

# Equivariant Structured Positional Rotations

## Integrated Writeup

Throughout we work inside the universe  $\{\mathbb{R}, \mathbb{C}, \mathbb{N}, \text{matrices}, \text{vectors}\}$  and adhere to the conventions stated below. Every claim is tied to an explicit boxed statement. Cleanups `cleanup` add the missing boxes [2.4], [2.7], [2.9], [2.16], [3.6a], [5.4a], [5.11] and expand the algebra in [2.1], [2.2], [3.11]. None of these touch the hypotheses [1.2], [1.3] or constructions [2.12], [3.7], [4.1], so the main theorem (statement [5.9]) remains unchanged.

## Global Conventions

All matrices are real unless noted otherwise. Vectors are real column vectors. Transpose is  $^\top$ ; conjugate transpose is  $^*$ . We write  $[A, B] := AB - BA$ , use direct sums  $\oplus$ , and denote block-diagonal concatenation over  $u \in \{1, \dots, m\}$  by  $\bigoplus_{u=1}^m(\cdot)$ . The spectrum of  $M$  is  $\sigma(M)$ , the number of structural nonzeros is  $\text{nnz}(M)$ , and the dimension of a subspace  $V$  is  $\dim(V)$ . Orthogonal and special orthogonal groups are  $O(n)$  and  $SO(n)$ .

## 1 BEG-0: Preliminaries

The base objects `base` and identities `facts0` now follow the bookkeeping format requested by the user.

$$[0.1] \quad \mathbb{N} \stackrel{\varnothing}{=} \{1, 2, 3, \dots\} \quad () \quad (1)$$

$$[0.2] \quad d_h, d_c, D \in \underbrace{\mathbb{N}}_{[0.1]} \quad ([0.1]) \quad (2)$$

$$[0.3] \quad I_n \stackrel{\varnothing}{=} \{A \in \mathbb{R}^{n \times n} : A_{ij} = \delta_{ij}\} \quad () \quad (3)$$

$$[0.4] \quad (\cdot)^\top : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{m \times n}, \quad (\cdot)^* : \mathbb{C}^{n \times m} \rightarrow \mathbb{C}^{m \times n}, \quad (4)$$

$$A^\top \stackrel{\varnothing}{=} \text{transpose}(A), \quad B^* \stackrel{\varnothing}{=} \overline{B}^\top \quad () \quad (5)$$

$$[0.5] \quad [A, B] \stackrel{\varnothing}{=} AB - BA \quad () \quad (6)$$

$$[0.6] \quad \mathbb{R}^n \stackrel{\varnothing}{=} \{x \in \mathbb{R}^{n \times 1}\}, \quad \mathbb{C}^n \stackrel{\varnothing}{=} \{x \in \mathbb{C}^{n \times 1}\} \quad ([0.1]) \quad (7)$$

$$[0.7] \quad O(n) \stackrel{\varnothing}{=} \left\{ Q \in \mathbb{R}^{n \times n} \mid \underbrace{Q^\top Q}_{[0.4]} \stackrel{[0.3]}{=} \underbrace{I_n}_{[0.3]} \right\} \quad ([0.3], [0.4]) \quad (8)$$

$$[0.8] \quad SO(n) \stackrel{\varnothing}{=} \left\{ Q \in O(n) \mid \det(Q) = +1 \right\} \quad ([0.7]) \quad (9)$$

$$[0.9] \quad \exp(M) \stackrel{\varnothing}{=} \sum_{t=0}^{\infty} \frac{1}{t!} M^t, \quad M \in \mathbb{R}^{n \times n} \quad ([0.1]) \quad (10)$$

$$[0.10] \quad [M, N] = 0 \implies \exp(M + N) \stackrel{[0.9]}{=} \exp(M) \exp(N) \quad ([0.5], [0.9]) \quad (11)$$

$$\text{If } M = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}, \quad \exp(M) \stackrel{[0.9]}{=} \begin{bmatrix} \exp(M_1) & 0 \\ 0 & \exp(M_2) \end{bmatrix} \quad ([0.9]) \quad (12)$$

