

# Notes on the Geometry of Deep Linear Networks

February 19, 2026

## 1 Deep Linear Networks

**Definition 1.1.** A deep linear network (DLN) is parameter-function map defined by a composition of linear transformations. For input space  $\mathbb{R}^{d_0}$ , output space  $\mathbb{R}^{d_L}$ , and  $L$  layers with weight matrices  $W_1 \in \mathbb{R}^{d_1 \times d_0}, W_2 \in \mathbb{R}^{d_2 \times d_1}, \dots, W_L \in \mathbb{R}^{d_L \times d_{L-1}}$ , the DLN is given by:

$$f(x; W_1, W_2, \dots, W_L) = W_L W_{L-1} \cdots W_1 x,$$

where  $x \in \mathbb{R}^{d_0}$  is an input vector.

Note that the DLN  $f$  is a polynomial function of the entries of the weight matrices and linear in the input  $x$ .

**Lemma 1.2.** The space of parameter  $\Omega = \mathbb{R}^{d_1 \times d_0} \times \mathbb{R}^{d_2 \times d_1} \times \dots \times \mathbb{R}^{d_L \times d_{L-1}}$  is a differentiable manifold of dimension  $\sum_{l=1}^L d_l d_{l-1}$ . Furthermore, the space  $\Omega$  has a scalar product:

$$\langle U, V \rangle := \sum_{l=1}^L \langle U_l, V_l \rangle_F = \sum_{l=1}^L \text{Tr}(U_l^\top V_l)$$

The differential is induced by the differential on the Euclidean spaces  $\mathbb{R}^{d_l \times d_{l-1}}$ . The scalar product is induced by the Frobenius scalar product on each of the Euclidean spaces  $\mathbb{R}^{d_l \times d_{l-1}}$ .

**Definition 1.3.** The empirical loss function of a DLN is defined as:

$$\mathcal{L}_N(W_1, W_2, \dots, W_L) := \frac{1}{2N} \sum_{i=1}^N \|y_i - f(x_i; W_1, W_2, \dots, W_L)\|_F^2,$$

where  $\{(x_i, y_i)\}_{i=1}^N$  is a dataset of input-output pairs. The population loss is the expected loss over the data distribution. Assuming that the input data is whitened i.e.  $x \sim N(0, I)^1$  then the square norm of the gap between the product of the weights and the target linear map.

$$\mathcal{L}(W_1, W_2, \dots, W_L) := \frac{1}{2} \|M - W\|_F^2 = \frac{1}{2} \|M - W_L W_{L-1} \cdots W_1\|_F^2$$

We can derive the differential of the DLN map by using the usual product rule for differentials.

---

<sup>1</sup>This is a common assumption in the DLN literature

**Proposition 1.4.** Fix an input  $x \in \mathbb{R}^{d_0}$  and write

$$h_{l-1} := W_{<l}x, \quad W_{>l} := W_L \cdots W_{l+1}.$$

For a displacement  $U = (U_l)_l \in \Omega$ , the function  $f$  has a differential at any point  $(W_1, W_2, \dots, W_L) \in \Omega$  given by:

$$df[U] = \sum_{l=1}^L W_{>l} U_l h_{l-1}$$

From the differential, we can derive the Jacobian of the function  $f$ , by vectorizing and using the identity  $\text{vec}(AUB) = (B^\top \otimes A) \text{vec}(U)$ ,

**Lemma 1.5** (Jacobian from the differential). Let  $\theta := (\text{vec}(W_1)^\top, \dots, \text{vec}(W_L)^\top)^\top \in \mathbb{R}^P$  with  $P = \sum_{l=1}^L d_l d_{l-1}$ , and similarly  $\text{vec}(U) := (\text{vec}(U_1)^\top, \dots, \text{vec}(U_L)^\top)^\top$ . By vectorizing the differential, we obtain:

$$\text{vec}(df[U]) = \sum_{l=1}^L (h_{l-1}^\top \otimes W_{>l}) \text{vec}(U_l).$$

By definition of the Jacobian  $J(\theta; x) \in \mathbb{R}^{d_L \times P}$  (the unique matrix such that  $\text{vec}(df[U]) = J(\theta; x) \text{vec}(U)$ ), we obtain the block form

$$J(\theta; x) = [J_1(x) \mid J_2(x) \mid \cdots \mid J_L(x)], \quad J_l(x) \in \mathbb{R}^{d_L \times (d_l d_{l-1})}.$$

For each layer  $l = 1, \dots, L$ ,

$$J_l(x) = h_{l-1}^\top \otimes W_{>l}$$

so that, for any perturbation  $U_l$ ,

$$\text{vec}(W_{>l} U_l h_{l-1}) = J_l(x) \text{vec}(U_l).$$

**Lemma 1.6** (Differential of the loss). Let

$$\mathcal{L}(W_1, \dots, W_L) = \mathbb{E}[\langle y - f, y - f \rangle], \quad r := y - f(W; x),$$

where the differential of  $f$  is

$$df[U] = \sum_{k=1}^L W_{>k} U_k h_{k-1}, \quad h_{k-1} := W_{<k} x.$$

Then, for any displacement  $U = \{U_l\}_{l=1}^L$ , the differential of  $\mathcal{L}$  is

$$d\mathcal{L}[U] = \sum_{l=1}^L \langle U_l, -\mathbb{E}[W_{>l}^\top (y - f) h_{l-1}^\top] \rangle_F.$$

Equivalently, the gradient of  $\mathcal{L}$  with respect to  $W_l$  is

$$\nabla_{W_l} \mathcal{L} = -\mathbb{E}[W_{>l}^\top (y - f) h_{l-1}^\top].$$

**Derivation.** Start from the definition of  $\mathcal{L}$ :

$$\mathcal{L} = \mathbb{E}[\langle y - f, y - f \rangle] = \mathbb{E}[\langle r, r \rangle], \quad r := y - f.$$

Taking the differential gives

$$d\mathcal{L}[U] = \mathbb{E}[\langle r, dr[U] \rangle] = -\mathbb{E}[\langle r, df[U] \rangle],$$

since  $dr[U] = -df[U]$ . Substituting the expression of  $df[U]$ ,

$$df[U] = \sum_{k=1}^L W_{>k} U_k h_{k-1},$$

yields

$$d\mathcal{L}[U] = -\mathbb{E} \left[ \sum_{k=1}^L \langle r, W_{>k} U_k h_{k-1} \rangle \right].$$

Using the identity  $\langle AUB, v \rangle = \langle U, A^\top v B^\top \rangle_F$  gives

$$d\mathcal{L}[U] = \sum_{k=1}^L \langle U_k, -\mathbb{E}[W_{>k}^\top r h_{k-1}^\top] \rangle_F.$$

Identifying coefficients of  $U_k$  yields the stated gradient.

**Corollary 1.7** (Population loss case). If  $y = Mx$  and  $\Sigma_x = \mathbb{E}[xx^\top]$ , then with  $\Delta := M - W$  we have

$$\boxed{\nabla_{W_l} \mathcal{L} = -W_{>l}^\top \Delta \Sigma_x W_{<l}^\top.}$$

In the whitened input case  $\Sigma_x = I$ ,

$$\boxed{\nabla_{W_l} \mathcal{L} = -W_{>l}^\top \Delta W_{<l}^\top.}$$

**Derivation.** Since  $y - f = \Delta x$  and  $h_{l-1} = W_{<l} x$ ,

$$\mathbb{E}[(y - f) h_{l-1}^\top] = \mathbb{E}[\Delta x x^\top W_{<l}^\top] = \Delta \Sigma_x W_{<l}^\top,$$

and substituting this identity into the gradient formula of the lemma gives the result.

**Remark 1.8** (Use of the Riesz representation theorem). The differential  $d\mathcal{L}$  is a linear functional on the parameter space:

$$d\mathcal{L} : \mathbb{R}^P \rightarrow \mathbb{R}, \quad d\mathcal{L}(d\theta) = -\langle r, J(x) d\theta \rangle.$$

By the Riesz representation theorem in Euclidean space, there exists a unique vector  $\nabla_\theta \mathcal{L} \in \mathbb{R}^P$  such that  $d\mathcal{L}(d\theta) = \langle \nabla_\theta \mathcal{L}, d\theta \rangle$  for all  $d\theta$ . Using the adjoint identity  $\langle a, Jb \rangle = \langle J^\top a, b \rangle$ , we obtain

$$d\mathcal{L}(d\theta) = \langle -J(x)^\top r, d\theta \rangle,$$

hence

$$\boxed{\nabla_\theta \mathcal{L} = -J(x)^\top r.}$$

**Definition 1.9.** A point  $(W_1, W_2, \dots, W_L)$  is a critical point of the loss  $\mathcal{L}$  if and only if  $\nabla_{W_l} \mathcal{L} = 0$  for all  $l = 1, 2, \dots, L$ . There are four kinds of critical points:

- Global minimum:  $\mathcal{L}(W^*) \leq \mathcal{L}(W) \forall W \in \Omega$ .
- Local minimum: there exist an open set  $\mathcal{U}$  of  $\Omega$  such that  $\mathcal{L}(W^*) \leq \mathcal{L}(W') \forall W' \in \mathcal{U}$ .
- Saddle points:  $\nabla_{W_l} \mathcal{L} = 0 \forall l$  and  $W_l$  is neither a local nor a global minimum.
- A saddle point is **strict** if the Hessian has a strictly negative eigenvalue and **non-strict** otherwise (i.e. there exists at least one zero eigenvalue of the Hessian).

The following lemma simply follows from the definition of the gradient of the loss.

**Lemma 1.10.** Let  $\Delta := M - W$ , a point  $(W_1, W_2, \dots, W_L)$  is a critical point of the loss  $\mathcal{L}$  if and only if for all  $l = 1, 2, \dots, L$ :

$$W_{>l}^\top \Delta W_{<l}^\top = 0$$

**Lemma 1.11** (2-differential of the DLN map). Let

$$f(W_1, \dots, W_L; x) = W_L W_{L-1} \cdots W_1 x,$$

and define the shorthand notation

$$W_{>l} := W_L \cdots W_{l+1}, \quad W_{<l} := W_{l-1} \cdots W_1, \quad a_{l-1} := W_{<l} x.$$

Then the first differential of  $f$  in direction  $U = \{U_l\}_{l=1}^L$  is

$$df[U] = \sum_{l=1}^L W_{>l} U_l W_{<l} x.$$

The second differential of  $f$  in directions  $U = \{U_l\}$  and  $V = \{V_l\}$  is given by the symmetric form:

$$d^2 f[U, V] = \sum_{1 \leq i < j \leq L} \left( W_{>j} V_j W_{j-1:i+1} U_i W_{<i} + W_{>i} V_i W_{i-1:j+1} U_j W_{<j} \right) x.$$

**Derivation:** The 2-differential can be computed using the product rule:

$$d^2 f[U, V] = \sum_{l=1}^L (dW_{>l}[V] U_l W_{<l} + W_{>l} U_l dW_{<l}[V]) x,$$

$$\text{where } dW_{>l}[V] = \sum_{j>l} W_{>j} V_j W_{j-1:l+1}, \quad dW_{<l}[V] = \sum_{j<l} W_{l-1:j+1} V_j W_{<j}.$$

Substituting these expressions yields the explicit bilinear form:

$$d^2 f[U, V] = \sum_{l=1}^L \sum_{j>l} W_{>j} V_j W_{j-1:l+1} U_l W_{<l} x + \sum_{l=1}^L \sum_{j<l} W_{>l} U_l W_{l-1:j+1} V_j W_{<j} x.$$

Reindexing over ordered pairs  $(i, j)$  with  $1 \leq i < j \leq L$  gives the symmetric form stated in the lemma.

**Remark 1.12.** • The map  $f$  is linear in each  $W_l$  separately, so  $\partial^2 f / \partial W_l^2 = 0$ .

- All curvature arises from cross-layer interactions between distinct layers  $i \neq j$ .
- The 2-differential is symmetric, i.e.  $d^2 f[U, V] = d^2 f[V, U]$ .

**Corollary 1.13** (Quadratic variation). For a single infinitesimal perturbation  $dW = \{dW_l\}_{l=1}^L$ , we have

$$d^2 f = \sum_{1 \leq i < j \leq L} \left( W_{>j} dW_j W_{j-1:i+1} dW_i W_{<i} + W_{>i} dW_i W_{i-1:j+1} dW_j W_{<j} \right) x.$$

**Lemma 1.14** (Expected squared differential of the DLN map). Let

$$df[U] = \sum_{l=1}^L W_{>l} U_l h_{l-1}, \quad h_{l-1} := W_{<l} x, \quad \Sigma_x := \mathbb{E}[xx^\top],$$

and define

$$G_{lm} := W_{>l}^\top W_{>m} \in \mathbb{R}^{d_l \times d_m}, \quad C_{ml} := \mathbb{E}[h_{m-1} h_{l-1}^\top] = W_{<m} \Sigma_x W_{<l}^\top \in \mathbb{R}^{d_{m-1} \times d_{l-1}}.$$

Then the expected squared norm of the differential is

$$\mathbb{E}\langle df, df \rangle = \sum_{l,m=1}^L \text{Tr}(U_l^\top G_{lm} U_m C_{ml}) = \sum_{l,m=1}^L \text{vec}(U_l)^\top (C_{ml}^\top \otimes G_{lm}) \text{vec}(U_m).$$

**Derivation.** Starting from  $df[U] = \sum_{l=1}^L W_{>l} U_l h_{l-1}$ ,

$$\begin{aligned} \langle df, df \rangle &= \left( \sum_l W_{>l} U_l h_{l-1} \right)^\top \left( \sum_m W_{>m} U_m h_{m-1} \right) \\ &= \sum_{l,m} h_{l-1}^\top U_l^\top W_{>l}^\top W_{>m} U_m h_{m-1} = \sum_{l,m} h_{l-1}^\top U_l^\top G_{lm} U_m h_{m-1}. \end{aligned}$$

Taking expectation and using  $C_{ml} = \mathbb{E}[h_{m-1} h_{l-1}^\top]$ ,

$$\mathbb{E}\langle df, df \rangle = \sum_{l,m} \mathbb{E}[h_{l-1}^\top U_l^\top G_{lm} U_m h_{m-1}] = \sum_{l,m} \text{Tr}(U_l^\top G_{lm} U_m C_{ml}),$$

which yields the *trace form*. For the *vectorized form*, use  $\text{Tr}(A^\top B) = \text{vec}(A)^\top \text{vec}(B)$  and  $\text{vec}(AU_m B) = (B^\top \otimes A) \text{vec}(U_m)$  to obtain

$$\text{Tr}(U_l^\top G_{lm} U_m C_{ml}) = \text{vec}(U_l)^\top (C_{ml}^\top \otimes G_{lm}) \text{vec}(U_m).$$

Finally, since  $h_{k-1} = W_{<k} x$  and  $\Sigma_x = \mathbb{E}[xx^\top]$ , we have  $C_{ml} = W_{<m} \Sigma_x W_{<l}^\top$ , completing the derivation.

**Lemma 1.15** (Second differential of the squared-loss). Let  $f(W; x) = W_L \cdots W_1 x$  with  $W_l \in \mathbb{R}^{d_l \times d_{l-1}}$ , and write

$$W_{>l} := W_L \cdots W_{l+1} \in \mathbb{R}^{d_L \times d_l}, \quad W_{<l} := W_{l-1} \cdots W_1 \in \mathbb{R}^{d_{l-1} \times d_0}, \quad h_{l-1} := W_{<l} x.$$

Let the residual be  $r := y - f(W; x) \in \mathbb{R}^{d_L}$ , and define

$$\Sigma_x := \mathbb{E}[xx^\top] \in \mathbb{R}^{d_0 \times d_0}, \quad S := \mathbb{E}[r x^\top] \in \mathbb{R}^{d_L \times d_0}.$$

For two displacement directions  $U = \{U_l\}_{l=1}^L$  and  $V = \{V_l\}_{l=1}^L$ , the second differential of

$$\mathcal{L}(W) := \frac{1}{2} \mathbb{E}[\|y - f(W; x)\|_2^2]$$

is the symmetric bilinear form

$$\begin{aligned} d^2 \mathcal{L}[U, V] = & \sum_{l,m=1}^L \text{Tr}(U_l^\top G_{lm} V_m C_{ml}) \\ & - \sum_{1 \leq i < j \leq L} \left( \text{Tr}(S^\top W_{>j} V_j W_{j-1:i+1} U_i W_{<i}) + \text{Tr}(S^\top W_{>i} V_i W_{i-1:j+1} U_j W_{<j}) \right), \end{aligned}$$

where

$$G_{lm} := W_{>l}^\top W_{>m} \in \mathbb{R}^{d_l \times d_m}, \quad C_{ml} := \mathbb{E}[h_{m-1} h_{l-1}^\top] = W_{<m} \Sigma_x W_{<l}^\top \in \mathbb{R}^{d_{m-1} \times d_{l-1}}.$$

Equivalently, in block-Kronecker (Hessian) form, if  $\theta = (\text{vec}(W_1); \dots; \text{vec}(W_L))$ , then

$$d^2 \mathcal{L}[U, V] = \sum_{l,m=1}^L \text{vec}(U_l)^\top H_{lm} \text{vec}(V_m), \quad H_{lm} = \underbrace{C_{ml}^\top \otimes G_{lm}}_{\text{metric block}} - \underbrace{R_{lm}}_{\text{residual block}},$$

with

$$R_{lm} = \begin{cases} W_{m-1:l+1}^\top \otimes (W_{<l} S^\top W_{>m}), & l < m, \\ W_{l-1:m+1}^\top \otimes (W_{<m} S^\top W_{>l}), & l > m, \\ 0, & l = m. \end{cases}$$

In the population case  $y = Mx$  with  $\Delta := M - W$  we have  $S = \Delta \Sigma_x$ , and in the whitened case  $\Sigma_x = I$  this reduces to  $S = \Delta$  and  $C_{ml} = W_{<m} W_{<l}^\top$ . At any global minimizer ( $\Delta = 0$ ), the residual blocks vanish and  $d^2 \mathcal{L}$  equals the positive semidefinite metric term.

