RANK CHANGES WHAT OVERLAP MEANS: DIMENSION-MATCHED GEOMETRY OF LORA ADAPTERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Subspace overlap is routinely read as evidence that two LoRA adapters share a task-relevant direction. Rank moves the reference point of that measurement: independent q-dimensional subspaces in a d-dimensional module have expected normalized overlap q/d. We call the resulting rank-induced occupancy component *phantom overlap* and measure it with a module-local Haar residual, followed by a marginal-preserving null for pair structure. In a controlled Llama-3.1-8B library, raw overlap rises 186% from ranks 24 to 512 while the Haar residual falls 40%; at three measured ranks, pair-interaction dispersion remains 12.2–$13.1\times$ its null expectation. Positive residuals recur across language and vision libraries, model scales, and training seeds, while the rank trajectory remains library-specific. Controlled two-task compositions show source acquisition and positive joint utility at moderate source strength, with target retention and coefficient sensitivity measured separately. A preregistered study over all 28 task pairs separates measurement from selection: dimension matching corrects cross-rank interpretation, whereas composition utility depends on the evaluated operating condition.

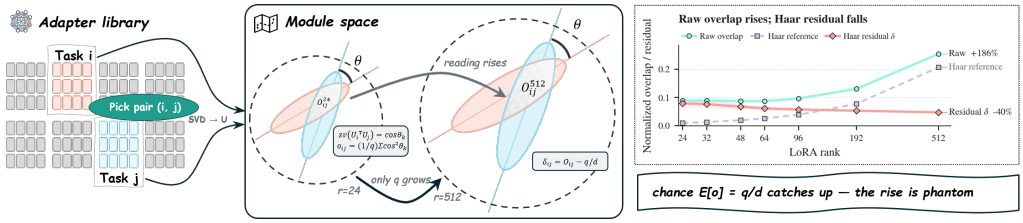

Figure 1: Measurement path for phantom overlap. We reconstruct the realized update subspaces by SVD for a task pair drawn from a completed adapter library, and compare their normalized overlap at low and high rank. The local Haar reference grows as q/d, so the raw overlap can rise mechanically even as the residual $\delta_{ij} = o_{ij} - q/d$ falls. The right panel shows the retained library trajectory: raw overlap rises while the Haar residual declines.

1 INTRODUCTION

Adapters are routinely compared by subspace overlap when libraries are merged or routed (Lu et al., 2026; Chowdhury et al., 2026; Tang & Yang, 2026). In one public LoRA-merging analysis, output-side overlap increases from 0.0166 at rank 8 to 0.0839 at rank 64, about five-fold, and the increase is read as accumulating shared directions (Tang & Yang, 2026). The dimension-matched independent-subspace reference rises eight-fold over the same range, so the observed five-fold trajectory is below the mechanical growth rate: relative crowding declines rather than accumulates. Rank sets the meter's zero point: independent q-dimensional subspaces in a d-dimensional module have expected normalized overlap q/d, so increasing rank mechanically raises overlap before trained orientation is considered, and the reference moves with the quantity being compared. The measurement question is whether the same number can be compared across ranks without conditioning on the dimensions that generate it.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We call this rank-induced occupancy component *phantom overlap*. The measurement path and rank reversal are summarized in Figure 1. In the Llama-3.1-8B library, raw overlap rises 186% from rank 24 to 512 while the Haar residual falls 40%, producing a raw-overlap/Haar-residual reversal; all 28 task pairs remain positive at each rank (Figure 2(a)).

We measure completed libraries from stored LoRA factors with a module-local Haar reference, then separate occupancy, orientation, and pair structure. Rank sweeps and controls test the effect across modules; cross-scale, cross-family, seed, and vision libraries test positive residuals rather than a universal residual trajectory. A marginal-preserving null checks whether pair interactions persist after additive task effects are removed. We connect the corrected measurement to behavior through two controlled composition studies: a four-pair endpoint analysis under declared protocols and a preregistered 28-pair selection study that tests whether geometry supports pair-specific rank selection.

We make four contributions. **First**, we identify and characterize *phantom overlap*: rank increases the overlap expected from independent subspaces. **Second**, we provide a module-local, dimension-matched measurement that separates this occupancy from the Haar residual in completed adapter libraries. **Third**, we show that a marginal-preserving null exposes pair interactions and that positive residuals recur across libraries while their rank trajectories differ. **Fourth**, we provide a controlled two-sided composition evaluation that identifies an operating region where moderate source strength acquires source capability while preserving positive joint utility. Before comparing adapter overlap across ranks, use a dimension-matched reference.

2 Related Work

Model merging and adapter composition. Model merging combines independently trained updates through arithmetic or averaging (Ilharco et al., 2022; Wortsman et al., 2022), sparsification and sign resolution (Yadav et al., 2023; Yu et al., 2024), regression (Jin et al., 2022), or adaptive coefficients (Yang et al., 2024). LoRA-specific methods use sparse or orthogonal adapters, magnitude–direction separation, and task projectors (Zhang et al., 2025; Zheng et al., 2025; Ryum & Kang, 2026); continual methods constrain update directions or filter occupied spaces (Wang et al., 2023; Liang & Li, 2024; Tang et al., 2026a; Cheng et al., 2025; Qiu et al., 2026). These lines use geometry to alter training or composition; we study the complementary measurement problem: how pairwise geometry in a completed adapter library should be read when rank and module dimension vary, and how pair structure persists after single-task marginals are removed.

Predicting mergeability. Demystifying Mergeability combines weight-, subspace-, activation-, and gradient-based metrics and finds method- and partner-specific success signatures (Zhou et al., 2026). M-Loss and MergeProbe use examples or training traces to estimate retention and select merge actions (Wang et al., 2026a; Tang et al., 2026c). Weight-only and post-training Fisher diagnostics require calibration to the adapter regime and objective (Paul, 2026; Kukanov et al., 2026), while write-direction controls show that effective LoRA directions can be sparse and behaviorally consequential (Sato et al., 2026). Angular alignment has weak links to composition performance, and functional interference varies with inputs and prompt format (Sivaramakrishnan et al., 2026; Zhu et al., 2026). Metric choice can likewise dissolve an apparent qualitative conclusion (Schaeffer et al., 2023). We therefore treat composition as a behavioral endpoint, measuring target retention and source acquisition separately before forming equal-weight utility.

Rank-conditioned subspace measurement. Principal angles and projector statistics compare linear subspaces without choosing a basis. Geometry-aware methods use them to alter training or merging, while gradient geometry can trigger capacity separation (Feng, 2026; Lu et al., 2026; Wang et al., 2026b; Chowdhury et al., 2026); related work also measures adapter redundancy and activation-feature geometry across rank settings (Dieudonné et al., 2026; Prasanth, 2026). Rank-performance comparisons are sensitive to the training-budget rule (Tang et al., 2026b). In LoRA merging, Tang and Yang report rising output-side overlap across ranks (Tang & Yang, 2026), while the free-position law of Xie and Alanwar calibrates update-versus-pretrained overlap (Xie & Alanwar, 2026). We supply the corresponding module-local, dimension-matched reference for pairwise adapter overlap, separate Haar occupancy from Haar residual, and test pair interactions after additive task effects are removed.

3 PROBLEM SETUP AND ADAPTER GEOMETRY

Measurement Protocol. We measure a completed adapter library directly from its stored LoRA factors, without access to training data. For each adapter we reconstruct the realized update, compute module-level subspace statistics, and aggregate them into pair- and library-level quantities: the Haar reference, Haar residual, and pair-interaction test (Sections 4–5). Separately, a controlled composition stage evaluates a fixed target adapter before and after adding one fixed source update on both constituent tasks (Section 8).

Uncertainty reporting. Mean-based summaries use mean $\pm$ sample SD as the primary dispersion report. Bracketed values are 95% intervals: pair-cluster bootstrap intervals for library geometry and composition summaries, paired-example bootstrap intervals for output-change measures, task-node jackknife intervals for cross-library residuals, Student-t intervals across training seeds, and permutation intervals for null draws. Counts, ranks, correlations, and single archived trajectories remain descriptive point quantities.

3.1 ADAPTERS AS SUBSPACES

Let θ_0 denote a pretrained model and let $\mathcal{L} = \{\Delta\theta_1, \ldots, \Delta\theta_n\}$ be a library of independently trained LoRA adapters (Hu et al., 2021). At linear module l, adapter i contributes an update

$$\Delta W_i^l = B_i^l A_i^l = U_i^l \Sigma_i^l (V_i^l)^\top, \tag{1}$$

where the final equality is a compact singular value decomposition. The columns of $U_i^l \in \mathbb{R}^{d_{\text{out},l} \times q_i^l}$ span the *adapter subspace* $\mathcal{U}_i^l$, and $q_i^l = r_{\text{eff},i}^l$ is its effective rank. Because $\Delta W_i^l x \in \mathcal{U}_i^l$ for every module input x, this image space contains all output-perturbation directions available to the adapter. This weight-space object is directly computable from the realized update matrix.

The update matrix determines which output directions a task adapter can change, while the input-side factor and hidden activations determine how strongly those directions are reached on a particular example. The column space gives a data-free, basis-invariant description of available directions, and the singular values retain their update magnitudes. The pair (U_i^l, Σ_i^l) therefore provides the geometric and spectral information retained by this measurement; input-conditioned behavior is measured in the composition study.

The library view is important because merging is performed after the individual adapters have already been trained: we compare the realized updates at the same module and aggregate their pairwise descriptors across the targeted modules. Rank is thereby a property of the trained objects being combined. Nominal rank controls the maximum dimension, while effective rank and the module output dimension determine the local occupancy that the comparison actually sees.

3.2 PAIRWISE GEOMETRY

For two rank-q subspaces with orthonormal bases $U_i, U_j \in \mathbb{R}^{d \times q}$, the singular values of $U_i^\top U_j$ are $\{\cos\theta_k\}_{k=1}^q$, where $\theta_k \in [0, \pi/2]$ are their principal angles. Each subspace is a point on the Grassmann manifold $\mathrm{Gr}(q, d)$. Let $P_i = U_i U_i^\top$ and $P_j = U_j U_j^\top$ be the corresponding orthogonal projectors. We use

$$o_{ij} = \frac{1}{q} \operatorname{tr}(P_i P_j) = \frac{1}{q} \sum_{k=1}^{q} \cos^2 \theta_k \tag{2}$$

as normalized squared overlap. Experiments index configurations by the nominal global ratio r/d for comparability with LoRA practice. The exact local quantity is $q_i^l / d_{\text{out},l}$; consequently, global r/d is a coarse summary, while module-level ratios preserve architecture-specific detail.

The principal-angle profile is more informative than its mean alone. An angle near 0 indicates a shared output direction, an angle near $\pi/2$ indicates directions that are orthogonal in the linear update space, and intermediate angles describe partial overlap. The overlap o_{ij} averages the squared cosine and therefore measures how much of one subspace is visible through the other. It is insensitive to the choice of basis within either subspace, which is essential because different LoRA

factorizations can represent the same update. We retain the full angle profile in the analysis and use scalar summaries only to compare rank settings or to summarize a completed library reproducibly.

These geometric quantities describe the update space. Two adapters can share a direction while applying updates with different signs or magnitudes, and orthogonal output directions can still interact after nonlinear layers or through later modules. Geometry exposes the operating conditions presented to a merger; the application study measures the downstream score separately.

4 When Raw Overlap Misleads

We connect rank to subspace occupancy and distinguish partial overlap from its aligned and orthogonal endpoints. Complete proofs are in Appendix E.

A Random-Subspace Reference Independent Haar subspaces provide a rotationally invariant dimensional reference for trained adapters.

Proposition 4.1 (Expected Haar overlap)**.** *Let P_1 and P_2 project onto independent Haar-random q-dimensional subspaces of $\mathbb{R}^d$. Then*

$$\mathbb{E}[\mathrm{tr}(P_1P_2)] = \mathbb{E}\left[\sum_{k=1}^{q}\cos^2\theta_k\right] = \frac{q^2}{d}, \qquad \mathbb{E}[o_{12}] = \frac{q}{d}. \tag{3}$$

Rotational invariance gives $\mathbb{E}[P_i] = (q/d)I_d$, from which Equation 3 follows. The expected normalized squared overlap therefore increases linearly with the rank ratio. Departures of trained adapters from this reference capture task and optimization structure.

