Equivariant Structured Positional Rotations: Complete Algebraic Derivation

Abstract

This document provides a complete, boxed derivation of the relative-position property for equivariant structured positional rotations in attention mechanisms. Each step is recorded as a labelled statement $[T\cdot]$ with explicit dependencies so that the argument can be followed mechanically.

Contents

1 Tier 0: Linear-Algebra Primitives

$$\boxed{[T0.1 : \mathbb{N}, \mathrm{Idx}]} \qquad \mathbb{N} \overset{\varnothing}{=} \{1, 2, 3, \dots\} \qquad [\mathrm{deps} : \varnothing]$$
$$d_h, d_c, D \in \mathbb{N}$$
$$H \overset{\varnothing}{=} \{1, \dots, d_h\}, \quad K \overset{\varnothing}{=} \{1, \dots, d_c\}, \quad J_D \overset{\varnothing}{=} \{1, \dots, D\}$$

$$\boxed{[T0.2 : I_n, \delta_{ij}]} \; \delta_{ij} \overset{\varnothing}{=} \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \qquad [\mathrm{deps} : T0.1]$$
$$I_n \overset{\varnothing}{=} [\delta_{ij}]_{i,j=1}^{n}$$

$$\boxed{[T0.3 : \mathbb{R}^{m\times n}, (\cdot)^\top, (\cdot)^*, \mathrm{tr}]} \qquad \mathbb{R}^{m\times n} \overset{\varnothing}{=} \text{set of } m \times n \text{ real matrices} \qquad [\mathrm{deps} : T0.1]$$
$$\mathbb{C}^{m\times n} \overset{\varnothing}{=} \text{set of } m \times n \text{ complex matrices}$$
$$A^\top \in \mathbb{R}^{n\times m} \quad (A \in \mathbb{R}^{m\times n})$$
$$B^* \overset{\varnothing}{=} \overline{B}^\top \quad (B \in \mathbb{C}^{m\times n})$$
$$\mathrm{tr}(M) \overset{\varnothing}{=} \sum_{i=1}^{n} M_{ii} \quad (M \in \mathbb{C}^{n\times n})$$

$$\boxed{[T0.4 : \det]} \qquad \det : \mathbb{C}^{n\times n} \to \mathbb{C} \qquad [\mathrm{deps} : T0.3]$$
$$\det(AB) = \det(A)\det(B)$$
$$\det(A^\top) = \det(A)$$
$$U \text{ invertible} \Rightarrow \det(UAU^{-1}) = \det(A)$$
$$\text{(multiplicativity, transpose invariance, similarity invariance)}$$

$$\boxed{[T0.5 : \oplus, \mathrm{blk}, \det \oplus]} \qquad \bigoplus_{u=1}^{m} M_u \overset{\varnothing}{=} \mathrm{diag}(M_1, \dots, M_m) \qquad [\mathrm{d}$$
$$\det\left(\bigoplus_{u=1}^{m} M_u\right) = \prod_{u=1}^{m} \det(M_u)$$
$$\text{(block triangular matrices have determinant equal to product of block determinants)}$$

$$\boxed{[T0.6 : O, SO, U]} \quad O(n) \overset{\varnothing}{=} \{Q \in \mathbb{R}^{n\times n} \mid Q^\top Q = I_n\} \qquad [\mathrm{deps} : T0.2, T0.3, T0.4]$$
$$SO(n) \overset{\varnothing}{=} \{Q \in O(n) \mid \det Q = 1\}$$
$$U(n) \overset{\varnothing}{=} \{W \in \mathbb{C}^{n\times n} \mid W^* W = I_n\}$$
$$Q \in O(n) \Rightarrow Q^{-1} = Q^\top$$

$$\boxed{[T0.7 : \det O]} \; Q \in O(n) \Rightarrow (\det Q)^2 = \det(Q^\top Q) = \det(I_n) = 1 \qquad [\mathrm{deps} : T0.6, T0.4]$$
$$\Rightarrow \det Q \in \{\pm 1\}$$

$$\boxed{[T0.8 : \sigma, \sigma_{\min}, \text{normal}]} \qquad \sigma(M) \overset{\varnothing}{=} \{\lambda \in \mathbb{C} \mid \exists x \neq 0 : Mx = \lambda x\} \qquad [\text{deps} : T0.3]$$

$$M^*M = MM^* \Rightarrow M \text{ normal}$$

$$\text{Singular values } s_i(M) \overset{\varnothing}{=} \sqrt{\sigma_i(M^*M)}$$

$$\sigma_{\min}(M) \overset{\varnothing}{=} \min_i s_i(M)$$

$$M \text{ normal} \Rightarrow s_i(M) = |\lambda_i| \text{ for } \lambda_i \in \sigma(M)$$