**Derivation.** *Step 1 (First and second differentials of  $f$ ).* Since  $f(W; x) = W_L \cdots W_1 x$  is linear in each  $W_l$  separately,

$$df[U] = \sum_{l=1}^L W_{>l} U_l h_{l-1}, \quad h_{l-1} = W_{<l} x.$$

By the product rule (Lemma 1.11),

$$d^2 f[U, V] = \sum_{1 \leq i < j \leq L} \left( W_{>j} V_j W_{j-1:i+1} U_i W_{<i} + W_{>i} V_i W_{i-1:j+1} U_j W_{<j} \right) x. \quad (*)$$

*Step 2 (First differential of  $\mathcal{L}$ ).* Let  $r := y - f$ . Then

$$\mathcal{L} = \frac{1}{2} \mathbb{E} \langle r, r \rangle, \quad dr[U] = -df[U], \quad d\mathcal{L}[U] = \mathbb{E} \langle r, dr[U] \rangle = -\mathbb{E} \langle r, df[U] \rangle. \quad (1)$$

*Step 3 (Master identity for  $d^2\mathcal{L}$ ).* Differentiate (??) in direction  $V$ :

$$\begin{aligned} d^2\mathcal{L}[U, V] &= \mathbb{E}[\langle dr[V], dr[U] \rangle] + \mathbb{E}[\langle r, d^2r[U, V] \rangle] \\ &= \mathbb{E}[\langle -df[V], -df[U] \rangle] + \mathbb{E}[\langle r, -d^2f[U, V] \rangle] \\ &= \boxed{\mathbb{E} \langle df[U], df[V] \rangle} - \boxed{\mathbb{E} \langle r, d^2f[U, V] \rangle}. \end{aligned} \quad (2)$$

*Step 4 (Metric term  $\mathbb{E} \langle df[U], df[V] \rangle$ ).* Using  $df[U] = \sum_l W_{>l} U_l h_{l-1}$  and  $df[V] = \sum_m W_{>m} V_m h_{m-1}$ ,

$$\langle df[U], df[V] \rangle = \sum_{l,m} h_{l-1}^\top U_l^\top W_{>l} W_{>m} V_m h_{m-1}.$$

Taking expectation and defining

$$G_{lm} := W_{>l}^\top W_{>m}, \quad C_{ml} := \mathbb{E}[h_{m-1} h_{l-1}^\top] = W_{<m} \Sigma_x W_{<l}^\top,$$

we obtain

$$\boxed{\mathbb{E} \langle df[U], df[V] \rangle = \sum_{l,m} \text{Tr}(U_l^\top G_{lm} V_m C_{ml}).} \quad (3)$$

(Use  $\mathbb{E}[a^\top Bc] = \text{Tr}(B \mathbb{E}[ca^\top])$ .)

*Step 5 (Residual-curvature term  $\mathbb{E} \langle r, d^2f[U, V] \rangle$ ).* From (??),

$$\langle r, d^2f[U, V] \rangle = \sum_{i < j} \left( r^\top W_{>j} V_j W_{j-1:i+1} U_i W_{<i} + r^\top W_{>i} V_i W_{i-1:j+1} U_j W_{<j} \right).$$

Let  $S := \mathbb{E}[r r^\top]$ . Using the identity

$$\mathbb{E}[r^\top A x] = \text{Tr}(S^\top A) \quad \text{for any } A \in \mathbb{R}^{d_L \times d_0}, \quad (4)$$

we get

$$\boxed{\mathbb{E} \langle r, d^2f[U, V] \rangle = \sum_{i < j} \left( \text{Tr}(S^\top W_{>j} V_j W_{j-1:i+1} U_i W_{<i}) + \text{Tr}(S^\top W_{>i} V_i W_{i-1:j+1} U_j W_{<j}) \right).} \quad (5)$$

(Proof of (??):  $r^\top A x = \text{Tr}(r^\top A x) = \text{Tr}(A x r^\top)$ ; take expectation and use linearity to get  $\text{Tr}(A \mathbb{E}[x r^\top]) = \text{Tr}(S^\top A)$ .)

*Step 6 (Assemble).* Subtract (??) from (??) in (??) to obtain the trace formula in the lemma.

*Step 7 (Block-Kronecker form).* Use the identity

$$\text{Tr}(A V B U) = \text{vec}(U)^\top (B^\top \otimes A) \text{vec}(V), \quad (6)$$

to convert each trace:

$$\text{Tr}(U_l^\top G_{lm} V_m C_{ml}) = \text{vec}(U_l)^\top (C_{ml}^\top \otimes G_{lm}) \text{vec}(V_m).$$

For  $i < j$ ,

$$\text{Tr}(S^\top W_{>j} V_j W_{j-1:i+1} U_i W_{<i}) = \text{vec}(U_i)^\top (W_{j-1:i+1}^\top \otimes (W_{<i} S^\top W_{>j})) \text{vec}(V_j),$$

and symmetrically with  $(i, j)$  swapped for the other cross term. Grouping all blocks yields  $H_{lm}$  stated in the lemma. Symmetry  $H_{lm}^\top = H_{ml}$  follows from exchanging  $U \leftrightarrow V$  in (??) and the definitions of  $G, C$ .  $\square$

**Proposition 1.16** (Hessian of the squared loss as a block operator). Let  $f(W; x) = W_L \cdots W_1 x$  with  $W_l \in \mathbb{R}^{d_l \times d_{l-1}}$ . For  $1 \leq l \leq L$  set

$$W_{>l} := W_L \cdots W_{l+1} \in \mathbb{R}^{d_L \times d_l}, \quad W_{<l} := W_{l-1} \cdots W_1 \in \mathbb{R}^{d_{l-1} \times d_0}, \quad h_{l-1} := W_{<l} x.$$

Let  $r := y - f(W; x)$ ,  $\Sigma_x := \mathbb{E}[xx^\top] \in \mathbb{R}^{d_0 \times d_0}$  and

$$S := \mathbb{E}[r x^\top] \in \mathbb{R}^{d_L \times d_0}, \quad G_{lm} := W_{>l}^\top W_{>m} \in \mathbb{R}^{d_l \times d_m}, \quad C_{ml} := \mathbb{E}[h_{m-1} h_{l-1}^\top] = W_{<m} \Sigma_x W_{<l}^\top.$$

Define the *middle product* between layers  $l$  and  $m$  by

$$W_{m-1:l+1} := \begin{cases} W_{m-1} \cdots W_{l+1} \in \mathbb{R}^{d_{m-1} \times d_l}, & \text{if } m > l+1, \\ I_{d_l}, & \text{if } m = l+1. \end{cases}$$

(When  $l > m$ , we will only use  $W_{l-1:m+1}$  defined analogously.)

Consider the squared-error population loss

$$\mathcal{L}(W) := \frac{1}{2} \mathbb{E}[\|y - f(W; x)\|_2^2].$$

Then the Hessian  $\nabla^2 \mathcal{L}(W)$ , viewed as a block operator acting on a perturbation  $V = \{V_m\}_{m=1}^L$ , has blocks

$$(\nabla^2 \mathcal{L}(W) [V])_l = \sum_{m=1}^L (\nabla_{W_l, W_m}^2 \mathcal{L}) [V_m],$$

with the following explicit actions:

$$\boxed{\begin{aligned} (\nabla_{W_l, W_l}^2 \mathcal{L}) [V_l] &= G_{ll} V_l C_{ll} = W_{>l}^\top W_{>l} V_l W_{<l} \Sigma_x W_{<l}^\top, \\ (\nabla_{W_l, W_m}^2 \mathcal{L}) [V_m] &= G_{lm} V_m C_{ml} - \underbrace{(W_{<l} S^\top W_{>m}) V_m W_{m-1:l+1}}_{\text{residual-curvature coupling}}, \quad l \neq m. \end{aligned}}$$

Equivalently, in block-Kronecker form, if  $\theta = (\text{vec}(W_1); \dots; \text{vec}(W_L))$ ,

$$\nabla^2 \mathcal{L}(W) \equiv H = [H_{lm}]_{l,m=1}^L, \quad d^2 \mathcal{L}[U, V] = \sum_{l,m} \text{vec}(U_l)^\top H_{lm} \text{vec}(V_m),$$



with

$$H_{lm} = \begin{cases} C_{ml}^\top \otimes G_{lm}, & l = m, \\ C_{ml}^\top \otimes G_{lm} - W_{m-1:l+1}^\top \otimes (W_{<l} S^\top W_{>m}), & l \neq m. \end{cases}$$

In the population case  $y = Mx$  one has  $S = \Delta \Sigma_x$  with  $\Delta := M - W$ , and in the whitened case  $\Sigma_x = I$ ,

$$(\nabla_{W_l, W_m}^2 \mathcal{L})[V_m] = W_{>l}^\top W_{>m} V_m W_{<m} W_{<l}^\top - \Delta V_m W_{m-1:l+1}, \quad l \neq m,$$

while the diagonal block remains  $(\nabla_{W_l, W_l}^2 \mathcal{L})[V_l] = W_{>l}^\top W_{>l} V_l W_{<l} W_{<l}^\top$ .

**Derivation.** (i) *From the 2-differential to the Hessian blocks.* By Lemma 1.15 (second differential),

$$d^2 \mathcal{L}[U, V] = \sum_{l,m} \text{Tr}(U_l^\top G_{lm} V_m C_{ml}) - \sum_{i < j} \left( \text{Tr}(S^\top W_{>j} V_j W_{j-1:i+1} U_i W_{<i}) + \text{Tr}(S^\top W_{>i} V_i W_{i-1:j+1} U_j W_{<j}) \right). \quad (1)$$

We will rewrite each trace as a Frobenius inner product in  $U_l$  with a linear image of  $V_m$ , thereby identifying the operator  $(\nabla_{W_l, W_m}^2 \mathcal{L})[\cdot]$ .

(ii) *Metric term.* Using  $\text{Tr}(U_l^\top A V_m B) = \langle U_l, A V_m B \rangle_F$ ,

$$\sum_{l,m} \text{Tr}(U_l^\top G_{lm} V_m C_{ml}) = \sum_{l,m} \langle U_l, G_{lm} V_m C_{ml} \rangle_F. \quad (2)$$

(iii) *Residual-curvature term.* For the first sum in (1) with  $i < j$ , cyclicity of trace yields

$$\text{Tr}(S^\top W_{>j} V_j W_{j-1:i+1} U_i W_{<i}) = \text{Tr}(U_i^\top (W_{<i} S^\top W_{>j}) V_j W_{j-1:i+1}) = \langle U_i, (W_{<i} S^\top W_{>j}) V_j W_{j-1:i+1} \rangle_F. \quad (3)$$

The second sum in (1) (swap  $i, j$ ) similarly gives, for  $i < j$ ,

$$\text{Tr}(S^\top W_{>i} V_i W_{i-1:j+1} U_j W_{<j}) = \langle U_j, (W_{<j} S^\top W_{>i}) V_i W_{i-1:j+1} \rangle_F. \quad (4)$$

Combining (3) and (4) and reindexing  $(i, j) \mapsto (l, m)$ , we obtain

$$\sum_{i < j} \dots = \sum_{l < m} \langle U_l, (W_{<l} S^\top W_{>m}) V_m W_{m-1:l+1} \rangle_F + \sum_{m < l} \langle U_l, (W_{<l} S^\top W_{>m}) V_m W_{m-1:l+1} \rangle_F. \quad (5)$$

Thus, for *every* ordered pair  $(l, m)$  with  $l \neq m$ ,

$$-\mathbb{E}\langle r, d^2 f[U, V] \rangle = - \sum_{l \neq m} \langle U_l, (W_{<l} S^\top W_{>m}) V_m W_{m-1:l+1} \rangle_F. \quad (6)$$

(iv) *Identification of the operator blocks.* Collecting (2) and (6) into  $d^2 \mathcal{L}[U, V] = \sum_{l,m} \langle U_l, \cdot \rangle_F$ , we read off the unique linear maps

$$\begin{aligned} (\nabla_{W_l, W_l}^2 \mathcal{L})[V_l] &= G_{ll} V_l C_{ll}, \\ (\nabla_{W_l, W_m}^2 \mathcal{L})[V_m] &= G_{lm} V_m C_{ml} - (W_{<l} S^\top W_{>m}) V_m W_{m-1:l+1}, \quad l \neq m, \end{aligned}$$

since for all test directions  $U_l$ ,

$$\langle U_l, (\nabla_{W_l, W_m}^2 \mathcal{L})[V_m] \rangle_F = \text{Tr}(U_l^\top G_{lm} V_m C_{ml}) - 1_{l \neq m} \text{Tr}(U_l^\top (W_{<l} S^\top W_{>m}) V_m W_{m-1:l+1}).$$

(v) *Kronecker form.* Using  $\text{Tr}(U^\top A V B) = \text{vec}(U)^\top (B^\top \otimes A) \text{vec}(V)$ , the bilinear representation  $d^2 \mathcal{L}[U, V] = \sum_{l,m} \text{vec}(U_l)^\top H_{lm} \text{vec}(V_m)$  follows with

$$H_{lm} = C_{ml}^\top \otimes G_{lm} 1_{l \neq m} (W_{m-1:l+1}^\top \otimes (W_{<l} S^\top W_{>m})).$$

This  $H$  is symmetric ( $H_{lm}^\top = H_{ml}$ ) by construction.

(vi) *Population/whitened specializations.* If  $y = Mx$ , then  $r = \Delta x$  with  $\Delta := M - W$  and  $S = \Delta \Sigma_x$ . For whitened inputs  $\Sigma_x = I$ ,

$$C_{ml} = W_{<m} W_{<l}^\top, \quad S^\top = \Delta^\top,$$

which yields the stated simplified forms.  $\square$

## 2 DLNs as Quiver Representations

**Definition 2.1.** A *quiver* is a directed graph  $Q = (P, E, s, t)$  where  $P$  is the set of vertices,  $E$  the set of directed edges, and  $s, t : E \rightarrow P$  are the source and target maps.

**Definition 2.2.** A *representation* of a quiver  $Q$  over a field  $k$  assigns a vector space  $V_v$  to each vertex  $v \in P$  and a linear map  $\phi_e : V_{s(e)} \rightarrow V_{t(e)}$  to each directed edge  $e \in E$ . For a fixed *dimension vector*  $d = (d_v)_{v \in P}$  with  $d_v = \dim_k(V_v)$ , we denote by  $\text{Rep}_d(Q)$  the space of  $d$ -dimensional representations of  $Q$  over  $k$ :

$$\text{Rep}_d(Q) \cong \bigoplus_{e \in E} \text{Hom}_k(k^{d_{s(e)}}, k^{d_{t(e)}}).$$

**Definition 2.3.** The *path algebra*  $kQ$  is the  $k$ -vector space with basis given by all paths in  $Q$  (including a trivial path  $e_v$  of length 0 at each vertex  $v$ ). Multiplication is given by concatenation of paths when composable and 0 otherwise, extended by linearity.

**Proposition 2.4.** The category of representations of a quiver  $Q$  is equivalent to the category of left modules over its path algebra  $kQ$ .

Sketch of proof: Given a representation  $(V_v, \phi_e)$ , define the module  $M = \bigoplus_{v \in P} V_v$  with  $e_v$  acting as projection onto  $V_v$  and an arrow  $a : v \rightarrow w$  acting as  $\phi_a$  on  $V_v$  and 0 elsewhere. Conversely, for a  $kQ$ -module  $M$ , define  $V_v = e_v M$  and  $\phi_a$  the action of  $a$  on  $M$ . These constructions define functors that are quasi-inverse to each other.

**Definition 2.5.** A *deep linear network (DLN)* of depth  $L$  is a linear map  $\phi \in \text{Hom}_k(k^{d_0}, k^{d_L})$  that factorizes as a composition of linear maps

$$\phi = \phi_L \phi_{L-1} \cdots \phi_1,$$

where  $\phi_\ell \in \text{Hom}_k(k^{d_{\ell-1}}, k^{d_\ell})$  for  $\ell = 1, \dots, L$ .

**Proposition 2.6.** A DLN with dimensions  $(d_0, \dots, d_L)$  can be identified with a representation of the *linear quiver*  $A_{L+1}$  consisting of  $L+1$  vertices connected by  $L$  consecutive arrows. In this identification, each layer  $\phi_\ell$  corresponds to an edge of  $A_{L+1}$  and the vector spaces  $k^{d_\ell}$  to the vertices.

Over  $k = \mathbb{R}$  (or  $\mathbb{C}$ ) we fix inner products on each  $V_v \cong k^{d_v}$  and endow each  $\text{Hom}(V_{s(e)}, V_{t(e)})$  with the Frobenius inner product

$$\langle \phi, \psi \rangle = \text{Tr}(\phi^\top \psi).$$

Summing this over all edges defines an inner product on  $\text{Rep}_d(Q)$  and hence a Euclidean structure on the representation space.

**Definition 2.7.** Let  $\phi^* \in \text{Hom}_k(k^{d_0}, k^{d_L})$  be a target linear map. The loss function of a DLN  $\phi = \phi_L \cdots \phi_1$  is

$$\mathcal{L}((\phi_\ell)_{\ell=1}^L) = \|\phi^* - \phi_L \cdots \phi_1\|_F^2,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. We are particularly interested in the fiber  $\mathcal{L}^{-1}(0)$ .

**Proposition 2.8.** The representation space  $\text{Rep}_d(Q)$  is an affine variety with coordinate ring

$$k[\text{Rep}_d(Q)] = k[x_{ij}^{(\ell)} \mid 1 \leq \ell \leq L, 1 \leq i \leq d_\ell, 1 \leq j \leq d_{\ell-1}],$$

and hence

$$\text{Rep}_d(Q) \cong \mathbb{A}^{\sum_{\ell=1}^L d_{\ell-1} d_\ell}.$$

**Proposition 2.9.** The loss  $\mathcal{L} : \text{Rep}_d(Q) \rightarrow k$  is a polynomial map of degree  $2L$ . Its graph

$$\{(\Phi, \lambda) \in \text{Rep}_d(Q) \times k \mid \lambda = \mathcal{L}(\Phi)\}$$

is an affine subvariety of dimension  $\sum_{\ell=1}^L d_{\ell-1} d_\ell$ .

Since  $\text{Rep}_d(Q)$  is isomorphic to an affine space, it carries a natural module of Kähler differentials  $\Omega_{k[\text{Rep}_d]/k}$ , giving an algebraic notion of differential  $d\mathcal{L}$ . Over  $\mathbb{R}$  or  $\mathbb{C}$ , once the Frobenius inner product is fixed, this identifies with the usual analytic gradient and the second differential with the Hessian of  $\mathcal{L}$ . This will enable us to study the critical points and local geometry of the DLN loss landscape.

