Manufactured Rank Is a Spectator: A Causal Test of Nonlinear Low-Rank Adapters in Task Adaptation

Peter Hollows

Abstract

A growing family of parameter-efficient adapters proposes to improve on LoRA by inserting an element-wise nonlinearity between the low-rank factors, *manufacturing* rank beyond the nominal ceiling without adding parameters. The premise is correlational: the nonlinearity provably raises the realized update's stable rank, and higher rank co-occurs with higher accuracy. We test it causally and find the manufactured rank is a *spectator*. A taxonomy first narrows the target: some methods advertised under "nonlinearity as rank" do not raise the realized rank at all (genLoRA proves $\mathrm{rank}(\Delta W) \leq r$ by its own construction), while those that genuinely fold the spectrum are sinusoidal (sine-LoRA, loran), and even there the extra rank is overwhelmingly low-energy. For these we introduce **Rank-Content Ablation (RCA)**, a within-checkpoint causal test that ablates the manufactured spectrum of a *trained* update while separating rank content from energy. Across 40 sine-LoRA checkpoints the manufactured rank is inert by a pre-registered load-bearing criterion: truncating back to the nominal rank is lossless (η=1.00), and the manufactured tail alone sits at the no-adapter floor. The verdict replicates on loran and across the full folding spectrum. From the training side, the first matched-per-arm-tuning evaluation finds no surviving advantage (a few tenths of a point at most, reproduced by zero-rank-lift controls, accuracy flat-to-decreasing in the rank-controlling frequency), on GLUE (DeBERTaV3) and commonsense (Llama-3.2-1B QLoRA) alike, where no adapter beats the no-adapter base. A rank-demanding probe confirms the test *can* fire (a genuine rank-64 adapter is catastrophic to truncate, η=0), while sine cannot use rank at all: it trains only where it manufactures no rank, and folds only where it cannot train. Our claim concerns nonlinear low-rank adapters *for task adaptation*; we neither test nor contest the signal-fitting regimes (e.g. implicit neural representations) where the construction originates and high-frequency structure may make rank genuinely load-bearing. Manufacturing rank with a nonlinearity inflates the spectrum, but on these tasks that rank is a spectator, out of the optimizer's reach.

1 Introduction

Low-rank adaptation (LoRA) [11] fine-tunes a frozen weight W_0 with an additive low-rank update $\Delta W = \frac{\alpha}{r} BA$, trading expressivity for a tiny parameter budget. A prominent line of recent work argues that the rank-r ceiling is the binding limitation and sets out to *manufacture* additional rank without adding parameters, by inserting an element-wise nonlinearity between the low-rank factors: $\Delta W = \gamma\, \phi(\omega BA)$. Sine-LoRA [13], which takes $\phi = \sin$, is the flagship instance and reports consistent gains at matched parameter count; the same "nonlinearity as rank" rationale appears in adapters using radial-basis and structured nonlinear constructions [5, 22]. The shared rationale is a theorem and a correlation: a sinusoid provably raises the rank of a low-rank matrix, and across configurations higher realized rank co-occurs with better downstream accuracy. This family is, however, heterogeneous (§2): some members genuinely fold the spectrum and raise the realized rank, while others reparameterize the factors and provably leave it at the nominal r. Our causal test concerns the former; for the latter the rank-lift premise is contradicted by the method's own construction.

This rationale is *correlational*. That a reparameterization both raises stable rank and improves accuracy does not establish that the former causes the latter: the nonlinearity simultaneously rescales the effective learning rate (near the origin $\sin(\omega x) \approx \omega x$), bounds the update ($|\sin| \leq 1$), and gates gradients by a data-dependent factor $\cos(\omega BA)$, any of which could account for an optimization-side benefit while the elevated rank rides along as a byproduct [15, 25]. No prior work has intervened on the realized rank to test whether it is load-bearing.

We close this gap with a causal test we call **Rank-Content Ablation (RCA)**: given a *trained* adapter, we operate directly on the singular value spectrum of its realized update and re-evaluate, separating the contribution of the rank the nonlinearity manufactured (the spectrum beyond the nominal rank r) from its energy and from the top-r subspace. Because RCA is applied post hoc within a single checkpoint, it asks a clean question that is independent of how large the method's headline effect is: *is the manufactured rank carrying task signal?*

Applied to sine-LoRA, the answer is no. Truncating each trained update back to its nominal rank r is lossless; a matched-energy control shows this reflects empty rank content rather than lost magnitude; the manufactured spectrum evaluated in isolation sits exactly at the no-adapter floor; and grafting that spectrum onto a genuine linear adapter is inert. The conclusion holds on every trained checkpoint, on FP16 GLUE adaptation of DeBERTaV3 and on 4-bit QLoRA adaptation of Llama-3.2-1B. It is reinforced from the training side: under per-arm-matched hyperparameters the published advantage over a well-tuned linear adapter does not replicate, its accuracy is flat-to-decreasing in the frequency that controls rank manufacture, and three zero-rank-lift controls that reproduce only the optimization-side channels match or exceed it. And no form of rank beats a well-tuned baseline in either setting: on the Llama tier even a genuine linear high-rank construction only matches the no-adapter base, while sine's manufactured rank degrades it.

Contributions.

1. **A taxonomy of the "nonlinearity as rank" family** that separates methods which genuinely raise the realized rank (sine-LoRA, loran) from those whose construction provably leaves it at the nominal r (the generative RBF adapter genLoRA, $\text{rank}(\Delta W) \leq r$), sharpening which methods a rank-mechanism claim even applies to.

2. **Rank-Content Ablation**, a within-checkpoint causal test that adjudicates whether the rank a nonlinear adapter manufactures is load-bearing, with a matched-energy control separating rank content from energy. The test is method-agnostic: it requires only a materialized, rank-inflated update.

3. **A causal refutation of the rank-lift mechanism across two published methods**: on sine-LoRA (40 trained checkpoints, the full folding spectrum) and replicated on loran, the manufactured rank is inert, against the pre-registered load-bearing criterion.

4. **The first matched-per-arm-tuning evaluation of sine-LoRA**, showing the headline gain does not survive and localizing any residual effect to optimization, with cross-architecture confirmation on a quantized 1B LM.

Our analysis is pre-registered; we report all arms, seeds, frequencies, and the one outcome (an underpowered equivalence test) the data could not resolve.

Table 1: What each construction does to the realized rank of a fixed trained product $M = BA$, means over the 24 adapted projections (DeBERTaV3, nominal r=4). We report two notions of rank: *numerical rank* (count of singular values exceeding an absolute tolerance 10^{-5}, threshold-sensitive) and energy-weighted *stable rank* $\rho = \|M\|_F^2/\sigma_1^2$. genLoRA leaves both at nominal by its own bound. Where sine and loran fold, the numerical rank inflates by up to 166× (4→665) while the stable rank stays near 1 at usable frequencies: the "manufactured rank" is almost entirely low-energy directions, foreshadowing their causal inertness.

construction	stable rank	numerical rank
genLoRA $\sum_i f_i^B f_i^{A\top}$ ($\leq r$ by Prop. 3.1)	1.10	4
sine $\gamma \sin(\omega M)$, ω=10	1.10	25
sine $\gamma \sin(\omega M)$, ω=100	1.32	127
sine $\gamma \sin(\omega M)$, ω=1000	17.19	665
loran $\gamma(\beta \sin(\omega M) \odot M + M)$, β=1, ω=100	1.71	89

2 Which adapters actually manufacture rank?

The "nonlinearity as rank" slogan promises more than the constructions deliver. Measuring what each method actually does to the realized update ΔW it bakes into the frozen weights (Table 1), the program splits into two ways the rank story fails: in some methods the realized rank is not raised at all, and in the rest it is raised only in energetically negligible directions.

