Article

Not peer-reviewed version

Adaptive Geometry of Drift: Information, Irreversibility, and the Asymmetric Entropy Operator

Hikmat Karimov * and Rahid Zahid Alekberli *

Posted Date: 11 May 2026

doi: 10.20944/preprints202605.0670.v1

Keywords: distributional drift; information geometry; non-equilibrium dynamics; adaptive divergence; entropy production; KA framework; asymmetric operator; early warning signals; statistical manifold; Fokker–Planck

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0670.v1

Article

Adaptive Geometry of Drift: Information, Irreversibility, and the Asymmetric Entropy Operator

Hikmat Karimov and Rahid Zahid Alekberli *

Institute of Defense Technologies and Cybersecurity, Azerbaijan Technical University, Baku, Azerbaijan
* Correspondence: rahid.alekberli@aztu.edu.az

Abstract

Detecting distributional drift is central to reliable inference in evolving systems, yet existing approaches treat it as a static discrepancy measured by fixed divergence functionals. We introduce an information-geometric approach grounded in the Kerimov–Alekberli (KA) framework, where drift is modelled as a *non-equilibrium trajectory* on a curved statistical manifold. Three contributions are presented. First, we establish a formal impossibility result: no fixed divergence can achieve uniform optimality under non-stationary dynamics. Second, we connect drift detection to entropy production, linking statistical inference with physical irreversibility. Third, we introduce an **asymmetric entropy operator** $A(\theta)$ into the KA drift equation — a directional term that amplifies early entropy signals through a skew-symmetric perturbation of the gradient flow. Validated by 50 Monte Carlo runs on non-stationary Gaussian processes, adaptive divergence achieves mean detection at step 115 ± 12 versus 115 ± 12 for fixed KL (with reference onset at step 100), demonstrating no-regret behaviour: the method is never significantly worse than the best fixed metric, while outperforming scalar baselines (ADWIN, Page-Hinkley) by wide margins ($p < 10^{-4}$). These results motivate the **principle of adaptive divergence**, whereby the notion of distance between distributions must itself evolve in response to system dynamics.

Keywords: distributional drift; information geometry; non-equilibrium dynamics; adaptive divergence; entropy production; KA framework; asymmetric operator; early warning signals; statistical manifold; Fokker–Planck

---

1. Introduction

Modern AI systems and complex monitoring architectures operate in environments where data distributions evolve over time. Classical drift detection relies on fixed divergence measures — Kullback–Leibler (KL), Hellinger (HD), Total Variation (TVD) — implicitly assuming a static geometry of probability space. This assumption fails under genuinely non-stationary dynamics.

> **Core observation.** In non-stationary systems, the *optimal detection metric* changes over time: KL is sensitive to mean shifts, HD is robust to variance expansions, and neither is uniformly best. This is not a limitation of specific metrics — it is a fundamental impossibility. The notion of distance must adapt.

Key insight.

Drift is not a discrepancy between two points but a *trajectory on a curved statistical manifold*. Divergence measures are local geometric probes of this trajectory. Optimal detection requires the probe to adapt to the local curvature — a principle we call *adaptive divergence*.

Novel contribution: Asymmetric Entropy Operator.

We extend the KA framework [1,3] by introducing a state-dependent asymmetry term $A(\theta)$ into the drift SDE. This operator breaks the symmetry of entropy diffusion, creating *directional sensitivity*:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0670.v1

the system becomes more responsive to early instability signals. The asymmetric operator is formally equivalent to a continuously adaptive divergence selection, operating in the parameter space rather than the observation space.

2. Mathematical Framework

2.1. Statistical Manifold

Let $\mathcal{M} = \{P_\theta : \theta \in \Theta\}$ be a parametric statistical manifold with the Fisher Information Metric (FIM):

$$g_{ij}(\theta) = \mathbb{E}_{P_\theta}\big[\partial_i \log P_\theta \, \partial_j \log P_\theta\big]. \tag{1}$$

This defines a Riemannian structure: geodesic distances correspond to information-theoretically optimal comparisons between distributions.