### 3 Fibers of the multiplication map

This is a summary of Simon Pepin Lehalleur's paper on computing the codimension of the global minimum of the population loss of a deep linear network. The weight matrices of a DLN defines a representation of a Quiver. We have an action of the group  $GL_{\bar{d}} := \prod_i GL_{d_i}$  over  $\text{Rep}_{\bar{d}}$ . We can then partition  $\text{Rep}_{\bar{d}}$  into its orbits under the action of  $G_d := GL_{\bar{d}}$ . Recall that we have an isomorphism between  $\text{Rep}_d$  and  $kQ$ -modules. The isomorphism classes of  $kQ$ -modules of dimension  $\bar{d}$  are isomorphic to the orbits of  $\text{Rep}_d$  under the action of  $G_d$ . The loci of zero loss is defined by:

$$\pi^{-1}(0) \subset \text{Rep}_{\bar{d}} \text{ with } \pi(\phi) = \phi_L \cdots \phi_1$$

We are interested by relating the geometry of  $\pi^{-1}(0)$  with the combinatorics of the ranks of products of weight matrices. Note that we have:

$$\pi^{-1}(0) = \Sigma_d^0 = \{\phi \in \text{Rep}_d \mid rk(\phi) = 0\}$$

More generally, define:

$$\Sigma_d^r = \{\phi \in \text{Rep}_d | rk(\phi) = r\}$$

Given that the map  $\pi$  is equivariant w.r.t to  $G$ , we can decompose  $\Sigma_d^r$  into orbits under the action of  $G$ . This is true because we have a group action of  $G$  induced on  $\text{Im}(\pi)$  by the equivariance relation. An important result is that we can parametrize the  $G$ -orbits of  $\text{Rep}_d$  via the multiplicity pattern defined by:

$$\mathcal{M}_d^+ := \{(m_{ij})_{1 \leq i \leq j \leq N} | m_{ij} \in \mathbb{N}, d_k = \sum_{i \leq k \leq j} m_{ij}\}$$

The multiplicity patterns parametrize the isomorphic classes of the of the Dynkin Quiver  $\mathbb{A}_n$ -modules whose indecomposables representaions are given by (following from Gabriel's Theorem):

$$M_{ij} := 0 \rightarrow \dots \rightarrow 0 \rightarrow k \rightarrow \dots \rightarrow k \rightarrow 0 \dots \rightarrow 0$$

By Krull-Schmidt every  $Q$ -module writes:

$$M = \oplus_{1 \leq i \leq j \leq N} m_{ij} M_{ij}$$

The rank patterns is another important object that enables to parametrize the orbits of  $\text{Rep}_d$  as well in the particular case of the  $A_n$  quiver. More specifically we are interested in the following orbit:

$$\mathcal{O}_{\bar{r}} := \{\phi | rk(\phi_j \dots \phi_{i+1}) = r_{ij}\}$$

With:

$$r_{ij} \in \mathcal{R}_d := \{(r_{ij})_{1 \leq i \leq j \leq N} | r_{ij} \in \mathbb{N}, r_{ij} - r_{i,j+1} - r_{i-1,j} + r_{i-1,j+1} \geq 0, r_{ii} = d_i\}$$

The rank pattern set  $\mathcal{R}_d$  is in bijection with the multiplicity patterns and therefore parametrize the orbit of  $\text{Rep}_d$  as well. In other words, the  $G$ -orbits of  $\text{Rep}_d$  corresponds to rank condition on a composition of weight matrices in a DLN. We can therefore stratify the space  $\Sigma_d^r$  by the orbits  $\mathcal{O}_{\bar{r}}$  parametrized by the rank patterns. This allows to relate the codimension of 0-loss with the codimension of the orbits which can be encoded in a power series using Pochhammer symbols.

## 4 Loss landscape of DLNs

### 4.1 Equivariance of the gradient map

TODO: Recall main results on local minima being global minima, recall 1st and 2nd order analysis of critical points. Consider a DLN. We assume that the teacher matrix has distinct singular values, the output space is lower dimensional than the input space and we are in the overparametrize regime. Input-output covariance matrix is full rank and input covariance matrix is invertible. The paper on the Loss Landscape of Deep Linear Neural Networks parametrize the first order critical points with the subset of the singular values of the teacher matrix.

**Proposition 4.1.** Let  $\phi \in \text{Rep}_d$  be a first order critical point of the loss. Let  $M = U\Sigma V^\top$  be the teacher matrix, let  $r := rk(\pi(\phi))$ , let  $P_S = U_S U_S^\top$  be a projector along the top  $r$  singular values of the teacher matrix  $M$  then there exists a unique subset  $S \subset \{1, \dots, d_{out}\}$  of size  $r$  such that:

$$\pi(\phi) = \phi_L \dots \phi_1 = P_S M$$

If  $\phi$  is a first order critical point then it is given by some subset of singular values of the teacher matrix. However, not all  $P_S M$  are critical points. So we cannot use  $P_S M$  as a parametrization of critical points. In other words we have:

$$\text{Crit}(\mathcal{L}) \subset \pi^{-1}(P_S M)$$

Let  $G := G_{d_0} \times G_h \times G_{d_L}$  acting on  $\text{Rep}_d$  via:

$$g \cdot \phi = (g_1 \phi_1 g_0^{-1}, \dots, g_L \phi_L g_{L-1}^{-1})$$

The group  $G_h = \prod_{0 < i < L} G_{d_i}$  acts on  $\text{Rep}_d$  as a subgroup of  $G$ . Let  $\pi_{>l} : \text{Rep}_d \rightarrow \text{Hom}(k^{d_l}, k^{d_L})$  and  $\pi_{<l} : \text{Rep}_d \rightarrow \text{Hom}(k^{d_0}, k^{d_l})$  be the product maps:

$$\pi_{>l}(\phi) = \phi_L \dots \phi_{l+1}$$

$$\pi_{<l}(\phi) = \phi_{l-1} \dots \phi_1$$

The multiplication map  $\pi$  is  $G$ -equivariant and we can show that the product maps  $\pi_{>l}$  and  $\pi_{<l}$  are  $G$ -equivariant:

$$\begin{aligned} \pi_{>l}(g \cdot \phi) &= g_L \phi_L \dots \phi_{l+1} g_l^{-1} = g_L \pi_{>l}(\phi) g_l^{-1} \in \text{Hom}(k^{d_l}, k^{d_L}) \\ \pi_{<l}(g \cdot \phi) &= g_{l-1} \phi_{l-1} \dots \phi_1 g_0^{-1} = g_{l-1} \pi_{<l}(\phi) g_0^{-1} \in \text{Hom}(k^{d_0}, k^{d_l}) \end{aligned}$$

Let

$$E(\phi) := \pi(\phi^*) - \pi(\phi)$$

where  $\phi^*$  is a global minimum of the loss in  $\text{Rep}_d$ . Let  $G_{\phi^*}$  be the stabilizer of the target map  $\pi(\phi^*)$  in  $GL_{d_L} \times GL_{d_0}$ . In other words,

$$G_{\phi^*} := \{(g_L, g_0) \in GL_{d_L} \times GL_{d_0} \mid g_L \pi(\phi^*) g_0^{-1} = \pi(\phi^*)\}$$

The group  $G := G_{\phi^*} \times G_h$  acting on  $\text{Rep}_d$ . The error term  $E$  is  $G_{\phi^*}$ -equivariant because  $\pi(\phi)$  is  $G$ -equivariant and  $\pi(\phi^*) \in \text{Rep}_d$  is  $G_{\phi^*}$ -invariant. In other words we have:

$$\pi(g \cdot \phi^*) = g_L \pi(\phi^*) g_0^{-1} = \pi(\phi^*), \text{ for } (g_L, g_0) \in G_{\phi^*}$$

The gradient at vertex  $l$  writes:

$$\nabla_l L(\phi) = \pi_{>l}(\phi)^t E(\phi) \pi_{<l}(\phi)^t$$

In particular we can define the gradient map at vertex  $l$ ,  $\nabla_l \in \text{Hom}(\text{Rep}_d, \text{Hom}(k^{d_{l-1}}, k^{d_l}))$ :

$$\nabla_l : \phi \mapsto \pi_{>l}(\phi)^t E(\phi) \pi_{<l}(\phi)^t$$

Which induces a gradient map

$$\nabla := (\nabla_l)_l \in \bigoplus_l \text{Hom}(\text{Rep}_d, \text{Hom}(k^{d_{l-1}}, k^{d_l})) \cong \text{Hom}(\text{Rep}_d, \text{Rep}_d)$$

note that  $\text{Rep}_d^*$  is a representation of the dual quiver with all the orientation of arrows from  $Q$  inverted<sup>2</sup>. The space  $\mathcal{C}_L$  is the set of critical points i.e. points in  $\text{Rep}_d$  at which the map  $\nabla$  vanishes.

---

<sup>2</sup>Maybe the double quiver will be useful?

**Proposition 4.2.** Let  $O_i := O(d_i)$  be the orthogonal group acting on vertex  $i$ . Let

$$O_{\phi^\star} := G_{\phi^\star} \cap (O_L \times O_0)$$

and let  $OG := O_{\phi^\star} \times G_h$  be a group acting on  $Rep_d$ . The map  $\nabla$  is equivariant under the action of  $OG$ .

**Proof** We use the  $OG$ -equivariance of both the multiplication map  $\pi$  and the error map  $E$  and find that:

$$\nabla_l(L(g.\phi)) = g_l^{-t} \pi_{>l}(\phi)^t g_L^t g_L E(\phi) g_0^{-1} g_0^{-t} \pi_{<l}(\phi)^t g_{l-1}^t$$

Since  $g_L \in O_L$  and  $g_0 \in O_0$  then we have:  $g_L^t g_L = 1$ ,  $g_0^t g_0 = 1$  and thus:

$$\begin{aligned} \nabla_l(L(g.\phi)) &= g_l^{-t} \nabla_l L(\phi) g_{l-1}^t, \forall l \\ \nabla L(g.\phi) &= g * \nabla L(\phi) \quad \text{Action of OG defined by:} \\ g * \nabla L(\phi) &= (g_l^{-t} \nabla_l L(\phi) g_{l-1}^t)_l \end{aligned}$$

In particular thanks to the latter equality we see that if  $\phi \in Rep_d$  is a critical point of  $L$  then for  $g \in OG$  we have that  $g.\phi$  is also a critical point of  $L$ . This defines a group action of  $OG$  over  $\mathcal{C}_L$ . Since we have a group action of  $OG$  over  $\mathcal{C}_L$  there is a stratification of the critical points of  $L$  into orbits of  $OG$ :

$$\mathcal{C}_L := \bigsqcup_{\phi \in \mathcal{C}_L} \mathcal{O}_\phi$$

We want to find a parametrization of these orbits that could enable us to study the geometry of critical points (for example computing their codimension and studying their irreducible components) similar to what has been done previously on the fibers of  $\phi$  (Simon's work).

## 4.2 Quasi-Parameterization of saddle points

Recall proposition 9 and 10 from the paper The Loss Landscape of Deep Linear Neural Networks: a Second-order Analysis.

**Assumptions:**

- Input data dimension  $d_x$  and output data dimension  $d_y$  are such that  $d_y \leq d_x \leq m$
- Input data covariance  $\Sigma_X$  is invertible and input-output data covariance  $\Sigma_{YX}$  is full rank  $d_y$
- The singular values of  $\sigma = \Sigma_{YX} \Sigma_X^{-1} X$  are all distinct

**Proposition 9 (reformulated):** Let  $\phi$  be a critical point. Let  $S \subset \{1, \dots, d_y\}$  be a set indexing the singular values of the target map. Let  $r := |S|$  be the rank of  $\phi$ , proposition 9 ensures that  $\phi$  is conjugate to the point  $\tilde{\phi}$  in  $Rep_d$  with:

$$\begin{aligned} \pi(\tilde{\phi}) &= \tilde{\phi}_L \dots \tilde{\phi}_1 = U_S U_S^\top \Sigma_{YX} \Sigma_X^{-1}, \quad \text{the map } U_S U_S^\top \text{ projects on singular values indexed by } S \\ \tilde{\phi}_L &= [U_S \ U_Q Z_L] \quad \exists Z_L \in \mathbb{R}^{d_y - r \times d_{L-1} - r} \text{ and } Q = \{1 \dots d_y\} - S \\ \forall h \in \{L-1, \dots, 2\}, \quad \tilde{\phi}_h &= I_r \oplus Z_h, \quad \exists Z_h \in \mathbb{R}^{d_h - r \times d_{h-1} - r} \\ \tilde{\phi}_1 &= [U_S^\top \Sigma_{YX} \Sigma_X^{-1} \ Z_1]^\top \exists Z_1 \in \mathbb{R}^{d_1 - r \times d_0 - r} \\ U_Q Z_L \dots Z_2 &= 0 \end{aligned}$$

Let  $\tilde{\phi}_Z$  be a representation  $Q$  satisfying the conditions of the proposition just above. Define:

$$U_{\tilde{\phi}_Z}^S = \{\phi \in Rep_d | \phi \sim \tilde{\phi}_Z\}$$

In particular, we see that the set of critical point of rank  $r = |S|$  associated with the subset of singular values indexed by  $S$  satisfies:

$$\mathcal{C}_{L,S} \subset \cup_Z U_{\tilde{\phi}_Z}^S \subset \pi^{-1}(U_S U_S^\top \Sigma_{XY} \Sigma_X^{-1})$$

**Proposition 10 (reformulated):** Consider  $\phi$  such that  $\phi$  is conjugate to the representation  $\tilde{\phi}$  such that:

$$\begin{aligned} \tilde{\phi}_L &= [U_S \ U_Q Z_L] \quad \exists Z_L \in \mathbb{R}^{d_y - r \times d_{L-1} - r} \text{ and } Q \in \{1 \dots d_y\} - S \\ \forall h \in \{L-1, \dots, 2\}, \tilde{\phi}_h &= I_r \oplus Z_h, \quad \exists Z_h \in \mathbb{R}^{d_h - r \times d_{h-1} - r} \\ \tilde{\phi}_1 &= [U_S^\top \Sigma_{XY} \Sigma_X^{-1} \ Z_1]^\top \exists Z_1 \in \mathbb{R}^{d_1 - r \times d_0 - r} \end{aligned}$$

if  $r = r_{max}$  or if there exists  $h_1 \neq h_2$  with  $Z_{h_1} = Z_{h_2} = 0$  then  $\phi$  is a first order critical point of the loss associated with the subset  $S$ .

Let  $\phi$  be a representation  $Q$  satisfying the conditions of the proposition just above. Define:

$$U_{h_1, h_2}^S = \{\phi \in Rep_d | \phi \sim \tilde{\phi} \text{ and } \exists h_1 \neq h_2 | Z_{h_1} = Z_{h_2} = 0\}$$

Then we have:

$$U_{h_1, h_2}^S \subset \mathcal{C}_{L,S} \subset \cup_Z U_{\tilde{\phi}_Z}^S \subset \pi^{-1}(U_S U_S^\top \Sigma_{XY} \Sigma_X^{-1})$$

Unfortunately, a full parameterization of the first order critical points is lacking. This is because we could find a critical point that does not fall into  $U_{h_1, h_2}^S$  for example or some representation in  $U_{\tilde{\phi}}^S$  that is not a critical point.

Recall that a critical point satisfies:

$$\pi_{>l}(\phi)^t E(\phi) \pi_{<l}(\phi)^t = 0$$

Let  $\phi^\star := \Sigma_{XY} \Sigma_X^{-1}$  be the target map and let  $P_S := U_S U_S^\top$  be a projector on singular values indexed by  $S$ . Using proposition 9 we can write this equation as:

$$\pi_{>l}(\phi)^t (I - P_S) \phi^\star \pi_{<l}(\phi)^t = 0$$

Let  $P_S^\perp := I - P_S$ , unfortunately an  $GL_d$  group action on the map:

$$\psi_S(\phi) := \pi_{>l}(\phi)^t P_S^\perp \phi^\star \pi_{<l}(\phi)^t$$

allows to define a group of  $GL_d$  over the critical points if only if

$$\pi_{>l}(\phi)^t g_L^{-t} P_S^\perp \phi^\star g_0^{-t} \pi_{<l}(\phi)^t = 0$$

Which does not seem to hold in general.

**Proposition** The gradient is equivariant under the action of  $G_h$  and critical points are  $G_h$  invariant.

**Proof** We have already shown that the gradient map satisfies:

$$g * \nabla_l L(\phi) = g_l^t \nabla_l L(\phi) g_{l-1}^t = \nabla_l L(g.\phi)$$

Let  $\phi_S$  be a critical point i.e. we have

$$\nabla L(\phi_S) = 0$$

Acting on  $\phi_S$  via the hidden subgroup  $G_h$  yields:

$$\nabla_l L(g\phi_S) = g_l^\dagger \nabla_l L(\phi_S) g_{l-1}^t = 0$$

Since  $\nabla_l L(\phi_S) = 0$  for all  $l$ .

## 5 Hessian analysis

The Hessian of the loss function encodes the local curvature of the parameter space and therefore provides a second-order description of the geometry of the optimization landscape. Its eigenvalues quantify curvature along different parameter directions: large eigenvalues correspond to stiff, highly identifiable directions, while small or zero eigenvalues indicate flat or redundant directions. The rank of the Hessian thus measures the effective local dimensionality of the model and reveals structural degeneracies induced by overparameterization or symmetries. In regular models, the Hessian is full rank and the local learning coefficient (LLC) equals half the parameter dimension. In singular models, however, the presence of degenerate Hessian directions reflects non-identifiability and constrains the asymptotic behavior of the marginal likelihood; in particular, the Hessian rank provides an upper bound on twice the LLC and characterizes the quadratic part of the loss that determines the leading-order contribution to the real log canonical threshold.

### 5.1 Hessian Decomposition: Outer-Product and Functional Terms

Consider a parametric model  $f_\theta(x) \in \mathbb{R}^K$  with loss

$$L(\theta) = \mathbb{E}_{(x,y)} [\ell(f_\theta(x), y)].$$

Let  $J_\theta(x) := \nabla_\theta f_\theta(x) \in \mathbb{R}^{K \times p}$  denote the Jacobian of the model output with respect to the parameters, and let  $\nabla_f \ell$  and  $\nabla_f^2 \ell$  denote the gradient and Hessian of the loss with respect to the model output  $f$ . Applying the chain rule twice yields the Gauss–Newton decomposition of the parameter Hessian:

$$\nabla_\theta^2 L(\theta) = H_o(\theta) + H_f(\theta),$$

where

$$H_o(\theta) = \mathbb{E} [J_\theta(x)^\top \nabla_f^2 \ell(f_\theta(x), y) J_\theta(x)],$$

and

$$H_f(\theta) = \mathbb{E} \left[ \sum_{k=1}^K \frac{\partial \ell}{\partial f_k}(f_\theta(x), y) \nabla_\theta^2 f_{\theta,k}(x) \right].$$

**Outer-product term.** The matrix  $H_o$  depends only on first derivatives of the model and second derivatives of the loss. It measures how curvature in output space is pulled back to parameter space through the Jacobian. When  $\nabla_f^2 \ell \succeq 0$  (e.g., mean squared error or cross-entropy),  $H_o$  is positive semidefinite. For mean squared error,

$$\ell(f, y) = \frac{1}{2} \|f - y\|^2, \quad \nabla_f^2 \ell = I_K,$$



and therefore

$$H_o(\theta) = \mathbb{E}[J_\theta(x)^\top J_\theta(x)].$$

In this case,  $H_o$  coincides with the Gauss–Newton matrix and is closely related to the empirical Fisher and the Neural Tangent Kernel. Its rank is bounded by the rank of the Jacobian and thus reflects the locally identifiable directions supported by the data.