The advertised rank is often not even raised. The generative RBF adapter genLoRA [22], named for the principle "nonlinearity as rank," synthesizes the columns of B and rows of A from latent vectors through nonlinear RBF generators f_i^B, f_i^A and forms $\Delta W = \sum_{i=1}^r f_i^B(Z_B)\, f_i^A(Z_A)^\top$, a sum of r rank-one terms. The nonlinearity acts on the *generation* of the factors, never on their product, and the authors themselves prove $\text{rank}(\Delta W) \leq r$ (their Prop. 3.1). The realized rank is exactly nominal (Table 1): the headline variable the method is named for is, by its own theorem, left untouched. We make no claim about *what* genLoRA's mechanism is; we claim only that the supra-nominal-rank account is *inapplicable to it by construction*, much as for a generative reparameterization like OP-LoRA [25]. No intervention is needed to see that realized rank cannot be its lever, because its own bound fixes that rank at nominal.

When the rank is raised, it is almost all noise. Sine-LoRA and loran do fold the spectrum, applying an element-wise nonlinearity directly to the rank-r product $M = BA$:

$$\Delta W_{\text{sine}} = \gamma \sin(\omega M), \qquad \Delta W_{\text{loran}} = \gamma\big(\beta \sin(\omega M) \odot M + M\big),$$

(loran's "Sinter" map, amplitude β its "A", reducing to plain LoRA as $\beta \to 0$ [5]). Here the realized rank genuinely exceeds r, but Table 1 exposes the catch. The *numerical* rank explodes (sine reaches 665 of a possible 768), while the energy-weighted *stable* rank stays near nominal at every frequency that trains well (1.10→1.32 as ω goes 10→100), rising materially only at extreme ω=1000 (17.2), a regime that *degrades* accuracy (§8). The "manufactured rank" is thus overwhelmingly low-energy ripple: a wall of tiny singular values that inflates any rank count but carries almost no Frobenius mass. Whether that ripple carries task signal is what a rank–accuracy correlation cannot decide and what our causal test (§3) settles.

In short, the rank-mechanism premise is unsupported on both branches before any fine-tuning is even run: where the rank is raised it is energetically negligible, and where it is claimed loudest (genLoRA) it is not raised at all. The rest of the paper converts "energetically negligible" into "causally inert," on sine-LoRA (§5, §8) and loran (§6).

3 A causal test for manufactured rank

3.1 Setup: nominal vs. realized rank

We consider adapters that, for a frozen $W_0 \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$, learn an additive update ΔW through a low-rank parameterization. A standard LoRA update $\Delta W = \frac{\alpha}{r} BA$ with $B \in \mathbb{R}^{d_{\text{out}} \times r}$, $A \in \mathbb{R}^{r \times d_{\text{in}}}$ has rank at most the *nominal* rank r. A growing design line instead inserts a nonlinearity ϕ between the factors, $\Delta W = \gamma\, \phi(\omega BA)$, so the *realized* update has stable rank far exceeding r. Writing the SVD $\Delta W = U \operatorname{diag}(\sigma)\, V^\top$ with $\sigma_1 \geq \sigma_2 \geq \cdots$, the realized stable rank $\rho(\Delta W) = \left(\sum_i \sigma_i^2\right)/\sigma_1^2$ can be many times r. These methods are motivated by a correlation: ϕ raises ρ, and across hyperparameters higher ρ co-occurs with better performance. The inference drawn, that the manufactured rank (σ_i for $i > r$) improves adaptation, has never been tested by intervention. We ask, of a *single trained adapter*, whether the realized rank beyond r is load-bearing or a byproduct.

3.2 Rank-Content Ablation

We call the test **Rank-Content Ablation (RCA)**: a battery of spectral interventions on a trained adapter that separates the *rank content* of the realized update from its energy/norm. We materialize ΔW, compute its SVD, and construct modified updates $\widetilde{\Delta W}$. Each is baked into the frozen weights $(W_0 + \widetilde{\Delta W})$, the trained task head is restored unchanged, and the model is re-evaluated on held-out data. Let $U_{1:r}, V_{1:r}$ denote the leading r singular subspaces and $P = \sum_{i \leq r} \sigma_i u_i v_i^\top$, $T = \sum_{i > r} \sigma_i u_i v_i^\top$ the top-r and tail components ($\Delta W = P + T$).

I1, truncation to nominal rank. $\widetilde{\Delta W} = P$. If the manufactured rank is load-bearing, discarding T should degrade performance. Because T also carries energy, we report a norm-rescaled variant $\widetilde{\Delta W} = P \cdot \|\Delta W\|_F / \|P\|_F$, so a drop cannot be attributed to lost magnitude alone.

I1b, matched-energy random-direction removal. A truncation result could reflect lost *energy* rather than lost *rank content*. We remove an equal Frobenius mass from *random* directions while preserving the exact top-r subspace: draw N Gaussian, project out the top-r row and column spaces ($N \leftarrow N - U_{1:r} U_{1:r}^\top N - N V_{1:r} V_{1:r}^\top$), rescale to $\|N\|_F = \|T\|_F$, and set $\widetilde{\Delta W} = P + T - N$. The contrast between I1 and I1b isolates rank content from energy. Reported over several noise seeds.

I3, tail-only. $\widetilde{\Delta W} = T$. The manufactured rank in isolation: if it encodes task-relevant structure it beats the no-adapter baseline; if it is a byproduct it sits at the floor.

I2, tail-spectrum placebo. For a paired plain-LoRA checkpoint with update ΔW^{lin}, we add a random matrix with the sine adapter's tail singular values $\{\sigma_i\}_{i > r}$ but random orthonormal U, V, and re-evaluate. If the elevated spectrum were intrinsically useful, injecting it should help. (This control supplies evidence *for* the null only and is excluded from the affirmative criterion.)

Table 2: Pre-registered RCA signatures in retained gain fraction η, with the declared decision bands. The I1b criterion is the *gap* to truncation, $\Delta{=}\eta_{\text{I1b}}{-}\eta_{\text{I1}}$ (a positive gap means removing the manufactured directions hurts more than removing equal random energy, i.e. content matters); the I3 criterion is a ± 0.05 band around the floor rather than a point at 0, so noise-level positives do not count as signal.

intervention	rank is **load-bearing**	rank is a **byproduct**
I1 truncate-to-r (raw & rescaled)	$\eta \leq 0.5$	$\eta \geq 0.9$
I1b matched-energy random removal	$\Delta > 0.1$ (content)	$\|\Delta\| \leq 0.1$ (energy only)
I3 tail-only	$\eta > 0.05$ (signal)	$\|\eta\| \leq 0.05$ (floor)
I2 tail placebo	—	≤ 0.1 pt change

3.3 Evaluation and decision rule

All interventions are evaluated on a held-out split the adapter did not train on, with the trained head fixed; only ΔW is manipulated. This held-out split is disjoint from the carved split used for hyperparameter selection, so no intervention is evaluated on data that informed tuning. We report the *retained gain fraction* $\eta = \big(m(\widetilde{\Delta W}) - m_0\big)/\big(m(\Delta W) - m_0\big)$, where m_0 is the no-adapter score: $\eta = 1$ is lossless, $\eta = 0$ is the base floor. The denominator is the full adaptation gain over the frozen base, typically tens of points (e.g. 18.2→66.1 MCC on CoLA) rather than a marginal edge, so η is a ratio of large quantities; the rare checkpoint with near-zero over-base gain is excluded. The two hypotheses give mutually exclusive, pre-registered signatures (Table 2). The truncation result is the primary contrast; the matched-energy control adjudicates the energy-vs-content confound. The test is method-agnostic: it applies to any adapter producing a materialized ΔW whose realized stable rank exceeds its nominal rank.