(unitary diagonalisation of normal matrices)

$$\boxed{[T0.9 : |\cdot|, |\cdot|_1]} \qquad |x| \overset{\varnothing}{=} \left(\sum_u x_u^2\right)^{1/2} \quad (x \in \mathbb{R}^m) \qquad [\text{deps} : T0.8]$$

$$|x|_1 \overset{\varnothing}{=} \sum_u |x_u|$$

$$|M| \overset{\varnothing}{=} \sup_{v \neq 0} \frac{|Mv|}{|v|} \quad (M \in \mathbb{R}^{m \times n})$$

$$|AB| \leq |A|\,|B|, \quad |A + B| \leq |A| + |B|$$

$$|a| \overset{\varnothing}{=} |a| \quad (a \in \mathbb{R})$$

(overloaded notation chooses scalar, vector, or operator norm from context)

$$\boxed{[T0.10 : [A, B]]}\ [A, B] \overset{\varnothing}{=} AB - BA \qquad [\text{deps} : T0.3]$$

$$\boxed{[T0.11 : \exp, \log, \exp\ \text{ props}]} \qquad \exp(M) \overset{\varnothing}{=} \sum_{t \geq 0} \frac{1}{t!} M^t \qquad [\text{deps} : T0.3, T0.4, T0.5, T$$

$$\log(X) \text{ locally defined as } \exp^{-1}(X) \text{ near } I_n$$

$$[A, B] = 0 \Rightarrow \exp(A + B) = \exp(A)\exp(B)$$

$$(\exp M)^\top = \exp(M^\top)$$

$$U \text{ invertible} \Rightarrow \exp(UMU^{-1}) = U \exp(M) U^{-1}$$

$$\det(\exp M) = \exp(\operatorname{tr} M)$$

$$\exp\left(\bigoplus_{u=1}^m M_u\right) = \bigoplus_{u=1}^m \exp(M_u)$$

(all properties follow from the power-series definition)

$$\boxed{[T0.12 : \mathcal{C}, O(\cdot), \approx_\varepsilon]}\ \mathcal{C} \overset{\varnothing}{=} \{C, C', C'', \cdots \mid C \geq 0 \text{ finite constants}\} \qquad [\text{deps} : T0.9]$$

$$f = O(g) \Leftrightarrow \exists C \in \mathcal{C} : f \leq Cg$$

$$X \approx_\varepsilon Y \Leftrightarrow |X - Y| \leq \varepsilon$$

$$\boxed{[T0.13 : \text{BCH}]}\ \exp(M)\exp(N) = \exp\big(M + N + \tfrac{1}{2}[M, N] + R(M, N)\big) \qquad [\text{deps} : T0.11, T0.10, T0.12, T0.9]$$

$$|R(M, N)| = O(|[M, N]|^2)$$

(Baker–Campbell–Hausdorff series (local version))

$$\boxed{[T0.14 : \ker, \mathrm{im}, \mathrm{rank}]}\ \ker(M) \stackrel{\varnothing}{=} \{x \mid Mx = 0\} \qquad [\text{deps} : T0.3, T0.1]$$

$$\mathrm{im}(M) \stackrel{\varnothing}{=} \{Mx \mid x\}$$

$$\mathrm{rank}(M) \stackrel{\varnothing}{=} \dim(\mathrm{im}(M))$$

$$\boxed{[T0.15 : \mathrm{proj}]} \qquad P^2 = P \text{ and } P^\top = P \Rightarrow P \text{ orthogonal projector} \qquad [\text{deps} : T0.14, T0.3]$$

$$P, Q \text{ orthogonal projectors and } PQ = 0 \Rightarrow \mathrm{im}(P) \perp \mathrm{im}(Q)$$

$$(\text{characterisation via images})$$

$$\boxed{[T0.16 : \mathrm{skew}_{\mathrm{spec}}]} \qquad S^\top = -S \Rightarrow S \text{ normal} \qquad [\text{deps} : T0.3, T0.8]$$