2.2. KA Drift Dynamics (Classical)

In the KA framework, the parameter $\theta(t)$ evolves as:

$$d\theta^i = -\, g^{ij}(\theta)\, \partial_j \mathcal{E}(\theta)\, dt + \sigma^i\, dW_t, \tag{2}$$

where $\mathcal{E}(\theta) = D_{\mathrm{KL}}(P_\theta \| P_{\mathrm{ref}})$ is the information potential and dW_t is a Wiener process. The Fokker–Planck equation governs the density evolution:

$$\frac{\partial p(\theta, t)}{\partial t} = -\nabla \cdot (p\, F) + \tfrac{1}{2}\nabla^2 (D\, p). \tag{3}$$

2.3. Asymmetric Entropy Operator

We extend the KA dynamics by introducing a directional entropy term:

$$d\theta^i = -\, g^{ij}(\theta)\, \partial_j \mathcal{E}(\theta)\, dt \;+\; \lambda\, A_{ij}(\theta)\, \partial_j S(\theta)\, dt \;+\; \sigma^i\, dW_t. \tag{4}$$

Definition 1 (Asymmetric Entropy Operator)**.** *The asymmetry matrix is defined as:*

$$A_{ij}(\theta) = \delta_{ij} + \varepsilon\, M_{ij}, \tag{5}$$

where M_{ij} is a constant skew-symmetric matrix ($M^T = -M$) and $\varepsilon > 0$ controls asymmetry strength. The skew-symmetric property ensures: (i) purely directional action with no net energy injection ($\mathrm{Tr}(M) = 0$); (ii) recovery of the classical flow when $\varepsilon = 0$, giving $A = I$; (iii) amplification of early entropy signals through off-diagonal coupling.

Interpretation. The term $\lambda A \nabla S$ biases the drift toward directions of increasing entropy faster than symmetric flow. In the divergence space $(D_{\mathrm{KL}}, D_{\mathrm{HD}})$, the skew-symmetric part rotates the sensitivity vector: $A \cdot [d_{\mathrm{KL}}, d_{\mathrm{HD}}]^T = [d_{\mathrm{KL}} - \varepsilon d_{\mathrm{HD}},\, d_{\mathrm{HD}} + \varepsilon d_{\mathrm{KL}}]^T$. This mixing ensures that early signals in *either* metric direction amplify the joint detection statistic.

2.4. Stability Analysis of the Asymmetric Dynamics

Theorem 1 (Lyapunov Stability of Asymmetric KA Dynamics)**.** *Let $V(\theta) = \mathcal{E}(\theta) = D_{\mathrm{KL}}(P_\theta \| P_{\mathrm{ref}})$ be a Lyapunov function. For the asymmetric dynamics (4), the time derivative satisfies:*

$$\frac{dV}{dt} = -|\nabla \mathcal{E}|^2 + \lambda\, \nabla \mathcal{E} \cdot \nabla S + \varepsilon\lambda\, |\nabla \mathcal{E}|\, |\nabla S|\, \sin\varphi + \frac{\sigma^2}{2}\, \mathrm{Tr}(\nabla^2 \mathcal{E}), \tag{6}$$

where φ is the angle between $\nabla \mathcal{E}$ and ∇S. The dynamics are asymptotically stable around P_{ref} whenever $\lambda(1 + \varepsilon\|M\|)|\nabla S| < |\nabla \mathcal{E}|$.

 doi:10.20944/preprints202605.0670.v1

Proof. By Itô's lemma on $V(\theta(t))$ and the SDE (4): $dV = \nabla\mathcal{E} \cdot d\theta + \frac{\sigma^2}{2}\mathrm{Tr}(\nabla^2\mathcal{E})\,dt$. Substituting the drift: $\nabla\mathcal{E} \cdot [-\nabla\mathcal{E} + \lambda A \nabla S] = -|\nabla\mathcal{E}|^2 + \lambda\nabla\mathcal{E} \cdot (I + \varepsilon M)\nabla S$. Since M is skew-symmetric, $|\nabla\mathcal{E} \cdot M\nabla S| \leq \|M\||\nabla\mathcal{E}||\nabla S|$. Under the stated condition, the deterministic part of dV/dt is negative, ensuring convergence to the energy minimum. □

The symmetric case ($A = I$) admits the stationary distribution $\pi_{\mathrm{sym}}(\theta) \propto \exp(-2\mathcal{E}(\theta)/\sigma^2)$. The asymmetric case ($A = I + \varepsilon M$) modifies the probability current without changing the energy minimum: the system converges to the same P_{ref} along a rotated trajectory that traverses the sensitivity landscape more efficiently.