Why within-checkpoint. Arm-level comparisons confound the mechanism question with tuning and seed variance and presuppose a sizable headline effect to dissect. RCA instead operates post hoc on each trained adapter, so its conclusion is robust to whether the method's headline gain is large, small, or absent, a property we exploit in §8.

What RCA identifies (and what it does not). RCA's estimand is the *inference-time* marginal contribution of the manufactured tail subspace to task performance, holding fixed everything training produced (the factors, the head, the top-r subspace). It establishes that the supra-nominal rank in the *converged* update carries no task signal. It does *not*, on its own, rule out that the manufactured rank played a transient role *during* optimization (e.g. as a scaffold that shapes the trajectory and is then safely discardable at convergence). We address that training-dynamics alternative not with a single experiment but with a conjunction: the straight-through cos-gate control isolates sine's training-time gradient channel with zero forward rank (§8), the frequency dose-response shows accuracy falling as trained rank rises, and the matched-tuning null shows there is no gain for any such scaffold to have produced. "Causal" throughout refers to this converged-update interventional sense.

Samples and exclusions. Table 3 reconciles every sample size in the paper to its design. The sole exclusion rule is pre-registered and mechanical: η is undefined when a checkpoint's over-base gain $m(\Delta W){-}m_0 \leq 10^{-6}$ (the adapter did not move the metric), which drops two MRPC $\omega{=}1$

Table 3: Samples and exclusions across all analyses. η-undefined exclusions apply only where noted (over-base gain $\leq 10^{-6}$). ‡Post-registration additions (memorization, calibration, and the loran replication) are exploratory and were not part of the pre-registered plan.

analysis	unit	n	design (pre-exclusion)
RCA primary (sine)	checkpoint	40	4 tasks $\times$ {headline ω, ω=400} $\times$ 5 seeds
RCA full sweep (sine)	checkpoint	138	4 tasks $\times$ 7 ω $\times$ 5 seeds = 140, -2 undefined
RCA replication (loran)	checkpoint	12	4 tasks $\times$ 3 seeds
RCA positive control (MoRA, GLUE)	checkpoint	20	4 tasks $\times$ 5 seeds
Memorization: recall vs rank‡	run	$\sim$30	$r \in \{4, 16, 64, 256\}$ over $N \in \{10^3..10^4\}$; r4/r64@N=4096 $\times$3 seeds
Memorization: sine ω-sweep‡ (Fig. 2)	run	29	sine r=4, N=1024, 13 freqs $\omega \in [1, 2000]$, 3 seeds through boundary
Memorization RCA pos. control‡	run	3	r64@N=4096, truncate→r4 (η=0)
Calibration probe‡ (NLL/ECE)	checkpoint	6	sine, 3 classification tasks $\times$ 2 seeds
I2 tail-graft placebo	pair	20	4 tasks $\times$ 5 seeds (sine→plain)
Matched-tuning CR2 (Tab. 6)	run	440	sine/controls/plain-r4 @ 20 seeds; plain-r8/MoRA @ 5
Matched GLUE means (Tab. 5)	cell	5	seeds per (arm, task) cell
Frequency dose-response (Fig. 3)	run	15	sine, dev split, per ω
Llama confirmation (Tab. 7)	run	24	8 arm-conditions $\times$ 3 seeds

checkpoints from the full sweep; all denominators are post-exclusion. No run is dropped for its outcome.

4 Case study: sine-LoRA, premise and scope

RCA is only meaningful on a checkpoint whose realized update exceeds its nominal rank; we verify this premise, which also reveals a boundary on when the sine mechanism engages.

The premise holds on DeBERTaV3-base [10]. Trained at the method's published GLUE recipe, the sine adapter enters its folding regime: ωBA reaches 2.9π in magnitude (well past the $\pi/2$ at which sin becomes non-monotone), the realized update carries $\sim$2% of its Frobenius energy beyond the nominal rank, and stable rank reaches 1.8$\times$ that of plain LoRA. There is genuine supra-nominal rank to interrogate; §5 shows it is inert.

A scope condition: dormant on RoBERTa-base [20]. The same recipe on RoBERTa-base never reaches the folding regime. At every stable learning rate the trained argument stays small ($|\omega BA| \leq 0.29\pi$, 99.9% of entries below 0.14π), so $\sin(\omega BA) \approx \omega BA$ to within a few percent and stable rank is only 1.16$\times$ plain LoRA, failing the pre-registered premise threshold. On this backbone the sine adapter is, to numerical precision, a linearly-rescaled LoRA: no manufactured rank, because the nonlinearity operates in its linear regime; raising ω to force folding destabilizes training first. This boundary is a scope condition on the method (manufacturing rank requires driving ωBA into the nonlinear regime, which the optimizer does not always reach) and dictates that the causal test be conducted on DeBERTaV3-base.[1]

[1]The original sine-LoRA GLUE backbone is unidentifiable from the paper: it is named "RoBERTa V3" (not a standard model), the accompanying citation points to Sentence-BERT [23], and neither RoBERTa nor DeBERTa is referenced anywhere in the paper. The reported GLUE scores instead fingerprint a DeBERTaV3-base recipe (e.g.

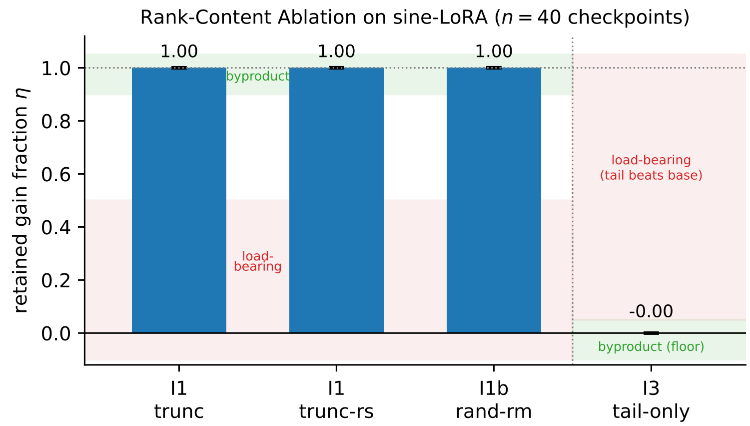

Figure 1: Rank-Content Ablation on trained sine-LoRA adapters (n=40, held-out dev). Truncating to the nominal rank and the matched-energy random removal are lossless (η=1); the manufactured tail alone is at the no-adapter floor (η=0). Shaded bands are the pre-registered decision regions (green: byproduct, red: load-bearing); the rule *inverts* for I3 tail-only (byproduct is at the floor, load-bearing would beat base), as marked by the divider. All four interventions land in the byproduct region.