The Haar calculation provides the dimensional occupancy reference for trained adapters. It answers the baseline question: how much overlap would be present if the two subspaces had no preferred orientation beyond their dimensions? A measured departure from this reference is the pair-level Haar residual analyzed in the rank sweep; its magnitude reflects the combined effect of task similarity, shared initialization, optimization, and architecture.

For adapters with possibly unequal effective ranks, the pair shares $q^l_{ij} = \min(q^l_i, q^l_j)$ principal angles, and the normalized overlap in Equation 2 becomes $o^l_{ij} = \mathrm{tr}(P^l_i P^l_j)/q^l_{ij}$. The corresponding dimension-only reference under this normalization is

$$o^l_{ij,\mathrm{Haar}} = \frac{\max(q^l_i, q^l_j)}{d_{\mathrm{out},l}}, \qquad \delta^l_{ij} = o^l_{ij} - o^l_{ij,\mathrm{Haar}}. \tag{4}$$

For equal-rank pairs this reduces to $q/d_{\mathrm{out},l}$. A positive δ^l_{ij} means that the realized pair is more aligned than independent Haar subspaces with the same local dimensions. The Haar residual is a weight-space measure of departure from the dimensional null. The measurement in Appendix C.1 applies this comparison to the existing adapter weights. The mean-angle envelope, unavoidable-intersection result, and projection-commutator identity provide additional descriptors of high occupancy and partial overlap. Appendix E states and proves these results. The empirical argument in the main text uses the directly observed overlap and its local Haar residual.

5 Dimension-Matched Measurement

5.1 Adapter-Only Measurements

For each adapter pair and linear module, we compute the compact SVD, principal angles, normalized overlap o^l_{ij}, and angular non-commutativity u^l_{ij}. Write $q^l_{ij} = \min(q^l_i, q^l_j)$ for the number of principal angles available to a pair. We distinguish the arithmetic mean angle $\bar{\theta}$ from the rank-level mean-cosine summary $\theta_c = \arccos(\bar{c})$, where $\bar{c}$ is the pooled mean of $\cos\theta^l_k$ over the same adapter-pair and module units. For a library-level threshold summary, we pool the near-orthogonal fraction across all adapter pairs and modules:

$$p_\perp(\phi) = \frac{\sum_{i<j}\sum_{l=1}^{L}\sum_{k=1}^{q^l_{ij}}\mathbf{1}[\theta^l_k \geq \phi]}{\sum_{i<j}\sum_{l=1}^{L} q^l_{ij}}, \tag{5}$$

Table 1: Rank-conditioned geometry for Llama-3.1-8B ($d = 4096$). $\theta_c = \arccos(\bar{c})$ is the pooled mean-cosine angle, $p_\perp$ is the fraction of principal angles above $80°$, and $\Delta = p_\perp - p_\perp^{\text{Haar}}$ is in percentage points. The Haar quantities use each module's effective rank and output dimension; uncertainty details are in Appendix C.4.

Ranks 16–64						Ranks 96–512					
r	r/d	θ_c	$p_\perp$	$p_\perp^{\text{Haar}}$	Δ	r	r/d	θ_c	$p_\perp$	$p_\perp^{\text{Haar}}$	Δ
16	0.39%	75.9°	46.6%	94.5%	−47.9	96	2.34%	76.3°	49.3%	67.5%	−18.2
24	0.59%	76.8°	49.9%	90.9%	−40.9	128	3.12%	75.5°	47.0%	61.7%	−14.8
32	0.78%	76.9°	50.6%	88.2%	−37.6	192	4.69%	73.6°	41.6%	52.0%	−10.4
48	1.17%	77.1°	51.7%	81.1%	−29.3	256	6.25%	72.3°	37.9%	45.5%	−7.6
64	1.56%	77.1°	51.5%	75.3%	−23.8	512	12.50%	66.0°	27.2%	31.7%	−4.4

using the fixed reference $\phi = 80°$ in every experiment. The full principal-angle distribution remains the primary measurement; $p_\perp$ provides one consistent scalar summary.

The measurement has three levels. Module-level angle profiles record orientation; pair-level overlap, local Haar residual, non-commutativity, and magnitude summarize interactions; library-level pooling gives a rank-conditioned view of the collection. This separation preserves dimensional occupancy $q_i^l/d_{\text{out},l}$, realized overlap o_{ij}^l, and Haar residual δ_{ij}^l as distinct quantities. Because the SVD is applied to the realized update ΔW_i^l, the measurements are invariant to rotations of the internal LoRA factors. They are computed after training and before composition, so they describe the stored library that a merger consumes.

5.2 Marginal-Preserving Pair-Interaction Null

The Haar residual removes expected overlap induced by local dimension. Its pair variation can still arise when one task has systematically high or low residual overlap with every partner. We separate this single-task contribution inside each exact module by fitting the complete task-pair matrix as

$$\delta_{ij}^l = \mu_l + a_i^l + a_j^l + e_{ij}^l, \qquad \sum_i a_i^l = 0, \tag{6}$$

where μ_l is the module mean, a_i^l is task i's additive marginal, and e_{ij}^l is the remaining pair interaction. Least squares on all 28 edges makes every task-incidence sum of e_{ij}^l zero.

The null holds $\mu_l + a_i^l + a_j^l$ fixed and independently relabels the eight task nodes of the interaction matrix e^l within each module. Node relabeling preserves the zero incidence sums, so every fitted marginal, module, and interaction multiset remains fixed. Let $\bar{e}_{ij} = L^{-1}\sum_l e_{ij}^l$ and let s_{int} be its sample standard deviation over the 28 pairs. Excess s_{int} tests whether a stable pair interaction recurs across modules after both dimensional occupancy and additive task effects are removed.

5.3 Composed-Task Utility

The controlled intervention adds one fixed source update to one fixed target at each rank,

$$\Delta W_{t+s}(r) = \Delta W_t(r) + \gamma(r)\Delta W_s(r), \qquad \gamma(r) = c\,\frac{\|\Delta W_t(r)\|_F}{\|\Delta W_s(r)\|_F}. \tag{7}$$

The primary protocol sets $c = 0.5$ globally. Coefficient controls use $c \in \{0.25, 1.0\}$, and a module-wise protocol scales every source module to half the norm of its target module. Exact factor concatenation represents each sum. We hold this exact additive composition operator fixed throughout the behavioral study, so the intervention isolates rank selection and source strength.

For target data D_t, source data D_s, target-only model M_t, composed model M_{t+s}, and source specialist M_s, where M_s is the source-task adapter evaluated on D_s, define target retention $R_t = m(M_{t+s}, D_t) - m(M_t, D_t)$ and source acquisition $A_s = m(M_{t+s}, D_s) - m(M_t, D_s)$. Write $S_s = m(M_s, D_s)$ for the source-specialist score. The primary endpoint is equal-weight joint utility $U = (R_t + A_s)/2$; S_s supplies a recovery reference. We use normalized exact match for non-generation tasks and ROUGE-L for generation tasks.

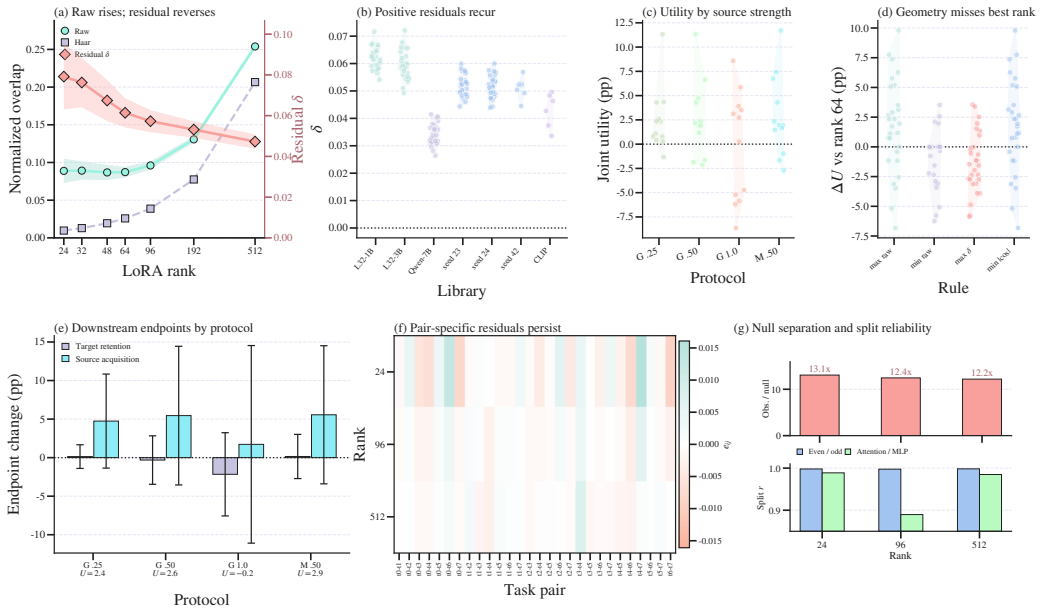

Figure 2: Main experimental panels. (a) Raw overlap and Haar residual across ranks. (b) Residual values across libraries and seeds. (c) Joint utility across composition protocols. (d) Utility under frozen rank-selection rules. (e) Target retention and source acquisition. (f) Pair-interaction residuals. (g) Observed and null interaction dispersion. Uncertainty units follow the corresponding result sections.

5.4 Interpreting the Quantities

At rank r, we retain the local occupancy, realized overlap, dimension-matched residual, and pair-interaction residual:

$$\mathcal{A}_r = \left\{ \left(\rho_i^l, o_{ij}^l, \delta_{ij}^l, e_{ij}^l\right) : \rho_i^l = \frac{q_i^l}{d_{\text{out},l}},\ i < j,\ l \right\}. \tag{8}$$

It also stores the full principal-angle profile and the module identifiers used in every aggregate. This representation preserves the distinction between a change in the available dimensions, a change in pair orientation, and a change in the way a particular task pair recurs across modules.

The first calibration stage pairs each observed overlap with the Haar reference computed from that module's effective ranks and output dimension, so cross-rank comparisons retain o, o_{Haar}, and δ together and modules with different output dimensions stay on a common interpretive scale. The second stage tests pair-specific recurrence: the node-relabeled null keeps each module, its fitted marginals, and its interaction multiset fixed, asking whether the same task-pair pattern recurs across modules rather than whether the library has a nonzero average Haar residual.

From geometry to composition. We compare ranks using o and δ, test pair recurrence with the interaction statistic, and then evaluate a target–source pair under a declared composition coefficient and allocation rule. Target retention and source acquisition remain separate endpoints. The preregistered selection study in Section 8.2 applies frozen rules to all 28 task pairs; its complete rows are in Appendix D.

6 Rank-Conditioned Geometry Reveals the Reversal

The experiments proceed as a four-step argument: rank-conditioned geometry establishes the raw trajectory; a dimension-matched null identifies the reversal; marginal-preserving tests establish stable pair structure and cross-library recurrence; controlled composition quantifies utility and tests

Table 2: Pair-interaction dispersion versus the marginal-preserving null.

r	Observed s_{int}	Null mean $\pm$ SD	95% null interval	Ratio
24	5.735	0.439 ± 0.070	[0.311, 0.585]	$13.05\times$
96	2.173	0.175 ± 0.028	[0.123, 0.231]	$12.45\times$
512	1.973	0.162 ± 0.025	[0.116, 0.213]	$12.18\times$

Note. All s_{int} values are scaled by 10^3. The null column is mean $\pm$ sample SD over 5,000 marginal-preserving permutation draws; the 95% interval is a null-draw reference interval rather than a trained-library confidence interval. At every rank, the observed statistic exceeds all 5,000 permutation draws; with the standard plus-one correction, $p = 1/5{,}001 \approx 0.0002$.

geometry-based rank selection. SuperNI supplies the principal Llama libraries and composition test, while CLIP/VTAB supplies a cross-modal stored-weight measurement. Dataset, training, metric, and provenance details appear in Appendix B.

6.1 Nominal Rank Organizes Geometry

We sweep ten nominal ranks from $r = 16$ to 512 for a model with hidden dimension $d = 4096$, covering $r/d = 0.39\%$ to 12.5%. The mean-cosine angle summary θ_c remains near 76° at low ratios and then falls to 66.0° at the largest measured ratio. The fraction of angles above the fixed 80° reference shows a gradual low-rank drift before a sustained high-rank decline, falling from 51.7% at $r = 48$ to 27.2% at $r = 512$. Near-orthogonality persists as nominal rank grows through $r = 64$; sustained geometric change emerges at larger ranks.