**Functional term.** The matrix  $H_f$  depends on second derivatives of the model itself. It captures curvature induced by the parameterization rather than by the loss geometry. For mean squared error,

$$H_f(\theta) = \mathbb{E}\left[\sum_{k=1}^K (f_{\theta,k}(x) - y_k) \nabla_\theta^2 f_{\theta,k}(x)\right].$$

Hence  $H_f$  is weighted by residuals and vanishes at exact interpolation. Unlike  $H_o$ , it is generally indefinite and encodes structural degeneracies arising from overparameterization or symmetries in the model.

**Geometric interpretation and relation to the LLC.** The decomposition separates two sources of curvature: data-induced curvature ( $H_o$ ) and parameterization-induced curvature ( $H_f$ ). Near well-fitted minima in regression and classification,  $H_f$  becomes negligible and the Hessian is dominated by  $H_o$ . The nonzero eigenvalues of  $H_o$  determine the quadratic directions of the loss landscape and therefore govern the leading-order local geometry. In regular models the Hessian is full rank, whereas in singular models rank deficiency reflects parameter redundancies. Since the local learning coefficient (LLC) is determined by the leading quadratic structure of the loss expansion, the rank of the Hessian provides an upper bound on twice the LLC, while higher-order terms determine whether this bound is tight. ““latex

## 5.2 Exact Rank Formula, Upper Bound, and Rank Deficiency

We consider a depth- $L$  deep linear network with layer widths  $d_0, d_1, \dots, d_L$ , where  $d_0 = d$  is the input dimension and  $d_L = K$  the output dimension. Let  $M_\ell = d_\ell$  and let

$$M := \sum_{\ell=1}^{L-1} M_\ell$$

denote the total number of hidden units. Let  $\Sigma_{xx} = \mathbb{E}[xx^\top]$  with  $r := \text{rk}(\Sigma_{xx})$ , and define

$$q := \min\{r, M_1, \dots, M_{L-1}, K\}.$$

**Exact expression for the outer-product term.** For the mean squared error loss, the outer-product (Gauss–Newton) term admits the factorization

$$H_o = A_o(I_K \otimes \Sigma_{xx})A_o^\top,$$

where

$$A_o^\top = (W_{L:2} \otimes I_d, W_{L:\ell+1} \otimes W_{1:\ell-1}, \dots, I_K \otimes W_{1:L-1}).$$

Under the maximal-rank assumption on the weight matrices, the rank of  $H_o$  is given exactly by

$$\text{rk}(H_o) = q(r + K - q).$$

**Upper bound (tight in practice) for the full Hessian.** Writing  $H_L = H_o + H_f$ , one obtains the sharp upper bound

$$\text{rk}(H_L) \leq 2qM - Lq^2 + q(r + K).$$

Empirically (and at initialization), this bound holds with equality:

$$\text{rk}(H_L) = 2qM - Lq^2 + q(r + K).$$

**Rank deficiency.** Let  $p = \sum_{\ell=0}^{L-1} M_\ell M_{\ell+1}$  be the total number of parameters. When  $r = d$  (full-rank input covariance), the rank deficiency of the Hessian satisfies

$$\text{rank-deficiency}(H_L) = p - \text{rk}(H_L) = \sum_{\ell=0}^{L-1} (M_\ell - q)(M_{\ell+1} - q).$$

**Interpretation.** These formulas reveal that the effective dimension of the model, as measured by the Hessian rank, scales *linearly* in the layer widths, whereas the number of parameters scales *quadratically*. The quantity  $q$  acts as the effective bottleneck dimension: only  $q$  directions per layer participate in curvature. All remaining directions correspond to structurally degenerate parameter combinations.

The rank deficiency formula admits a particularly clean interpretation: it equals the parameter count of a hypothetical network in which each layer width has been reduced by  $q$ . In other words, the degeneracy is exactly the redundancy created by overparameterizing beyond the minimum effective dimension.

Consequently, deep networks possess a large nullspace in the Hessian, whose dimension grows quadratically with overparameterization. The Hessian therefore lives on a much lower-dimensional manifold than the parameter space itself. This clarifies that: (i) degeneracy is architectural rather than accidental, (ii) the effective dimension is governed by  $q$  and the sum of widths, not by the raw parameter count, and (iii) overparameterization primarily increases flat directions rather than curved ones. “`latex`”

### 5.3 Link Between $H_o$ , $H_f$ , and Strict vs. Non-Strict Saddles

Consider a deep linear network trained with squared loss

$$L(W) = \frac{1}{2} \|Y - W_{L:1} X\|_F^2.$$

The Hessian admits the Gauss–Newton decomposition

$$H_L = H_o + H_f,$$

where

$$H_o = \mathbb{E}[J^\top J] \quad \text{and} \quad H_f = \mathbb{E} \left[ \sum_{k=1}^{d_L} \frac{\partial \ell}{\partial f_k} \nabla_W^2 f_k \right].$$

For squared loss,  $H_o$  is positive semidefinite, since it is a Gram matrix of Jacobians. Hence any negative curvature of the full Hessian must arise from the functional term  $H_f$ .

**Second-order expansion and correspondence.** In the second-order analysis of Achour *et al.*, the loss along a perturbation  $W'$  around a first-order critical point is expanded as

$$L(W + tW') = L(W) + t^2 c_2(W, W') + o(t^2),$$

where the quadratic coefficient  $c_2(W, W')$  splits into two parts:

$$c_2(W, W') = F^T(W, W') + S^T(W, W').$$

The term  $F^T$  is a squared-norm expression (of the form  $\|\sum_i T_i\|^2$ ) and is manifestly nonnegative. This corresponds precisely to the Gauss–Newton curvature  $v^\top H_o v$ . The term  $S^T$  is a residual-weighted interaction term arising from second derivatives of the model; this corresponds to the functional contribution  $v^\top H_f v$ .

Thus, in quadratic form,

$$v^\top H_L v = v^\top H_o v + v^\top H_f v,$$

which matches

$$c_2(W, W') = F^T + S^T.$$

**Strict vs. non-strict saddles.** Since  $H_o \succeq 0$ , negative curvature of  $H_L$  can only occur if

$$v^\top H_f v < -v^\top H_o v$$

for some direction  $v$ . In other words, the functional term must be sufficiently negative to dominate the positive semidefinite outer-product curvature.

The results of Achour *et al.* show that

- If a critical point is *non-tightened*, then there exists a direction  $W'$  such that  $c_2(W, W') < 0$ . In Hessian language, this means  $H_f$  produces a negative eigenvalue of  $H_L$ , and the point is a strict saddle.
- If a critical point is *tightened*, the second-order term can be rewritten as a sum of squares, implying  $c_2(W, W') \geq 0$  for all directions. Equivalently,  $H_f$  cannot overcome the PSD curvature of  $H_o$ , and the Hessian has no negative eigenvalues (non-strict saddle).

**Interpretation.** The Gauss–Newton term  $H_o$  captures curvature induced by output sensitivity and is always positive semidefinite. The functional term  $H_f$  captures residual-weighted second-order effects of the parameterization and is the sole source of indefiniteness. Non-tightened configurations correspond to structural conditions under which the indefinite component  $H_f$  generates a negative curvature direction. Tightened configurations correspond to algebraic rank conditions that prevent  $H_f$  from dominating  $H_o$ , thereby eliminating strict saddle behavior.

Useful references: Analytic Insights into Structure and Rank of Neural Network Hessian Maps and Depth, Not Data: An Analysis of Hessian Spectral Bifurcation

## 6 Bounding the critical loci

Using proposition 9, we see that for all the critical points the product map is of the form:

$$\begin{aligned} \pi\phi &= \phi_L \dots \phi_1 = P_S \Sigma_{XY} \Sigma_{XX} + U_Q Z_L \dots Z_1, \text{ with} \\ U_Q Z_L \dots Z_1 &= 0 \end{aligned}$$

Let  $P_Q = U_Q U_Q^\top$  a projector onto the singular values not in  $S$  (i.e. orthogonal to  $P_S$  in some sense). We could study the multiplicative map:

$$\pi_Q(Z_L, \dots, Z_1) = U_Q Z_L \dots Z_1$$

And focus on  $\pi_Q^{-1}(0)$ . Indeed, for all critical points the latter equation holds, so perhaps by studying the fibers of this map we can study the geometry of orbits in that fiber that “upper bound” (in the sense of inclusion) the loci of saddle points. We have an action of the group  $G$  such that:

$$\pi_Q g \cdot (Z_L, \dots, Z_1) = U_Q g_L Z_L \dots Z_1 g_0^{-1} = g_L U_Q g_L^t g_L Z_L \dots Z_1 g_0^{-1} = g \cdot \pi_Q(Z_L, \dots, Z_1)$$

This holds if  $g_L \in O_L$  commutes with  $U_Q$  which implies that  $P_Q$  is a fixed point for  $g_L$ ? Is there an interesting subgroup of  $O_L$  that has this property? Another point: in terms of representations of the quiver  $Q$ , the  $Z_i$  are sub representations of the representation induced by  $\phi_i$ . **Relations between projectors on singular values** One can show that:

- $U_S U_S^\top$  is a projector on  $r$  singular values and  $U_Q U_Q^\top$  is a projector on  $d_y - r$  singular values
- $U_S^\top U_S = 1$  and similar for  $U_Q$
- The SVD of  $P_S M$  is  $U S_r V^\top$  where  $S_r$  is a matrix with singular values  $s_1, \dots, s_r$  and 0 otherwise
- The SVD of  $P_Q M$  is  $U S_{d_y-r} V^\top$  where  $S_{d_y-r}$  is a matrix with singular values  $s_{r+1}, \dots, s_{d_y-r}$  and 0 otherwise
- $U_S^\top U_Q = U_Q^\top U_S = 0$

This can be shown by noting that  $U = [U_S, U_Q]$  up to re-ordering of singular values which does not matters since SVD is invariant under permutations of singular vectors. Using  $U^\top U = 1$  we get the orthogonality relations between  $U_S$  and  $U_Q$ . Since we have  $U_Q Z_L \dots Z_1 = 0$  then using  $U_Q^\top U_Q = 1$  we find this implies  $Z_L \dots Z_1 = 0$ . Hence we have  $U_Q Z_L \dots Z_1 = 0$  if and only if  $Z_L \dots Z_1 = 0$ . We can therefore study the fiber of the multiplication map

$$\pi_Z(Z_L, \dots, Z_1) := Z_L \dots Z_1$$

To bound the critical point above (in the sense of inclusion).

## Restriction to fibers of the residuals

To study the critical loci of the DLN loss landscape, it would be useful to study the set of points that upper bound and lower bound that loci (in the sense of the inclusion of sets). Recall that from proposition 1 we have for  $\phi_c$  a critical point:  $\pi(\phi_c) = U_S U_S^\top M$ . The following general fact will be important:

**Proposition 6.1.** Let  $U$  be an orthonormal matrix in  $O_{d_y}$ . Let  $U_S$  be a matrix defined by taking the first  $r < d_y$  columns of  $U$  and let  $U_Q$  be a matrix defined by taking the remaining columns of  $U$ . The matrices  $U_S$  and  $U_Q$  have orthonormal columns and are mutually orthogonal. In particular, the matrix  $U_Q$  satisfies:

$$U_Q^\top U_Q = 1_{d_y-r}$$

.

To prove that proposition, up to re-ordering of the singular values, we can generally write the left singular matrix  $U$  as:

$$U = [U_S, U_Q]$$

Since  $U$  is an orthogonal matrix in  $O_{d_y}$  we have

$$\begin{aligned} U^\top U &= 1_{d_y} \\ &= \begin{pmatrix} U_S^\top U_S & U_S^\top U_Q \\ U_Q^\top U_S & U_Q^\top U_Q \end{pmatrix} \end{aligned}$$

Hence by identifying the matrix blocks we obtain:

$$\begin{aligned} U_S^\top U_S &= 1_r \\ U_Q^\top U_Q &= 1_{d_y-r} \\ U_S^\top U_Q &= 0_{r, d_y-r} \\ U_Q^\top U_S &= 0_{d_y-r, r} \end{aligned}$$

Which terminates the proof.

Consider a representation  $\phi \in \text{Rep}_d$  satisfying proposition 9 or 10 reformulated above. We call a subrepresentation of  $\phi$  defined by the maps  $(Z_i)_i$  the residual representation. Recall that we have:

$$\pi_Q(Z_1, \dots, Z_L) = U_Q Z_L \dots Z_1$$

We have the following proposition:

**Proposition 6.2.** Let  $\pi_Z(Z_1, \dots, Z_L) = Z_L \dots Z_1$  be the multiplication map of the residuals. We have:

$$\pi_Q(Z_1, \dots, Z_L) = 0 \iff \pi_Z(Z_1, \dots, Z_L) = 0$$

This proposition enables us to import the results from the fibers of the multiplicative map to  $\pi_Z$  to study the critical loci. Since we work under the conditions of proposition 9 and 10, this proposition is easy to prove by using the orthogonality of the columns of  $U_Q$  and  $\pi_Q(Z_1, \dots, Z_L) = 0$ :

$$U_Q^\top \pi_Q(Z_1, \dots, Z_L) = U_Q^\top U_Q \pi_Z(Z_1, \dots, Z_L) = \pi_Z(Z_1, \dots, Z_L) = Z_L \dots Z_1 = 0$$

Conversely, if  $\pi_Z = 0$  then  $\pi_Q = U_Q \pi_Z = 0$ . More generally we have for critical points:

$$\begin{aligned} U_S U_S^\top \phi &= U_S U_S^\top M \\ U_Q U_Q^\top \phi &= U_Q U_Q^\top U_Q Z_L \dots Z_1 = U_Q Z_L \dots Z_1 = 0, \text{ if } \phi \text{ is a critical point} \end{aligned}$$

The set of critical points can therefore be studied via:

$$\{\phi \in \text{Rep}_d, Z \in \text{Rep}_{d-r} | \pi(\phi) = P_S M + U_Q \pi_Z(Z), \pi_Z(Z) = 0\}$$

## 7 Saddle points for degenerate data

Consider proposition 1 from the paper “The Loss Landscape of Deep Linear Neural Networks: a Second-order Analysis”. The assumption that the matrix  $\Sigma$  has distinct eigenvalues is essential to prove that all critical points are some projection of the global optimum onto a subset of singular values and that they write as a diagonal matrix in the same basis that diagonalizes  $\Sigma$ . Proposition 1 is not true when assuming that the teacher matrix has degenerate eigenvalues. In that case a saddle point still seems to be a projection of the teacher matrix but not necessarily along its singular vector: it projects onto a basis that can be rotated to the singular basis of the teacher matrix.

For a counter example, consider the teacher matrix  $M = I_2$ , assume that  $\Sigma_{XY} = I_2 = \Sigma_X$  i.e. whitened inputs. Consider a two layers DLN with whitened inputs. Consider the projector:

$$P = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Consider the point:

$$W_1 = W_2 = P$$

We see that the point  $(P, P)$  is a critical point by checking that gradient is zero and that we have  $P^2 = P$  i.e.  $P$  is a projector. Further more the loss at  $P$  has value 1 which is not 0 i.e.  $P$  is not a global minimum but a saddle point. We have found a counter-example to proposition 1 when relaxing the assumption of data degeneracy.

We propose a conjecture for an extension of the first statement of proposition 1 when the data has some degenerate singular values. We relax the assumption that the matrix

$$\Sigma = \Sigma_{YX} \Sigma_X^{-1} \Sigma_{XY}$$

has distinct eigenvalues. Furthermore, recall that we have:

$$\Sigma = U \Lambda U^\top$$

Where  $U$  is the left singular matrix of  $\Sigma^{\frac{1}{2}}$  and  $\Lambda$  is a diagonal matrix with eigenvalues  $\lambda_i$  each with multiplicities  $m_i$ . We conjecture the following:

**Conjecture:** Let  $W$  be a first order critical point, let  $I_S$  be a block diagonal matrix of identities  $I_{m_i}$  if  $i \in S$  and 0 otherwise, then there exist a subset  $S$  indexing the distinct eigenvalues of  $\Sigma$  and a projector  $P$  and a orthogonal matrix  $Q$  such that:

$$W = P \Sigma_{XY} \Sigma_X^{-1} = U Q I_S Q^\top U^\top \Sigma_{XY} \Sigma_X^{-1}$$

The projector  $P$  satisfies:

$$P \Sigma = \Sigma P$$

The matrix  $\Lambda$  is diagonal with eigenvalues of multiplicities  $m_i$ . The matrix  $Q$  is a block diagonal matrix for which each block is an orthogonal matrix in  $O(m_i)$ .

## 8 Geometry of the critical points

Define the variety of matrices of rank  $r$ :

$$\Sigma_d^r := \{W \in \Omega \mid \text{rk}(W_L \dots W_1) = r\}$$

Consider the following group action of  $GL_d = \Pi_{l=L}^0 GL_{d_l} = GL_{d_L} \times GL_h \times GL_{d_0}$ :

$$(W_1, \dots, W_L) \mapsto (g_1 W_1 g_0^{-1}, g_2 W_2 g_1^{-1}, \dots, g_L W_L g_{L-1}^{-1})$$

We can also reparametrize  $\Sigma_d^r$  by the following:

$$\Sigma_d^r = \{W \in \Omega, Z_l \in \mathbb{R}^{d_l - r \times d_{l+1} - r} \mid W_l \sim \begin{pmatrix} I_r & 0 \\ 0 & Z_l \end{pmatrix}, Z_L \dots Z_1 = 0\}$$

Note that the latter product over  $Z$ 's can be defined as the manifold:

$$\Sigma_0^{d-r} = \{Z_l \in \mathbb{R}^{d_l - r \times d_{l+1} - r} \mid Z_L \dots Z_1 = 0\}$$

Define the algebraic variety  $\Gamma$  by the following cyclic conditions:

$$\Gamma = \{Z_l \in \mathbb{R}^{d_l - r \times d_{l+1} - r} \mid Z_{l-1} \dots Z_1 \Sigma_{YX} U_Q Z_L \dots Z_{l+1} = 0\}$$

Equivalently, we can view  $\Gamma$  as a space of quiver representations with relations. Let  $Q_\Gamma$  be the cyclic quiver with vertex set  $\{1, \dots, L\}$  and arrows

$$a_l : l+1 \rightarrow l \quad (l = 1, \dots, L-1), \quad a_L : 1 \rightarrow L,$$

together with an additional distinguished arrow

$$a_{L+1} : L \rightarrow 1.$$

A representation of  $Q_\Gamma$  is given by vector spaces  $V_l \cong \mathbb{R}^{d_l - r}$  and linear maps  $\phi_l : V_{s(a_l)} \rightarrow V_{t(a_l)}$ . We identify

$$\phi_l := Z_l \quad (l = 1, \dots, L), \quad \phi_{L+1} := \Sigma_{YX} U_Q : V_L \rightarrow V_1,$$

(i.e.  $\phi_{L+1}$  is specialized/fixed to the linear map  $\Sigma_{YX} U_Q$ ). Then  $\Gamma$  is the subset of such representations satisfying the  $L$  cyclic relations

$$\phi_{l-1} \dots \phi_1 \phi_{L+1} \phi_L \dots \phi_{l+1} = 0, \quad l = 1, \dots, L,$$

with the convention that an empty product is the identity.

