PREDICTION LOSSES AND THE VALUE OF HINTS IN OPTIMISTIC MIRROR DESCENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Prediction-dependent regret bounds motivate training losses for online decision systems, but statistical calibration of a regret remainder does not ensure task improvement. We study this gap for fixed-step optimistic mirror descent with full feedback and exogenous costs. The exact positive remainder is a Bregman loss calibrated to the conditional cost mean; subtracting the prediction-dependent stability term recovers task loss. We characterize policy differences through inverse-Hessian sensitivity. At any polytope vertex exposed by a cumulative linear margin, including nonsimple vertices, all bounded hint policies have absolutely convergent loss differences and share a logarithmic regret coefficient equal to the total active barrier weight divided by the step size. On a simplex approaching a face, the controller instead retains a lower-dimensional prediction problem, with a uniform logarithmic cumulative approximation error. A paired-cost construction then yields a linear task advantage for a learned saturated policy over a learned mean predictor, even as the latter's excess remainder tends to zero. We also establish a scalar boundary-curvature transition and matching estimation bounds under shrinking prediction geometry. Together, these results identify when calibration is statistically sufficient and when the controller's surviving decision directions make the prediction target consequential.

1 INTRODUCTION

Decision-focused learning trains predictions for a downstream optimization task (Donti et al., 2017; Elmachtoub & Grigas, 2022; Mandi et al., 2024). In online optimization, a natural source of training objectives is a prediction-dependent regret bound: estimate the next cost so that the bound's error term is small. This approach raises two different questions. Does the resulting loss elicit the right statistical quantity, and does improving that loss improve the decisions made by the specified controller? The second question depends on parts of the regret decomposition that a positive surrogate may omit.

We study fixed-step optimistic mirror descent with full cost feedback. The controller maintains a base state updated from observed costs and uses a hint to choose the current action. Under exogenous costs, every hint policy has the same base trajectory. This permits direct comparisons of the policies' actual losses, including when the controller approaches the boundary of its decision set. The inverse Hessian of the mirror potential determines which changes in predictions can still change those actions.

Contributions. We first connect the exact regret decomposition to two classical Bregman prediction losses. The positive remainder is calibrated to the conditional mean; reversing its arguments gives a convex loss that matches the expected cost-induced action. The stability correction, rather than either positive loss alone, recovers the conditional task objective. Our main results then characterize the consequences for policies and learning:

- At any polytope vertex exposed by a cumulative linear margin, all bounded hint policies have absolutely convergent task-loss differences. Their common logarithmic coefficient is determined by the active barrier weights, even at nonsimple vertices and for changing cumulative cost directions (Theorem 4).
- On a simplex approaching a face with bounded cumulative tangential costs, the full controller reduces to a controller on that face with logarithmic cumulative error. Normal hint changes have summable effects, while tangential sensitivity remains (Theorem 5).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- A learned policy that saturates the sign of a sample-mean estimate has a linear task advantage over directly using that estimate as a hint. The mean predictor nevertheless attains vanishing excess reverse loss under a realizable model (Theorem 6).

Scalar power-curvature results complement this directional picture: every pair of allowed policies has a convergent loss difference even as total loss changes from bounded to logarithmic to polynomial. Matching statistical bounds show how a shrinking prediction metric can accelerate excess-risk decay without changing the ordinary rate of mean estimation. Sharp divergence-orientation comparisons and the scalar specializations are developed in the appendices.

Relation to prior work. Optimistic online learning supplies prediction-dependent regret bounds and methods for learning predictable processes (Rakhlin & Sridharan, 2013). The mean and dual-centroid identities used here are classical Bregman prediction theory (Banerjee et al., 2005; Nielsen & Nock, 2007); the convex orientation is also a Fenchel–Young loss (Blondel et al., 2020). Differentiating a regularized action map is standard in optimization layers (Amos & Kolter, 2017). SPO calibration concerns a hard optimization oracle (Elmachtoub & Grigas, 2022; Liu & Grigas, 2021), while online decision-focused learning develops regret guarantees for learned predictors under evolving decision objectives (Capitaine et al., 2025). Our contribution is the uniform comparison of actual hint policies through boundary geometry, and its connection to a learned task separation for a fixed regularized controller. The vertex proof uses log-barrier homogeneity from central-path analysis (Boyd & Vandenberghe, 2004), with cumulative exposing directions that need not converge.

2 Controller and exact regret

Let $d \geq 1$ and let $\mathcal{Z} \subseteq \mathbb{R}^d$ be nonempty, open, and convex. We use $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ for the Euclidean inner product and norm. The notation $A \succeq B$ means that the symmetric matrix $A - B$ is positive semidefinite, and I denotes the identity matrix of the required dimension. For a differentiable strictly convex potential $R : \mathcal{Z} \to \mathbb{R}$ and $u, v \in \mathcal{Z}$, define the Bregman divergence by

$$D_R(u, v) = R(u) - R(v) - \langle \nabla R(v), u - v \rangle.$$

Unless stated otherwise, $R \in C^2(\mathcal{Z})$, $\nabla^2 R(z) \succ 0$, and $\nabla R : \mathcal{Z} \to \mathbb{R}^d$ is a diffeomorphism. These assumptions ensure that the conjugate $R^*(y) = \sup_{z \in \mathcal{Z}} \{\langle y, z \rangle - R(z)\}$ is finite and C^2 on $\mathbb{R}^d$, with $\nabla R^* = (\nabla R)^{-1}$. They include quadratic potentials and the quadratic-plus-log-barrier potentials below.

At round t, the learner observes a context x_t, produces a hint $m_t \in \mathbb{R}^d$, and plays $z_t \in \mathcal{Z}$. It then observes the entire cost vector $c_t \in \mathbb{R}^d$ and incurs $\langle c_t, z_t \rangle$. Let $\mathcal{G}_t$ be the sigma-field containing all information available when the hint is chosen, including the context, base state, and any predictor randomization. The hint is $\mathcal{G}_t$-measurable. Statistical claims concern the conditional law of c_t given $\mathcal{G}_t$. All pathwise identities hold for arbitrary realized sequences. For statistical and policy comparisons, costs are exogenous to hints and actions; varying a hint leaves the conditional cost law fixed. Comparisons between policies use the same cost sequence.

For a fixed step size $\eta > 0$ and initial base $b_1 \in \mathcal{Z}$, the base states (b_t) and actions are generated by

$$z_t = \operatorname*{arg\,min}_{z \in \mathcal{Z}} \{\eta \langle m_t, z \rangle + D_R(z, b_t)\}, \qquad b_{t+1} = \operatorname*{arg\,min}_{z \in \mathcal{Z}} \{\eta \langle c_t, z \rangle + D_R(z, b_t)\}. \tag{1}$$

Equivalently, with $y_t = \nabla R(b_t)$ and $S_t = \sum_{i=1}^{t} c_i$ (with $S_0 = 0$),

$$\nabla R(z_t) = y_t - \eta m_t, \qquad y_{t+1} = y_t - \eta c_t = y_1 - \eta S_t. \tag{2}$$

Thus the base trajectory is determined by costs and initialization. Hints affect current actions, but, for a fixed cost sequence, do not affect future base states. This distinction is essential when comparing predictors.

For polyhedral decisions, affine equalities can be eliminated without changing regret. In the resulting open domain, let $s_1, \ldots, s_J$ be the positive affine constraint slacks. An admissible potential is

$$R(z) = \tfrac{1}{2} \|z\|^2 - \sum_{j=1}^{J} \mu_j \log s_j(z), \qquad \mu_j > 0. \tag{3}$$

Appendix A gives the strict-feasibility assumption, affine reduction, and proof of the mirror properties.

Let $\mathrm{cl}\,\mathcal{Z}$ denote the closure of $\mathcal{Z}$. For a horizon $T \geq 1$ and comparator $v \in \mathrm{cl}\,\mathcal{Z}$, define $\mathrm{Reg}_T(v) = \sum_{t=1}^{T} \langle c_t, z_t - v \rangle$ and

$$F_t = D_R(b_{t+1}, z_t), \qquad B_t = D_R(z_t, b_{t+1}), \qquad V_t = D_R(z_t, b_t). \tag{4}$$

We reserve *prediction-error remainder* for B_t. Neither F_t nor B_t includes the comparator or stability contribution.

Theorem 1 (Exact regret identity)**.** *For every $v \in \mathcal{Z}$,*

$$\eta\,\mathrm{Reg}_T(v) = D_R(v, b_1) - D_R(v, b_{T+1}) + \sum_{t=1}^{T} (B_t - V_t). \tag{5}$$

In particular, $\mathrm{Reg}_T(v) \leq D_R(v, b_1)/\eta + \eta^{-1} \sum_t B_t$. *For every finite boundary comparator $v \in \mathrm{cl}\,\mathcal{Z}$, the exact extension is*

$$\eta\,\mathrm{Reg}_T(v) = P_T(v) + \sum_{t=1}^{T} (B_t - V_t), \qquad P_T(v) = R^*(y_1) - R^*(y_{T+1}) - \eta \langle S_T, v \rangle. \tag{6}$$

Proof. The three-point Bregman identity and (2) give

$$\eta \langle c_t, z_t - v \rangle = D_R(v, b_t) - D_R(v, b_{t+1}) + B_t - V_t.$$

Summing proves (5). Equivalently, the usual optimistic cross term is $\eta \langle c_t - m_t, z_t - b_{t+1} \rangle = F_t + B_t$; its F_t part cancels the forward divergence in the one-step identity. Finally, the difference of comparator divergences equals $P_T(v)$ for interior v. Both sides after this cancellation are continuous affine functions of v, proving (6). No value of R at the boundary is required. □

The theorem is an exact specialization of the optimistic mirror-descent analysis. Dropping nonnegative divergences produces an upper bound, not another identity. For example, if $\nabla^2 R \succeq \lambda I$ on $\mathcal{Z}$, conjugate smoothness yields $B_t \leq \eta^2 \|m_t - c_t\|^2/(2\lambda)$. This recovers the familiar squared-error regret bound for interior comparators. For boundary comparators, the finite term $P_T(v)$ must be retained.