This trajectory separates nominal capacity growth from the realized geometric shift (Table 1). Through $r = 64$ the observed fraction stays within 46.6–51.7%, while the matched Haar expectation falls from 94.5% to 75.3% and the angle-domain residual Δ rises from -47.9 to -23.8 percentage points. Across the full sweep Δ ends at -4.4 percentage points at $r = 512$; the overlap-domain Haar residual in Section 6.2 remains positive for all 28 pairs at every rank.

6.2 A Local Dimension-Matched Null Reverses the Cross-Rank Ordering

We re-measure one consistent eight-task Llama library at ranks $24, 32, 48, 64, 96, 192,$ and 512, comparing every module with an independent Haar pair having the same effective ranks and output dimension.

Figure 2(a) contains the paper's central reversal. Across the full seven-rank trajectory, raw overlap rises from 0.08886 ± 0.01610 $[0.08326, 0.09489]$ to 0.25390 ± 0.00353 $[0.25262, 0.25517]$ $(+186\%)$, while the Haar residual falls from 0.07918 ± 0.01610 $[0.07358, 0.08521]$ to 0.04726 ± 0.00353 $[0.04599, 0.04854]$ (-40%); intervals are 95% pair-bootstrap intervals. The larger high-rank overlap therefore corresponds to a smaller departure from the local dimensional reference. All 28 pairs have positive residuals throughout the measured interval.

7 Stable Pair Structure Persists After Dimension Matching

7.1 Marginal-Preserving Tests Reveal Pair Interactions

We fit Equation 6 in each of 224 modules at ranks 24, 96, and 512, then apply 5,000 marginal-preserving interaction permutations. The additive task terms explain 85.8%, 83.4%, and 71.8% of within-module pair variance on average, making their removal a substantive control. The test statistic is the dispersion of the 28 interaction means across modules.

Observed pair-interaction dispersion is 12.2–$13.1\times$ the null mean and exceeds every permutation draw (Table 2; Figure 2(f)–(g)). Disjoint module splits recover the same 28-pair pattern: even versus odd layers have Pearson $r = 0.998$ at all three ranks, and attention versus MLP projections give $r = 0.890$–0.989; exhaustive eight-node QAP tests give exact $p = 1/40{,}320$ throughout. Separate replication tests establish positive residuals across Llama-3.2-1B/3B, Llama-3.1-8B, and Qwen2.5-7B libraries, three Llama-8B training seeds at rank 256, and five CLIP ranks (Figure 2(b)). Appendix C.4 lists module distributions, task-node jackknife intervals, the original task-label null, and cross-library controls.

Uncertainty estimates support this result. The grand Haar residual is positive at all three ranks. Its leave-one-task-out jackknife 95% Student-t intervals are $[0.05901, 0.09935]$, $[0.05062, 0.06398]$, and $[0.04333, 0.05119]$ for ranks 24, 96, and 512, respectively. The paired contrast between ranks 192 and 24 also changes sign between raw overlap and the residual. Raw overlap increases by 0.04179, with a 95% task-node interval of $[0.02466, 0.05892]$; the residual decreases by 0.02601, with an interval of $[-0.04314, -0.00888]$. The standard deviation across the 28 paired contrasts is 0.01358 for both quantities.

7.2 Positive Residual Recurs Across Libraries

The cross-library check is evaluated at the level of positive residual. Figure 2(b) shows the aligned-layer residuals across model scale, family, seed replicates, and modality; every Llama library shows 28/28 positive task pairs, while all 30 CLIP rank–pair conditions have positive residuals (rank-specific means, intervals, and shuffled controls in Appendix C.4). The rank trajectory is library-specific: the main Llama library declines from 0.0792 to 0.0473, Llama-3.2-3B rises from 0.0380 to 0.0659, and CLIP rises through rank 128 before decreasing at rank 256. The reusable result is the local dimension-matched reference and its positive-departure test.

8 Controlled Composition Across Ranks and Protocols

8.1 Moderate Source Strength Reveals a Positive Joint-Utility Window

We evaluate four target–source pairs spanning all eight SuperNI domains across seven ranks (24–512), with the pair set fixed before cross-rank behavior is assessed. Each composed model is evaluated on 128 target and 128 source examples. The same-rank target-only model provides both anchors, so source acquisition measures skill added to the deployed target model. Table 11 (Appendix D) reports means $\pm$ sample SDs and pair-bootstrap 95

The primary intervention acquires source-task skill at every rank in the four-pair mean (Figure 2(e)), and joint utility remains positive at all seven ranks. Global $c = 0.25$ produces positive joint utility in 11/12 matched conditions, while unit strength gives negative aggregate utility (Figure 2(c)). The point estimate for the paired $c = 0.5$ minus $c = 1.0$ joint-utility difference is 2.79 ± 3.46 points with a pair-cluster bootstrap 95% CI of $[0.00, 5.58]$; the four-pair exact sign test gives $p = 0.375$, so we report this coefficient contrast as a descriptive operating-region comparison. Appendix D lists absolute scores, gaps, and condition-level results.

8.2 Prespecified Geometry Rules Do Not Select the Utility-Optimal Rank

Under this fixed additive operator, the preregistered study enumerates all 28 directed task pairs across the seven ranks, with every selection rule frozen before utility inspection (Figure 2(d)): the prespecified max-δ rule selects rank 24 in 21/28 pairs and rank 32 in 7/28, and falls 1.314 percentage points below nominal rank 64 (-1.314 ± 2.561, pair-bootstrap 95% CI $[-2.23, -0.37]$, Cliff's $\delta = -0.500$, W/L/T $= 7/21/0$); the min-raw-overlap rule also falls below nominal rank 64, by 0.955 points (-0.955 ± 2.380, 95% CI $[-1.84, -0.12]$, $\delta = -0.321$, W/L/T $= 4/13/11$); and the min-$|\cos \Delta W|$ rule gains 1.703 points (1.703 ± 3.936, 95% CI $[0.28, 3.12]$, $\delta = 0.393$, W/L/T $= 19/8/1$, Holm-adjusted $p = 0.078$) but selects rank 512 in all 28 pairs, exactly matching max raw overlap and the always-largest-rank rule; this gain follows the rank-utility trend rather than pair-specific geometry, with mean per-pair Spearman correlations with utility of $+0.289$ for raw overlap, -0.293 for δ, -0.306 for $|\cos \Delta W|$, and $+0.306$ for nominal rank; the idealized model in Appendix D.2 explains why projector geometry omits sign and update-magnitude information, Section 8.3 summarizes the argument, and Appendix D lists the complete selection table, fixed-rule regret, and secondary output-change measures.

8.3 An Idealized Account of the Selection Boundary

Why do the tested projector-statistic rules fail to recover pair-specific utility? In the quadratic idealization of Appendix D.2, each task has displacement $e_t = \theta_t^* - \theta_0$ and loss $L_t(\theta) = \frac{1}{2}(\theta_t^* - \theta)^\top H(\theta_t^* - \theta)$. A converged rank-$r$ adapter writes $\Delta_t = \Pi_t e_t$, and the composed model is

$\theta_0 + \Delta_t + \gamma\Delta_s$. Expanding both losses at the deployment anchor $\theta_0 + \Delta_t$, with $\langle u, v\rangle_H = u^\top H v$, gives

$$R_t = \gamma \langle \Delta_s, (I - \Pi_t)\, e_t\rangle_H - \tfrac{\gamma^2}{2} \|\Delta_s\|_H^2, \tag{9}$$

$$A_s = \gamma \langle \Delta_s,\, e_s - \Delta_t\rangle_H - \tfrac{\gamma^2}{2} \|\Delta_s\|_H^2, \tag{10}$$

Three properties follow. **Residual orthogonality**: the benefit term of R_t is the alignment of Δ_s with $(I - \Pi_t)e_t$, the target displacement left outside the target adapter. Pairwise overlap measures shared directions, which enter the cost and interference terms. **Sign blindness**: the interference $-\gamma\langle\Delta_s, \Delta_t\rangle_H$ depends on the sign of the alignment, while squared-cosine overlap is invariant to $\pm\Delta_s$. **Magnitude dominance**: **Magnitude dominance**: under the planted scaling used here, $\|\Delta_s\|_H^2 = \|\Pi_s e_s\|_H^2$ grows as $O(r)$ with rank, whereas the corresponding alignment terms grow as $O(\sqrt{r})$. Update magnitude can therefore dominate cross-pair utility variation even when the projector geometry is held fixed. The planted instantiation reproduces the measurement pattern: with k=3 shared directions and isotropic H, raising rank moves raw overlap from 0.23 to 0.50 while the Haar residual falls from 0.22 to 0.002, so the raw-overlap/Haar-residual reversal in Section 6.2 arises from the dimension-matched geometry alone; the rank-monotone update cosine likewise collapses the frozen min-$|\cos \Delta W|$ rule onto rank 512, matching Section 8.2. Projector geometry is therefore a valid measurement of occupancy and orientation, but it is incomplete as a standalone predictor of two-sided composition utility.

9 Discussion, Limitations, and Future Work

Interpretation. The measurement separates the dimensional occupancy of a realized adapter pair from its Haar residual, then separates additive task marginals from pair-interaction residuals. The Llama library supplies the central raw-overlap/Haar-residual reversal; the cross-library measurements show positive residuals along library-specific trajectories, keeping empirical recurrence separate from a universal rank law.

Practical value and guidance. Composition provides the behavioral endpoint: target retention and source acquisition. The positive-utility regime occurs at moderate source strength, rather than at unit strength. These results suggest three practical guidelines. (i) Report the dimension-matched reference q/d alongside every cross-rank overlap claim: rising raw overlap is the expected occupancy trajectory of independent subspaces, not evidence of newly shared directions. (ii) Before attributing structure to a task pair, use the module-local Haar residual δ and the marginal-preserving interaction null; both can be computed from stored factors alone. (iii) Treat geometry-based rank or pair selection as a hypothesis to be tested against a fixed-rank behavioral baseline. In our preregistered study, every projector-statistic rule reduced to a global rank preference, whereas moderate-strength composition delivered positive utility without consulting downstream scores. In the composition study, we hold the merger operator fixed to exact addition to isolate rank selection and examine how geometry-based selection relates to behavioral utility.

Limitations and future work. Our measurements concern the realized linear update subspaces; their behavioral effects also depend on input-conditioned activations, update signs, network curvature, and the merger rule. We use exact additive composition throughout the composition study to provide a controlled setting for examining how rank-conditioned geometry relates to composition utility. Our behavioral evaluation covers four SuperNI composition pairs and one Llama adapter family. Future work can extend this measurement and two-sided evaluation to a broader range of merger protocols, modalities, and activation-level mechanisms.

10 Conclusion

Rank changes what adapter overlap means. Phantom overlap is the dimensional occupancy component of a rank-conditioned comparison; the module-local Haar residual and the marginal-preserving interaction test separate it from trained structure in stored adapter libraries, with positive residuals across model scales, families, seeds, and a vision library. Controlled composition identifies a moderate-strength region of positive joint utility, while the complete 28-pair study shows that geometry-based selectors reduce to global rank preferences rather than pair-specific choices.

MANDATORY STATEMENTS

AI USE STATEMENT

Generative AI tools supported this work in five limited roles: (i) improving manuscript readability by polishing selected passages and condensing verbose text while preserving its information; (ii) suggesting organizational alternatives for the paper; (iii) locating and identifying literature relevant to the study, with the authors independently assessing its relevance, novelty, and relationship to prior work; (iv) drafting and refining plotting code for figures based on experiments executed by the authors; and (v) drafting, debugging, and refactoring selected code fragments in experiment runners and reproducibility utilities. The authors designed and controlled the core implementation and experimental logic. Before use, each AI-assisted code fragment was checked against the intended protocol and exercised through the relevant experiment or unit-test pipeline. AI tools did not determine the theoretical model, methodology, experimental protocol, reported results, or scientific interpretation. All experimental data and figures derive from author-executed runs and recorded outputs. The authors reviewed and verified all cited references, checked the behavior of AI-assisted code and outputs, and take responsibility for the final manuscript, claims, and artifacts produced with AI assistance.

REPRODUCIBILITY STATEMENT

Appendix B documents datasets, training configurations, adapter provenance, metrics, and thresholds. Appendices C–D specify the null constructions, uncertainty procedures, controls, and condition-level results. The accompanying artifact includes deterministic analysis scripts, machine-readable outputs, execution logs, file hashes, and contract tests for the pair-interaction and composition analyses. The released package regenerates the reported statistics from the included outputs; base-model checkpoints, original adapter weights, and private training infrastructure are not redistributed.

