Yi's Framework: Structured Loss Analysis Under Time and Order Constraints

Anonymous

Abstract

We present *Yi's Framework*, a operational framework for decision systems that face two ubiquitous structural constraints: *finite-time latency* and *order-sensitivity* (noncommutativity) under reflexive dynamics. We model realized loss as a Bregman regret to an ideal benchmark and prove a *calibrated, additive lower bound*

$$L \geq L_{\text{ideal}} + g_1(\lambda) + g_2(\varepsilon_\star) + g_{12}(\lambda, \varepsilon_\star) - \Delta_{\text{ncx}},$$

Definition (Interaction term g_{12}). Let $a \in \arg\min_{r \in \mathcal{C}_\varepsilon} D_\Phi(p^\star \| r)$ and $b \in \arg\min_{r \in \mathcal{C}_\lambda} D_\Phi(a \| r)$ be any measurable selections (selection-free convention). We define

$$g_{12}(\lambda, \varepsilon) \ := \inf_{q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda} D_\Phi(b \| q).$$

By definition $g_{12} \geq 0$; throughout we assume $\mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda \neq \emptyset$ (feasible intersection).

where g_1 quantifies the latency (time-window) penalty, g_2 quantifies the noncommutative (order) penalty, and $g_{12} \geq 0$ captures their empirical interaction. Here $\Delta_{\text{ncx}} \geq 0$ is a nonconvexity/approximation penalty that *vanishes* under convex or Φ-geodesic-convex conditions; in practice we report the safe nonnegative component $\max\{0,\, g_1 + g_2 + g_{12} - \Delta_{\text{ncx}}\}$. The bound follows from a two-stage Bregman projection with a generalized Pythagorean inequality. We prove monotonicity, calibration properties, conditions for $g_{12} = 0$ (orthogonality), and give estimation procedures and examples. The result turns diverse domain "impossibilities" into a single geometric statement with actionable diagnostics.

Feasible intersection (Assumption I). We assume $\mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda \neq \emptyset$; otherwise g_{12} is undefined and the bound degenerates (use convexification or redesign).

1 Introduction

Decision systems in products, platforms, science, finance, and policy must act within *finite windows*, wait for *delayed, partial verification*, and operate in *reflexive* environments where actions change the future evidence. Empirically, three tensions recur: (i) one cannot wait forever to verify, (ii) the order of predict/verify/act affects outcomes, (iii) the two tensions can amplify each other.

Goal. We formalize these constraints and prove an *inevitable* additive lower bound on achievable loss, decomposed into a time-like component, an order-like component, and a nonnegative interaction. The proof employs proper scoring rules, Bregman divergences, and two-stage convex projections—yielding a precise and portable calculus.

Contributions.

- A *unified loss decomposition* under time and order constraints.
- A *constructive proof* via Bregman geometry (two projections + Pythagorean inequality).
- *Monotonicity* and *orthogonality* theorems; when and why interaction vanishes.
- Practical *diagnostics* for estimating g_1, g_2, g_{12} with uncertainty.

2 Setup: Signals, Loss, and Ideals

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. At decision epoch t_0 the agent chooses an action/policy based on information $\mathcal{I}_{t_0}$ and a predictive object Q (e.g., distribution, score, or control). Outcomes Y are realized later; verification arrives after a lag $\tau \geq 0$. The *action window* has length $\Delta > 0$.

Definition 1 (Ideal benchmark)**.** *L_{ideal} is the minimal expected loss in a counterfactual world with* zero verification lag *($\tau = 0$),* unbounded window *($\Delta \to \infty$), and* order-commuting operators *(no reflexivity): the Bayes risk given true conditionals and full information.*

Operational note. In strongly reflexive systems where operator non-commutativity is fundamental, the counterfactual limit defining L_{ideal} may not have a unique physical realization. In such cases L_{ideal} serves as an idealized reference point; our decomposition (g_1, g_2, g_{12}) quantifies structural penalties relative to this reference, whether or not it corresponds to an attainable state in a given deployment.

Loss as regret. We assume a strictly proper scoring rule (or a convex surrogate) inducing a Bregman divergence D_Φ (Sec. 4). The *excess loss* (regret) w.r.t. the ideal is

$$L - L_{\text{ideal}} = \mathbb{E}\big[D_\Phi(P^\star \| Q)\big],$$

where $P^\star$ denotes the ideal predictive object (true conditional law or Bayes-optimal score).

3 Constraints as Convex Sets

We encode two constraint *families* as closed convex sets in the prediction space.

Definition 2 (Latency-feasible set $\mathcal{C}_\lambda$)**.** *For a latency parameter $\lambda \equiv \lambda(\tau, \Delta)$ (increasing in τ/Δ), the set $\mathcal{C}_\lambda$ contains predictive objects Q that are measurable with respect to the restricted information σ-algebra available before decisions must be finalized.*

Definition 3 (Order-feasible set $\mathcal{C}_\varepsilon$)**.** *For an order-sensitivity index $\varepsilon_\star \geq 0$, the set $\mathcal{C}_\varepsilon$ contains predictive objects Q achievable under operator orderings that satisfy specified reflexivity/disclosure constraints.*

The *feasible region* is $\mathcal{C}(\lambda, \varepsilon_\star) = \mathcal{C}_\lambda \cap \mathcal{C}_\varepsilon$ (assumed nonempty).

