} \mathbb{C}^\star \\ \lambda \otimes z &\longmapsto \varphi^\vee(\lambda) \otimes z \end{aligned}$$

is a homomorphism of tori, keeping in mind our previous identifications. The two assignments are inverse to each other, so that

$$\text{Hom}_{\text{tori}}(T^n, T^m) \cong \text{Hom}_{\mathbb{Z}}(X(T^m), X(T^n)).$$

(As mentioned before, [6], Chapter III, §1, contains this and more results on algebraic tori.)

1.3. Group actions. For $n \geq 1$, we have the general linear group $\mathrm{GL}_n(\mathbb{C})$. By Leibniz's rule, the determinant

$$\det: \mathrm{Mat}_n(\mathbb{C}) \longrightarrow \mathbb{C}$$

is a polynomial function, i.e., a regular function. Furthermore, $\mathrm{GL}_n(\mathbb{C})$ is the open subset

$$D(\det) := \{ m \in \mathrm{Mat}_n(\mathbb{C}) \mid \det(m) \neq 0 \}.$$

So, we view $\mathbb{C}[x_{ij}, i, j = 1, \dots, n, \det^{-1}]$ as the algebra of regular functions on $\mathrm{GL}_n(\mathbb{C})$. As for tori, one sees that the group law and inversion are morphisms.³ For this reason, we say that $\mathrm{GL}_n(\mathbb{C})$ is an algebraic group.

For $m, n \geq 1$, a map

$$\varrho: T^n \longrightarrow \mathrm{GL}_m(\mathbb{C})$$

is a *homomorphism of algebraic groups* or a *representation* of T^n on $\mathbb{C}^m$, if it is both a group homomorphism and a morphism, i.e., writing

$$\varrho: g \longmapsto (f_{ij}(g))_{i,j=1,\dots,m},$$

the function f_{ij} is regular on T^n , $i, j = 1, \dots, m$.

Remark 1.3.1. Often, it is more convenient to work without coordinates. For a finite dimensional complex vector space V , let $\mathrm{GL}(V)$ be the group of linear automorphisms of V . We use the usual identification⁴ of $\mathrm{GL}(\mathbb{C}^m)$ with the group $\mathrm{GL}_m(\mathbb{C})$ of invertible $(m \times m)$ -matrices, $m \geq 1$. The choice of a basis B for V may be viewed as an isomorphism $\varphi_B: \mathbb{C}^m \longrightarrow V$, $m := \dim_{\mathbb{C}}(V)$. The latter induces the isomorphism

$$\begin{aligned} \psi_B: \mathrm{GL}(V) &\longrightarrow \mathrm{GL}_m(\mathbb{C}) \\ L &\longmapsto \varphi_B^{-1} \circ L \circ \varphi_B. \end{aligned}$$

Given $m, n \geq 1$ and an m -dimensional complex vector space V , a *representation* of T^n on V is a map

$$\varrho: T^n \longrightarrow \mathrm{GL}(V)$$

for which there exists a basis B , such that $\psi_B \circ \varrho$ is a representation of T^n on $\mathbb{C}^m$. In that case, $\psi_C \circ \varrho$ will be a representation of T^n on $\mathbb{C}^m$, for every basis C of V .

For a representation

$$\varrho: T^n \longrightarrow \mathrm{GL}(V),$$

we get the action

$$\begin{aligned} \alpha: T^n \times V &\longrightarrow V \\ (g, v) &\longmapsto \varrho(g)(v). \end{aligned}$$

This is also a morphism of algebraic varieties.

A basic fact is that ϱ is diagonalizable. For this, let

$$\Phi_n := \{ g \in T^n \mid \exists l \geq 1 : g^l = e = (1, \dots, 1) \}$$

be the subgroup of elements of finite order. For $g \in \Phi_n$, the image $\varrho(g)$ is an element of finite order in $\mathrm{GL}(V)$. The theory of the Jordan normal form ([15], XIV, §2) shows that an element of finite order in $\mathrm{GL}(V)$ is diagonalizable, so that there exists a basis B for V , such that $\psi_B(\varrho(g))$ is a diagonal matrix. Moreover, any set of diagonalizable elements of $\mathrm{GL}(V)$, such that any two elements of the set commute with each other, may be

³For the group law, it is a bit more difficult to figure out what this should mean, for the inversion, one uses Cramer's rule ([15], Chapter XIII, Proposition 4.16).

⁴Since we are working with row vectors, the identification is $\mathrm{GL}_m(\mathbb{C}) \longrightarrow \mathrm{GL}(\mathbb{C}^m)$, $m \longmapsto (v \longmapsto v \cdot m^t)$.

simultaneously diagonalized (compare [15], Chapter XIV, Exercise 13 (d)). In particular, we find a basis B for V , such that $\psi_B(\varrho(g))$ is a diagonal matrix, for every $g \in \Phi_n$. Observe that

- The subgroup $D_m \subset \mathrm{GL}_m(\mathbb{C})$ is a closed subset.⁵
- The subset $\Phi_n \subset T^n$ is dense.⁶

For reasons of continuity, $\psi_B(\varrho(T^n)) \subset D_m$.

Remark 1.3.2. We may identify D_m with the m -dimensional torus T^m . So, at the end, we have arrived at a homomorphism $T^n \longrightarrow T^m$ of tori.

For a character $\chi: T^n \longrightarrow \mathbb{C}^\star$ of T^n , its *eigenspace* is given as

$$V_\chi := \{ v \in V \mid \forall g \in T^n : \alpha(g, v) = \varrho(g)(v) = \chi(g) \cdot v \}.$$

As usual, eigenvectors for different characters are $\mathbb{C}$ -linearly independent ([15], Chapter XIV, Theorem 3.3). So, the above observation tells us that there are $s \geq 1$ and distinct characters $\chi_1, \dots, \chi_s$ of T^n , such that

- $V_{\chi_i} \neq 0, i = 1, \dots, s,$
- $V = V_{\chi_1} \oplus \dots \oplus V_{\chi_s}.$

Given $m, n \geq 1$, an m -dimensional complex vector space V , a representation $\varrho: T^n \longrightarrow \mathrm{GL}(V)$, and the corresponding action $\alpha: T^n \times V \longrightarrow V$, we are interested in forming the quotient as an algebraic variety. This is a non-trivial problem. Look at

$$\begin{aligned} \varrho: \mathbb{C}^\star &\longrightarrow \mathrm{GL}_m(\mathbb{C}) \\ z &\longmapsto z \cdot E_m. \end{aligned}$$

It corresponds to the action

$$\begin{aligned} (1.3.2.1) \quad \alpha: \mathbb{C}^\star \times \mathbb{C}^m &\longrightarrow \mathbb{C}^m \\ (z, v) &\longmapsto z \cdot v. \end{aligned}$$

For every line $\ell \subset \mathbb{C}^m$ through the origin, $\ell \setminus \{0\}$ is an orbit. In addition, $\{0\}$ is an orbit. Let $\mathcal{Q}$ be the set of orbits endowed with the quotient topology and $\pi: \mathbb{C}^m \longrightarrow \mathcal{Q}$ the (continuous) quotient map. Then, for a line ℓ through the origin, $\pi^{-1}(\{\ell \setminus \{0\}\}) = \ell \setminus \{0\}$ is not a closed subset, so that $\{\ell \setminus \{0\}\}$ is not a closed subset of $\mathcal{Q}$. This means that not all points of $\mathcal{Q}$ are closed, so that we cannot endow the topological space $\mathcal{Q}$ with the structure of an algebraic variety.

So, in order to figure out what a suitable quotient could be, we try to determine its ring of regular functions. For this, let $\mathbb{C}[V]$ be the ring of regular functions on V that we considered in Section 1.1, and define the action⁷

$$\begin{aligned} \alpha^\star: T^n \times \mathbb{C}[V] &\longrightarrow \mathbb{C}[V] \\ (g, f) &\longmapsto \left(g \star f: v \longmapsto f(\alpha(g^{-1}, v)) \right) \end{aligned}$$

as well as the *invariant ring*

$$(1.3.2.2) \quad \mathbb{C}[V]^{T^n} := \{ f \in \mathbb{C}[V] \mid \forall g \in T^n : g \star f = f \}.$$

⁵You may think of the usual euclidean topology or the Zariski topology. The subgroup D_m is defined by the vanishing of the off diagonal elements, i.e., by the vanishing of certain functions which are continuous in both the euclidean and the Zariski topology.

⁶Again, this is true in both topologies.

⁷This kind of problems can also be considered for non-commutative groups, such as $\mathrm{GL}_n(\mathbb{C})$. In this case, we need to multiply by g^{-1} in order to get a left action. Our convention has been chosen to match this more general situation.

It is the ring consisting of those regular functions on V that are constant on all T^n -orbits. This should be the ring of regular functions on the quotient.

Example 1.3.3. For the action in (1.3.2.1), the only $\mathbb{C}^\star$ -invariant functions which are constant on all $\mathbb{C}^\star$ -orbits are the constant functions. In fact, by continuity, a $\mathbb{C}^\star$ -invariant function is not only constant on an orbit, but also on the closure of that orbit. So, we are dealing with functions which are constant on all lines through the origin. This suggests that, in this example, the quotient we are looking for is just a point.

Remark 1.3.4. The ring $\mathbb{C}[V]$ of regular functions on V is, of course, infinite dimensional as complex vector space. Yet, the action $\alpha^\star$ of T^n on $\mathbb{C}[V]$ is locally finite, i.e., for any $f \in \mathbb{C}[V]$, there exist a finite dimensional linear subspace $f \in W \subset \mathbb{C}[V]$ which is T^n -invariant, that is,

$$\forall g \in T^n \forall h \in W : g \star h \in W,$$

and a representation $\varrho_W : T^n \rightarrow \mathrm{GL}(W)$, such that

$$\forall g \in T^n \forall h \in W : g \star h = \varrho_W(g)(h).$$

For this, we refer to [6], Chapter I, 1.9, Proposition. For a character $\chi : T^n \rightarrow \mathbb{C}^\star$, we let

$$W^\chi := \{ h \in \mathbb{C}[W] \mid \forall g \in T^n : g \star h = \chi(g) \cdot h \}$$

be the corresponding eigenspace. Our previous result, therefore, implies

$$\mathbb{C}[V] = \bigoplus_{\chi \in X(T^n)} W^\chi.$$

Letting $0 : T^n \rightarrow \mathbb{C}^\star, g \mapsto 1$, be the trivial character, we find

$$\mathbb{C}[V]^{T^n} = W^0.$$

It is a fundamental result of Hilbert's ([12], Theorem I, [8], Chapter 3, Section 1, [19], Theorem 3.4) that $\mathbb{C}[V]^{T^n}$ is finitely generated, i.e., there are finitely many T^n -invariant polynomials $f_1, \dots, f_s$, such that every invariant polynomial $f \in \mathbb{C}[V]^{T^n}$ can be written as a polynomial in $f_1, \dots, f_s$. We define the morphism

$$\begin{aligned} \pi : V &\rightarrow \mathbb{C}^s \\ v &\mapsto (f_1(v), \dots, f_s(v)) \end{aligned}$$

which is T^n -invariant, i.e., constant on all T^n -orbits in V . Its image is the quotient we are looking for. It is called the *categorical quotient* and is denoted by $V //_{\varrho} T^n$.⁸

Now, we return to the observation we made in Remark 1.3.2. We assume $V = \mathbb{C}^m$ and that the image of ϱ lies in the subgroup $D_m \subset \mathrm{GL}_m(V)$ of diagonal matrices. Then, each monomial in $\mathbb{C}[x_1, \dots, x_m]$ is an eigenvector for the action of T^n on $\mathbb{C}[x_1, \dots, x_m]$. This implies that, for each character, the eigenspace W^χ has a basis of monomials. In particular, we may choose the generators $f_1, \dots, f_s$ to be monomials. We see that $\pi : \mathbb{C}^m \rightarrow \mathbb{C}^s$ restricts to a homomorphism $\pi^\circ = \pi|_{D_m} : T^m = D_m \rightarrow T^s$ of tori.

Remark 1.3.5. a) At this stage, we recognize one of the problems in forming quotients. This process is not compatible with open embeddings. We look again at the representation

$$\begin{aligned} \varrho : \mathbb{C}^\star &\rightarrow \mathrm{GL}_m(\mathbb{C}) \\ z &\mapsto z \cdot \mathbb{E}_m. \end{aligned}$$

⁸The set $f_1, \dots, f_s$ of generators for the invariant ring is not uniquely determined. However, a different choice of generators leads to a quotient which is “canonically isomorphic” to the given one. See [19], Chapter 2, §4, for more details.

