# Blocking the Leakage: Manifold-Aware Gradient Projection for Long-Horizon Test-Time Adaptation

A geometric intervention for long-horizon test-time adaptation: protect the low-rank knowledge subspace, adapt in the residual.

Yunnan University

# Can TTA stay stable over long horizons?

Short episodes reward fast adaptation. Lifelong deployment punishes small biased updates that accumulate across revisited domains, natural shifts, and blind spots.

TTA should learn at test time without slowly rewriting the model it depends on.

MAIN REFRAMING

# Collapse is directed.

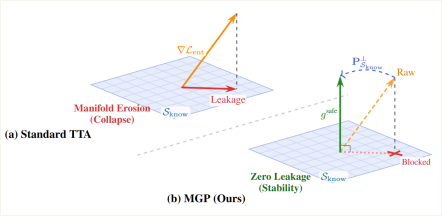
Filtering decides whether to update. It does not decide where the admitted gradient points. Confidently wrong samples can carry a stable destructive direction.

# Spurious gradients leak into the protected subspace.

01 **Core knowledge** concentrates in a stable low-rank subspace.

02 **Confident e rrors** a re h igh-rank, b ut not harmless isotropic noise.

03 **Leakage** accumulates coherently and erodes representations.



SPECTRAL EVIDENCE

# Gradient space splits into protected knowledge and residual plasticity.

## Spectral asymmetry

Reliable gradients reveal a low-rank subspace.

## Leakage builds

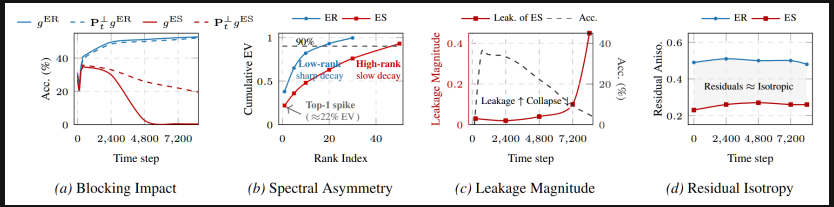
Spurious updates drift coherently.

## Residual useful

Orthogonal directions keep plasticity.

## Project leakage

Block drift while still adapting.




*(a)* Blocking Impact *(b)* Spectral Asymmetry *(c)* Leakage Magnitude *(d)* Residual Isotropy

LEAKAGE HYPOTHESIS

# The dangerous component is the projection into $\mathcal{S}_{\text{know}}$.

Even small per-step leakage dominates when its mean direction is nonzero. High-dimensional residual components can cancel; protected drift grows with horizon.

$$g_t = \mathbf{P}_{\mathcal{S}_{\text{know}}} g_t + \mathbf{P}_{\mathcal{S}_{\text{know}}}^{\perp} g_t$$

Standard TTA cannot distinguish leakage from plasticity in the raw update stream.

MANIFOLD-AWARE GRADIENT PROJECTION

# Track the dominant subspace; update only orthogonally.

01

## Collect raw gradients

FIFO buffer of recent entropy gradients.

02

## Distill spikes

MP bulk-edge separates low-rank structure.

03

## Fuse inertially

Admit only novel stable directions.

04

## Project update

$g^{safe} = (I - U_t U_t^T) g^{raw}$.

ALGORITHM

MGP

# Estimate $U_t$ online; protect stable spikes.

01 BUFFER

## Raw-gradient window

$G_t = [g_{t-n+1}^{raw}, \cdots, g_t^{raw}]^\top$

Use the centered FIFO buffer; eigenvalues are $\lambda_i = s_i^2/n$.

02 MP EDGE

## Select spike rank

$$\hat{\sigma}^2 = \frac{\mathrm{med}(\lambda_i)}{m_\gamma}$$

$$\tau_e = \hat{\sigma}^2(1+\sqrt{\gamma})^2$$

$$\hat{r} = \#\{i : \lambda_i > \tau_e\} \le r_{\max}$$

03 FUSION

## Inertial update

$$R = (I - U_t U_t^\top)\tilde{U}_t$$

$$U_{t+1} = \mathrm{QR}([U_t, R_{:,\mathcal{I}}])$$

$\mathcal{I}$: directions with novel residual energy.

$$g_t^{safe} = (I - U_t U_t^\top) g_t^{raw}, \qquad \theta_{t+1} = \theta_t - \eta\, \mathbb{I}_t g_t^{safe}$$

GUARANTEE

# Protection is exact for $\mathcal{S}_t$; residual drift depends on tracking error.

**Tracked protection:** $\mathbf{P}_t(\theta_{t+1} - \theta_t) = 0$ at every step.

**Reference drift:** bounded by misalignment-weighted gradient norms.

**Spectral tracking:** eigengap plus covariance estimation controls subspace error.

$$F_T(\mathcal{S}_\star) \leq \eta \sum_{t<T} \delta_t \|g_t^{raw}\|_2$$

With residual cancellation, drift becomes sublinear; with coherent leakage, unconstrained TTA drifts linearly.