Standing assumptions.

Assumption 1. *$\mathcal{C}_\lambda$ and $\mathcal{C}_\varepsilon$ are nonempty, closed, and convex with respect to the affine structure of the prediction space; Φ is Legendre (strictly convex, essentially smooth).*

Domain guidance. The convexity assumption is natural in:

- Probability forecast spaces (distributions form a convex set).
- Risk-neutral pricing (loss is linear in probabilities).
- Randomized policies (mixtures across episodes).

It may fail in:

- Hard sequential decisions with discrete action trees.
- Strongly path-dependent systems with irreversible state changes.

When convexity fails, we recommend using Section 8.2 (convexification) or Section A.3 (mirror geometry) with explicit tracking of δ_{relax} as part of Δ_{ncx}.

4 Bregman Geometry Preliminaries

Let Φ be a strictly convex, differentiable potential on a finite-dimensional open convex set $\mathcal{X}$. The Bregman divergence

$$D_\Phi(p\|q) = \Phi(p) - \Phi(q) - \langle \nabla\Phi(q), p - q\rangle$$

is nonnegative and equals zero iff $p = q$. For a nonempty, closed, convex set $\mathcal{C} \subseteq \mathcal{X}$ and any $p \in \mathcal{X}$, the *Bregman projection* is

$$\Pi_{\mathcal{C}}(p) \in \arg\min_{q\in\mathcal{C}} D_\Phi(p\|q).$$

Lemma 1 (Generalized Pythagorean inequality)**.** *If $q^\star = \Pi_{\mathcal{C}}(p)$, then for any $q \in \mathcal{C}$,*

$$D_\Phi(p\|q) \geq D_\Phi(p\|q^\star) + D_\Phi(q^\star\|q).$$

Proof. First-order optimality for $q^\star$ yields $\langle \nabla\Phi(q^\star) - \nabla\Phi(q), q - q^\star\rangle \geq 0$ for all $q \in \mathcal{C}$. Expanding $D_\Phi(p\|q) - D_\Phi(p\|q^\star) - D_\Phi(q^\star\|q)$ and applying this variational inequality gives the claim. □

5 Main Result: Yis Structured Lower-Bound Decomposition

Define the *ideal* predictive object $p^\star \equiv P^\star$. Let

$$q_\varepsilon = \Pi_{\mathcal{C}_\varepsilon}(p^\star), \qquad q_{\lambda,\varepsilon} = \Pi_{\mathcal{C}_\lambda}(q_\varepsilon), \quad q_{\text{feas}} = \Pi_{\mathcal{C}(\lambda,\varepsilon_\star)}(p^\star).$$

Theorem 1 (Yis Structured Lower-Bound Decomposition)**.** *Under Assumption 1, for any feasible $q \in \mathcal{C}(\lambda, \varepsilon_\star)$,*

$$D_\Phi(p^\star\|q) \;\geq\; \underbrace{D_\Phi(p^\star\|q_\varepsilon)}_{g_2(\varepsilon_\star)} + \underbrace{D_\Phi(q_\varepsilon\|q_{\lambda,\varepsilon})}_{g_1(\lambda)} + \underbrace{D_\Phi(q_{\lambda,\varepsilon}\|q)}_{g_{12}(\lambda,\varepsilon_\star)}. \tag{1}$$

Consequently,

$$L - L_{\text{ideal}} \;\geq\; g_1(\lambda) + g_2(\varepsilon_\star) + g_{12}(\lambda, \varepsilon_\star),$$

where g_1 and g_2 are nonnegative and monotone in their respective arguments, and $g_{12} \geq 0$ by definition, with $g_{12} = 0$ if and only if $q_{\lambda,\varepsilon} = q_{\text{feas}}$ (orthogonality / commuting projections).

Proof. Apply Lemma 1 with $\mathcal{C} = \mathcal{C}_\varepsilon$ to get $D_\Phi(p^\star\|q) \geq D_\Phi(p^\star\|q_\varepsilon) + D_\Phi(q_\varepsilon\|q)$ for any $q \in \mathcal{C}_\varepsilon$ (in particular any feasible q). Then apply Lemma 1 again on $\mathcal{C} = \mathcal{C}_\lambda$ with $p = q_\varepsilon$ and $q \in \mathcal{C}_\lambda$ (again any feasible q qualifies), to obtain $D_\Phi(q_\varepsilon\|q) \geq D_\Phi(q_\varepsilon\|q_{\lambda,\varepsilon}) + D_\Phi(q_{\lambda,\varepsilon}\|q)$. Summing the two inequalities yields Eq. (1). Taking expectations over contexts (if any) transfers to $L - L_{\text{ideal}}$. Monotonicity:

if $\lambda_1 \leq \lambda_2$ then $\mathcal{C}_{\lambda_1} \supseteq \mathcal{C}_{\lambda_2}$, so the projection distance cannot decrease when shrinking the set; similarly for $\varepsilon_\star$. Finally $g_{12} \geq 0$ is $D_\Phi(q_{\lambda,\varepsilon} \| q) \geq 0$, with equality iff $q = q_{\lambda,\varepsilon}$. When the intersection projection equals the sequential projection ($q_{\text{feas}} = q_{\lambda,\varepsilon}$), the minimal feasible q attains $g_{12} = 0$. □

