

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	2
2	Tier 1: Structured Generators	4
3	Tier 2: Joint Block Structure	5
4	Tier 3: Cayley Post-Rotations	6
5	Tier 4: Attention Queries, Keys, and Scores	6
6	Tier 5: Stability Estimates	7
7	Tier 6: Conclusions	8

1 Tier 0: Linear-Algebra Primitives

$$\boxed{[T0.1 : \mathbb{N}, \text{Idx}] \quad \mathbb{N} \stackrel{\varnothing}{=} \{1, 2, 3, \dots\} \quad [\text{deps} : \varnothing]$$

$$d_h, d_c, D \in \mathbb{N}$$

$$H \stackrel{\varnothing}{=} \{1, \dots, d_h\}, \quad K \stackrel{\varnothing}{=} \{1, \dots, d_c\}, \quad J_D \stackrel{\varnothing}{=} \{1, \dots, D\}$$

$$\boxed{[T0.2 : I_n, \delta_{ij}] \quad \delta_{ij} \stackrel{\varnothing}{=} \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad [\text{deps} : T0.1]$$

$$I_n \stackrel{\varnothing}{=} [\delta_{ij}]_{i,j=1}^n$$

$$\boxed{[T0.3 : \mathbb{R}^{m \times n}, (\cdot)^\top, (\cdot)^*, \text{tr}] \quad \mathbb{R}^{m \times n} \stackrel{\varnothing}{=} \text{set of } m \times n \text{ real matrices} \quad [\text{deps} : T0.1]$$

$$\mathbb{C}^{m \times n} \stackrel{\varnothing}{=} \text{set of } m \times n \text{ complex matrices}$$

$$A^\top \in \mathbb{R}^{n \times m} \quad (A \in \mathbb{R}^{m \times n})$$

$$B^* \stackrel{\varnothing}{=} \overline{B}^\top \quad (B \in \mathbb{C}^{m \times n})$$

$$\text{tr}(M) \stackrel{\varnothing}{=} \sum_{i=1}^n M_{ii} \quad (M \in \mathbb{C}^{n \times n})$$

$$\boxed{[T0.4 : \det] \quad \det : \mathbb{C}^{n \times n} \rightarrow \mathbb{C} \quad [\text{deps} : T0.3]$$

$$\det(AB) = \det(A) \det(B)$$

$$\det(A^\top) = \det(A)$$

$$U \text{ invertible} \Rightarrow \det(UAU^{-1}) = \det(A)$$

(multiplicativity, transpose invariance, similarity invariance)

$$\boxed{[T0.5 : \oplus, \text{blk}, \det \oplus] \quad \bigoplus_{u=1}^m M_u \stackrel{\varnothing}{=} \text{diag}(M_1, \dots, M_m) \quad [\text{d} : \dots]$$

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

(block triangular matrices have determinant equal to product of block determinants)

$$\boxed{[T0.6 : O, SO, U] \quad O(n) \stackrel{\varnothing}{=} \{Q \in \mathbb{R}^{n \times n} \mid Q^\top Q = I_n\} \quad [\text{deps} : T0.2, T0.3, T0.4]$$

$$SO(n) \stackrel{\varnothing}{=} \{Q \in O(n) \mid \det Q = 1\}$$

$$U(n) \stackrel{\varnothing}{=} \{W \in \mathbb{C}^{n \times n} \mid W^* W = I_n\}$$