$$[0.11] \quad J \stackrel{\varnothing}{=} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad R_2(\theta) \stackrel{\varnothing}{=} \exp(\theta J) \stackrel{[0.9]}{=} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \in SO(2) \quad ([0.8], [0.9]) \quad (13)$$

$$[0.12] \quad R_2(a)^\top R_2(b) \stackrel{[0.11]}{=} R_2(-a) R_2(b) \stackrel{[0.10]}{=} R_2(b - a) \quad ([0.10], [0.11]) \quad (14)$$

$$[0.13] \quad S^\top = -S \implies S_{ii} = 0, \quad S_{ji} = -S_{ij} \quad ([0.4]) \quad (15)$$

$$[0.14] \quad \det \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix} = \det(M_1) \det(M_2) \quad () \quad (16)$$

$$[0.15] \quad U(n) \stackrel{\varnothing}{=} \left\{ \mathcal{U} \in \mathbb{C}^{n \times n} : \mathcal{U}^* \mathcal{U} \stackrel{[0.4]}{=} I_n \right\} \quad ([0.3], [0.4]) \quad (17)$$

$$[0.16] \quad \sigma(M) \stackrel{\varnothing}{=} \{\lambda \in \mathbb{C} : \exists x \in \mathbb{C}^n \setminus \{0\}, \quad Mx = \lambda x\} \quad ([0.6]) \quad (18)$$

We also use  $\ker(M) := \{x \in \mathbb{R}^n : Mx = 0\}$ ,  $\text{im}(M) := \{Mx : x \in \mathbb{R}^n\}$ ,  $\dim(V)$  for the dimension of a subspace,  $\text{diag}(\lambda_1, \dots, \lambda_n) := [\delta_{ij} \lambda_i]_{i,j=1}^n$ , and  $\bigoplus_{u=1}^m M_u := \text{diag}(M_1, \dots, M_m)$ .

## 2 BEG-1: Hypotheses and Structured Generator

We fix the index sets  $K := \{1, \dots, d_c\}$  and  $H := \{1, \dots, d_h\}$  together with generators  $\{L_k\}_{k \in K}$  and the coordinate vector  $r \in \mathbb{R}^{d_c}$ .

$$\boxed{[1.2]} \quad \forall k \in K : L_k \in \mathbb{R}^{d_h \times d_h}, L_k^\top \stackrel{\varnothing}{=} -L_k \quad ([0.2], [0.4]) \quad (19)$$

$$\boxed{[1.3]} \quad \forall a, b \in K : [L_a, L_b] \stackrel{[0.5]}{=} L_a L_b - L_b L_a \stackrel{\varnothing}{=} 0_{d_h \times d_h} \quad ([0.5], [0.2]) \quad (20)$$

$$\boxed{[1.4]} \quad \forall r \in \mathbb{R}^{d_c} : A(r) \stackrel{\varnothing}{=} \sum_{k=1}^{d_c} \underbrace{L_k}_{[1.2]} \underbrace{[r]_k}_{[0.6]} \quad ([1.2], [0.6], [0.2]) \quad (21)$$

$$\boxed{[1.5]} \quad R_{\text{STR}}(r) \stackrel{\varnothing}{=} \exp \left( \underbrace{A(r)}_{[1.4]} \right) \quad ([1.4], [0.9]) \quad (22)$$

These hypotheses ensure  $A(r)$  is skew-symmetric and that the  $L_k$  commute, paving the way for the simultaneous block-rotation structure in BEG-2.