Remark 1 (Interpretation)**.** *g_2 is the* order penalty*: how far the ideal is from any order-feasible object. g_1 is the* latency penalty given *we have already respected order constraints. g_{12} captures the residual gap between sequentially enforcing the constraints and jointly enforcing both—a nonnegative "interaction" that vanishes when projections commute.*

Remark 2 (Selection and the decomposition)**.** *While g_{12} is selection-free by construction (see the definition on p. 1), g_1 and g_2 may depend on which projection $a \in \arg\min_{r \in \mathcal{C}_\varepsilon} D_\Phi(p^\star \| r)$ is chosen when uniqueness fails. However, for any optimal feasible point $q^\star \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$, the* total *lower bound*

$$g_1(\lambda) + g_2(\varepsilon_\star) + g_{12}(\lambda, \varepsilon_\star)$$

is selection-invariant, because it equals $D_\Phi(p^\star \| q^\star)$; both sides are defined as minima over the same feasible set.

6 Units, Calibration, and Monotonicity

The divergence D_Φ inherits the *loss units* of the underlying strictly proper score. Thus g_1, g_2, g_{12} are already in actionable units (regret, dollars, risk points, service-level debt).

Proposition 1 (Boundary conditions)**.** *$g_1(0) = 0$, $g_2(0) = 0$; if $\tau = 0$ or $\Delta \to \infty$, then $\lambda = 0$ thus $g_1 = 0$; if operators commute (no reflexivity/disclosure sensitivity), then $\varepsilon_\star = 0$ thus $g_2 = 0$.*

Proposition 2 (Monotonicity)**.** *If λ increases (shrinking information or shortening windows), $g_1(\lambda)$ is nondecreasing; if $\varepsilon_\star$ increases (stronger reflexivity/order sensitivity), $g_2(\varepsilon_\star)$ is nondecreasing.*

7 When Does the Interaction Vanish?

Theorem 2 (Orthogonality / commuting projections)**.** *Suppose the constraint sets are* Bregman-orthogonal *at $q_{\lambda,\varepsilon}$, i.e., the normal cones $N_{\mathcal{C}_\lambda}(q_{\lambda,\varepsilon})$ and $N_{\mathcal{C}_\varepsilon}(q_{\lambda,\varepsilon})$ span complementary subspaces under the dual geometry. Then $q_{\lambda,\varepsilon} = \Pi_{\mathcal{C}(\lambda,\varepsilon)}(p^\star)$ and $g_{12} = 0$.*

Proof sketch. Under complementary normal cones, sequential KKT conditions at $q_{\lambda,\varepsilon}$ satisfy the joint projection KKT system for the intersection. Hence the sequential projection equals the joint projection, and $D_\Phi(q_{\lambda,\varepsilon} \| q_{\text{feas}}) = 0$. □

Remark 3. *Operationally: if* enforcing order constraints *does not perturb the* gradient of *the* latency constraint *at the solution (and vice versa), interaction disappears. Otherwise, the constraints "push" in coupled directions, inducing $g_{12} > 0$.*

8 Extensions Beyond Convexity and Legendre Assumptions

8.1 Beyond Convexity and Legendre Assumptions

Scope. Yi's FrameworkYi's Framework decomposes

$$L \;\geq\; L_{\text{ideal}} \;+\; g_1(\lambda) \;+\; g_2(\varepsilon_\star) \;+\; g_{12}(\lambda, \varepsilon_\star),$$

under closed convex feasibility and a Legendre potential Φ that induces the Bregman divergence $D_\Phi(\cdot \,\|\, \cdot)$. Real systems are often nonconvex and non-Legendre. We provide four robust paths that keep a valid lower bound and clear operational guidance.

Unified penalized form. In general deployments we will track

$$L \;\geq\; L_{\text{ideal}} \;+\; g_1(\lambda) \;+\; g_2(\varepsilon_\star) \;+\; g_{12}(\lambda, \varepsilon_\star) \;-\; \Delta_{\text{ncx}},$$

where $\Delta_{\text{ncx}} \geq 0$ aggregates nonconvexity/approximation penalties (e.g., convex-hull relaxation gaps, local-curvature remainder ζ, and empirical truncation), with $\Delta_{\text{ncx}} = 0$ under convex or Φ-geodesic-convex conditions.

Nonconvexity penalty decomposition. We write $\Delta_{\text{ncx}} := \delta_{\text{relax}} + \zeta + \delta_{\text{emp}} \geq 0$, where δ_{relax} is the convex-hull relaxation gap from replacing $(\mathcal{C}_\varepsilon, \mathcal{C}_\lambda)$ by $(\widehat{\mathcal{C}}_\varepsilon, \widehat{\mathcal{C}}_\lambda)$, ζ is the local curvature remainder from prox-regular/geodesic analysis, and δ_{emp} is the empirical truncation/monitoring component induced by the 2×2 or DR evaluation. Each term is either estimable or admits an operational upper bound reported with the lower bound.