REFERENCES

Runxi Cheng, Feng Xiong, Yongxian Wei, Wanyun Zhu, and Chun Yuan. Whoever started the interference should end it: Guiding data-free model merging via task vectors. *arXiv preprint arXiv:2503.08099*, 2025.

Mehreen Hossain Chowdhury, Nowshin Mahjabin, Ahmed Shafin Ruhan, Md Azam Hossain, Abu Raihan Mostofa Kamal, and Md Tahmid Rahman Laskar. Routing is not enough: Diagnosing intra-adapter subspace contention in moe+ lora fine-tuning. *arXiv preprint arXiv:2609.03150*, 2026.

Tanguy Dieudonné, Giulia Lanzillotta, Enis Simsar, Louis Barinka, and Thomas Hofmann. When one adapter speaks for many: Discovering low-rank redundancy in continual fine-tuning. *arXiv preprint arXiv:2606.28117*, 2026.

Yibo Feng. Geo-lora: Geometry-aware subspace evolution for low-rank adaptation in continual learning. *arXiv preprint arXiv:2608.26960*, 2026.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022.

Xisen Jin, Xiang Ren, Daniel Preotiuc-Pietro, and Pengxiang Cheng. Dataless knowledge fusion by merging weights of language models. *arXiv preprint arXiv:2212.09849*, 2022.

Ivan Kukanov, Janne Laakkonen, and Ville Hautamäki. How meta-learning shapes lora adapter geometry in speech deepfake detection. *arXiv preprint arXiv:2607.22010*, 2026.

Yan-Shuo Liang and Wu-Jun Li. Inflora: Interference-free low-rank adaptation for continual learning. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23638–23647. IEEE, 2024.

Chenyang Lu, Jiaru Li, Jinman Zhao, Xinran Chen, Yining Wang, Renyi Cai, Yuchen Li, and Chao He. On the representation geometry of LoRA model merging. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 5289–5304, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.261. URL https://aclanthology.org/2026.findings-acl.261/.

Roi Paul. Spectral geometry of lora adapters encodes training objective and predicts harmful compliance. *arXiv preprint arXiv:2604.08844*, 2026.

KK Prasanth. Feature geometry of lora adapters: A sparse autoencoder analysis of representational divergence in fine-tuned language models. *arXiv preprint arXiv:2605.28896*, 2026.

Zihuan Qiu, Lei Wang, Yang Cao, Runtong Zhang, Bing Su, Yi Xu, Fanman Meng, Linfeng Xu, Qingbo Wu, and Hongliang Li. Null-space filtering for data-free continual model merging: Preserving stability, promoting plasticity. In *International Conference on Learning Representations*, volume 2026, pp. 128492–128518, 2026.

Keumseo Ryum and Joonhyuk Kang. Ct-merging: Consensus directions and task-level scaling for lora adapter merging. *arXiv preprint arXiv:2607.20561*, 2026.

Haruto Sato, Yuki Tanaka, Ren Nakamura, Aoi Kobayashi, and Mei Ito. Behaviorally effective lora writes are sparse and structured. *arXiv preprint arXiv:2609.01374*, 2026.

Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo. Are emergent abilities of large language models a mirage? *Advances in neural information processing systems*, 36:55565–55581, 2023.

Gowtham Sivaramakrishnan, Sarvesha Kumar Kombaiah Seetha, Kishan Gupta Balaji, and Santhosh Baradwaj Vaduvur Ranganathan. Permdora–understanding adapter interference in language models: Limits of parameter-space geometry. *arXiv preprint arXiv:2606.11262*, 2026.

Anke Tang, Enneng Yang, Li Shen, Yong Luo, Han Hu, Lefei Zhang, Bo Du, and Dacheng Tao. Merging on the fly without retraining: A sequential approach to scalable continual model merging. *Advances in Neural Information Processing Systems*, 38:82081–82118, 2026a.

Haolun Tang, Jingyi Zhan, Yan Feng, and Zhipeng Chen. An artifact audit of budget-dependent LoRA rank comparisons, 2026b. URL https://doi.org/10.21203/rs.3.rs-10426794/v1. Research Square preprint, under review at Scientific Reports.

Lin Tang, Wei Zhang, Jing Li, Hongyu Chen, Ming Zhao, and Yuxuan Wang. Predicting mergeability of parameter-efficient fine-tuning updates. *arXiv preprint arXiv:2606.19549*, 2026c.

Yixuan Tang and Yi Yang. Crowded in b-space: Calibrating shared directions for lora merging. *arXiv preprint arXiv:2604.16826*, 2026.

Tiantong Wang, Yiyang Duan, Haoyu Chen, Tiantong Wu, and Wei Yang Bryan Lim. M-loss: Quantifying model merging compatibility with limited unlabeled data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26471–26479, 2026a.

Xiao Wang, Tianze Chen, Qiming Ge, Han Xia, Rong Bao, Rui Zheng, Qi Zhang, Tao Gui, and Xuan-Jing Huang. Orthogonal subspace learning for language model continual learning. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 10658–10671, 2023.

Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, et al. Super-naturalinstructions: Generalization via declarative instructions on 1600+ nlp tasks. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pp. 5085–5109, 2022.

Zhongsheng Wang, Zhedong Lin, Qian Liu, Xinyu Zhang, and Jiamou Liu. Sdo: Subspace deconflicting operator for multi-adapter composition. *arXiv preprint arXiv:2608.13820*, 2026b.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International conference on machine learning*, pp. 23965–23998. Pmlr, 2022.

Peng Xie and Amr Alanwar. The intruder threshold: A spectral law for lora fine-tuning. *arXiv preprint arXiv:2607.23711*, 2026.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. *Advances in neural information processing systems*, 36:7093–7115, 2023.

Enneng Yang, Zhenyi Wang, Li Shen, Shiwei Liu, Guibing Guo, Xingwei Wang, and Dacheng Tao. Adamerging: Adaptive model merging for multi-task learning. In *International Conference on Learning Representations*, volume 2024, pp. 22743–22763, 2024.

Linyan Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In *Icml*, volume 2, pp. 21, 2024.

Xiaohua Zhai, Joan Puigcerver, Alexander Kolesnikov, Pierre Ruyssen, Carlos Riquelme, Mario Lucic, Josip Djolonga, Andre Susano Pinto, Maxim Neumann, Alexey Dosovitskiy, et al. A large-scale study of representation learning with the visual task adaptation benchmark. *arXiv preprint arXiv:1910.04867*, 2019.

Juzheng Zhang, Jiacheng You, Ashwinee Panda, and Tom Goldstein. Lori: Reducing cross-task interference in multi-task low-rank adaptation. *arXiv preprint arXiv:2504.07448*, 2025.

Shenghe Zheng, Hongzhi Wang, Chenyu Huang, Xiaohui Wang, Tao Chen, Jiayuan Fan, Shuyue Hu, and Peng Ye. Decouple and orthogonalize: A data-free framework for lora merging. *arXiv preprint arXiv:2505.15875*, 2025.

Luca Zhou, Bo Zhao, Rose Yu, and Emanuele Rodolà. Demystifying mergeability: Interpretable properties to predict model merging success. *arXiv preprint arXiv:2601.22285*, 2026.

Chencheng Zhu, Xiaoyang Li, and Taotao Cai. When do task vectors interfere? mapping the validity boundaries of weight-space composition. *arXiv preprint arXiv:2608.09490*, 2026.

A Notation and Measurement Construction

Table 3: Notation used in the paper.

Symbol	Definition
$\Delta W_i^l = B_i^l A_i^l$	Realized LoRA update for adapter i and module l
q_i^l	Effective rank of ΔW_i^l
U_i^l	Orthonormal basis for $\mathrm{col}(\Delta W_i^l)$
$P_i^l = U_i^l (U_i^l)^\top$	Orthogonal projector onto the adapter subspace
$q_i^l / d_{\mathrm{out},l}$	Local effective rank ratio
θ_k^l	Principal angle between two local adapter subspaces
o_{ij}^l	Normalized squared subspace overlap
$o_{ij,\mathrm{Haar}}^l$	Module-local dimension-matched Haar reference
δ_{ij}^l	Haar residual $o_{ij}^l - o_{ij,\mathrm{Haar}}^l$
a_i^l	Additive single-task marginal in module l
e_{ij}^l	Pair-interaction residual after removing task marginals
$p_\perp(\phi)$	Fraction of pooled principal angles at least ϕ

The measurement reads stored LoRA factors after training. For every common LoRA module, it reconstructs the realized update, computes its compact SVD, retains the nonzero left singular vectors, and computes principal angles between task adapters. All pairwise comparisons are module

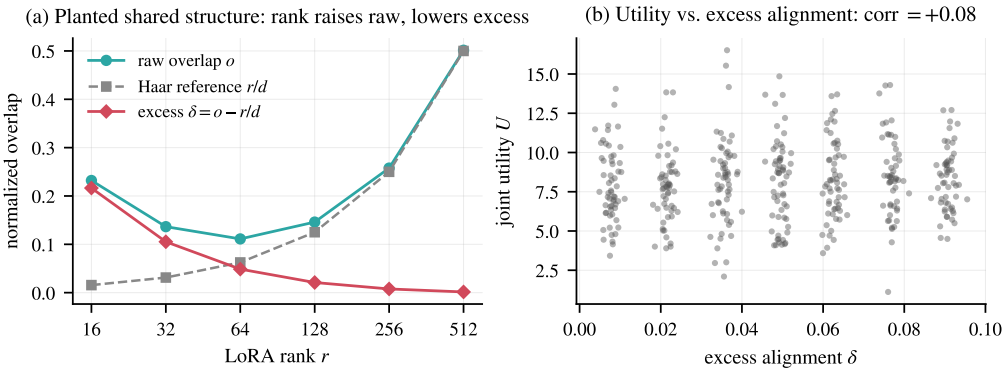

Figure 3: Idealized mechanism. (a) With a fixed set of planted shared directions (k=3, d=1024), raising the rank moves raw overlap up while moving the Haar residual down; the raw-overlap/Haar-residual reversal in Figure 2(a) arises from geometry alone. (b) Joint utility against the Haar residual over 400 planted pairs (r=64): within-strip variance dominates and the correlation is $+0.08$.

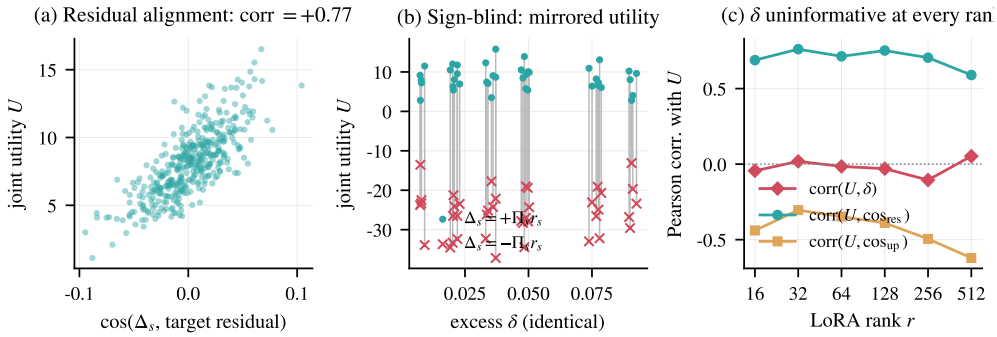

Figure 4: Utility information omitted by projector statistics in the idealized model. (a) At rank 64 over 400 planted pairs, utility tracks the Haar residual (corr $= +0.77$). (b) Sign blindness: at identical geometry, $\pm\Delta_s$ produce mirrored utilities. (c) Across ranks 16–512 with 300 planted pairs per rank, the Haar residual is uninformative ($|\text{corr}| \leq 0.11$) while residual-alignment correlations range from 0.590 to 0.761 and update cosine becomes increasingly negative at higher ranks. Panels (a) and (c) use different planted samples, so their rank-64 values are not expected to coincide.

matched. Module-level measurements are pooled with equal module weight; $p_\perp(80°)$ pools every available principal angle across modules and task pairs.

