FEDBRACKET: CONTROLLING FEDERATED DRIFT THROUGH CLIENT-UPDATE NONCOMMUTATIVITY

Anonymous authors
Paper under double-blind review

ABSTRACT

First-order summaries of federated heterogeneity cannot distinguish client updates whose instantaneous disagreement is similar but whose local trajectories interact differently through curvature. We introduce the leading order-sensitive Hessian–gradient bracket as a control signal for local computation. FEDBRACKET estimates this interaction from sparse client pairs and selected-layer Hessian-vector products, then shortens local trajectories in high-bracket rounds. On federated CIFAR-100 with Dirichlet concentration 0.1, FEDBRACKET reaches 75.3% final accuracy in 368 rounds to 70% accuracy with normalized drift 0.52, 3.9 mean local steps, and 3.6% training overhead. A gradient-variance controller reaches 74.0% in 409 rounds with drift 0.66 and 3.8 steps, while a random controller reaches 72.9% in 461 rounds with drift 0.78. These mean-step-matched controls support update noncommutativity as decision-useful information beyond first-order variance or arbitrary step reduction.

1 INTRODUCTION

FedAvg composes several nonlinear client update fields before synchronization. Gradient disagreement describes their directions at the current model, but not the order dependence created when one direction is transported through another client's local geometry. Consequently, similar first-order disagreement need not imply similar round-level drift. This matters operationally because the local-step count controls both communication efficiency and how far each client follows its own geometry before aggregation.

FedAvg introduced iterative local model averaging (McMahan et al., 2017); FedProx and SCAFFOLD subsequently addressed heterogeneous optimization and client drift (Li et al., 2020; Karimireddy et al., 2020). These methods motivate robust federated optimization, but their standard diagnostics do not directly measure the cross-client curvature–direction interaction considered here. Second-order heterogeneity has been studied through bounded Hessian differences (Mishchenko et al., 2024), and Lie-bracket order effects with HVP diagnostics have been analyzed for sequential learning (Sweeney, 2026). The key insight is decision-facing: pairwise update noncommutativity can be estimated sparsely enough to control local computation within a federated round schedule.

We make three testable contributions. First, we isolate the leading Hessian–gradient bracket as a local measure of client-update order sensitivity. Second, we estimate its energy with sparse client pairs and selected-layer Hessian-vector products (HVPs), avoiding the cost of an exact all-pair audit. Third, we use that estimate to control local steps and test the mechanism against fixed-step optimization, a first-order gradient-variance controller, a mean-step-matched random controller, and an exact-bracket reference.

2 CLIENT-UPDATE NONCOMMUTATIVITY

Let one local gradient step on client i be

$$U_i(\theta) = \theta - \eta g_i(\theta), \qquad g_i = \nabla F_i(\theta). \tag{1}$$

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.

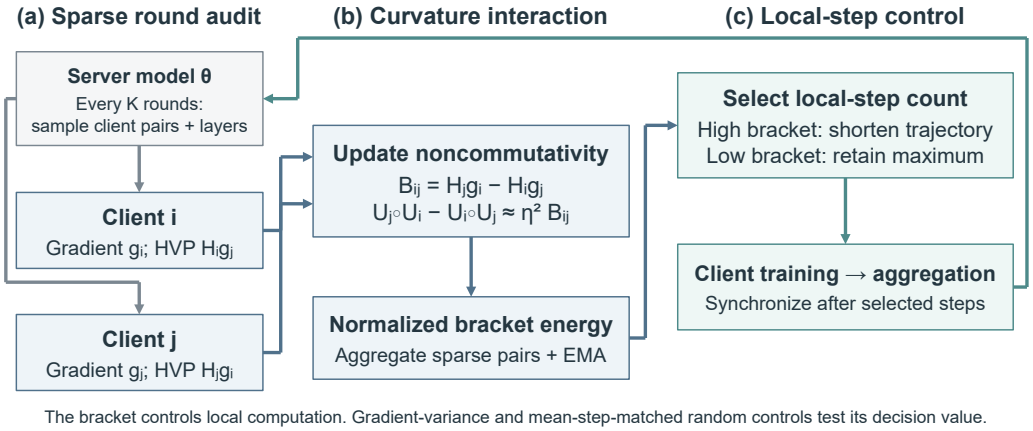

Figure 1: FedBracket control loop. Periodic audits sample client pairs and layers to estimate $B_{ij} = H_j g_i - H_i g_j$ using Hessian–vector products. Normalized bracket energy is smoothed and used to shorten local trajectories in high-bracket rounds while preserving the maximum step count in low-bracket rounds. The signal controls local-step allocation before synchronization.