8.2 Convex Relaxation, Always-Valid Conservative Bounds

Let $\widehat{\mathcal{C}}_\varepsilon := \operatorname{cl}\operatorname{conv}(\mathcal{C}_\varepsilon)$ and $\widehat{\mathcal{C}}_\lambda := \operatorname{cl}\operatorname{conv}(\mathcal{C}_\lambda)$ and define $\hat{a} = \Pi^\Phi_{\widehat{\mathcal{C}}_\varepsilon}(p^\star)$, $\hat{b} = \Pi^\Phi_{\widehat{\mathcal{C}}_\lambda}(\hat{a})$, with

$$\widehat{g}_2 = D_\Phi(p^\star \,\|\, \hat{a}), \quad \widehat{g}_1 = D_\Phi\Big(\hat{a} \,\Big\|\, \hat{b}\Big), \quad \widehat{g}_{12} = \inf_{q \in \widehat{\mathcal{C}}_\varepsilon \cap \widehat{\mathcal{C}}_\lambda} D_\Phi\Big(\hat{b} \,\Big\|\, q\Big).$$

Theorem (short). For any $q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$, $L - L_{\text{ideal}} \geq \widehat{g}_2 + \widehat{g}_1 + \widehat{g}_{12}$. *Use:* safety & audits; transfers directly to OEC if L aligns with D_Φ (proper scoring/calibrated proxy).

Remark 4 (Conservativeness of convexification)**.** *Since $\widehat{\mathcal{C}}_\varepsilon \supseteq \mathcal{C}_\varepsilon$ and $\widehat{\mathcal{C}}_\lambda \supseteq \mathcal{C}_\lambda$, we also have $\widehat{\mathcal{C}}_\varepsilon \cap \widehat{\mathcal{C}}_\lambda \supseteq \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$. Each of $\widehat{g}_2, \widehat{g}_1, \widehat{g}_{12}$ is defined via a minimization over these convexified sets, so the convexified geometric gap $\widehat{g}_2 + \widehat{g}_1 + \widehat{g}_{12}$ cannot exceed the original gap $g_2 + g_1 + g_{12}$ when the latter is well-defined. Thus the convexified lower bound is* smaller *(more conservative) while remaining valid for all $q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$.*

8.3 Local Analysis under Prox-Regularity

With Φ α-strongly convex and smooth near $\{p^\star, a, b\}$, and $\mathcal{C}_\varepsilon, \mathcal{C}_\lambda$ prox-regular locally,

$$L - L_{\text{ideal}} \;\geq\; g_2 + g_1 + g_{12}^{\text{rob}} - \zeta, \qquad \zeta \geq 0.$$

A practical upper bound:

$$\boxed{\zeta \;\leq\; \frac{\rho}{\alpha}\Big(D_\Phi(p^\star \,\|\, a) + D_\Phi(a \,\|\, b)\Big) + c \cdot \widehat{\kappa}\,(\delta\lambda\,\delta\varepsilon)^2},$$

where ρ is weak nonconvexity (prox-regular modulus), and $\widehat{\kappa}$ a local curvature density measured by a micro $2{\times}2$ toggle. *Unified view:* the remainder ζ is subsumed into Δ_{ncx}.

8.4 Assumption-Light Empirical Decomposition

Define four regimes $(L_{00}, L_{01}, L_{10}, L_{11})$ via toggled/staggered latency & order. Estimators:

$$\widehat{g}_2 = L_{01} - L_{00}, \quad \widehat{g}_1 = L_{11} - L_{01}, \quad \widehat{g}_{12} = L_{11} - L_{01} - L_{10} + L_{00}, \quad [\widehat{g}_{12}]_+ = \max\{0, \widehat{g}_{12}\}.$$

Lemma (short). $L - L_{\text{ideal}} \ \geq\ \widehat{g}_2 + \widehat{g}_1 + [\widehat{g}_{12}]_+$. *Practice:* estimate L_{ij} by DR/IPW (selection) + IPCW (censoring); report ESS & clipping%; clustered CIs.

Weights, truncation, and ESS. Inverse-propensity and censoring weights can be heavy-tailed; we therefore enforce explicit truncation and reporting rules. Fix a truncation threshold $c \geq 1$ and let w denote the product of selection and censoring weights. A standard bound for the truncation bias is

$$|\text{Bias}| \ \lesssim\ \mathbb{P}(w > c) \cdot \sup |Y| \ +\ \mathcal{O}(c^{-1}),$$

so we always report: (i) the chosen threshold c, (ii) the fraction of mass with $w > c$ (clipping%), (iii) the effective sample size $\text{ESS} = \big(\sum_i w_i\big)^2 / \sum_i w_i^2$, and (iv) the untruncated estimate side by side with the truncated one. In deployments where ESS falls below a minimum threshold (e.g. 100–200), we recommend using the estimates only for monitoring rather than for hard guarantees. We also adopt doubly robust estimation with cross-fitting, stabilized weights, and optionally TMLE/DR-learner style targeted regressions to improve numerical stability; the residual uncertainty is absorbed into the empirical component of Δ_{ncx}.

Unified view: the truncation control contributes to Δ_{ncx} (empirical truncation component).