The targeted Llama-3.1-8B modules have output dimension 1,024 for `k_proj` and `v_proj`, 4,096 for `q_proj`, `o_proj`, and `down_proj`, and 14,336 for `gate_proj` and `up_proj`. The global index $r/4096$ therefore summarizes several local occupancy ratios. Every Haar reference uses the actual module output dimension and effective rank.

B Experimental Setup and Provenance

Super-NaturalInstructions. The language study uses eight SuperNI domain categories (Wang et al., 2022): question answering (D1), text classification (D2), text generation (D3), structure-to-text (D4), text-to-structure (D5), sequence tagging (D6), text modification (D7), and commonsense reasoning (D8). Each category contributes 1,000 training and 200 test examples. The composition evaluation uses the first 128 deterministic test examples on each selected task side.

Llama adapters. SuperNI adapters use Meta-Llama-3.1-8B-Instruct with LoRA on `q_proj`, `k_proj`, `v_proj`, `o_proj`, `gate_proj`, `up_proj`, and `down_proj`. The retained

Table 4: Null-adjusted overlap in one consistent eight-task Llama library. Raw overlap and residual are reported as pair-level mean ± sample SD [95% pair-bootstrap CI]; the Haar reference is the pooled module-level dimensional reference. Each row contains 224 modules and all 28 task pairs.

r	Raw overlap o	Haar reference	Residual δ	Positive pairs
24	0.08886 ± 0.01610 $[0.08326, 0.09489]$	0.00969	0.07918 ± 0.01610 $[0.07358, 0.08521]$	28/28
32	0.08919 ± 0.01214 $[0.08497, 0.09372]$	0.01291	0.07627 ± 0.01214 $[0.07205, 0.08081]$	28/28
48	0.08681 ± 0.01008 $[0.08327, 0.09063]$	0.01937	0.06744 ± 0.01008 $[0.06390, 0.07126]$	28/28
64	0.08724 ± 0.00717 $[0.08481, 0.09001]$	0.02583	0.06141 ± 0.00717 $[0.05898, 0.06418]$	28/28
96	0.09604 ± 0.00544 $[0.09412, 0.09809]$	0.03874	0.05730 ± 0.00544 $[0.05538, 0.05935]$	28/28
192	0.13065 ± 0.00404 $[0.12920, 0.13215]$	0.07749	0.05317 ± 0.00404 $[0.05172, 0.05466]$	28/28
512	0.25390 ± 0.00353 $[0.25262, 0.25517]$	0.20663	0.04726 ± 0.00353 $[0.04599, 0.04854]$	28/28

Table 5: Direct task-label null before removing additive task marginals. Null pair SD is mean ± sample SD over 5,000 draws. At every rank, the observed statistic exceeds all permutation draws; plus-one empirical $p = 1/5{,}001 \approx 0.0002$.

r	Observed pair SD	Null pair SD	95% interval	Ratio
24	0.016102	0.001176 ± 0.000269	[0.000698, 0.001732]	$13.7\times$
96	0.005436	0.000468 ± 0.000108	[0.000275, 0.000697]	$11.6\times$
512	0.003533	0.000366 ± 0.000079	[0.000225, 0.000531]	$9.7\times$

`intermediate_ranks` family fixes 1,000 examples per domain, three epochs, batch size 2, gradient accumulation 8, learning rate 10^{-4}, weight decay 0.01, LoRA dropout 0.05, and seed 42. It sets $\alpha_{\mathrm{LoRA}} = r$, so PEFT scaling equals one at every rank. Data exposure, optimizer steps, and target modules are shared across the measured ranks.

CLIP adapters. The vision measurement reads stored CLIP ViT-B/16 adapters for CIFAR-100, DTD, SVHN, and RESISC45 from VTAB-1k (Zhai et al., 2019). Ranks are 16, 32, 64, 128, and 256. LoRA targets `q_proj`, `k_proj`, `v_proj`, `out_proj`, `fc1`, and `fc2`; all factors have full nominal rank. The measurement covers 72 common modules and six task pairs per rank directly from the saved factors.

Composed-task utility. The four target–source pairs are D2←D5, D4←D7, D6←D1, and D8←D3. They cover every adapter exactly once; the pair set is fixed before cross-rank behavior is evaluated. The primary global-norm protocol covers ranks $24, 32, 48, 64, 96, 192$, and 512 with source ratio $c = 0.5$. Protocol controls use $c \in \{0.25, 0.5, 1.0\}$ under global norm matching and $c = 0.5$ under module-wise norm matching at ranks $24, 96$, and 512. Greedy decoding uses left padding, batch size 16, 128 maximum new tokens, and 128 fixed test examples on each task side. Random seeds and deterministic kernels are fixed to 42.

For each rank, source scaling is

$$\gamma(r) = c\,\frac{\|\Delta W_t(r)\|_F}{\|\Delta W_s(r)\|_F}. \tag{11}$$

The concatenated factors

$$B_{t+s} = [B_t, \sqrt{\gamma} B_s], \qquad A_{t+s} = [A_t^\top, \sqrt{\gamma} A_s^\top]^\top \tag{12}$$

represent $\Delta W_t + \gamma \Delta W_s$ exactly. The algebra self-test has relative error below 10^{-7}, all 448 LoRA tensors load in every condition, and the realized source-to-target update-norm ratio equals c. Module-wise matching uses $\gamma_l = 0.5\|\Delta W_t^l\|_F / \|\Delta W_s^l\|_F$ and retains a global realized ratio of 0.5. All behavioral comparisons retain this exact additive composition operator to isolate rank selection; alternative merger operators are outside the selection study.

Let C_t and C_s be the composed model's primary scores on target and source data, T_t and T_s the target-only model's scores on those data, and S_s the source specialist's source score. Target retention is $R_t = C_t - T_t$, source acquisition is $A_s = C_s - T_s$, and joint task utility is

$$U = \frac{1}{2}(R_t + A_s). \tag{13}$$

Table 6: Reliability of the 28-pair interaction pattern across disjoint module splits. Cells give Pearson r / Spearman ρ; every exhaustive two-sided eight-node QAP test has exact $p = 1/40{,}320$.

r	Even vs. odd layers	Attention vs. MLP
24	0.998 / 0.988	0.989 / 0.981
96	0.998 / 0.984	0.890 / 0.898
512	0.998 / 0.990	0.985 / 0.976

Table 7: Cross-modal stored-weight measurement on four CLIP/VTAB adapters. Raw overlap and residual are reported as pair-level mean ± sample SD [95% pair-bootstrap CI]; the Haar reference is the pooled module-level dimensional reference. Each rank contains 72 modules and six task pairs.

r	Raw overlap o	Haar reference	Residual δ	Positive pairs
16	0.04543 ± 0.00562 $[0.04171, 0.04983]$	0.01823	0.02720 ± 0.00562 $[0.02348, 0.03160]$	6/6
32	0.07270 ± 0.00746 $[0.06773, 0.07857]$	0.03646	0.03625 ± 0.00746 $[0.03127, 0.04212]$	6/6
64	0.11773 ± 0.00829 $[0.11228, 0.12425]$	0.07292	0.04482 ± 0.00829 $[0.03937, 0.05133]$	6/6
128	0.19617 ± 0.00785 $[0.19062, 0.20189]$	0.14583	0.05034 ± 0.00785 $[0.04479, 0.05605]$	6/6
256	0.33465 ± 0.00637 $[0.32982, 0.33903]$	0.29167	0.04298 ± 0.00637 $[0.03816, 0.04736]$	6/6

The source specialist supplies a recovery reference. Recoverable source-gap fraction is $(C_s - T_s)/(S_s - T_s)$ when $S_s > T_s$; zero and negative denominators are marked ineligible. A Pareto improvement has $R_t \geq 0$, $A_s \geq 0$, and a strict gain on at least one side. Target-loss-constrained acquisition requires $A_s > 0$ and $R_t \geq -\tau$ for a declared tolerance τ.

Task-side summaries use sample mean ± sample SD with a paired-example bootstrap 95% CI. The four frozen pairs are task-disjoint and cover all eight tasks once. Rank and protocol aggregates first average repeated ranks within each pair and then give mean ± sample SD with a pair-cluster bootstrap 95% CI. Geometry summaries over all 28 dependent pairs additionally use a leave-one-task-out jackknife; QAP tests relabel all eight task nodes together.

Confirmatory selection provenance. The 28-pair rank-selection study follows the frozen protocol in `SELECTION_CONFIRMATORY_PROTOCOL.md`. It uses the retained Llama-3.1-8B adapters, ranks $24, 32, 48, 64, 96, 192$, and 512, the `global_add_050` composition, 128 deterministic test examples per task side, greedy decoding, and seed 42. The execution log records 168 target-on-source anchor evaluations and 336 composition-side evaluations; the locked artifact aggregates them into 196 complete pair–rank rows with zero failed conditions. The factor-composition self-test has relative error 7.12×10^{-8}, and the reproduction gate matches the four retained pilot pairs. The run resumed across two instances, changed from PRO 6000 to RTX 5090, and used 196 protocol-authorized batch-size fallbacks; all deviations are disclosed in protocol Section 9. The locked result artifact is `stored_selection_result.json`; its SHA-256, the selection-analysis hash, and the retained-source hash are recorded in protocol Section 9. The frozen analysis script is `selection_confirmatory_analysis.py`.

C Additional Geometry Evidence

C.1 Llama Null-Adjusted Overlap

Each row pairs the realized overlap with the module-local Haar reference and the resulting residual. Effective rank equals nominal rank for every pooled unit; all rows use the same training family, module set, and rank definition.

C.2 Task Marginals and Pair Interactions

The direct module-wise task-label null preserves each module's realized subspaces and full overlap multiset. It gives a first test of named-pair coherence, but it also randomizes any stable single-task marginal. Table 5 retains this nested control.

Table 8: Cross-scale Llama measurements. Orth is $p_\perp(80°)$. Each row is a single retained library trajectory and is therefore reported descriptively; the five-seed uncertainty calculation is given separately below.

Model	d	r	r/d	Orth%
Llama-3.2-1B	2048	16	0.78%	40.8
		64	3.12%	41.2
		128	6.25%	35.5
		256	12.50%	26.7
Llama-3.2-3B	3072	16	0.52%	65.7
		64	2.08%	46.9
		128	4.17%	40.9
		256	8.33%	31.5

Table 9: Five-seed replication of $p_\perp(80°)$. Mean and sample SD are percentages; intervals are Student-t 95% CIs over the five training seeds. The displayed seed values are rounded, while the mean and SD use the higher-precision analysis aggregates.

r	r/d	Mean	SD	95% t CI
16	0.39%	46.9515	0.3603	[46.5041, 47.3989]
32	0.78%	50.5101	0.2293	[50.2254, 50.7948]
64	1.56%	51.3470	0.1640	[51.1434, 51.5506]
128	3.12%	46.7885	0.1027	[46.6610, 46.9160]
256	6.25%	37.5029	0.2256	[37.2228, 37.7830]

The main test fits Equation 6 in every exact module and relabels only the zero-marginal interaction matrices. The fitted additive terms explain 0.858 ± 0.050, 0.834 ± 0.088, and 0.718 ± 0.214 of within-module pair variance at ranks 24, 96, and 512. Removing this large marginal component leaves interaction dispersion 12.2–13.1$\times$ above the corresponding null. This conditional residual-label test preserves the fitted task marginals and each module's interaction multiset; the synthetic residual matrices serve as randomization references. The grand Haar residual remains positive under a leave-one-task-out jackknife using Student-t critical values with seven degrees of freedom: its 95% intervals are $[0.05901, 0.09935]$, $[0.05062, 0.06398]$, and $[0.04333, 0.05119]$. The paired rank-192 minus rank-24 contrasts also retain the reversal: raw overlap changes by 0.04179 ± 0.01358 with task-node interval $[0.02466, 0.05892]$, while residual changes by -0.02601 ± 0.01358 with interval $[-0.04314, -0.00888]$. SD describes the 28 pair contrasts; these intervals quantify finite-library task sensitivity.

C.3 CLIP Null-Adjusted Overlap

All 30 rank–pair conditions have positive residuals. The rows preserve the CLIP trajectory separately from the Llama trajectory; raw overlap and residual are listed together so the two libraries can be read under the same local null.

C.4 Cross-Scale and Cross-Family Evidence

Across libraries, the rank-specific residual point means span 0.0340 (Qwen2.5-7B, r=256) to 0.1029 (Llama-3.2-1B, r=16); every Llama library shows 28/28 positive pairs, and all 30 CLIP rank–pair conditions have positive residuals. Tables 8–10 give the per-rank summaries and controls.