5 The manufactured rank is causally inert

We apply RCA to every trained sine-LoRA checkpoint on the primary tier (the per-task headline frequency and the high-frequency ω=400 condition, four GLUE tasks, five seeds: 40 adapters), evaluated on held-out development data.

Result. Figure 1 and Table 4 report η pooled over the 40 checkpoints. Truncating each trained update back to its nominal rank is *lossless* ($\eta = 1.000 \pm 0.002$), and remains lossless after norm rescaling (1.000 ± 0.002): the supra-nominal rank can be deleted without measurable cost. The matched-energy random removal removes the *same* Frobenius mass from content-free directions and is likewise lossless (1.000 ± 0.001), so the truncation result reflects the **absence of rank content, not of energy**: there is no energy confound to explain away because removing equal energy at random costs nothing either. The manufactured tail in isolation collapses to the no-adapter floor ($\eta = -0.000 \pm 0.001$). Grafting the sine tail spectrum onto a trained plain-LoRA update (I2) moves the metric by -0.001 points on average (max $|\Delta| = 0.099$ over 20 paired checkpoints).

The signature is unanimous: per-checkpoint truncation η never falls below 0.995, and tail-only η never exceeds 0.001. A representative case (CoLA, seed 13): the intact adapter scores 66.1 MCC, truncation to rank r scores 66.1, and the tail alone scores 18.2, identical to the frozen base's 18.2.

Robust across the folding spectrum. The result is not an artifact of weakly-folded adapters. Extending RCA to the full frequency sweep (n=138, after excluding two zero-gain MRPC ω=1 checkpoints for which η is undefined; $\omega \in \{1, \dots, 1000\}$) spans trained stable rank from 1.18 to 1.75, and the signature is unchanged: the per-frequency mean truncation η stays ≥ 0.999 and mean tail-only $\eta \leq 0.004$ at every ω, and at ω=1000 (where the manufactured rank is largest) they are

CoLA ≈68.6, a DeBERTaV3 number, versus ≈63 for RoBERTa-base), and no GLUE code is released. We therefore adopt DeBERTaV3-base, and note the published recipe fails to train on actual RoBERTa-base.

Table 4: Rank-Content Ablation on trained sine-LoRA adapters (DeBERTaV3-base, held-out dev, n=40). η = fraction of the adapter's over-base gain retained.

intervention	η (mean $\pm$ sd)	reading
I1 truncate-to-r (raw)	1.000 ± 0.002	manufactured rank removable at no cost
I1 truncate-to-r (norm-rescaled)	1.000 ± 0.002	not a magnitude effect
I1b matched-energy random removal	1.000 ± 0.001	not an energy effect; content empty
I3 tail-only (r+1..d)	-0.000 ± 0.001	manufactured rank carries no signal
I2 tail-spectrum placebo (pt)	-0.001 (max 0.099)	elevated spectrum inert on a real adapter

0.9997 and 0.0006. More manufactured rank does not make any of it load-bearing.

Against the pre-registered criteria. The load-bearing hypothesis required truncation to destroy $\geq$ 50% of the gain, a positive content gap (Δ>0.1), and the isolated tail above the floor band (η>0.05); the byproduct hypothesis required near-lossless truncation ($\eta \geq 0.9$), random and content removal to agree ($|\Delta| \leq 0.1$), and the tail within the floor band ($|\eta| \leq 0.05$). Every value falls on the byproduct side, none close to the load-bearing threshold: truncation destroys 0% where the mechanism account required $\geq$ 50%, the I1/I1b gap is Δ=0.000, and the tail sits at η=−0.000. The manufactured rank is an epiphenomenon of the nonlinear parameterization. Because RCA is paired within each checkpoint, this holds regardless of the size (or existence) of any headline advantage (§8).

6 The verdict replicates on a second method: loran

If the inertness were an artifact of the sine activation specifically, it need not transfer to another rank-folding construction. It does. We apply RCA to loran's Sinter adapter (§2), the second published method that genuinely raises realized rank, trained on the four GLUE tasks at a premise-satisfying setting (ω=100, amplitude β=1), three seeds: 12 adapters. Each folds to trained stable rank 1.34–1.52 (comparable to sine at its operating point) and adapts strongly (e.g. CoLA MCC 0.18 → 0.66, STS-B Spearman 0.90 over the frozen base), so there is genuine supra-nominal rank to interrogate.

Result. The RCA signature is identical to sine-LoRA's (pooled over the 12 checkpoints). Truncating each trained loran update back to its nominal rank is lossless (η=1.010 ± 0.018, never below 1.000), the matched-energy random removal is likewise lossless (η=1.001 ± 0.004), and the manufactured tail in isolation collapses to the no-adapter floor (η=0.003 ± 0.008). loran manufactures rank by a different element-wise nonlinearity, and that rank is just as inert: the spectator finding is a property of the "fold the product" construction, not of the sine function in particular.

7 Can the test fire? A rank-demanding probe

Both folding methods come back inert. Before reading that as a fact about the methods, we check that it is not a fact about our instrument: can RCA register a load-bearing positive at all? It can, and a task that genuinely demands rank then shows sine cannot reach the regime where its rank would help.

The instrument is not rigged. A fair worry is that $\eta\approx 1$ is guaranteed by the operating point: with the tail at $\sim$2% of Frobenius energy, truncating it is cheap as linear algebra, independent of content. Two checks rule this out. First, η is unmoved even where the tail carries *real* energy: at ω=1000 the manufactured spectrum reaches stable rank 17.2, yet truncation remains lossless (η=0.9997). Second, the identical truncation applied to MoRA, a genuine high-rank *linear* adapter (trained stable rank 6–13, n=20), retains η=0.94±0.06, and even truncating to *rank one* retains 62–97% of its gain. That the test does not fire on GLUE is a property of the tasks, not the instrument: GLUE adaptation is intrinsically near rank-one, so there is no load-bearing rank for *any* method to lose. To make the test fire we need a task that genuinely demands rank.

A rank-demanding positive control, and sine on it. We construct one: associative memorization of N=4096 random key→value token pairs on a frozen Llama-3.2-1B, where the number of associations a low-rank update can store is rank-limited. Empirically the capacity scales with rank at roughly a few hundred associations per unit rank: plain LoRA at rank 4 saturates near $N\approx$1000 (it recalls 99.9% at N=1024 but only 2% at N=4096), rank 64 recalls 100% of the 4096 pairs, and rank 256 saturates near $N\approx 10^4$. At N=4096 the demand thus far exceeds the nominal r=4. This furnishes the positive control RCA lacked: truncating the rank-64 adapter's update back to rank 4 collapses recall from 1.00 to 0.00, η=0.000. The test fires when rank is load-bearing. η is not pinned at 1; it drops to the floor when the content beyond rank r genuinely carries the task, which settles the circularity worry by demonstration.