8.5 Φ-Geodesic Convexity (mirror/g-convex) Path

Some sets are not Euclidean-convex but are convex in the mirror geometry induced by a Legendre Φ. For closed g-convex $\mathcal{C}_\varepsilon, \mathcal{C}_\lambda$, mirror projections are unique and satisfy a mirror Pythagorean identity, yielding

$$L - L_{\text{ideal}} \ \geq\ g_2 + g_1 + g_{12}^{\Phi} \quad (\text{no } \zeta).$$

Use: choose Φ aligned with the domain (entropy on simplex; quadratic on subspaces; log-partition for exponential families). *Unified view:* under Φ-geodesic convexity, $\Delta_{\text{ncx}} = 0$ and the clean additive law is recovered.

8.6 Guidance Matrix (When to Use Which Path)

Scenario	**Path**	**Output**	**KPI**
Safety & audit	Section 8.2 + Section 8.4	Conservative bound + empirical check	$\widehat{\kappa}\downarrow$, ESS$\uparrow$
Performance tuning	Section 8.3 + Section 8.4	Tight bound, ζ monitor, A/B confirm	$\zeta\downarrow$, OEC$\uparrow$
Routine monitor	Section 8.4	Real-time components	$[\widehat{g}_{12}]_+$, clipping%
Geometry-aligned	Section 8.5 + Section 8.4	Clean additive (mirror)	$[\widehat{g}_{12}]_+\downarrow$
Diagnosis	Section 8.3 (ζ)	Geometry vs. interaction bottlenecks	$\widehat{\kappa}\downarrow$, $\zeta\downarrow$

Remark 5 (Semantics of penalized lower bound)**.** *The nonconvexity penalty only loosens the* lower *bound. A safe report is* $\text{LB}_{\text{safe}} = \max\{0,\ g_1 + g_2 + g_{12} - \Delta_{\text{ncx}}\}$. *When* $\Delta_{\text{ncx}} > g_1 + g_2 + g_{12}$*, the bound is vacuous but still valid; then default to Section 8.4/Section 8.5 to reduce the penalty.*

9 Relation to Classical Decompositions

Biasvariance (static). For point estimation with quadratic loss and a static target, bias and variance are orthogonal in expectation, yielding no cross-term. That corresponds to $\varepsilon_\star = 0$ (no reflexivity) and $\lambda = 0$ (no lag), hence $g_1 = g_2 = g_{12} = 0$ relative to the static Bayes risk. Yi's FrameworkYi's Framework generalizes to dynamic, reflexive, finite-window pipelines where the additive penalties are strictly positive.

CAP-like limits (latency extreme). Under effective partition/communication failure, $\tau \to \infty$ $\Rightarrow \lambda \to \infty$ and g_1 dominates; the theory subsumes distributed and experimental constraints as special cases of the latency term.

10 Estimation of g_1, g_2, g_{12}

Let $L(\cdot)$ denote measured loss/regret under controlled regimes.

Two-stage projection emulation. Construct four regimes: Unconstrained, Order-only, Latency-only, Both. Let L_{00}, L_{01}, L_{10}, L_{11} be corresponding losses above L_{ideal} (estimated via gold references or extrapolation). Define

$$\widehat{g}_2 = L_{01}, \quad \widehat{g}_1 = L_{11} - L_{01}, \quad \widehat{g}_{12} = L_{11} - L_{01} - L_{10} + L_{00} \geq 0.$$

The nonnegativity of $\widehat{g}_{12}$ is implied by Eq. (1) when the regimes emulate sequential and joint projections. Confidence intervals follow from block bootstraps or delta methods.

Interference, reflexivity, and SUTVA. The 2×2 layout is an *emulation* of the two-stage projections, not a literal switch that turns reflexivity or feedback off. In practice we approximate the ideal regimes by: (i) using shadow evaluation or off-policy logging windows in which the deployed policy is held fixed while outcomes are recorded, so that current decisions do not feed back into the state during the evaluation window; (ii) randomizing at the level of clusters, buckets, or episodes so that interference is allowed within clusters but assumed negligible across clusters (partial interference); and (iii) using encouragement designs where an assignment Z nudges the use of a constrained or unconstrained policy while sequential order acts as a mediator. These design choices make the working SUTVA/partial-interference assumptions explicit.

SUTVA sensitivity. We introduce $\Delta_{\text{SUTVA}} \geq 0$ as a sensitivity radius that upper-bounds possible bias in $\widehat{g}_{12}$ due to violations of these assumptions. Alongside the point estimate we report an interval

$$\widehat{g}_{12} \in [\widehat{g}_{12}^{\text{naive}} - \Delta_{\text{SUTVA}},\ \widehat{g}_{12}^{\text{naive}} + \Delta_{\text{SUTVA}}],$$

with Δ_{SUTVA} calibrated by domain knowledge (cluster sizes, leakage rates) or simulation. Correspondingly, the empirical lower bound may be further relaxed as

$$L - L_{\text{ideal}} \geq \widehat{g}_2 + \widehat{g}_1 + [\widehat{g}_{12}]_+ - \Delta_{\text{ncx}} - \Delta_{\text{SUTVA}},$$

when a conservative accounting for interference is required; in this view, Δ_{SUTVA} can be treated as part of the empirical component of Δ_{ncx}.

Continuous calibration. If λ and $\varepsilon_\star$ are graded, fit shape-constrained monotone regressions $g_1(\lambda)$, $g_2(\varepsilon_\star)$ (e.g., isotonic or convex regression) and report g_{12} as the residual consistent with subadditivity bounds.