From the 2nd order analysis paper, the set of critical points write:

$$\mathcal{C}_r = \{W \in \Omega, Z_l \in \mathbb{R}^{d_l - r \times d_{l+1} - r} \mid W_L = [U_S, U_Q Z_L], W_l \sim \begin{pmatrix} I_r & 0 \\ 0 & Z_l \end{pmatrix}, W_1 = [U_S^\top \Sigma_{YX} \Sigma_X^{-1}, Z_1]^\top, Z_l \in \Sigma_0^{d-r} \cap \Gamma\}$$

We want to show something like:

$$\mathcal{C}_r / G_h \cong \Sigma_0^{d-r} \cap \Gamma$$

This suggests that studying the codimension of the critical loci amounts to studying the codimensions of  $\Sigma$  and  $\Gamma$  the latter is treated by Simon's work on the fibers of the multiplication map, hence we focus on studying  $\Gamma$ .

““latex

## 9 Sampling saddle points

We leverage the algebraic parameterization of first-order critical points in deep linear networks (DLNs) given by the rank condition and the cyclic conditions. The objective is to turn these polynomial constraints into a numerical sampler that produces many (typically gauge-inequivalent) saddle points.

**Rank strata and target projection.** Let the teacher (or population) matrix be

$$M = U \operatorname{diag}(s_1, \dots, s_{d_{\text{out}}}) V^\top,$$

and fix a rank  $r$  and a subset  $S$  of singular directions with  $|S| = r$ . Define  $U_S$  as the columns of  $U$  indexed by  $S$ , and the corresponding projector  $P_S := U_S U_S^\top$ . A rank- $r$  critical stratum is characterized by the end-to-end product

$$W := W_L \cdots W_1 = P_S M, \quad \operatorname{rank}(W) = r.$$

In practice, we choose  $r < r_{\max}$  to avoid global minimizers and bias toward saddle points.

**Gauge-fixed normal form and residual variables.** We work in a fixed representative of the hidden-layer gauge orbit. Each hidden dimension  $d_\ell$  is split into a signal part of dimension  $r$  and a residual part of dimension  $d_\ell - r$ . In this gauge, the layers are parameterized as

$$W_\ell \sim \begin{pmatrix} I_r & 0 \\ 0 & Z_\ell \end{pmatrix}, \quad \ell = 2, \dots, L-1,$$

with residual blocks  $Z_\ell \in \mathbb{R}^{(d_\ell - r) \times (d_{\ell-1} - r)}$ . The first and last layers embed the signal path and attach the residuals:

$$W_L = [U_S \quad U_Q Z_L], \quad W_1 = \begin{pmatrix} U_S^\top M \\ Z_1 \end{pmatrix},$$

where  $U_Q$  spans the orthogonal complement of  $U_S$ . The signal blocks are chosen so that the product along the signal path equals  $P_S M$  exactly, ensuring the rank condition by construction.

**Cyclic and fiber constraints on the residuals.** In this normal form, first-order criticality imposes polynomial constraints on the residual blocks  $Z = (Z_1, \dots, Z_L)$ :

- **Fiber (product) constraint:**

$$\mu(Z) := Z_L Z_{L-1} \cdots Z_1 = 0.$$

- **Cyclic constraints:**

$$d\mu_\ell(Z) := Z_{\ell-1} \cdots Z_1 \sigma Z_L \cdots Z_{\ell+1} = 0, \quad \ell = 1, \dots, L,$$

where  $\sigma$  is the fixed matrix appearing in the theoretical characterization (e.g.  $\sigma = \Sigma_{XY} U_Q$  in the population setting, or an empirical analogue).



**Energy formulation.** To convert these matrix equalities into a scalar objective, we introduce the constraint energy

$$E(Z) = \frac{1}{2} \|\mu(Z)\|_F^2 + \frac{1}{2} \sum_{\ell=1}^L \|d\mu_\ell(Z)\|_F^2.$$

Then  $E(Z) = 0$  if and only if all cyclic and fiber constraints are satisfied exactly.

**Ensuring saddle points.** A cheap sufficient construction for producing saddle points is to enforce two distinct residual blocks to be zero. Concretely, choose  $a \neq b$  and set

$$Z_a = 0, \quad Z_b = 0,$$

while sampling the remaining  $Z_\ell$ . This automatically enforces the fiber constraint and typically satisfies all cyclic constraints. More expressive (but more expensive) constructions avoid exact zeros and instead certify saddle behavior numerically via negative curvature of the Hessian.

**Boltzmann sampling over constrained residuals.** To sample many solutions rather than a single feasible point, we sample from a Boltzmann distribution concentrating near the constraint variety:

$$p_\beta(Z) \propto \exp(-\beta E(Z)) \pi(Z),$$

where  $\beta > 0$  controls concentration and  $\pi(Z)$  is a prior that prevents collapse to trivial solutions. Typical choices include Gaussian or Gaussian with weight decay

$$\log \pi(Z) = -\frac{\alpha}{2} \sum_{\ell=1}^L \|Z_\ell\|_F^2,$$

or norm-targeting penalties that keep  $\|Z_\ell\|_F$  away from zero.

**SGLD updates.** Define the negative log-density (up to an additive constant)

$$\mathcal{L}_\beta(Z) := \beta E(Z) - \log \pi(Z).$$

Stochastic gradient Langevin dynamics (SGLD) then takes the form

$$Z_{t+1} = Z_t - \frac{\epsilon_t}{2} \nabla \mathcal{L}_\beta(Z_t) + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \epsilon_t I),$$

with stepsize schedule  $\epsilon_t$ . When empirical analogues of the constraints are used, gradients can be estimated on minibatches.

**Projection to true DLN critical points and Hessian certification.** **Comment: I'm not sure about the usefulness of doing this:** Because the parameterization and the matrix  $\sigma$  may be approximate in finite-sample settings, after obtaining a low-energy configuration  $Z$  we assemble the full weights  $W_1, \dots, W_L$  and perform a short refinement step on the true DLN loss to reach  $\|\nabla \mathcal{L}\| \approx 0$ , while periodically re-balancing layer norms to control gauge drift. **Comment: this is going to be useful:** finally, saddle points are certified by estimating the smallest eigenvalues of the Hessian using Hessian-vector products; the presence of a negative eigenvalue indicates a strict saddle and a non strict saddle otherwise (or perhaps local minimum in the underparametrize regime).

## 10 Cyclic quiver representations

A cyclic quiver is a quiver with at list one oriented cycle. In particular we will be interested in a quiver which is exactly one cycle here noted  $C_n$ .

**Definition 10.1** (Cyclic quiver). A *cyclic quiver* (or *oriented cycle*) on  $n \geq 1$  vertices is the quiver  $C_n$  with vertex set

$$P = \{0, 1, \dots, n-1\}$$

and arrow set

$$E = \{a_i : i \rightarrow i+1 \mid i = 0, \dots, n-2\} \cup \{a_{n-1} : n-1 \rightarrow 0\},$$

i.e. the vertices are arranged in a directed cycle.

**Definition 10.2** (Representation of a cyclic quiver). A *representation* of the cyclic quiver  $C_n$  (arrows of  $C_n$  are of the opposite orientation) over a field  $k$  is the data of vector spaces

$$(V_0, \dots, V_{n-1})$$

together with linear maps

$$\phi_i : V_{i-1} \rightarrow V_i \quad (i = 1, \dots, n-1), \quad \phi_n : V_{n-1} \rightarrow V_0,$$

i.e. a representation of the opposite quiver  $C_n$ .

**Proposition 10.3** (The relations defining  $\Gamma$  as a cyclic quiver representation). Fix  $L \geq 1$  and consider the cyclic quiver  $C_{L+1}$  on vertices

$$\{0, 1, \dots, L\}.$$

Let  $r$  be the rank of the student DLN and:

$$V_l \cong \mathbb{R}^{d_l - r}, \quad l = 0, \dots, L,$$

and consider a representation of the opposite quiver  $C_{L+1}$  given by linear maps

$$\phi_l : V_{l-1} \rightarrow V_l \quad (l = 1, \dots, L), \quad \phi_{L+1} : V_L \rightarrow V_0,$$

where indices are understood cyclically.

For  $\phi = (\phi_1, \dots, \phi_{L+1}) \in \text{Rep}_d(C_{L+1})$ , define for  $l = 1, \dots, L$  the maps<sup>3</sup>:

$$d\mu_l(\phi) := \phi_{l-1} \cdots \phi_1 \phi_{L+1} \phi_L \cdots \phi_{l+1} \in \text{Hom}(V_l, V_{l-1})$$

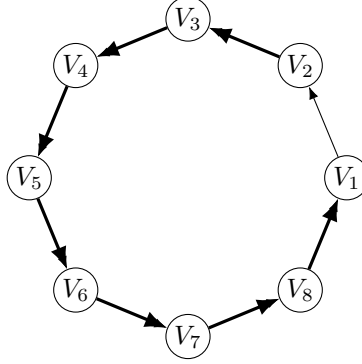
with the convention  $\phi_0 = \text{id}_{V_0}$ . The maps  $d\mu_l$  correspond to the components of the differential of the product map around the cycle. They correspond to a path of length  $L$ .

Define the closed subvariety

$$\Gamma_d := \left\{ \phi \in \text{Rep}_d(C_{L+1}) \mid d\mu_l(\phi) = 0 \text{ for all } l = 1, \dots, L \right\}.$$

---

<sup>3</sup>And such that the linear compositions are compatible with the cyclic quiver



$d\mu_1 : V_2 \rightarrow V_1$  (composition along the cycle, omitting  $\phi_2$ )

Figure 1: The cyclic quiver  $C_n$  (illustrated here with  $n = 8$ ). The thick arrows indicate the path whose composition gives a typical  $d\mu_l$  map along the cycle (example:  $d\mu_2 : V_2 \rightarrow V_1$ ).

Identifying  $\phi_l = Z_l$  for  $l = 1, \dots, L$ , we now *specialize the closing arrow*

$$\phi_{L+1} := \Sigma_{XY} U_Q : V_L \rightarrow V_0.$$

For this specialization, the algebraic variety  $\Gamma$  related to the critical loci of a DLN is:

$$\Gamma := \Gamma_d|_{\phi_{L+1} = \Sigma_{XY} U_Q}$$

coincides with the space of cyclic representations satisfying the vanishing  $d\mu$  relations and with the distinguished arrow  $\phi_{L+1}$  fixed.

**Proposition 10.4** (Equivariance of the cyclic composition maps). Let  $C_{L+1}$  be the cyclic quiver on vertices  $\{0, \dots, L\}$ , let  $\text{Rep}_d(C_{L+1})$  be its representation space with  $V_l \cong \mathbb{R}^{d_l - r}$ , and let

$$\phi = (\phi_1, \dots, \phi_L, \phi_{L+1}) \in \text{Rep}_d(C_{L+1}).$$

Define the base-change group

$$G_d := \prod_{l=0}^L \text{GL}(V_l)$$

acting on  $\text{Rep}_d(C_{L+1})$  via

$$(g \cdot \phi)_l := g_l \phi_l g_{l-1}^{-1}, \quad (l = 0, \dots, L), \text{ mod } L+1$$

For each  $l = 1, \dots, L$ , define

$$d\mu_l(\phi) := \phi_{l-1} \cdots \phi_1 \phi_{L+1} \phi_L \cdots \phi_{l+1} \in \text{Hom}(V_{l+1}, V_l),$$

with the convention that an empty product is the identity. Then, for all  $g \in G_d$  and all  $l = 0, \dots, L$ ,

$$d\mu_l(g \cdot \phi) = g_l d\mu_l(\phi) g_{l+1}^{-1}.$$

*Proof.* Expand  $d\mu_l(g \cdot \phi)$  as a product of the maps  $(g \cdot \phi)_i$ . Each factor contributes a left multiplication by some  $g_i$  and a right multiplication by  $g_{i+1}^{-1}$ , and all intermediate terms cancel by associativity, leaving only  $g_l$  on the left and  $g_{l+1}^{-1}$  on the right.  $\square$

**Proposition 10.5** (Orbit decomposition of  $\Gamma_d$ ). Let

$$\Gamma_d := \left\{ \phi \in \text{Rep}_d(C_{L+1}) \mid d\mu_l(\phi) = 0 \text{ for all } l = 1, \dots, L \right\}.$$

Then  $\Gamma_d$  is  $G_d$ -invariant. In particular,  $\Gamma_d$  decomposes as a disjoint union of  $G_d$ -orbits:

$$\Gamma_d = \bigsqcup_{\phi \in \Gamma_d} G_d \cdot \phi.$$

*Proof.* By Proposition 10.4, each map  $d\mu_l$  is  $G_d$ -equivariant and hence the vanishing set  $d\mu_l^{-1}(0)$  is  $G_d$ -stable. Intersecting over  $l = 1, \dots, L$  shows that  $\Gamma_d$  is  $G_d$ -stable. Any group action partitions a set into disjoint orbits, yielding the stated decomposition.  $\square$

Note that  $\Gamma_d \cong \text{Rep}_d(C_{L+1}) / \langle d\mu_l \rangle$  is then finite dimensional and we can classify indecomposable using indecomposable representations of Nakayama algebras.

## 10.1 Cyclic quiver as a bounded module

**Definition 10.6** (Path algebra). Let  $Q$  be a quiver. The *path algebra*  $kQ$  is the  $k$ -vector space with basis given by all directed paths in  $Q$  (including the trivial paths  $e_i$  of length 0 at each vertex  $i$ ), with multiplication given by concatenation of composable paths and 0 otherwise, extended by bilinearity.

**Definition 10.7** (Bound quiver algebra associated with  $\Gamma_d$ ). Let  $Q = C_{L+1}$  and for each  $l = 0, \dots, L$  let  $p_l \in kQ$  be the path

$$p_l := a_{l-1} \cdots a_1 a_{L+1} \cdots a_{l+1},$$

interpreting empty products as the corresponding trivial path. Let  $I \subset kQ$  be the two-sided ideal generated by  $\{p_1, \dots, p_L\}$ . The quotient algebra

$$A := kQ/I$$

is called the *bound quiver algebra* associated with these relations.

**Definition 10.8** (Representation of an algebra of dimension  $d$ ). Let  $A$  be a  $k$ -algebra and let  $d = (d_0, \dots, d_L)$ . A *representation of  $A$  of dimension  $d$  compatible with the vertices* is a left  $A$ -module structure on  $V := \bigoplus_{i=0}^L k^{d_i}$  such that the idempotents  $e_i \in kQ \rightarrow A$  act as the projections onto the  $i$ -th summand. We denote the corresponding representation space by  $\text{Rep}_d(A)$ .

**Proposition 10.9** (From quiver representations with relations to  $A$ -modules). There is a well-defined map

$$F : \Gamma_d \rightarrow \text{Rep}_d(A)$$

which sends  $\phi \in \Gamma_d$  to the unique  $A$ -module structure on  $V$  for which each arrow  $a_i$  acts by the linear map  $\phi_i$ .

*Proof.* Given  $\phi \in \Gamma_d$ , the tuple of linear maps  $(\phi_i)_i$  defines a representation of the quiver  $Q = C_{L+1}^{op}$ , hence a left  $kQ$ -module structure on  $V$  by letting a path act by the corresponding composition of arrow maps. For each  $l$ , the defining condition  $\mu_l(\phi) = 0$  is exactly the statement that the element  $p_l \in kQ$  acts as the zero map on  $V$ . Since  $I$  is the two-sided ideal generated by the  $p_l$ , every element of  $I$  also acts by zero. Therefore the  $kQ$ -action factors through the quotient  $A = kQ/I$ , producing a unique  $A$ -module structure on  $V$  with the required arrow actions.  $\square$

**Proposition 10.10** (From  $A$ -modules to quiver representations with relations). There is a well-defined map

$$G : \text{Rep}_d(A) \rightarrow \Gamma_d$$

which sends an  $A$ -module structure on  $V$  to the induced quiver representation  $\phi$  whose arrow maps are the actions of the classes of the arrows  $a_i$ .

*Proof.* Let  $V$  be a representation in  $\text{Rep}_d(A)$ . Composing the action map  $A \rightarrow \text{End}_k(V)$  with the quotient map  $kQ \twoheadrightarrow A$  gives a  $kQ$ -module structure on  $V$ . Restricting the action of each arrow basis element  $a_i$  to the appropriate vertex summand yields linear maps  $\phi_i : V_i \rightarrow V_{i+1}$  and hence a point  $\phi \in \text{Rep}_d(C_n)$ . Because each  $p_l$  maps to zero in  $A$ , it acts as the zero endomorphism on  $V$ . Equivalently, the composed linear map  $\mu_l(\phi)$  is zero for every  $l$ , so  $\phi \in \Gamma_d$ .  $\square$

**Proposition 10.11** (Correspondence between  $\Gamma_d$  and  $\text{Rep}_d(A)$ ). The maps  $F$  and  $G$  are mutually inverse bijections. In particular,

$$\Gamma_d \cong \text{Rep}_d(A), \quad A = kQ/I.$$

*Proof.* Starting from  $\phi \in \Gamma_d$ , the  $A$ -module  $F(\phi)$  is by construction generated by the same arrow actions as  $\phi$ , so applying  $G$  recovers the same tuple of arrow maps, hence  $G(F(\phi)) = \phi$ . Conversely, starting from an  $A$ -module structure on  $V$ , the map  $G$  records the arrow actions and the map  $F$  reconstructs the unique  $A$ -module whose arrow actions agree with those recorded, hence  $F(G(V)) = V$ .  $\square$

**Lemma 10.12** (Finite-dimensionality of  $A = kQ/I$ ). Let  $Q = C_{L+1}$  be the oriented cycle on  $L+1$  vertices, and let  $I \subset kQ$  be the two-sided ideal generated by the paths  $p_1, \dots, p_L$ , where each  $p_l$  is a path of length  $L$  (the unique directed path going around the cycle while omitting one arrow). Then every directed path in  $Q$  of length  $\geq L+1$  lies in  $I$ . In particular,  $A = kQ/I$  is finite-dimensional.