### 3 BEG-2: Simultaneous Block Rotation Structure

#### 3.1 Normality and Spectrum

$$\boxed{[2.1]} \quad L^\top = -L \implies L^\top L = (-L)L = -L^2, LL^\top = L(-L) = -L^2 \implies L \text{ is normal} \quad ([1.2], [0.4]) \quad (23)$$

$$\boxed{[2.2]} \quad L^\top = -L, Lx = \lambda x, x \neq 0 \implies \lambda \in i\mathbb{R} \quad ([0.4]) \quad (24)$$

#### 3.2 Simultaneous Diagonalisation

$$\boxed{[2.3]} \quad \{L_k\}_{k \in K} \text{ normal and commuting} \implies \exists \mathcal{U} \in U(d_h) : \mathcal{U}^* L_k \mathcal{U} = \text{diag}(i\lambda_{k,1}, \dots, i\lambda_{k,d_h}) \quad ([1.2], [1.3], [2.1], [0.]) \quad (25)$$

#### 3.3 Real Block Form and Projectors

$$\boxed{[2.4]} \quad \exists U \in O(d_h), m, d_{\text{null}} \in \mathbb{N} : 2m + d_{\text{null}} = d_h, \quad (26)$$

$$U^\top L_k U = \left( \bigoplus_{u=1}^m B_{k,u} \right) \oplus 0_{d_{\text{null}} \times d_{\text{null}}}, B_{k,u}^\top = -B_{k,u} \quad ([1.2], [2.3], [0.7], [0.16]) \quad (27)$$

$$\boxed{[2.5]} \quad \forall u : B_{k,u} = \lambda_{k,u} J_u, J_u = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad ([2.4], [0.11]) \quad (28)$$

$$\boxed{[2.6]} \quad \Pi_{\text{act}} \stackrel{\varnothing}{=} U \begin{bmatrix} I_{2m} & 0 \\ 0 & 0 \end{bmatrix} U^\top, \Pi_{\text{null}} \stackrel{\varnothing}{=} U \begin{bmatrix} 0 & 0 \\ 0 & I_{d_{\text{null}}} \end{bmatrix} U^\top, \quad (29)$$

$$\Pi_{\text{act}}^2 = \Pi_{\text{act}}, \Pi_{\text{null}}^2 = \Pi_{\text{null}}, \Pi_{\text{act}} \Pi_{\text{null}} = 0 \quad ([2.4], [0.3], [0.7]) \quad (30)$$

$$\boxed{[2.7]} \quad \beta_u \stackrel{\varnothing}{=} (\lambda_{1,u}, \dots, \lambda_{d_c,u})^\top, [\beta_u]_k = \lambda_{k,u} \quad ([2.5]) \quad (31)$$

$$\boxed{[2.8]} \quad \theta_u(r) \stackrel{\varnothing}{=} \beta_u^\top r = \sum_{k=1}^{d_c} \underbrace{\lambda_{k,u}}_{[2.5]} \underbrace{[r]_k}_{[1.4]} \quad ([2.7], [1.4]) \quad (32)$$

$$\boxed{[2.9]} \quad U^\top A(r)U = \sum_{k=1}^{d_c} \underbrace{U^\top L_k U}_{[2.4]} [r]_k \quad (33)$$

$$= \left( \bigoplus_{u=1}^m \theta_u(r) J_u \right) \oplus 0_{d_{\text{null}} \times d_{\text{null}}} \quad ([1.4], [2.4], [2.5], [2.8]) \quad (34)$$

$$\boxed{[2.10]} \quad \exp(U^\top A(r)U) = \left( \bigoplus_{u=1}^m R_2(\theta_u(r)) \right) \oplus I_{d_{\text{null}}} \quad ([2.9], [0.10], [0.11]) \quad (35)$$

$$\boxed{[2.11]} \quad R_{\text{STR}}(r) = U \left[ \left( \bigoplus_{u=1}^m R_2(\theta_u(r)) \right) \oplus I_{d_{\text{null}}} \right] U^\top \quad ([1.5], [2.9], [2.10], [0.10], [0.7]) \quad (36)$$

$$\boxed{[2.12]} \quad R_{\text{STR}}(r)^\top R_{\text{STR}}(r) = I_{d_h}, \quad \det(R_{\text{STR}}(r)) = 1 \Rightarrow R_{\text{STR}}(r) \in SO(d_h) \quad ([2.11], [0.7], [0.8], [0.14]) \quad (37)$$

$$\boxed{[2.13]} \quad R_{\text{STR}}(r_i)^\top R_{\text{STR}}(r_j) = R_{\text{STR}}(r_j - r_i) \quad ([2.11], [2.12], [0.12], [2.8]) \quad (38)$$