11 Worked Example: Gaussian Control with Lag and Reflexivity

Consider $Y \in \mathbb{R}$, $Y \sim \mathcal{N}(\mu, \sigma^2)$. The agent outputs $Q = \mathcal{N}(m, v)$, scored by log-loss; then $D_\Phi(P^\star \| Q) = \frac{(m-\mu)^2}{2v} + \frac{1}{2}\Big(\frac{\sigma^2}{v} - 1 - \log\frac{\sigma^2}{v}\Big)$.

Latency constraint. Suppose m must be formed from a *lagged proxy* $\tilde{\mu}$ with $\tilde{\mu} \sim \mathcal{N}(\mu, \sigma_\lambda^2)$ independent of μ and $v = \sigma^2$. Then $g_1(\lambda) = \mathbb{E}\big[\frac{(\tilde{\mu}-\mu)^2}{2\sigma^2}\big] = \frac{\sigma_\lambda^2}{2\sigma^2}$, monotone in the proxy MSE induced by lag.

Order constraint (reflexivity). Suppose actions shift μ by δa and exposure mixes distributions (disclosure), limiting feasible (m, v) to a convex set that shrinks with $\varepsilon_\star$. The projection onto this set increases the mean error by $\Delta m(\varepsilon_\star)$, yielding $g_2(\varepsilon_\star) = \frac{(\Delta m(\varepsilon_\star))^2}{2\sigma^2}$.

Interaction. If the proxy noise correlates with action-induced shifts (e.g., the same channels both delay labels and cause exposure), the joint-feasible projection is stricter than the sequential one, generating $g_{12} > 0$. This toy model makes the abstract terms computable and illustrates monotonicity.

12 Proof Details

We collect key proofs for completeness.

Lemma 2 (Existence/uniqueness of projections)**.** *Under Assumption 1,* $\Pi_{\mathcal{C}}(p)$ *exists and is unique for* D_Φ *with Legendre* Φ*.*

Proof. $D_\Phi(p\|\cdot)$ is strictly convex and lower semicontinuous on closed convex $\mathcal{C}$; a unique minimizer exists by standard convex analysis. ☐

Lemma 3 (Two-stage projection bound)**.** *Let* $\mathcal{A}, \mathcal{B}$ *be closed convex sets with nonempty intersection,* $p \in \mathcal{X}$*, and define* $a = \Pi_{\mathcal{A}}(p)$*,* $b = \Pi_{\mathcal{B}}(a)$*,* $c = \Pi_{\mathcal{A}\cap\mathcal{B}}(p)$*. Then for any* $q \in \mathcal{A} \cap \mathcal{B}$*,*

$$D_\Phi(p\|q) \geq D_\Phi(p\|a) + D_\Phi(a\|b) + D_\Phi(b\|q).$$

Proof. Apply Lemma 1 with $(p, \mathcal{A})$ to split off $D_\Phi(p\|a)$ and with $(a, \mathcal{B})$ to split off $D_\Phi(a\|b)$; the remainder is $D_\Phi(b\|q) \geq 0$. ☐

Proof of Theorem 1. Identify $\mathcal{A} = \mathcal{C}_\varepsilon$, $\mathcal{B} = \mathcal{C}_\lambda$, $p = p^\star$, $a = q_\varepsilon$, $b = q_{\lambda,\varepsilon}$, q feasible; then invoke the lemma and take expectations over contexts. ☐

Proof of Theorem 2. Under complementary normal cones, sequential KKT multipliers jointly satisfy the intersection KKT system; thus $b = c$, implying $g_{12} = D_\Phi(b\|c) = 0$. ☐

13 Operational Playbook (Synopsis)

Measure & report. Instrument τ (lag) and Δ (window) to derive λ. Measure order-sensitivity $\varepsilon_\star$ via counterfactual replay and order perturbations. Publish $(L_{\text{ideal}}, \widehat{g}_1, \widehat{g}_2, \widehat{g}_{12})$ with CIs.

Route by smallest structural penalty. When λ dominates, prefer many *reversible* steps (fast VA). When $\varepsilon_\star$ dominates, prefer conservative, well-verified PV. Use PA when models are robust and actions reversible. Always maintain rollbacks and blast-radius limits.

Invest to shrink penalties. Parallel/stream verification lowers g_1; partial disclosure, decoupled interfaces, and shadow evaluation lower g_2. Seek orthogonality to reduce g_{12}.

14 Discussion and Limitations

Yi's FrameworkYi's Framework provides a geometry that unifies disparate "impossibility" phenomena under time and order constraints. The main limitation is *calibration*: mapping $\lambda, \varepsilon_\star$ to concrete regimes requires domain-specific design. The framework diagnoses structural floors; it complements, not replaces, ethical, legal, and human-centric constraints.

15 Conclusion

We proved a single inequality that decomposes unavoidable loss into time, order, and interaction components via Bregman projections. This converts abstract trade-offs into measurable, optimizable quantities, enabling principled routing and investment across domains. In deployments with nonconvex feasibility or approximate projections, we adopt the penalized additive bound with $\Delta_{\text{ncx}} \geq 0$ and report LB_{safe}; under convex or Φ-geodesic-convex regimes, $\Delta_{\text{ncx}} = 0$ and the clean additive law holds.