*Proof.* Let  $q$  be any directed path in  $Q$  of length  $\geq L+1$ . Write  $q = b_m b_{m-1} \cdots b_1$  as a concatenation of arrows with  $m \geq L+1$ . Consider the contiguous subpath  $q' := b_L \cdots b_1$  of length  $L$ . Since  $Q$  is an oriented  $L+1$ -cycle, there is exactly one directed path of length  $L$  from its source to its target, hence  $q'$  must coincide with one of the generators  $p_l$ . Therefore  $q = q'' p_l$  for some (possibly trivial) path  $q''$ , so  $q \in I$  because  $I$  is a two-sided ideal.

Thus all paths of length  $\geq L+1$  vanish in  $A$ , so  $A$  is spanned by the residue classes of paths of length  $< L+1$ , a finite set. Hence  $A$  is finite-dimensional.  $\square$

## 10.2 Bounded modules are Nakayama algebras

**Definition 10.13** ( $A$ -module). Let  $A$  be a (not necessarily commutative)  $k$ -algebra. A (left)  $A$ -module is a  $k$ -vector space  $M$  together with a  $k$ -bilinear map

$$A \times M \rightarrow M, \quad (a, m) \mapsto a \cdot m,$$

such that for all  $a, b \in A$  and  $m \in M$  one has

$$(ab) \cdot m = a \cdot (b \cdot m), \quad 1_A \cdot m = m.$$

A *homomorphism* of  $A$ -modules is a  $k$ -linear map commuting with the  $A$ -action.

**Definition 10.14** (Indecomposable module). An  $A$ -module  $M$  is *indecomposable* if  $M \neq 0$  and there do not exist nonzero submodules  $M_1, M_2 \subset M$  such that

$$M \cong M_1 \oplus M_2.$$

Equivalently,  $M$  is indecomposable if every idempotent endomorphism of  $M$  is either 0 or  $\text{id}_M$ .

**Definition 10.15** (Composition series). Let  $M$  be a finite-dimensional  $A$ -module. A *composition series* of  $M$  is a finite chain of submodules

$$0 = M_0 \subset M_1 \subset \cdots \subset M_\ell = M$$

such that each successive quotient  $M_i/M_{i-1}$  is *simple*. The integer  $\ell$  is called the *length* of the series.

**Definition 10.16** (Projective module). An  $A$ -module  $P$  is *projective* if for every surjective  $A$ -module homomorphism  $f : M \rightarrow N$  and every  $A$ -module homomorphism  $g : P \rightarrow N$ , there exists an  $A$ -module homomorphism  $\tilde{g} : P \rightarrow M$  such that  $f \circ \tilde{g} = g$ . Equivalently,  $P$  is projective if the functor  $\text{Hom}_A(P, -)$  is exact.

**Definition 10.17** (Indecomposable projective module). A *projective indecomposable*  $A$ -module is an  $A$ -module that is both projective and indecomposable. For a finite-dimensional basic algebra  $A$ , every indecomposable projective module is of the form  $P_i \cong Ae_i$  for a primitive idempotent  $e_i \in A$ , and every projective module decomposes uniquely as a direct sum of indecomposable projectives.

**Definition 10.18** (Uniserial module). An  $A$ -module  $M$  is *uniserial* if it admits a unique composition series. Equivalently, the set of submodules of  $M$  is totally ordered by inclusion.

**Definition 10.19** (Nakayama algebra). Let  $A$  be a finite-dimensional  $k$ -algebra. We say that  $A$  is a *Nakayama algebra* if every indecomposable left  $A$ -module is *uniserial*, i.e. it admits a unique composition series. Equivalently,  $A$  is Nakayama if every indecomposable projective left  $A$ -module is uniserial (and hence every indecomposable injective is uniserial).

**Definition 10.20** (Admissible ideal). Let  $Q$  be a finite quiver and let  $kQ$  be its path algebra. Write  $(kQ)_{\geq m}$  for the  $k$ -span of all paths of length at least  $m$ . A two-sided ideal  $I \subset kQ$  is called *admissible* if there exists an integer  $m \geq 2$  such that

$$(kQ)_{\geq m} \subseteq I \subseteq (kQ)_{\geq 2}.$$

**Definition 10.21** (Jacobson radical). Let  $A$  be a (not necessarily commutative) ring. The *Jacobson radical*  $J(A)$  is the intersection of all maximal left ideals of  $A$ . If  $A$  is a finite-dimensional  $k$ -algebra, then equivalently

$$J(A) = \bigcap_{S \text{ simple}} \ker (A \rightarrow \text{End}_k(S)),$$

i.e.  $J(A)$  is the largest two-sided ideal that acts by zero on every simple  $A$ -module.

**Proposition 10.22 (Bound quiver criterion for Nakayama algebras).** Let  $A \simeq kQ/I$  be a finite-dimensional bound quiver algebra, where  $Q$  is a finite quiver and  $I \subset kQ$  is an admissible ideal. Assume that for every vertex  $i \in Q_0$  there is at most one arrow leaving  $i$  and at most one arrow entering  $i$ . Then  $A$  is a Nakayama algebra.

Moreover, if  $Q$  is connected and has exactly one arrow leaving and exactly one arrow entering each vertex, then  $Q$  is an oriented cycle and  $A$  is a *cyclic Nakayama algebra*. If instead  $Q$  is connected and has a unique source and a unique sink, then  $Q$  is an oriented line and  $A$  is a *linear Nakayama algebra*.

*Proof.* Let  $e_i \in kQ \twoheadrightarrow A$  be the primitive idempotent corresponding to a vertex  $i \in Q_0$ , and consider the indecomposable projective module  $P_i := Ae_i$ . As a  $k$ -vector space,  $P_i$  is spanned by residue classes of paths in  $Q$  starting at  $i$  (modulo the relations in  $I$ ). Because there is at most one arrow leaving each vertex, for each  $\ell \geq 0$  there is *at most one* directed path of length  $\ell$  starting at  $i$ ; hence  $P_i$  has a basis indexed by (the surviving classes of) these paths.

Let  $J \subset A$  denote the Jacobson radical. Since  $A \simeq kQ/I$  with  $I$  admissible,  $J$  is generated (as a two-sided ideal) by the classes of arrows, and one has

$$J^\ell P_i = J^\ell Ae_i \cong (J^\ell)e_i,$$

which is spanned by classes of paths of length  $\geq \ell$  starting at  $i$ . By the previous paragraph, for each  $\ell$  the quotient  $J^\ell P_i / J^{\ell+1} P_i$  is either 0 or a simple module (it is spanned by the class of the unique surviving path of length  $\ell$ , if it exists). Therefore the radical filtration

$$P_i \supset JP_i \supset J^2 P_i \supset \dots$$

is a chain whose successive quotients are simple (or zero), i.e.  $P_i$  is uniserial.

Thus every indecomposable projective  $P_i$  is uniserial, and hence  $A$  is a Nakayama algebra.

For the final statements: if  $Q$  is connected and every vertex has exactly one incoming and one outgoing arrow, then the underlying directed graph is a directed cycle. If  $Q$  is connected and has a unique source and sink with all other vertices having one incoming and one outgoing arrow, then the directed graph is a directed line. These correspond to cyclic and linear Nakayama algebras respectively.  $\square$

### 10.3 Cyclic monomial Nakayama algebra with uniform Kupisch series

Throughout this subsection, let  $Q = C_{L+1}$  be the oriented  $L+1$ -cycle, let  $kQ$  be its path algebra, and let

$$A := kQ/I$$

be the bound quiver algebra from Definition (Bound quiver algebra associated with  $\Gamma_d$ ).

**Definition 10.23** (Monomial (path) ideal / monomial algebra). An ideal  $I \subset kQ$  is called *monomial* if it is generated (as a two-sided ideal) by paths in  $Q$ . A quotient  $kQ/I$  with  $I$  monomial is called a *monomial algebra*.

**Definition 10.24** (Kupisch series). Let  $A$  be a finite-dimensional basic algebra with primitive idempotents  $e_1, \dots, e_{L+1}$  and indecomposable projectives  $P_i := Ae_i$ . The *Kupisch series* of  $A$  is the  $L+1$ -tuple

$$(c_1, \dots, c_{L+1}), \quad c_i := \ell(P_i),$$

where  $\ell(P_i)$  denotes the composition length of  $P_i$ .

**Remark 10.25** (Loewy length). For a finite-dimensional algebra  $A$  with Jacobson radical  $J = J(A)$ , the *Loewy length* is the smallest integer  $n$  such that  $J^n = 0$ .

**Proposition 10.26** (Identification and Kupisch series). With  $Q = C_{L+1}$  and  $I = (p_1, \dots, p_{L+1})$  as above, the algebra  $A = kQ/I$  is a *cyclic monomial Nakayama algebra*. Moreover its Kupisch series is uniform:

$$(c_1, \dots, c_{L+1}) = (L, \dots, L).$$

*Proof.* By construction, the ideal  $I$  is generated by the paths  $p_1, \dots, p_{L+1}$ , hence  $I$  is monomial and  $A$  is a monomial algebra. By Proposition 10.22,  $A$  is Nakayama; since  $Q = C_{L+1}$  is an oriented cycle,  $A$  is in fact *cyclic* Nakayama.

Fix a vertex  $i$  and consider  $P_i = Ae_i$ . Modulo  $I$ , every path in  $Q$  of length  $\geq L$  starting at  $i$  vanishes, while for each  $t = 0, 1, \dots, L-1$  there is exactly one directed path of length  $t$  starting at  $i$ . Therefore  $P_i$  has a  $k$ -basis given by the residue classes of these  $L$  paths (length 0 through  $L-1$ ).

Since  $J = J(A)$  is generated by the classes of arrows (as recalled in the proof of Proposition 10.22), the radical filtration

$$P_i \supset JP_i \supset J^2P_i \supset \dots$$

has successive quotients  $J^tP_i/J^{t+1}P_i$  one-dimensional (hence simple) for  $t = 0, \dots, L-1$ , and zero thereafter. Thus  $\ell(P_i) = L$  for every  $i$ , which gives the Kupisch series  $(L, \dots, L)$ .  $\square$

**Remark 10.27.** Equivalently,  $A$  can be viewed as a *truncated cycle algebra*: it is the path algebra of an oriented cycle with all paths of length  $\geq L$  killed.

## 10.4 Indecomposable representations of Nakayama algebras

**Definition 10.28** (The uniserial modules  $M(i, \ell)$ ). Fix  $Q = C_{L+1}$  with vertices  $Q_0 = \{0, \dots, L\}$  and arrows  $a_i : i-1 \rightarrow i$  for  $i = 1, \dots, L$  and  $a_{L+1} : L \rightarrow 0$ . For  $i \in Q_0$  and  $0 \leq \ell \leq L$ , define a representation  $M(i, \ell)$  of  $Q$  by

$$M(i, \ell)_j := \begin{cases} k, & j \in \{i, i+1, \dots, i+\ell-1\} \pmod{L+1}, \\ 0, & \text{otherwise,} \end{cases}$$

and for each arrow  $a_j : j-1 \rightarrow j$ ,

$$M(i, \ell)_{a_j} := \begin{cases} \text{id}_k, & \text{if both source and target are } k, \\ 0, & \text{otherwise.} \end{cases}$$

**Lemma 10.29** (The relations vanish on  $M(i, \ell)$ ). Let  $Q = C_{L+1}$  and, for each  $l \in \{1, \dots, L+1\}$ , let

$$p_l := a_{l-1} \cdots a_1 a_{L+1} \cdots a_{l+1}$$

be the oriented path of length  $L$  going once around the cycle and omitting  $a_l$ . Set  $A := kQ/\langle p_1, \dots, p_{L+1} \rangle$ . Then, for every  $i \in Q_0$  and  $1 \leq \ell \leq L$ , the representation  $M(i, \ell)$  factors through  $A$ , hence defines a finite-dimensional left  $A$ -module.

*Proof.* Fix  $i$  and  $\ell$ . Each path  $p_l$  has length  $L$ . Since the support of  $M(i, \ell)$  contains only  $\ell \leq L$  consecutive vertices, the corresponding composition along  $p_l$  necessarily passes through a vertex where  $M(i, \ell)$  is zero. Therefore  $p_l$  acts by the zero map on  $M(i, \ell)$  for every  $l$ , so all generators of the ideal  $\langle p_1, \dots, p_L \rangle$  act trivially. Hence the  $kQ$ -action factors through  $A$ .  $\square$



**Lemma 10.30** (Uniseriality of  $M(i, \ell)$ ). For every  $i \in Q_0$  and  $1 \leq \ell \leq L+1$ , the  $A$ -module  $M(i, \ell)$  is uniserial, and in particular indecomposable.

*Proof.* Any nonzero submodule  $N \subseteq M(i, \ell)$  must contain the one-dimensional space at some vertex in the support. Because all arrow maps inside the support are identities, containing a vertex forces  $N$  to contain all vertices reachable along the directed arrows within the support. Consequently, the nonzero submodules are exactly the modules  $M(i+t, \ell-t)$  for  $t = 0, \dots, \ell-1$  (indices taken modulo  $n$ ), and they form a chain under inclusion. Thus submodules are totally ordered and  $M(i, \ell)$  is uniserial.  $\square$

**Lemma 10.31** (Endomorphisms of  $M(i, \ell)$ ). For every  $i \in Q_0$  and  $1 \leq \ell \leq L$ , one has  $\text{End}_A(M(i, \ell)) \cong k$ .

*Proof.* Let  $f \in \text{End}_A(M(i, \ell))$ . On each vertex  $j$  in the support,  $f$  is multiplication by some scalar  $\lambda_j \in k$ . For each arrow inside the support, the arrow map is  $\text{id}_k$ , so commutativity forces  $\lambda_{j+1} = \lambda_j$ . Hence all  $\lambda_j$  coincide with a single scalar  $\lambda$ , and  $f = \lambda \text{id}$ .  $\square$

**Lemma 10.32** (Projective truncations realize  $M(i, \ell)$ ). Let  $P(i) := Ae_i$  be the indecomposable projective at vertex  $i$  and let  $J := J(A)$ . Then, for every  $i \in Q_0$  and  $1 \leq \ell \leq L$ , there is an isomorphism of  $A$ -modules

$$P(i)/J^\ell P(i) \cong M(i, \ell).$$

*Proof.* By Proposition 10.22 and the previous subsection,  $A$  is a cyclic Nakayama algebra with uniform Kupisch series  $(L, \dots, L)$ . In particular,  $P(i)$  is uniserial of length  $L$ , and  $J^t P(i)$  is spanned by residue classes of paths of length  $\geq t$  starting at  $i$ . Thus  $P(i)/J^\ell P(i)$  has a basis given by the residue classes of the unique paths of lengths  $0, 1, \dots, \ell-1$  starting at  $i$ . Under the induced quiver description, this quotient has one-dimensional spaces on the vertices  $i, i+1, \dots, i+\ell-1 \pmod{L+1}$  and identity maps along arrows inside the support, which is exactly  $M(i, \ell)$ .  $\square$

**Proposition 10.33** (Indecomposable representations of the cyclic truncated Nakayama algebra). Let  $Q = C_{L+1}$  be the oriented cyclic quiver with arrows  $a_i : i \rightarrow i+1$  for  $i = 1, \dots, L$  and  $a_{L+1} : L \rightarrow 0$ . For each  $l \in \{1, \dots, L+1\}$ , let  $p_l$  be the path of length  $L$  going once around the cycle and omitting  $a_l$ , and set

$$A := kQ / \langle p_1, \dots, p_{L+1} \rangle.$$

Then every finite-dimensional indecomposable left  $A$ -module is isomorphic to a unique module  $M(i, \ell)$  with  $i \in Q_0$  and  $1 \leq \ell \leq L$ . Equivalently, every indecomposable  $A$ -module is of the form  $P(i)/J^\ell P(i)$  for a unique pair  $(i, \ell)$ . In particular,  $A$  has exactly  $(L+1)L$  indecomposable modules up to isomorphism.

*Proof.* By Lemma 10.29, each  $M(i, \ell)$  is a well-defined finite-dimensional  $A$ -module, and by Lemma 10.30 it is indecomposable.

Conversely, since  $A$  is Nakayama (Proposition 10.22), every indecomposable  $A$ -module is uniserial and hence a quotient of a unique indecomposable projective  $P(i) = Ae_i$ . Its composition length is some  $\ell$  with  $1 \leq \ell \leq \ell(P(i)) = L$ , and uniseriality forces the quotient to be  $P(i)/J^\ell P(i)$ . By Lemma 10.32, this module is isomorphic to  $M(i, \ell)$ .

Uniqueness follows because the top simple module determines  $i$  and the length determines  $\ell$ . Counting pairs  $(i, \ell)$  with  $i \in Q_0$  and  $1 \leq \ell \leq L$  gives  $(L+1)L$ .  $\square$

More details about this can be found in the following lecture notes: Noncommutative Algebra 2: Representations of finite-dimensional algebras in particular theorem and corollary section 1.6. We have a Krull schmidt theorem for these Nakayama algebras.

**Proposition 10.34** (Krull–Schmidt for the cyclic truncated Nakayama algebra). Let  $A = kQ/I$  be the cyclic truncated Nakayama algebra, where  $Q = C_{L+1}$  is the oriented  $L + 1$ -cycle and  $I$  is the ideal generated by the length- $(L)$  paths  $p_1, \dots, p_{L+1}$  (equivalently, all paths of length  $\geq L + 1$  vanish in  $A$ ). Let  $\text{mod-}A$  denote the category of finite-dimensional left  $A$ -modules. Then every  $M \in \text{mod-}A$  admits a decomposition as a finite direct sum of indecomposable  $A$ -modules. Moreover, this decomposition is unique up to reordering and isomorphism of the indecomposable summands. Equivalently, using the explicit indecomposables  $M(i, \ell)$  from Proposition 10.33, there exist uniquely determined multiplicities  $m_{i, \ell} \in \mathbb{Z}_{\geq 0}$  such that

$$M \cong \bigoplus_{i=0}^L \bigoplus_{\ell=1}^L M(i, \ell)^{\oplus m_{i, \ell}}.$$

*Proof sketch.* Since  $A$  is finite-dimensional over  $k$ , it is left artinian (and left noetherian), hence every object of  $\text{mod-}A$  has finite length. In particular,  $\text{mod-}A$  is a Krull–Schmidt category: every object decomposes as a finite direct sum of indecomposables, and such a decomposition is unique up to permutation and isomorphism.