## 4 BEG-3: Cayley Post-Rotation

We parameterise  $P_{\text{sp}}$  via an upper-triangular mask and Cayley transform.

$$\boxed{[3.1]} \quad U_{\text{mask}} \in \{0, 1\}^{d_h \times d_h}, \quad [U_{\text{mask}}]_{ij} = 0 \text{ for } i \geq j \quad ([0.2]) \quad (39)$$

$$\boxed{[3.2]} \quad \tilde{S} \in \mathbb{R}^{d_h \times d_h} \quad ([0.2]) \quad (40)$$

$$\boxed{[3.3]} \quad S_{ij} \stackrel{\emptyset}{=} [U_{\text{mask}}]_{ij} \tilde{S}_{ij} \quad (i < j), \quad S_{ji} \stackrel{\emptyset}{=} -S_{ij}, \quad S_{ii} \stackrel{\emptyset}{=} 0 \quad ([3.1], [3.2], [0.13]) \quad (41)$$

$$\boxed{[3.4]} \quad S^\top = -S \quad ([3.3], [0.4]) \quad (42)$$

$$\boxed{[3.5]} \quad \text{nnz}(S) \leq 2 \text{nnz}(U_{\text{mask}}) \quad ([3.1], [3.3]) \quad (43)$$

$$\boxed{[3.6]} \quad S^\top = -S \implies \sigma(S) \subseteq \{0\} \cup i\mathbb{R} \quad ([3.4], [0.16]) \quad (44)$$

$$\boxed{[3.6a]} \quad -1 \notin \sigma(S) \implies (I_{d_h} + S)^{-1} \text{ exists} \quad ([3.6], [0.3]) \quad (45)$$

$$\boxed{[3.7]} \quad P_{\text{sp}} \stackrel{\emptyset}{=} (I_{d_h} - S)(I_{d_h} + S)^{-1} \quad ([3.6], [0.3]) \quad (46)$$

$$\boxed{[3.8]} \quad P_{\text{sp}}^\top P_{\text{sp}} = I_{d_h} \quad ([3.4], [3.7], [0.4], [0.3]) \quad (47)$$

$$\boxed{[3.9]} \quad Sw = i\mu w \Rightarrow P_{\text{sp}} w = \frac{1 - i\mu}{1 + i\mu} w, \quad \left| \frac{1 - i\mu}{1 + i\mu} \right| = 1 \quad ([3.4], [3.7], [0.16]) \quad (48)$$

$$\boxed{[3.10]} \quad P_{\text{sp}} \in SO(d_h) \iff P_{\text{sp}}^\top P_{\text{sp}} = I_{d_h}, \quad \det(P_{\text{sp}}) = 1 \quad ([3.8], [3.9], [0.8]) \quad (49)$$

Let  $\mathbb{R}^{d_h} = \text{act} \oplus \text{null}$  be the  $U$ -aligned decomposition.

$$\boxed{[3.11]} \quad \dim(\text{act}) = 2m, \dim(\text{null}) = d_{\text{null}} = d_h - 2m \quad ([2.4], [2.6], [0.16]) \quad (50)$$

$$\boxed{[3.12]} \quad U_{\text{mask}} \text{ supported in } \text{null} \times \text{null} \quad (51)$$

$$\implies S = \begin{bmatrix} 0 & 0 \\ 0 & S_{\text{null}} \end{bmatrix}, P_{\text{sp}} = \begin{bmatrix} I_{2m} & 0 \\ 0 & R_{\text{null}} \end{bmatrix}, R_{\text{null}} \in SO(d_{\text{null}}) \quad ([3.1], [3.3], [3.7], [3.10], [2.4]) \quad (52)$$