The same task lets us probe sine where rank is the binding constraint. Sine's premise is that the nonlinearity buys capacity beyond the nominal rank, so a task that rewards such capacity is a natural place to test it. We hold the demand fixed at N=1024 and sweep the frequency across thirteen values (three seeds through the boundary, n=29); at this N recall reads out *trainability* (a plain rank-4 adapter already attains it), so the sweep isolates whether sine can *reach* a high-realized-rank regime as ω varies, separately from the capacity *demand* probed by the N=4096 control above. The result is a clean trade-off (Fig. 2). Where sine trains, its update folds almost linearly: through $\omega\leq$16 recall is $\approx$100%, the fold argument is tiny ($|\omega BA| \leq 0.013$, so $\sin(\omega BA) \approx \omega BA$), and the *realized* rank (stable rank, the energy-weighted measure) holds between 1.6 and 3.0, the range of a plain rank-4 adapter, even as the *advertised* numerical rank climbs from 9 to 72. Trainability then crumbles over a stochastic band ($\omega \approx$ 20 to 48; seeds split, e.g. recall 0.19 and 0.92 at ω=20) and is gone by ω=48. Realized rank appears only at the far extreme: stable rank 140 at ω=2000, where the fold is finally deep ($|\omega BA|$=6.6) and Proposition 1 of Ji et al. [13] bites; there, recall is 0. Crucially, the failure at rank-manufacturing frequencies is genuine non-optimization, not undertraining: the final training loss there sits at the no-learning ceiling ($\approx \ln V \approx 10.3$ nats for a V-way next-token target; measured 10.5–11.3 for $\omega \geq 38$), against ≤ 0.03 wherever sine trains, so the optimizer makes no measurable progress rather than incomplete progress. The two regimes do not overlap: no frequency makes sine both trainable and high in *realized* rank. The manufactured rank is thus not merely inert but *unreachable by training*: producing energy-bearing rank requires the frequencies at which the sinusoid cannot be optimized, the training-side shadow of the scope condition in §4. This boundary holds for the standard recipe we use (AdamW, fixed learning rate); we do not claim no optimizer can cross it, only that gradient descent as commonly run does not, which is the condition under which the method is deployed. Consequently, on a task that genuinely rewards exceeding rank 4, sine never does, while a genuine rank-64 linear adapter succeeds: where sine can be trained its fold is near-linear and its realized rank is that of plain LoRA, and where it has real rank it cannot be trained. This is the quantitative form of the spectator result, now in the regime where rank is load-bearing. (These memorization runs are exploratory; we rely on the

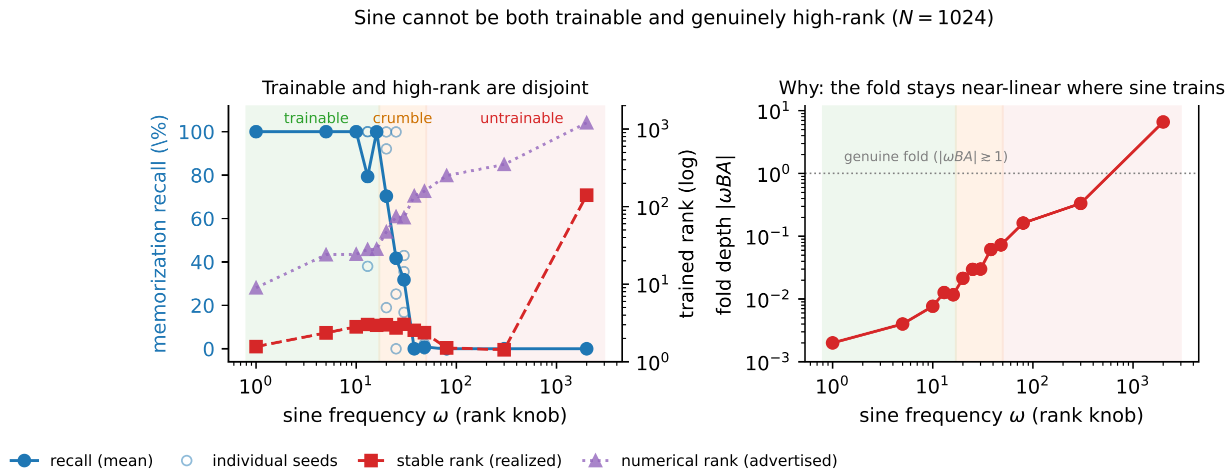

Figure 2: Sine cannot occupy the regime its premise requires (N=1024; three seeds through the boundary). *Left:* as ω rises the *advertised* numerical rank climbs steadily (purple), but the *realized* stable rank (red) stays at plain-rank-4 levels until ω=2000; memorization recall (blue) holds near 100% only while sine trains and crumbles to 0 over $\omega \approx 20$–48 (dots: individual seeds). *Right:* the mechanism, the fold argument $|\omega BA|$, stays far below 1 wherever sine trains, so $\sin(\omega BA) \approx \omega BA$ (near-linear, no genuine fold); a real fold appears only at ω=2000, deep in the untrainable region. The trainable and genuinely-high-rank regimes are disjoint.

categorical separations, not on precise means.)

8 Matched tuning: little gain to explain

The manufactured rank is inert where it exists and unreachable where it would help. We now ask whether there is any headline gain for it to explain.

A caveat, foregrounded: this tier is reproduction-dependent. The original sine-LoRA GLUE setup is not fully reproducible (the authors name only "RoBERTa V3," release no GLUE code, and report hyperparameters that fingerprint a DeBERTaV3-base recipe, which we adopt; §4). The matched-tuning comparison below therefore rests on a reconstructed recipe, and we treat it as *corroborative*, not load-bearing. Crucially, "we could not reproduce a sine *advantage* under matched tuning" is a weaker, separable claim from "the manufactured rank is inert," and only the former is exposed to the reproduction caveat: the RCA result (§5, §6) is paired *within* each trained checkpoint and so is robust to baseline mis-reproduction by construction. The explicit falsifier for the central claim is therefore not a reproduced gain per se: it is a faithful reproduction whose over-base gain *RCA traces to the manufactured rank* (truncation to nominal rank destroying it). Absent that signature, a larger headline gain would not overturn the mechanism finding.

Matched-tuning design. We compare nine arms at identical adapter-parameter budget on W_q, W_v of DeBERTaV3-base over four GLUE tasks [26]: plain LoRA at r=4 and r=8; a block-diagonal slice and an OP-LoRA-style overparameterization (descriptive); a linear high-rank construction at matched parameters (MoRA [14]); the sine activation across a frequency grid; and three **zero-rank-lift controls** that each isolate one optimization-side channel: a bounded trust

Table 5: Matched-tuning GLUE development scores (mean over 5 seeds). Sine at the frozen headline ω; controls at carved-val-best ω. CoLA=MCC, MRPC=F1, STS-B=Spearman, RTE=acc. Best per task in bold (differences n.s., see text); the bordered column (sine) is the arm under test. †OP-LoRA's MRPC mean reflects a convergence failure: it collapses to majority-class prediction (all-positive F1 = 81.2) on 4 of 5 seeds, with the one converged seed scoring 92.3 in line with the other arms. OP-LoRA is a descriptive arm and bears no weight in our conclusions.

task	plain r4	plain r8	sine	tanh	lin-scale	cos-gate	MoRA	block	OP-LoRA
RTE	85.6	**86.0**	83.3	84.5	85.0	85.5	79.9	84.3	**86.0**
CoLA	68.6	**68.7**	67.4	**68.7**	68.3	67.2	66.6	67.5	67.0
MRPC	92.5	**92.9**	92.6	92.3	92.7	92.6	91.7	**92.9**	83.4†
STS-B	**91.5**	91.3	**91.5**	91.4	91.4	**91.5**	90.8	**91.5**	90.9

region ($\gamma \tanh(\omega BA)$), an effective-learning-rate scale ($\gamma\omega BA$, rsLoRA-style [15]), and a straight-through arm whose forward pass is exactly linear but whose backward pass is gated by $\cos(\omega BA)$, reproducing sine's gradient modulation with provably zero forward rank lift. Each arm is tuned on its own coarse-to-fine grid in effective-scale ($\mathrm{LR} \times \gamma \times \omega$) units on a carved validation split; the per-task headline frequency is frozen before any test-split evaluation. Five seeds per cell; development metrics are touched once.