$$\sigma(S) \subset i\mathbb{R}$$

$$\lambda = i\mu \in \sigma(S) \Rightarrow -\lambda = -i\mu \in \sigma(S)$$

$$(\text{skew-symmetric spectra occur in conjugate-sign pairs})$$

$$\boxed{[T0.17 : J, R_2]}\ J \stackrel{\varnothing}{=} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad R_2(\theta) \stackrel{\varnothing}{=} \exp(\theta J) \qquad [\text{deps} : T0.11]$$

$$\boxed{[T0.18 : R_2 \in SO(2)]} \qquad R_2(\theta)^\top R_2(\theta) = I_2 \qquad [\text{deps} : T0.17, T0.16, T0.11,$$

$$\det R_2(\theta) = 1$$

$$\Rightarrow R_2(\theta) \in SO(2)$$

$$(\text{uses } J^\top = -J \text{ and properties of exp and the determinant})$$

$$\boxed{[T0.19 : \mathrm{Weyl}]}\ \left|\sigma_{\min}(A + E) - \sigma_{\min}(A)\right| \le |E| \qquad [\text{deps} : T0.8, T0.9]$$

$$\boxed{[T0.20 : (X^{-1})^\top]}\ X \text{ invertible } \Rightarrow (X^{-1})^\top = (X^\top)^{-1} \qquad [\text{deps} : T0.3]$$

2 Tier 1: Structured Generators

$$\boxed{[T1.1 : L_k]} \qquad \forall k \in K : \ L_k \in \mathbb{R}^{d_h \times d_h}, \quad L_k^\top = -L_k, \quad [L_a, L_b] = 0 \qquad [\text{deps} : T0.1, T0.3, T0.10, T0.16]$$

$$(\text{model hypothesis: commuting skew-symmetric generators})$$

$$\boxed{[T1.2 : \text{normal } L_k]}\ L_k^\top = -L_k \Rightarrow L_k \text{ normal} \qquad [\text{deps} : T1.1, T0.16]$$

$$\boxed{[T1.3 : A(r)]}\ r \in \mathbb{R}^{d_c} \Rightarrow A(r) \stackrel{\varnothing}{=} \sum_{k=1}^{d_c} r_k L_k \qquad [\text{deps} : T1.1, T0.1]$$

$$\boxed{[T1.4 : A(r)^\top]}\ A(r)^\top = -A(r) \qquad [\text{deps} : T1.3, T1.1, T0.3]$$

$$\boxed{[T1.5 : [A(r), A(s)]]}\ [A(r), A(s)] = 0 \qquad [\text{deps} : T1.3, T1.1, T0.10]$$

$$\boxed{[T1.6 : R_{\mathrm{STR}}]}\ R_{\mathrm{STR}}(r) \stackrel{\varnothing}{=} \exp(A(r)) \qquad [\text{deps} : T1.3, T0.11]$$

3 Tier 2: Joint Block Structure

$\boxed{[T2.1:\text{sim-unitary}]}$ $\{L_k\}_{k\in K}$ commute and are normal $[\text{deps}: T1.2, T1.1, T0.6, T0.16]$

$$\Rightarrow \exists \mathcal{U}_c \in U(d_h):\ \mathcal{U}_c^* L_k \mathcal{U}_c = \operatorname{diag}(i\lambda_{k,1},\dots,i\lambda_{k,d_h})$$
$$\lambda_{k,j}\in\mathbb{R}$$

$\boxed{[T2.2:\text{real 2D planes}]}$ Fix $j \in H:\ \mathcal{U}_c^* L_k \mathcal{U}_c e_j = i\lambda_{k,j} e_j$ $[\text{deps}: T2.1, T0.17, T0.16, T0.3]$

$$\Rightarrow L_k(\Re e_j) = \lambda_{k,j} J(\Re e_j, \Im e_j)$$
$$L_k(\Im e_j) = \lambda_{k,j} J(\Re e_j, \Im e_j)$$
$\Rightarrow \operatorname{span}\{\Re e_j, \Im e_j\}$ is a real, 2-D, L_k-invariant plane

