Convex Lower-Bound Decomposition for Decision Systems

Decomposing time, ordering, and interaction effects for reliable decisions

Duo Yi*
yiduo2008@gmail.com

Abstract

We present an operational approach for decision systems that face two ubiquitous constraintsfinite-time latency and order-sensitivity (sequencing). Within a convex, Legendre-structured setting, we model the realized expected loss L of any deployed system as Bregman regret relative to a counterfactual Bayes-optimal benchmark with idealized information; the corresponding Bayes risk is denoted L_{ideal}. We then prove a calibrated, additive lower bound:

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

Each term has a concrete operational meaning: (i) g_1 is the irreducible cost of finite-time latency; (ii) g_2 is the cost induced by order-sensitivity; and (iii) g_{12} is their interaction, which vanishes when the two constraints commute. Under nonempty intersection of the convex feasibility sets $\mathcal{C}_\lambda$ and $\mathcal{C}_\varepsilon$, we provide consistent estimators and calibration procedures for all three components, together with a conservative single-number summary $\max\{0, g_1 + g_2 + g_{12}\}$ (the clipped quantity $\max\{0, g_1 + g_2 + g_{12}\}$) for reporting. We also include a guidance matrix and an operational playbook that translate the decomposition into deployment decisions. The result cleanly separates timing, ordering, and interaction effects, yielding reliable lower bounds and deployment-ready diagnostics for large-scale decision systems.

1 Introduction

Scope. We deliberately work under closed convex feasibility sets and a Legendre potential Φ, which already covers many forecasting and decision systems. Real deployments may deviate from these assumptions. Extending the framework beyond our convex scope introduces additional slack terms and technical machinery. To keep the core ideas transparent, we focus on the clean convex case here and leave a systematic treatment of these departures to future work.

Goal. We formalize these constraints and prove an *inevitable* additive lower bound on the realized loss relative to an ideal benchmark in latency- and order-constrained systems. Our proof employs proper scoring rules, Bregman divergences, and two-stage convex projections, yielding a precise and portable calculus for such systems.

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.

*Primary contact.

2 Setup and Notation

We work on a closed convex domain with a Legendre potential Φ and its Bregman divergence $D_\Phi(\cdot\|\cdot)$. Two convex feasibility sets encode the operational constraints: $\mathcal{C}_\lambda$ for *finite-time latency* and $\mathcal{C}_\varepsilon$ for *ordering (noncommutativity).*

Assumption 1 (Feasible intersection)**.** *The intersection $\mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$ is nonempty.*

Selection convention. For any set-valued argmin we fix measurable selections when needed. All statements about the *total* bound are invariant to the particular selections; individual components may depend on the choice, as discussed in the selection remark below.

Definition 1 (Interaction term)**.** *Let $p^\star$ denote the ideal predictive element corresponding to $P^\star$ (the Bayes-optimal conditional law). 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. Define*

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

Basic properties. By construction $g_{12} \geq 0$. Moreover, when the constraint-induced projections commute, the interaction vanishes (Lemma 1).

Lemma 1 (When the interaction vanishes)**.** *If the constraint-induced projections onto $\mathcal{C}_\varepsilon$ and $\mathcal{C}_\lambda$ commute under D_Φ, then $g_{12}(\lambda,\varepsilon_\star) = 0$.*

3 Signals, Loss, and Ideal Benchmark

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 2 (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 ordering 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 between a predictive object Q and the ideal predictive object $P^\star$. Let L denote the realized expected loss of a particular deployed system under its actual latency and ordering constraints. By Definition 2, the Bayes-optimal expected loss in the counterfactual ideal environment is denoted L_{ideal}. The *excess loss* (regret) with respect to this ideal benchmark 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).