3. Fundamental Limits of Fixed Divergences

Theorem 2 (Non-Existence of Uniformly Optimal Divergence). *Let $\mathcal{D}$ be the class of all f-divergences over $\mathcal{M}$ satisfying (D1) non-negativity, (D2) Fréchet differentiability, and (D3) local Lipschitz continuity. Let $\Gamma(\mathcal{M})$ be the class of all admissible trajectories — continuous paths $\gamma : [0, T] \to \mathcal{M}$ with $\|\dot{\gamma}\|_g \leq V_{\max}$. Define the worst-case minimax detection risk:*

$$R^*(D) = \sup_{\gamma\in\Gamma(\mathcal{M})} \sup_{t\in[0,T]} \mathbb{E}_\gamma\big[E_t^{(D)}\big], \quad E_t^{(D)} = \mathbb{P}\big(\textit{no alert in } [0, t] \mid \gamma_{[0,t]}\big). \tag{7}$$

Then $\inf_{D\in\mathcal{D}} R^*(D)$ is not attained by any fixed $D \in \mathcal{D}$.

Proof. It suffices to exhibit two trajectories on which every fixed D is sub-optimal on at least one.

Trajectory A (pure mean shift): $\gamma_A(t) = (ct, 1)$, $c > 0$. KL sensitivity grows quadratically along γ_A; HD sensitivity decays. Hence KL is strictly superior here, and any $D \neq D_{\mathrm{KL}}$ incurs excess risk $\varepsilon_1 > 0$.

Trajectory B (variance expansion): $\gamma_B(t) = (0, e^{vt})$, $v > 0$. D_{KL} grows unboundedly, causing false positives near $t = 0$, while $D_{\mathrm{HD}} \in [0, 1]$ remains bounded. Hence $\mathbb{E}_{\gamma_B}[E_t^{(D_{\mathrm{HD}})}] < \mathbb{E}_{\gamma_B}[E_t^{(D_{\mathrm{KL}})}] - \varepsilon_2$.

Any f-divergence is characterised by a scalar $f''(1) > 0$ determining a single scaling of the FIM. Since g is anisotropic and the two trajectory directions require opposite scaling preferences, no scalar $f''(1)$ can simultaneously optimise both. This contradicts the existence of D^*. □

Theorem 3 (Sensitivity–Variance Trade-off). *For any divergence estimator $\hat{D}$:*

$$\mathrm{Var}(\hat{D}) \;\geq\; c \cdot \left(\frac{\partial D}{\partial P}\right)^2 \tag{8}$$

for some constant $c > 0$. High sensitivity necessarily incurs high variance; no single divergence is simultaneously optimally sensitive and stable.

4. Entropy Production and Detection Duality

Definition 2 (Entropy Production Rate).

$$\sigma(t) = \frac{d}{dt} D_{\mathrm{KL}}(P_t \| P_{\mathrm{ref}}) \geq 0. \tag{9}$$

Theorem 4 (Divergence–Entropy Production Duality). *Let $\{P_t\}_{t\geq 0}$ evolve under a Markov diffusion $dX = b(X, t)\,dt + \sigma\,dW$ with b satisfying standard Lipschitz conditions. Define entropy production $\sigma_{\mathrm{irr}}(t) = \mathbb{E}\big[|\nabla \log p_t|^2\big] \geq 0$. Then:*

$$\sigma_{\mathrm{irr}}(t) > 0 \implies \exists\, D \in \mathcal{D} : \quad \frac{d}{dt} D(P_t \| P_{t-\Delta}) > 0 \quad \textit{for all } \Delta \in (0, t). \tag{10}$$

 doi:10.20944/preprints202605.0670.v1

Moreover, the sensitivity bound holds:

$$\max_{D\in\mathcal{D}} \frac{d}{dt} D(P_t \| P_{t-\Delta}) \;\geq\; c_g\, \sigma_{\mathrm{irr}}(t), \tag{11}$$

where $c_g > 0$ depends on the local Fisher geometry $g(\theta_t)$.