$\boxed{[T2.3:\text{JointBlockDiag}]}$ $\exists m, d_{\text{null}} \in \mathbb{N}:\ 2m + d_{\text{null}} = d_h$ $[\text{deps}: T2.2, T0.6, T0.16, T0.5]$

$$\exists U \in O(d_h):\ U^\top L_k U = \left(\bigoplus_{u=1}^{m} \lambda_{k,u} J\right) \oplus 0_{d_{\text{null}}\times d_{\text{null}}}$$

$\boxed{[T2.4:\beta,\theta]}$ $\beta_u \overset{\varnothing}{=} (\lambda_{1,u},\dots,\lambda_{d_c,u})^\top \in \mathbb{R}^{d_c}$ $[\text{deps}: T2.3, T1.1, T0.1]$

$$\theta_u(r) \overset{\varnothing}{=} \beta_u^\top r \in \mathbb{R}$$

$\boxed{[T2.5:U^\top A(r)U]}$ $U^\top A(r) U = \left(\bigoplus_{u=1}^{m} \theta_u(r) J\right) \oplus 0_{d_{\text{null}}\times d_{\text{null}}}$ $[\text{deps}: T1.3, T2.3, T2.4, T0.17, T0.5]$

$\boxed{[T2.6:\exp\text{blk}]}$ $\exp\big(U^\top A(r) U\big) = \left(\bigoplus_{u=1}^{m} R_2(\theta_u(r))\right) \oplus I_{d_{\text{null}}}$ $[\text{deps}: T2.5, T0.17, T0.11, T0.5]$

$\boxed{[T2.7:R_{\text{STR}}\text{ form}]}$ $R_{\text{STR}}(r) = U\left[\left(\bigoplus_{u=1}^{m} R_2(\theta_u(r))\right) \oplus I_{d_{\text{null}}}\right] U^\top$ $[\text{deps}: T1.6, T2.6, T0.11, T2.3]$

$\boxed{[T2.8:R_{\text{STR}}\in SO]}$ $R_{\text{STR}}(r) \in SO(d_h)$ $[\text{deps}: T2.7, T$

(uses $R_2(\theta)\in SO(2)$, block determinant multiplicativity, and $U \in O(d_h)$)

$\boxed{[T2.9:\text{rel } R_{\text{STR}}]}$ $R_{\text{STR}}(r_i)^\top R_{\text{STR}}(r_j) = R_{\text{STR}}(r_j - r_i)$ $[\text{deps}: T1.4, T1.5, T1.6, T0.11]$

$\boxed{[T2.10:\Pi_{\text{act}},\Pi_{\text{null}},d_{\text{act}}]}$ $\Pi_{\text{act}} \overset{\varnothing}{=} U \operatorname{diag}(I_{2m}, 0)\, U^\top$ $[\text{deps}: T2.3, T0.15, T0.2, T0.14, T0.6]$

$$\Pi_{\text{null}} \overset{\varnothing}{=} U \operatorname{diag}(0, I_{d_{\text{null}}})\, U^\top$$
$$\Pi_{\text{act}}^\top = \Pi_{\text{act}},\ \Pi_{\text{act}}^2 = \Pi_{\text{act}}$$
$$\Pi_{\text{null}}^\top = \Pi_{\text{null}},\ \Pi_{\text{null}}^2 = \Pi_{\text{null}}$$
$$\Pi_{\text{act}}\Pi_{\text{null}} = 0 \Rightarrow \operatorname{im}(\Pi_{\text{act}}) \perp \operatorname{im}(\Pi_{\text{null}})$$
$$d_{\text{act}} \overset{\varnothing}{=} \operatorname{rank}(\Pi_{\text{act}}) = 2m$$

$\boxed{[T2.11:\Pi_{\text{act}}\text{ comm}]}$ $R_{\text{STR}}(r)\Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r)$ $[\text{deps}: T2.7, T2.10, T0.5, T0.6]$