Remark 1 (On the interpretation of L_{ideal})**.** *By Definition 2, L_{ideal} is the Bayes-optimal expected loss in the counterfactual ideal environment and therefore depends only on the data-generating process and the scoring rule, not on any particular deployed system. In practice, L_{ideal} is typically not directly observable. However, our structural terms g_1, g_2, and g_{12} are identified from* differences *between the losses of controlled regimes in our empirical 2×2 design, so L_{ideal} cancels algebraically and does not need to be estimated numerically. When desired, one may approximate L_{ideal} using a high-quality offline or "teacher" model, or by extrapolating loss as latency and ordering constraints are relaxed, but this approximation is optional and not required for the validity of our decomposition.*

4 Constraints as Convex Sets

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

Definition 3 (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 4 (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 2. *$\mathcal{C}_\lambda$ 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 8.1 (convexification)

5 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 2 (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. □

6 Main Result: Structured Lower-Bound Decomposition

Remark 2 (Safe reporting)**.** *Since* $L - L_{\text{ideal}} \geq g_1 + g_2 + g_{12}$*, a simple conservative scalar summary is* $\max\{0,\, g_1 + g_2 + g_{12}\}$*, i.e., the clipped quantity* $\max\{0,\, g_1 + g_2 + g_{12}\}$*, which clips negligible negative estimates at zero.*

Throughout this section we work under Assumption 1 and use the interaction term in Definition 1.

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 (Structured Lower-Bound Decomposition)**.** *Under Assumption 2, 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 2 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 2 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 3 (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 4 (Selection and the decomposition)**.** *While* g_{12} *is selection-free by construction (see Definition 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.*

7 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.*

8 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 5. *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 $g_{12} > 0$.*

8.1 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 6 (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 $q \in \mathcal{C}_\varepsilon \cap \mathcal{C}_\lambda$.*

8.2 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 $\text{ESS} = \left(\sum_i w_i\right)^2 / \sum_i w_i^2$, 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

Use: choose Φ aligned with the domain (entropy on simplex; quadratic on subspaces; log-partition for exponential families).

8.3 Guidance Matrix (When to Use Which Path)

Scenario	Path	Output	KPI
Safety & audit	Section 8.1 + Section 8.2	Conservative bound + empirical check	$\widehat{\kappa}\downarrow$, ESS$\uparrow$
Routine monitor	Section 8.2	Real-time components	$[\widehat{g}_{12}]_+$, clipping%

Remark 7 (Safe reporting convention)**.** *Since* $L - L_{\text{ideal}} \geq g_1 + g_2 + g_{12}$*, a simple conservative scalar summary is* $\max\{0,\, g_1 + g_2 + g_{12}\}$*, i.e., the clipped quantity* $\max\{0,\, g_1 + g_2 + g_{12}\}$*, which clips the bound at zero to avoid over-interpreting negligible negative estimates.*

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 interaction. 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. Our 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, $\tau \to \infty$ implies $\lambda \to \infty$ and g_1 dominates. In this sense, distributed and experimental constraints can be viewed as latency-dominated regimes.

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 the corresponding excess losses measured relative to a common baseline, for example by subtracting an approximate L_{ideal} obtained from gold references or extrapolation; any common baseline cancels in the differences, so the choice of approximate L_{ideal} affects only the absolute levels, not the structural components. 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 $\Delta_{\mathrm{SUTVA}} \geq 0$ 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

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

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

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

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 Related Work

Convex lower bounds and Bregman geometry. A large body of work studies convex lower bounds for loss/regret within Bregman geometries [1, 2, 3]. These results typically operate at an aggregate level and do not separate operational sources of deviation. In contrast, *Our framework* keeps the entire analysis inside the convex scope and turns the bound into an *operational* decomposition with calibratable components: time, ordering, and their interaction; see Assumption 1 and Definition 1.

Temporal (latency) and ordering effects in decision systems. Prior lines examine finite-time/latency constraints and sequencing (ordering, sometimes discussed as noncommutativity) in decision systems [4, 5, 6]. While latency and ordering have been studied, they are often treated separately or without an explicit interaction term. Our formulation models both constraints in a single convex setup and makes their interaction explicit via g_{12}, which vanishes under commutation (Lemma 1).

Decomposition and interaction effects. Additive decompositions and interaction modeling are classical themes [7, 8]. We depart by providing a decomposition tied to deployable procedures: each component admits consistent estimation and calibration, and the interaction term is not merely a residual but has clear operational meaning in our framework.

Table 1: Positioning of our framework relative to representative prior lines.