3 Divergence orientation determines the statistical target

Freeze the pre-round information and define

$$z_t(m) = \nabla R^*(y_t - \eta m), \qquad \Psi_t(m) = \eta^{-2} R^*(y_t - \eta m). \tag{7}$$

Bregman duality and the affine change of coordinates give two losses:

$$\begin{aligned} f_t(m, c) &= D_{\Psi_t}(m, c) = \eta^{-2} D_R(z_t(c), z_t(m)), \\ q_t(m, c) &= D_{\Psi_t}(c, m) = \eta^{-2} D_R(z_t(m), z_t(c)). \end{aligned} \tag{8}$$

Thus $F_t = \eta^2 f_t(m_t, c_t)$ and $B_t = \eta^2 q_t(m_t, c_t)$. The forward loss f_t is strictly convex in m. It is a Fenchel–Young loss with target state $z_t(c)$, scaled by η^{-2}:

$$f_t(m, c) = \eta^{-2} \big[R^*(y_t - \eta m) + R(z_t(c)) - \langle y_t - \eta m, z_t(c) \rangle \big].$$

Its gradient matches the state discrepancy. The reverse loss has a different gradient:

$$\nabla_m f_t(m, c) = -\eta^{-1} (z_t(m) - z_t(c)), \qquad \nabla_m q_t(m, c) = [\nabla^2 R(z_t(m))]^{-1} (m - c). \tag{9}$$

Theorem 2 (Conditional Bayes acts and excess risks)**.** *Suppose the conditional cost law given $\mathcal{G}_t$ has compact support. Write $\bar{c}_t = \mathbb{E}[c_t \mid \mathcal{G}_t]$. The unique unrestricted minimizer of $\mathbb{E}[q_t(m, c_t) \mid \mathcal{G}_t]$ is $\bar{c}_t$, and*

$$\mathbb{E}[q_t(m, c_t) - q_t(\bar{c}_t, c_t) \mid \mathcal{G}_t] = D_{\Psi_t}(\bar{c}_t, m). \tag{10}$$

The unique unrestricted minimizer $m^\star_{F,t}$ of the forward risk satisfies

$$\nabla \Psi_t(m^\star_{F,t}) = \mathbb{E}[\nabla \Psi_t(c_t) \mid \mathcal{G}_t], \quad \textit{or equivalently} \quad z_t(m^\star_{F,t}) = \mathbb{E}[z_t(c_t) \mid \mathcal{G}_t], \tag{11}$$

and its conditional excess risk is $D_{\Psi_t}(m, m^\star_{F,t})$.

Proof outline. Expand both conditional risks as functions of the hint and subtract their respective minima. For the forward loss, the expected state lies in the convex hull of a compact subset of $\mathcal{Z}$, hence in $\mathcal{Z}$; invertibility of z_t gives existence and uniqueness. Appendix B gives the expansions. □

The statistical identities are the classical Bregman mean and centroid identities (Banerjee et al., 2005; Nielsen & Nock, 2007). Here they identify the target of the *exact positive term in regret*. They do not identify the task-optimal hint: V_t also depends on m_t. For any predictor class containing $\bar{c}_t$, squared error and q_t share this conditional minimizer. If the class cannot represent the mean, minimizing the expected divergence in (10) can select a different approximation from least squares. Nonzero conditional variance by itself is not such misspecification.

The two orientations coincide for quadratic geometry. More generally, Proposition 7 characterizes when forward mean calibration or global reverse convexity is possible: both require a quadratic prediction potential under their stated domain assumptions. Proposition 8 bounds either loss between squared error divided by twice the largest and smallest primal Hessian eigenvalues along the corresponding dual segment. Such interior comparisons need not have uniform constants near a boundary, and do not equate restricted-class minimizers. Both statements and proofs appear in Appendix B.

4 When hints change task performance

Calibration of a positive remainder does not by itself compare policies. For that comparison, define the counterfactual task loss $\ell_t(m, c) = \langle c, z_t(m)\rangle$ and the inverse Hessian $J_t(m) = [\nabla^2 R(z_t(m))]^{-1}$. For a hint sequence $m = (m_t)_{t\geq 1}$, write $L_T(m) = \sum_{t=1}^{T} \ell_t(m_t, c_t)$. All comparisons below use the same exogenous costs and initialization.

Theorem 3 (Task sensitivity and the stability correction)**.** *For two hints m, m', put $m_u = m' + u(m - m')$ for $u \in [0, 1]$. Then*

$$\ell_t(m, c) - \ell_t(m', c) = -\eta \int_0^1 c^\top J_t(m_u)(m - m')\, du. \tag{12}$$

Let $\mathcal{M}_t$ be a nonempty compact convex hint set, let d_t be its Euclidean diameter, and let $\kappa_t = \sup_{m\in\mathcal{M}_t} \|J_t(m)\|_{\mathrm{op}}$, where $\|\cdot\|_{\mathrm{op}}$ is the Euclidean operator norm. Any two policies taking values in $\mathcal{M}_t$ satisfy

$$|L_T(m) - L_T(m')| \leq \eta \sum_{t=1}^{T} \|c_t\|\kappa_t d_t. \tag{13}$$

Moreover, with $V_t(m) = D_R(z_t(m), b_t)$,

$$q_t(m, c) - \eta^{-2} V_t(m) = \eta^{-1}\ell_t(m, c) + \eta^{-2}\big[R^*(y_t - \eta c) - R^*(y_t)\big]. \tag{14}$$

Consequently, whenever the conditional mean $\bar{c}_t$ is finite and nonzero, conditional expected task loss has no stationary point in the interior of $\mathcal{M}_t$. In particular, an interior conditional mean is not task-optimal for this controller.

Proof. Differentiating the action map gives $\nabla_m z_t(m) = -\eta J_t(m)$. Integration along the hint segment proves (12); taking norms and summing proves (13). Expanding the two primal divergences proves (14). Finally, the conditional task gradient is $-\eta J_t(m)\bar{c}_t$, which is nonzero because $J_t(m)$ is positive definite. □

Thus $\sum_t \|c_t\|\kappa_t d_t < \infty$ is a sufficient condition for every pair of allowed policies to have an absolutely convergent cumulative loss difference. The directional identity (12) can be sharper than this norm bound. The signed loss in (14) is exactly task-aligned, but need not be convex or nonnegative. Dropping stability changes its statistical target. This distinction holds even for quadratic geometry and perfectly representable conditional means.

Theorems 4 and 5 turn this criterion into a geometric distinction between a fully contracting vertex and a face with persistent tangential sensitivity. Appendix E also gives an exact deterministic face example.

5 POLICY EQUIVALENCE AT ARBITRARY POLYTOPE VERTICES

Let $P \subset \mathbb{R}^d$ be a bounded full-dimensional polytope, described by affine slacks $s_j(z) \geq 0$, $j = 1, \ldots, J$, with a strictly feasible point. Use $\mathcal{Z} = \operatorname{int} P$ and

$$R(z) = \tfrac{1}{2} z^\top Q z + \langle h, z \rangle - \sum_{j=1}^{J} \mu_j \log s_j(z), \qquad Q \succeq 0, \quad h \in \mathbb{R}^d, \quad \mu_j > 0.$$

Fix a vertex v and let $I(v) = \{j : s_j(v) = 0\}$ be its active constraints. Define their total barrier weight by $\nu_v = \sum_{j \in I(v)} \mu_j$. The following condition requires cumulative costs to expose v with a linear margin: for some $\delta > 0$ and integer $t_0 \geq 1$,

$$\langle S_t, z - v \rangle \geq \delta t \|z - v\| \qquad (z \in P,\ t \geq t_0). \tag{15}$$

Individual costs may change sign, and S_t/t need not converge.

Theorem 4 (Uniform equivalence at an exposed vertex)**.** *Suppose $\|c_t\| \leq G$ and $\|m_t\| \leq H$ for fixed finite G, H, and assume* (15)*. For fixed $\eta > 0$ and interior initialization, uniformly over all such sequences,*

$$\operatorname{Reg}_T(v) = \frac{\nu_v}{\eta} \log T + O(1), \qquad F_t + B_t = O(t^{-2}), \qquad \sup_{\|m\| \leq H} \|J_t(m)\|_{\mathrm{op}} = O(t^{-2}). \tag{16}$$

For any two allowed policies and $T > n \geq 1$,

$$\sum_{t=n+1}^{T} |\langle c_t, z_t - z'_t \rangle| \leq C/n. \tag{17}$$

Constants depend only on $P, R, v, G, H, \delta, t_0, \eta, b_1$. The vertex need not be simple; redundant active inequalities are permitted.

Proof outline. Translate v to zero, and let a_j be the gradient of active slack s_j. Write $R = R_0 + \phi$, where $R_0(x) = -\sum_{j \in I(v)} \mu_j \log(a_j^\top x)$ and ϕ is smooth near zero. For every $\lambda > 0$, logarithmic homogeneity gives $R_0(\lambda x) = R_0(x) - \nu_v \log \lambda$ and $\langle \nabla R_0(x), x \rangle = -\nu_v$.

To cover actions, base updates, and the dual segments between them, set $K = \max\{G, H\}$ and consider any hint u with $\|u\| \leq K$. Define $\theta_t(u) = \eta(S_{t-1} + u) - y_1$ and write $x_t = z_t(u)$ for its action. For a fixed $\kappa > 0$ and all sufficiently large t, the margin implies $\langle \theta_t(u), x \rangle \geq \kappa t \|x\|$ for all $x \in P$, uniformly over these hints. Choose an interior point e. The objective $\langle \theta_t(u), x \rangle + R(x)$ at e/t is $\nu_v \log t + O(1)$. Since R is bounded below, optimality first gives $\|x_t\| = O(\log t/t)$. This places x_t in a neighborhood where $\nabla\phi$ is bounded. Taking the inner product of the stationarity equation with x_t sharpens the bound:

$$\kappa t \|x_t\| \leq \langle \theta_t(u), x_t \rangle = \nu_v - \langle \nabla\phi(x_t), x_t \rangle \leq \nu_v + C\|x_t\|, \qquad \|x_t\| = O(t^{-1}).$$

The rescaled state tx_t is therefore bounded. Comparison with e/t also bounds $R_0(tx_t)$ above. Every active slack of tx_t is bounded above, so no single negative logarithm can diverge while their sum stays bounded. Consequently there are constants $\alpha_j, \beta_j > 0$ such that

$$\alpha_j \leq t s_j(x_t) \leq \beta_j \quad (j \in I(v)), \qquad \nabla^2 R(x_t) \succeq t^2 \sum_{j \in I(v)} \frac{\mu_j a_j a_j^\top}{\beta_j^2}.$$

The active normals span $\mathbb{R}^d$, including at a nonsimple vertex, so the matrix on the right is positive definite and $J_t(u) = O(t^{-2})$. Dual Taylor expansion now bounds F_t, B_t, V_t by $O(t^{-2})$. The slack bounds give $R(b_{T+1}) = \nu_v \log T + O(1)$, while $\langle y_{T+1}, b_{T+1} \rangle = O(1)$. Conjugacy and (6) yield the logarithmic regret law, and sensitivity gives the absolutely summable policy difference. Appendix F supplies the uniform constants and finite-prefix bounds. □

Logarithmic homogeneity is familiar from central-path analysis (Boyd & Vandenberghe, 2004, Chapter 11). Here the exposing direction may change every round, and the conclusion compares every bounded hint policy on that shared trajectory. The leading coefficient depends on the specified barrier, including every active weight, rather than just the dimension. The simplex and simple-vertex expansions in Appendix N provide explicit special cases.

