
When Are Shared Observations Enough? Task-Relative Calibration for Neuro-Symbolic Robot Teammates

David Erman

ERMANDAVID613@GMAIL.COM

Department of Computer Science, Bar-Ilan University, Ramat Gan 5290002, Israel

Abstract

Robotic systems increasingly combine separately trained encoders with symbolic planners. The encoders may use different latent coordinates, so the planner must know whether shared observations suffice for its own readouts. Let Q_* be the unknown rotation or reflection from source to target coordinates. If X and Y collect the source and target encodings of shared observations, then $Y = Q_*X + E$. Let U span the source-space directions read by the task. In zero noise, every orthogonal map fitting the anchors gives the same task readouts if and only if $U \subseteq \text{span}(X)$; otherwise a reflection fixes all anchors but changes a task direction. We also give a directional noisy bound and an idealized resource law: if $k = \dim U < r$, then k chosen exact anchors are necessary and sufficient, whereas r generic passive anchors are needed almost surely. This yields an auditable accept/abstain/request-more interface. A preliminary non-robot test shows that orthogonality is only approximate.

1. Why Full Alignment Can Be Wasteful

Suppose two learned modules encode the same scene in three dimensions. Their paired observations occupy only a plane. Two source-to-target rotations/reflections can then agree on every observed pair while differing on the unseen third direction. If the symbolic planner reads only directions in the observed plane, that ambiguity is harmless; if a safety rule reads the third direction, it is not. The question is therefore not whether the whole latent spaces are aligned, but whether the remaining ambiguity can change a deployed readout.

2. Model and Exact Task-Level Boundary

Choose one encoder as the *source* and the other as the *target*; both use r -dimensional latent vectors. The main unknown is Q_* , the source-to-target coordinate map in our model. We restrict this map to rotations and reflections, so

$$Q_*^\top Q_* = I.$$

For the i th shared observation, or *anchor*, let $x_i \in \mathbb{R}^r$ and $y_i \in \mathbb{R}^r$ be its source and target encodings. We model $y_i = Q_*x_i + e_i$, where e_i is the mismatch for that pair. Stacking m anchors as columns gives

$$X = [x_1, \dots, x_m], \quad Y = [y_1, \dots, y_m] = Q_*X + E.$$

The task is specified in source coordinates. Let $U \subseteq \mathbb{R}^r$ be the *task subspace*: the span of the source-space directions the symbolic layer may read. If the source uses the scalar readout $u^\top x$, then

the corresponding target-space direction is Q_*u , because $(Q_*u)^\top(Q_*x) = u^\top x$. Thus only the action of the alignment on U can affect these readouts. Let $k = \dim U$.

Noisy certificate. Orthogonal Procrustes chooses $\widehat{Q} \in \arg \min_{Q^\top Q = I} \|Y - QX\|_F$. For each task direction $u \in \text{span}(X)$,

$$\|(\widehat{Q} - Q_*)u\|_2 \leq 2\|E\|_F \|X^\dagger u\|_2, \quad (1)$$

where $X^\dagger$ is the Moore–Penrose pseudoinverse. Thus X need not span the whole latent space; it only needs to support the directions the task uses. Procrustes optimality bounds $\|(\widehat{Q} - Q_*)X\|_F$ by $2\|E\|_F$, and $u = XX^\dagger u$ yields (1).

Exact boundary. Now set $E = 0$. Call an orthogonal map Q an *exact fit* when $QX = Y$. Then

$$\text{all exact fits map every } u \in U \text{ identically} \iff U \subseteq \text{span}(X). \quad (2)$$

For any exact fit Q , the map $S = Q_*^\top Q$ satisfies $SX = X$ and therefore fixes $\text{span}(X)$ pointwise. Hence $U \subseteq \text{span}(X)$ forces $Qu = Q_*u$ for every $u \in U$. Conversely, if the inclusion fails, reflection across the anchor span fixes every anchor but moves some task direction; composing it with Q_* gives a second exact fit. The obstruction is information-theoretic, not a Procrustes failure.