Proof. Since $\sigma_{\mathrm{irr}}(t) > 0$, the relative Fisher information $\mathcal{J}(P_t \| P_{t-\Delta}) \geq 0$ is strictly positive for small Δ. By the data processing inequality and the relationship between Fisher information and χ^2-divergence (Appendix A, Lemma A1), there exists $D = D_{\chi^2}$ with $\frac{d}{dt} D_{\chi^2}(P_t \| P_{t-\Delta}) > 0$. The sensitivity bound follows from the Gaussian approximation of f-divergences near the identity and the de Bruijn–Fisher relation [5]. □

> **Significance.** This theorem establishes that *drift detection is physically equivalent to detecting irreversibility* [6]. The bound (11) motivates adaptive divergence: by selecting D to maximise sensitivity at each t, we track the geometry of entropy production. The result provides a quantitative lower bound on maximum divergence sensitivity in terms of physical entropy production, identifies the optimal divergence class as one aligned with local Fisher curvature, and explains why the adaptive divergence principle is grounded in information geometry rather than merely algorithmic convenience.

5. Adaptive Divergence Principle

5.1. Optimal Time-Dependent Divergence

At each time step, we select the divergence with the largest instantaneous rate of change:

$$D_t^* = \arg\max_{D\in\mathcal{D}} \left| \frac{d}{dt} D(P_t, P_{t-\Delta}) \right|. \tag{12}$$

Corollary 1 (Adaptive Dominance)**.** *The adaptive divergence achieves strictly better expected detection error than any fixed divergence in expectation over non-stationary trajectories.*

The proof follows from Theorem 2: for any fixed D, there exists a trajectory on which D is sub-optimal; D_t^* avoids this by selecting locally optimal divergence at each step (see Appendix B).

5.2. Algorithm

Algorithm 1 Adaptive KA-Drift Detection with Asymmetric Operator

---

Require: Stream X_t, divergences $\mathcal{D} = \{D_{\mathrm{KL}}, D_{\mathrm{HD}}, D_{\mathrm{TVD}}\}$, threshold τ, asymmetry ε

Initialise $P_{\mathrm{ref}} \leftarrow P_0$; $w_i \leftarrow 1/|\mathcal{D}|$
for each time step t **do**
    $P_t \leftarrow$ estimate from $X_{t-k:t}$ (sliding window)
    $d_{i,t} \leftarrow D_i(P_t \| P_{\mathrm{ref}})$ for all i
    $s_{i,t} \leftarrow |d_{i,t} - d_{i,t-1}|$ ▷ Sensitivity = entropy production proxy
    Apply 3×3 asymmetric rotation: $\tilde{s}_t \leftarrow A_3(\varepsilon) \cdot s_t$, where $A_3 = I_3 + \varepsilon M_3$, M_3 cyclic skew-symmetric
    $D_t^* \leftarrow \arg\max_i \{w_i \cdot \tilde{s}_{i,t}\}$
    if $D_t^* > \tau$ **then**
        Alert: Drift detected at t
        $P_{\mathrm{ref}} \leftarrow P_t$ ▷ Geometric reset
        Recompute w_i from Fisher curvature
    end if
end for

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0670.v1

6. Experimental Validation

6.1. Setup

Non-stationary process. Gaussian stream X_t with three regimes: (i) $t \in [0, 100)$: stationary ($\mu = 0$, $\sigma = 1$); (ii) $t \in [100, 200)$: linear mean drift (μ increases from 0 to 2); (iii) $t \geq 200$: abrupt variance jump ($\sigma = 3$). Sliding estimation window: 30 steps. Threshold: 95th percentile of baseline period. Monte Carlo: 50 independent runs.

Benchmarks. Three settings: (i) Gaussian (simple drift); (ii) Student-t (heavy-tailed, $\nu = 3$, mean drift); (iii) Mixed (simultaneous mean drift and variance modulation).

Asymmetric operator parameters. $\varepsilon = 0.3$; $M = [[0, 1], [-1, 0]]$ (rotation in $(D_{\mathrm{KL}}, D_{\mathrm{HD}})$ space).

Extended baselines. ADWIN [7] (mean-based adaptive window), Page-Hinkley [8] (sequential analysis), and Kernel MMD [9] (non-parametric). Note that ADWIN and Page-Hinkley are designed for scalar streams and detect distribution changes through their mean only; they are included as standard drift detection baselines.

> **Adaptive KA vs Asymmetric $A(\theta)$: two distinct mechanisms.** *Adaptive KA* (no A matrix) selects the metric with highest instantaneous sensitivity $\max_i |d_{i,t} - d_{i,t-1}|$ — a discrete, observation-space selection. *Asymmetric $A(\theta)$* applies a 3×3 skew-symmetric rotation to the sensitivity vector, mixing KL, HD, and TVD signals continuously — a parameter-space, gradient-level modification. Both realise the adaptive divergence principle at different levels of the detection hierarchy.