The five-seed sweep varies the training seed for the $p_\perp$ summary (seeds 42–46); the residual replication in Table 10 uses seeds 23, 24, and 42. Both use the same training family as the main library.

The shuffled controls preserve trained subspaces and module dimensions while removing the original layer correspondence. The resulting comparison isolates the layer-correspondence-sensitive component without introducing a new training protocol.

Table 10: Stored-weight residual replication and layer-permutation control. Residual cells report mean ± pair SD [95% task-node jackknife CI] with Student-t critical values ($\mathrm{df} = n_{\mathrm{tasks}} - 1$). T/P is tasks/pairs. Control values are mean residual after deranging layer correspondences within projection type.

Library	r	T/P	Residual mean ± pair SD	95% CI	Shuffled control
Llama-3.1-8B	24	8/28	0.07918 ± 0.01610	[0.05901,0.09935]	0.00049
	96	8/28	0.05730 ± 0.00544	[0.05062,0.06398]	0.00061
	512	8/28	0.04726 ± 0.00353	[0.04333,0.05119]	0.00134
Llama-3.2-1B	16	8/28	0.10287 ± 0.01985	[0.07734,0.12840]	0.00077
	128	8/28	0.07150 ± 0.00582	[0.06433,0.07866]	0.00093
	256	8/28	0.06183 ± 0.00388	[0.05723,0.06642]	0.00161
Llama-3.2-3B	16	8/28	0.03804 ± 0.00842	[0.02902,0.04705]	0.00022
	128	8/28	0.06589 ± 0.00564	[0.05896,0.07282]	0.00113
	256	8/28	0.05929 ± 0.00519	[0.05300,0.06557]	0.00204
Qwen2.5-7B	256	8/28	0.03400 ± 0.00379	[0.02968,0.03833]	0.00036
Llama seed 23	256	8/28	0.05203 ± 0.00390	[0.04761,0.05646]	0.00123
Llama seed 24	256	8/28	0.05154 ± 0.00397	[0.04695,0.05614]	0.00118
Llama seed 42	256	4/6	0.05096 ± 0.00409	[0.03913,0.06279]	0.00126

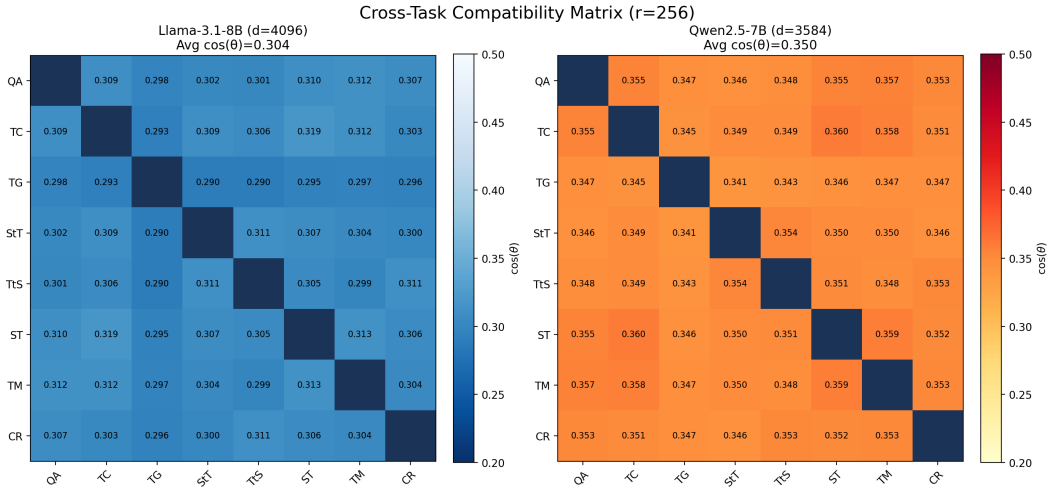

Figure 5: Pairwise mean-cosine matrices at $r = 256$ for Llama-3.1-8B and Qwen2.5-7B. Both libraries contain task-pair variation at different absolute levels.

D Composed-Task Utility and Output-Change Analysis

The primary endpoint is equal-weighted joint task utility. Target retention compares the composed model with the target specialist on target examples; source acquisition compares the same two models on source examples. Table 11 summarizes the rank and protocol panels, and Table 12 lists all 28 conditions from the primary global $c = 0.5$ protocol. The full analysis records each task-side delta as mean ± sample SD with a paired-example bootstrap 95% CI.

For global $c = 0.5$, the full pair-cluster summaries are $T_t = 64.52 \pm 20.29$ [43.94, 77.86], $C_t = 64.20 \pm 18.84$ [45.25, 75.60], $T_s = 38.69 \pm 14.67$ [27.90, 52.88], $C_s = 44.14 \pm 18.39$ [29.33, 58.95], and $S_s = 43.69 \pm 18.35$ [28.87, 58.50]. Among the nine conditions with positive recoverable gaps, six recover at least half and two meet or exceed the source-specialist gap; the other three conditions have $S_s \leq T_s$ and receive no recovery ratio. On the same 12 rank–pair conditions, the point estimate for global $c = 0.5$ minus global $c = 1.0$ joint utility is 2.79 ± 3.46 points with pair-cluster bootstrap 95% CI $[0.00, 5.58]$; the four-cluster exact sign test gives $p = 0.375$, so this coefficient contrast is descriptive rather than a superiority test.

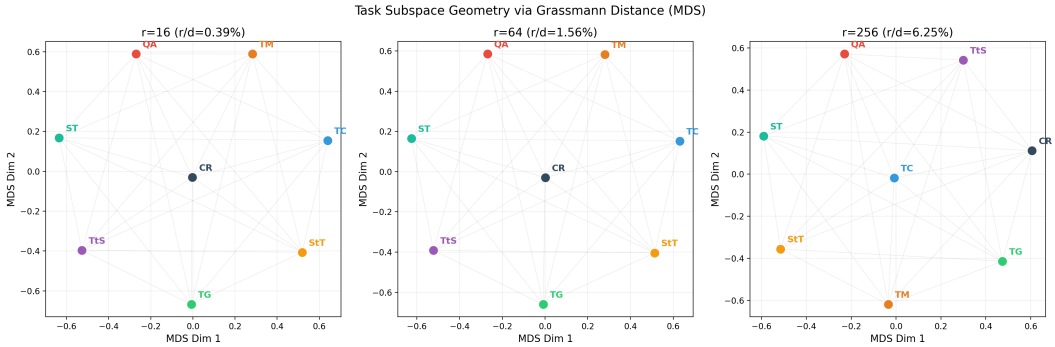

Figure 6: Descriptive MDS projections of pairwise Grassmann distances at ranks 16, 64, and 256. Axes are arbitrary up to rotation and reflection.

Table 11: Two-sided composition utility in percentage points. Cells report mean ± sample SD [95% pair-bootstrap CI]. Panel A uses the primary global $c = 0.5$ protocol. Panel B compares all protocols on the common ranks $24, 96, 512$; values are first averaged over ranks within each pair. $+U$ counts positive joint-utility conditions, and Bal. additionally requires non-negative target retention.

Panel	Configuration	Target retention	Source acquisition	Joint utility U	$+U$	Bal.
A	$r = 24$	-0.22 ± 3.17 [$-2.79, 2.34$]	6.25 ± 6.45 [0.69, 11.81]	3.01 ± 1.76 [1.51, 4.51]	4/4	2/4
	$r = 32$	-1.49 ± 2.68 [$-3.44, 1.17$]	7.52 ± 6.11 [2.57, 12.47]	3.02 ± 2.72 [0.40, 4.75]	3/4	1/4
	$r = 48$	-2.34 ± 1.92 [$-3.91, -0.77$]	8.71 ± 9.69 [0.96, 16.60]	3.19 ± 4.27 [$-0.26, 6.63$]	3/4	0/4
	$r = 64$	0.78 ± 1.92 [$-0.79, 2.34$]	7.10 ± 6.08 [2.10, 11.78]	3.94 ± 3.68 [0.65, 7.03]	3/4	2/4
	$r = 96$	-2.53 ± 2.74 [$-4.69, -0.37$]	4.92 ± 9.53 [-1.31, 14.22]	1.20 ± 4.06 [-1.90, 4.57]	2/4	1/4
	$r = 192$	0.75 ± 1.97 [$-1.08, 2.54$]	6.25 ± 10.19 [-0.39, 15.82]	3.50 ± 5.01 [0.39, 8.58]	4/4	3/4
	$r = 512$	1.81 ± 2.48 [0.25, 4.30]	5.19 ± 12.76 [-5.72, 17.25]	3.50 ± 5.57 [-0.81, 9.04]	3/4	3/4
B	Global $c = 0.25$	0.13 ± 0.46 [$-0.26, 0.53$]	4.74 ± 5.12 [1.30, 9.90]	2.44 ± 2.42 [0.91, 4.82]	11/12	6/12
	Global $c = 0.50$	-0.31 ± 1.95 [$-1.95, 1.17$]	5.45 ± 8.66 [-1.02, 14.00]	2.57 ± 3.41 [0.04, 5.96]	9/12	6/12
	Global $c = 1.00$	-2.16 ± 4.73 [$-6.15, 1.82$]	1.72 ± 12.97 [-8.77, 13.15]	-0.22 ± 5.76 [-5.06, 4.62]	7/12	3/12
	Module $c = 0.50$	0.15 ± 1.60 [$-1.24, 1.30$]	5.56 ± 8.50 [-0.63, 14.00]	2.85 ± 3.49 [0.23, 6.34]	9/12	7/12

D.1 Rank-selection study

Confirmatory design. The frozen design uses one prespecified target–source direction for each of the 28 unordered SuperNI pairs, candidate ranks $24, 32, 48, 64, 96, 192$, and 512, and the `global_add_050` composition. Rules, endpoints, the one-point SESOI, and the analysis script are fixed before utility inspection. Appendix B records the execution provenance, reproduction gate, hardware changes, batch-size fallbacks, and artifact hashes.

The earlier four-pair pilot used the same frozen statistics and composition endpoint. Its nominal-64 mean utility was 3.94, maximum-residual utility was 2.31, raw-overlap utility was 3.70, update-cosine utility was 3.70 (the two rules make identical selections on the pilot pairs), and the descriptive oracle was 5.44. These values remain as pilot context; the complete 28-pair study below is the confirmatory result.

The confirmatory enumeration evaluates all 28 directed task pairs at seven ranks. Rules are computed from stored adapter factors before utility is read. Table 14 lists the selected rank and joint utility for every rule; the oracle is downstream and descriptive.

The corrected max-δ rule is below the fixed default with a large negative paired effect. The min-raw-overlap rule has the same direction. The min-$|\cos \Delta W|$ rule is the only active rule with a positive mean difference, but its 28/28 collapse to rank 512 is identical to max raw overlap and therefore carries no pair-specific selection content. The rank-wise Spearman means are $+0.289$ for raw overlap, -0.293 for δ, -0.306 for $|\cos \Delta W|$, and $+0.306$ for nominal rank. The result establishes the prespecified behavioral boundary under `global_add_050`; it does not rank merger algorithms or identify a universal composition predictor.

Table 12: Condition-level utility for the primary global $c = 0.5$ composition. Each cell is a percentage-point mean ± sample SD with a paired-example bootstrap 95% CI shown on the second line. Values are changes from the same-rank target-only model.