We have explained that the categorical quotient is just a point, $\mathbb{C}^m //_{\varrho} \mathbb{C}^{\star} = \{0\}$, in this case. On the other hand, we may form the quotient of T^m , and the result is an $(m-1)$ -dimensional torus. Recall that $N := \Lambda(T^m) \cong \mathbb{Z}^m$. Then, $\varphi := \Lambda(\varrho): \Lambda(\mathbb{C}^{\star}) \longrightarrow \Lambda(T^m)$ corresponds to

$$\begin{aligned} \mathbb{Z} &\longrightarrow \mathbb{Z}^m \\ k &\longmapsto (k, \dots, k). \end{aligned}$$

So, the quotient $\overline{N} := N/\text{Im}(\varphi)$ is clearly a free abelian group, and our quotient torus is given as

$$\overline{N} \otimes_{\mathbb{Z}} \mathbb{C}^{\star}.$$

b) It is known that the quotient morphism $\pi: V \longrightarrow V //_{\varrho} T^n$ induces a bijection between the set of closed T^n -orbits in V and the set of points of the categorical quotient $V //_{\varrho} T^n$ ([8], Chapter 4, Section 1, Lemma 1, [19], Corollary 3.5.2).

Example 1.3.6. We look at the representation

$$\begin{aligned} \varrho: \mathbb{C}^{\star} &\longrightarrow \text{GL}_2(\mathbb{C}) \\ z &\longmapsto \begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}. \end{aligned}$$

For any point $\underline{a} = (a_1, a_2) \in (\mathbb{C}^{\star})^{\times 2}$, the orbit is a hyperbola and closed inside $\mathbb{C}^2$. The orbit of $(1, 0)$ is the x_1 -axis minus the origin and is not closed. Likewise, the orbit of $(0, 1)$ is the non-closed set $(x_2\text{-axis}) \setminus \{0\}$. Finally, the orbit of the origin is the closed set $\{(0, 0)\}$. It is easy to see that the invariant ring $\mathbb{C}[x_1, x_2]^{\mathbb{C}^{\star}}$ is generated by the monomial $x_1 \cdot x_2$. So, the categorical quotient is isomorphic to $\mathbb{C}$, and the quotient morphism is given by

$$\begin{aligned} \pi: \mathbb{C}^2 &\longrightarrow \mathbb{C} \\ (a_1, a_2) &\longmapsto a_1 \cdot a_2. \end{aligned}$$

If we look at the action

$$\begin{aligned} \varrho': \mathbb{C}^{\star} &\longrightarrow \text{GL}_2(\mathbb{C}) \\ z &\longmapsto \begin{pmatrix} z^2 & 0 \\ 0 & z^{-2} \end{pmatrix}, \end{aligned}$$

we get the same orbits, the same invariant ring, the same categorical quotient, and the same quotient morphism. However, the homomorphisms $\Lambda(\varrho)$ and $\Lambda(\varrho')$ are different. The cokernel of the first homomorphism is a free abelian group, and the one of second has torsion. This reflects the fact that ϱ' has a non-trivial finite kernel, namely the subgroup $\{\pm 1\}$. We will come back to this example in Example 2.4.3, b), and Example 3.3.1, b).

1.4. Semistability and stability. In order to analyze the properties of a categorical quotient and the corresponding quotient map, we will briefly discuss two common notions which go back to Hilbert [13].

Let $m, n \geq 1$, V an m -dimensional complex vector space, and $\varrho: T^n \longrightarrow \text{GL}(V)$ a representation of T^n on V . For a point $v \in V$, the following conditions are equivalent (see [8], Chapter 4, Section 3, Lemma 2):

- 0 is not contained in the closure⁹ $\overline{T^n \cdot v}$ of the orbit of v .

⁹The orbit, denoted by $T^n \cdot v$, is the image of the morphism $T^n \longrightarrow V, g \longmapsto \alpha(g, v) = \varrho(g)(v)$. For this reason, it is constructible, by Chevalley's theorem ([6], Chapter AG, 10.2 Corollary, [17], Chapter I, §8, Corollary 2). This implies that the closures in the strong and the Zariski topology do agree ([17], Chapter I, §10, Corollary 1).

- There exist $d \geq 1$ and an invariant function $f \in \mathbb{C}[V]^{T^n}$ which is homogeneous of degree d and satisfies $f(v) \neq 0$.

Remark 1.4.1. a) This is the way to phrase semistability for arbitrary reductive linear algebraic groups. In view of previous observations, the second condition is equivalent to the existence of a non-constant invariant monomial f with $f(v) \neq 0$.

b) Let V_ϱ^{ss} be the set of semistable points in V . This set is open. In fact, for a point $v \in V_\varrho^{\text{ss}}$, there is a non-constant homogeneous function $f \in \mathbb{C}[V]^{T^n}$ with $f(v) \neq 0$. The open subset $D(f) := \{w \in V \mid f(w) \neq 0\}$ is clearly contained in V_ϱ^{ss} . So, V_ϱ^{ss} is an open set. It is also invariant under the T^n -action.

c) Let $\pi: V \rightarrow V_\varrho^{\text{ss}}$ be the quotient morphism. Then, V_ϱ^{ss} is the complement of $\pi^{-1}(\pi(0))$. This also shows that V_ϱ^{ss} is a T^n -invariant open subset of V .

A point $v \in V$ is *stable*, if the orbit $T^n \cdot v$ is closed and the stabilizer of v is finite.

Remark 1.4.2. Since we assume that $n \geq 1$, a stable point is different from the origin, and the second assumption implies that a stable point is semistable.

Theorem 1.4.3. i) *The set*

$$V_\varrho^{\text{s}} := \{v \in V \mid v \text{ is stable}\}$$

is open and T^n -invariant.

ii) *For every T^n -invariant open subset $U \subset V_\varrho^{\text{s}}$, the image $\pi(U)$ of U under the quotient map $\pi: V \rightarrow V //_\varrho T^n$ is an open subset of $V //_\varrho T^n$, and any fiber of the restricted morphism*

$$\pi|_U: U \rightarrow \pi(U)$$

consists of exactly one orbit,

Proof. For i), see [19], Lemma 3.12, and, for ii), [19], Theorem 3.14, (ii), and Proposition 3.10, (a). $\square$

So, the concept of stability provides us with the subset V_ϱ^{s} , such that we can endow the set of T^n -orbits in this open subset with a natural structure of a quasi-affine variety. This means that, on this open subset, the quotient can be taken in the best possible way. We will denote it by V_ϱ^{s}/T^n . In addition, for any T^n -invariant open subset $U \subset V_\varrho^{\text{s}}$, the quotient U/T^n exists as an open subset of V_ϱ^{s}/T^n , by Part ii) of the above theorem. However, as we know from Example 1.3.3, V_ϱ^{s} can be empty.

Remark 1.4.4. A point $v \in V \setminus \{0\}$ is *polystable*, if the orbit $T^n \cdot v$ is a closed subset of V . The set V_ϱ^{ps} of polystable points is, in general, not open nor does it carry a natural structure of an algebraic variety.

Example 1.4.5. In Example 1.3.6, we have

$$V_\varrho^{\text{s}} = V_\varrho^{\text{ss}} = \{(a_1, a_2) \in \mathbb{C}^2 \mid a_1 \cdot a_2 \neq 0\} = \mathbb{C}^2 \setminus ((x_1\text{-axis}) \cup (x_2\text{-axis})).$$

Let $m \geq 1$. Identifying T^m with the subgroup $D_m \subset \text{GL}_m(\mathbb{C})$ of diagonal matrices, we arrive at the action

$$\begin{aligned} \beta: T^m \times \mathbb{C}^m &\longrightarrow \mathbb{C}^m \\ ((\varepsilon_1, \dots, \varepsilon_m), (a_1, \dots, a_m)) &\longmapsto (\varepsilon_1 \cdot a_1, \dots, \varepsilon_m \cdot a_m). \end{aligned}$$

For $n \geq 1$ and a homomorphism $\varrho: T^n \rightarrow T^m$, the action $\alpha: T^n \times \mathbb{C}^m \rightarrow \mathbb{C}^m$ is induced by ϱ and β , that is,

$$\forall g \in T^n \forall v \in \mathbb{C}^m: \quad \alpha(g, v) = \beta(\varrho(g), v).$$

Lemma 1.4.6. i) If, in the above situation, $V_\varrho^{\text{ss}} \neq \emptyset$, then $T^m \subset V_\varrho^{\text{ss}}$.
 ii) Likewise, if $V_\varrho^{\text{s}} \neq \emptyset$, then $T^m \subset V_\varrho^{\text{s}}$.

Proof. Any set which is non-empty and open in the Zariski topology is dense (both in the euclidean and in the Zariski topology; see [11], Chapter I, Example 1.1.3, [17], Page 26 and §10, Theorem 1). Furthermore, the intersection of two dense open subsets is non-empty. So, under the respective assumption, we find

$$T^m \cap V_\varrho^{(\text{s})\text{s}} \neq \emptyset.$$

Pick a point v_0 in this intersection. Then, $T^m \cdot v_0 = T^m$. It remains to show that V_ϱ^{ss} and V_ϱ^{s} are T^m -invariant. From our observation about the compatibility of the actions α and β and the commutativity of the group T^m , we infer

$$\forall v \in \mathbb{C}^m \forall g \in T^m : \quad T^n \cdot (g \cdot v) = g \cdot (T^n \cdot v).$$

Since the map

$$\begin{aligned} \beta_g : \mathbb{C}^m &\longrightarrow \mathbb{C}^m \\ v &\longmapsto \beta(g, v) \end{aligned}$$

is an isomorphism, $g \in T^m$, the assertions now follow immediately from the characterization of semistable and stable points by properties of their T^n -orbits. $\square$

2. MODULI SPACES OF NEURAL NETWORKS

In this part, we will first review the formalism of quivers and framed quivers and discuss the relevant group actions and quotient spaces. In the language of [2] and [1], the latter are moduli spaces for thin representations. We will describe the class of quivers that we allow for neural networks and give an elementary proof of a basic result about generators for invariant rings from [10] and [1] in the special situation relevant for neural networks. We will also recall how to determine relations among those generators. These give equations for the moduli spaces. Finally, we will make an easy but quite interesting observation regarding the action on and the quotient of the dense torus in the parameter space of thin framed quiver representations.

2.1. Quivers. A *quiver* is a quadruple $Q = (V, E, s, t)$, consisting of finite sets V and E and maps $s, t : E \longrightarrow V$. We will call the elements of V *vertices* and the elements of E *arrows* or *oriented edges*. For an arrow $e \in E$, the vertex $s(e)$ is the *source* and the vertex $t(e)$ the *target* of e . So, we will think of e as an arrow pointing from $s(e)$ to $t(e)$. Let $n \geq 1$. Then, a *path* of *length* n is a tuple $p = (e_1, \dots, e_n)$ of arrows, such that $t(e_i) = s(e_{i+1})$, $i = 1, \dots, n-1$.¹⁰ We call $s'(p) := s(e_1)$ the *source* of p and $t'(p) := t(e_n)$ the *target* of p . An *oriented cycle* is a path p with $s'(p) = t'(p)$ and a *loop* an oriented cycle of length one. For a quiver $Q = (V, E, s, t)$, we let “ $\sim$ ” be the equivalence relation on V generated by $s(e) \sim t(e)$, $e \in E$. The equivalence classes of “ $\sim$ ” are called the *connected components* of Q , and we say that Q is *connected*, if there is only one connected component.

Remark 2.1.1. The quiver Q is connected if and only if, for any two distinct vertices $a, b \in V$, there exist $n \geq 1$, a tuple $(e_1, \dots, e_n)$ of arrows, and vertices $v_1 := a, v_2, \dots, v_n, v_{n+1} := b$ with $\{s(e_i), t(e_i)\} = \{v_i, v_{i+1}\}$, $i = 1, \dots, n$.

¹⁰So, a path of length one is just an arrow.

A *framed quiver* is a quadruple $Q = (V_u, V_h, E, s, t)$ in which V_u and V_h are disjoint sets and $s, t: E \rightarrow V := V_u \sqcup V_h$ are maps. So, (V, E, s, t) is a quiver as before. The vertices $v \in V_u$ are said to be *unmarked* and the vertices $v \in V_h$ *hidden*. In pictures, we will draw unmarked vertices as circles and hidden vertices as dots. The *hidden subquiver* of Q is $\tilde{Q} = (V_h, E_h, s|_{E_h}, t|_{E_h})$ with $E_h := \{e \in E \mid s(e) \in V_h \wedge t(e) \in V_h\}$.