A local Taylor expansion gives

$$U_j \circ U_i(\theta) = \theta - \eta(g_i + g_j) + \eta^2 H_j g_i + O(\eta^3), \tag{2}$$

$$U_i \circ U_j(\theta) = \theta - \eta(g_i + g_j) + \eta^2 H_i g_j + O(\eta^3). \tag{3}$$

Their leading order-sensitive difference is

$$U_j \circ U_i - U_i \circ U_j = \eta^2 B_{ij} + O(\eta^3), \qquad B_{ij} = H_j g_i - H_i g_j. \tag{4}$$

This is a local characterization of update-order sensitivity, not a global convergence theorem.

Mechanistic hypothesis. When bracket energy is high, additional local steps should amplify order-sensitive disagreement that is invisible to an instantaneous gradient-variance statistic; reducing local steps in those rounds should therefore lower drift. This yields two falsifiable controls. If first-order disagreement is sufficient, the gradient-variance controller should match FEDBRACKET. If only the average reduction in local computation matters, the random controller should match it. The bracket claim requires both alternatives to be separated under comparable mean local steps.

3 BRACKET-AWARE LOCAL-STEP CONTROL

Every K rounds, the server samples client pairs and selected layers. Each paired client evaluates an HVP using a quantized direction supplied through the server, and the server aggregates normalized bracket energy with an exponential moving average. The controller decreases local steps only in high-bracket rounds and preserves the maximum step count when bracket energy is low. Thus, the bracket is not an additional optimizer objective: it selects how much local composition is allowed before synchronization. Disabling the audit recovers the underlying federated optimizer exactly. The protocol assumes the update visibility of ordinary non-secure FedAvg; secure-aggregation compatibility is outside the evaluated interface.

The core causal comparison is not simply adaptive versus fixed local steps. A gradient-variance controller tests a first-order signal, and a random controller matches the average local computation without using bracket information. The exact-bracket oracle uses substantially more computation and serves only as an information-rich reference.

Figure 1 summarizes how the sparse bracket audit controls local computation before federated aggregation.

Table 1: Federated CIFAR-100 results with Dirichlet concentration 0.1 and five nominal local steps. Drift is normalized to FedAvg and overhead is additional training time. Bold marks the best non-oracle result.

Method	Final acc. ↑	Rounds to 70 ↓	Drift ↓	Mean steps	Overhead ↓
FedAvg	71.8	520	1.00	5.0	0.0%
FedProx	72.6	480	0.84	5.0	3.0%
SCAFFOLD	73.4	442	0.71	5.0	8.4%
FedDyn	74.1	421	0.69	5.0	5.8%
FedAdam	73.8	430	0.74	5.0	2.9%
Gradient-variance controller	74.0	409	0.66	3.8	1.8%
Random local-step controller	72.9	461	0.78	3.8	0.1%
FEDBRACKET	**75.3**	**368**	**0.52**	3.9	3.6%
Exact-bracket oracle	75.9	346	0.45	3.8	14.9%

4 EXPERIMENTAL PROTOCOL

The broader protocol defines a controlled geometry suite that varies Hessian eigenspaces while matching first-order gradient disagreement, together with a federated CIFAR-100 evaluation at Dirichlet concentration 0.1 and five nominal local steps. The quantitative evidence reported here is only the federated evaluation in Table 1; no result from the controlled suite is used to support the present conclusion. The table records final accuracy, rounds to 70% accuracy, drift normalized to FedAvg, mean local steps, and additional training time. FedAvg, FedProx, SCAFFOLD, FedDyn, and FedAdam provide standard optimization references (McMahan et al., 2017; Li et al., 2020; Karimireddy et al., 2020; Acar et al., 2021; Reddi et al., 2021). The gradient-variance and random controllers test the two alternative explanations identified above. All controllers use the same client partitions, model, round budget, and evaluation schedule.