Sine has no matched-tuning advantage. At its pre-registered frozen frequency, sine does not beat plain LoRA on any task (Table 5): it trails on RTE and CoLA and ties on MRPC and STS-B, and plain LoRA at r=8 matches or exceeds it everywhere. The pre-registered inference (a per-task z-standardized model with CR2 cluster-robust standard errors clustered by task and Holm correction; Table 6, 20 seeds) finds *no* contrast significant: sine is statistically indistinguishable from plain LoRA and from all three zero-rank-lift controls (every p_{Holm}=1.0), and the point estimates place plain LoRA and the controls slightly *above* sine, not below. A restricted wild cluster bootstrap-t, the recommended inference at few clusters, agrees (Table 6, p_{WCB}): enumerating all 2^4=16 Rademacher sign vectors, no contrast clears the few-cluster floor of 2/16=0.125, and even MoRA's large negative contrast (−1.9 standardized units) sits exactly at that floor. With only four task clusters the design simply cannot reject at conventional levels, by either method. The descriptive paired-by-seed differences trend the same way (sine −0.4 to −0.9 pt vs. the linear arms) but do not reach significance under task-clustered inference. The published ∼0.6-pt edge thus does not reproduce under matched per-arm tuning, echoing the analogous finding for vanilla LoRA variants [18].

This frozen frequency is, moreover, not sine's most favorable: the carved-val selection that froze it predicts the development optimum poorly (rank correlation between carved-val and dev orderings of ω is near zero). At sine's *development-best* ω (a post-hoc, sine-favorable choice), it marginally exceeds plain LoRA on CoLA, MRPC and STS-B by ≤ 0.3 pt and trails on RTE, never becoming a robust advantage. Under either tuning the conclusion is the same: any sine effect is at most a few tenths of a point and is reproduced by the zero-rank-lift controls.

We are careful not to over-read this null. The data establish that sine confers *no advantage* (it is not significantly better than plain LoRA on any contrast, and its point estimate sits below the linear baseline), but they do not positively *certify* equivalence: the pre-registered TOST does not reach the ±0.5-pt margin ($p \geq 0.93$ vs. plain LoRA and the controls), and with four task clusters the design is underpowered for a tight equivalence claim. "No detectable advantage," not "proven identical," is the honest reading, and it is all the matched-tuning evidence needs to carry: there is

Table 6: Pre-registered CR2 cluster-robust mixed-model contrasts (per-task z-standardized outcome, treatment-coded against sine, clustered by task, Holm-corrected). *: $p_{\text{Holm}} < 0.05$. Full dev set ($N = 440$, 20 seeds). p_{WCB} is the restricted wild cluster bootstrap-t (enumerated over all 2^4 Rademacher sign vectors; few-cluster floor 2/16=0.125). Equivalence (TOST, ± 0.5 pt) and the precision gate shown for the optimization-side contrasts.

contrast (X − sine)	est. (z)	CR2 SE	p_{Holm}	p_{WCB}	TOST p	gate fired
plain-r4	+0.383	0.426	1.000	0.375	0.93	no
plain-r8	+0.528	0.565	1.000	0.500	—	—
tanh	+0.440	0.396	1.000	0.500	0.98	no
lin-scale	+0.455	0.389	1.000	0.375	0.98	no
cos-gate	+0.182	0.353	1.000	0.875	0.28	no
MoRA	−1.899	0.644	0.360	0.125	—	—

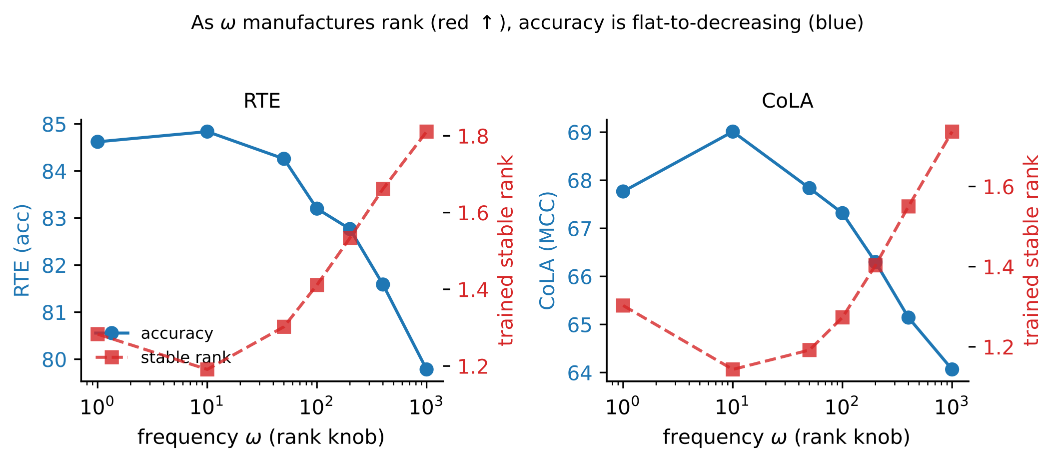

Figure 3: Frequency dose-response for sine-LoRA. As ω raises the trained stable rank (red, right axis), development accuracy (blue, left axis) is flat-to-decreasing: the optimum is the *least* nonlinear setting, and pushing into the rank-manufacturing regime hurts.

no positive sine effect for the manufactured rank, shown causally inert in §5, to explain.

Realized rank does not predict performance. Three convergent analyses agree (Figure 3). *(i) Frequency dose-response:* increasing ω monotonically raises stable rank, but accuracy is flat-to-*decreasing* (15 seeds): RTE runs 84.6 (ω=1) → 84.8 (ω=10) → 82.8 (ω=200) → 79.8 (ω=1000); CoLA 67.8 → 69.0 → 66.3 → 64.1. The optimum is the *least* nonlinear setting. *(ii) Treatment-independent variation:* holding ω=200 and varying realized rank through the factor rank $r \in \{2, 4, 8\}$ and initialization scale, stable rank rises (RTE 1.32 → 1.58 → 1.70) while accuracy does not (82.2 → 83.4 → 82.5); CoLA is flat. *(iii) Cross-arm:* MoRA realizes by far the highest stable rank (8.1 vs. 1.53 for sine at its frozen RTE optimum) yet is worst on three of four tasks; sine's own trained stable rank there (1.53) barely exceeds plain LoRA's (1.50); matched tuning selects the regime where sine scarcely folds. Whatever small, task-dependent effect the reparameterization has is reproduced by zero-rank-lift controls, locating it in optimization rather than rank.