6 A FACE RETAINS A LOWER-DIMENSIONAL PREDICTION PROBLEM

On the simplex, write $z_0 = 1 - \sum_{i=1}^{d} z_i$ and use $R(z) = -\sum_{i=0}^{d} \mu_i \log z_i$ with positive weights. Choose a proper index set $A \subset \{0, \dots, d\}$ with $0 \in A$ and $|A| \geq 2$. Let $I = A \setminus \{0\}$ denote the retained reduced coordinates and $N = \{0, \dots, d\} \setminus A$ the excluded coordinates. The face consists of distributions supported on A. Costs are reduced relative to alternative 0. Assume bounded costs and hints, and for some finite D and $\delta > 0$,

$$|S_{t,i}| \leq D \quad (i \in I), \qquad S_{t,j} \geq \delta t \quad (j \in N,\ t \geq t_0). \tag{18}$$

On the open face simplex define $R_A(u) = -\mu_0 \log(1 - \sum_{i \in I} u_i) - \sum_{i \in I} \mu_i \log u_i$. Let $y_{1,I}$ be the restriction of the initial dual state to I and define the face action and its inverse Hessian by

$$u_t(m_I) = \nabla R_A^*(y_{1,I} - \eta(S_{t-1,I} + m_I)), \qquad J_{A,t}(m_I) = [\nabla^2 R_A(u_t(m_I))]^{-1}.$$

Theorem 5 (Uniform reduction to the exposed face)**.** *Under* (18)*, uniformly over bounded hints and allowed histories,*

$$z_{t,j}(m) = \frac{\mu_j}{\eta S_{t-1,j}} + O(t^{-2})\ (j \in N), \qquad z_{t,I}(m) = u_t(m_I) + O(t^{-1}) \tag{19}$$

for sufficiently large t. In the coordinate order (I, N),

$$J_t(m) = \begin{pmatrix} J_{A,t}(m_I) + O(t^{-1}) & O(t^{-2}) \\ O(t^{-2}) & O(t^{-2}) \end{pmatrix}, \qquad L_T(m) = \sum_{t=1}^{T} \langle c_{t,I}, u_t(m_{t,I}) \rangle + O(\log(1+T)). \tag{20}$$

Matrix errors are in operator norm; all constants depend only on the fixed bounds, weights, dimension, initialization, and step size. Two policies with identical tangential hints $m_{t,I}$ have an absolutely summable loss difference, with tail $O(n^{-1})$ after round n.

Proof outline. Fix a round and hint, and define $\theta_i = \eta(S_{t-1,i} + m_i) - y_{1,i}$ for $i \geq 1$, with $\theta_0 = 0$. Let λ and λ_A be the full and face normalization multipliers, respectively. Writing $u_0 = 1 - \sum_{i \in I} u_i$ for the face probability of alternative 0, stationarity gives

$$z_i = \frac{\mu_i}{\theta_i + \lambda}, \quad \sum_{i=0}^{d} \frac{\mu_i}{\theta_i + \lambda} = 1, \qquad u_i = \frac{\mu_i}{\theta_i + \lambda_A}, \quad \sum_{i \in A} \frac{\mu_i}{\theta_i + \lambda_A} = 1.$$

Each multiplier is the unique root above its denominator poles. The retained coefficients are bounded, whereas excluded coefficients grow linearly in t. Both multipliers remain bounded, and the excluded alternatives have total probability $O(t^{-1})$. The derivative of the face normalization sum stays bounded away from zero in magnitude, giving $\lambda - \lambda_A = O(t^{-1})$. Substitution proves the action estimates in (19).

For sensitivity, partition the full Hessian into tangent block A_t, cross block C_t, and normal block D_t. The action estimates imply $A_t = \nabla^2 R_A(u_t) + O(t^{-1})$, $C_t = O(1)$, and $D_t^{-1} = O(t^{-2})$. The Schur complement $A_t - C_t D_t^{-1} C_t^\top$ differs from the face Hessian by $O(t^{-1})$. Block inversion yields (20); bounded costs turn the action error into an $O(t^{-1})$ one-round loss error. If only normal hints differ, sensitivity uses only the normal columns of J_t, all of order t^{-2}. Summing gives the sharper policy tail. Appendix G provides the uniform bounds. □

The tangent block remains uniformly positive definite because the face dual states stay bounded. Thus a linear task advantage on the face survives in the coupled full simplex, up to logarithmic error. This result identifies the surviving controller, rather than inferring predictive value solely from proximity to the boundary. The bounded tangential cumulative-cost assumption is essential to this uniform interior reduction; it does not describe every possible approach to a face.

7 A LEARNED TASK POLICY CAN OUTPERFORM A CALIBRATED PREDICTOR

We now compare two learners with the same observations and admissible hint interval. Use the product domain $(0, \infty) \times (-1, 1)$, initial state $(1, 0)$, and

$$R(s, u) = \tfrac{1}{2}(s-1)^2 - \mu \log s - \log(1 - u^2), \qquad \mu > 0.$$

Fix $M, \delta > 0$ and an unknown $\theta \in (-1, 1) \setminus \{0\}$. Let $Y_k \in \{-1, 1\}$ be independent with $\Pr(Y_k = 1) = (1 + \theta)/2$. Pair the exogenous costs as

$$c_{2k-1} = (M, \delta Y_k), \qquad c_{2k} = (M, -\delta Y_k). \tag{21}$$

Admissible policies use normal hint M and tangential hints in $[-\delta, \delta]$. Both learners use the known tangential cost $-\delta Y_k$ on even rounds. Before odd round $2k - 1$, set $\widehat{\theta}_1 = 0$ and $\widehat{\theta}_k = (k - 1)^{-1} \sum_{j<k} Y_j$ for $k \geq 2$. The *mean policy* uses tangential hint $\delta\widehat{\theta}_k$. The *saturated policy* uses $\delta \operatorname{sgn}(\widehat{\theta}_k)$, with $\operatorname{sgn}(0) = 1$. Neither knows θ; both observe each Y_k only after their odd-round action.

Define $g(x) = x/(1 + \sqrt{1 + x^2})$, the inverse gradient of $-\log(1 - u^2)$. Write L_{2K}^{mean} and L_{2K}^{sat} for their task losses over K pairs.

Theorem 6 (Learnable separation from mean calibration)**.** *For the preceding model, define*

$$\Delta_\theta = \delta|\theta|\big[g(\eta\delta) - g(\eta\delta|\theta|)\big] > 0. \tag{22}$$

Then

$$\mathbb{E}[L_{2K}^{\text{mean}} - L_{2K}^{\text{sat}}] = K\Delta_\theta + O(\log(1 + K)). \tag{23}$$

On odd rounds the mean policy has expected excess reverse loss $\Theta(k^{-1})$ relative to the conditional mean, whereas the saturated policy has bounded cumulative expected task excess relative to the best admissible predictable policy that knows θ. The linear separation also holds on a coupled three-alternative simplex.

Proof outline. Each pair resets the tangential base dual state to zero. The odd-round expected task loss at hint m is $-\delta\theta g(\eta m)$, minimized at $m = \delta \operatorname{sgn}(\theta)$. The conditional mean $\delta\theta$ instead minimizes the reverse loss. The difference between these two oracle losses is (22). The sample mean is unbiased with variance $O(k^{-1})$, so a second-order expansion bounds its expected task-risk difference from the mean oracle by $O(k^{-1})$. Its sign error probability is at most $e^{-(k-1)\theta^2/2}$. Summation proves (23) and bounded task excess; compact curvature gives the reverse-loss rate. For the simplex, take weights $(1, 1, \mu)$, equal initial probabilities on alternatives $0, 1$, reduced costs $(2\delta Y_k, M)$ and $(-2\delta Y_k, M)$, and the corresponding doubled tangential hints. Theorem 5 transfers the face comparison with an $O(\log K)$ error. Appendix H gives the complete argument. □

This separation holds despite realizability of the conditional mean. The task policy can itself be implemented by postprocessing a regression estimate. Thus the result identifies a mismatch between the calibrated target and its direct use by a fixed controller; it does not imply that regression cannot support optimal decisions.

8 Boundary curvature and the common task-loss rate

The vertex/face distinction concerns directions of sensitivity. A separate question is how fast the common loss grows when every direction contracts. Consider the half-line with initial state 1, comparator 0, and potential $\mathcal{R}(s) = \frac{1}{2}(s - 1)^2 + \mu\mathcal{B}(s)$. Assume $\mathcal{B} \in C^2(0, \infty)$ is convex and $\mathcal{B}''(s) \sim c_0 s^{-q}$ as $s \downarrow 0$, with $\mu, c_0 > 0$ and $q > 1$. Fix $0 < a < b$, let $c_t, m_t \in [aM, bM]$, and define $r = 1/(q - 1)$. For sufficiently large M, Theorems 16 and 17 establish uniformly that

$$\begin{aligned} F_t, B_t = O(M^{1-r} t^{-r-1}), \qquad L_T = \Theta\left(M^{1-r} \sum_{t=1}^{T} t^{-r}\right), \\ \sum_{t>n} c_t |z_t - z_t'| = O(M^{1-r}(1 + n)^{-r}) \qquad (n \geq 0). \end{aligned} \tag{24}$$

Thus every pair of allowed policies has an absolutely convergent loss difference for every $q > 1$. Nevertheless, total loss is bounded for $1 < q < 2$, logarithmic for $q = 2$, and polynomial of order $T^{(q-2)/(q-1)}$ for $q > 2$. These conclusions retain arbitrary admissible cost sequences and fixed controller parameters.

The difference between the loss and sensitivity exponents follows directly from the scalar state equation. For cumulative argument $H \geq 0$, define $s(H) = (\mathcal{R}')^{-1}(\mathcal{R}'(1) - \eta H)$. Integrating the

curvature assumption gives $\mathcal{R}'(s) \sim -\mu c_0 s^{-(q-1)}/(q-1)$ at zero. With $C_* = (\mu c_0/((q-1)\eta))^r$, inversion and differentiation of the state equation yield

$$s(H) \sim C_* H^{-r}, \qquad -s'(H) = \frac{\eta}{\mathcal{R}''(s(H))} = \Theta(H^{-r-1}) \qquad (H \to \infty).$$

The action is $z_t = s(S_{t-1} + m_t)$. Its argument is of order Mt, whereas two allowed hints differ by at most $(b-a)M$. The derivative therefore contributes an additional factor t^{-1} to policy differences compared with the loss itself.