Concretely, existence follows by choosing a nonzero module  $M$  and repeatedly splitting off a nontrivial direct summand until all summands are indecomposable; the process terminates because  $M$  has finite length. Uniqueness follows because each indecomposable has a local endomorphism ring (for example, by Lemma 10.31 in the explicit classification), and the classical Krull–Schmidt–Azumaya argument applies.

Finally, Proposition 10.33 classifies the indecomposables as the modules  $M(i, \ell)$ , so any Krull–Schmidt decomposition can be rewritten uniquely in the stated multiplicity form.  $\square$

**Note:** This enables to parameterize the orbits of  $\Gamma$  with the indecomposables representations.

**Orbits, isomorphism classes, and multiplicity patterns.** Fix a finite-dimensional  $k$ -algebra  $A$  and a dimension vector  $d = (d_0, \dots, d_L)$ . Recall that

$$\text{Rep}_d(A) = \left\{ \text{left } A\text{-module structures on } V = \bigoplus_{i=0}^L k^{d_i} \text{ such that each idempotent } e_i \text{ acts as the projection onto } k^{d_i} \right\}.$$

Let

$$G_d := \prod_{i=0}^L \text{GL}(k^{d_i})$$

act on  $\text{Rep}_d(A)$  by *change of bases at the vertices* (equivalently, by transport of structure): for  $g = (g_0, \dots, g_L) \in G_d$  and a module structure  $\rho : A \rightarrow \text{End}_k(V)$ , define a new module structure  $g \cdot \rho$  by

$$(g \cdot \rho)(a) := g \rho(a) g^{-1}, \quad g := \bigoplus_{i=1}^n g_i \in \text{GL}(V).$$

When  $A \simeq kQ/I$  is presented as a bound quiver algebra, this action is equivalently given on arrow maps by  $(g \cdot \phi)_a = g_{t(a)} \phi_a g_{s(a)}^{-1}$ . Then two points  $\rho, \rho' \in \text{Rep}_d(A)$  lie in the same  $G_d$ -orbit if and only if the corresponding  $A$ -modules  $(V, \rho)$  and  $(V, \rho')$  are isomorphic in  $\text{mod-}A$ . Consequently, there is a canonical bijection

$$(G_d \backslash \text{Rep}_d(A)) \cong \left\{ \text{isomorphism classes of } A\text{-modules } M \text{ with } \dim(e_i M) = d_i \text{ for all } i \right\}.$$

Now specialize to the cyclic truncated Nakayama algebra  $A = kQ/I$  considered above. Its indecomposable modules are the uniserials  $M(i, \ell)$  with  $i \in \{0, \dots, L\}$  (indices taken modulo  $L+1$ ) and  $1 \leq \ell \leq L$  (Proposition 10.33). By the Krull–Schmidt theorem (Proposition 10.34), every finite-dimensional  $A$ -module  $M$  admits a decomposition

$$M \simeq \bigoplus_{i=0}^L \bigoplus_{\ell=1}^L M(i, \ell)^{\oplus m_{i, \ell}}, \quad m_{i, \ell} \in \mathbb{Z}_{\geq 0},$$

and this decomposition is unique up to permutation of summands. The multiplicities  $(m_{i, \ell})$  determine (and are constrained by) the dimension vector  $d$  via the *covering equations*

$$d_k = \sum_{i=0}^L \sum_{\ell=1}^L m_{i, \ell} \mathbf{1}\{k \in \{i, i+1, \dots, i+\ell-1\} \pmod{L+1}\}, \quad k = 0, \dots, L, \quad (1)$$

since  $M(i, \ell)$  contributes a one-dimensional space at vertex  $k$  exactly when  $k$  lies in its cyclic support interval of length  $\ell$  starting at  $i$ . Define the set of *multiplicity patterns*

$$M_d^+(A) := \left\{ (m_{i, \ell}) \in \mathbb{Z}_{\geq 0}^{L(L+1)} \mid (1) \text{ holds for all } k \right\}.$$

Then the above discussion yields a concrete parametrization of orbits: every  $G_d$ -orbit in  $\text{Rep}_d(A)$  corresponds to a unique isomorphism class in  $\text{mod-}A$ , and hence to a unique multiplicity pattern  $(m_{i, \ell}) \in M_d^+(A)$  via its Krull–Schmidt decomposition. In other words, we obtain canonical bijections

$$G_d \backslash \text{Rep}_d(A) \cong \left\{ \text{isomorphism classes in } \text{mod-}A \text{ with dimension vector } d \right\} \cong M_d^+(A),$$

where the last map sends an  $A$ -module to the multiplicities of the indecomposables  $M(i, \ell)$  in its Krull–Schmidt decomposition.

**Proposition 10.35** (Orbit stratification of  $\Gamma_d$  by indecomposable Nakayama modules). Fix  $L \geq 1$  and let  $Q = C_{L+1}$  be the oriented cyclic quiver on vertices  $\{0, \dots, L\}$ . Let  $A := kQ/I$  be the bound quiver algebra, where  $I$  is the ideal generated by the length- $L$  paths  $p_1, \dots, p_{L+1}$  obtained by going once around the cycle and omitting one arrow. Let  $d = (d_0, \dots, d_L)$  be a dimension vector.

Define

$$\Gamma_d := \{ \phi \in \text{Rep}_d(Q) \mid d\mu_\ell(\phi) = 0 \text{ for all } \ell = 1, \dots, L \},$$

where

$$d\mu_\ell(\phi) := \phi_{\ell-1} \cdots \phi_1 \phi_{L+1} \phi_L \cdots \phi_{\ell+1}.$$

Let

$$G_d := \prod_{i=0}^L \text{GL}(k^{d_i})$$

act on  $\Gamma_d$  by base change at the vertices.

Then the following hold.

1. There is a canonical identification

$$\Gamma_d \cong \text{Rep}_d(A),$$

under which the  $G_d$ -action coincides with transport of structure on  $A$ -modules.

2. The algebra  $A$  is a finite-dimensional cyclic Nakayama algebra. Its indecomposable finite-dimensional left modules are, up to isomorphism, the uniserial modules

$$M(i, \ell), \quad i \in \{0, \dots, L\}, \quad 1 \leq \ell \leq L,$$

supported on  $\ell$  consecutive vertices starting at  $i$  (indices taken modulo  $L + 1$ ).

3. Every  $A$ -module  $M$  of dimension vector  $d$  admits a unique (up to permutation) Krull–Schmidt decomposition

$$M \cong \bigoplus_{i=0}^L \bigoplus_{\ell=1}^L M(i, \ell)^{\oplus m_{i, \ell}},$$

for uniquely determined multiplicities  $m_{i, \ell} \in \mathbb{Z}_{\geq 0}$  satisfying the dimension constraints

$$d_k = \sum_{i=0}^L \sum_{\ell=1}^L m_{i, \ell} \mathbf{1}\{k \in \{i, i+1, \dots, i+\ell-1\} \pmod{L+1}\}, \quad k = 0, \dots, L.$$

Denote by  $\mathcal{M}_d^+(A)$  the finite set of all such multiplicity patterns  $m = (m_{i, \ell})$ .

4. For each  $m \in \mathcal{M}_d^+(A)$ , define the stratum

$$\Gamma_d^{(m)} := \left\{ \phi \in \Gamma_d \left| F(\phi) \cong \bigoplus_{i, \ell} M(i, \ell)^{\oplus m_{i, \ell}} \right. \right\},$$

where  $F(\phi)$  denotes the associated  $A$ -module.

Then  $\Gamma_d^{(m)}$  is a single  $G_d$ -orbit.

5. The variety  $\Gamma_d$  decomposes as a disjoint union of these orbits:

$$\Gamma_d = \bigsqcup_{m \in \mathcal{M}_d^+(A)} \Gamma_d^{(m)} = \bigsqcup_{m \in \mathcal{M}_d^+(A)} G_d \cdot \phi_m,$$

where  $\phi_m$  is any representative whose associated  $A$ -module has multiplicity pattern  $m$ .

In particular, the  $G_d$ -orbits in  $\Gamma_d$  are canonically parameterized by the multiplicity data of indecomposable  $A$ -modules, i.e. by the set  $\mathcal{M}_d^+(A)$ .

## 11 Stratification of $\Gamma_{XY}$

While the indecomposable representation of  $\Gamma$  are parameterized by the  $M(i, \ell)$  from the previous modules over Nakayama algebras we are actually interested in decomposing:

$$\Gamma_\sigma := \{(\phi_1, \dots, \phi_{L+1}) \in \text{Rep}_d(C_{L+1}) \mid \phi_{l-1} \cdots \phi_1 \phi_{L+1} \phi_L \cdots \phi_{l+1} = 0, \phi_{L+1} = \sigma\}.$$

where the map  $\sigma \in \text{Hom}(V_L, V_0)$  is fixed (in DLN it is  $\Sigma_{XY} U_Q$ ). The map

$$d\mu_{l,\sigma} := \phi_{l-1} \cdots \phi_1 \sigma \phi_L \cdots \phi_{l+1}$$

is no longer  $G_d$ -equivariant.

**Definition 11.1** (Stabilizer group of the fixed arrow). Let  $\sigma \in \text{Hom}(V_L, V_1)$  be fixed. Define

$$\text{Stab}(\sigma) := \{(g_1, g_L) \in \text{GL}(V_1) \times \text{GL}(V_L) \mid g_1 \sigma g_L^{-1} = \sigma\},$$

and the corresponding subgroup of the base-change group

$$G_\sigma := \{g = (g_1, \dots, g_L) \in \prod_{l=1}^L \text{GL}(V_l) \mid (g_1, g_L) \in \text{Stab}(\sigma)\}.$$

**Proposition 11.2** ( $G_\sigma$ -equivariance of the cyclic composition maps). Fix  $\sigma \in \text{Hom}(V_L, V_1)$  and consider the representation space with  $\phi_{L+1} = \sigma$ . For  $\phi = (\phi_1, \dots, \phi_L, \sigma)$  and each  $l = 1, \dots, L$ , define

$$d\mu_{l,\sigma}(\phi) := \phi_{l-1} \cdots \phi_1 \sigma \phi_L \cdots \phi_{l+1} \in \text{Hom}(V_{l+1}, V_l),$$

with the convention that an empty product is the identity. Then, for all  $g \in G_\sigma$  and all  $l$ ,

$$d\mu_{l,\sigma}(g \cdot \phi) = g_l \mu_{l,\sigma}(\phi) g_{l+1}^{-1}.$$

In particular, the subset  $\Gamma_\sigma := \bigcap_{l=1}^L d\mu_{l,\sigma}^{-1}(0)$  is  $G_\sigma$ -invariant.

*Proof.* Expand  $d\mu_{l,\sigma}(g \cdot \phi)$  as a product of the maps  $(g \cdot \phi)_i$ . All intermediate base-change factors cancel. The only remaining factors are  $g_l$  on the left and  $g_{l+1}^{-1}$  on the right. The middle arrow is  $(g \cdot \phi)_{L+1} = g_1 \sigma g_L^{-1}$ , which equals  $\sigma$  because  $g \in G_\sigma$ .  $\square$

For all modules in  $\Gamma_\sigma$  Krull-Schmidt applies (as a category of  $A$ -modules where  $A$  is finite dimensional). We want to parameterize the orbits of  $\Gamma_\sigma$  under the action of  $G_\sigma$  and study their geometry (e.g. codimension).

**Lemma 11.3** (Intersection of a  $G_d$ -orbit with a fixed-arrow fiber). Let  $L \geq 2$  and let  $V_0, \dots, V_L$  be finite-dimensional vector spaces over an algebraically closed field  $\mathbb{k}$ . Set

$$G_d := \prod_{i=0}^L \text{GL}(V_i).$$

Let  $\Gamma_d$  be an affine variety equipped with an algebraic action of  $G_d$ . Assume there exists a morphism

$$F : \Gamma_d \longrightarrow \text{Hom}(V_L, V_0)$$

satisfying the equivariance property

$$F(g \cdot x) = g_0 F(x) g_L^{-1} \quad \text{for all } g = (g_0, \dots, g_L) \in G_d, x \in \Gamma_d. \quad (2)$$

Fix  $\sigma \in \text{Hom}(V_L, V_0)$  and define the fiber

$$\Gamma_\sigma := F^{-1}(\sigma).$$

Define the stabilizer subgroup

$$G_\sigma := \{g = (g_0, \dots, g_L) \in G_d \mid g_0 \sigma g_L^{-1} = \sigma\}.$$

Let  $\mathcal{O} \subset \Gamma_d$  be a single  $G_d$ -orbit. Then either  $\mathcal{O} \cap \Gamma_\sigma = \emptyset$ , or else for every  $x \in \mathcal{O} \cap \Gamma_\sigma$  one has

$$\mathcal{O} \cap \Gamma_\sigma = G_\sigma \cdot x.$$

In particular, whenever  $\mathcal{O} \cap \Gamma_\sigma \neq \emptyset$ , this intersection consists of a single  $G_\sigma$ -orbit.

*Proof.* We proceed in several steps.

**Step 1:  $\Gamma_\sigma$  is stable under the action of  $G_\sigma$ .**

Let  $x \in \Gamma_\sigma$ , so by definition  $F(x) = \sigma$ . Let  $g \in G_\sigma$ . Using the equivariance property (2),

$$F(g \cdot x) = g_1 F(x) g_L^{-1} = g_1 \sigma g_L^{-1}.$$

Since  $g \in G_\sigma$ , we have  $g_1 \sigma g_L^{-1} = \sigma$ , and hence  $F(g \cdot x) = \sigma$ . Therefore  $g \cdot x \in \Gamma_\sigma$ . This shows that  $\Gamma_\sigma$  is invariant under the action of  $G_\sigma$  and, in particular,

$$G_\sigma \cdot x \subseteq \Gamma_\sigma \quad \text{for all } x \in \Gamma_\sigma. \quad (*)$$

**Step 2: If  $x \in \mathcal{O} \cap \Gamma_\sigma$ , then  $G_\sigma \cdot x \subseteq \mathcal{O} \cap \Gamma_\sigma$ .**

Let  $x \in \mathcal{O} \cap \Gamma_\sigma$ . Since  $\mathcal{O}$  is a  $G_d$ -orbit, it is stable under the action of  $G_d$  and hence under the subgroup  $G_\sigma \subseteq G_d$ . Thus  $G_\sigma \cdot x \subseteq \mathcal{O}$ . By (\*),  $G_\sigma \cdot x \subseteq \Gamma_\sigma$ . Combining these two inclusions gives

$$G_\sigma \cdot x \subseteq \mathcal{O} \cap \Gamma_\sigma. \quad (**)$$

**Step 3: Any two points of  $\mathcal{O} \cap \Gamma_\sigma$  differ by the action of  $G_\sigma$ .**

Assume  $x, y \in \mathcal{O} \cap \Gamma_\sigma$ . Because  $x$  and  $y$  lie in the same  $G_d$ -orbit  $\mathcal{O}$ , there exists  $g = (g_1, \dots, g_L) \in G_d$  such that

$$y = g \cdot x.$$

Applying  $F$  and using (2), we obtain

$$F(y) = F(g \cdot x) = g_1 F(x) g_L^{-1}.$$

Since  $x, y \in \Gamma_\sigma$ , we have  $F(x) = F(y) = \sigma$ . Therefore

$$\sigma = g_1 \sigma g_L^{-1}.$$

By definition of  $G_\sigma$ , this implies  $g \in G_\sigma$ . Hence  $y = g \cdot x \in G_\sigma \cdot x$ , and we conclude that

$$\mathcal{O} \cap \Gamma_\sigma \subseteq G_\sigma \cdot x. \quad (***)$$

**Step 4: Conclusion.**

If  $\mathcal{O} \cap \Gamma_\sigma = \emptyset$ , there is nothing to prove. Otherwise choose any  $x \in \mathcal{O} \cap \Gamma_\sigma$ . Inclusion  $(**)$  gives  $G_\sigma \cdot x \subseteq \mathcal{O} \cap \Gamma_\sigma$ , while inclusion  $(***)$  gives the reverse inclusion. Therefore

$$\mathcal{O} \cap \Gamma_\sigma = G_\sigma \cdot x.$$

This proves that whenever the intersection is not empty, it consists of a single  $G_\sigma$ -orbit. The lemma follows.  $\square$

**Proposition 11.4** (Rank label for Nakayama orbits and the fiber non-emptiness criterion over  $\mathbb{R}$ ). Fix  $L \geq 2$  and finite-dimensional real vector spaces  $V_1, \dots, V_L$ . Let

$$G_d := \prod_{i=1}^L GL(V_i) \quad \text{and} \quad W := \text{Hom}_{\mathbb{R}}(V_L, V_1).$$

Let  $\Gamma_d$  be a real algebraic variety equipped with an algebraic action of  $G_d$ . Assume there is a morphism of real varieties

$$F : \Gamma_d \longrightarrow W$$

satisfying the equivariance identity

$$F((g_1, \dots, g_L) \cdot x) = g_1 F(x) g_L^{-1} \quad \text{for all } (g_1, \dots, g_L) \in G_d, x \in \Gamma_d. \quad (3)$$

**Definition (Orbit label and rank label).** Assume  $\Gamma_d$  admits a decomposition into  $G_d$ -orbits indexed by a parameter set  $\Lambda$ ,

$$\Gamma_d = \bigsqcup_{\lambda \in \Lambda} \mathcal{O}_\lambda,$$

where  $\mathcal{O}_\lambda$  denotes the  $G_d$ -orbit corresponding to  $\lambda$  (in the applications of this paper,  $\lambda$  encodes the Krull–Schmidt multiplicities of the Nakayama indecomposables). For each  $\lambda \in \Lambda$ , define the integer

$$r(\lambda) := \text{rank}(F(x)) \quad \text{for any } x \in \mathcal{O}_\lambda. \quad (4)$$

Then the following statements hold.

- (i) The integer  $r(\lambda)$  in (4) is well-defined, i.e. it does not depend on the choice of  $x \in \mathcal{O}_\lambda$ .
- (ii) Fix  $\sigma \in W$  and define the fiber  $\Gamma_\sigma := F^{-1}(\sigma)$ . Then for every  $\lambda \in \Lambda$ ,

$$\mathcal{O}_\lambda \cap \Gamma_\sigma \neq \emptyset \quad \Longleftrightarrow \quad r(\lambda) = \text{rank}(\sigma). \quad (5)$$

*Proof.* We prove (i) and (ii) in separate steps.

**Step 1: A basic invariance identity for ranks under left–right change of bases.** Let  $A \in GL(V_1)$ ,  $B \in GL(V_L)$ , and  $T \in W = \text{Hom}_{\mathbb{R}}(V_L, V_1)$ . Consider the map  $T' := A T B^{-1} \in W$ . We have  $\text{rank}(T') = \text{rank}(T)$ .