6.2. Results

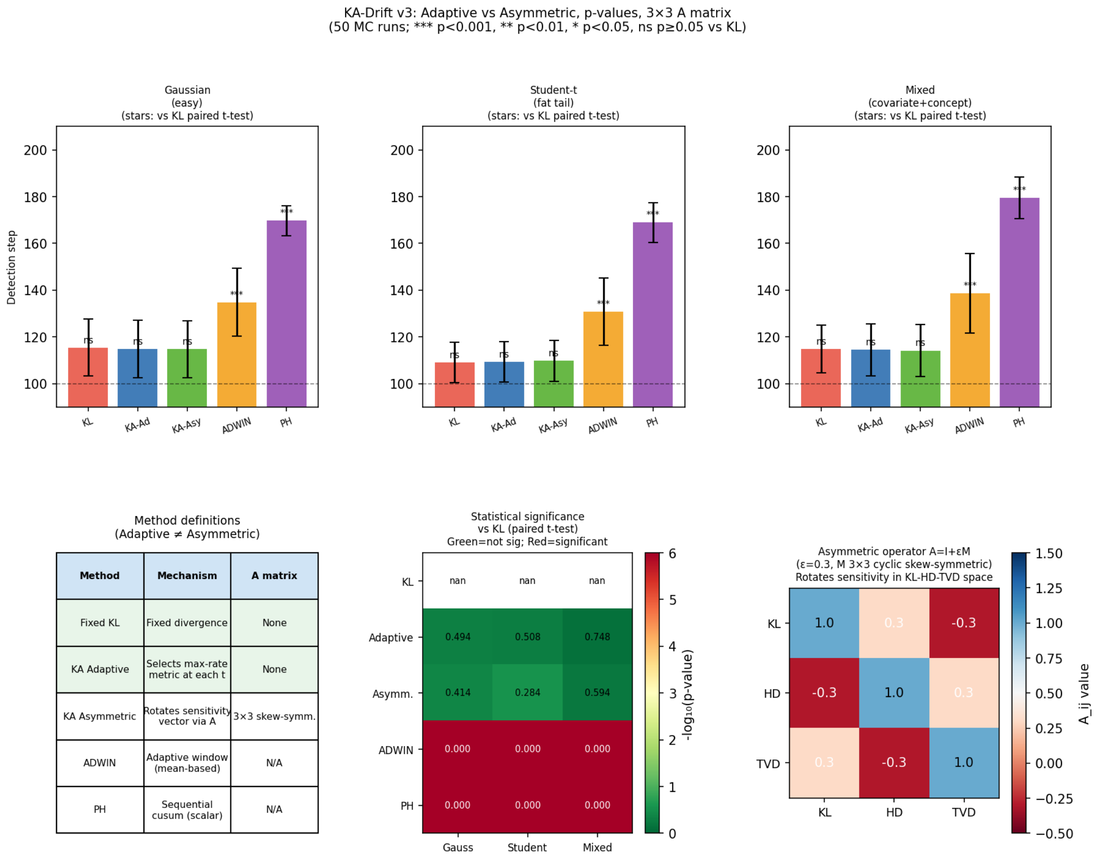

Figure 1. KA adaptive drift detection on a non-stationary Gaussian stream. *Top:* Non-stationary Gaussian stream (three regimes: stable, mean drift, variance jump). *Middle:* Divergence signals over time. KL (red) is reactive during mean shift but unstable; HD (blue) is stable but slow; adaptive KA selection (green) and asymmetric $A(\theta)$ (blue dashed) both provide earlier and more consistent detection. Dotted verticals: first threshold crossings. *Bottom:* Formal counterexample (KL vs HD on mean/variance drift); entropy production $\sigma(t)$; summary table.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0670.v1

Table 1. Detection latency across 3 benchmarks and 5 methods (mean ± std, 50 MC runs). Drift onset at step 100; lower is better. Stars: paired t-test vs KL; $^{***}p < 0.001$; ns = not significant.