Table 7: Llama-3.2-1B QLoRA commonsense accuracy (mean over 3 seeds). The *no-adapter base* row is the frozen model with no update; on this subset every adapter at best matches it, and most degrade it. Best *adapter* per metric in bold (italic base = reference, not counted); * = arm under test (sine).

arm	mean	BoolQ	PIQA	ARC-e	ARC-c
no-adapter base	*59.29*	*63.8*	*74.8*	*61.9*	*36.7*
plain LoRA r4	56.72	55.8	73.9	60.1	**37.2**
plain LoRA r8	56.21	55.1	73.8	59.0	36.9
sine @headline*	55.94	53.7	73.8	59.5	36.7
sine @400*	56.43	56.0	73.9	59.2	36.6
tanh	56.51	55.3	73.8	60.0	36.9
lin-scale	56.46	55.2	73.8	59.9	36.9
cos-gate	55.90	53.5	73.8	59.4	36.8
MoRA	**59.42**	**62.9**	**75.1**	**62.4**	**37.2**

9 Cross-architecture confirmation: Llama-3.2-1B QLoRA

To test dependence on the primary setting, we repeat the matched-arm comparison on a generative 1B decoder under 4-bit quantization. We fine-tune Llama-3.2-1B [8] with NF4 QLoRA [6] on a seeded 15k-example commonsense subset and evaluate zero-shot on BoolQ, PIQA, ARC-easy, ARC-challenge [2–4] via a pinned harness [7]; eight arm-conditions $\times$ 3 seeds, adapters on W_q, W_v. (RCA is run on the FP16 tier only: quantization perturbs the realized spectrum, so a post-hoc truncation of a dequantized update would not cleanly isolate the trained rank content.)

Both findings replicate. The two primary-tier conclusions hold (Table 7). Sine does not beat plain LoRA (-0.78 pts paired vs. $r{=}4$), and the zero-rank-lift controls again match it (within 0.6 pts). Both tiers agree in sign, so the pre-registered "tiers disagree" ambiguous outcome does not arise.

No form of rank beats the base. The no-adapter reference (Table 7, top) reframes the comparison: on this commonsense subset, fine-tuning mostly *degrades* the base (59.3). Plain LoRA falls to 56.7 and sine to 55.9; MoRA, the *linear* high-rank construction, is the only arm that holds the base (59.4, +0.1). That margin is negligible: no arm shows genuine high rank *improving* adaptation. So even genuine linear high rank confers no over-base gain here whose source could be rank, and the manufactured-rank sine is among the *worst* arms, giving up 3.5 points. Across both architectures, then, no form of rank, manufactured or genuine linear, beats a well-tuned baseline; the arms differ only in how much they sacrifice, and sine sacrifices the most. This is the cross-architecture face of §5: in every setting we measured, rank is a spectator.

10 Related work

Manufacturing rank with nonlinearity. Sine-LoRA [13] inserts a sinusoid between low-rank factors and proves it raises the realized update's rank; loran [5] applies a structured sine map ("Sinter") to the product to the same end. Both genuinely fold the spectrum (§2), and our causal test applies to both: we show the manufactured rank, rather than a co-occurring optimization-side effect, drives no gain. The generative RBF adapter genLoRA [22] is advertised under the same "nonlinearity as rank" banner but is mechanically different: it generates the factors nonlinearly and proves $\mathrm{rank}(\Delta W) \le r$, so it raises no supra-nominal rank to interrogate; the rank-mechanism account is inapplicable to it by construction, whatever its actual source of benefit.

Rank and scaling in linear LoRA. MoRA [14] and HiRA [12] raise rank *linearly*, providing our high-rank baseline. The rsLoRA factor [15] isolates the effective-learning-rate channel that ω also modulates. OP-LoRA [25] shows over-parameterizing the *optimization* of a low-rank update, rather than its rank, can improve adaptation, a precedent for the optimization-side hypothesis and our straight-through control. A broader PEFT line improves LoRA without manufacturing rank nonlinearly (DoRA [19] decomposes magnitude and direction, PiSSA [21] initializes from principal singular components, VeRA [16] shares frozen random factors, and LoRA+ [9] retunes per-factor learning rates), and we do not re-evaluate these. AdaLoRA [27] is the closest to our concern: it allocates a learned, SVD-parameterized spectrum across layers, to which RCA applies directly (its singular values are explicit), a natural next target.

Spectral geometry and tuning confounds. Work on "intruder dimensions" [24] cautions that matching a spectral metric need not match the mechanism, motivating intervention over readout. Reproductions show apparent LoRA-variant gains often dissolve under per-method tuning [17, 18]; this motivates our matched effective-scale protocol, and §8's null replication is, to our knowledge, the first for sine-LoRA.

11 Discussion

What the sine activation does. Our results locate sine-LoRA's effect, where it has one, on the optimization side rather than in rank. The three zero-rank-lift controls each reproduce one channel of the periodic reparameterization (a bounded trust region, an effective-learning-rate rescaling, and data-dependent gradient gating), and each matches sine. Accuracy is flat-to-decreasing in the rank-controlling frequency, and the manufactured rank is inert within every checkpoint. A sinusoid between low-rank factors is, in its useful regime, a conditioned linear adapter; in its rank-manufacturing regime it is no better and often worse.

Implications for adapter design. The "manufacture rank nonlinearly" program rests on a premise our test does not support: that supra-nominal realized rank is the operative variable. Where rank genuinely is the lever, explicit linear high-rank methods capture it; a nonlinearity that inflates the spectrum as a side effect does not. We suggest targeting the optimization-side quantities directly (conditioning, trust region, gradient shaping) and validating any rank-based adapter causally rather than by a rank-vs-accuracy correlation.

Standard benchmarks cannot discriminate the premise. That no setting we measured exhibits load-bearing rank is not only a fact about our adapters; it is a fact about the benchmarks. Standard PEFT evaluations (GLUE, commonsense QA) are intrinsically low-rank adaptation problems: the weight change that solves them occupies a few directions, the low intrinsic dimensionality that makes plain LoRA effective in the first place [1]. We see this directly in our positive control, where even a genuine high-rank *linear* adapter is losslessly truncated to rank one (§7). To be clear, this does not make prior reported gains illusory; those gains are real, but on such tasks they are not *identifiable* as rank effects, because rank is not the binding constraint and any small benefit is attributable to optimization. To make the rank mechanism identifiable at all we built a rank-*demanding* probe (controlled memorization, §7): there the manufactured rank supplies no usable capacity while a genuine linear high-rank adapter supplies all of it. We suggest that any rank-boosting method claiming the rank-r ceiling is the bottleneck be evaluated on at least one task whose adaptation demonstrably demands high rank, not only on benchmarks whose low intrinsic dimensionality makes the rank question moot.

RCA as a general instrument. Rank-Content Ablation needs only a materialized update with realized rank above nominal, so it transfers across nonlinear-rank adapters: we apply it unchanged to both sine-LoRA and loran (§6), and it carries over to future constructions of the same form. The matched-energy control is the essential ingredient, converting an ambiguous "truncation hurts/helps" observation into a clean separation of rank content from energy. We release the implementation and the pre-registered decision rules; all reported analyses regenerate from the released code and seeds, and we deposit a versioned artifact (per-run scores, η values, and spectral and rank summaries) at Zenodo (DOI: 10.5281/zenodo.20887092) for direct reuse.