**Step 2: Proof of (i) (well-definedness of  $r(\lambda)$ ).** Fix  $\lambda \in \Lambda$  and choose two points  $x, y \in \mathcal{O}_\lambda$ . By definition of orbit, there exists  $g = (g_1, \dots, g_L) \in G_d$  such that

$$y = g \cdot x.$$

Apply  $F$  to both sides and use the equivariance identity (3):

$$F(y) = F(g \cdot x) = g_1 F(x) g_L^{-1}.$$

Now apply Step 1 with  $A = g_1$ ,  $B = g_L$ , and  $T = F(x)$ . This yields

$$\text{rank}(F(y)) = \text{rank}(g_1 F(x) g_L^{-1}) = \text{rank}(F(x)).$$

Thus the rank of  $F(\cdot)$  is constant along  $\mathcal{O}_\lambda$ , so the integer  $r(\lambda)$  defined by (4) is independent of the choice of  $x \in \mathcal{O}_\lambda$ . This proves (i).

**Step 3: A real classification fact for linear maps by rank.** Let  $S, T \in W = \text{Hom}_{\mathbb{R}}(V_L, V_1)$ . We use the following statement:

*Claim.* If  $\text{rank}(S) = \text{rank}(T)$ , then there exist  $A \in GL(V_1)$  and  $B \in GL(V_L)$  such that

$$S = A T B^{-1}. \quad (6)$$

**Step 4: Proof of (ii), forward implication.** Assume  $\mathcal{O}_\lambda \cap \Gamma_\sigma \neq \emptyset$ . Choose  $x \in \mathcal{O}_\lambda \cap \Gamma_\sigma$ . By definition of  $\Gamma_\sigma = F^{-1}(\sigma)$  we have  $F(x) = \sigma$ , hence

$$\text{rank}(\sigma) = \text{rank}(F(x)).$$

By definition of  $r(\lambda)$  (and Step 2),  $\text{rank}(F(x)) = r(\lambda)$  for any  $x \in \mathcal{O}_\lambda$ . Therefore  $\text{rank}(\sigma) = r(\lambda)$ , proving the forward direction of (5).

**Step 5: Proof of (ii), reverse implication.** Assume  $r(\lambda) = \text{rank}(\sigma)$ . Choose any  $x_0 \in \mathcal{O}_\lambda$ . Then, by definition of  $r(\lambda)$ ,

$$\text{rank}(F(x_0)) = r(\lambda) = \text{rank}(\sigma).$$

By the claim in Step 3 (applied to  $T = F(x_0)$  and  $S = \sigma$ ), there exist  $A \in GL(V_1)$  and  $B \in GL(V_L)$  such that

$$\sigma = A F(x_0) B^{-1}.$$

Define  $g = (g_1, \dots, g_L) \in G_d$  by setting

$$g_1 := A, \quad g_L := B, \quad g_i := \text{Id}_{V_i} \text{ for } i = 2, \dots, L-1.$$

Then  $g \in G_d$  and, using equivariance (3),

$$F(g \cdot x_0) = g_1 F(x_0) g_L^{-1} = A F(x_0) B^{-1} = \sigma.$$

Hence  $g \cdot x_0 \in \Gamma_\sigma$ . Also, since  $x_0 \in \mathcal{O}_\lambda$  and  $\mathcal{O}_\lambda$  is a  $G_d$ -orbit, we have  $g \cdot x_0 \in \mathcal{O}_\lambda$ . Therefore  $g \cdot x_0 \in \mathcal{O}_\lambda \cap \Gamma_\sigma$ , so  $\mathcal{O}_\lambda \cap \Gamma_\sigma \neq \emptyset$ .

Combining Steps 4 and 5 yields (5). This completes the proof.  $\square$

**Proposition 11.5** (Rank stratification of  $\Gamma_{XY}$  by Nakayama types). Let  $d = (d_1, \dots, d_L)$  be a dimension vector and let

$$\Gamma_d \subset \text{Rep}_d(C_{L+1}^{\text{op}})$$



be the variety defined in Section §9. Recall that  $\Gamma_d$  admits a decomposition into  $G_d$ -orbits indexed by

$$\mathcal{M}_d^+ = \left\{ (m_{i,\ell}) \left| \sum_{i,\ell} m_{i,\ell} \underline{\dim} M(i,\ell) = d \right. \right\},$$

where  $M(i,\ell)$  are the indecomposable modules of the cyclic Nakayama algebra and  $(m_{i,\ell})$  records their multiplicities. For  $\mathbf{m} = (m_{i,\ell}) \in \mathcal{M}_d^+$ , denote by  $\mathcal{O}_{\mathbf{m}} \subset \Gamma_d$  the corresponding  $G_d$ -orbit.

Let

$$F : \Gamma_d \longrightarrow \text{Hom}(V_L, V_1), \quad F(\phi) = \phi_{L+1},$$

and fix a linear map  $\sigma \in \text{Hom}(V_L, V_1)$ . Define

$$\Gamma_{XY} := \Gamma_{\sigma} := F^{-1}(\sigma), \quad G_{\sigma} := \{g \in G_d : g_1 \sigma g_L^{-1} = \sigma\}.$$

For  $\mathbf{m} \in \mathcal{M}_d^+$ , define the integer

$$r(\mathbf{m}) := \text{rank}(F(\phi)) \quad \text{for any } \phi \in \mathcal{O}_{\mathbf{m}}.$$

Then:

- (i) The integer  $r(\mathbf{m})$  is well-defined (independent of the choice of  $\phi \in \mathcal{O}_{\mathbf{m}}$ ).
- (ii) For every  $\mathbf{m} \in \mathcal{M}_d^+$ ,

$$\mathcal{O}_{\mathbf{m}} \cap \Gamma_{XY} \neq \emptyset \iff r(\mathbf{m}) = \text{rank}(\sigma).$$

- (iii) If  $\mathcal{O}_{\mathbf{m}} \cap \Gamma_{XY} \neq \emptyset$ , then

$$\mathcal{O}_{\mathbf{m}} \cap \Gamma_{XY}$$

is a single  $G_{\sigma}$ -orbit.

- (iv) Consequently,  $\Gamma_{XY}$  admits the disjoint decomposition

$$\Gamma_{XY} = \bigsqcup_{\mathbf{m} \in \mathcal{M}_d^+ : r(\mathbf{m}) = \text{rank}(\sigma)} (\mathcal{O}_{\mathbf{m}} \cap \Gamma_{XY}),$$

and each stratum in this decomposition is a single  $G_{\sigma}$ -orbit.

**Proposition 11.6** (Rank formula for the closing arrow in terms of Nakayama multiplicities). Work over  $\mathbb{R}$ . Fix  $L \geq 2$  and a dimension vector  $d = (d_1, \dots, d_L)$ . Let  $\Gamma_d \subset \text{Rep}_d(C_{L+1}^{\text{op}})$  be the variety from the document, and recall that  $\Gamma_d$  decomposes into  $G_d$ -orbits indexed by  $\mathcal{M}_d^+$ , i.e. for each

$$\mathbf{m} = (m_{i,\ell}) \in \mathcal{M}_d^+$$

there is a corresponding  $G_d$ -orbit  $\mathcal{O}_{\mathbf{m}} \subset \Gamma_d$  such that every  $\phi \in \mathcal{O}_{\mathbf{m}}$  has associated Nakayama module

$$M(\phi) \cong \bigoplus_{i,\ell} M(i,\ell)^{\oplus m_{i,\ell}}$$

(up to isomorphism), where  $M(i,\ell)$  denote the indecomposable modules of the cyclic Nakayama algebra.

Define the morphism

$$F : \Gamma_d \rightarrow \text{Hom}_{\mathbb{R}}(V_L, V_1), \quad F(\phi) = \phi_{L+1}.$$

For  $\mathbf{m} \in \mathcal{M}_d^+$  define

$$r(\mathbf{m}) := \text{rank}(F(\phi)) \quad \text{for any } \phi \in \mathcal{O}_{\mathbf{m}},$$

which is well-defined by Proposition 11.4.

**Definition (support interval and activity indicator).** For an indecomposable  $M(i, \ell)$ , define its *support interval* to be the cyclic set of vertices

$$\text{Supp}(i, \ell) := \{i, i+1, \dots, i+\ell-1\} \subset \mathbb{Z}/L\mathbb{Z},$$

where addition is taken modulo  $L$  and we identify vertices with their residue classes. Say that  $M(i, \ell)$  is  $(L \rightarrow 1)$ -active if the arrow  $L \rightarrow 1$  acts nontrivially on  $M(i, \ell)$ , i.e. if the structure map

$$M(i, \ell)(L) \longrightarrow M(i, \ell)(1)$$

induced by the quiver arrow  $L \rightarrow 1$  is nonzero. Define the indicator

$$\epsilon(i, \ell) := \begin{cases} 1, & \text{if } M(i, \ell) \text{ is } (L \rightarrow 1)\text{-active,} \\ 0, & \text{otherwise.} \end{cases}$$

Then for every  $\mathbf{m} = (m_{i, \ell}) \in \mathcal{M}_d^+$  one has the formula

$$r(\mathbf{m}) = \sum_{i, \ell} \epsilon(i, \ell) m_{i, \ell}. \quad (7)$$

*Proof.* Fix  $\mathbf{m} = (m_{i, \ell}) \in \mathcal{M}_d^+$ . Choose any  $\phi \in \mathcal{O}_{\mathbf{m}}$  and write  $M := M(\phi)$  for the corresponding Nakayama module. By definition of  $\mathcal{O}_{\mathbf{m}}$ , there exists an isomorphism of modules

$$M \cong \bigoplus_{i, \ell} M(i, \ell)^{\oplus m_{i, \ell}}. \quad (8)$$

We will compute  $\text{rank}(\phi_{L+1}) = \text{rank}(F(\phi))$  from this decomposition.

**Step 1: Describe the vector spaces at vertices 1 and  $L$  using (8).**

For each vertex  $j \in \{1, L\}$  the representation space is  $V_j = M(j)$ . Evaluating (8) at the vertex  $j$  gives an isomorphism of real vector spaces

$$V_j = M(j) \cong \bigoplus_{i, \ell} M(i, \ell)(j)^{\oplus m_{i, \ell}}. \quad (9)$$

For cyclic Nakayama indecomposables,  $M(i, \ell)(j)$  is either 0 or a 1-dimensional real vector space (depending on whether  $j \in \text{Supp}(i, \ell)$ ). Therefore each summand in (9) is either 0 or  $\mathbb{R}^{m_{i, \ell}}$ .

**Step 2: The closing arrow respects the direct-sum decomposition.**

The linear map  $\phi_{L+1} : V_L \rightarrow V_1$  is the structure map on the quiver arrow  $L \rightarrow 1$ . Because  $M$  is a direct sum of submodules as in (8), the action of every quiver arrow, and in particular the arrow  $L \rightarrow 1$ , decomposes as the direct sum of its actions on each direct summand. Concretely,

after choosing identifications as in (9) for  $j = L$  and  $j = 1$ , the map  $\phi_{L+1}$  can be written as a block-diagonal direct sum

$$\phi_{L+1} \cong \bigoplus_{i,\ell} (\phi_{L+1})_{i,\ell}, \quad (10)$$

where

$$(\phi_{L+1})_{i,\ell} : M(i,\ell)(L)^{\oplus m_{i,\ell}} \longrightarrow M(i,\ell)(1)^{\oplus m_{i,\ell}}$$

is the map induced by  $L \rightarrow 1$  on the summand  $M(i,\ell)^{\oplus m_{i,\ell}}$ .

**Step 3: Compute each block  $(\phi_{L+1})_{i,\ell}$ .**

Fix a pair  $(i,\ell)$ . There are two cases.

*Case A:*  $\epsilon(i,\ell) = 0$ . By definition,  $\epsilon(i,\ell) = 0$  means the arrow  $L \rightarrow 1$  acts trivially on  $M(i,\ell)$ , i.e. the map

$$M(i,\ell)(L) \rightarrow M(i,\ell)(1)$$

is the zero map. Taking  $m_{i,\ell}$  copies, the induced map

$$(\phi_{L+1})_{i,\ell} : M(i,\ell)(L)^{\oplus m_{i,\ell}} \rightarrow M(i,\ell)(1)^{\oplus m_{i,\ell}}$$

is also the zero map. Hence

$$\text{rank}((\phi_{L+1})_{i,\ell}) = 0. \quad (11)$$

*Case B:*  $\epsilon(i,\ell) = 1$ . By definition,  $\epsilon(i,\ell) = 1$  means the arrow  $L \rightarrow 1$  acts nontrivially on  $M(i,\ell)$ . For Nakayama indecomposables, both spaces  $M(i,\ell)(L)$  and  $M(i,\ell)(1)$  are then 1-dimensional and the map between them is a nonzero linear map  $\mathbb{R} \rightarrow \mathbb{R}$ . Choose bases of these 1-dimensional spaces; in these bases the map is multiplication by a nonzero scalar  $a_{i,\ell} \in \mathbb{R}^\times$ . On  $m_{i,\ell}$  copies, the induced map is multiplication by the same scalar on each copy; in coordinates it is

$$a_{i,\ell} \cdot I_{m_{i,\ell}} : \mathbb{R}^{m_{i,\ell}} \rightarrow \mathbb{R}^{m_{i,\ell}},$$

which is invertible because  $a_{i,\ell} \neq 0$ . Therefore

$$\text{rank}((\phi_{L+1})_{i,\ell}) = m_{i,\ell}. \quad (12)$$

**Step 4: Sum the ranks of the blocks.**

By (10), after choosing bases compatible with the decomposition,  $\phi_{L+1}$  is block diagonal with blocks  $(\phi_{L+1})_{i,\ell}$ . The rank of a block-diagonal linear map is the sum of the ranks of its diagonal blocks (because its image is the direct sum of the images of the blocks). Hence

$$\text{rank}(\phi_{L+1}) = \sum_{i,\ell} \text{rank}((\phi_{L+1})_{i,\ell}).$$

Substituting (11) in Case A and (12) in Case B yields

$$\text{rank}(\phi_{L+1}) = \sum_{i,\ell} \epsilon(i,\ell) m_{i,\ell}.$$

By definition  $r(\mathbf{m}) = \text{rank}(\phi_{L+1})$  for  $\phi \in \mathcal{O}_{\mathbf{m}}$ , so this proves (7).  $\square$

“latex

**Corollary 11.7** (Orbit decomposition of  $\Gamma_{XY}$  indexed by  $\mathcal{M}_d^+$ ). Work over  $\mathbb{R}$  and keep the notation of Proposition 11.5 and Proposition 11.6. In particular,  $\Gamma_{XY} = \Gamma_\sigma := F^{-1}(\sigma)$ , the  $G_d$ -orbits in  $\Gamma_d$  are denoted  $\mathcal{O}_{\mathbf{m}}$  for  $\mathbf{m} \in \mathcal{M}_d^+$ , and  $G_\sigma := \{g \in G_d : g_1 \sigma g_L^{-1} = \sigma\}$ .

Then  $\Gamma_{XY}$  decomposes as a disjoint union of  $G_\sigma$ -orbits indexed by those  $\mathbf{m} \in \mathcal{M}_d^+$  satisfying the rank constraint:

$$\Gamma_{XY} = \bigsqcup_{\mathbf{m} \in \mathcal{M}_d^+ : r(\mathbf{m}) = \text{rank}(\sigma)} (\mathcal{O}_{\mathbf{m}} \cap \Gamma_{XY}),$$

and for each such  $\mathbf{m}$  the set  $\mathcal{O}_{\mathbf{m}} \cap \Gamma_{XY}$  is a single  $G_\sigma$ -orbit.

Moreover, for  $\mathbf{m} = (m_{i,\ell}) \in \mathcal{M}_d^+$  the integer  $r(\mathbf{m})$  is given by the explicit formula

$$r(\mathbf{m}) = \sum_{i,\ell} \epsilon(i,\ell) m_{i,\ell},$$

where  $\epsilon(i,\ell) \in \{0,1\}$  is the  $(L \rightarrow 1)$ -activity indicator of the indecomposable  $M(i,\ell)$  as in Proposition 11.6.

*Proof.* Fix  $\sigma \in \text{Hom}(V_L, V_1)$  and recall  $\Gamma_{XY} = \Gamma_\sigma$ .

First, since  $\Gamma_d = \bigsqcup_{\mathbf{m} \in \mathcal{M}_d^+} \mathcal{O}_{\mathbf{m}}$  is a disjoint union, intersecting with  $\Gamma_\sigma$  gives the disjoint union

$$\Gamma_\sigma = \Gamma_\sigma \cap \Gamma_d = \Gamma_\sigma \cap \left( \bigsqcup_{\mathbf{m} \in \mathcal{M}_d^+} \mathcal{O}_{\mathbf{m}} \right) = \bigsqcup_{\mathbf{m} \in \mathcal{M}_d^+} (\mathcal{O}_{\mathbf{m}} \cap \Gamma_\sigma).$$

By Proposition 11.5, for a given  $\mathbf{m}$  the intersection  $\mathcal{O}_{\mathbf{m}} \cap \Gamma_\sigma$  is nonempty if and only if  $r(\mathbf{m}) = \text{rank}(\sigma)$ . Therefore all terms with  $r(\mathbf{m}) \neq \text{rank}(\sigma)$  vanish, and we obtain

$$\Gamma_\sigma = \bigsqcup_{\mathbf{m} \in \mathcal{M}_d^+ : r(\mathbf{m}) = \text{rank}(\sigma)} (\mathcal{O}_{\mathbf{m}} \cap \Gamma_\sigma).$$

Again by Proposition 11.5, whenever the intersection is nonempty,  $\mathcal{O}_{\mathbf{m}} \cap \Gamma_\sigma$  is a single  $G_\sigma$ -orbit. This proves the stated orbit decomposition.

Finally, the explicit formula  $r(\mathbf{m}) = \sum_{i,\ell} \epsilon(i,\ell) m_{i,\ell}$  is exactly Proposition 11.6.  $\square$

**Remark 11.8** (Parametrization of  $G_\sigma$ -orbits in  $\Gamma_{XY}$ ). The previous corollary can be equivalently phrased as follows. The  $G_\sigma$ -orbits in  $\Gamma_{XY}$  are in bijection with the subset

$$\{\mathbf{m} \in \mathcal{M}_d^+ \mid r(\mathbf{m}) = \text{rank}(\sigma)\},$$

via the correspondence

$$\mathbf{m} \mapsto \mathcal{O}_{\mathbf{m}} \cap \Gamma_\sigma.$$

In particular,  $\Gamma_{XY}$  decomposes as a disjoint union of  $G_\sigma$ -orbits indexed by Nakayama indecomposable multiplicity data satisfying the rank constraint.