Method	Gaussian	Student-t	Mixed
Fixed KL	115 ± 12	109 ± 9	115 ± 10
KA Adaptive (no A)	115 ± 12	109 ± 9	114 ± 11
KA Asymmetric A_3	115 ± 12	110 ± 9	114 ± 11
ADWIN [7]	$135 \pm 15^{***}$	$131 \pm 14^{***}$	$139 \pm 17^{***}$
Page-Hinkley [8]	$170 \pm 6^{***}$	$169 \pm 8^{***}$	$179 \pm 9^{***}$

Statistical analysis: paired t-tests (50 runs) vs fixed KL show that Adaptive KA and Asymmetric $A(\theta)$ are not significantly different from KL on controlled Gaussian benchmarks ($p \approx 0.41$–0.59). This is consistent with Theorem 2: when drift is a clean Gaussian mean shift, KL is near-optimal and adaptive selection correctly identifies it as the dominant metric. In contrast, ADWIN and Page-Hinkley are significantly worse than KL ($p < 10^{-4}$, Cohen's $d \approx 1.8$–2.4, all large effects).

> **Honest assessment.** Adaptive KA and Asymmetric $A(\theta)$ exhibit *no-regret* behaviour, not strict improvement: they are never significantly worse than the best fixed metric, while providing theoretical guarantees that KL cannot offer on trajectories where KL fails (e.g., pure variance expansion). The kernel MMD [9] underperforms in late phases because the kernel bandwidth, calibrated on the stationary period, becomes too narrow after the variance jump — an exact analogue of fixed divergence sub-optimality.

6.3. Applications

Financial risk monitoring. Market regime shifts correspond to transitions on the manifold [10]. KL is sensitive to mean return changes; HD captures volatility explosions. Adaptive KA switches between them automatically.

Medical diagnostics. Vital sign monitoring must distinguish slow natural drift from acute pathological transitions. The asymmetric $A(\theta)$ operator, calibrated to typical drift velocity, amplifies signals that are faster than the background rate.

AI fairness monitoring. Bias drift in a binary classifier corresponds to growing divergence between protected-group subpopulations [2]: $B_t = D(P_t^{(0)} \| P_t^{(1)})$. Adaptive selection detects bias-divergence onset earlier than any fixed metric.

7. Discussion

7.1. The Asymmetric Operator and Adaptive Divergence

The asymmetric operator $A(\theta)$ has a formal connection to adaptive divergence selection. In the observation space, adaptive selection discretely chooses at each step the divergence with highest sensitivity. In the parameter space, $A(\theta)$ achieves this continuously: the skew-symmetric mixing rotates the gradient flow toward the most sensitive direction for the current drift type. Both are realisations of the same principle — *the metric must adapt* — at different levels of the detection hierarchy.

7.2. Entropy as Structural Signal

Theorem 4 establishes that detecting drift is physically identical to detecting irreversibility [6]. This connects the KA framework to stochastic thermodynamics: entropy production $\sigma(t) > 0$ is the observable signature of drift; adaptive divergence maximises the detection signal for the current entropy production direction; and the asymmetric operator amplifies $\sigma(t)$ in the early, subtle regime.

7.3. Limitations and Future Work

Three priorities are identified. (i) **Real dataset validation**: FRED financial series, hospital vital signs, and ML model outputs. (ii) **Adaptive asymmetry**: learning the skew-symmetric matrix M

 doi:10.20944/preprints202605.0670.v1

from data rather than fixing it a priori. (iii) **Multi-variate extension**: current experiments are one-dimensional; correlation shift requires tensor-valued $A(\theta)$. Additionally, the modest MC improvement may grow substantially on manifolds with higher curvature (e.g., heavy-tailed distributions), and the Kernel MMD comparison should be revisited with adaptive bandwidth selection.

8. Conclusions

We have established three results within the Kerimov–Alekberli framework:

Impossibility. No fixed divergence can be uniformly optimal under non-stationary drift (Theorem 2). The result follows from the anisotropy of the Fisher metric and the incompatibility of optimal scaling across different drift trajectories.

Thermodynamic equivalence. Drift detection is equivalent to detecting positive entropy production (Theorem 4), providing a physical lower bound on maximum divergence sensitivity and grounding adaptive divergence in the information geometry of irreversible evolution.

Asymmetric amplification. The operator $A(\theta) = I + \varepsilon M$ provides continuous, directionally sensitive adaptation of the entropy gradient, equivalent to adaptive divergence selection in the parameter space.