Let $L \geq 3$. A *quiver with L layers* is a tuple $Q = (V_1, \dots, V_L, E, s, t)$ in which $V_1, \dots, V_L$ are pairwise disjoint non-empty sets and $s, t: E \rightarrow V_1 \sqcup \dots \sqcup V_L$, such that

- for every element $e \in E$, there are $1 \leq i < j \leq L$ with $s(e) \in V_i$ and $t(e) \in V_j$,
- the *associated quiver* (V, E, s, t) is connected.

We call $I := V_1$ the *input layer* and $O := V_L$ the *output layer*. Setting $V_u := I \sqcup O$, $V_h := V_2 \sqcup \dots \sqcup V_{L-1}$, the *associated framed quiver* is (V_u, V_h, s, t) .

Remark 2.1.2. Obviously, there do not exist any oriented cycles in the associated quiver of a quiver with L layers.

Convention 2.1.3. To ease notation, for a quiver $Q = (V_1, \dots, V_L, E, s, t)$ with L layers, we will denote the associated framed quiver as well as the associated quiver also by Q .

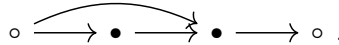
Note that, given a quiver $Q = (V_1, \dots, V_L, E, s, t)$ with L layers, we may redefine the sets $V_1, \dots, V_L$, such that all sources of Q lie in V_1 and all sinks in V_L . This will, however, change the associated framed quiver. Next, given a quiver $Q = (V_1, \dots, V_L, E, s, t)$ with L layers, we may remove all vertices from V_1 and all arrows which start at a vertex in V_1 . The resulting quiver may be viewed as a disjoint union of quivers with at most $L - 1$ layers.

We say that a quiver $Q = (V_1, \dots, V_L, E, s, t)$ with L layers *satisfies Condition (M)*, if all sources of Q lie in V_1 and all sinks in V_L . Now, we recursively define *Condition (N)* for quivers with L layers:

- For a quiver with three layers, it is Condition (M).
- For a quiver with $L + 1$ layers, it means that the connected components of the quiver obtained by removing all vertices in V_1 and all arrows with source in V_1 are layered quivers which satisfy Condition (N).

Remark 2.1.4. a) Given a quiver $Q = (V_1, \dots, V_L, E, s, t)$ with L layers which satisfies Condition (M), it is possible to rearrange $V_2, \dots, V_{L-1}$ in such a way that Q satisfies Condition (N).

b) Condition (N) allows for arrows skipping a layer as in the following example:



c) Proposition 3.2.1 shows that Condition (M) is kind of a minimal requirement for having reasonable moduli spaces.

2.2. Moduli spaces attached to framed quivers. Let $Q = (V_u, V_h, E, s, t)$ be a framed quiver. We introduce the vector space

$$R_Q := \bigoplus_{e \in E} \mathbb{C}$$

and the torus

$$T_Q := \bigtimes_{v \in V} \mathbb{C}^\star.$$

An element of R_Q will be written in the form $(r_e, e \in E)$ and an element of T_Q as $(\lambda_v, v \in V)$. Using this notation, we define the group action¹¹

$$\begin{aligned} \alpha_Q: T_Q \times R_Q &\longrightarrow R_Q \\ ((\lambda_v, v \in V), (r_e, e \in E)) &\longmapsto (\lambda_{t(e)} \cdot r_e \cdot \lambda_{s(e)}^{-1}, e \in E). \end{aligned}$$

The *change of basis group* is

$$T_h := \bigotimes_{v \in V_h} \mathbb{C}^\star.$$

Define

$$\iota(\lambda_v, v \in V_h) := (\kappa_v, v \in V) \quad \text{with} \quad \kappa_v := \begin{cases} \lambda_v, & \text{if } v \in V_h \\ 1, & \text{if } v \in V_u \end{cases}, \quad (\lambda_v, v \in V) \in T_h.$$

Then,

$$\begin{aligned} \iota: T_h &\longrightarrow T_Q \\ (\lambda_v, v \in V_h) &\longmapsto \iota(\lambda_v, v \in V_h) \end{aligned}$$

is a group homomorphism, and

$$\begin{aligned} \alpha: T_h \times R_Q &\longrightarrow R_Q \\ ((\lambda_v, v \in V_h), (r_e, e \in E)) &\longmapsto \alpha_Q(\iota(\lambda_v, v \in V_h), (r_e, e \in E)) \end{aligned}$$

is the action we are interested in.

Remark 2.2.1. Set

$$D_Q := \bigotimes_{e \in E} \mathbb{C}^\star.$$

Then, the action α is associated with a homomorphism

$$(2.2.1.1) \quad \varrho: T_h \longrightarrow D_Q$$

of tori (compare Section 1.2). In fact, setting $\varepsilon_e := 1, e \in E$, and $\varepsilon = (\varepsilon_e, e \in E)$, we get

$$(2.2.1.2) \quad \varrho(\lambda_v, v \in V_h) = \alpha((\lambda_v, v \in V_h), \varepsilon) = (\lambda_{t(e)} \cdot \lambda_{s(e)}^{-1}, e \in E).$$

The categorical quotient (see Page 8)

$$\mathcal{M}_Q := R_Q // T_h$$

is the *moduli space* of the framed quiver Q .

2.3. The invariant ring. In this section, we will explain how to compute the invariant ring $\mathbb{C}[R_Q]^{T_h}$ (cf. (1.3.2.2)) in the case that $Q = (V_1, \dots, V_L, E, s, t)$ is a quiver in layers. We already know from Page 8 that it may be generated by monomials. We introduce a generator x_e , for every arrow $e \in E$. In this way, we may identify $\mathbb{C}[R_Q]$ with the polynomial ring $\mathbb{C}[x_e, e \in E]$. Accordingly, a monomial will be written in the form $x_{\underline{\mu}}$ with $\underline{\mu} = (\mu_e, e \in E)$ a collection of natural numbers. For a path $p = (e_1, \dots, e_u)$, we define

$$\underline{\mu}_p = (\mu_e, e \in E) \quad \text{with} \quad \mu_e := \begin{cases} 1, & \text{if } e \in \{e_1, \dots, e_u\} \\ 0, & \text{if } e \notin \{e_1, \dots, e_u\} \end{cases},$$

as well as the monomial

$$M_p := x_{\underline{\mu}_p}.$$

Exercise 2.3.1. Let p be a path in Q whose source lies in V_1 and whose target lies in V_L . Prove that the monomial M_p is invariant under the action of T_h on R_Q .

¹¹See, e.g., [3] for the general set-up which explains the conventions.

Theorem 2.3.2. *The monomials M_p , p a path in Q with $s'(p) \in V_1$ and $t'(p) \in V_L$, generate the invariant ring $\mathbb{C}[R_Q]^{T_h}$.*

Proof. This is a special case of a more general theorem ([10], Theorem 1.1, see also [1], Theorem 3.3). Assume that $\underline{\mu} = (\mu_e, e \in E)$ is a collection of natural numbers, not all zero, such that $\underline{x}^\mu$ is an invariant monomial. We define $S := \{e \in E \mid \mu_e \neq 0\}$. We will show that there is a path in Q with source in V_1 and target in V_L which passes only through vertices in S . Then, the monomial M_p divides the monomial $\underline{x}^\mu$. The quotient $\underline{x}^\mu/M_p$ is still an invariant monomial. Iterating the construction, we see that $\underline{x}^\mu$ is a product of monomials of the form M_p , p a path in Q with source in V_1 and target in V_L .

Let $j_0 := \min\{j \in \{1, \dots, L\} \mid \exists e \in S : s(e) \in V_j\}$. Assume that $j_0 > 1$ and pick an arrow $e_0 \in S$ with $s(e_0) \in V_{j_0}$. For a complex number $t \in \mathbb{C}^\star$, consider

$$(\lambda_v, v \in V_h) \in T_h \quad \text{with} \quad \lambda_v := \begin{cases} t, & \text{if } v = s(e_0) \\ 1, & \text{if } v \neq s(e_0) \end{cases}.$$

This element acts on the monomial $\underline{x}^\mu$ by $t^{-\mu_{s(e_0)}}$. Since $\mu_{s(e_0)} \geq 1$, the monomial $\underline{x}^\mu$ cannot be invariant, a contradiction.

Next, we consider

$$j_1 := \max\{j \in \{1, \dots, L\} \mid \exists \text{ path } p \text{ in } S \text{ with } s'(p) \in V_1 \text{ and } t'(p) \in V_j\}.$$

We already know that $j_1 \geq 2$, and we would like to show $j_1 = L$. So, suppose $j_1 < L$. Pick a path p_0 in S with $s'(p_0) \in V_1$ and $t'(p_0) \in V_{j_1}$, and define, for a complex number t ,

$$(\lambda_v, v \in V_h) \in T_h \quad \text{with} \quad \lambda_v := \begin{cases} t, & \text{if } v = t'(p_0) \\ 1, & \text{if } v \neq t'(p_0) \end{cases}.$$

This element acts on the monomial $\underline{x}^\mu$ by $t^{\mu_{t'(p_0)}}$. Since $\mu_{t'(p_0)} \geq 1$, the monomial $\underline{x}^\mu$ cannot be invariant. Again, we have arrived at a contradiction. $\square$

Exercise 2.3.3. In (1.2.0.1), we have established an isomorphism $X(D_Q) \longrightarrow \mathbb{Z}^{\#E}$. This means that, for every character $\chi : D_Q \longrightarrow \mathbb{C}^\star$, there is a unique tuple $\underline{\mu} = (\mu_e, e \in E)$ of integers, such that $\chi = \chi^\mu$ with

$$\begin{aligned} \chi^\mu : D_Q &\longrightarrow \mathbb{C}^\star \\ (x_e, e \in E) &\longmapsto \prod_{e \in E} x_e^{\mu_e}. \end{aligned}$$

The homomorphism $\varrho : T_h \longrightarrow D_Q$ from (2.2.1.2) induces the homomorphism

$$\begin{aligned} X(\varrho) : X(D_Q) &\longrightarrow X(T_h) \\ \chi &\longmapsto \chi \circ \rho. \end{aligned}$$

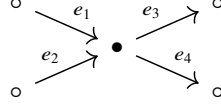
Check that $M_p \in \ker(X(\varrho))$, for every path p in Q with $s'(p) \in V_1$ and $t'(p) \in V_L$, and mimic the proof of Theorem 2.3.2 to show that, for every character $\chi \in \ker(X(\varrho))$, there are paths $p_1, \dots, p_r$ in Q with $s'(p_j) \in V_1$ and $t'(p_j) \in V_L$, $j = 1, \dots, r$, as well as integers $\alpha_1, \dots, \alpha_r$, such that

$$\chi = \prod_{j=1}^r M_{p_j}^{\alpha_j},$$

or, in additive notation,

$$\chi = \sum_{j=1}^r \alpha_j \cdot M_{p_j}.$$

Example 2.3.4. In Figure 1 of [16], the quiver



has been considered. According to Theorem 2.3.2, we need to look at the four paths $p_1 := (e_1, e_3)$, $p_2 := (e_1, e_4)$, $p_3 := (e_2, e_3)$, and $p_4 := (e_2, e_4)$, yielding generators $M_{p_1}, M_{p_2}, M_{p_3}, M_{p_4}$ for the invariant ring. Note that there is the relation

$$M_{p_1} \cdot M_{p_4} = M_{p_2} \cdot M_{p_3}.$$

This shows that the quotient $\mathcal{M}(Q)$ is isomorphic to a quadric cone in $\mathbb{C}^4$ (see also Examples 2.4.3 and 3.1.5).

There are general algorithms for determining relations among the generators from Theorem 2.3.2 in concrete examples based on Gröbner basis techniques (see [23]). Suppose we are given $\underline{\mu}_1, \dots, \underline{\mu}_s \in \mathbb{N}^{\times n}$. We then look at the homomorphism

$$\begin{aligned} h: \mathbb{C}[t_1, \dots, t_s] &\longrightarrow \mathbb{C}[x_1, \dots, x_n] \\ t_i &\longmapsto \underline{x}^{\underline{\mu}_i}, \quad i = 1, \dots, s. \end{aligned}$$

It corresponds to the morphism

$$\begin{aligned} F: \mathbb{C}^n &\longrightarrow \mathbb{C}^s \\ (x_1, \dots, x_n) &\longmapsto (\underline{x}^{\underline{\mu}_1}, \dots, \underline{x}^{\underline{\mu}_s}). \end{aligned}$$

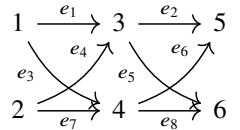
The kernel I of h is a *toric ideal* in the sense of [23], Page 31. It describes the image of F . Algorithm 4.5 in [23] describes an easy way to compute a reduced Gröbner basis of I . Using the remarks at the bottom of Page 32, we get the following procedure:

- Fix an order “ $\prec$ ” on the set of monomials in $x_1, \dots, x_n, t_1, \dots, t_s$, such that $t_i \prec x_j$, $i = 1, \dots, s$, $j = 1, \dots, n$.
- Compute the reduced Gröbner basis $\mathcal{G}$ with respect to “ $\prec$ ” of the ideal

$$\langle t_i - \underline{x}^{\underline{\mu}_i} : i = 1, \dots, s \rangle.$$

- The elements of $\mathcal{G} \cap \mathbb{C}[t_1, \dots, t_s]$ will form the reduced Gröbner basis for I with respect to “ $\prec$ ”.

