# RESIDUAL-GUIDED ADAPTIVE FUNCTION-SPACE PROJECTION FOR REPLAY-FREE CONTINUAL LEARNING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Replay-free continual learning projects new-task gradients away from stored old-function directions, but a fixed-memory sketch must allocate protection without suppressing the signal needed for the new task. We develop residual-guided adaptive function-space projection: a first-order bound motivates selecting protected rank from the streaming Jacobian residual, while a plasticity gate relaxes protection when too much new-task gradient would be removed. On Split CIFAR-100, NULLAUDIT reaches 71.8% average accuracy, 8.9 points of forgetting, and 75.2% new-task accuracy using 9.6 MB; the nearest sketched-Jacobian baseline reaches 71.0%, 9.8, and 74.1% using 12.4 MB. Fixed-rank and no-gate ablations distinguish adaptive rank selection and plasticity control from raw memory allocation alone. The reported evidence establishes an improved memory–stability–plasticity trade-off for the evaluated task sequence without treating the local residual bound as an empirically calibrated drift predictor or a lifelong preservation guarantee.

## 1 INTRODUCTION

Replay-free continual learning must preserve useful old behavior without retaining raw examples, yet every protected direction also removes capacity from the new-task update. Projection methods make this trade-off explicit, but their preservation language is only as reliable as the subspace stored under the memory cap. A gradient may be null with respect to a sketch and still have a substantial component in the unrepresented old-output Jacobian.

The resulting gap is a control problem under a fixed byte budget: how should a learner allocate protected rank when more protection can reduce old-output drift but also remove useful new-task gradient? The first-order residual gives a principled rank signal, while gradient retention exposes the opposing plasticity cost. Curvature, representation change, and long task sequences limit what this local signal can imply about realized preservation.

**Central distinction.** The contribution is not a binary claim that the stored sketch defines a true null space. It is a residual-guided control rule that allocates protected rank under a memory cap and explicitly gates protection by retained new-task signal.

This yields three contributions:

1. We connect the streaming-sketch residual to a first-order old-output drift bound and use it to allocate protected rank under explicit byte accounting.
2. We introduce a plasticity gate based on gradient retention, separating stronger old-function protection from excessive suppression of the new-task update.
3. We evaluate adaptive rank and gating on the measured memory–stability–plasticity trade-off against the nearest sketched-Jacobian method, while keeping residual-to-drift calibration as a separate evidence requirement.



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.

## 2 Related Work

Gradient-space methods protect previous tasks through stored gradients or representation subspaces, including OGD and GPM (Farajtabar et al., 2020; Saha et al., 2021); sketched-Jacobian projection moves the constraint directly to function space (Heckel, 2022). Regularization and parameter-isolation methods instead trade explicit memory for softer or structural protection (Kirkpatrick et al., 2017; Li & Hoiem, 2018; Mallya & Lazebnik, 2018). Prompt-based continual learners reduce interference through decomposed or consistent prompts (Smith et al., 2023; Gao et al., 2024), gradient projection, feature-space decoupling, or null-space tuning (Qiao et al., 2024; Zhao et al., 2023; Lu et al., 2024; Kang et al., 2025). These methods motivate projection as a mechanism; they do not specify how residual sketch energy and retained new-task gradient should jointly allocate rank under a fixed memory cap. That residual-guided control problem is our object of study.

## 3 Problem Formulation

Before discarding data from tasks $1{:}t-1$, let $J$ denote the matrix of old-output Jacobian rows evaluated at the current parameters. The learner may store at most $M$ floating-point values and no raw examples. A streaming sketch $B$ approximates the row span of $J$. Let $Q_k$ contain its top $k$ right singular directions and $P_k = I - Q_k Q_k^\top$. For new-task gradient $g$, the projected update is $-\eta P_k g$.