Anchor counts. Assume $0 < k < r$. If exact source-space anchors may be chosen freely, k are necessary and sufficient: fewer cannot span U , while a basis of U does. If anchors are instead independent draws from a distribution with a density on $\mathbb{R}^r$, any $m < r$ contain the fixed nonzero U in their span with probability zero, while r span $\mathbb{R}^r$ almost surely. Thus

$$\text{task-directed exact anchors} = k, \quad \text{generic passive threshold} = r. \quad (3)$$

Physical reachability can only make the active problem harder.

3. Meaning, Prior Work, and Limits

The symbolic layer can turn these statements into an audit rule. It declares U and a tolerated error in the mapped task directions. If a validated upper bound $\bar{e} \geq \|E\|_F$ is available, (1) certifies each mapped direction; in zero noise, (2) is exact. A failed certificate means *insufficient evidence*, not that a learned module is wrong: the system may abstain or request more shared observations.

Procrustes is classical (Schönemann, 1966; Arias-Castro et al., 2018). Relative representations, latent translation, semantic alignment, and robot stitching already connect learned spaces from shared observations (Moschella et al., 2022; Maiorca et al., 2023; Ricciardi et al., 2025; Jian et al., 2024); task-identifiability, task-sufficient world models, and neuro-symbolic task abstractions are further neighbors (Xi and Bloem-Reddy, 2022; Feng et al., 2026; Liang et al., 2024). We claim neither the first anchor method nor a new Procrustes algorithm. Our contribution is the exact task-restricted sufficiency/impossibility boundary, its idealized anchor-count consequence, and the symbolic accept/abstain contract.

A preliminary non-robot check on two independently trained MLP encoders for handwritten digits found that orthogonal alignment helped at small anchor counts, but a general linear/affine map fit better at large anchor counts. Orthogonality is therefore a falsifiable structural prior, not a universal property of learned representations. The next empirical step is a robot/world-model or cross-embodiment benchmark that measures task-readout agreement and restricts anchor selection to physically reachable states.

References

- Arias-Castro, E.; Javanmard, A.; and Pelletier, B. 2018. Perturbation bounds for Procrustes, classical scaling, and trilateration, with applications to manifold learning. *arXiv:1810.09569*.
- Feng, F.; Zheng, Y.; Fu, M.-H.; Chen, Y.; Chen, G.; Murphy, K. P.; Huang, B.; and Zhang, K. 2026. Learning task-sufficient world models by synergizing agentic exploration and structured modeling. *arXiv:2607.04409*.
- Jian, P.-C.; Lee, E.; Bell, Z. I.; Zavlanos, M. M.; and Chen, B.-Y. 2024. Perception stitching: Zero-shot perception encoder transfer for visuomotor robot policies. *Transactions on Machine Learning Research*.
- Liang, Y.-C.; Kumar, N.; Tang, H.; Weller, A.; Tenenbaum, J. B.; Silver, T.; and Ellis, K. 2024. VisualPredicator: Learning abstract world models with neuro-symbolic predicates for robot planning. *arXiv:2410.23156*.
- Maiorca, V.; Moschella, L.; Norelli, A.; Fumero, M.; Locatello, F.; and Rodolà, E. 2023. Latent space translation via semantic alignment. *arXiv:2311.00664*.
- Moschella, L.; Maiorca, V.; Fumero, M.; Norelli, A.; Locatello, F.; and Rodolà, E. 2022. Relative representations enable zero-shot latent space communication. *arXiv:2209.15430*.
- Ricciardi, A. P.; Maiorca, V.; Moschella, L.; Marin, R.; and Rodolà, E. 2025. Mapping representations in reinforcement learning via semantic alignment for zero-shot stitching. *arXiv:2503.01881*.
- Schönemann, P. H. 1966. A generalized solution of the orthogonal Procrustes problem. *Psychometrika* 31:1–10.
- Xi, Q.; and Bloem-Reddy, B. 2022. Indeterminacy in generative models: Characterization and strong identifiability. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, 6912–6939.