Target ← source	Rank	Target retention	Source acquisition	Joint utility
D2 ← D5	24	−3.91±23.15 [−7.81, 0.00]	12.50±43.47 [5.47, 20.31]	4.30±35.72 [0.00, 8.59]
	32	−3.91±26.33 [−8.59, 0.78]	14.06±41.12 [7.03, 21.09]	5.08±35.61 [0.78, 9.38]
	48	−3.91±29.17 [−9.38, 0.78]	20.31±49.18 [11.72, 28.91]	8.20±42.14 [3.12, 13.28]
	64	3.12±24.90 [−0.78, 7.81]	13.28±44.14 [5.47, 21.09]	8.20±36.12 [3.91, 12.50]
	96	−5.47±22.83 [−9.38, −1.56]	18.75±43.02 [11.72, 26.56]	6.64±36.45 [2.34, 10.94]
	192	0.78±15.35 [−1.56, 3.12]	21.09±48.04 [13.28, 29.69]	10.94±37.01 [6.64, 15.23]
	512	0.78±23.46 [−3.12, 4.69]	21.88±48.50 [13.28, 30.47]	11.33±39.46 [6.64, 16.02]
D4 ← D7	24	−1.67±17.22 [−4.77, 1.20]	11.12±29.47 [6.17, 16.44]	4.72±24.93 [1.81, 7.71]
	32	−2.04±21.24 [−5.80, 1.64]	10.87±34.64 [5.05, 16.96]	4.41±29.40 [0.89, 7.98]
	48	−1.53±12.12 [−3.77, 0.38]	11.66±33.83 [5.94, 17.70]	5.06±26.21 [2.04, 8.18]
	64	−0.80±13.83 [−3.24, 1.60]	10.29±38.55 [3.72, 17.03]	4.74±29.43 [1.18, 8.36]
	96	−1.52±15.20 [−4.22, 1.10]	−2.79±43.99 [−10.55, 4.91]	−2.16±32.85 [−6.17, 1.87]
	192	−1.70±11.68 [−3.95, 0.10]	4.69±39.79 [−2.19, 11.49]	1.49±29.44 [−2.08, 5.08]
	512	1.00±10.42 [−0.62, 2.94]	3.37±44.94 [−4.45, 11.28]	2.18±32.58 [−1.92, 6.21]
D6 ← D1	24	1.56±35.46 [−4.69, 7.81]	0.78±29.42 [−3.91, 6.25]	1.17±32.52 [−2.73, 5.08]
	32	2.34±26.52 [−2.34, 7.03]	4.69±30.38 [−0.78, 10.16]	3.52±28.48 [0.00, 7.03]
	48	−3.91±29.17 [−9.38, 0.78]	5.47±26.05 [1.56, 10.16]	0.78±27.99 [−2.34, 4.30]
	64	1.56±30.70 [−3.91, 7.03]	5.47±31.52 [0.00, 10.94]	3.52±31.11 [−0.39, 7.42]
	96	0.78±26.61 [−3.91, 5.47]	3.12±24.90 [−0.78, 7.81]	1.95±25.75 [−1.17, 5.08]
	192	3.12±21.51 [0.00, 7.03]	0.00±17.75 [−3.12, 3.12]	1.56±19.74 [−0.78, 3.91]
	512	0.00±17.75 [−3.12, 3.12]	4.69±27.66 [0.00, 9.38]	2.34±23.31 [−0.39, 5.47]
D8 ← D3	24	3.12±24.90 [−0.78, 7.81]	0.59±11.35 [−1.39, 2.52]	1.86±19.36 [−0.43, 4.31]
	32	−2.34±26.52 [−7.03, 2.34]	0.46±12.74 [−1.66, 2.66]	−0.94±20.81 [−3.51, 1.57]
	48	0.00±12.55 [−2.34, 2.34]	−2.61±14.78 [−5.29, −0.21]	−1.30±13.74 [−3.01, 0.36]
	64	−0.78±31.98 [−6.25, 4.69]	−0.63±13.71 [−2.96, 1.73]	−0.71±24.56 [−3.72, 2.32]
	96	−3.91±23.15 [−7.81, 0.00]	0.61±12.28 [−1.66, 2.65]	−1.65±18.63 [−4.00, 0.50]
	192	0.78±19.83 [−2.34, 4.69]	−0.78±15.65 [−3.51, 1.90]	0.00±17.84 [−2.12, 2.23]
	512	5.47±22.83 [1.56, 9.38]	−9.18±17.64 [−12.32, −6.25]	−1.86±21.64 [−4.30, 0.72]

Table 13: Absolute scores and constrained utility on matched ranks 24, 96, and 512. Panel A reports pair-cluster means ± sample SD [95% bootstrap CI] over the four task-disjoint pairs, averaged over ranks within each pair. Panel B reports the eligible source gap, recovery ratio, and two constrained counts.

Panel A: Absolute scores

Protocol	T_t	C_t	T_s	C_s	S_s
	64.52±20.29	64.20±18.84	38.69±14.67	44.14±18.39	43.69±18.35
global norm, 0.50	[43.94, 77.86]	[45.25, 75.60]	[27.90, 52.88]	[29.33, 58.95]	[28.87, 58.50]
	64.52±20.29	64.65±20.01	38.69±14.67	43.43±15.98	43.69±18.35
global norm, 0.25	[43.94, 77.86]	[44.47, 77.60]	[27.90, 52.88]	[31.00, 55.99]	[28.87, 58.50]
	64.52±20.29	62.35±18.53	38.69±14.67	40.41±20.99	43.69±18.35
global norm, 1.00	[43.94, 77.86]	[44.73, 74.09]	[27.90, 52.88]	[22.77, 56.23]	[28.87, 58.50]
	64.52±20.29	64.67±19.33	38.69±14.67	44.25±18.11	43.69±18.35
module norm, 0.50	[43.94, 77.86]	[45.18, 76.38]	[27.90, 52.88]	[29.62, 58.87]	[28.87, 58.50]

Panel B: Constrained utility

Protocol	Gap	Rec.	Pareto / τ=2
global norm, 0.50	9/12	0.82	6/12 / 7/12
global norm, 0.25	9/12	0.60	5/12 / 9/12
global norm, 1.00	9/12	0.68	3/12 / 3/12
module norm, 0.50	9/12	0.82	7/12 / 7/12

Normalized-output agreement and character similarity are retained as secondary measures of target-side behavioral change. They do not replace target or source task scores. Table 17 preserves the 28 previously measured conditions.

The residual has stronger absolute association with both output-change endpoints for three pairs, with pair-dependent signs. Rank correlations remain descriptive because each trajectory contains

Table 14: Prespecified rank-selection rules under global $c = 0.5$ across all 28 directed task pairs. Each cell is selected rank / joint utility in percentage points. Nominal-64 is the fixed default; Oracle uses downstream utility and is descriptive.

Pair	Nominal-64	Max. raw o	Min. raw o	Max. δ	Min. rank	Min. $\lvert\cos \Delta W\rvert$	Oracle
D2 ← D5	64 / 8.20	512 / 11.33	48 / 8.20	24 / 4.30	24 / 4.30	512 / 11.33	512 / 11.33
D4 ← D7	64 / 4.74	512 / 2.18	32 / 4.41	24 / 4.72	24 / 4.72	512 / 2.18	48 / 5.06
D6 ← D1	64 / 3.52	512 / 2.34	64 / 3.52	24 / 1.17	24 / 1.17	512 / 2.34	32 / 3.52
D8 ← D3	64 / −0.71	512 / −1.86	24 / 1.86	32 / −0.94	24 / 1.86	512 / −1.86	24 / 1.86
D2 ← D1	64 / 2.73	512 / 3.52	64 / 2.73	24 / 0.78	24 / 0.78	512 / 3.52	512 / 3.52
D1 ← D3	64 / −0.63	512 / 1.10	64 / −0.63	24 / 0.72	24 / 0.72	512 / 1.10	32 / 1.81
D1 ← D4	64 / 10.02	512 / 19.82	32 / 9.97	24 / 8.90	24 / 8.90	512 / 19.82	512 / 19.82
D5 ← D1	64 / 1.17	512 / 3.13	48 / −1.17	24 / −1.17	24 / −1.17	512 / 3.13	512 / 3.13
D7 ← D1	64 / −4.32	512 / 0.88	64 / −4.32	24 / −2.78	24 / −2.78	512 / 0.88	512 / 0.88
D8 ← D1	64 / 1.95	512 / 3.13	48 / 0.39	32 / −0.78	24 / 1.95	512 / 3.13	512 / 3.13
D3 ← D2	64 / 4.01	512 / 9.78	64 / 4.01	24 / −0.85	24 / −0.85	512 / 9.78	512 / 9.78
D2 ← D4	64 / 17.12	512 / 23.39	24 / 11.34	24 / 11.34	24 / 11.34	512 / 23.39	512 / 23.39
D2 ← D6	64 / 9.38	512 / 12.89	64 / 9.38	24 / 5.47	24 / 5.47	512 / 12.89	512 / 12.89
D7 ← D2	64 / −0.85	512 / 6.90	64 / −0.85	24 / 2.69	24 / 2.69	512 / 6.90	512 / 6.90
D2 ← D8	64 / 0.39	512 / 0.39	48 / −2.73	32 / −1.17	24 / −3.91	512 / 0.39	64 / 0.39
D3 ← D4	64 / 12.00	512 / 19.37	32 / 14.10	24 / 9.33	24 / 9.33	512 / 19.37	512 / 19.37
D5 ← D3	64 / 0.93	512 / 2.04	48 / 0.16	24 / 0.38	24 / 0.38	512 / 2.04	192 / 4.05
D6 ← D3	64 / 0.55	512 / 2.90	64 / 0.55	24 / 2.52	24 / 2.52	512 / 2.90	512 / 2.90
D7 ← D3	64 / 0.96	512 / −2.51	64 / 0.96	24 / 1.93	24 / 1.93	512 / −2.51	24 / 1.93
D4 ← D5	64 / 7.97	512 / 10.61	24 / 4.88	24 / 4.88	24 / 4.88	512 / 10.61	512 / 10.61
D4 ← D6	64 / 7.30	512 / 8.04	32 / 5.07	24 / 6.15	24 / 6.15	512 / 8.04	512 / 8.04
D8 ← D4	64 / 16.46	512 / 19.37	24 / 13.04	32 / 19.85	24 / 13.04	512 / 19.37	32 / 19.85
D6 ← D5	64 / 8.98	512 / 14.06	32 / 12.50	24 / 5.86	24 / 5.86	512 / 14.06	512 / 14.06
D5 ← D7	64 / 4.26	512 / −2.56	48 / 1.38	24 / 3.61	24 / 3.61	512 / −2.56	32 / 6.10
D8 ← D5	64 / 8.98	512 / 11.33	24 / 2.73	32 / 6.25	24 / 2.73	512 / 11.33	512 / 11.33
D6 ← D7	64 / −0.37	512 / −0.78	64 / −0.37	24 / 2.15	24 / 2.15	512 / −0.78	192 / 5.02
D6 ← D8	64 / 4.69	512 / 1.56	24 / −0.39	32 / −1.17	24 / −0.39	512 / 1.56	64 / 4.69
D7 ← D8	64 / 1.53	512 / −3.64	24 / 3.51	32 / 0.07	24 / 3.51	512 / −3.64	24 / 3.51

Table 15: Confirmatory rank-selection summary over 28 task pairs. Utility, paired Δ from nominal rank 64, and oracle regret are reported as mean $\pm$ sample SD [95% pair-bootstrap CI]. Cliff's δ is the paired-difference dominance effect; W/L/T gives strict wins, losses, and ties; Holm p is reported for the prespecified active hypotheses. Oracle regret is $U_{\text{oracle}} - U_{\text{rule}}$.

Rule	U mean $\pm$ SD [95% CI]	Δ mean $\pm$ SD [95% CI]	Cliff δ	W/L/T	Holm p	Regret mean $\pm$ SD [95% CI]
Nominal-64	4.678±5.297 [2.792, 6.595]	–	–	–	–	3.138±2.565 [2.242, 4.097]
Max raw o	6.382±7.650 [3.745, 9.301]	1.703±3.936 [0.287, 3.119]	0.393	19/8/1	–	1.434±2.435 [0.607, 2.377]
Min raw o	3.723±5.020 [1.981, 5.601]	−0.955±2.380 [−1.842, −0.118]	−0.321	4/13/11	1.000	4.093±3.038 [3.022, 5.238]
Max δ	3.364±4.828 [1.717, 5.279]	−1.314±2.561 [−2.232, −0.370]	−0.500	7/21/0	1.000	4.452±3.448 [3.238, 5.735]
Min rank	3.246±3.973 [1.834, 4.720]	−1.432±2.776 [−2.435, −0.424]	−0.393	8/19/1	–	4.570±3.650 [3.254, 5.949]
Min $\lvert\cos \Delta W\rvert$	6.382±7.650 [3.689, 9.297]	1.703±3.936 [0.278, 3.123]	0.393	19/8/1	0.078	1.434±2.435 [0.623, 2.401]
Oracle	7.816±6.505 [5.543, 10.290]	3.138±2.565 [2.247, 4.081]	0.893	25/0/3	–	–

seven shared rank points. Word overlap, length drift, empty-output rate, and repeated-trigram change remain in the stored analysis artifact as additional output-change measures. The utility analysis uses task scores on both constituent tasks as its decision endpoint.