4 Tier 3: Cayley Post-Rotations

$$\boxed{[T3.1 : \text{Cayley}]}\; S \in \mathbb{R}^{d_h \times d_h},\; S^\top = -S \Rightarrow (I+S) \text{ invertible} \qquad [\text{deps} : T0.16, T0.3, T0.2]$$
$$P_{\text{sp}} \overset{\varnothing}{=} (I-S)(I+S)^{-1}$$
$$(I-S)(I+S) = (I+S)(I-S) = I - S^2$$
$$(\text{uses } \sigma(S) \subset i\mathbb{R} \text{ so } 1 + i\mu \neq 0)$$

$$\boxed{[T3.2 : P_{\text{sp}} \in SO, \mathcal{B}]} \qquad P_{\text{sp}}^\top = ((I+S)^{-1})^\top (I-S)^\top = I \qquad [\text{deps} : T3.1, T0.16, T0.20, T0.7, T0.6, T0.4$$
$$\det(P_{\text{sp}}) = \prod_u \frac{1 - i\mu_u}{1 + i\mu_u} = 1$$
$$\mathcal{B} \overset{\varnothing}{=} \{(I-S)(I+S)^{-1} \mid S^\top = -S\} \subseteq SO(d_h)$$

$$\boxed{[T3.3 : \text{Cayley surj}]} \qquad Q \in SO(d_h),\; -1 \notin \sigma(Q) \Rightarrow S_Q \overset{\varnothing}{=} (I-Q)(I+Q)^{-1} \qquad [\text{deps} : T0.16, T0.20, T0.$$
$$S_Q^\top = -S_Q, \quad (I - S_Q)(I + S_Q)^{-1} = Q$$
$$(\text{Cayley transform bijection on } SO(d_h) \text{ with no } -1 \text{ eigenvalues})$$

$$\boxed{[T3.4 : \mathcal{A}]} \qquad \mathcal{A} \overset{\varnothing}{=} \left\{ U \operatorname{diag}(I_{2m}, R_{\text{null}})\, U^\top \,\middle|\, R_{\text{null}} \in SO(d_{\text{null}}) \right\} \subseteq SO(d_h) \qquad [\text{deps} : T2.3, T0.6, T0.7, T0.5, T0.4, T$$
$$R_{\text{null}} \in SO(d_{\text{null}}),\; -1 \notin \sigma(R_{\text{null}}) \Rightarrow U \operatorname{diag}(I_{2m}, R_{\text{null}})\, U^\top \in \mathcal{B}$$

$$\boxed{[T3.5 : R_{\text{sp}}]} \qquad R_{\text{sp}}(r) \overset{\varnothing}{=} R_{\text{STR}}(r) P_{\text{sp}} \qquad [\text{deps} : T2.8, T3.2, T0.6, T0.4]$$
$$R_{\text{STR}}(r) \in SO(d_h),\; P_{\text{sp}} \in SO(d_h) \Rightarrow R_{\text{sp}}(r) \in SO(d_h)$$

$$\boxed{[T3.6 : \Pi_{\text{act}} R_{\text{sp}} \Pi_{\text{act}}]} \quad P_{\text{sp}} \in \mathcal{A} \Rightarrow \Pi_{\text{act}} P_{\text{sp}} = \Pi_{\text{act}} = P_{\text{sp}} \Pi_{\text{act}} \qquad [\text{deps} : T3.5, T2.10, T2.11, T3.4, T0.5, T0.6]$$
$$R_{\text{STR}}(r) \Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r)$$
$$\Rightarrow \Pi_{\text{act}} R_{\text{sp}}(r) \Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r) \Pi_{\text{act}}$$

$$\boxed{[T3.7 : \Pi\text{-rel } R_{\text{sp}}]} \; P_{\text{sp}} \in \mathcal{A} \Rightarrow \Pi_{\text{act}} R_{\text{sp}}(r_i)^\top R_{\text{sp}}(r_j) \Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \qquad [\text{deps} : T3.6, T2.9, T3.5, T3.2,$$

5 Tier 4: Attention Queries, Keys, and Scores

$$\boxed{[T4.1 : W, q, k, v]} \qquad W_Q, W_K, W_V \in \mathbb{R}^{d_h \times D} \qquad [\text{deps} : T0.1, T0.3]$$
$$x_i \in \mathbb{R}^D,\; r_i \in \mathbb{R}^{d_c}$$
$$q_i = W_Q x_i, \quad k_j = W_K x_j, \quad v_j = W_V x_j$$

$$\boxed{[T4.2 : q^{(\text{act})}, k^{(\text{act})}, d_{\text{act}}]} \; q_i^{(\text{act})} \overset{\varnothing}{=} \Pi_{\text{act}} q_i, \quad k_j^{(\text{act})} \overset{\varnothing}{=} \Pi_{\text{act}} k_j \qquad [\text{deps} : T2.10, T4.1]$$
$$d_{\text{act}} \overset{\varnothing}{=} 2m$$