Example 2.3.5. Consider the quiver



The moduli space is the spectrum of the corresponding invariant ring. By Theorem 2.3.2, the invariant ring is

$$\mathbb{C}[x_1x_2, x_1x_5, x_2x_4, x_3x_6, x_3x_8, x_4x_5, x_6x_7, x_7x_8] \subset \mathbb{C}[x_1, \dots, x_8].$$

We choose the lexicographic ordering $t_1 \prec \dots \prec t_8 \prec x_1 \prec \dots \prec x_8$, and we form the ideal

$$\langle t_1 - x_1x_2, t_2 - x_1x_5, t_3 - x_2x_4, t_4 - x_3x_6, t_5 - x_3x_8, t_6 - x_4x_5, t_7 - x_6x_7, t_8 - x_7x_8 \rangle.$$

The above procedure yields¹²

$$I = \langle t_1 t_6 - t_2 t_3, t_4 t_8 - t_5 t_7 \rangle.$$

So, the moduli space is the complete intersection of two quadric cones in $\mathbb{C}^6$.

We can partially generalize these observations. For this, we will first produce some obvious equations for the moduli space. We set

$$(2.3.5.1) \quad \mathcal{P} := \{ p \mid p \text{ a path in } Q \text{ with } s'(p) \in V_1 \text{ and } t'(p) \in V_L \}.$$

For a path $p = (e_1, \dots, e_t) \in \mathcal{P}$ and an index $\ell \in \{2, \dots, L-1\}$, such that p passes through a vertex in the layer V_ℓ , we let $k \in \{1, \dots, t-1\}$ be the index with $t(e_k) \in V_\ell$ and define

$${}_\ell p := (e_1, \dots, e_k), \quad {}^\ell p := (e_{k+1}, \dots, e_t),$$

so that, using the symbol “ $\bullet$ ” for the usual concatenation of paths, we have

$$p = {}^\ell p \bullet {}_\ell p.$$

Next, let $p_1, p_2 \in \mathcal{P}$ and $\ell \in \{2, \dots, L-1\}$. We say that p_1 and p_2 *meet* in the (hidden) layer V_ℓ , if there is a vertex $v_0 \in V_\ell$, such that both p_1 and p_2 pass through v_0 . We then define

$$p_1 \sqcup_\ell p_2 := {}^\ell p_1 \bullet {}_\ell p_2.$$

Note that

$$M_{p_1} \cdot M_{p_2} = M_{p_1 \sqcup_\ell p_2} \cdot M_{p_2 \sqcup_\ell p_1}.$$

For a path $p \in \mathcal{P}$, we let $t_p \in \{t_1, \dots, t_s\}$ be the variable with $h(t_p) = M_p$. By the above observation, we have

$$t_{p_1} \cdot t_{p_2} - t_{p_1 \sqcup_\ell p_2} \cdot t_{p_2 \sqcup_\ell p_1} \in \ker(h).$$

If ${}_\ell p_1 = {}_\ell p_2$ or ${}^\ell p_1 = {}^\ell p_2$, then this relation will be just zero. There are further redundancies. In fact, if p_1 and p_2 share a path p' , then the above construction will produce the same result, for every $\ell \in \{2, \dots, L-1\}$, such that p' contains a vertex of V_ℓ . So, for $p_1, p_2 \in \mathcal{P}$ and $\ell \in \{2, \dots, L-1\}$, we say that p_1 and p_2 *properly meet* in the (hidden) layer V_ℓ , if there is a vertex $v_0 \in V_\ell$, such that¹³

$$t'({}_\ell p_1) = t'({}_\ell p_2), \quad {}_\ell p_1 \neq {}_\ell p_2, \quad \text{and} \quad t'({}^{\ell+1} p_1) \neq t'({}^{\ell+1} p_2).$$

There is one further redundancy which we will mention in Remark 2.3.7, a). Based on our discussion, we set

$$\mathcal{E} := \{ t_{p_1} t_{p_2} - t_{p_1 \sqcup_\ell p_2} t_{p_2 \sqcup_\ell p_1} \mid p_1, p_2 \in \mathcal{P}, \ell \in \{2, \dots, L-1\} : p_1, p_2 \text{ properly meet in } V_\ell \}.$$

The equations in $\mathcal{E}$ cut out an algebraic subset $X \subset \mathbb{C}^s$, and we clearly have

$$\mathcal{M}_Q \subset X.$$

We would like to show that we actually do have equality. More precisely, the following holds true.

Theorem 2.3.6. *The ideal $I = \ker(h)$ is generated by the elements of $\mathcal{E}$.*

¹²Of course, one readily checks that these two relations do hold.

¹³The last condition says that p_1 and p_2 split at the layer V_ℓ , i.e., continue to different vertices in the quiver.

Proof. Let us make some preliminary observations. The polynomial ring $\mathbb{C}[x_1, \dots, x_n]$ is an integral domain, i.e., for polynomials $F, G \in \mathbb{C}[x_1, \dots, x_n]$, the identity $F \cdot G = 0$ implies that $F = 0$ or $G = 0$. For this reason, I is a prime ideal, i.e., for $f, g \in I$, the relation $f \cdot g \in I$ implies $f \in I$ or $g \in I$.

Furthermore, by [23], Lemma 4.3, a toric ideal is generated by elements of the form "monomial minus monomial", i.e., elements of the form

$$(2.3.6.1) \quad t_{p_1} \cdots t_{p_a} - t_{q_1} \cdots t_{q_b},$$

for appropriate natural numbers a, b and elements $p_1, \dots, p_a, q_1, \dots, q_b \in \mathcal{P}$. Since I is a prime ideal, we may assume that the sets $\{p_1, \dots, p_a\}$ and $\{q_1, \dots, q_b\}$ are disjoint. Finally, note that we must have $a = b$. In fact, a and b count (with multiplicities) the number of variables associated with arrows $e \in E$ with $s(e) \in V_1$ in $t_{p_1} \cdots t_{p_a}$ and $t_{q_1} \cdots t_{q_b}$, respectively.

We will show by induction on a that an element from I as in (2.3.6.1) is contained in the ideal $\tilde{I}$ that is generated by the elements of $\mathcal{E}$. First, note that there is no element of the form $t_p - t_q$ with $p \neq q$ in I . If we have an element as in (2.3.6.1) with $a = b \geq 2$ and write $q_1 = (f_1, \dots, f_u)$, then there must be an index $i \in \{1, \dots, a\}$, such that p_i starts with f_1 . So, it makes sense to define

$$(2.3.6.2) \quad j_0 := \max \{ j \in \{1, \dots, u\} \mid \exists i \in \{1, \dots, a\} : p_i = (f_1, \dots, f_j, e_{j+1}, \dots, e_t) \}.$$

Since $q_1 \notin \{p_1, \dots, p_a\}$, we have $j_0 < u$. We may assume that p_1 realizes the maximum. Then, p_1 does not contain the arrow f_{j_0+1} . However, there must be an index $i_0 \in \{2, \dots, a\}$, such that p_{i_0} contains f_{j_0+1} . Without loss of generality, we may assume $i_0 = 2$. Then, p_1 and p_2 properly meet in the layer V_ℓ . Here, V_ℓ contains $s(f_{j_0+1})$. Then,

$$(t_{p_1} \cdot t_{p_2} - t_{p_1 \sqcup_\ell p_2} \cdot t_{p_2 \sqcup_\ell p_1}) \cdot t_{p_3} \cdots t_{p_a} \in \tilde{I}.$$

We subtract this element from the one in (2.3.6.1). The result is

$$t_{p_1 \sqcup_\ell p_2} \cdot t_{p_2 \sqcup_\ell p_1} \cdot t_{p_3} \cdots t_{p_a} - t_{q_1} \cdots t_{q_a}.$$

Now,

$$p_2 \sqcup_\ell p_1 = (f_1, \dots, f_{j_0+1}, h_{j_0+2}, \dots, h_v).$$

This shows that we may modify our original element by an element from $\tilde{I}$, such that the result has the form

$$t_{q_1} \cdot (t_{w_2} \cdots t_{w_a} - t_{z_2} \cdots t_{z_a}),$$

for appropriate paths $w_2, \dots, w_a, z_2, \dots, z_a \in \mathcal{P}$. This element still belongs to I . Since $t_{q_1} \notin I$, we must have $t_{w_2} \cdots t_{w_a} - t_{z_2} \cdots t_{z_a} \in I$. If $a = 2$, this element must be zero, by our initial observation. Otherwise, we conclude by induction that it must belong to $\tilde{I}$. This ends the proof. $\square$

Remark 2.3.7. a) Let $p_1, p_2 \in \mathcal{P}$ and $\ell \in \{2, \dots, L-1\}$ be such that p_1 and p_2 meet in the layer V_ℓ . Then, $q_1 := p_1 \sqcup_\ell p_2$ and $q_2 := p_2 \sqcup_\ell p_1$ are paths properly meeting in the layer V_ℓ , such that

$$p_1 = q_1 \sqcup_\ell q_2 \quad \text{and} \quad p_2 = q_2 \sqcup_\ell q_1.$$

So, for every equation in $\mathcal{E}$, also minus that equation belongs to $\mathcal{E}$.

b) It is a natural question to determine the number $\#\mathcal{E}/2$. Given $L \geq 3$ and sets $V_1, \dots, V_L$, the associated (multilayer) perceptron is $Q = (V_1, \dots, V_L, E, s, t)$ where, for each $\ell \in \{1, \dots, L-1\}$, each $v \in V_\ell$, and each $w \in V_{\ell+1}$, there is exactly one arrow pointing

from v to w . Examples 2.3.4, 2.3.5, and 3.3.6 all display multilayer perceptrons. For a three layer perceptron, one computes

$$\frac{\#\mathcal{E}}{2} = \binom{\#V_1}{2} \cdot \#V_2 \cdot \binom{\#V_3}{2}.$$

For perceptrons with more layers, the counting becomes more difficult, because distinct paths may meet in more than one layer, including the input layer and the output layer. For a four layer perceptron $Q = (V_1, V_2, V_3, V_4, E, s, t)$, one finds

$$\frac{\#\mathcal{E}}{2} = \binom{\#V_2}{2} \cdot \binom{\#V_4}{2} \cdot (\#V_1)^2 \cdot \#V_3 + \binom{\#V_1}{2} \cdot \binom{\#V_3}{2} \cdot \#V_2 \cdot (\#V_4)^2 + \binom{\#V_1}{2} \cdot \binom{\#V_4}{2} \cdot \#V_1 \cdot \#V_2.$$

In Example 3.3.6, we indeed get $\#\mathcal{E}/2 = 20$. In general, the right hand sides in the above equations provide sharp upper bounds for quivers in layers contained in the corresponding multilayer perceptron.

We conclude this section by an exercise which we will use in Section 3.1.

Exercise 2.3.8. Let $Q = (V_1, \dots, V_L, s, t)$ be a quiver with L layers which satisfies Condition (N).

a) Show that, for every vertex $v_0 \in V_h = V_2 \sqcup \dots \sqcup V_{L-1}$, there exists a path p in Q which starts at a vertex in V_1 , passes through v_0 , and ends at a vertex in V_L .

b) Let $e \in E$ be an arrow. Show that there exists a path $p = (e_1, \dots, e_t)$ in Q with $s'(p) \in V_1$, $t'(p) \in V_L$, and $e \in \{e_1, \dots, e_t\}$.

2.4. The quotient of the torus D_Q . Next, we will present an important observation which will eventually enable us to understand the real points of the moduli spaces. Here, we will work with a quiver $Q = (V_1, \dots, V_L, E, s, t)$ in layers.