D.2 Idealized Model of Composition Utility

The idealized model isolates why a pairwise projector statistic cannot carry all information needed by two-sided composition utility. Each task has a quadratic loss $L_t(\theta) = \frac{1}{2}(\theta_t^* - \theta)^\top H(\theta_t^* - \theta)$ around the pretrained point θ_0, a converged rank-r adapter writes inside its own subspace, $\Delta_t = \Pi_t r_t$ with task displacement $r_t = \theta_t^* - \theta_0$, and the merged model is $\theta_0 + \Delta_t + \gamma\Delta_s$ for source strength γ. Expanding the two task losses at the deployment anchor $\theta_0 + \Delta_t$ gives the endpoints used above:

$$R_t = \gamma \langle \Delta_s, (I - \Pi_t) r_t \rangle_H - \tfrac{\gamma^2}{2} \|\Delta_s\|_H^2, \tag{14}$$

$$A_s = \gamma \langle \Delta_s, r_s - \Delta_t \rangle_H - \tfrac{\gamma^2}{2} \|\Delta_s\|_H^2, \tag{15}$$

with $\langle u, v \rangle_H = u^\top H v$. Three properties separate utility from pairwise projector statistics:

Table 16: Selected-rank frequencies in the 28-pair confirmatory study. The two rules that select rank 512 in every pair have the same mean utility because they make identical selections.

Rule	$r = 24$	$r = 32$	$r = 48$	$r = 64$	$r = 96$	$r = 192$	$r = 512$
Nominal-64	0	0	0	28	0	0	0
Max raw o	0	0	0	0	0	0	28
Min raw o	7	5	6	10	0	0	0
Max δ	21	7	0	0	0	0	0
Min rank	28	0	0	0	0	0	0
Min $\lvert\cos \Delta W\rvert$	0	0	0	0	0	0	28
Oracle	3	4	1	2	0	2	16

Table 17: Secondary target-side output-change measures for the 28 primary conditions. Agreement is a proportion and character similarity uses normalized sequence matching; both are reported as mean ± sample SD [95% paired-example bootstrap CI]. Task-score Δ is a deterministic task-score point estimate because no per-example task-score vector was retained.

Target ← source	Rank	Task-score Δ	Agreement mean ± SD [95% CI]	Character mean ± SD [95% CI]
text-classification ← text-to-structure	24	-3.906	0.844 ± 0.365 $[0.781, 0.906]$	0.901 ± 0.282 $[0.850, 0.947]$
	32	-3.906	0.914 ± 0.281 $[0.859, 0.961]$	0.940 ± 0.223 $[0.898, 0.975]$
	48	-3.906	0.883 ± 0.323 $[0.828, 0.938]$	0.930 ± 0.227 $[0.888, 0.967]$
	64	+3.125	0.898 ± 0.303 $[0.844, 0.945]$	0.946 ± 0.198 $[0.909, 0.977]$
	96	-5.469	0.906 ± 0.293 $[0.852, 0.953]$	0.950 ± 0.192 $[0.913, 0.979]$
	192	+0.781	0.945 ± 0.228 $[0.906, 0.984]$	0.972 ± 0.134 $[0.946, 0.993]$
	512	+0.781	0.914 ± 0.281 $[0.867, 0.961]$	0.937 ± 0.221 $[0.896, 0.971]$
structure-to-text ← text modification	24	-1.675	0.758 ± 0.430 $[0.680, 0.828]$	0.884 ± 0.279 $[0.831, 0.930]$
	32	-2.042	0.781 ± 0.415 $[0.703, 0.852]$	0.916 ± 0.238 $[0.872, 0.954]$
	48	-1.535	0.828 ± 0.379 $[0.758, 0.891]$	0.956 ± 0.143 $[0.929, 0.978]$
	64	-0.804	0.867 ± 0.341 $[0.805, 0.922]$	0.963 ± 0.139 $[0.937, 0.985]$
	96	-1.522	0.875 ± 0.332 $[0.812, 0.930]$	0.959 ± 0.138 $[0.934, 0.980]$
	192	-1.698	0.859 ± 0.349 $[0.797, 0.914]$	0.961 ± 0.135 $[0.936, 0.983]$
	512	+0.996	0.891 ± 0.313 $[0.836, 0.945]$	0.958 ± 0.164 $[0.927, 0.982]$
sequence tagging ← question answering	24	+1.562	0.531 ± 0.501 $[0.445, 0.617]$	0.694 ± 0.376 $[0.629, 0.755]$
	32	+2.344	0.641 ± 0.482 $[0.562, 0.719]$	0.784 ± 0.345 $[0.722, 0.841]$
	48	-3.906	0.641 ± 0.482 $[0.555, 0.727]$	0.740 ± 0.383 $[0.675, 0.805]$
	64	+1.562	0.602 ± 0.492 $[0.516, 0.680]$	0.734 ± 0.371 $[0.669, 0.793]$
	96	+0.781	0.617 ± 0.488 $[0.531, 0.695]$	0.774 ± 0.344 $[0.714, 0.833]$
	192	+3.125	0.789 ± 0.410 $[0.719, 0.859]$	0.914 ± 0.190 $[0.881, 0.944]$
	512	+0.000	0.742 ± 0.439 $[0.664, 0.820]$	0.886 ± 0.226 $[0.846, 0.924]$
commonsense reasoning ← text generation	24	+3.125	0.547 ± 0.500 $[0.461, 0.633]$	0.739 ± 0.336 $[0.678, 0.796]$
	32	-2.344	0.438 ± 0.498 $[0.352, 0.523]$	0.700 ± 0.347 $[0.639, 0.759]$
	48	+0.000	0.453 ± 0.500 $[0.367, 0.539]$	0.672 ± 0.355 $[0.610, 0.732]$
	64	-0.781	0.320 ± 0.468 $[0.242, 0.398]$	0.559 ± 0.373 $[0.496, 0.622]$
	96	-3.906	0.539 ± 0.500 $[0.453, 0.625]$	0.718 ± 0.357 $[0.654, 0.779]$
	192	+0.781	0.469 ± 0.501 $[0.383, 0.555]$	0.669 ± 0.369 $[0.603, 0.732]$
	512	+5.469	0.266 ± 0.443 $[0.188, 0.344]$	0.480 ± 0.382 $[0.414, 0.543]$

1. **Residual orthogonality.** The benefit term of target retention is the alignment of Δ_s with $(I - \Pi_t)r_t$, the part of the target displacement left outside the target adapter. Pairwise subspace overlap measures the directions shared by the two adapters, which enter the cost and interference terms.
2. **Sign blindness.** The interference contribution $-\gamma\langle\Delta_s, \Delta_t\rangle_H$ inside source acquisition depends on the sign of alignment, while overlap is built from squared cosines: $\pm\Delta_s$ give identical geometry and mirrored utility.
3. **Dominance of update-magnitude terms.** The source's learnable signal $\|\Delta_s\|_H^2 = \|\Pi_s r_s\|_H^2$ grows as $O(r)$, while the geometric alignment terms grow as $O(\sqrt{r})$, so cross-pair utility variance is dominated by update magnitude rather than subspace geometry.

The planted model reproduces the measurement pattern: with a fixed set of shared directions and isotropic H, raw overlap rises from 0.23 to 0.50 while the Haar residual falls from 0.22 to 0.002. Figure 3 shows the resulting raw-overlap/Haar-residual reversal. The model also predicts a rank-monotone absolute update cosine, so the frozen minimum-$\lvert\cos \Delta W\rvert$ rule collapses to the largest candidate rank. The main text reports the corresponding confirmatory selection result; the present

Table 18: Exact-permutation Spearman results for the secondary output-change endpoints. The table contrasts raw overlap with the local-Haar residual. Parentheses contain two-sided exact p-values.

	Raw overlap o		Pair residual δ	
Target ← source	Agreement	Character	Agreement	Character
text-classification ← text-to-structure	0.775 (.053)	0.429 (.354)	−0.667 (.116)	−0.571 (.200)
structure-to-text ← text modification	0.786 (.048)	0.500 (.267)	−0.857 (.024)	−0.643 (.139)
sequence tagging ← question answering	0.571 (.200)	0.607 (.167)	−0.643 (.139)	−0.714 (.088)
commonsense reasoning ← text generation	−0.214 (.662)	−0.571 (.200)	0.250 (.595)	0.643 (.139)

model isolates the sign and update-magnitude terms that make this collapse possible. Individual pair trajectories remain heterogeneous.

E Additional Geometric Identities

E.1 Expected Haar Overlap

Proof of Proposition 4.1. Let P be the projector onto a Haar-random q-dimensional subspace of $\mathbb{R}^d$. Rotational invariance gives $\mathbb{E}[P] = cI_d$. Since $\operatorname{tr}(P) = q$, $c = q/d$. Conditioning on P_1 and averaging over the independent P_2 yields

$$\mathbb{E}[\operatorname{tr}(P_1P_2) \mid P_1] = \operatorname{tr}(P_1\mathbb{E}[P_2]) = \frac{q}{d}\operatorname{tr}(P_1) = \frac{q^2}{d}. \tag{16}$$

The singular values of $U_1^\top U_2$ give $\operatorname{tr}(P_1P_2) = \sum_k \cos^2\theta_k$. Dividing by q proves $\mathbb{E}[o_{12}] = q/d$. □

E.2 Haar Mean-Angle Reference

Let $h(x) = \arccos(\sqrt{x})$ and let $\theta_\star \in (\pi/4, \pi/2)$ solve $\tan\theta_\star = 2\theta_\star$. Define $x_\star = \cos^2\theta_\star \approx 0.1554$ and

$$g(x) = \begin{cases} h(x), & 0 \le x \le x_\star, \\ h(x_\star)(1-x)/(1-x_\star), & x_\star < x \le 1. \end{cases} \tag{17}$$

For independent Haar subspaces and $\bar\theta = q^{-1}\sum_k \theta_k$,

$$\mathbb{E}[\bar\theta] \ge g(q/d), \qquad \mathbb{E}[\bar\theta] \ge \arccos\sqrt{q/d} \text{ when } q/d \le x_\star. \tag{18}$$

Proof. The second derivative

$$h''(x) = \frac{1-2x}{4[x(1-x)]^{3/2}} \tag{19}$$

changes sign at $1/2$. The line from $(x_\star, h(x_\star))$ to $(1,0)$ is tangent to h at $x_\star$ and lies below h on $[x_\star, 1]$, so g is the lower convex envelope. With $X_k = \cos^2\theta_k$, Jensen's inequality gives

$$\frac{1}{q}\sum_k h(X_k) \ge \frac{1}{q}\sum_k g(X_k) \ge g\left(\frac{1}{q}\sum_k X_k\right). \tag{20}$$

Taking expectations and applying Jensen again, followed by the Haar trace identity, gives $\mathbb{E}[\bar\theta] \ge g(q/d)$. □

E.3 Unavoidable Intersection

Proposition E.1. *For q-dimensional subspaces $\mathcal{U}_1, \mathcal{U}_2 \subseteq \mathbb{R}^d$,*

$$\dim(\mathcal{U}_1 \cap \mathcal{U}_2) \ge 2q - d. \tag{21}$$

Hence $2q > d$ forces at least $2q - d$ zero principal angles.

Proof. The dimension identity gives

$$\dim(\mathcal{U}_1 \cap \mathcal{U}_2) = 2q - \dim(\mathcal{U}_1 + \mathcal{U}_2) \ge 2q - d. \tag{22}$$

A k-dimensional intersection contributes k zero principal angles. □

E.4 PROJECTION-COMMUTATOR IDENTITY

Theorem E.2. *For projectors P_1, P_2 onto two q-dimensional subspaces with principal angles $\{\theta_k\}_{k=1}^q$,*

$$\|[P_1, P_2]\|_F^2 = \frac{1}{2} \sum_{k=1}^{q} \sin^2(2\theta_k). \tag{23}$$

Proof. The commutator is skew-symmetric, so

$$\|[P_1, P_2]\|_F^2 = -\operatorname{tr}([P_1, P_2]^2) \tag{24}$$

$$= 2\operatorname{tr}(P_1 P_2) - 2\operatorname{tr}(P_1 P_2 P_1 P_2). \tag{25}$$

The compression $P_1 P_2 P_1$ has eigenvalues $\cos^2 \theta_k$. Substitution gives

$$2 \sum_k \cos^2 \theta_k (1 - \cos^2 \theta_k) = \frac{1}{2} \sum_k \sin^2(2\theta_k). \tag{26}$$

□