$$\boxed{[T4.3 : \tilde{q}, \tilde{k}]} \quad \begin{aligned} \tilde{q}_i &= \Pi_{\text{act}} R_{\text{sp}}(r_i) q_i^{(\text{act})} \\ \tilde{k}_j &= \Pi_{\text{act}} R_{\text{sp}}(r_j) k_j^{(\text{act})} \end{aligned} \qquad [\text{deps} : T3.5, T4.2, T2.10]$$

$$\boxed{[T4.4 : \alpha_{ij} \text{ def}]} \quad \begin{aligned} \alpha_{ij} &\overset{\varnothing}{=} \frac{\tilde{q}_i^\top \tilde{k}_j}{\sqrt{d_{\text{act}}}} \\ &= \frac{1}{\sqrt{d_{\text{act}}}} (q_i^{(\text{act})})^\top \big(\Pi_{\text{act}} R_{\text{sp}}(r_i) \Pi_{\text{act}}\big)^\top \big(\Pi_{\text{act}} R_{\text{sp}}(r_j) \Pi_{\text{act}}\big) k_j^{(\text{act})} \\ &\qquad (\text{uses } \Pi_{\text{act}}^\top = \Pi_{\text{act}} = \Pi_{\text{act}}^2) \end{aligned} \qquad [\text{deps} : T4.3, T4.2, T2.10, T0.3]$$

$$\boxed{[T4.5 : \alpha_{ij} \text{ rel}]} \; P_{\text{sp}} \in \mathcal{A} \Rightarrow \alpha_{ij} = \frac{1}{\sqrt{2m}} (q_i^{(\text{act})})^\top \Big(\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}}\Big) k_j^{(\text{act})} \qquad [\text{deps} : T4.4, T3.6, T3.7, T4.2, T2$$

6 Tier 5: Stability Estimates

$$\boxed{[T5.1 : \varepsilon \text{ comm}]} \quad \begin{aligned} \varepsilon_{ab} \overset{\varnothing}{=} |[L_a, L_b]|, \quad \varepsilon \overset{\varnothing}{=} \max_{a,b} \varepsilon_{ab} \\ \big|[A(r), A(s)]\big| \le \sum_{a,b} |r_a||s_b| \varepsilon_{ab} \le |r|_1 |s|_1 \, \varepsilon \end{aligned} \qquad [\text{deps} : T1.3, T1.1, T0.10, T0.9]$$

$$\boxed{[T5.2 : \text{BCH err}]} \; \big|\log\big(\exp(A(r)) \exp(A(s))\big) - (A(r) + A(s))\big| \le \tfrac{1}{2} |[A(r), A(s)]| + C |[A(r), A(s)]|^2 \qquad [\text{deps} : T0.1$$

$$\boxed{[T5.3 : R_{\text{STR}} \text{ rel approx}]} \; \big|R_{\text{STR}}(r)^\top R_{\text{STR}}(s) - R_{\text{STR}}(s - r)\big| \le C \, \varepsilon \, |r|_1 |s|_1 + O(\varepsilon^2) \qquad [\text{deps} : T1.6, T1.4, T1.5, T5$$

$$\boxed{[T5.4 : S = S_- + E]} \; S = S_- + E, \quad S_-^\top = -S_-, \quad |E| \le \eta, \quad 0 \le \eta < 1 \qquad [\text{deps} : T0.16, T0.9]$$

$$\boxed{[T5.5 : \sigma_{\min}(I + S_-)]} \quad \begin{aligned} S_-^\top = -S_- \Rightarrow \sigma(S_-) \subset i\mathbb{R} \\ \Rightarrow |1 + i\mu| \ge 1 \text{ for } \lambda = i\mu \in \sigma(S_-) \\ \sigma_{\min}(I + S_-) \ge 1 \end{aligned} \qquad [\text{deps} : T5.4, T0.16, T0.8]$$

$$\boxed{[T5.6 : \sigma_{\min}(I + S)]} \; \begin{aligned} \sigma_{\min}(I + S) \ge \sigma_{\min}(I + S_-) - |E| \ge 1 - \eta > 0 \\ \Rightarrow I + S \text{ invertible} \end{aligned} \qquad [\text{deps} : T5.4, T5.5, T0.19, T0.9]$$