5 RESULTS

Table 1 tests the mechanism at comparable mean local computation. FEDBRACKET improves final accuracy by 1.3 points over the gradient-variance controller, reaches 70% accuracy 41 rounds earlier, and lowers normalized drift from 0.66 to 0.52. The similar mean step counts make a purely compute-volume explanation less plausible. Relative to the random controller, it gains 2.4 accuracy points and 93 rounds, showing that merely reducing local steps is insufficient. The exact-bracket oracle improves a further 0.6 points but increases overhead from 3.6% to 14.9%. The oracle gap indicates that the sparse estimator does not recover all bracket information, while the overhead gap explains why exact auditing is not the proposed operating point.

6 DISCUSSION

The two adaptive controls separate complementary alternatives. Gradient variance asks whether a first-order heterogeneity signal is enough; random control asks whether any schedule with fewer local steps suffices. Their separation from FEDBRACKET, together with the lower normalized drift, is consistent with the predicted role of cross-client curvature–direction interaction. Because Table 1 does not report bracket trajectories or sparse-to-exact agreement, it establishes decision utility for the audited signal rather than causal mediation by the bracket. It also does not prove that the bracket is the unique source of federated drift or that the local expansion remains accurate throughout training. The fixed-step federated optimizers show compatibility with standard optimization comparisons, but they do not isolate the mechanism because their local computation differs.

7 SCOPE AND FAILURE MODES

The bracket expansion is local and its reliability decreases for large steps or rapidly changing curvature. HVP auditing adds computation and requires visibility beyond secure aggregation. Only one dataset, one reported partition concentration, and one nominal local-step setting appear in the

main result; seed-level uncertainty, sparse-to-exact fidelity, and controller-sensitivity evidence are not shown in the current manuscript. The reported evidence therefore supports the evaluated controller on federated CIFAR-100; it does not establish a universal convergence advantage for every client objective, heterogeneity regime, or privacy protocol.

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

Federated learning does not itself guarantee privacy. The evaluated audit requires non-secure visibility of update-derived quantities and should not be represented as compatible with secure aggregation. Adaptive local computation can also distribute compute unevenly across clients; deployment therefore requires explicit privacy, participation, and resource-allocation policies.

8 Conclusion

Federated drift is shaped not only by how client updates disagree, but by how their local optimization flows fail to commute. FedBracket converts the leading Hessian–gradient bracket into a sparse control signal for local-step allocation. Mean-step-matched comparisons reduce drift and rounds to target while improving final accuracy, showing that the controller uses information absent from gradient variance and arbitrary step reduction. On federated CIFAR-100, the contribution is a curvature-mediated decision principle for adapting local computation, together with an explicit measurement of its overhead.

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

Durmus Alp Emre Acar, Yue Zhao, Ramon Matas, Matthew Mattina, Paul Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=B7v4QMR6Z9w.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. SCAFFOLD: Stochastic controlled averaging for federated learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5132–5143. PMLR, 2020. URL https://proceedings.mlr.press/v119/karimireddy20a.html.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems*, volume 2, pp. 429–450, 2020. URL https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the*

20th International Conference on Artificial Intelligence and Statistics, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Konstantin Mishchenko, Rui Li, Hongxiang Fan, and Stylianos Venieris. Federated learning under second-order data heterogeneity. OpenReview, 2024. URL https://openreview.net/forum?id=jkhVrIllKg.

Sashank J. Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečný, Sanjiv Kumar, and H. Brendan McMahan. Adaptive federated optimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=LkFG3lB13U5.

John Sweeney. The geometry of sequential learning: Lie-bracket prediction of transfer order. *arXiv preprint arXiv:2606.24993*, 2026. URL https://arxiv.org/abs/2606.24993.