The same estimate gives an integral approximation that preserves the realized cumulative costs. For integers $T > n \geq 1$, uniformly over allowed policies,

$$L_T - L_n = \int_{S_n}^{S_T} s(H)\, dH + O\big(M^{1-r} n^{-r}\big)\,.$$

Indeed, on each cumulative-cost interval the rectangle used by the controller differs from the integral by $O(M^{1-r}t^{-r-1})$. Summing these errors proves the claim. At fixed M, it yields the leading equivalences $L_T \sim C_* S_T^{1-r}/(1-r)$ for $r < 1$ and $L_T \sim C_* \log S_T$ for $r = 1$. For $r > 1$, define $L_\infty = \sum_{t \geq 1} c_t z_t$; the remaining loss satisfies $L_\infty - L_T \sim C_* S_T^{1-r}/(r-1)$. Thus even without convergence of average costs, all allowed policies share the leading loss or tail along the same cumulative-cost path.

For the exact log potential $\mathcal{B}(s) = -\log s$, Theorem 12 strengthens the order statement to $L_T = (\mu/\eta)\log T + O(1)$, with a constant uniform in the allowed cost scale. The positive remainder and every policy difference are summable even when ordinary cumulative squared prediction error grows linearly. The boundary comparator contribution supplies the common growth; discarding the first stability term would incorrectly introduce a logarithmic scale penalty. Appendices I–M retain the explicit constants, integral approximation, curvature transition, and sharp divergence-orientation comparisons in Appendix C.

9 Statistical optimality under a shrinking loss

Return to the exact half-line log potential with initial state 1. Let $c_t = MX_t$, where $X_t \in [a, b]$ are independent and identically distributed with mean $\bar{x}$ and variance $v > 0$. Assume $\eta a M \geq 2\max\{|\mu - 1|, \sqrt{\mu}\}$. For the history $\mathcal{F}_{t-1} = \sigma(X_1, \ldots, X_{t-1})$, define the conditional reverse risk $\mathcal{Q}_t(m) = \mathbb{E}[q_t(m, c_t) \mid \mathcal{F}_{t-1}]$. The sample-mean hint $m_t = S_{t-1}/(t-1)$, for $t \geq 2$, satisfies

$$\frac{\mu v}{16\eta^2 b^2 t^2 (t-1)} \leq \mathbb{E}[\mathcal{Q}_t(m_t) - \mathcal{Q}_t(M\bar{x})] \leq \frac{2\mu v}{\eta^2 a^2 t^2 (t-1)}. \tag{25}$$

Theorem 13 proves these constants, and Theorem 14 proves a matching minimax order over unknown two-point laws on $\{a, b\}$. The cumulative excess after round n is $\Theta(n^{-2})$. The argument sandwiches the random prediction curvature pathwise by constants times $(Mt)^{-2}$ and combines it with squared estimation error of order M^2/t. This avoids any independence assumption between the estimator and the history-dependent geometry. The statistical rate t^{-3} therefore reflects ordinary mean estimation under a shrinking metric. The face example retains order-one tangent curvature, where mean estimation instead has the usual t^{-1} excess rate and the choice of target changes leading task performance.

10 Numerical evidence and reproducibility

We test the mechanisms on controlled instances with known mathematical predictions. The supplement contains a single command to reproduce all figures, trajectory data, and machine-readable checks; Appendices P and Q specify every protocol. All comparisons keep costs, initialization, and controller parameters fixed.

Figure 1 checks three distinct consequences: the active-weight coefficient beyond simple vertices, the surviving face controller, and a linear difference between two learned policies. For the vertex experiment, four facets meet at the apex of a three-dimensional pyramid. Their barrier weights sum to 2.6, and $\eta = 0.4$, so Theorem 4 predicts the coefficient 6.5. We use 4096 rounds with alternating

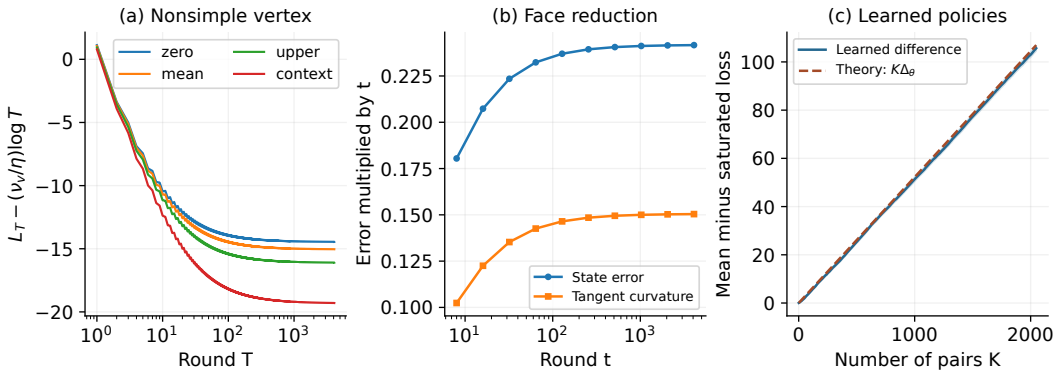

Figure 1: From geometric contraction to learned task improvement. Left: task-loss residual at a nonsimple pyramid vertex after subtracting the predicted logarithmic term. Middle: reduction error on a coupled simplex face, scaled by the horizon, for bounded tangential cost sums. Right: learned mean minus saturated-policy task loss on the coupled simplex; the line is the analytically predicted leading term and the band is two standard errors across independent cost streams. No asymptotic coefficient is fitted.

signed costs satisfying the cumulative margin. Four hint policies, including a parity-aware predictor that knows the current cost, give final doubling-block slopes between 6.458 and 6.482. Their bounded residuals are consistent with a shared leading term despite different predictive information.

The face experiment uses a three-alternative simplex with weights $(1, 1, 0.7)$ and $\eta = 0.5$. Tangential cumulative costs stay fixed while normal cumulative costs grow linearly. Doubling the horizon approximately halves the action-reduction error and quarters both cross-block sensitivity and the response to a change in the normal hint. These different rates distinguish approximation by the face controller from sensitivity to excluded alternatives.

For learned policies, we use the same simplex weights and step size, with $\theta = 0.35$, $\delta = 1$, and $M = 2$ in the paired-cost construction. Across 64 independent streams of 2048 pairs, the mean policy incurs 105.673 more task loss than the saturated policy, with standard error 0.540. The theoretical leading difference is $\Delta_\theta = 0.05222974$ per pair. Both learners see the same costs within each stream and estimate the unknown mean from previous pairs only; neither receives its true sign. Independent binomial enumeration in the product model checks the predicted gap without Monte Carlo error. The plotted uncertainty band is pointwise, rather than simultaneous across horizons. No hyperparameter is selected using these outcomes. The supplement includes the stream-level losses and random seeds, allowing the observed finite-horizon difference and its uncertainty to be recomputed independently. Additional scalar, simplex, and exact-arithmetic checks appear in the appendix. These experiments illustrate the scope of the theorems and check implementations; they use synthetic costs rather than an application benchmark.

11 Conclusion

A prediction loss extracted from a regret bound can be correctly calibrated while targeting a different hint from the one that minimizes task loss. Inverse-Hessian sensitivity determines the consequence of this difference. A cumulative vertex margin suppresses every direction, yielding a common leading regret for all bounded hints. A face can retain a learning problem on its tangent space, where target selection produces a linear task difference. The statistical rate must therefore be read together with the controller geometry and the full regret decomposition.

The analysis assumes fixed step size, full feedback, and shared exogenous costs. The face reduction requires bounded tangential cumulative costs; the learned separation uses a specified paired-cost process. Costs affected by actions, partial feedback, adaptive controller parameters, and more general approaches to a face require additional analysis. The experiments test the derived mechanisms on controlled instances, and application-level performance remains a separate empirical question.

REPRODUCIBILITY STATEMENT

The appendices provide complete supporting proofs and deterministic numerical protocols. One command reproduces all three figures and the checks described in Appendices P and Q; no application data are used.

AI USE STATEMENT

OpenAI Codex assisted with mathematical derivation and checking, literature searches, writing, LaTeX, and numerical code. The authors are responsible for verifying the final claims, references, and numerical descriptions.

REFERENCES

Brandon Amos and J. Zico Kolter. OptNet: Differentiable optimization as a layer in neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 136–145. PMLR, 2017. URL https://proceedings.mlr.press/v70/amos17a.html.

Arindam Banerjee, Xin Guo, and Hui Wang. On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, 2005. doi: 10.1109/TIT.2005.850145. URL https://www.dam.brown.edu/people/huiwang/research/bgw.pdf.

Heinz H. Bauschke and Jonathan M. Borwein. Joint and separate convexity of the Bregman distance. In Dan Butnariu, Yair Censor, and Simeon Reich (eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, volume 8 of *Studies in Computational Mathematics*, pp. 23–36. Elsevier, 2001. doi: 10.1016/S1570-579X(01)80004-5.

Heinz H. Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2017. doi: 10.1287/moor.2016.0817.

Mathieu Blondel, André F. T. Martins, and Vlad Niculae. Learning with Fenchel–Young losses. *Journal of Machine Learning Research*, 21(35):1–69, 2020. URL https://www.jmlr.org/papers/v21/19-021.html.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL https://web.stanford.edu/~boyd/cvxbook/.

Aymeric Capitaine, Maxime Haddouche, Eric Moulines, Michael I. Jordan, Etienne Boursier, and Alain Durmus. Online decision-focused learning, 2025. URL https://arxiv.org/abs/2505.13564. Version 3, revised March 2026.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Adam N. Elmachtoub and Paul Grigas. Smart predict, then optimize. *Management Science*, 68(1): 9–26, 2022. doi: 10.1287/mnsc.2020.3922.

Heyuan Liu and Paul Grigas. Risk bounds and calibration for a smart predict-then-optimize method. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80:1623–1701, 2024. doi: 10.1613/jair.1.15320.

Frank Nielsen and Richard Nock. On the centroids of symmetrized Bregman divergences, 2007. URL https://arxiv.org/abs/0711.3242.

Max Nilsson and Pontus Giselsson. The symmetry coefficient of positively homogeneous functions. *Optimization Letters*, 20:917–953, 2026. doi: 10.1007/s11590-025-02266-6.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 993–1019. PMLR, 2013. URL https://proceedings.mlr.press/v30/Rakhlin13.html.

A CONTROLLER GEOMETRY AND EXACT IDENTITIES

This appendix verifies the controller assumptions and records the finite comparator calculation. Appendix B treats calibration and convexity; Appendix D proves the log-barrier comparison; Appendix K gives the uniform half-line and learning bounds; Appendix M proves the general-curvature results and history-dependent targets. Appendices E–H prove task sensitivity, the arbitrary-vertex theorem, face reduction, and the learned separation. Appendix O supplies explicit simplex and simple-vertex expansions. Appendices P and Q describe the numerical checks.