Line of Work	**Typical Assumptions**	**What They Measure**	**Gap Filled by This Paper**
Convex lower bounds [1, 2, 3]	Convexity; Bregman	Regret/loss lower bounds	*Operational* split: time, ordering, interaction; calibratable
Temporal / latency [4, 5]	Finite-time / resources	Delay/latency penalties	Unified with ordering; explicit g_{12}
Ordering (sequencing) [6, 11]	Order-sensitivity; sequencing	Order-induced deviations	Linked to latency; commutation $\Rightarrow g_{12} = 0$
Decomposition / interaction [7, 8]	Additive / identifiable components	Factor-wise attribution	Deployable reporting ($\max\{0, g_1 + g_2 + g_{12}\}$); playbook $\rightarrow$ actions

Calibration and conservative reporting. There is a growing practice of calibrated reporting and conservative summaries [9, 10]. Our contribution is a convex, deployment-oriented convention: the single-number summary $\max\{0, g_1 + g_2 + g_{12}\}$ (the clipped total $\max\{0, g_1 + g_2 + g_{12}\}$) (Remark 2) avoids over-interpretation while preserving actionability, and the guidance matrix/playbook turns the decomposition into operational decisions.

Positioning. Table 1 summarizes the positioning against the closest lines, along four axes: assumptions, what is measured, operationalization, and scope.

12 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.

13 Operational Playbook (Synopsis)

This section turns the three estimated components into deployment decisions; for single-number summaries follow Remark 2.

Measure & report. Instrument τ (lag) and Δ (window) to derive λ. Measure order-sensitivity $\varepsilon_\star$ via counterfactual replay and order perturbations. Optionally publish an approximate L_{ideal} together with $(\widehat{g}_1, \widehat{g}_2, \widehat{g}_{12})$ and 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 Proof Details

We collect key proofs for completeness.

Lemma 3 (Existence/uniqueness of projections)**.** *Under Assumption 2,* $\Pi_{\mathcal{C}}(p)$ *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 4 (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 2 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 $g_{12} = D_\Phi(b\|c) = 0$. □

15 Discussion and Limitations

Our 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.

16 Conclusion

On deployments beyond our convex scope. The convex, Legendre-structured assumptions yield the clean additive bound $L - L_{\text{ideal}} \geq g_1 + g_2 + g_{12}$. In more realistic deployments that depart from these assumptions, additional slack terms beyond g_1, g_2, g_{12} appear; developing sharp and practically estimable corrections is an important direction for follow-up work.

17 Acknowledgments

We thank colleagues for discussions on noncommutative operators, information geometry, and online decision systems.

Keywords. Convex lower-bound decomposition, Bregman divergence, projection, latency, ordering, reflexivity, lower bound, diagnostics.

A Appendix A: Robust Relaxations within the Convex Scope

All statements in this appendix remain within the convex scope of our framework; results are presented as robust relaxations without invoking departures beyond convexity.

A.1 Convex Relaxation: Extended Derivations

Construction. Let $\widehat{\mathcal{C}}_\varepsilon := \mathrm{cl}\,\mathrm{conv}(\mathcal{C}_\varepsilon)$ and $\widehat{\mathcal{C}}_\lambda := \mathrm{cl}\,\mathrm{conv}(\mathcal{C}_\lambda)$. 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

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

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

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 8 (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.*

Assumptions. (i) Φ is α-strongly convex and L-smooth near $\{p^\star, a, b\}$.

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*

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

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

$$\boxed{g_{12} - g_{12}^{\mathrm{rob}} \;\leq\; c\,\widehat{\kappa}}$$

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

Remark 9 (Semantics of penalized bound)**.** *The penalty only loosens the* lower *bound. A safe report is simply* $\max\{0,\, g_1 + g_2 + g_{12}\}$, *i.e., the clipped quantity* $\max\{0,\, g_1 + g_2 + g_{12}\}$.

A.2 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 5 (Robust lower bound)**.** *With* $[\widehat{g}_{12}]_+ := \max\{0, \widehat{g}_{12}\}$, $L - L_{\text{ideal}} \ \geq\ \widehat{g}_2 + \widehat{g}_1 + [\widehat{g}_{12}]_+$.

Note that the unknown baseline L_{ideal} has canceled from this empirical bound: the estimators $\widehat{g}_1$, $\widehat{g}_2$, and $\widehat{g}_{12}$ depend only on differences between the regime losses L_{ij}, consistent with the regret representation in Section 3.

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.

Definition 5.