# Baselines collapse; MGP keeps improving.

**+1.06**
IN-C
R40 minus R1.

**+4.50**
IN-R
Natural shifts.

**7.4M**
Long horizon
Prolonged streams.

REPRESENTATIVE LONG-HORIZON RESULTS

# MGP keeps positive Round-40 drift across highlighted settings.

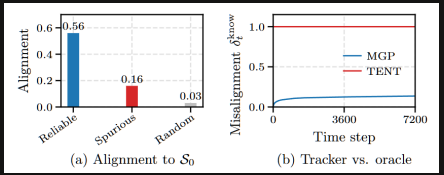
| Setting | Strong failing baseline | Baseline Δ | Ours Δ |
|---|---|---|---|
| IN-C revisits | PTTA | -42.08 | +1.06 |
| IN-3DCC revisits | DeYO | -41.06 | +0.55 |
| IN-C → 3DCC → C-Bar | SAR | -37.75 | +1.45 |
| IN-R natural shift | DeYO / PTTA | -48.32 / -46.50 | +4.50 |

BLIND-SPOT REGIME

# When every stream sample starts misclassified, filtering loses its compass.

Entropy confidence no longer identifies reliable updates, but residual gradients still contain adaptation signal.

| Dataset | Ours R1 | Ours R40 | Δ |
|---|---|---|---|
| IN-C | 38.87 | 39.69 | **+0.82** |
| IN-K | 22.97 | 27.01 | **+4.04** |
| IN-R | 25.29 | 32.85 | **+7.56** |



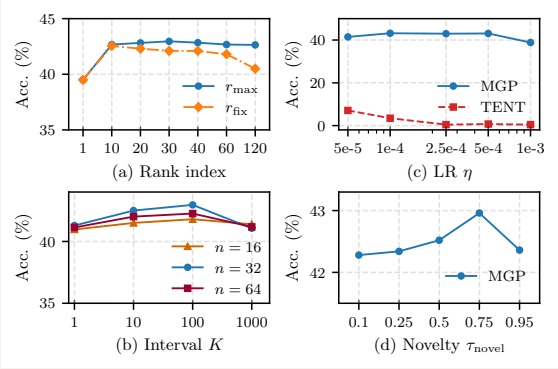

LEAKAGE STORY

# Diagnostics match the mechanism.

- Spurious gradients align with $\mathcal{S}_0$.
- Unsupervised $\mathcal{S}_t$ tracks oracle subspace.
- Projected residuals stay less anisotropic.
- Protected drift remains bounded.

# Failure is governed by residual structure and tracking quality.

| Factor | Condition | Δ Acc. |
|---|---|---|
| Residual | Low anisotropy | -0.14% |
| Residual | High anisotropy | -1.80% |
| Tracking | Frequent refresh | -0.46% |
| Tracking | Sparse refresh | -1.89% |



BEFORE

## TTA asks: which samples should update the model?

Confidence thresholds · entropy gates · reset heuristics

WITH MGP

## TTA also asks: where may updates move?

Protect low-rank knowledge · adapt in residual space

ONE SENTENCE

# Stability is a direction problem.

MGP blocks collapse by removing the protected-subspace component of every test-time gradient, while leaving residual directions available for adaptation.

FOR QUESTIONS

# Projection is lightweight and gate-compatible.

$$\mathbf{P}_t = \mathbf{U}_t \mathbf{U}_t^\top, \quad \mathbf{P}_t^\perp = \mathbf{I} - \mathbf{P}_t$$

$$g_t^{safe} = \mathbf{P}_t^\perp g_t^{raw}, \quad \theta_{t+1} = \theta_t - \eta \mathbb{I}_t g_t^{safe}$$

WHAT TO REMEMBER

## Directed collapse

Errors leak into protected directions.

## Leakage blocked

Projection moves updates to residual space.

## Plasticity kept

Long horizons keep improving.

BLOCKING THE LEAKAGE

# Thank you.

Protect knowledge · keep plasticity · adapt safely over long horizons

ICML 2026 · Test-Time Adaptation