A.1 EXISTENCE, CONJUGACY, AND DIFFERENTIATION

Under the assumptions of Section 2, for every $y \in \mathbb{R}^d$ there is a unique $z_y \in \mathcal{Z}$ such that $\nabla R(z_y) = y$. Strict convexity implies

$$R(z) \geq R(z_y) + \langle y, z - z_y \rangle \qquad (z \in \mathcal{Z}).$$

Thus z_y attains the finite conjugate value $R^*(y) = \langle y, z_y \rangle - R(z_y)$. The inverse function theorem yields

$$\nabla R^*(y) = z_y, \qquad \nabla^2 R^*(y) = [\nabla^2 R(z_y)]^{-1} \succ 0.$$

The objectives in (1) differ by a constant from $R(z) - \langle y_t - \eta m_t, z \rangle$ and $R(z) - \langle y_t - \eta c_t, z \rangle$. Their unique minimizers therefore satisfy (2). In particular, for every counterfactual hint m,

$$\nabla_m z_t(m) = -\eta[\nabla^2 R(z_t(m))]^{-1}, \qquad \nabla^2 \Psi_t(m) = [\nabla^2 R(z_t(m))]^{-1}. \tag{26}$$

For the potential (3), write $s_j(z) = \sigma_j - a_j^\top z$. Direct differentiation gives

$$\nabla^2 R(z) = I + \sum_j \frac{\mu_j a_j a_j^\top}{s_j(z)^2} \succeq I.$$

For any fixed y, the function $R(z) - \langle y, z \rangle$ tends to infinity as $\|z\| \to \infty$ within $\mathcal{Z}$: each affine slack is bounded above by a constant times $1 + \|z\|$, so the negative logarithms grow in magnitude at most logarithmically, while the positive quadratic term grows quadratically. On bounded subsets, approaching a finite boundary point sends at least one slack to zero, while the other slacks are bounded above. Thus $R(z) - \langle y, z \rangle$ also tends to infinity at the finite boundary. Its sublevel sets are compact subsets of $\mathcal{Z}$, giving an interior minimizer. This proves surjectivity of ∇R; strict convexity gives injectivity and the inverse function theorem gives the required diffeomorphism. The same argument allows any added affine term in R, which does not change its Bregman divergence.

A.2 AFFINE REDUCTION AND BASIS INVARIANCE

For $\bar{W} = \{w : Aw = d_0,\ Hw \leq r_0\}$, assume a point w° satisfies the equalities and every displayed inequality strictly. Let U have orthonormal columns spanning $\ker A$. Writing $w = w^\circ + Uz$ gives an open reduced domain $\mathcal{Z} = \{z : s_j(z) > 0\ \forall j\}$ with J affine slacks $s_j(z) = (r_0 - H(w^\circ + Uz))_j$.

Under this strict-feasibility assumption, suppose $\ker A$ has positive dimension. Every point in the equality subspace has a unique representation $w = w^\circ + Uz$. For $u = w^\circ + Uv$,

$$\langle c_t, w_t - u \rangle = \langle U^\top c_t, z_t - v \rangle.$$

Hence cumulative regret depends only on the reduced cost. A component of the hint orthogonal to $\ker A$ adds a constant to the optimization objective and has no effect on the action. If $U' = UO$ with O orthogonal, then $z' = O^\top z$, the transformed potential is $R'(z') = R(Oz')$, and

$$D_{R'}(u', v') = D_R(Ou', Ov').$$

The action in the original coordinates, both divergence orientations, and the regret identity are invariant under this change of basis. When $\ker A = \{0\}$, the feasible equality subspace is a singleton and regret is identically zero; no reduced prediction problem remains.

A.3 Complete one-step identity

For interior v, expand the comparator difference using conjugacy:

$$D_R(v, b_t) - D_R(v, b_{t+1}) = R^*(y_t) - R^*(y_{t+1}) - \langle y_t - y_{t+1}, v\rangle.$$

Similarly,

$$B_t - V_t = R^*(y_{t+1}) - R^*(y_t) + \langle y_t - y_{t+1}, z_t\rangle.$$

Addition and $y_t - y_{t+1} = \eta c_t$ give the one-step equality in Theorem 1. The optimistic form is recovered from

$$\begin{aligned} F_t + B_t &= \langle \nabla R(z_t) - \nabla R(b_{t+1}), z_t - b_{t+1}\rangle \\ &= \eta\langle c_t - m_t, z_t - b_{t+1}\rangle. \end{aligned}$$

Thus the forward divergence cancels exactly before any norm inequality is applied.

For $v \in \operatorname{cl}\mathcal{Z}$, take $v_\epsilon = (1-\epsilon)v + \epsilon b_1$ with $0 < \epsilon \le 1$. Open convexity implies $v_\epsilon \in \mathcal{Z}$. The *difference*

$$D_R(v_\epsilon, b_1) - D_R(v_\epsilon, b_{T+1}) = R^*(y_1) - R^*(y_{T+1}) - \eta\langle S_T, v_\epsilon\rangle$$

has a finite limit. The left side of the regret identity also converges, being a finite sum of linear functions of v_ϵ. Taking $\epsilon \downarrow 0$ proves (6). Individual divergences at v are neither defined nor subtracted in this argument. The useful boundary upper bound is

$$\operatorname{Reg}_T(v) \le \frac{P_T(v)}{\eta} + \frac{1}{\eta}\sum_t B_t.$$

In particular, a summable error remainder does not imply bounded regret unless the comparator contribution is also controlled.

A.4 Optimistic FTRL and the scope of the surrogate connection

Equation (2) gives

$$z_t = \operatorname*{arg\,min}_{z\in\mathcal{Z}} \left\{ \langle S_{t-1} + m_t, z\rangle + \frac{R(z) - \langle y_1, z\rangle}{\eta} \right\}.$$

This is optimistic follow-the-regularized-leader with the indicated fixed regularizer. The Fenchel–Young representation in Section 3 follows from $R(z_t(c)) + R^*(y_t - \eta c) = \langle y_t - \eta c, z_t(c)\rangle$. It establishes a connection to regularized prediction losses (Blondel et al., 2020). It does not identify this particular loss with SPO+, nor does it establish a perturbation distribution realizing an arbitrary barrier map. Those are additional representation questions not used by any result here.

B Calibration, convexity, and approximation

Proposition 7 (Limits of simultaneous calibration and convexity)**.** *Let Ψ have positive-definite Hessian. On an open convex domain, $D_\Psi(m, c)$ has the arithmetic mean as Bayes act for every finitely supported cost distribution if and only if Ψ is quadratic, up to an affine term. If $\Psi \in C^3(\mathbb{R}^d)$, then $D_\Psi(c, m)$ is convex in m for every $c \in \mathbb{R}^d$ if and only if Ψ is quadratic.*

For the first claim, two-point distributions force $\nabla\Psi$ to preserve convex combinations, hence to be affine. For the second, $\nabla^2_m D_\Psi(c, m) = \nabla^2\Psi(m) + \nabla^3\Psi(m)[m - c]$. Allowing unrestricted c forces the third derivative to vanish. The proofs below supply the details. The unrestricted cost domain is essential: a nonquadratic reverse loss can be convex on a bounded hint and cost range.

Proposition 8 (Comparison on a compact interior region)**.** *Fix t, m, c and suppose $\lambda I \preceq \nabla^2 R(z_t((1-u)c + um)) \preceq \Lambda I$ for all $u \in [0, 1]$, with $0 < \lambda \le \Lambda$. Then both prediction losses satisfy*

$$\frac{\|m - c\|^2}{2\Lambda} \le f_t(m, c),\ q_t(m, c) \le \frac{\|m - c\|^2}{2\lambda}. \tag{27}$$

If all endpoint states belong to a common compact $K \subset \mathcal{Z}$, these bounds hold uniformly with the Hessian extrema on $K_{\mathrm{dual}} = (\nabla R)^{-1}(\operatorname{conv}(\nabla R(K)))$.

The proof below uses the inverse-Hessian identity and integral Taylor expansion. These are comparisons of loss values, not equality of restricted-class minimizers. Horizon-dependent compact sets need not give uniform constants near a boundary.

B.1 Duality and gradients

For $u, v \in \mathcal{Z}$, expansion of the conjugates gives

$$D_R(u, v) = D_{R^*}(\nabla R(v), \nabla R(u)).$$

If $\Psi(m) = \eta^{-2} R^*(y - \eta m)$, then for any m, c,

$$D_\Psi(m, c) = \eta^{-2} D_{R^*}(y - \eta m, y - \eta c).$$

Substitution of the played and updated states proves both equalities in (8). Since $\nabla\Psi(m) = -\eta^{-1} z_t(m)$,

$$\nabla_m D_\Psi(m, c) = \nabla\Psi(m) - \nabla\Psi(c).$$

For the reverse orientation, differentiating the definition cancels the two first-derivative terms:

$$\nabla_m D_\Psi(c, m) = \nabla^2\Psi(m)(m - c).$$

Equation (26) gives (9). These formulas hold on the frozen pre-round state; no differentiation through a counterfactual future trajectory is being taken.

B.2 Bayes risks and existence of the forward minimizer

Fix a realization of $\mathcal{G}_t$ for which the conditional support is compact. All expectations below are with respect to that conditional distribution. Let $\bar{c} = \mathbb{E}c$ and $a = \mathbb{E}\nabla\Psi(c)$. Continuity on the compact support makes the relevant moments finite. Direct expansion yields

$$\mathbb{E}D_\Psi(c, m) = \mathbb{E}\Psi(c) - \Psi(m) - \langle \nabla\Psi(m), \bar{c} - m\rangle,$$
$$\mathbb{E}D_\Psi(m, c) = \Psi(m) - \langle a, m\rangle + \mathbb{E}\big[\langle \nabla\Psi(c), c\rangle - \Psi(c)\big].$$

The first line minus its value at $\bar{c}$ equals $D_\Psi(\bar{c}, m)$. It is strictly positive unless $m = \bar{c}$.

For the second line, the map z_t sends the compact support to a compact set $K \subset \mathcal{Z}$. In finite dimension, $\operatorname{conv} K$ is compact and is contained in $\mathcal{Z}$ by convexity. The expectation $\mathbb{E}z_t(c)$ belongs to $\operatorname{conv} K$. Since $z_t : \mathbb{R}^d \to \mathcal{Z}$ is a bijection, there is a unique $m_F^\star$ with $z_t(m_F^\star) = \mathbb{E}z_t(c)$. Equivalently $\nabla\Psi(m_F^\star) = a$. The forward risk difference is

$$\Psi(m) - \Psi(m_F^\star) - \langle \nabla\Psi(m_F^\star), m - m_F^\star\rangle = D_\Psi(m, m_F^\star).$$

This proves Theorem 2, including existence and uniqueness. The argument is conditional and therefore remains valid with random contexts and random pre-round geometry. It is the application of established Bregman prediction and centroid identities (Banerjee et al., 2005; Nielsen & Nock, 2007) to the exact regret remainder.