Remark 2.4.1. The invariant open subset D_Q is clearly characterized by the inequality

$$\tilde{M}_0 \neq 0 \quad \text{with} \quad \tilde{M}_0 := \prod_{e \in E} x_e.$$

The monomial $\tilde{M}_0$ need not be invariant under the action of T_h . However, by Exercise 2.3.8, b), the open subset D_Q may also be described by the inequality

$$M_0 \neq 0, \quad M_0 := \prod_{p \in \mathcal{P}} M_p, \quad \mathcal{P} \text{ the set of paths from a vertex in } V_1 \text{ to a vertex in } V_L.$$

The monomial M_0 is invariant under the T_h -action, by Exercise 2.3.1. Accordingly, the ring of T_h -invariant regular functions is given by the localization

$$(\mathbb{C}[R_Q]^{T_h})_{M_0} = \left\{ \frac{f}{M_0^k} \mid f \in \mathbb{C}[R_Q]^{T_h}, k \in \mathbb{N} \right\}.$$

First, let $r \geq 1$ and $N \cong \mathbb{Z}^{\oplus r}$ a finitely generated free abelian group. A subgroup $A \subset N$ is *saturated*, if the quotient $B := N/A$ is torsion free. In that case, B is also a finitely generated free abelian group, and there exists a group homomorphism $\sigma: B \rightarrow N$ with $s \circ \sigma = \text{id}_B$, $s: N \rightarrow B$ the projection. This yields an isomorphism $N \cong A \oplus B$.

Proposition 2.4.2. *In the above situation, the homomorphism*

$$\Lambda(Q): \Lambda(T_h) \rightarrow \Lambda(D_Q)$$

is injective, and the image is a saturated subgroup.

Proof. For $v \in V_h$, we define the one parameter subgroup

$$\begin{aligned} \varepsilon_v : \mathbb{C}^\star &\longrightarrow T_h \\ z &\longmapsto (z^{\delta_{v,w}}, w \in V). \end{aligned}$$

Then, $(\varepsilon_v, v \in V)$ is a basis for the free abelian group $\Lambda(T_h)$. In a similar fashion, we define the basis $(\eta_e, e \in E)$ for the free abelian group $\Lambda(D_Q)$.

Suppose we are given $\eta = \sum_{e \in E} \beta_e \cdot \eta_e \in \Lambda(D_Q)$ and $\alpha_v \in \mathbb{Q}$ with

$$\eta = \sum_{v \in V_h} \alpha_v \cdot \Lambda(\varrho)(\varepsilon_v).$$

We need to show that the coefficients $\alpha_v, v \in V_h$, are uniquely determined integers.

First, we look at an arrow $e_0 \in E$, such that $s(e_0)$ is a source of Q . Then, $\Lambda(\varrho)(\varepsilon_{t(e_0)})$ is the only element which can contribute to η_{e_0} , so that¹⁴

$$\alpha_{t(e_0)} = \beta_{e_0}.$$

For the element

$$\eta' := \eta - \beta_{e_0} \cdot \Lambda(\varrho)(\varepsilon_{t(e_0)}) = \sum_{e \in E} \beta'_e \cdot \eta_e,$$

we have $\beta'_{e_0} = 0$, and also $\beta'_e = 0$, for any other arrow e with $s(e) \in V_1$ and $t(e) = t(e_0)$.

We set

$$\tilde{V}_2 := \{v \in V \mid \exists e \in E : s(e) \in V_1 \wedge t(e) = v\},$$

and pick, for every vertex $v \in \tilde{V}_2$, an arrow $e(v) \in E$ with $s(e(v)) \in V_1$ and $t(e(v)) = v$.

So, if we replace η by

$$\eta - \sum_{v \in \tilde{V}_2} \beta_{e(v)} \cdot \Lambda(\varrho)(\varepsilon_v),$$

we may assume that $\eta_e = 0$, for all $e \in E$ with $s(e) \in V_1$.

We may now prove the assertion by induction on the number of hidden layers. If there is only one hidden layer, the above consideration immediately does the trick. In general, we also carry out the above construction. Let $Q' = (V', E', s', t')$ be the quiver that is obtained from Q by removing the layer V_1 and all arrows $e \in E$ with $s(e) \in V_1$. We define $D_{Q'}$ and T'_h as above. Then,

$$D_Q = D_{Q'} \times \bigtimes_{\substack{e \in E: \\ s(e) \in V_1}} \mathbb{C}^\star \quad \text{and} \quad T_h = T'_h \times \bigtimes_{v \in V_2} \mathbb{C}^\star.$$

Therefore, the natural maps

$$\Lambda(D_{Q'}) \longrightarrow \Lambda(D_Q) \quad \text{and} \quad \Lambda(T'_h) \longrightarrow \Lambda(T_h)$$

are injective and their images are saturated subgroups of $\Lambda(D_Q)$ and $\Lambda(T_h)$, respectively. Furthermore, there is the commutative diagram

$$\begin{array}{ccc} \Lambda(T'_h) & \longrightarrow & \Lambda(D_{Q'}) \\ \downarrow & & \downarrow \\ \Lambda(T_h) & \longrightarrow & \Lambda(D_Q). \end{array}$$

By the initial construction, we are reduced to the case that η lies in $\Lambda(D_{Q'})$ and is the image of an element of $\Lambda(T'_h)$. Let $Q^{(1)}, \dots, Q^{(c)}$ be the connected components of the quiver Q . By

¹⁴Compare Equation (2.2.1.2).

assumption, $Q^{(i)}$ satisfies Condition (N) and we define $T_h^{(i)}$ as well as $\varrho^{(i)} : T_h^{(i)} \longrightarrow D_{Q^{(i)}}$ as before, $i = 1, \dots, c$. Then,

$$D_{Q'} = \bigtimes_{i=1}^c D_{Q^{(i)}} \quad \text{and} \quad T'_h = \bigtimes_{i=1}^c T_h^{(i)},$$

so that

$$\Lambda(D_{Q'}) = \bigoplus_{i=1}^c \Lambda(D_{Q^{(i)}}) \quad \text{and} \quad \Lambda(T'_h) = \bigoplus_{i=1}^c \Lambda(T_h^{(i)})$$

and the map $\Lambda(T'_h) \longrightarrow \Lambda(D_{Q'})$ considered before is just

$$\bigoplus_{i=1}^c \Lambda(\varrho^{(i)}).$$

Now, the quiver $Q^{(i)}$ has at least one layer less than Q , $i = 1, \dots, c$. So, the claim follows from the inductive hypothesis for $Q^{(i)}$, $i = 1, \dots, c$. $\square$

By the proposition, the quotient

$$\overline{\Lambda} := \Lambda(D_Q) / \Lambda(T_h) := \Lambda(D_Q) / \text{Im}(\Lambda(\varrho))$$

is torsion free, as well, and we get a splitting

$$(2.4.2.1) \quad \Lambda(D_Q) \cong \Lambda(T_h) \oplus \overline{\Lambda}.$$

As we explained in (1.2.0.2), we have isomorphisms

$$\Lambda(T_h) \otimes_{\mathbb{Z}} \mathbb{C}^\star \cong T_h \quad \text{and} \quad \Lambda(D_Q) \otimes_{\mathbb{Z}} \mathbb{C}^\star \cong D_Q.$$

Now, we may define the torus

$$\overline{T} := \overline{\Lambda} \otimes_{\mathbb{Z}} \mathbb{C}^\star.$$

Then, we have a splitting

$$(2.4.2.2) \quad D_Q \cong T_h \times \overline{T}$$

of tori. So, if we take the quotient of D_Q by the action of T_h , we will get $\overline{T}$ as the result. One of the problems of forming quotients in algebraic geometry is that the process does not necessarily commute with open embeddings. In Remark 3.1.4, we will explain that, in the present example, $\overline{T}$ is a dense open part of the moduli space we are considering and the projection $D_Q \longrightarrow \overline{T}$ is a model for the quotient over D_Q .

Example 2.4.3. a) Let us have another look at Example 2.3.4. The displayed equation shows

$$M_{p_4} = M_{p_1}^{-1} \cdot M_{p_2} \cdot M_{p_3}.$$

So, we can set

$$\overline{\Lambda} := \langle M_{p_1}, M_{p_2}, M_{p_3} \rangle.$$

b) The first action described in Example 1.3.6 is modeled on the level of one parameter subgroups by the homomorphism

$$\begin{aligned} \mathbb{Z} &\longrightarrow \mathbb{Z} \oplus \mathbb{Z} \\ k &\longmapsto (k, -k), \end{aligned}$$

and the second one by

$$\begin{aligned}\mathbb{Z} &\longrightarrow \mathbb{Z} \oplus \mathbb{Z} \\ k &\longmapsto (2 \cdot k, -2 \cdot k).\end{aligned}$$

The image of the first homomorphism spans a direct summand of $\mathbb{Z} \oplus \mathbb{Z}$, and the cokernel is isomorphic to $\mathbb{Z}$. It may be spanned by the class of, e.g., $(1, 0)$. The image of the second homomorphism does not split off, and the cokernel is isomorphic to $\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})$, and, so, has torsion. The cokernel is generated, for example, by the classes of $(1, 0)$ and $(1, -1)$. The latter defines a 2-torsion element.

3. STABLE, POLYSTABLE, AND REAL POINTS

Stable tuples in R_Q were determined in [1] and [3] in a more general setting. We first rephrase the stability criterion from those papers. Combining the result with Proposition 2.4.2, we get an algebraic proof and a generalization of a result from the paper [16]. Afterwards, we will also characterize the polystable tuples in R_Q . The final section applies this characterization and Proposition 2.4.2 in order to describe the moduli space of a neural network over the real numbers.

3.1. Stable points and a theorem of Meng et al. In geometric invariant theory, the Hilbert–Mumford criterion is an important tool for characterizing semistable and stable points by intrinsic properties (see [22], Theorem 1.5.1.2). It can be applied to representations of network quivers. We need to introduce some more notation. Let $Q = (V_1, \dots, V_L, E, s, t)$ be a quiver in layers. We set

$$A_{\text{in}} := \bigoplus_{u \in V_1} \mathbb{C}, \quad A_{\text{out}} := \bigoplus_{v \in V_L} \mathbb{C}, \quad \text{and} \quad A_h := \bigoplus_{w \in V_h} \mathbb{C}.$$

A tuple $\underline{r} = (r_e, e \in E) \in R_Q$ and a path $p = (e_1, \dots, e_t) \in \mathcal{P}$ define the $\mathbb{C}$ -linear map

$$\begin{aligned}\varphi_p^{\underline{r}} &: A_{\text{in}} \longrightarrow A_{\text{out}} \\ (\lambda_u, u \in V_1) &\longmapsto (\nu_v, v \in V_L)\end{aligned}$$

with

$$\nu_v := \begin{cases} r_1 \cdots r_t \cdot \lambda_{s'(p)}, & \text{if } v = t'(p) \\ 0, & \text{if } v \neq t'(p) \end{cases}.$$

Altogether, we get, for a tuple $\underline{r} = (r_e, e \in E) \in R_Q$, the $\mathbb{C}$ -linear map

$$\varphi^{\underline{r}} := \sum_{p \in \mathcal{P}} \varphi_p^{\underline{r}} : A_{\text{in}} \longrightarrow A_{\text{out}}.$$

Next, we consider $E_- := \{e \in E \mid s(e) \in V_1\}$ and $E_+ := \{e \in E \mid t(e) \in V_L\}$. With a similar construction as before, we associate $\mathbb{C}$ -linear maps

$$\varphi_-^{\underline{r}} : A_{\text{in}} \longrightarrow A_h \quad \text{and} \quad \varphi_+^{\underline{r}} : A_h \longrightarrow A_{\text{out}}$$

with E_- and E_+ , respectively, and a tuple $\underline{r} = (r_e, e \in E) \in R_Q$.

Finally, if we are given a tuple $\underline{r} = (r_e, e \in E) \in R_Q$, we call a subset $W \subset V_h$ *admissible*, if

$$\forall e \in E : \quad (s(e) \in W \wedge t(e) \in V_h \wedge r_e \neq 0) \implies t(e) \in W.$$

For an admissible subset $W \subset V_h$ and $w \in V_h$, we set

$$B_w := \begin{cases} \mathbb{C}, & \text{if } w \in W \\ \{0\}, & \text{if } w \notin W \end{cases},$$

and get the $\mathbb{C}$ -linear subspace

$$B_W := \bigoplus_{w \in V_h} B_w \subset A_h.$$

Theorem 3.1.1. i) A tuple $\underline{r} = (r_e, e \in E)$ is semistable if and only if the associated $\mathbb{C}$ -linear map $\varphi^{\underline{r}}$ is non-zero.

ii) A tuple $\underline{r} = (r_e, e \in E)$ is stable if and only if the following two properties hold true:

1. There is no admissible subset $W \neq \emptyset$ with $B_W \subset \ker(\varphi_+^{\underline{r}})$.
2. There is no admissible subset $W \neq V_h$ with $\text{im}(\varphi_-^{\underline{r}}) \subset B_W$.