$$\boxed{[3.13]} \quad \mathcal{A} \stackrel{\varnothing}{=} \left\{ P \in SO(d_h) \mid P = \begin{bmatrix} I_{2m} & 0 \\ 0 & R_{\text{null}} \end{bmatrix}, R_{\text{null}} \in SO(d_{\text{null}}) \right\} \quad ([0.8], [3.11]) \quad (53)$$

$$\boxed{[3.14]} \quad \mathcal{B} \stackrel{\varnothing}{=} \left\{ P_{\text{sp}} : P_{\text{sp}} = (I_{d_h} - S)(I_{d_h} + S)^{-1}, S^{\top} = -S \right\} \quad ([3.7]) \quad (54)$$

$$\boxed{[3.15]} \quad \mathcal{A} \subseteq \mathcal{B} \subseteq SO(d_h) \quad ([3.10], [3.12], [3.13]) \quad (55)$$

## 5 BEG-4: Composed Rotation and Relative/Absolute Behaviour

$$\boxed{[4.1]} \quad R_{\text{sp}}(r) \stackrel{\varnothing}{=} \underbrace{R_{\text{STR}}(r)}_{[2.11]} \underbrace{P_{\text{sp}}}_{[3.7]} \quad ([2.11], [3.7]) \quad (56)$$

$$\boxed{[4.2]} \quad R_{\text{sp}}(r)^{\top} R_{\text{sp}}(r) = I_{d_h}, \det(R_{\text{sp}}(r)) = 1 \implies R_{\text{sp}}(r) \in SO(d_h) \quad ([4.1], [2.12], [3.10], [0.8]) \quad (57)$$

When  $P_{\text{sp}} \in \mathcal{A}$  the active block is untouched.

$$\boxed{[4.3]} \quad P_{\text{sp}} \in \mathcal{A} \implies \Pi_{\text{act}} R_{\text{sp}}(r) \Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r) \Pi_{\text{act}} \quad ([4.1], [3.13], [2.6]) \quad (58)$$

$$\boxed{[4.4]} \quad P_{\text{sp}} \in \mathcal{A} \implies \Pi_{\text{act}} R_{\text{sp}}(r_i)^{\top} R_{\text{sp}}(r_j) \Pi_{\text{act}} = \Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \quad ([4.1], [4.3], [2.13], [3.13], [2.6]) \quad (59)$$

If  $P_{\text{sp}} \in \mathcal{B} \setminus \mathcal{A}$  we can write  $P_{\text{sp}} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$  with  $B \neq 0$  or  $C \neq 0$ .

$$\boxed{[4.5]} \quad P_{\text{sp}} \in \mathcal{B} \setminus \mathcal{A} \implies \Pi_{\text{act}} R_{\text{sp}}(r_i)^{\top} R_{\text{sp}}(r_j) \Pi_{\text{act}} \neq \Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}} \quad ([3.14], [4.1], [2.13], [2.6]) \quad (60)$$

## 6 BEG-5: Attention Logits

$$\boxed{[5.1]} \quad W_Q, W_K, W_V \in \mathbb{R}^{d_h \times D}, \quad x_i \in \mathbb{R}^D, \quad r_i \in \mathbb{R}^{d_c} \quad ([0.2]) \quad (61)$$

$$\boxed{[5.2]} \quad q_i \stackrel{\mathcal{O}}{=} W_Q x_i, \quad k_i \stackrel{\mathcal{O}}{=} W_K x_i, \quad v_i \stackrel{\mathcal{O}}{=} W_V x_i \quad ([5.1]) \quad (62)$$

$$\boxed{[5.3]} \quad q_i^{(\text{act})} \stackrel{\mathcal{O}}{=} \Pi_{\text{act}} q_i, \quad k_j^{(\text{act})} \stackrel{\mathcal{O}}{=} \Pi_{\text{act}} k_j \quad ([5.2], [2.6]) \quad (63)$$