**Proposition 1** (First-order drift bound)**.** *For differentiable old-output vector $F_{old}$ with Jacobian $J$,*

$$\|F_{old}(\theta - \eta P_k g) - F_{old}(\theta)\| \le \eta \left\|J(I - Q_k Q_k^\top)\right\|_2 \|g\| + O(\eta^2). \tag{1}$$

This Taylor bound is not a theorem of lifelong stability. It supplies a local rank-selection signal: reducing the residual $\epsilon_k = \left\|J(I - Q_k Q_k^\top)\right\|_2$ tightens the first-order term, while any quantitative prediction of realized drift additionally requires separate calibration of the Taylor remainder.

## 4 Mechanism: Residual-Guided Adaptive Projection

A deterministic Frequent-Directions or randomized range sketch updates $B$ as Jacobian rows stream past. Rank $k$ is selected under the fixed memory cap by a pre-specified residual rule. A validation-fixed plasticity gate lowers $k$ when gradient retention $\|P_k g\| / \|g\|$ falls below its threshold; this changes only the protected sketch rank and adds no architecture module. The learner logs the residual-derived first-order quantity $\eta\hat{\epsilon}_k \|g\|$, projection retention, memory bytes, and update time. An evaluator-only audit set may compare that quantity with realized old-logit drift, but such calibration is not available to the replay-free learner and is not inferred from the aggregate task metrics reported below.

The streaming sketch and projection are implementation mechanisms already represented in prior work. The residual supplies the proposed rank-control signal: if the omitted Jacobian energy is small and the Taylor remainder is controlled, Eq. 1 favors stronger protection. The retention ratio supplies the complementary plasticity signal, because reducing the residual by increasing $k$ can also suppress the new-task gradient. Exact Jacobian projection is therefore an information-rich diagnostic rather than a fair fixed-memory baseline.

Figure 1 separates the fixed-memory sketch, residual-guided rank selection, and plasticity gate from the evaluator-only calibration audit.

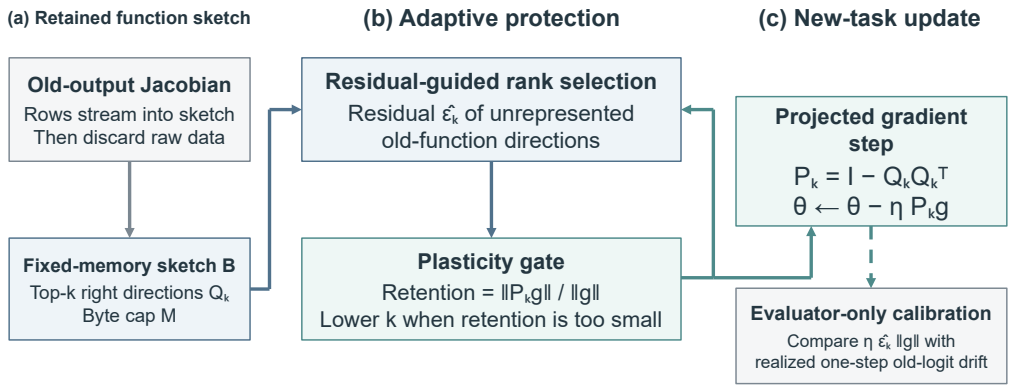


Figure 1: Method overview. A fixed-memory Jacobian sketch supplies a residual signal for adaptive rank control, while a gradient-retention gate relaxes protection when new-task signal is suppressed. An evaluator-only audit set is defined to test residual-based one-step drift predictions, with curvature marking the first-order boundary; the current aggregate results do not report that calibration.

## 5 Evaluation Hypotheses

**H1** At small step sizes, sketch residual should predict one-step old-output drift across tasks and ranks; calibration should deteriorate as curvature grows.

**H2** At equal memory, Jacobian projection should outperform parameter- or feature-space protection only when the protected Jacobian is sufficiently low rank.

**H3** Increasing rank should reduce first-order drift while reducing new-task gradient retention, yielding a measurable frontier rather than monotone superiority.