Proof. This is [1], Lemma 4.1. In the proof, a procedure called “deframing” was applied. A direct proof without deframing is contained in [3]. $\square$

In our setting, we may reformulate the criterion as follows.

Theorem 3.1.2. Let $Q = (V_1, \dots, V_L, E, s, t)$ be a quiver in layers, satisfying Condition (N).

i) A tuple $\underline{r} = (r_e, e \in E)$ is semistable if and only if there exists a path $p = (e_1, \dots, e_t)$ in Q with $s'(p) \in V_1$, $t'(p) \in V_L$, and $r_{e_1} \cdots r_{e_t} \neq 0$.

ii) A tuple $\underline{r} = (r_e, e \in E)$ is stable if and only if, for every hidden vertex $v_0 \in V_h$, there exists a path $p = (e_1, \dots, e_t)$ in Q with $s'(p) \in V_1$, $t'(p) \in V_L$, and $r_{e_1} \cdots r_{e_t} \neq 0$ which passes through v_0 .

Proof. i) This is immediate from Theorem 3.1.1, i).

ii) Assume that $\underline{r} = (r_e, e \in E)$ is unstable. First, suppose that $W \subsetneq V_h$ is an admissible subset with $\text{im}(\varphi_-^{\underline{r}}) \subset B_W$. Let $v_0 \in V_h \setminus W$ be a vertex. For every path $p = (e_1, \dots, e_u)$ with $s'(p) \in V_1$ and $t'(p) = v_0$, we must have $r_{e_1} \cdots r_{e_u} = 0$. Second, let $\emptyset \neq W \subset V_h$ be an admissible subset with $B_W \subset \ker(\varphi_+^{\underline{r}})$. Pick a vertex $v_0 \in W$. It is obvious that, for every path $p = (e_1, \dots, e_u)$ in Q with $s'(p) = v_0$ and $t'(p) \in V_L$, we must have $r_{e_1} \cdots r_{e_u} = 0$.

Conversely, let $v_0 \in V_h$ be a hidden vertex for which the stated conditions fails. Note that this mean that one of the following properties holds true:

- For every path $p = (e_1, \dots, e_u)$ with $s'(p) \in V_1$ and $t'(p) = v_0$, we have $r_{e_1} \cdots r_{e_u} = 0$.
- For every path $p = (e_1, \dots, e_u)$ in Q with $s'(p) = v_0$ and $t'(p) \in V_L$, we have $r_{e_1} \cdots r_{e_u} = 0$.

If the first property is satisfied, then we let $W \subset V_h$ be the set of vertices $w \in V_h$ for which there exists a path $p = (e_1, \dots, e_u)$ in Q with $s'(p) \in V_1$, $t'(p) = w$, and $r_{e_1} \cdots r_{e_u} \neq 0$. This is an admissible subset with $v_0 \notin W$ and $\text{im}(\varphi_-^{\underline{r}}) \subset B_W$.

If the second condition holds, then we let $W \subset V_h$ be the set formed by v_0 and the vertices $w \in V_h$ for which there exists a path $p = (e_1, \dots, e_u)$ in Q with $s'(p) = v_0$, $t'(p) = w$, and $r_{e_1} \cdots r_{e_u} \neq 0$. It is readily checked that W is a non-empty admissible subset with $B_W \subset \ker(\varphi_+^{\underline{r}})$. In both cases, we conclude by Theorem 3.1.1. $\square$

Corollary 3.1.3. Let $Q = (V_1, \dots, V_L, E, s, t)$ be a quiver in layers, satisfying Condition (N). Then, $D_Q \subset R^s(Q)$.

Proof. A tuple $\underline{r} = (r_e, e \in E)$ satisfies, by definition, $r_e \neq 0, e \in E$. So, the result follows directly from Theorem 3.1.2, ii), and Exercise 2.3.8, a). $\square$

Remark 3.1.4. a) Theorem 4.2 in [1] also implies the existence of stable tuples. Together with Lemma 1.4.6, ii), we get $D_Q \subset R_Q^s$ from this.

b) Since D_Q is a T_h -invariant open subset of the subset $R_Q^s \subset R_Q$ of stable points, its image $\overline{U} \subset \mathcal{M}_Q$ is an open subset, and the induced map $D_Q \longrightarrow \overline{U}$ is a categorical quotient

for D_Q with respect to the induced T_h -action and an orbit space. Since $\mathcal{M}_Q$ is irreducible, $\overline{U}$ is also dense. Finally, by the universal property of categorical quotients, there is a unique isomorphism $\eta: \overline{T} \rightarrow \overline{U}$, such that the diagram

$$\begin{array}{ccc} D_Q & \longrightarrow & \overline{T} \\ & \searrow & \downarrow \eta \\ & & \overline{U} \end{array}$$

commutes. The unlabeled arrows represent the canonical quotient maps.

c) If $\underline{r} = (r_e, e \in E) \in R_Q$ is an unstable tuple, then there is a hidden vertex $v_0 \in V_h$, such that $r_{e_1} \cdots r_{e_t} = 0$, for every path $p = (e_1, \dots, e_t)$ in Q which connects a vertex in V_1 to a vertex in V_L and passes through v_0 . In terms of networks, this means that the vertex does not pass any information from the input to the output. So, it can be removed. We will describe the precise construction in Section 3.2.

Example 3.1.5. We continue Example 2.3.4 and Example 2.4.3, a). The torus $(\mathbb{C}^\star)^3 \cong \overline{T} = \overline{\Lambda} \otimes_{\mathbb{Z}} \mathbb{C}^\star$ acts on $\mathbb{C}^4$ by

$$\begin{aligned} (\mathbb{C}^\star)^3 \times \mathbb{C}^4 &\longrightarrow \mathbb{C}^4 \\ ((z_1, z_2, z_3), (a_1, a_2, a_3, a_4)) &\longmapsto (z_1 \cdot a_1, z_2 \cdot a_2, z_3 \cdot a_3, z_1^{-1} \cdot z_2 \cdot z_3 \cdot a_4). \end{aligned}$$

The quotient $\mathcal{M}_Q \subset \mathbb{C}^4$ is cut out by the equation $a_1 \cdot a_4 - a_2 \cdot a_3$. Note that the orbit of $(1, 1, 1, 1)$ under the above group action lies in $\mathcal{M}_Q$, and this embeds $\overline{T}$ as an open subset into $\mathcal{M}_Q$. Note that the equation of $\mathcal{M}_Q$ is invariant under the group action, and, therefore, the multiplication on $\overline{T}$ extends to an action of $\overline{T}$ on $\mathcal{M}_Q$.

Proposition 3.1.6. *Let $Q = (V_1, \dots, V_L, E, s, t)$ be a quiver in layers, satisfying Condition (N). Then, the dimension of the moduli space $\mathcal{M}_Q$ is $\#E - \#V_h$.*

Proof. Let $R_Q^s \subset R_Q$ be the non-empty open subset of stable tuples. The image $\mathcal{M}_Q^s$ of R_Q^s under the quotient map $R_Q \rightarrow \mathcal{M}_Q$ is a non-empty open subset. Since R_Q is irreducible, so are $\mathcal{M}_Q$ and $\mathcal{M}_Q^s$, and $\mathcal{M}_Q$ and $\mathcal{M}_Q^s$ have the same dimension. Furthermore, the fibers of the map $R_Q^s \rightarrow \mathcal{M}_Q^s$ are the orbits of stable tuples. For a stable tuple, the orbit has the same dimension as the acting group T_h . The group T_h obviously has dimension $\#V_h$. Moreover, the dimension of R_Q is $\#E$, and as before, $\dim(R_Q^s) = \dim(R_Q)$. By general properties of the dimension ([6], 10.1 Theorem), we find

$$\dim(\mathcal{M}_Q^s) = \dim(R_Q^s) - \dim(T_h) = \#E - \#V_h.$$

This concludes the proof. $\square$

Next, we explain the notation of the paper [16]. For this, we let $m := \#E$ be the number of arrows and enumerate the arrows, i.e., we write $E = \{e(1), \dots, e(m)\}$. Recall from (2.3.5.1) that $\mathcal{P}$ is the set of paths in Q starting at a vertex in V_1 and ending at a vertex in V_L . For a path $p \in \mathcal{P}$, we define the column vector $c_p \in \mathbb{F}_2^m$ whose j -th entry is one, if $e(j)$ appears in the path p , and zero otherwise. Let $s := \#\mathcal{P}$ and write $\mathcal{P} = \{p(1), \dots, p(s)\}$. The $(m \times s)$ -matrix

$$A := (c_{p(1)}, \dots, c_{p(s)}) \in \text{Mat}(m, s; \mathbb{F}_2)$$

is called the *structure matrix* of the network quiver Q .

Theorem 3.1.7. *The rank of the structure matrix A is $\#E - \#V_h$.*

In [16], this theorem is stated and proved for multilayer perceptrs (see loc. cit., Theorem 3.4).

Proof. We recall the splitting from (2.4.2.2). It leads to the short exact sequence

$$0 \longrightarrow X(\bar{T}) \longrightarrow X(D_Q) \xrightarrow{X(\varrho)} X(T_h) \longrightarrow 0.$$

According to Exercise 2.3.3, the monomials M_p , $p \in \mathcal{P}$, generate the kernel of $X(\varrho)$. By the above exact sequence, the latter is isomorphic to $X(\bar{T})$. Note that $X(\bar{T})$ is a free abelian group of rank

$$\dim(\bar{T}) = \dim(D_Q) - \dim(T_h) = \#E - \#V_h.$$

We can tensorize the above sequence over the integers with the field $\mathbb{F}_2 = \{0, 1\}$ and get, in particular, the surjection

$$X(D_Q) \otimes_{\mathbb{Z}} \mathbb{F}_2 \xrightarrow{\bar{X}(\varrho)} X(T_h) \otimes_{\mathbb{Z}} \mathbb{F}_2.$$

The classes $\bar{M}_p$, $p \in \mathcal{P}$, generate the kernel of $\bar{X}(\varrho)$. On the other hand, the enumeration of E gives the isomorphism $X(D_Q) \cong \mathbb{Z}^m$ from (1.2.0.1). We can also tensorize this isomorphism over the integers with the field $\mathbb{F}_2$. This leads to the map

$$X(D_Q) \longrightarrow X(D_Q) \otimes_{\mathbb{Z}} \mathbb{F}_2 \xrightarrow{\cong} \mathbb{F}_2^m.$$

This map sends M_p to c_p , $p \in \mathcal{P}$. It is now clear that the rank of the structure matrix A equals the dimension (over $\mathbb{F}_2$) of the kernel of $\bar{X}(\varrho)$ and, so, the dimension of $X(\bar{T}) \otimes_{\mathbb{Z}} \mathbb{F}_2$, i.e., the rank of $X(\bar{T})$. As we have seen above, that number equals $\#E - \#V_h$. $\square$

3.2. Polystable points. We let $Q = (V_1, \dots, V_L, E, s, t)$ be a quiver in layers which satisfies Condition (N) and $\underline{r} = (r_e, e \in E) \in R_Q$. Then, the *support* of $\underline{r}$ is the subquiver $\text{Supp}(\underline{r}) = (V'_1, \dots, V'_L, E', s', t')$ of Q that is obtained by first removing all arrows $e \in E$ with $r_e = 0$ and then all isolated vertices from the resulting quiver. Using this concept, we obtain the following characterization of polystable tuples (compare Page 10).

