FIBER: A Differentially Private Optimizer with Filter-Aware Innovation Bias Correction

Duc Dm[1] Thao Do[1] Minh Son Hoang[1] Tran Le Duc Anh[2] Daeyoung Kim[1] Huy L. Nguyen[3]

[1] KAIST [2] VinUniversity (CEI) [3] Northeastern University • ICML 2026, Seoul

Scan: code & paper

KAIST · Northeastern · VinUniversity

Abstract

DP training protects examples by adding noise to gradients, but the noise interacts nontrivially with adaptive optimizers. Recent DP methods **temporally filter** privatized gradients to reduce variance - but filtering also changes the DP noise statistics seen by AdamW's second-moment accumulator. Corrections derived for unfiltered noise (subtracting $\sigma^2 w$) become miscalibrated.

We propose **FIBER**, a DP optimizer for temporally filtered gradients that (i) denoises in **innovation space**, (ii) **decouples** the two-point geometry from the innovation gain, and (iii) applies a **filter-aware second-moment calibration** subtracting the attenuated DP noise $A(\omega)\sigma^2 w$.

Across vision and language benchmarks, FIBER consistently improves DP optimizers and surpasses state of the art under equivalent privacy constraints.

Contributions

- **Innovation-space filtering.** Denoise the residual stream $\delta_t = g_t - \hat{g}_{t-1}$ with a lightweight second-order recursion, tracking nonstationary gradient drift under DP noise.
- **Decoupled control.** First explicit separation of observation geometry (α, β) from innovation gain ω, enabling independent tuning of estimator geometry and temporal smoothing.
- **Filter-aware DP-AdamW calibration.** Rigorous analysis of how filtering attenuates DP noise; closed-form A(ω) correction of AdamW's second moment under filtered noise.
- **Extensive validation.** Outperforms DP-Adam(W) and temporal-filtering baselines on vision & language; SOTA under tight privacy budgets and long horizons.

Problem: Filtering Miscalibrates AdamW

Privatized gradient: $g_t = \hat{g}_t + w_t \,, \quad w_t \sim N(0, \sigma^2 w\, I)$

AdamW accumulates the filtered gradient $\hat{g}_t$ into its second moment v_t. For any stable linear filter, the DP noise variance is attenuated:

$$\mathrm{Var}(\hat{g}_{t,i} \mid \hat{g}) = A \cdot \sigma^2 w \,, \quad A \in (0,1]$$

So $E[\hat{v}_{t,i}] \approx E[s^2_{t,i}] + A \cdot \sigma^2 w$. Two consequences:

(i) uncorrected $\hat{v}_t$ keeps a positive DP-noise term that shrinks adaptive steps (preconditioner inflation);

(ii) classic bias corrections (subtract $\sigma^2 w$) are miscalibrated once filtering is present.

FIBER fix: $\bar{v}_t = \max(\hat{v}_t - A(\omega)\sigma^2 w \,, \varepsilon_v)$

Method: FIBER

Three components on top of DP-AdamW. All ops are deterministic functions of the privatized gradients - so FIBER preserves DP-SGD's (ε,δ)-DP by post-processing.

Algorithm — *Filter-aware Innovation Bias-corrected OptimizER*

Input: θ_0, B, C, σ_{DP}, η, $(\beta_1,\beta_2,\varepsilon)$, λ; FIBER: $(\kappa,\gamma,\omega,\varepsilon_v)$
Init: $\tilde{g}_{-1}=r_{-1}=m_{-1}=v_{-1}=0$; $\sigma^2_w=(\sigma_{DPC}/B)^2$;
$A(\omega)=(2-\omega)/(4-3\omega)$
for t = 0, …, T−1 do

(1) DP two-point observation

$u_t(\xi) = a\cdot\nabla f(\theta_t+d_{t-1};\xi) + (1-a)\cdot\nabla f(\theta_t;\xi), \quad a=(1-\gamma)/\kappa$
$g_t = {}^1/_B \sum_\xi \mathrm{clip}(u_t(\xi),C) + w_t, \quad w_t \sim \mathcal{N}(0,\sigma^2_w\, I)$

(2) Innovation (residual) filtering

$\delta_t \leftarrow g_t - \hat{g}_{t-1}$
$r_t \leftarrow (1-\omega)\, r_{t-1} + \omega\, \delta_t$
$\hat{g}_t \leftarrow \hat{g}_{t-1} + r_t$