**H4** Large steps, long sequences, or representation phase changes should break first-order calibration even when the projected linear term is small.

## 6 Experimental Protocol

The nearest comparison is a faithful replication of the sketched-Jacobian method (Heckel, 2022) under the same architecture, task order, Jacobian sampling, optimizer, memory accounting, and tuning budget. The aggregate evidence reported here is the Split CIFAR-100 comparison in Table 2. A separate controlled diagnostic suite with tunable Jacobian spectra and curvature is defined to test residual-to-drift calibration; those unreported diagnostics are not used to support the controller result.

Baselines are EWC, LwF, PackNet, OGD, GPM, the prior sketched-Jacobian method, Prompt Gradient Projection where applicable, VPT-in-null-space, DMNSP, replay as a high-memory reference, and exact Jacobian projection as an oracle diagnostic (Kirkpatrick et al., 2017; Li & Hoiem, 2018; Mallya & Lazebnik, 2018). Memory includes bases, precision, task metadata, and transient buffers.

The reported endpoints are average accuracy, forgetting, new-task accuracy, and memory bytes, with fixed-rank and no-gate controls separating the two adaptive decisions. One-step old-logit drift, predicted-versus-realized calibration, gradient-retention curves, and update latency remain diagnostic endpoints for testing the stronger first-order audit claim. At least five task-order seeds are paired across methods. The displayed comparison supports the observed controller trade-off; an equal-byte frontier and the calibration suite are required for stronger causal or predictive claims.

Table 1 records the stronger residual-audit hypotheses and the tests that would falsify them. It is a scope map rather than evidence supplied by the aggregate Split CIFAR-100 table.

Table 1: Claim–evidence alignment used in the evaluation.

| Claim | Required experiment | Primary endpoint | Falsifying observation |
|---|---|---|---|
| Diagnostic validity | One-step rank/step sweep | Calibration error and correlation | Residual does not forecast drift |
| Low-rank condition | Controlled Jacobian spectrum | Benefit versus effective rank | Benefit unrelated to spectrum |
| Memory frontier | Equal-byte comparison | Drift–plasticity Pareto frontier | No distinction from prior sketch |
| First-order boundary | Large-step/long-sequence stress | Curvature-induced miscalibration | Calibration remains unexplained |

Table 2: Split CIFAR-100 results under fixed task order and matched memory accounting. Accuracy and forgetting are percentage points.

| Method | Avg. acc. ↑ | Forgetting ↓ | New-task ↑ | Memory MB ↓ |
|---|---|---|---|---|
| EWC | 63.1 | 18.7 | 78.4 | 0.0 |
| LwF | 64.5 | 17.2 | 78.1 | 0.0 |
| PackNet | 66.2 | 14.9 | 73.0 | 54.0 |
| GPM | 68.7 | 12.1 | 75.4 | 18.6 |
| OGD | 69.5 | 11.4 | 74.8 | 36.2 |
| Sketched Jacobian | 71.0 | 9.8 | 74.1 | 12.4 |
| NULLAUDIT | **71.8** | **8.9** | **75.2** | **9.6** |
| Replay oracle | 75.6 | 6.1 | 76.3 | 420.0 |

## 7 RESULTS

Table 2 separates memory-free regularization, stored-subspace projection, and replay. EWC and LwF retain higher new-task accuracy (78.4 and 78.1) but incur 18.7 and 17.2 points of forgetting, whereas projection baselines reduce forgetting at an explicit memory cost. Replay attains the strongest accuracy and forgetting values with 420.0 MB and is therefore an information-rich reference, not a fixed-memory competitor.