$$Q \in O(n) \Rightarrow Q^{-1} = Q^\top$$

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

$$\Rightarrow \det Q \in \{\pm 1\}$$

$$\begin{aligned}
& \boxed{[T0.8 : \sigma, \sigma_{\min}, \text{normal}]} \quad \sigma(M) \stackrel{\mathcal{O}}{=} \{\lambda \in \mathbb{C} \mid \exists x \neq 0 : Mx = \lambda x\} \quad [\text{deps} : T0.3] \\
& \quad M^*M = MM^* \Rightarrow M \text{ normal} \\
& \quad \text{Singular values } s_i(M) \stackrel{\mathcal{O}}{=} \sqrt{\sigma_i(M^*M)} \\
& \quad \sigma_{\min}(M) \stackrel{\mathcal{O}}{=} \min_i s_i(M) \\
& \quad M \text{ normal} \Rightarrow s_i(M) = |\lambda_i| \text{ for } \lambda_i \in \sigma(M) \\
& \quad (\text{unitary diagonalisation of normal matrices})
\end{aligned}$$

$$\begin{aligned}
& \boxed{[T0.9 : |\cdot|, |\cdot|_1]} \quad |x| \stackrel{\mathcal{O}}{=} \left(\sum_u x_u^2 \right)^{1/2} \quad (x \in \mathbb{R}^m) \quad [\text{deps} : T0.8] \\
& \quad |x|_1 \stackrel{\mathcal{O}}{=} \sum_u |x_u| \\
& \quad |M| \stackrel{\mathcal{O}}{=} \sup_{v \neq 0} \frac{|Mv|}{|v|} \quad (M \in \mathbb{R}^{m \times n}) \\
& \quad |AB| \leq |A| |B|, \quad |A + B| \leq |A| + |B| \\
& \quad |a| \stackrel{\mathcal{O}}{=} |a| \quad (a \in \mathbb{R}) \\
& \quad (\text{overloaded notation chooses scalar, vector, or operator norm from context})
\end{aligned}$$

$$\boxed{[T0.10 : [A, B]]} \quad [A, B] \stackrel{\mathcal{O}}{=} AB - BA \quad [\text{deps} : T0.3]$$

$$\begin{aligned}
& \boxed{[T0.11 : \exp, \log, \exp \text{ props}]} \quad \exp(M) \stackrel{\mathcal{O}}{=} \sum_{t \geq 0} \frac{1}{t!} M^t \quad [\text{deps} : T0.3, T0.4, T0.5, T] \\
& \quad \log(X) \text{ locally defined as } \exp^{-1}(X) \text{ near } I_n \\
& \quad [A, B] = 0 \Rightarrow \exp(A + B) = \exp(A) \exp(B) \\
& \quad (\exp M)^\top = \exp(M^\top) \\
& \quad U \text{ invertible} \Rightarrow \exp(UMU^{-1}) = U \exp(M) U^{-1} \\
& \quad \det(\exp M) = \exp(\text{tr } M) \\
& \quad \exp\left(\bigoplus_{u=1}^m M_u\right) = \bigoplus_{u=1}^m \exp(M_u) \\
& \quad (\text{all properties follow from the power-series definition})
\end{aligned}$$

$$\begin{aligned}
& \boxed{[T0.12 : \mathcal{C}, O(\cdot), \approx_\varepsilon]} \quad \mathcal{C} \stackrel{\mathcal{O}}{=} \{C, C', C'', \dots \mid C \geq 0 \text{ finite constants}\} \quad [\text{deps} : T0.9] \\
& \quad f = O(g) \Leftrightarrow \exists C \in \mathcal{C} : f \leq Cg \\
& \quad X \approx_\varepsilon Y \Leftrightarrow |X - Y| \leq \varepsilon
\end{aligned}$$

$$\begin{aligned}
& \boxed{[T0.13 : \text{BCH}]} \quad \exp(M) \exp(N) = \exp\left(M + N + \frac{1}{2}[M, N] + R(M, N)\right) \quad [\text{deps} : T0.11, T0.10, T0.12, T0.9] \\
& \quad |R(M, N)| = O(|[M, N]|^2) \\
& \quad (\text{Baker–Campbell–Hausdorff series (local version)})
\end{aligned}$$