Keywords. Yi's FrameworkYi's Framework; Bregman divergence; projection; latency; noncommutativity; reflexivity; lower bound; diagnostics.

A Appendix A: Robust Extensions

A.1 Convex Relaxation, Always-Valid Conservative Bounds

Construction. Let $\widehat{\mathcal{C}}_\varepsilon := \operatorname{cl}\operatorname{conv}(\mathcal{C}_\varepsilon)$ and $\widehat{\mathcal{C}}_\lambda := \operatorname{cl}\operatorname{conv}(\mathcal{C}_\lambda)$. Define relaxed projections

$$\hat{a} := \Pi^{\Phi}_{\widehat{\mathcal{C}}_\varepsilon}(p^\star), \qquad \hat{b} := \Pi^{\Phi}_{\widehat{\mathcal{C}}_\lambda}(\hat{a}),$$

and relaxed penalties

$$\widehat{g}_2 := D_\Phi(p^\star \,\|\, \hat{a}), \qquad \widehat{g}_1 := D_\Phi\left(\hat{a} \,\middle\|\, \hat{b}\right), \qquad \widehat{g}_{12} := \inf_{q \in \widehat{\mathcal{C}}_\varepsilon \cap \widehat{\mathcal{C}}_\lambda} D_\Phi\left(\hat{b} \,\middle\|\, q\right).$$

Theorem 3 (Convexified lower bound — always valid)**.** *For any realized $q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$ (hence $q \in \widehat{\mathcal{C}}_\varepsilon \cap \widehat{\mathcal{C}}_\lambda$) and any Legendre Φ,*

$$D_\Phi(p^\star \,\|\, q) \;\geq\; \widehat{g}_2 + \widehat{g}_1 + \widehat{g}_{12}.$$

Consequently, $L - L_{\text{ideal}} \;\geq\; \widehat{g}_2 + \widehat{g}_1 + \widehat{g}_{12}$.

Proof sketch. $\widehat{\mathcal{C}}_\varepsilon, \widehat{\mathcal{C}}_\lambda$ are closed convex; the Bregman Pythagorean identity gives $D_\Phi(p^\star \,\|\, q) \geq D_\Phi(p^\star \,\|\, \hat{a}) + D_\Phi\Big(\hat{a} \,\Big\|\, \hat{b}\Big) + D_\Phi\Big(\hat{b} \,\Big\|\, q\Big)$ for all $q \in \widehat{\mathcal{C}}_\varepsilon \cap \widehat{\mathcal{C}}_\lambda$. Taking $\inf_q$ on the RHS yields the claim. Any actual feasible q lies in the relaxed intersection, so the bound is conservative. □

Remark 6 (OEC alignment). *When L is aligned with D_Φ (proper scoring or calibrated proxy), Theorem 3 transfers directly to the operational objective; otherwise read it as a geometric audit bound.*

A.2 Local Prox-Regular Analysis with Curvature Penalty

Assumptions. (i) Φ is α-strongly convex and L-smooth near $\{p^\star, a, b\}$. (ii) $\mathcal{C}_\varepsilon, \mathcal{C}_\lambda$ are prox-regular near a, b (local single-valued projections; bounded set curvature).

Proposition 3 (Local robust decomposition with curvature penalty). *Let $a \in \arg\min_{u \in \mathcal{C}_\varepsilon} D_\Phi(p^\star \,\|\, u)$, $b \in \arg\min_{v \in \mathcal{C}_\lambda} D_\Phi(a \,\|\, v)$ and*

$$g_{12}^{\mathrm{rob}} := \inf_{q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda} D_\Phi(b \,\|\, q)\,.$$

Then there exists $\zeta = \zeta(\alpha, L, \text{curvature at } a, b) \geq 0$ such that for all feasible $q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$,

$$D_\Phi(p^\star \,\|\, q) \;\geq\; \underbrace{D_\Phi(p^\star \,\|\, a)}_{g_2} \;+\; \underbrace{D_\Phi(a \,\|\, b)}_{g_1} \;+\; g_{12}^{\mathrm{rob}} \;-\; \zeta.$$

Hence $L - L_{\mathrm{ideal}} \geq g_2 + g_1 + g_{12}^{\mathrm{rob}} - \zeta$.

Operable upper bound for ζ. With weak nonconvexity ρ (prox-regular/weakly-convex modulus) and Φ-strong convexity α,

$$\boxed{\zeta \;\leq\; \frac{\rho}{\alpha}\Big(D_\Phi(p^\star \,\|\, a) + D_\Phi(a \,\|\, b)\Big) \;+\; c \cdot \widehat{\kappa}\,(\delta\lambda\,\delta\varepsilon)^2,}$$

where c is a problem-dependent constant and $\widehat{\kappa}$ is a local curvature density from a 2x2 micro-experiment.

Remark 7 (Semantics of penalized bound). *The penalty only loosens the* lower *bound. A safe report is* $\mathrm{LB}_{\mathrm{safe}} = \max\{0,\, g_1 + g_2 + g_{12} - \zeta\}$. *When ζ exceeds $g_1 + g_2 + g_{12}$, the bound is vacuous but valid; switch to Section 8.4/Section 8.5 to reduce the penalty.*

A.3 Assumption-Light Empirical 2x2 Decomposition

Four regimes (toggle/staggered design).

	order/NC on	order/NC off
latency on	L_{11}	L_{01}
latency off	L_{10}	L_{00}

Estimators. $\widehat{g}_2 = L_{01} - L_{00},\ \widehat{g}_1 = L_{11} - L_{01},\ \widehat{g}_{12} = L_{11} - L_{01} - L_{10} + L_{00}$.