Limitations. Our equivalence test does not certify equivalence between sine and the controls: the pre-registered TOST fails to reach the ± 0.5 pt margin ($p \geq 0.93$; the paired point estimates, -0.76 to -0.86 pt, themselves exceed it), and with four task clusters the design is underpowered for a tight equivalence claim. We therefore claim only that sine has no advantage over a well-tuned linear adapter, not that the controls are statistically equivalent; the conclusions rest on the interventions and the negative superiority contrast. Our RCA outcomes are in-distribution task metrics (mean accuracy/MCC/F1/Spearman): a low-energy tail that left mean accuracy untouched could in principle still shape calibration, out-of-distribution behavior, or alignment with a task-discriminative subspace. On the calibration axis the truncation is also inert: across six sine checkpoints, truncating the manufactured tail to nominal rank leaves both negative log-likelihood and expected calibration error unchanged ($|\Delta\text{NLL}| \leq 0.0005$, $|\Delta\text{ECE}| \leq 0.002$), and the tail-only result places the manufactured rank alone at the no-adapter floor; a hidden non-accuracy function is therefore unlikely, though we do not rule out out-of-distribution or subspace-alignment effects. RCA is run in full precision (FP16 GLUE classification and FP32 Llama memorization); we do not intervene on the 4-bit-quantized commonsense tier, since quantization perturbs the realized spectrum. The arm-comparison primary tier is one base model on four GLUE tasks, chosen because the premise is structurally absent on the alternative backbone, and we do not test the generative or $\geq$7B-scale regimes where sine-LoRA stakes its strongest claims: we test its rank *mechanism* on classification (GLUE) and 1B-scale commonsense, not its full performance envelope, and a folding-capable large-model checkpoint would extend the intervention. More consequentially, we do not probe the implicit-neural-representation and signal-fitting regimes (image and shape fitting, NeRF) that are sine-LoRA's lineage and the site of its strongest results, and we expect an inert verdict to be regime-specific by construction. Manufactured rank is inert here because an adaptation task does not reward a high-frequency update: gradient pressure keeps the fold argument ωBA small (§7: $|\omega BA| \leq 0.013$ wherever sine trains), so the realized rank stays low-energy (stable rank $\approx r$). Signal fitting inverts that incentive: the target is itself high-frequency, deep folds are rewarded rather than penalised, and the manufactured rank could then carry signal, just as genuine capacity demand makes rank load-bearing in our memorization control. The original fitting results are thus consistent with, not contradicted by, an inert verdict on adaptation; our claim is scoped to nonlinear low-rank *adapters* for task adaptation, the setting our interventions probe. The reproducibility of the original GLUE setup is treated in §4 and §8; all deviations from the reconstructed recipe are listed in `DEVIATIONS.md` and the pre-registered `specs/05_final_prereg_spec.json`.

Conclusion. Manufacturing rank with a nonlinearity raises a low-rank update's realized rank, but that rank carries no task signal. On the intervention tier (DeBERTaV3/GLUE) the manufactured rank is *causally* inert for both folding methods we can probe, sine-LoRA and loran, across the full folding spectrum; behaviorally, neither holds any advantage over a well-tuned linear adapter under matched tuning, and this picture is corroborated on a second architecture and modality (4-bit QLoRA commonsense on Llama-3.2-1B), where no adapter beats the no-adapter base: genuine linear high rank merely matches it, and the manufactured rank degrades it. Rank, here, is a spectator, not a mechanism.

References

[1] Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021. arXiv:2012.13255.

[2] Yonatan Bisk, Rowan Zellers, Ronan Le Bras, Jianfeng Gao, and Yejin Choi. PIQA: Reasoning about physical commonsense in natural language. In *AAAI*, 2020. arXiv:1911.11641.

[3] Christopher Clark, Kenton Lee, Ming-Wei Chang, Tom Kwiatkowski, Michael Collins, and Kristina Toutanova. BoolQ: Exploring the surprising difficulty of natural yes/no questions. In *NAACL*, 2019. arXiv:1905.10044.

[4] Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try ARC, the AI2 reasoning challenge. 2018. arXiv:1803.05457.

[5] Guanzhi Deng, Mingyang Liu, Dapeng Wu, Yinqiao Li, and Linqi Song. Enhancing low-rank adaptation with structured nonlinear transformations. 2025. arXiv:2509.21870.

[6] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. QLoRA: Efficient fine-tuning of quantized LLMs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2305.14314.

[7] Leo Gao, Jonathan Tow, Baber Abbasi, et al. A framework for few-shot language model evaluation. EleutherAI lm-evaluation-harness, 2023. https://github.com/EleutherAI/lm-evaluation-harness.

[8] Aaron Grattafiori et al. The Llama 3 herd of models. 2024. arXiv:2407.21783.

[9] Soufiane Hayou, Nikhil Ghosh, and Bin Yu. LoRA+: Efficient low rank adaptation of large models. In *International Conference on Machine Learning (ICML)*, 2024. arXiv:2402.12354.

[10] Pengcheng He, Jianfeng Gao, and Weizhu Chen. DeBERTaV3: Improving deBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2111.09543.

[11] 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. In *International Conference on Learning Representations (ICLR)*, 2022. arXiv:2106.09685.

[12] Qiushi Huang, Tom Ko, Zhan Zhuang, Lilian Tang, and Yu Zhang. HiRA: Parameter-efficient hadamard high-rank adaptation for large language models. In *International Conference on Learning Representations (ICLR)*, 2025. OpenReview:TwJrTz9cRS.

[13] Yiping Ji, Hemanth Saratchandran, Cameron Gordon, Zeyu Zhang, and Simon Lucey. Efficient learning with sine-activated low-rank matrices. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2403.19243.

[14] Ting Jiang, Shaohan Huang, Shengyue Luo, Zihan Zhang, Haizhen Huang, Furu Wei, Weiwei Deng, Feng Sun, Qi Zhang, Deqing Wang, and Fuzhen Zhuang. MoRA: High-rank updating for parameter-efficient fine-tuning. 2024. arXiv:2405.12130.

[15] Damjan Kalajdzievski. A rank stabilization scaling factor for fine-tuning with loRA. 2023. arXiv:2312.03732.

[16] Dawid J. Kopiczko, Tijmen Blankevoort, and Yuki M. Asano. VeRA: Vector-based random matrix adaptation. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2310.11454.

[17] Sangyoon Lee and Jaeho Lee. Beware of the batch size: Hyperparameter bias in evaluating loRA. 2026. arXiv:2602.09492.

[18] Yu-Ang Lee, Ching-Yun Ko, Pin-Yu Chen, and Mi-Yen Yeh. Learning rate matters: Vanilla loRA may suffice for LLM fine-tuning. 2026. arXiv:2602.04998.

[19] Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. DoRA: Weight-decomposed low-rank adaptation. In *International Conference on Machine Learning (ICML)*, 2024. arXiv:2402.09353.

[20] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. 2019. arXiv:1907.11692.

[21] Fanxu Meng, Zhaohui Wang, and Muhan Zhang. PiSSA: Principal singular values and singular vectors adaptation of large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2404.02948.

[22] Yihao Ouyang, Shiwei Li, Haozhao Wang, Xiandi Luo, Zhuoqi Hu, Yuetong Song, Qiyu Qin, Yichen Li, and Ruixuan Li. Nonlinearity as rank: Generative low-rank adapter with radial basis functions. 2026. arXiv:2602.05709.

[23] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. arXiv:1908.10084.

[24] Reece Shuttleworth, Jacob Andreas, Antonio Torralba, and Pratyusha Sharma. LoRA vs full fine-tuning: An illusion of equivalence. 2024. arXiv:2410.21228.

[25] Piotr Teterwak, Kate Saenko, Bryan A. Plummer, and Ser-Nam Lim. OP-LoRA: The blessing of dimensionality. 2024. arXiv:2412.10362.

[26] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *ICLR (also EMNLP BlackboxNLP Workshop 2018)*, 2019. arXiv:1804.07461.

[27] Qingru Zhang, Minshuo Chen, Alexander Bukharin, Nikos Karampatziakis, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2303.10512.