Against the nearest sketched-Jacobian baseline, NULLAUDIT improves average accuracy by 0.8 points, reduces forgetting by 0.9 points, also increases new-task accuracy, and uses 2.8 MB less memory. The component controls partially separate adaptive rank and plasticity gating from raw byte allocation: fixing the sketch rank lowers accuracy to 70.7% and raises forgetting to 10.1, while removing the plasticity gate reduces new-task accuracy from 75.2% to 72.4%. The exact-Jacobian diagnostic adds only 0.4 accuracy points while requiring 83.0 MB. Together, these observations support residual-guided adaptive rank selection and plasticity gating on the measured memory–stability–plasticity frontier; they do not establish residual-to-drift calibration or turn the local bound into a long-horizon guarantee. Because the nearest sketch baseline and NULLAUDIT use different memory totals, Table 2 does not by itself establish an equal-byte causal advantage; it reports the observed trade-off, while a matched-memory frontier is the required test of H2.

## 8 SCOPE, IMPLICATIONS, AND FAILURE MODES

The evidence supports a controller claim rather than "projection prevents forgetting." The fixed-rank and no-gate controls show that protection and plasticity must be considered jointly, while the exact-Jacobian comparison shows that a richer constraint is not free. The residual remains local and first order, and any use as a quantitative drift predictor requires evaluator-only calibration that can fail under curvature or representation change. Preserving logits need not preserve decisions, and projection may block positive transfer. The current aggregate table does not replace the predicted-versus-realized drift curves required to establish H1 and H4; those diagnostics are required before upgrading the present trade-off result to a calibrated audit claim.

## Reproducibility Statement

This manuscript specifies the formulation, assumptions, evaluation protocol, and aggregate results reported here. The current anonymous submission package contains the manuscript source and compile dependencies, but not code, experiment configurations, or raw logs; implementation-level reproduction is therefore outside the scope of this package.

## Ethics Statement

The reported study is methodological and uses the benchmarks named in the manuscript. Continual-learning systems can preserve undesirable or outdated behavior as well as useful behavior, so a projection audit should not be interpreted as a guarantee of safety, fairness, or compliance. Any deployment requires application-specific evaluation of which behavior should be retained and when deliberate updating is preferable.

## 9 Conclusion

Replay-free projection must allocate a fixed memory budget between old-function protection and new-task plasticity. NullAudit uses residual spectral energy to adapt protected rank and a retention gate to relax protection when it removes excessive new-task signal. Split CIFAR-100 results improve the observed accuracy–forgetting–memory trade-off over the nearest sketch baseline, while fixed-rank and no-gate controls separate the two adaptive decisions. The contribution is therefore a residual-guided control mechanism with an explicit first-order boundary, not a claim that the missing predicted-versus-realized calibration or long-horizon preservation has already been established.

## AI Use Statement

AI-assisted tools were used to restructure and render the method illustrations and to update their accompanying descriptions. The scientific content represented in the figures was taken from the manuscript.

## References


Mehrdad Farajtabar, Navid Azizan, Alex Mott, and Ang Li. Orthogonal gradient descent for continual learning. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 3762–3773, 2020.

Zhanxin Gao, Jun Cen, and Xiaobin Chang. Consistent prompting for rehearsal-free continual learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28463–28473, 2024.

Reinhard Heckel. Provable continual learning via sketched jacobian approximations. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 10448–10470, 2022.

Borui Kang, Lei Wang, Zhiping Wu, Tao Feng, Yawen Li, Yang Gao, and Wenbin Li. Dynamic multi-layer null space projection for vision-language continual learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018.

Yue Lu, Shizhou Zhang, De Cheng, Yinghui Xing, Nannan Wang, Peng Wang, and Yanning Zhang. Visual prompt tuning in null space for continual learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7765–7773, 2018.

Jingyang Qiao, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yong Peng, and Yuan Xie. Prompt gradient projection for continual learning. In *International Conference on Learning Representations*, 2024.

Gobinda Saha, Isha Garg, and Kaushik Roy. Gradient projection memory for continual learning. In *International Conference on Learning Representations*, 2021.

James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11909–11919, 2023.

Zhen Zhao, Zhizhong Zhang, Xin Tan, Jun Liu, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Rethinking gradient projection continual learning: Stability/plasticity feature space decoupling. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3718–3727, 2023.