(3) AdamW moments

$m_t \leftarrow \beta_1\, m_{t-1} + (1-\beta_1)\, \hat{g}_t$
$v_t \leftarrow \beta_2\, v_{t-1} + (1-\beta_2)\, (\hat{g}_t \odot \hat{g}_t)$

(4) Filter-aware bias correction

$\hat{m}_t \leftarrow m_t/(1-\beta_1^{t+1})$; $\hat{v}_t \leftarrow v_t/(1-\beta_2^{t+1})$
$\bar{v}_t \leftarrow \max(\hat{v}_t - A(\omega)\sigma^2_w \,, \varepsilon_v)$

(5) Parameter update

$\theta_{t+1} \leftarrow (1-\eta\lambda)\, \theta_t - \eta \cdot \hat{m}_t \oslash (\sqrt{\bar{v}_t} + \varepsilon)$
$d_t \leftarrow \theta_{t+1} - \theta_t$

end for

Key Takeaways

- ✓ Innovation filtering beats gradient-state EMA in high-noise regimes.
- ✓ Filter-aware correction (A(ω)) prevents preconditioner inflation; DiSK-CORR improves over DiSK.
- ✓ GLUE: FIBER tops DP-AdamW on all tasks, e.g. QNLI 90.5 vs 86.0 @ ε=1.
- ✓ Empirically measured attenuation η_t matches closed-form A(ω) (Fig 6).

Results

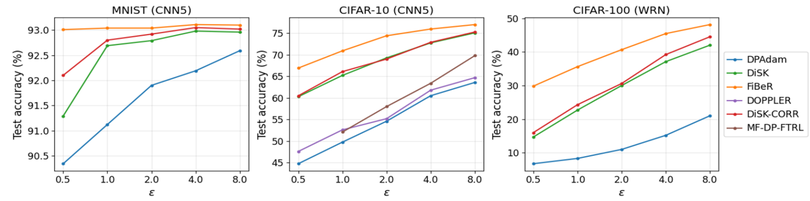

Fig 1. *Final test accuracy vs. privacy budget ε (train from scratch): MNIST, CIFAR-10, CIFAR-100. FIBER leads, with largest gains at low ε.*

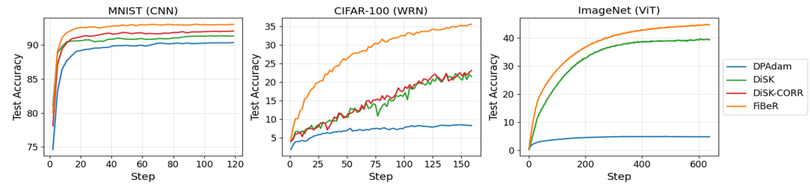

Fig 2. *Training dynamics at fixed ε: MNIST (0.5), CIFAR-100 (1), ImageNet-1k ViT (8). FIBER converges faster and higher.*

Table 1. *Fine-tuning ViT-small on CIFAR-100 (accuracy %).*

Method	ε=0.5	ε=1	ε=2	ε=4	ε=8
DP-AdamW	63.2	78.0	83.7	85.7	86.8
DiSK	83.5	85.4	86.8	87.6	88.5
FIBER	**88.6**	**89.4**	**90.0**	**90.4**	**90.5**

Table 2. *Component ablation on CIFAR-10 (avg Δ over DP-AdamW).*

Method	ε=0.5	ε=1	ε=4	ε=8	Avg Δ
DP-AdamW	44.8	49.8	60.5	63.6	-
1-pt FIBER	46.7	53.4	63.3	66.5	+2.6
DiSK	60.3	65.2	72.8	75.1	+13.9
Full FIBER	**66.9**	**70.9**	**76.0**	**77.0**	**+18.4**

Table 3. *Fine-tuning RoBERTa-base on GLUE (accuracy %).*

	ε = 1		ε = 6.7	
Task	DiSK	FIBER	DiSK	FIBER
MNLI	82.0	**83.0**	84.8	**84.7**
QNLI	88.7	**90.5**	88.9	**90.6**
SST-2	91.5	**93.1**	92.8	**94.0**
QQP	86.9	**88.6**	89.0	**89.5**

Correspondence: ducdm200158@kaist.ac.kr · Proceedings of ICML 2026, PMLR 306

ICML 2026 · Seoul