$$\begin{aligned}
\boxed{[T0.14 : \ker, \text{im}, \text{rank}]} \quad & \ker(M) \stackrel{\varnothing}{=} \{x \mid Mx = 0\} \quad [\text{deps} : T0.3, T0.1] \\
& \text{im}(M) \stackrel{\varnothing}{=} \{Mx \mid x\} \\
& \text{rank}(M) \stackrel{\varnothing}{=} \dim(\text{im}(M))
\end{aligned}$$

$$\begin{aligned}
\boxed{[T0.15 : \text{proj}]} \quad & P^2 = P \text{ and } P^\top = P \Rightarrow P \text{ orthogonal projector} \quad [\text{deps} : T0.14, T0.3] \\
& P, Q \text{ orthogonal projectors and } PQ = 0 \Rightarrow \text{im}(P) \perp \text{im}(Q) \\
& (\text{characterisation via images})
\end{aligned}$$

$$\begin{aligned}
\boxed{[T0.16 : \text{skew}_{\text{spec}}]} \quad & S^\top = -S \Rightarrow S \text{ normal} \quad [\text{deps} : T0.3, T0.8] \\
& \sigma(S) \subset i\mathbb{R} \\
& \lambda = i\mu \in \sigma(S) \Rightarrow -\lambda = -i\mu \in \sigma(S) \\
& (\text{skew-symmetric spectra occur in conjugate-sign pairs})
\end{aligned}$$

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

$$\begin{aligned}
\boxed{[T0.18 : R_2 \in SO(2)]} \quad & R_2(\theta)^\top R_2(\theta) = I_2 \quad [\text{deps} : T0.17, T0.16, T0.11, \\
& \det R_2(\theta) = 1 \\
& \Rightarrow R_2(\theta) \in SO(2) \\
& (\text{uses } J^\top = -J \text{ and properties of exp and the determinant})
\end{aligned}$$

$$\boxed{[T0.19 : \text{Weyl}]} \quad |\sigma_{\min}(A + E) - \sigma_{\min}(A)| \leq |E| \quad [\text{deps} : T0.8, T0.9]$$

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

2 Tier 1: Structured Generators

$$\begin{aligned}
\boxed{[T1.1 : L_k]} \quad & \forall k \in K : L_k \in \mathbb{R}^{d_h \times d_h}, \quad L_k^\top = -L_k, \quad [L_a, L_b] = 0 \quad [\text{deps} : T0.1, T0.3, T0.10, T0.16] \\
& (\text{model hypothesis: commuting skew-symmetric generators})
\end{aligned}$$

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

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

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

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

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

3 Tier 2: Joint Block Structure

$$\begin{aligned}
& \boxed{[T2.1 : \text{sim-unitary}]} \quad \{L_k\}_{k \in K} \text{ commute and are normal} \quad [\text{deps} : T1.2, T1.1, T0.6, T0.16] \\
& \Rightarrow \exists \mathcal{U}_c \in U(d_h) : \mathcal{U}_c^* L_k \mathcal{U}_c = \text{diag}(i\lambda_{k,1}, \dots, i\lambda_{k,d_h}) \\
& \quad \lambda_{k,j} \in \mathbb{R}
\end{aligned}$$

$$\begin{aligned}
& \boxed{[T2.2 : \text{real 2D planes}]} \quad \text{Fix } j \in H : \mathcal{U}_c^* L_k \mathcal{U}_c e_j = i\lambda_{k,j} e_j \quad [\text{deps} : T2.1, T0.17, T0.16, T0.3] \\
& \quad \Rightarrow L_k(\Re e_j) = \lambda_{k,j} J(\Re e_j, \Im e_j) \\
& \quad \quad L_k(\Im e_j) = \lambda_{k,j} J(\Re e_j, \Im e_j) \\
& \quad \Rightarrow \text{span}\{\Re e_j, \Im e_j\} \text{ is a real, 2-D, } L_k\text{-invariant plane}
\end{aligned}$$