B.3 Proof of Proposition 7

First suppose the forward loss has the arithmetic mean as Bayes act for every finitely supported distribution on an open convex set U. For $x, z \in U$ and $0 < \lambda < 1$, take probabilities $\lambda, 1 - \lambda$ at these points. The mean is in U, so the first-order condition is necessary:

$$\nabla\Psi(\lambda x + (1 - \lambda)z) = \lambda\nabla\Psi(x) + (1 - \lambda)\nabla\Psi(z).$$

The continuous map $\nabla\Psi$ preserves convex combinations on U and is therefore affine. Its Jacobian is a constant symmetric positive-definite matrix H. Integration gives $\Psi(m) = \frac{1}{2} m^\top H m + \langle \ell, m\rangle + d$ on U. Conversely, this potential gives $D_\Psi(m, c) = \frac{1}{2}(m - c)^\top H(m - c)$, whose Bayes act is the mean.

For the second claim, assume $\Psi \in C^3(\mathbb{R}^d)$ and $m \mapsto D_\Psi(c, m)$ is convex for every $c \in \mathbb{R}^d$. For fixed vectors v, w and scalar λ, take $c = m + \lambda w$. Positive semidefiniteness of the Hessian gives

$$v^\top \nabla^2\Psi(m) v - \lambda \nabla^3\Psi(m)[v, v, w] \geq 0 \qquad \text{for every } \lambda \in \mathbb{R}.$$

Hence $\nabla^3\Psi(m)[v, v, w] = 0$. For fixed w, polarization of the symmetric bilinear form in the first two arguments gives $\nabla^3\Psi(m)[u, v, w] = 0$ for all u, v. Thus $\nabla^3\Psi$ vanishes everywhere, and connectedness gives a constant Hessian and the asserted quadratic form. The converse follows from the same squared Mahalanobis expression. On a bounded cost domain λ cannot vary arbitrarily, so the necessity argument no longer applies. Separate convexity of Bregman divergences has a broader established literature (Bauschke & Borwein, 2001).

B.4 COMPACT-INTERIOR COMPARISONS

Proof. The Hessian of Ψ_t is $[\nabla^2 R(z_t(m))]^{-1}$. Inverting the assumed matrix inequalities bounds it between $\Lambda^{-1} I$ and $\lambda^{-1} I$ along the prediction segment. For either ordering of m, c, apply

$$D_{\Psi_t}(x, y) = \int_0^1 (1-u)(x-y)^\top \nabla^2 \Psi_t(y + u(x-y))(x-y)\, du.$$

The integral of $1-u$ is $1/2$, giving (27). For the uniform assertion, every interpolated dual point lies in $\mathrm{conv}(\nabla R(K))$. This convex hull is compact in finite dimension; continuity of $(\nabla R)^{-1}$ makes K_{dual} compact in $\mathcal{Z}$. Continuity and positive definiteness of $\nabla^2 R$ give the required extrema. □

The curvature condition concerns the inverse image of the dual segment, which need not be the straight segment between the primal endpoint states. This is why endpoint Hessian bounds alone are insufficient and the proposition uses K_{dual}. If $R(z) = \frac{1}{2} z^\top H z + \langle \ell, z \rangle + d$ on $\mathbb{R}^d$ with $H \succ 0$, then

$$f_t(m, c) = q_t(m, c) = \tfrac{1}{2}(m-c)^\top H^{-1}(m-c).$$

For $H = I$ both equal half squared error. Even in this quadratic case, equivalence of loss values does not ensure equality of restricted minimizers. For example, let $H^{-1} = \mathrm{diag}(1, 4)$, let $c = 0$ deterministically, and restrict hints to $\{(1, 0), (0, 3/4)\}$. Squared error selects $(0, 3/4)$, with error $9/16 < 1$, whereas the weighted loss selects $(1, 0)$, with value $1/2 < 9/8$. The common unrestricted minimizer remains 0.

B.5 WHAT MEAN CALIBRATION DOES AND DOES NOT IMPLY

The conditional excess risk of a predictor g is exactly

$$D_{\Psi_t}(\bar{c}_t, g(x_t)),$$

provided the predictor is measurable with respect to the conditioning information. The squared-error excess risk is $\frac{1}{2}\|\bar{c}_t - g(x_t)\|^2$. If the class contains the conditional mean, both are minimized there. If not, the two objectives project the target onto the class using different losses; neither projection equality nor a task-level ordering follows. Likewise, learning a conditional mean given only x_t need not represent the conditional mean given the full $\mathcal{G}_t$. The choice of predictor information is part of the realizability assumption.

For an unregularized linear oracle on a decision set with an attained minimum and finite expected losses, any minimizer of $\langle \mathbb{E}[c \mid x], w \rangle$ minimizes conditional expected task loss by linearity. That observation is distinct from Theorem 2, which concerns a regularized controller's positive regret remainder. In the half-line model, the fixed action 0 already minimizes every realized positive-cost linear objective over $[0, \infty)$. Theorems 11 and 12 describe the behavior of the specified interior controller relative to that action, rather than a minimax lower bound for online decision making.

C SHARP COMPARISONS BETWEEN THE TWO ORIENTATIONS

Using the convex forward loss to control regret requires a bound on B_t/F_t. We give local versions of a Bregman symmetry coefficient (Bauschke et al., 2017; Nilsson & Giselsson, 2026). They depend on the variation of curvature along the relevant states, rather than its absolute size.

Theorem 9 (Sharp directional-curvature comparison)**.** *Let $a, b \in \mathcal{Z}$, $a \neq b$, $\delta = a - b$, and suppose*

$$0 < \underline{h} \leq \delta^\top \nabla^2 R(b + u\delta)\delta \leq \overline{h} \quad \textit{for } 0 \leq u \leq 1.$$

Then

$$\sqrt{\underline{h}/\overline{h}} \leq \frac{D_R(a, b)}{D_R(b, a)} \leq \sqrt{\overline{h}/\underline{h}}. \tag{28}$$

The constants cannot be improved using only these two curvature bounds.

The square-root extremum also occurs for a piecewise-quadratic homogeneous potential in Nilsson & Giselsson (2026, Remark 3.1); the statement here bounds every segment from its directional curvature range.

Proof. Set $h(u) = \delta^\top \nabla^2 R(b + u\delta)\delta$. The two divergences are $\int_0^1 (1-u)h(u)\,du$ and $\int_0^1 uh(u)\,du$. For $K = \sqrt{\overline{h}/\underline{h}}$, the maximum of $\int_0^1 [1-(1+K)u]h(u)\,du$ under the stated bounds is $(\overline{h} - \underline{h}K^2)/(2(1+K)) = 0$. This proves the upper bound; reversing the segment proves the lower bound. A curvature profile equal to $\overline{h}$ before $1/(1+K)$ and $\underline{h}$ afterward attains the extremum in the bounded measurable class. Smooth approximations approach it within the C^2 class. □

For log barriers, endpoint slacks give an explicit comparison. Let $\psi(x) = x - 1 - \log x$ and define

$$\mathcal{A}(\rho) = \frac{\psi(\rho)}{\psi(1/\rho)} \quad (\rho > 1), \qquad \mathcal{A}(1) = 1. \tag{29}$$

Theorem 10 (Sharp log-barrier comparison)**.** *For $R(z) = \frac{1}{2}z^\top Qz - \sum_j \mu_j \log s_j(z)$ with $Q \succeq 0$, $\mu_j > 0$, and positive affine slacks, suppose $\rho^{-1} \le s_j(a)/s_j(b) \le \rho$ for every j. Then*

$$\mathcal{A}(\rho)^{-1} D_R(b,a) \le D_R(a,b) \le \mathcal{A}(\rho) D_R(b,a). \tag{30}$$

The factor is sharp over this class, and $\mathcal{A}(e^u) = 1 + \frac{2}{3}u + O(u^2)$ as $u \downarrow 0$.

Indeed, $D_R(a,b) = \frac{1}{2}(a-b)^\top Q(a-b) + \sum_j \mu_j \psi(s_j(a)/s_j(b))$. The quadratic term is symmetric, and $\psi(e^u)/\psi(e^{-u})$ increases with $u > 0$. A one-dimensional log barrier attains the factor; an added positive quadratic term preserves sharpness in the boundary limit. Appendix D gives the monotonicity and sharpness details. Consequently, (6) yields the forward-loss bound

$$\mathrm{Reg}_T(v) \le P_T(v)/\eta + \eta \sum_{t=1}^{T} \mathcal{A}(\rho_t) f_t(m_t, c_t) \tag{31}$$

whenever the played and updated slack ratios are bounded by ρ_t. The coefficients may depend on the hint and realized cost; (31) alone does not make their product with f_t a convex training objective.

D Sharp divergence comparisons

D.1 Integral representation and directional sharpness

For a C^2 potential and $\delta = a - b$,

$$D_R(a,b) = \int_0^1 (1-u)\delta^\top \nabla^2 R(b + u\delta)\delta\,du,$$

whereas changing variables in the reversed integral gives

$$D_R(b,a) = \int_0^1 u\,\delta^\top \nabla^2 R(b + u\delta)\delta\,du.$$

For $K = \sqrt{\overline{h}/\underline{h}}$, write $u_0 = 1/(1+K)$. The coefficient $1-(1+K)u$ is positive before u_0 and negative afterward, so

$$\begin{aligned} D_R(a,b) - K D_R(b,a) &\le \overline{h} \int_0^{u_0} (1-(1+K)u)\,du + \underline{h} \int_{u_0}^1 (1-(1+K)u)\,du \\ &= \frac{\overline{h} - \underline{h}K^2}{2(1+K)} = 0. \end{aligned}$$

Applying the same argument after replacing u by $1-u$ gives the lower bound.

For sharpness when $\overline{h} > \underline{h}$, work in one dimension with endpoints $0, 1$. Let $h_0(u)$ equal $\overline{h}$ on $u < u_0$ and $\underline{h}$ on $u > u_0$. Choose smooth functions h_ϵ taking values in $[\underline{h}, \overline{h}]$ and converging to h_0 almost everywhere on $[0,1]$. Extend them smoothly to $\mathbb{R}$ with the same positive bounds. Twice integrating gives smooth strictly convex potentials R_ϵ with $R_\epsilon'' = h_\epsilon$ and gradients onto $\mathbb{R}$. Dominated convergence shows that their divergence ratio tends to K. Thus the constant is optimal even within admissible smooth mirror potentials; it need not be attained by a continuous curvature profile when the bounds differ. If $\overline{h} = \underline{h}$, the two integrals are identical and the constant one is attained.

D.2 Exact log-barrier divergence

For a positive affine slack $s(z)$, affine linearity gives

$$D_{-\log s}(a,b) = -\log\frac{s(a)}{s(b)} + \frac{s(a)-s(b)}{s(b)} = \psi\left(\frac{s(a)}{s(b)}\right).$$

Bregman divergences are linear in the generating potential. Therefore

$$D_R(a,b) = \tfrac{1}{2}(a-b)^\top Q(a-b) + \sum_j \mu_j \psi\left(\frac{s_j(a)}{s_j(b)}\right). \tag{32}$$

The expression remains valid for $Q \succeq 0$ without strict convexity, and adding an affine term to R makes no difference.

