2 — Palindromic Mirror Eliminates One-Way Bias (and How to

A 1-page, hand-verifiable rationale with falsifiers, receipts, and a paper worksheet

One-line claim

If a workflow passes a palindromic mirror test—forward transform T, then inverse T^{-1}, returning to the same rest P within tolerance ε—then one-way bias cannot accumulate between views; failures localize as repairable deltas (Δ-lifts) without contaminating meaning.

Setup

Objects: input x, view operators $\{V_1 \dots V_8\}$, forward transform T, inverse T^{-1}, palindromic rest P, metric $\|\cdot\|$ (e.g., L_2, TV, KL). Invariants: determinism under replay, lockstep parity (two lanes), idempotent canonicalization $C(C(z))=C(z)$.

The palindromic test

Core check: $\| T \cdot P \cdot T^{-1} - P \| \le \varepsilon$. Edge checks (octet): for each view Vi, $\| V_i \cdot T \cdot P \cdot T^{-1} - V_i \cdot P \| \le \varepsilon_i$ with $\varepsilon_i \le \varepsilon$. Interpretation: forward∘back leaves the rest unchanged in every view.

Why bias is annihilated

Any asymmetric drift d inserted by T shows up as a commutator residual $[T, P] := TPT^{-1} - P$. If residual $\le \varepsilon$ (strict ratchet shrinks ε after each pass), then persistent one-way moves cannot accumulate. Non-zero residuals are localized and become Δ-lift targets.

Receipts (what to ledger)

• Residual norms: $r_0 = \|TPT^{-1}-P\|$, $r_i = \|V_iTPT^{-1}-V_iP\|$. • Votes: mirror_votes (e.g., 22/24 probes), view_votes ($\ge$ 6/8). • Four-bit commit: e.g., 1011 for {mirror, octet, Δ-lift, strict} passes. • Merkle-ish page hash over {P, T, thresholds, receipts} for replay.

Δ-lift cookbook (local repairs)

Admissible moves that reduce r_0 and all r_i without increasing any: • Re-parameterize T (units, scaling, pose). • Replace fragile step with identity-equivalent form (algebraic or numerical). • Insert parity guard (swap lanes) where asymmetry leaked. Monotone rule: accept only if all residuals non-increase; otherwise annihilate.

Minimal falsifiers (to keep us honest)

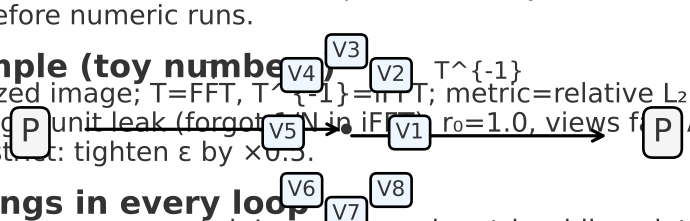
F1: Exhibit a pipeline that passes A/B accuracy but fails the palindromic check (large r_0). F2: Show a case where $r_0 \approx 0$ but r_i explode in $\ge$ 2 views (octet gap). F3: Show replay with identical seeds produces divergent receipts (breaks determinism). Any F1–F3 breaks the claim; fix requires tightening views or redefining P.

Paper worksheet (evening exercise)

1) Choose P (a canonical 'rest' object) and metric. 2) Draw the palindrome rail: P →(T)→ • →(T^{-1})→ P. 3) Compute r_0; then place 8 small boxes (views $V_1 \dots V_8$) around it and compute r_i. 4) If any $r_i > \varepsilon_i$: mark in red, propose one Δ-lift, recompute, and verify monotone decrease. 5) Ledger: write four-bit, votes, and thresholds used. Tip: start with symbolic identities (e.g., FFT/iFFT, encode/decode) before numeric runs.

Tiny live example (toy numbers)

Let P be a normalized image; T=FFT, T^{-1}=iFFT; metric=relative L_2. r_0=3.1e−12, $(r_1 \dots r_8) \le$5e−12 → pass. After injecting a unit leak (forgot 1/N in iFFT): r_0=1.0, views fail. Δ-lift: add 1/N → r_0 back to ~1e−12; strict: tighten ε by ×0.5.



Why this belongs in every loop

Without the palindrome, you can 'win' on a forward metric while quietly accumulating bias. With it, every improvement must survive the round-trip in all views, turning progress into stable receipts.

Receipt sketch: r_0, $\{r_i\}$, votes, four-bit, thresholds → hash → ledger row