$$\begin{aligned}
& \boxed{[T2.3 : \text{JointBlockDiag}]} \quad \exists m, d_{\text{null}} \in \mathbb{N} : 2m + d_{\text{null}} = d_h \quad [\text{deps} : T2.2, T0.6, T0.16, T0.5] \\
& \quad \exists U \in O(d_h) : U^\top L_k U = \left(\bigoplus_{u=1}^m \lambda_{k,u} J \right) \oplus 0_{d_{\text{null}} \times d_{\text{null}}}
\end{aligned}$$

$$\begin{aligned}
& \boxed{[T2.4 : \beta, \theta]} \quad \beta_u \stackrel{\mathcal{O}}{=} (\lambda_{1,u}, \dots, \lambda_{d_c,u})^\top \in \mathbb{R}^{d_c} \quad [\text{deps} : T2.3, T1.1, T0.1] \\
& \quad \theta_u(r) \stackrel{\mathcal{O}}{=} \beta_u^\top r \in \mathbb{R}
\end{aligned}$$

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

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

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

$$\begin{aligned}
& \boxed{[T2.8 : R_{\text{STR}} \in SO]} \quad R_{\text{STR}}(r) \in SO(d_h) \quad [\text{deps} : T2.7, T] \\
& \quad (\text{uses } R_2(\theta) \in SO(2), \text{ block determinant multiplicativity, and } U \in O(d_h))
\end{aligned}$$

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

$$\begin{aligned}
& \boxed{[T2.10 : \Pi_{\text{act}}, \Pi_{\text{null}}, d_{\text{act}}]} \quad \Pi_{\text{act}} \stackrel{\mathcal{O}}{=} U \text{diag}(I_{2m}, 0) U^\top \quad [\text{deps} : T2.3, T0.15, T0.2, T0.14, T0.6] \\
& \quad \Pi_{\text{null}} \stackrel{\mathcal{O}}{=} U \text{diag}(0, I_{d_{\text{null}}}) U^\top \\
& \quad \Pi_{\text{act}}^\top = \Pi_{\text{act}}, \quad \Pi_{\text{act}}^2 = \Pi_{\text{act}} \\
& \quad \Pi_{\text{null}}^\top = \Pi_{\text{null}}, \quad \Pi_{\text{null}}^2 = \Pi_{\text{null}} \\
& \quad \Pi_{\text{act}} \Pi_{\text{null}} = 0 \Rightarrow \text{im}(\Pi_{\text{act}}) \perp \text{im}(\Pi_{\text{null}}) \\
& \quad d_{\text{act}} \stackrel{\mathcal{O}}{=} \text{rank}(\Pi_{\text{act}}) = 2m
\end{aligned}$$

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

4 Tier 3: Cayley Post-Rotations

$$\boxed{[T3.1 : \text{Cayley}]} \quad S \in \mathbb{R}^{d_h \times d_h}, \quad S^\top = -S \Rightarrow (I + S) \text{ invertible} \quad [\text{deps} : T0.16, T0.3, T0.2]$$

$$\begin{aligned} P_{\text{sp}} &\stackrel{\varnothing}{=} (I - S)(I + S)^{-1} \\ (I - S)(I + S) &= (I + S)(I - S) = I - S^2 \\ &\text{(uses } \sigma(S) \subset i\mathbb{R} \text{ so } 1 + i\mu \neq 0) \end{aligned}$$

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

$$\begin{aligned} \boxed{[T3.3 : \text{Cayley surj}]} \quad & Q \in SO(d_h), \quad -1 \notin \sigma(Q) \Rightarrow S_Q \stackrel{\varnothing}{=} (I - Q)(I + Q)^{-1} \quad [\text{deps} : T0.16, T0.20, T0.7] \\ & S_Q^\top = -S_Q, \quad (I - S_Q)(I + S_Q)^{-1} = Q \\ & \text{(Cayley transform bijection on } SO(d_h) \text{ with no } -1 \text{ eigenvalues)} \end{aligned}$$