D.3 Monotonicity of the comparison factor

For $u > 0$, let $N(u) = e^u - 1 - u$ and $D(u) = e^{-u} - 1 + u$. Both are positive, vanish at zero, and

$$N'(u) = e^u D'(u), \qquad D'(u) = 1 - e^{-u} > 0.$$

It follows that

$$\frac{N(u)}{D(u)} = \frac{\int_0^u e^v D'(v)\,dv}{\int_0^u D'(v)\,dv}$$

is a weighted average of e^v. This average is strictly less than e^u. Differentiating the ratio gives

$$\left(\frac{N}{D}\right)'(u) = \frac{D'(u)}{D(u)}\left(e^u - \frac{N(u)}{D(u)}\right) > 0.$$

Its limit as $u \downarrow 0$ is one. For $x = e^u \geq 1$, this proves monotonicity of $\psi(x)/\psi(1/x)$; for $x < 1$ the ratio is the reciprocal of the one at $1/x$. Consequently, if $\rho^{-1} \leq x \leq \rho$, then

$$\mathcal{A}(\rho)^{-1}\psi(1/x) \leq \psi(x) \leq \mathcal{A}(\rho)\psi(1/x).$$

This also holds at $x = 1$, where both divergences vanish. Apply it to each term of (32). The symmetric nonnegative quadratic term satisfies the same inequalities because $\mathcal{A}(\rho) \geq 1$. Summing proves Theorem 10.

Taylor expansion gives

$$N(u) = \tfrac{1}{2}u^2 + \tfrac{1}{6}u^3 + O(u^4), \qquad D(u) = \tfrac{1}{2}u^2 - \tfrac{1}{6}u^3 + O(u^4),$$

and division yields $\mathcal{A}(e^u) = 1 + \frac{2}{3}u + O(u^2)$.

For $\rho > 1$, the pure one-dimensional potential $-\mu\log s$ at endpoints $\rho\epsilon, \epsilon$ attains $\mathcal{A}(\rho)$ exactly. To see sharpness while retaining a positive quadratic and the global mirror assumptions, use $\mathcal{R}(s) = \frac{1}{2}(s-1)^2 - \mu\log s$. At the same endpoints the common quadratic divergence is $\frac{1}{2}(\rho-1)^2\epsilon^2$, whereas the log divergences are $\mu\psi(\rho)$ and $\mu\psi(1/\rho)$. Letting $\epsilon \downarrow 0$ gives the same limiting factor. At $\rho = 1$ the factor one cannot be improved since nonzero purely quadratic divergences are symmetric.

E Task sensitivity and a boundary-face example

For fixed pre-round state, differentiate the dual update to obtain

$$\nabla_m z_t(m) = -\eta[\nabla^2 R(z_t(m))]^{-1}.$$

The segment joining two hints lies in $\mathcal{M}_t$ by convexity. Continuity on this compact set makes κ_t finite. Applying the fundamental theorem of calculus to $c^\top z_t(m)$ proves (12). The Cauchy–Schwarz inequality gives

$$|\ell_t(m,c) - \ell_t(m',c)| \leq \eta\|c\|\kappa_t\|m - m'\| \leq \eta\|c\|\kappa_t d_t.$$

This also bounds the sum of the absolute roundwise differences, so summability in Theorem 3 implies absolute convergence, not merely bounded partial sums. No independence of the hints and past costs is needed.

Writing $z = z_t(m)$ and $b^+ = z_t(c)$, expansion gives

$$\begin{aligned} D_R(z, b^+) - D_R(z, b_t) &= R^*(y_t - \eta c) - R^*(y_t) + \langle \nabla R(b_t) - \nabla R(b^+), z \rangle \\ &= R^*(y_t - \eta c) - R^*(y_t) + \eta \langle c, z \rangle. \end{aligned}$$

This proves (14). For a finite conditional cost mean, conditional task risk is $\langle \bar{c}_t, z_t(m) \rangle$. Its gradient is nonzero when $\bar{c}_t \neq 0$. An interior point admits a small feasible step in the negative gradient direction, so cannot be a local minimum. The assertion does not concern lower-dimensional hint sets, whose relative interiors need not admit that step. If $\bar{c}_t = 0$, every hint has zero conditional expected task loss.

E.1 Exact linear gap along a face

Use the product potential of Section 7, initial state $(1, 0)$, and deterministic costs $c_t = (M, \delta\epsilon_t)$ with $\epsilon_t = (-1)^{t+1}$ observed before the action. Compare hints $m_t^{(0)} = (M, 0)$ and $m_t^{(1)} = (M, \delta\epsilon_t)$. Define $h = \eta\delta/(1 + \sqrt{1 + \eta^2\delta^2})$. The controller separates into two coordinates, and we will show

$$L_T(m^{(0)}) - L_T(m^{(1)}) = \delta h T. \tag{33}$$

The positive coordinate has identical costs and hints under both policies, so it contributes zero to their loss difference and converges to zero as cumulative cost Mt diverges. For the second coordinate let $\phi(u) = -\log(1 - u^2)$ and let $g = (\phi')^{-1}$. The derivative $\phi'(u) = 2u/(1 - u^2)$ is strictly increasing and onto $\mathbb{R}$. For $v \in \mathbb{R}$ its inverse is

$$g(v) = \frac{v}{1 + \sqrt{1 + v^2}}.$$

In particular $g(0) = 0$ and $g(-\eta\delta) = -h$ with h as defined above. At odd rounds the preceding signed cost sum is zero, and at even rounds it is δ. Thus at odd rounds the uninformative and informative tangential actions are respectively $0, -h$, and at even rounds they are $-h, 0$. Multiplying their difference by the corresponding costs $\delta, -\delta$ gives δh on every round. The context is observed before the action, so the informative hint is predictable. Summation proves (33) for every T. The construction demonstrates persistence of task sensitivity in a tangent direction; it does not assert that a particular estimator must recover that information.

F Proof for arbitrary polytope vertices

We prove Theorem 4, including nonsimple vertices. Translation permits $v = 0$; costs then measure regret relative to v, and the translated potential has the same quadratic-plus-affine-plus-log form. For each $j \in I(v)$ write $s_j(x) = a_j^\top x$. Define the open tangent cone $\mathcal{C} = \{x : a_j^\top x > 0 \text{ for every } j \in I(v)\}$ and

$$R_0(x) = -\sum_{j \in I(v)} \mu_j \log(a_j^\top x), \qquad \phi(x) = R(x) - R_0(x).$$

All inactive slacks are positive at zero, so ϕ has a convex C^2 extension to a neighborhood of zero. The active normals span $\mathbb{R}^d$: otherwise a nonzero vector orthogonal to all of them would permit a feasible line segment through zero, contradicting that it is a vertex. Thus $\nabla^2 R_0(x) \succ 0$ on $\mathcal{C}$. The full potential tends to infinity at the boundary of the bounded domain, has positive-definite Hessian, and hence satisfies the standing mirror assumptions.

F.1 Localization uniformly over changing directions

Let $K = \max\{G, H\}$ and consider any counterfactual u with $\|u\| \leq K$. This includes actions, updated states, and their interpolating hints. Write $x_t(u) = z_t(u)$ and $\theta_t(u) = \eta(S_{t-1} + u) - y_1$. The action minimizes $\langle \theta_t(u), x \rangle + R(x)$ on $\operatorname{int} P$. By (15), for every sufficiently large t,

$$\langle \theta_t(u), x \rangle \geq c_1 t \|x\| \quad (x \in P), \qquad \|\theta_t(u)\| \leq c_2 t, \tag{34}$$

where $c_1, c_2 > 0$ and the entry time depend only on the stated parameters. Indeed, subtracting the bounded vector $y_1 - \eta u$ from ηS_{t-1} changes the exposure lower bound by at most a constant times $\|x\|$.

Choose a fixed strictly feasible vector $e \in \operatorname{int} P$. Then e/t is feasible for all $t \geq 1$, and its objective value is

$$\langle \theta_t(u), e/t \rangle + R(e/t) = \nu_v \log t + O(1).$$

The potential R is bounded below on its bounded domain, since every slack is bounded above and the quadratic and affine terms are bounded. Optimality and (34) give $\|x_t(u)\| = O(\log t/t)$, uniformly in u and the allowed histories. Consequently all sufficiently late states lie in the neighborhood where $\nabla\phi$ is bounded.

The first-order equation is $\theta_t(u) + \nabla R_0(x_t(u)) + \nabla\phi(x_t(u)) = 0$. Logarithmic homogeneity gives $\langle \nabla R_0(x), x \rangle = -\nu_v$. Taking the inner product with $x_t(u)$ therefore yields

$$c_1 t\|x_t(u)\| \leq \langle \theta_t(u), x_t(u) \rangle = \nu_v - \langle \nabla\phi(x_t(u)), x_t(u) \rangle \leq \nu_v + C\|x_t(u)\|.$$

For large t this improves localization to $\|x_t(u)\| \leq C'/t$.

F.2 Scaled slacks and inverse curvature

Set $U_t = tx_t(u)$, which now lies in a fixed bounded subset of $\mathcal{C}$. The comparison with e/t also gives

$$\langle \theta_t(u)/t, U_t \rangle + R_0(U_t) + \phi(U_t/t) \leq C.$$

The first term is nonnegative by exposure, and ϕ is bounded near zero. Hence $R_0(U_t)$ is bounded above. Every active slack $a_j^\top U_t$ is bounded above by boundedness of U_t. Each term $-\mu_j \log(a_j^\top U_t)$ consequently has a uniform lower bound. An upper bound on their sum then supplies an upper bound on each term, so there exist fixed $\alpha_j, \beta_j > 0$ such that

$$\alpha_j \leq t s_j(x_t(u)) \leq \beta_j \qquad (j \in I(v)). \tag{35}$$

It follows that $R(x_t(u)) = \nu_v \log t + O(1)$. Moreover, convexity of ϕ and the active Hessian give

$$\nabla^2 R(x_t(u)) \succeq t^2 \sum_{j \in I(v)} \frac{\mu_j a_j a_j^\top}{\beta_j^2} \succeq c_3 t^2 I,$$

where $c_3 > 0$ because the active normals span the space. Thus $J_t(u) \preceq c_3^{-1} t^{-2} I$. All finitely many early dual states lie in a common compact set, so continuity extends this bound to every round after increasing its constant.