Proposition 3.2.1. *In the above situation, the tuple $\underline{r} = (r_e, e \in E)$ is polystable if and only if the connected components of $\text{Supp}(\underline{r})$ satisfy Condition (M).*

Proof. First, assume that $\underline{r}$ is polystable and suppose $v_0 \in V'_2 \sqcup \dots \sqcup V'_{L-1}$ is a source. For a complex number $t \in \mathbb{C}^\star$, introduce

$$\lambda(t) := (\lambda_v^t, v \in V_h) \in T_h \quad \text{with} \quad \lambda_v^t := \begin{cases} t, & \text{if } v = v_0 \\ 1, & \text{if } v \neq v_0 \end{cases}.$$

Then, for $\underline{r}(t) := (r_e(t), e \in E) := \lambda(t) \cdot \underline{r}$, we have

$$r_e(t) = \begin{cases} t^{-1} \cdot r_e, & \text{if } s(e) = v_0 \\ r_e, & \text{if } s(e) \neq v_0 \end{cases}, \quad t \in \mathbb{C}^\star.$$

We set

$$\underline{r}' := (r'_e, e \in E) := \lim_{t \rightarrow \infty} \lambda(t) \cdot \underline{r}.$$

Then, $\underline{r}'$ is contained in the closure of the orbit of $\underline{r}$. Since we assume that the latter is closed, $\underline{r}'$ is actually contained in the orbit of $\underline{r}$. But, then, $r'_e = 0$, for every arrow $e \in E$ with $s(e) = v_0$, implies $r_e = 0$, for every arrow $e \in E$ with $s(e) = v_0$. So, v_0 is an isolated vertex of $\text{Supp}(\underline{r})$, and this contradicts our construction. A similar argument works, if $v_0 \in V'_2 \sqcup \dots \sqcup V'_{L-1}$ is a sink.

Now, let $\underline{r} \in R_Q$ be a tuple, such that the connected components of $Q' := \text{Supp}(\underline{r})$ satisfy Condition (M), and let $\underline{r}' \in R_Q$ be a tuple which is contained in the closure of the orbit of $\underline{r}$. Then, $\text{Supp}(\underline{r}')$ is a subquiver of Q' . We may view $R_{Q'}$ as a closed subset of R_Q . It contains both $\underline{r}$ and $\underline{r}'$ and is invariant under the action of T_h . More precisely, the group

$$T_h^c := \bigtimes_{v \in V_h \setminus V'_h} \mathbb{C}^\star, \quad V'_h := V'_2 \sqcup \cdots \sqcup V'_{L-1},$$

acts trivially on R_Q , and we may interpret the image of $R_{Q'}$ in $\mathcal{M}_Q$ as the quotient of $R_{Q'}$ by the action of

$$T'_h := \bigtimes_{v \in V'_h} \mathbb{C}^\star.$$

Altogether, we arrive at the commutative diagram

$$(3.2.1.1) \quad \begin{array}{ccc} R_{Q'} & \longrightarrow & R_Q \\ \downarrow & & \downarrow \\ \mathcal{M}_{Q'} & \longrightarrow & \mathcal{M}_Q \end{array}$$

in which the horizontal arrows are closed embeddings. Since the connected components of Q' satisfy Condition (M), we may rearrange the layers of Q' in such a way that its connected components satisfy Condition (N) (see Remark 2.1.4, a). However, the group T'_h and its action on $R_{Q'}$ depend only on the set V'_h of hidden vertices. This means that Corollary 3.1.3 applies to Q' as well, so that $\underline{r}$ is stable with respect to the T'_h -action on $R_{Q'}$. Our assumptions imply that $\underline{r}$ and $\underline{r}'$ are both mapped to the same point in $\mathcal{M}_{Q'}$. The stability of $\underline{r}$ now implies that $\underline{r}'$ lies in the T'_h -orbit of $\underline{r}$, and the latter equals the T_h -orbit of $\underline{r}$. We infer that the T_h -orbit of $\underline{r}$ in R_Q is closed. $\square$

Remark 3.2.2. Let $\underline{r} = (r_e, e \in E) \in R_Q$ be any element. The above proof gives a constructive procedure for determining an element $\underline{r}^P = (r_e^P, e \in E)$, such that the orbit of $\underline{r}^P$ is the unique closed orbit that is contained in the closure of the orbit of $\underline{r}$. In fact, let $v_0 \in V_2 \sqcup \cdots \sqcup V_{L-1}$ be a source or a sink of $\text{Supp}(\underline{r})$ and introduce $\underline{r}^1 := (r_e^1, e \in E)$ by

$$r_e^1 := \begin{cases} 0, & \text{if } s(e) = v_0 \\ r_e, & \text{if } s(e) \neq v_0 \end{cases}, \quad e \in E,$$

if v_0 is a source, and by

$$r_e^1 := \begin{cases} 0, & \text{if } t(e) = v_0 \\ r_e, & \text{if } t(e) \neq v_0 \end{cases}, \quad e \in E,$$

if v_0 is a sink. Then, replace $\underline{r}$ by $\underline{r}^1$ and iterate the construction. At some stage, you will arrive at a tuple $\underline{r}^P$, such that its support is empty or its connected components satisfy Condition (M). By the proposition, the orbit of $\underline{r}^P$ is closed,¹⁵ and the arguments in the proof show that $\underline{r}^P$ is contained in the closure of the orbit of $\underline{r}$. We will encounter this construction in Example 3.3.6, b).

¹⁵The case of the empty quiver corresponds to $\underline{r}^P = 0$.

3.3. Real points. In applications of neural networks, the set-up is real. To be more precise, given a network quiver $Q = (V_1, \dots, V_L, E, s, t)$, the weights attached to the arrows $e \in E$ are real. This means that

$$R_Q(\mathbb{R}) := \bigoplus_{e \in E} \mathbb{R}$$

should be the parameter space for the weights. We would then look at the groups

$$T_Q(\mathbb{R}) := \bigtimes_{v \in V} \mathbb{R}^\star, \quad T_h(\mathbb{R}) := \bigtimes_{v \in V_h} \mathbb{R}^\star, \quad \text{and} \quad D_Q(\mathbb{R}) := \bigtimes_{e \in E} \mathbb{R}^\star$$

and the actions analogous to those considered in Section 2.2. Recall that $\mathcal{P}$ is the set of paths in Q that connect a vertex in V_1 to a vertex in V_L . Set $s := \#\mathcal{P}$ and write $\mathcal{P} = \{p(1), \dots, p(s)\}$. Before, we considered the regular map

$$\begin{aligned} \psi: R_Q &\longrightarrow \mathbb{C}^s \\ (x_e, e \in E) &\longmapsto (M_{p(1)}, \dots, M_{p(s)}). \end{aligned}$$

Obviously, under this map, real points go to real points, so that we get an induced map

$$\psi_{\mathbb{R}}: R_Q(\mathbb{R}) \longrightarrow \mathbb{R}^s.$$

The relations among the monomials $M_{p(1)}, \dots, M_{p(s)}$ are of the form "monomial=monomial" ([23], Lemma 4.3). Those define the equations of $\mathcal{M}_Q$ in $\mathbb{C}^s$. We let $\mathcal{M}_Q(\mathbb{R})$ be the set of real solutions to those equations, i.e.,

$$\mathcal{M}_Q(\mathbb{R}) := \mathcal{M}_Q \cap \mathbb{R}^s.$$

We can ask the following questions about the "quotient" map

$$q_{\mathbb{R}}: R_Q(\mathbb{R}) \longrightarrow \mathcal{M}_Q(\mathbb{R}) :$$

- Is $q_{\mathbb{R}}$ surjective?
- Are the points in the image of $q_{\mathbb{R}}$ in bijection to the closed $T_h(\mathbb{R})$ -orbits in $R_Q(\mathbb{R})$?

We will show that both questions have positive answers (Theorem 3.3.4), so that $q_{\mathbb{R}}$ is the reasonable quotient to look at in the real case. Let us first discuss two examples from the more general area of quotient problems which illustrate that the answers to the above questions are not obvious.

Example 3.3.1. a) We study the action

$$\begin{aligned} \{\pm 1\} \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (\varepsilon, \lambda) &\longmapsto \varepsilon \cdot \lambda. \end{aligned}$$

It is easy to see that the ring of polynomials that are invariant under this action is generated by x^2 . This implies that the real quotient map is modeled by

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^2. \end{aligned}$$

The image of this map is $\mathbb{R}_{\geq 0} = \{y \in \mathbb{R} \mid y \geq 0\}$, and we see that the quotient map is not surjective.

b) We come back to the second action considered in Example 1.3.6. The real quotient map is modeled on

$$\begin{aligned} \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto x \cdot y. \end{aligned}$$

The orbits are $\{0\}$, the positive x -axis, the negative x -axis, the positive y -axis, the negative y -axis, the positive branch of the hyperbola $\{x \cdot y = \lambda\}$, and the negative branch of the hyperbola $\{x \cdot y = \lambda\}$, $\lambda \in \mathbb{R}^*$. The first five orbits are all mapped to zero. For $\lambda \in \mathbb{R}^*$, both the positive and the negative branch are mapped to λ . So, the quotient does not parameterize the closed orbits.

Remark 3.3.2. a) It is a general fact that the intersection of a closed $\mathbb{C}$ -orbit with the set of real points is the union of finitely many closed $\mathbb{R}$ -orbits (see [7], 2.3. Proposition).

b) The set of closed $\mathbb{R}$ -orbits carries the structure of a Hausdorff space [21]. We have an induced (proper) map of that space to the set of real points of the quotient. In Part a) of the above example, this is the inclusion $\mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$. In the second example, the quotient is the union of two copies of $\mathbb{R}$, intersecting at zero, and the map is two-to-one over $\mathbb{R}^*$ and one-to-one over $\{0\}$.

For the studying the separation properties of the real quotient map, we follow the strategy of the proof of Proposition 3.2.1. First note that, intrinsically, we have

$$D_Q(\mathbb{R}) = \Lambda(D_Q) \otimes_{\mathbb{Z}} \mathbb{R} \quad \text{and} \quad T_h(\mathbb{R}) = \Lambda(T_h) \otimes_{\mathbb{Z}} \mathbb{R}.$$

Likewise,

$$\bar{T}(\mathbb{R}) = \bar{\Lambda} \otimes_{\mathbb{Z}} \mathbb{R}.$$

Clearly, $D_Q(\mathbb{R})$, $T_h(\mathbb{R})$, and $\bar{T}(\mathbb{R})$ are the sets of real points of D_Q , T_h , and $\bar{T}$, respectively. Because of the splitting (2.4.2.1), we have the commutative diagram

$$\begin{array}{ccc} D_Q(\mathbb{R}) & \longrightarrow & D_Q \\ \downarrow & & \downarrow \\ \bar{T}(\mathbb{R}) & \longrightarrow & \bar{T}. \end{array}$$

By our previous discussion, we also have the commutative diagram

$$\begin{array}{ccc} D_Q(\mathbb{R}) & \longrightarrow & R_Q(\mathbb{R}) \\ \downarrow & & \downarrow q_{\mathbb{R}} \\ \bar{T}(\mathbb{R}) & \longrightarrow & \mathcal{M}_Q(\mathbb{R}). \end{array}$$

Remark 3.3.3. The quotient map $p: D_Q \rightarrow \bar{T}$ induces the homomorphism

$$\begin{aligned} X(p): X(\bar{T}) &\longrightarrow X(D_Q) \\ \chi &\longmapsto p \circ \chi \end{aligned}$$

on the level of groups of characters (see Section 1.2). We may pick characters $\chi_1, \dots, \chi_s$, such that

$$X(\chi_j) = M_{p_j}, \quad j = 1, \dots, s.$$

Then,

$$\begin{aligned} \bar{T} &\longrightarrow \mathcal{M}_Q \\ \tau &\longmapsto (\chi_1(\tau), \dots, \chi_s(\tau)). \end{aligned}$$

The above diagram tells us that the restriction of the quotient map $q_{\mathbb{R}}: R_Q(\mathbb{R}) \rightarrow \mathcal{M}_Q(\mathbb{R})$ to $D_Q(\mathbb{R})$ is injective and has the image

$$\mathcal{M}_Q(\mathbb{R}) \cap \bar{U}.$$

For the latter, recall Remark 3.1.4, b).