$$\boxed{[5.4]} \quad d_{\text{act}} \stackrel{\mathcal{O}}{=} \dim(\text{act}) = 2m \quad ([2.4], [3.11], [0.16]) \quad (64)$$

$$\boxed{[5.5]} \quad \tilde{q}_i \stackrel{\mathcal{O}}{=} \Pi_{\text{act}} R_{\text{sp}}(r_i) q_i^{(\text{act})}, \quad \tilde{k}_j \stackrel{\mathcal{O}}{=} \Pi_{\text{act}} R_{\text{sp}}(r_j) k_j^{(\text{act})} \quad ([4.1], [5.3], [2.6]) \quad (65)$$

$$\boxed{[5.6]} \quad \alpha_{ij} \stackrel{\mathcal{O}}{=} \frac{\tilde{q}_i^\top \tilde{k}_j}{\sqrt{d_{\text{act}}}} \quad ([5.5], [5.4]) \quad (66)$$

$$\boxed{[5.7]} \quad \alpha_{ij} = \frac{(q_i^{(\text{act})})^\top [(\Pi_{\text{act}} R_{\text{sp}}(r_i) \Pi_{\text{act}})^\top (\Pi_{\text{act}} R_{\text{sp}}(r_j) \Pi_{\text{act}})] k_j^{(\text{act})}}{\sqrt{d_{\text{act}}}} \quad ([5.6], [5.5], [5.3], [2.6]) \quad (67)$$

$$\boxed{[5.8]} \quad P_{\text{sp}} \in \mathcal{A} \implies (\Pi_{\text{act}} R_{\text{sp}}(r_i) \Pi_{\text{act}})^\top (\Pi_{\text{act}} R_{\text{sp}}(r_j) \Pi_{\text{act}}) = \Pi_{\text{act}} R_{\text{sp}}(r_j - r_i) \Pi_{\text{act}} \quad ([4.4], [4.1], [2.13], [2.6], [3.13]) \quad (68)$$

$$\boxed{[5.9]} \quad P_{\text{sp}} \in \mathcal{A} \implies \alpha_{ij} = \frac{(q_i^{(\text{act})})^\top [\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}}] k_j^{(\text{act})}}{\sqrt{d_{\text{act}}}} \quad ([5.7], [5.8], [4.4], [2.13], [5.4]) \quad (69)$$

$$\boxed{[5.10]} \quad P_{\text{sp}} \in \mathcal{B} \setminus \mathcal{A} \implies \alpha_{ij} \neq \frac{(q_i^{(\text{act})})^\top [\Pi_{\text{act}} R_{\text{STR}}(r_j - r_i) \Pi_{\text{act}}] k_j^{(\text{act})}}{\sqrt{d_{\text{act}}}} \quad ([4.5], [5.7], [4.1], [2.13], [3.14], [5.4]) \quad (70)$$

## 7 BEG-6: Summary and Fatal Check

Assuming [1.2] and [1.3], the statements [2.1] through [5.10] show:

- $R_{\text{STR}}(r)$  has the block-rotation form [2.11], lies in  $SO(d_h)$  [2.12], and satisfies the relative-offset identity [2.13].
- The Cayley transform parametrisation [3.7] yields  $P_{\text{sp}} \in SO(d_h)$  with controllable sparsity [3.5]; masks supported on the null block give  $\mathcal{A}$  [3.12] while all masks give  $\mathcal{B}$  [3.14].
- $R_{\text{sp}}(r)$  preserves purely relative offsets on the active block when  $P_{\text{sp}} \in \mathcal{A}$  [4.4], but leaks absolute position for  $P_{\text{sp}} \notin \mathcal{A}$  [4.5].
- Attention logits  $\alpha_{ij}$  inherit the relative-offset dependence when  $P_{\text{sp}} \in \mathcal{A}$  [5.9] and include absolute information otherwise [5.10].

The auxiliary cleanups `cleanup` add the explicit references requested by the reviewer without affecting hypotheses [1.2]–[1.3] or constructions [2.11], [3.7], [4.1], so statement [5.9] remains the main theorem.