$$\begin{aligned} \boxed{[T3.4 : \mathcal{A}]} \quad & \mathcal{A} \stackrel{\varnothing}{=} \left\{ U \text{diag}(I_{2m}, R_{\text{null}}) U^\top \mid R_{\text{null}} \in SO(d_{\text{null}}) \right\} \subseteq SO(d_h) \quad [\text{deps} : T2.3, T0.6, T0.7, T0.5, T0.4, T0.3] \\ & R_{\text{null}} \in SO(d_{\text{null}}), \quad -1 \notin \sigma(R_{\text{null}}) \Rightarrow U \text{diag}(I_{2m}, R_{\text{null}}) U^\top \in \mathcal{B} \end{aligned}$$

$$\begin{aligned} \boxed{[T3.5 : R_{\text{sp}}]} \quad & R_{\text{sp}}(r) \stackrel{\varnothing}{=} R_{\text{STR}}(r) P_{\text{sp}} \quad [\text{deps} : T2.8, T3.2, T0.6, T0.4] \\ & R_{\text{STR}}(r) \in SO(d_h), \quad P_{\text{sp}} \in SO(d_h) \Rightarrow R_{\text{sp}}(r) \in SO(d_h) \end{aligned}$$

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

$$\boxed{[T3.7 : \Pi\text{-rel } R_{\text{sp}}]} \quad P_{\text{sp}} \in \mathcal{A} \Rightarrow \Pi_{\text{act}} R_{\text{sp}}(r_i)^\top R_{\text{sp}}(r_j) \Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \quad [\text{deps} : T3.6, T2.9, T3.5, T3.2, T0.6]$$

5 Tier 4: Attention Queries, Keys, and Scores

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

$$\begin{aligned} \boxed{[T4.2 : q^{(\text{act})}, k^{(\text{act})}, d_{\text{act}}]} \quad & q_i^{(\text{act})} \stackrel{\varnothing}{=} \Pi_{\text{act}} q_i, \quad k_j^{(\text{act})} \stackrel{\varnothing}{=} \Pi_{\text{act}} k_j \quad [\text{deps} : T2.10, T4.1] \\ & d_{\text{act}} \stackrel{\varnothing}{=} 2m \end{aligned}$$

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

$$\boxed{[T4.4 : \alpha_{ij} \text{ def}]} \quad \alpha_{ij} \stackrel{\varnothing}{=} \frac{\tilde{q}_i^\top \tilde{k}_j}{\sqrt{d_{\text{act}}}} \quad [\text{deps} : T4.3, T4.2, T2.10, T0.3]$$

$$= \frac{1}{\sqrt{d_{\text{act}}}} (q_i^{(\text{act})})^\top (\Pi_{\text{act}} R_{\text{sp}}(r_i) \Pi_{\text{act}})^\top (\Pi_{\text{act}} R_{\text{sp}}(r_j) \Pi_{\text{act}}) k_j^{(\text{act})}$$

$$(\text{uses } \Pi_{\text{act}}^\top = \Pi_{\text{act}} = \Pi_{\text{act}}^2)$$

$$\boxed{[T4.5 : \alpha_{ij} \text{ rel}]} \quad P_{\text{sp}} \in \mathcal{A} \Rightarrow \alpha_{ij} = \frac{1}{\sqrt{2m}} (q_i^{(\text{act})})^\top (\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}}) k_j^{(\text{act})} \quad [\text{deps} : T4.4, T3.6, T3.7, T4.2, T0.3]$$