Theorem 3.3.4. *Let $Q = (V_1, \dots, V_L, E, s, t)$ be a quiver in layers which satisfies Condition (N). Then, the real quotient map $q_{\mathbb{R}}: R_Q(\mathbb{R}) \rightarrow \mathcal{M}_Q(\mathbb{R})$ is surjective and separates closed $T_h(\mathbb{R})$ -orbits in $R_Q(\mathbb{R})$.*

Proof. Let us first check the claim about the separation of closed orbits. Let $\underline{r}^i = (r_e^i, e \in E) \in R_Q(\mathbb{R})$, $i = 1, 2$, be two real tuples whose $T_h(\mathbb{R})$ -orbits in $R_Q(\mathbb{R})$ are closed and assume $q_{\mathbb{R}}(\underline{r}^1) = q_{\mathbb{R}}(\underline{r}^2)$. We point out that the argument in the first half of the proof of Proposition 3.2.1 can be performed exactly in the same way over the real numbers. So, the connected components of the supports $\text{Supp}(\underline{r}^1)$ and $\text{Supp}(\underline{r}^2)$ must satisfy Condition (M). So, we infer that the T_h -orbits of $\underline{r}^1$ and $\underline{r}^2$ are closed in R_Q without having to recur to the general theory. Since $\underline{r}^1$ and $\underline{r}^2$ map to the same point in $\mathcal{M}_Q$ and are both polystable, they must lie in the same T_h -orbit. So, we must actually have $\text{Supp}(\underline{r}^1) = \text{Supp}(\underline{r}^2)$. Let's denote this support quiver by Q' . Now, we can invoke Diagram (3.2.1.1). All the maps in that diagram are defined over the real numbers, so that we get the induced diagram

$$(3.3.4.1) \quad \begin{array}{ccc} R_{Q'}(\mathbb{R}) & \longrightarrow & R_Q(\mathbb{R}) \\ q'_{\mathbb{R}} \downarrow & & \downarrow q_{\mathbb{R}} \\ \mathcal{M}_{Q'}(\mathbb{R}) & \longrightarrow & \mathcal{M}_Q(\mathbb{R}) \end{array} \quad .$$

Now, both $\underline{r}^1$ and $\underline{r}^2$ lie in $D_{Q'}(\mathbb{R}) \subset R_{Q'}(\mathbb{R})$. As before, we see that the fact that the connected components of Q' satisfy Condition (M) suffices to infer that the quotient map $q'_{\mathbb{R}}: R_{Q'}(\mathbb{R}) \rightarrow \mathcal{M}_{Q'}(\mathbb{R})$ separates orbits in $D_{Q'}(\mathbb{R})$. So, $\underline{r}^1$ and $\underline{r}^2$ belong to the same $T'_h(\mathbb{R})$ -orbit which is the same as the corresponding $T_h(\mathbb{R})$ -orbit.

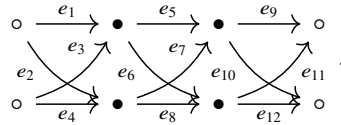
Finally, we turn to the surjectivity of $q_{\mathbb{R}}$. Let $y \in \mathcal{M}_Q(\mathbb{R})$ and $\underline{r} \in R_Q$ a polystable tuple with $\psi(\underline{r}) = y$. Then, the connected components of the quiver $Q' := \text{Supp}(\underline{r})$ satisfy Condition (M). We use Diagram (3.3.4.1). Note that $y \in \mathcal{M}_{Q'}(\mathbb{R}) \cap \underline{U}'$. By the previous discussion, there exists a real tuple $\underline{r}' \in D_{Q'}(\mathbb{R})$ with $q'_{\mathbb{R}}(\underline{r}') = y$. This finishes the proof. $\square$

Remark 3.3.5. a) If $\underline{r} \in R_Q$ is a polystable tuple with $\psi(\underline{r}) \in \mathcal{M}_Q(\mathbb{R})$, then it is easy to find an element in $g \in T_h$, such that $g \cdot \underline{r} \in R_Q(\mathbb{R})$. We will illustrate this in the following example.

b) If $\underline{r} \in R_Q$ is tuple with $\psi(\underline{r}) \in \mathcal{M}_Q(\mathbb{R})$ which is not polystable, then it might not be possible to find a real tuple in the T_h -orbit of $\underline{r}$. See Example 3.3.6, b), for this. In this case, we need to apply the procedure in Remark 3.2.2 to replace $\underline{r}$ by a polystable tuple.

c) The interested reader may find a more sophisticated approach to real points on moduli spaces of (unframed) quiver representations in [14].

Example 3.3.6. Let $Q = (V_1, V_2, V_3, V_4, E, s, t)$ be the following multilayer perceptron:



For convenience, we will write r_i instead of r_{e_i} , $i = 1, \dots, 12$.

a) Let $\underline{r} = (r_i, i = 1, \dots, 12)$ be a tuple in which all entries are non-zero which defines a real point in the moduli space. We rescale at the top left hidden vertex by r_1^{-1} and at the

bottom left hidden vertex by r_2^{-1} . We then get

$$\underline{r}' = \left(1, 1, \frac{r_3}{r_1}, \frac{r_4}{r_2}, r_1 \cdot r_5, r_1 \cdot r_6, r_2 \cdot r_7, r_2 \cdot r_8, r_9, r_{10}, r_{11}, r_{12} \right).$$

We assume that $r_1 \cdot r_5 \cdot r_9$ and $r_3 \cdot r_5 \cdot r_9$ are both real. So, r_3/r_1 is also real. Rescaling at the second layer yields

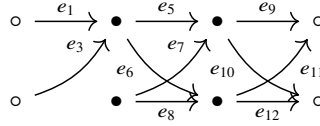
$$\underline{r}'' = \left(1, 1, \frac{r_3}{r_1}, \frac{r_4}{r_2}, 1, 1, \frac{r_2 \cdot r_7}{r_1 \cdot r_5}, \frac{r_2 \cdot r_8}{r_1 \cdot r_6}, r_1 \cdot r_5 \cdot r_9, r_1 \cdot r_5 \cdot r_{10}, r_1 \cdot r_6 \cdot r_{11}, r_1 \cdot r_6 \cdot r_{12} \right).$$

This will then be a real tuple in the orbit of $\underline{r}$.

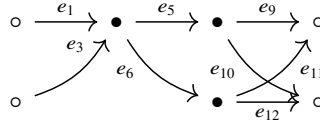
b) Consider $\underline{r} = (r_i, i = 1, \dots, 12)$ with

$$r_2 = r_4 = 0, \quad r_1 = r_3 = r_5 = r_6 = r_7 = \sqrt{-1}, \quad r_8 = 1 + \sqrt{-1}, \quad \text{and } r_9 = r_{10} = r_{11} = r_{12} = 1.$$

The tuple $\underline{r}$ defines a real point in the moduli space. It is not polystable. Its support is



and does not satisfy Condition (M). We can only rescale with real numbers at the second hidden layer in order to keep the weights at the arrows going to the output layer real, and we cannot rescale at the bottom left hidden vertex in such a way that we make both $\sqrt{-1}$ and $1 + \sqrt{-1}$ real. So, there is no real tuple in the orbit of $\underline{r}$. We must replace $\underline{r}$ by the tuple which assigns the weight zero to the arrows e_7 and e_8 while keeping the other weights. This has support



and is polystable.

REFERENCES

- [1] M.A. Armenta, Th. Brüstle, S. Hassoun, M. Reineke, *Double framed moduli spaces of quiver representations*, Linear Algebra App. **650** (2022), 98–131, <https://doi.org/10.1016/j.laa.2022.05.018>.
- [2] M.A. Armenta, P.-M. Jodoin, *The representation theory of neural networks*, Mathematics **2021**, 9, 3216, <https://doi.org/10.3390/math9243216>.
- [3] M.A. Armenta, A.H.W. Schmitt, *The Hilbert–Mumford criterion for representations of network quivers*, in U. Vandandoo et al. (eds.), *Proceedings of the International Conference on Applied Sciences and Engineering (ICASE 2023)*, 5–12, Atlantis Press, Highlights in Engineering, vol. 22, 2023, https://doi.org/10.2991/978-94-6463-330-6_2.
- [4] M.F. Atiyah, N.J. Hitchin, V.G. Drinfel'd, Yu.I. Manin, *Construction of instantons*, Phys. Lett. A **65** (1978), no. 3, 185–7, [https://doi.org/10.1016/0375-9601\(78\)90141-X](https://doi.org/10.1016/0375-9601(78)90141-X).
- [5] M. Bader, *Quivers, geometric invariant theory, and moduli of linear dynamical systems*, Linear Algebra Appl. **428** (2008), no. 11–12, 2424–54, <https://doi.org/10.1016/j.laa.2007.11.027>.
- [6] A. Borel, *Linear algebraic groups*, second edition, Grad. Texts in Math., vol. 126, Springer-Verlag, New York, 1991, xii+288 pp, <https://doi.org/10.1007/978-1-4612-0941-6>.
- [7] A. Borel, Harish-Chandra, *Arithmetic subgroups of algebraic groups*, Ann. Math. (2) **75** (1962), 485–535, <https://doi.org/10.2307/1970210>.
- [8] J.A. Dieudonné, J.B. Carrell, *Invariant theory, old and new*, Academic Press, New York–London, 1971, viii+85 pp.

- [9] W. Fulton, *Introduction to toric varieties*, the William H. Roever Lectures in Geometry, Annals of Mathematics Studies, vol. 131, Princeton University Press, Princeton, NJ, 1993, xii+157 pp, <https://doi.org/10.1515/9781400882526>.
- [10] M. Halic, M.-S. Stupariu, *Rings of invariants for representations of quivers*, C.R. Math. Acad. Sci. Paris **340** (2005), no. 2, 135-40, <https://doi.org/10.1016/j.crma.2004.12.012>.
- [11] R. Hartshorne, *Algebraic geometry*, corrected third printing, Grad. Texts in Math., vol. 52, Springer-Verlag, New York-Heidelberg, 1983, xvi+496 pp, <https://doi.org/10.1007/978-1-4757-3849-0>.
- [12] D. Hilbert, *Ueber die Theorie der algebraischen Formen*, Math. Ann. **36** (1890), no. 4, 473-534, <https://doi.org/10.1007/BF01208503>.
- [13] D. Hilbert, *Ueber die vollen Invariantensysteme*, Math. Ann. **42** (1893), 313-73, <https://doi.org/10.1007/BF01444162>.
- [14] V. Hoskins, F. Schaffhauser, *Rational points of quiver moduli spaces*, Ann. Inst. Fourier **70** (2020), no. 3, 1259-305, <https://doi.org/10.5802/aif.3334>.
- [15] S. Lang, *Algebra*, revised third edition, Grad. Texts in Math., vol. 211, Springer-Verlag, New York, 2002, xvi+914 pp, <https://doi.org/10.1007/978-1-4613-0041-0>.
- [16] Q. Meng, Sh. Zheng, H. Zhang, W. Chen, Q. Ye, Zh.-M. Ma, N. Yu, T.-Y. Liu, *G-SGD: Optimizing ReLU neural networks in its positively scale-invariant space*, in *International Conference on Learning Representations*, New Orleans, LA, USA, 2019, <http://openreview.net/pdf?id=SyxfEn09Y7>.
- [17] D. Mumford, *The red book of varieties and schemes*, second, expanded edition, includes the Michigan lectures (1974) on curves and their Jacobians, with contributions by Enrico Arbarello, Lecture Notes in Math., vol. 1358, Springer-Verlag, Berlin, 1999, x+306 pp, <https://doi.org/10.1007/b62130>.
- [18] H. Nakajima, *Varieties associated with quivers*, in *Representation theory of algebras and related topics* (Mexico City, 1994), 139-157, CMS Conf. Proc., vol. 19, Amer. Math. Soc., Providence, RI, 1996.
- [19] P.E. Newstead, *Introduction to moduli problems and orbit spaces*, reprint of the 1978 original printed in new typeset, Lectures on Mathematics and Physics, Mathematics, Tata Institute of Fundamental Research, vol. 51, Narosa, New Delhi, published for the Tata Institute of Fundamental Research, 2012, x+153 pp.
- [20] M. Reineke, *Framed quiver moduli, cohomology, and quantum groups*, J. Algebra **320** (2008), no. 1, 94-115, <https://doi.org/10.1016/j.jalgebra.2008.01.025>.
- [21] R.W. Richardson, P.J. Slodowy, *Minimum vectors for real reductive algebraic groups*, J. Lond. Math. Soc. (2) **42** (1990), no. 3, 409-29, <https://doi.org/10.1112/jlms/s2-42.3.409>.
- [22] A.H.W. Schmitt, *Geometric invariant theory and decorated principal bundles*, Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zürich, 2008, viii+389 pp, <https://doi.org/10.4171/065>.
- [23] B. Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series, vol. 8, American Mathematical Society, Providence, RI, 1996, xii+162 pp, <https://doi.org/10.1090/ulect/008>.

UNIVERSITÉ DE SHERBROOKE, INSTITUT QUANTIQUE, LOCAL D9-1025, 2500 BOULEVARD DE L'UNIVERSITÉ, SHERBROOKE, QUÉBEC, J1K 2R1, CANADA

Email address: Marco.Armenta@USherbrooke.ca

FREIE UNIVERSITÄT BERLIN, INSTITUT FÜR MATHEMATIK, ARNIMALLEE 3, D-14195 BERLIN, GERMANY

Email address: jorgeesqara@zedat.fu-berlin.de

FREIE UNIVERSITÄT BERLIN, INSTITUT FÜR MATHEMATIK, ARNIMALLEE 3, D-14195 BERLIN, GERMANY

Email address: alexander.schmitt@fu-berlin.de