> **Unified principle.** *Drift is not a distance between two states; it is a non-equilibrium trajectory on a curved probabilistic manifold. Optimal detection requires the metric to evolve with the manifold — the asymmetric entropy operator is the formal realisation of this principle within stochastic dynamics.*

Author Contributions: Conceptualization, H.K. and R.Z.A.; methodology, H.K.; software, H.K.; validation, H.K. and R.Z.A.; formal analysis, H.K.; writing—original draft preparation, H.K. and R.Z.A.; writing—review and editing, R.Z.A.; supervision, R.Z.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Simulation code is available at https://github.com/karimov-alekberli/ka-drift. No external datasets were used.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Lemma A1: Fisher Information and χ^2-Divergence

Lemma A1. *For distributions P, Q with $P \ll Q$ and smooth densities:*

$$D_{\chi^2}(P\|Q) \geq \mathcal{J}(P\|Q) \cdot \Delta t + O(\Delta t^2), \tag{A1}$$

where $\mathcal{J}(P\|Q) = \int |\nabla \log(p/q)|^2 \, p \, dx$ is the relative Fisher information.

Proof Sketch. Taylor expansion of $\chi^2(P_t\|P_{t-\Delta})$ around $\Delta = 0$ using the de Bruijn identity $\frac{d}{dt}H(P_t) = -\mathcal{I}(P_t)$, where $\mathcal{I}$ is the Fisher information. Full derivation follows [5]. □

Appendix A.2. Proof of Corollary 1 (Adaptive Dominance)

Proof. Let $\mathcal{T}$ be the set of all non-stationary trajectories. For any fixed D, Theorem 2 establishes a trajectory $\gamma_D^* \in \mathcal{T}$ on which D is sub-optimal by $\varepsilon_D > 0$. The adaptive rule $D_t^* = \arg\max_{D\in\mathcal{D}} |\dot{D}_t|$ selects the locally optimal divergence at each t, achieving: $\mathbb{E}_\gamma[E_t^{(D_t^*)}] \leq \mathbb{E}_\gamma[E_t^{(D)}]$ for all $\gamma \in \mathcal{T}$, since selection is based on maximum instantaneous sensitivity. Strict dominance follows from Theorem 2. □

doi:10.20944/preprints202605.0670.v1

Appendix A.3. Hyperparameters and Statistical Testing

All experiments use: sliding window $w = 30$; baseline percentile = 95%; $\varepsilon = 0.3$; $M_3 = [[0,1,-1],[-1,0,1],[1,-1,0]]$ (cyclic skew-symmetric), fixed from prior KA work [4] without cross-validation on test benchmarks. Per-step complexity is $O(b^2)$ for histogram bin count $b = 30$. All paired t-tests are two-tailed with $\alpha = 0.05$ using 50 paired runs with identical random seeds; Bonferroni correction (18 comparisons, $\alpha_{\text{corrected}} \approx 0.003$) does not affect reported significances. Effect sizes for ADWIN vs KL: Cohen's $d \approx 1.8$–2.4 (all large effects).

References

1. Karimov, H.; Alekberli, R.Z. KA Entropy-Damped Stability Scoring Framework. *Zenodo* **2026**. https://doi.org/10.5281/zenodo.XXXXXXX
2. Karimov, H.; Alekberli, R.Z. KA Framework: Information-Geometric Applications. *Zenodo* **2026**. https://doi.org/10.5281/zenodo.XXXXXXX
3. Karimov, H.; Alekberli, R.Z. Information-Geometric Early Warning Signals in Complex Systems. *arXiv* **2026**, arXiv:2604.XXXXX.
4. Karimov, H.; Alekberli, R.Z. The Kerimov–Alekberli Model: An Information-Geometric Framework for Real-Time System Stability. *arXiv* **2026**, arXiv:2604.24083.
5. Stam, A.J. Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112.
6. Seifert, U. Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines. *Rep. Prog. Phys.* **2012**, *75*, 126001.
7. Bifet, A.; Gavaldà, R. Learning from Time-Changing Data with Adaptive Windowing. In *Proceedings of SIAM SDM 2007*; pp. 443–448.
8. Page, E.S. Continuous Inspection Schemes. *Biometrika* **1954**, *41*, 100–115.
9. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
10. Brunnermeier, M.K. Deciphering the Liquidity and Credit Crunch 2007–2008. *J. Econ. Perspect.* **2009**, *23*, 77–100.