Lemma 4 (Robust lower bound). *With $[\widehat{g}_{12}]_+ := \max\{0, \widehat{g}_{12}\}$, $L - L_{\mathrm{ideal}} \;\geq\; \widehat{g}_2 + \widehat{g}_1 + [\widehat{g}_{12}]_+$.*

Practice. Estimate L_{ij} using DR/IPW (selection) and IPCW (censoring); report effective sample size (ESS) and weight-clipping rates (e.g., 99–99.5%). Provide clustered/bootstrap CIs. Use current $(\widehat{g}_1, \widehat{g}_2, [\widehat{g}_{12}]_+)$ to route, and invest to reduce them over horizons.

A.4 Φ-Geodesic Convexity (Mirror/g-Convex Path)

Mirror geometry. Let $\nabla\Phi$ be the mirror map and $\nabla\Phi^*$ its inverse. The Φ-geodesic segment between x, y is

$$\Gamma_\Phi(x,y;t) \;:=\; \nabla\Phi^*\big((1-t)\,\nabla\Phi(x) + t\,\nabla\Phi(y)\big)\,, \qquad t \in [0,1].$$

Definition 4 (Φ-geodesic convexity). *A set $C \subset X$ is Φ-geodesically convex (g-convex) if $\Gamma_\Phi(x,y;t) \in C$ for all $x, y \in C$, $t \in [0,1]$.*

Projection and Pythagorean in mirror geometry. For closed g-convex C, define $\Pi_C^\Phi(x) := \arg\min_{y\in C} D_\Phi(x \,\|\, y)$. Under Legendre Φ, the mirror projection is unique and satisfies a mirror Pythagorean inequality:

$$D_\Phi(x \,\|\, z) \;\geq\; D_\Phi\Big(x \,\Big\|\, \Pi_C^\Phi(x)\Big) \;+\; D_\Phi\Big(\Pi_C^\Phi(x) \,\Big\|\, z\Big), \qquad \forall z \in C.$$

Theorem 4 (Lower bound under Φ-geodesic convexity). *Suppose $\mathcal{C}_\varepsilon$ and $\mathcal{C}_\lambda$ are closed and g-convex. With $a = \Pi_{\mathcal{C}_\varepsilon}^\Phi(p^\star)$, $b = \Pi_{\mathcal{C}_\lambda}^\Phi(a)$ and $g_{12}^\Phi := \inf_{q\in\mathcal{C}_\varepsilon\cap\mathcal{C}_\lambda} D_\Phi(b \,\|\, q)$, we have*

$$D_\Phi(p^\star \,\|\, q) \;\geq\; D_\Phi(p^\star \,\|\, a) \;+\; D_\Phi(a \,\|\, b) \;+\; g_{12}^\Phi, \qquad \forall q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda,$$

hence $L - L_{\text{ideal}} \geq g_2 + g_1 + g_{12}^\Phi$ without the curvature penalty ζ.

Checks and examples. Pick Φ aligned with the domain: negative entropy on the simplex (probabilities), Mahalanobis/quadratic for Euclidean subspaces, log-partition families for exponential families. Empirically check approximate g-convexity by: (i) testing whether $\Gamma_\Phi(x,y;t)$ stays feasible for random x, y; (ii) testing near-orthogonality of normal directions of $\mathcal{C}_\varepsilon, \mathcal{C}_\lambda$ under the $G = \nabla^2\Phi$ metric.

A.5 Reporting Convention and Diagnostics

We report: (i) $\text{LB}_{\text{safe}} = \max\{0,\, g_1 + g_2 + g_{12} - \Delta_{\text{ncx}}\}$; (ii) a penalty ratio

$$r := \frac{\Delta_{\text{ncx}}}{g_1 + g_2 + g_{12} + \varepsilon}, \quad \varepsilon > 0 \text{ small},$$

with heuristics: $r \leq 0.5$ informative; $r > 1$ vacuous (default to Section 8.4/Section 8.5). We also report ESS, clipping%, and CIs (clustered bootstrap).

A.6 Guidance Matrix (Full)

Scenario	**Recommended Path**	**Output**	**KPI**
Safety & compliance audit	A.1 (convexify) + A.3	Conservative bound + empirical verification	$\widehat{\kappa}\downarrow$, ESS $\uparrow$
Performance optimization	A.2 (local) + A.3	Tight bound, ζ monitoring, A/B confirmation	$\zeta\downarrow$, OEC $\uparrow$
Routine monitoring	A.3 only	Real-time dashboard components	$[\hat{g}_{12}]_+$, clipping%
Geometry-aligned domains	A.4 (g-convex) + A.3	Clean additive bound in mirror geometry	$[\hat{g}_{12}]_+\downarrow$
Root-cause diagnosis	A.2 ζ analysis	Identify geometric vs. interaction bottlenecks	$\widehat{\kappa}\downarrow$, $\zeta\downarrow$