$$\boxed{[T5.7 : \text{Cayley Lip}]} \; \begin{aligned} P_{\text{sp}}(S) \overset{\varnothing}{=} (I - S)(I + S)^{-1}, \quad P_{\text{sp}}(S_-) \overset{\varnothing}{=} (I - S_-)(I + S_-)^{-1} \\ |P_{\text{sp}}(S) - P_{\text{sp}}(S_-)| \le C' \eta \end{aligned} \qquad [\text{deps} : T5.6, T5.4, T0.20, T0$$

$$\boxed{[T5.8:\eta_{\text{mix}}]}\ P_{\text{sp}} = U\begin{bmatrix} A & B \\ C & D \end{bmatrix}U^\top, \quad \eta_{\text{mix}} \overset{\varnothing}{=} |B| + |C| \qquad [\text{deps}: T3.2, T2.3, T2.10, T0.9, T0.6, T0.5, T0.7]$$

$$\begin{aligned} P_{\text{sp}} \in SO(d_h) \Rightarrow \begin{bmatrix} A & B \\ C & D \end{bmatrix}^\top \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= I \\ \Rightarrow |A - I_{2m}| &\le C''\eta_{\text{mix}} \\ \Rightarrow |\Pi_{\text{act}}P_{\text{sp}}\Pi_{\text{act}} - \Pi_{\text{act}}| &\le C''\eta_{\text{mix}} \end{aligned}$$

$$\boxed{[T5.9:\Pi R_{\text{sp}}\Pi\ \text{approx}]}\ \left|\Pi_{\text{act}}R_{\text{sp}}(r)\Pi_{\text{act}} - \Pi_{\text{act}}R_{\text{STR}}(r)\Pi_{\text{act}}\right| \le C''\eta_{\text{mix}} \qquad [\text{deps}: T3.5, T2.11, T5.8, T2.10, T0.9, T$$

$$\boxed{[T5.10:\alpha_{ij}\ \text{stab}]} \qquad \Delta_{ij} \overset{\varnothing}{=} C\big(\varepsilon|r_i|_1|r_j|_1 + \eta_{\text{mix}}\big)|q_i^{(\text{act})}|\,|k_j^{(\text{act})}| \qquad [\text{deps}: T4.4, T5.9, T5.3, T4.2, T5$$

$$\left|\alpha_{ij} - \frac{1}{\sqrt{2m}}(q_i^{(\text{act})})^\top\Big(\Pi_{\text{act}}R_{\text{STR}}(r_j - r_i)\Pi_{\text{act}}\Big)k_j^{(\text{act})}\right| \le \Delta_{ij}$$

7 Tier 6: Conclusions

$$\boxed{[T6.1:\text{GOAL}]}\ \big(T1.1 \wedge T2.3 \wedge P_{\text{sp}} \in \mathcal{A}\big) \Rightarrow \alpha_{ij} = \frac{1}{\sqrt{2m}}(q_i^{(\text{act})})^\top\Big(\Pi_{\text{act}}R_{\text{STR}}(r_j - r_i)\Pi_{\text{act}}\Big)k_j^{(\text{act})} \qquad [\text{deps}: T4.5, T3$$

$$\boxed{[T6.2:\text{GOAL}_{\text{rob}}]}\ \Big(|[L_a, L_b]| \le \varepsilon\ \forall a,b, \quad P_{\text{sp}} \in SO(d_h)\ \text{with mixing}\ \eta_{\text{mix}}\Big) \Rightarrow \alpha_{ij} \approx_{\Delta_{ij}} \frac{1}{\sqrt{2m}}(q_i^{(\text{act})})^\top\Big(\Pi_{\text{act}}R_{\text{STR}}($$

$$\Delta_{ij} \le C\big(\varepsilon|r_i|_1|r_j|_1 + \eta$$

Interpretation

- When $P_{\text{sp}} \in \mathcal{A}$ (no mixing between active and null subspaces), the attention weights α_{ij} depend only on the relative displacement $r_j - r_i$.
- Departures from commuting generators or unmixed post-rotations introduce a bounded deviation captured by Δ_{ij}.
- The construction leverages commuting skew generators to obtain an $SO(d_h)$-valued relative rotation, and Cayley transforms provide structured post-rotations that preserve orthogonality.