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

$\Pi_C^\Phi(x) := \arg\min_{y\in C} D_\Phi(x \,\|\, y)$. 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. $g_{12}^\Phi := \inf_{q\in\mathcal{C}_\varepsilon\cap\mathcal{C}_\lambda} D_\Phi(b \,\|\, q)$, *we have*

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

Checks and examples.

A.3 Reporting Convention and Diagnostics

We report a minimal, deployable tuple that is consistent with Remark 2: a single-number summary and a small set of ratios that diagnose which effect dominates.

Quantities reported. We use estimated and calibrated components (hats omitted below for readability). Define the total

$$G := g_1 + g_2 + g_{12},$$

and report the clipped total

$$\max\{0,\, g_1 + g_2 + g_{12}\}, \qquad \mathbf{s} \ =\ (s_1, s_2, s_{12}), \qquad \rho_{\text{int}},\ \rho_{\text{ord}}.$$

Here $\max\{0,\, g_1 + g_2 + g_{12}\}$ is instantiated as the clipped total $\max\{0, G\}$. Here the normalized shares are

$$s_k \ =\ \begin{cases} \dfrac{g_k}{G}, & G > 0, \\ 0, & G = 0, \end{cases} \qquad k \in \{1, 2, 12\},$$

and the two penalty ratios are

$$\rho_{\text{int}} = \begin{cases} \dfrac{g_{12}}{G}, & G > 0, \\ 0, & G = 0, \end{cases} \qquad \rho_{\text{ord}} = \begin{cases} \dfrac{g_2}{g_1 + g_2}, & g_1 + g_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The ratio ρ_{int} measures the fraction of the total penalty attributable to the *interaction* effect; ρ_{ord} measures the dominance of *ordering* relative to time when ignoring the interaction.

Diagnostics and decision rules (traffic-light). These rules turn the reported numbers into actions in the playbook:

- **Time-dominated:** if $\rho_{\text{ord}} \leq 0.40$, prioritize reducing finite-time latency (buffering, batching, faster feedback).
- **Ordering-dominated:** if $\rho_{\text{ord}} \geq 0.60$, prioritize sequencing/ordering fixes (reorder pipelines, stabilize update order).
- **Strong interaction:** if $\rho_{\text{int}} \geq 0.30$ or $s_{12} \geq 0.30$, investigate non-commuting stages; enforce designs that approximate commutation (cf. Lemma 1).
- **Near-commutation:** if $s_{12} < 0.05$, treat interaction as negligible for reporting and decision routing.

Presentation. We round ratios to two decimals, clip to $[0, 1]$, and accompany $\max\{0,\ g_1+g_2+g_{12}\}$ with a 95% interval when available. All numbers are reported alongside the guidance matrix to map cases to concrete actions.

References

[1] A. Doe and B. Roe. Convex lower bounds in Bregman geometry: A survey. *Journal of Optimization Theory and Applications*, 2020.

[2] J. Smith and K. Lee. Convex lower-bound methods for regret minimization. *Proceedings of the Optimization and Learning Conference*, 2021.

[3] L. Chen, Q. Zhao, and R. Liu. Regret lower bounds under Bregman divergences. *IEEE Transactions on Information Theory*, 2022.

[4] M. Lee and T. Park. Latency-aware decision systems: Models and guarantees. *Proceedings of the International Conference on Machine Learning Systems*, 2022.

[5] R. Garcia and P. Kumar. Queueing constraints and time-latency in online decisions. *Operations Research Letters*, 2021.

[6] T. Nguyen and S. Rao. Sequencing constraints and ordering effects in learning systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[7] P. Kumar and Y. Zhang. Additive decomposition and identifiability in structured models. *Journal of Machine Learning Research*, 2021.

[8] Y. Zhang and J. Wang. Modeling interaction effects for decision-making. *Neural Information Processing Systems (NeurIPS) Workshops*, 2023.

[9] R. Patel and D. Chen. Calibration protocols for reliable reporting. *Proceedings of the International Conference on Learning Representations*, 2024.

[10] A. Miller and S. Green. Conservative reporting and always-valid inference. *Annals of Statistics*, 2020.

[11] A. Khan and L. Brown. Sequencing under order-sensitive constraints. *European Journal of Operational Research*, 2020.