6 Tier 5: Stability Estimates

$$\boxed{[T5.1 : \varepsilon \text{ comm}]} \quad \varepsilon_{ab} \stackrel{\varnothing}{=} |[L_a, L_b]|, \quad \varepsilon \stackrel{\varnothing}{=} \max_{a,b} \varepsilon_{ab} \quad [\text{deps} : T1.3, T1.1, T0.10, T0.9]$$

$$|[A(r), A(s)]| \leq \sum_{a,b} |r_a| |s_b| \varepsilon_{ab} \leq |r|_1 |s|_1 \varepsilon$$

$$\boxed{[T5.2 : \text{BCH err}]} \quad |\log(\exp(A(r)) \exp(A(s))) - (A(r) + A(s))| \leq \frac{1}{2} |[A(r), A(s)]| + C |[A(r), A(s)]|^2 \quad [\text{deps} : T0.1]$$

$$\boxed{[T5.3 : R_{\text{STR}} \text{ rel approx}]} \quad |R_{\text{STR}}(r)^\top R_{\text{STR}}(s) - R_{\text{STR}}(s - r)| \leq C \varepsilon |r|_1 |s|_1 + O(\varepsilon^2) \quad [\text{deps} : T1.6, T1.4, T1.5, T0.9]$$

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

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

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

$$\boxed{[T5.7 : \text{Cayley Lip}]} \quad \begin{aligned} P_{\text{sp}}(S) &\stackrel{\varnothing}{=} (I - S)(I + S)^{-1}, \quad P_{\text{sp}}(S_-) \stackrel{\varnothing}{=} (I - S_-)(I + S_-)^{-1} \\ |P_{\text{sp}}(S) - P_{\text{sp}}(S_-)| &\leq C' \eta \end{aligned} \quad [\text{deps} : T5.6, T5.4, T0.20, T0.9]$$

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

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

$$\boxed{[T5.9 : \Pi R_{\text{sp}} \Pi \text{ approx}]} \quad |\Pi_{\text{act}} R_{\text{sp}}(r) \Pi_{\text{act}} - \Pi_{\text{act}} R_{\text{STR}}(r) \Pi_{\text{act}}| \leq C'' \eta_{\text{mix}} \quad [\text{deps} : T3.5, T2.11, T5.8, T2.10, T0.9, T0.6, T0.5, T0.7]$$

$$\begin{aligned} \boxed{[T5.10 : \alpha_{ij} \text{ stab}]} \quad &\Delta_{ij} \stackrel{\text{def}}{=} C(\varepsilon |r_i|_1 |r_j|_1 + \eta_{\text{mix}}) |q_i^{(\text{act})}| |k_j^{(\text{act})}| \quad [\text{deps} : T4.4, T5.9, T5.3, T4.2, T5.8, T5.7, T5.6, T5.5, T5.4, T5.3, T5.2, T5.1, T5.0] \\ &\left| \alpha_{ij} - \frac{1}{\sqrt{2m}} (q_i^{(\text{act})})^\top \left(\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \right) k_j^{(\text{act})} \right| \leq \Delta_{ij} \end{aligned}$$

7 Tier 6: Conclusions

$$\boxed{[T6.1 : \text{GOAL}]} \quad (T1.1 \wedge T2.3 \wedge P_{\text{sp}} \in \mathcal{A}) \Rightarrow \alpha_{ij} = \frac{1}{\sqrt{2m}} (q_i^{(\text{act})})^\top \left(\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \right) k_j^{(\text{act})} \quad [\text{deps} : T4.5, T5.8, T5.7, T5.6, T5.5, T5.4, T5.3, T5.2, T5.1, T5.0]$$

$$\begin{aligned} \boxed{[T6.2 : \text{GOAL}_{\text{rob}}]} \quad &\left(|[L_a, L_b]| \leq \varepsilon \forall a, b, \quad P_{\text{sp}} \in SO(d_h) \text{ with mixing } \eta_{\text{mix}} \right) \Rightarrow \alpha_{ij} \approx_{\Delta_{ij}} \frac{1}{\sqrt{2m}} (q_i^{(\text{act})})^\top \left(\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \right) k_j^{(\text{act})} \\ &\Delta_{ij} \leq C(\varepsilon |r_i|_1 |r_j|_1 + \eta_{\text{mix}}) \end{aligned}$$

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.