F.3 Regret and policy comparisons

Integral Taylor expansion for R^* bounds F_t and B_t by a constant times t^{-2}: their dual displacement is $\eta(m_t - c_t)$, uniformly bounded, and the whole segment is covered by the counterfactual estimate. The same argument between hints 0 and m_t bounds $V_t = O(t^{-2})$. For the base $b_{T+1} = x_{T+1}(0)$, (35) gives

$$R(b_{T+1}) = \nu_v \log(T+1) + O(1), \qquad \langle y_{T+1}, b_{T+1} \rangle = O(1),$$

since $\|y_{T+1}\| = O(T+1)$ and $\|b_{T+1}\| = O((T+1)^{-1})$. Conjugacy implies $R^*(y_{T+1}) = -\nu_v \log(T+1) + O(1)$. The exact boundary identity at comparator zero and the summable divergences prove $\eta \operatorname{Reg}_T(0) = \nu_v \log T + O(1)$, including small T by adjusting constants. Finally, the sensitivity identity gives

$$|\langle c_t, z_t - z_t' \rangle| \leq 2\eta G H C_4 t^{-2}.$$

Summing over $t > n$ proves (17). The argument never selects an independent subset of active constraints; all their weights enter the homogeneous degree ν_v.

G Proof of the reduction to a simplex face

We prove Theorem 5 for fixed coordinate bounds $|c_{t,i}| \leq G$ and $|m_{t,i}| \leq H$. Set $\theta_i = \eta(S_{t-1,i} + m_i) - y_{1,i}$ for $i \geq 1$ and $\theta_0 = 0$. For large t, the face coordinates satisfy $|\theta_i| \leq C_0$ for $i \in I$, while

$c_1 t \le \theta_j \le c_2 t$ for $j \in N$. Let λ be the full-simplex multiplier and λ_A the face multiplier. The first-order equations and normalization give

$$z_{t,i}(m) = \frac{\mu_i}{\theta_i + \lambda} \quad (0 \le i \le d), \qquad \sum_{i=0}^{d} \frac{\mu_i}{\theta_i + \lambda} = 1, \tag{36}$$

$$u_{t,i}(m_I) = \frac{\mu_i}{\theta_i + \lambda_A} \quad (i \in A), \qquad \sum_{i \in A} \frac{\mu_i}{\theta_i + \lambda_A} = 1. \tag{37}$$

Here $u_{t,0} = 1 - \sum_{i \in I} u_{t,i}$. Each scalar equation has a unique root above all its poles: its left side is continuous and strictly decreasing from infinity to zero.

Let $F_A(l) = \sum_{i \in A} \mu_i/(\theta_i + l)$ on the interval above its poles. The face pole locations are uniformly bounded. At either root, each summand is at most one, giving $\theta_i + \lambda_A \ge \mu_i$ for $i \in A$ and $\theta_i + \lambda \ge \mu_i$ for every coordinate. Both roots also have a fixed upper bound: above the largest relevant pole plus $\sum_i \mu_i + 1$, the full sum is less than one. For large t the normal-coordinate poles are negative and cannot enlarge this bound. Thus both roots lie in a fixed bounded interval, $\lambda \ge \lambda_A \ge \mu_0$, and

$$\sum_{j \in N} \frac{\mu_j}{\theta_j + \lambda} = O(t^{-1}), \qquad -F_A'(l) \ge \mu_0/l^2 \ge c_3 > 0 \quad (\lambda_A \le l \le \lambda).$$

Subtract the normalizations to obtain $F_A(\lambda_A) - F_A(\lambda) = \sum_{j \in N} \mu_j/(\theta_j + \lambda)$. The mean value theorem gives $\lambda - \lambda_A = O(t^{-1})$. The denominators for face coordinates are uniformly bounded above and away from zero, so $z_{t,I} - u_t = O(t^{-1})$ and $z_{t,0} - u_{t,0} = O(t^{-1})$. For excluded coordinates, $\theta_j + \lambda = \eta S_{t-1,j} + O(1)$, proving (19). This also places all face probabilities in a fixed compact subset of their simplex interior.

G.1 Block curvature and normal hint changes

Write the Hessian in the order (I, N) as

$$\nabla^2 R(z_t) = \begin{pmatrix} A_t & C_t \\ C_t^\top & D_t \end{pmatrix}.$$

The simplex Hessian is $\operatorname{diag}_{i \ge 1}(\mu_i/z_{t,i}^2) + (\mu_0/z_{t,0}^2)\mathbf{1}\mathbf{1}^\top$. Consequently $A_t = \nabla^2 R_A(u_t) + O(t^{-1})$, $C_t = O(1)$, and $D_t \succeq c_4 t^2 I$. The eigenvalues of $\nabla^2 R_A(u_t)$ stay bounded above and away from zero. The Schur complement therefore satisfies

$$A_t - C_t D_t^{-1} C_t^\top = \nabla^2 R_A(u_t) + O(t^{-1}).$$

Block inversion gives the tangent block $J_{A,t} + O(t^{-1})$, the off-diagonal blocks $O(t^{-2})$, and the normal block $O(t^{-2})$. All these bounds hold along any segment of bounded hints. If two policies have identical tangential hints, their hint difference is supported on N. Multiplication by the corresponding columns of J_t and the sensitivity identity show that their action difference is $O(t^{-2})$ in every coordinate. Bounded costs and summation prove the absolute policy-tail bound.

G.2 Task-loss reduction

For each sufficiently large round, bounded costs and (19) give

$$|\langle c_t, z_t(m)\rangle - \langle c_{t,I}, u_t(m_I)\rangle| \le \sum_{i \in I} |c_{t,i}|\,|z_{t,i} - u_{t,i}| + \sum_{j \in N} |c_{t,j}| z_{t,j} = O(t^{-1}).$$

The finite prefix is uniformly bounded by boundedness of the simplex and costs. Summing gives (20). In particular the theorem compares any two full policies to their two face policies with an $O(\log(1+T))$ discrepancy in the cumulative difference. It does not assert that this logarithmic error is optimal, or that tangential losses necessarily grow linearly.

H Proof of the learned task separation

All claims first concern the product controller in Section 7. The normal coordinates of the two policies coincide on every round and cancel from every comparison. Write $\phi(u) = -\log(1 - u^2)$.

Its derivative is $2u/(1-u^2)$, its inverse gradient is g, and

$$\phi''(u) = \frac{2(1+u^2)}{(1-u^2)^2} \geq 2.$$

Therefore $0 < g'(x) \leq 1/2$ for every finite x. The cumulative tangential cost is zero before each odd round and δY_k before the subsequent even round. The tangential action at odd-round hint m is $-g(\eta m)$. Both policies choose hint $-\delta Y_k$ on the even round, so their tangential action there is zero. Conditional on the previous pairs, the odd-round cost mean is $\delta\theta$. Its expected task loss is

$$f_\theta(m) = -\delta\theta g(\eta m), \qquad -\delta \leq m \leq \delta. \tag{38}$$

Strict monotonicity of g makes $m^* = \delta\,\mathrm{sgn}(\theta)$ its unique minimizer. On even rounds the known negative of the previous cost is also an endpoint hint and minimizes the task loss. Because hints do not affect future bases, these choices form the task-optimal policy among all policies using these hint intervals and knowing θ.

H.1 Learning the task-optimal sign

For $k \geq 2$, Hoeffding's inequality for observations in $[-1, 1]$ gives

$$\Pr_\theta\{\mathrm{sgn}(\widehat{\theta}_k) \neq \mathrm{sgn}(\theta)\} \leq \exp(-(k-1)\theta^2/2).$$

If the sign is wrong, the odd-round expected task excess is exactly $2\delta|\theta|g(\eta\delta)$; if it is right, the excess is zero. Let L^*_{2K} denote the cumulative loss of the task-optimal policy just defined. The first round contributes at most the same constant, hence

$$0 \leq \mathbb{E}_\theta[L^{\mathrm{sat}}_{2K} - L^*_{2K}] \leq 2\delta|\theta|g(\eta\delta)\left(1 + \frac{1}{e^{\theta^2/2}-1}\right). \tag{39}$$

The right side is finite for each fixed $\theta \neq 0$; no uniform bound as $\theta \to 0$ is claimed.

H.2 The mean predictor and its task gap

Let the mean oracle use odd-round hint $\delta\theta$ and the same even-round hint as the learners, and denote its loss by L^{Bayes}_{2K}. Oddness of g gives

$$\mathbb{E}_\theta[L^{\mathrm{Bayes}}_{2K} - L^*_{2K}] = K\delta|\theta|[g(\eta\delta) - g(\eta\delta|\theta|)] = K\Delta_\theta.$$

Since $0 < |\theta| < 1$ and g is strictly increasing, $\Delta_\theta > 0$. For $k \geq 2$, the sample mean is unbiased and has variance $(1-\theta^2)/(k-1)$:

$$\mathbb{E}_\theta(\widehat{\theta}_k - \theta) = 0, \qquad \mathbb{E}_\theta(\widehat{\theta}_k - \theta)^2 = (1-\theta^2)/(k-1).$$

Let $C_g = \sup_{|x|\leq\eta\delta}|g''(x)| < \infty$. Taylor expansion of $g(\eta\delta\widehat{\theta}_k)$ about $\eta\delta\theta$ cancels the linear term after expectation and gives

$$\left|\mathbb{E}_\theta g(\eta\delta\widehat{\theta}_k) - g(\eta\delta\theta)\right| \leq \frac{C_g\eta^2\delta^2(1-\theta^2)}{2(k-1)}.$$

Multiplying by $\delta|\theta|$, summing the harmonic series, and bounding the first round yields $|\mathbb{E}_\theta[L^{\mathrm{mean}}_{2K} - L^{\mathrm{Bayes}}_{2K}]| = O(\log(1+K))$. Combining this estimate with (39) proves (23).

For the statistical claim, the odd-round tangential prediction potential is $\Psi(m) = \eta^{-2}\phi^*(-\eta m)$ and has curvature $\Psi''(m) = 1/\phi''(-g(\eta m))$. On the fixed compact interval $[-\delta, \delta]$, let $0 < \lambda \leq \Lambda$ be its minimum and maximum curvature. The normal component contributes zero excess, and Theorem 2 gives conditional excess reverse loss $D_\Psi(\delta\theta, \delta\widehat{\theta}_k)$. Thus for $k \geq 2$,

$$\frac{\lambda\delta^2(1-\theta^2)}{2(k-1)} \leq \mathbb{E}_\theta D_\Psi(\delta\theta, \delta\widehat{\theta}_k) \leq \frac{\Lambda\delta^2(1-\theta^2)}{2(k-1)}.$$

The rate tends to zero even though direct use of that mean target has a nonzero per-pair task penalty.

H.3 EMBEDDING IN A COUPLED SIMPLEX

Take $d = 2$, weights $(\mu_0, \mu_1, \mu_2) = (1, 1, \mu)$, and any interior initialization b_1 satisfying $b_{1,0} = b_{1,1}$, where $b_{1,0} = 1 - b_{1,1} - b_{1,2}$. The initial first reduced dual coordinate is then zero. Choose $A = \{0, 1\}$ and $N = \{2\}$. For the costs specified in the theorem, $|S_{t,1}| \le 2\delta$ and $S_{t,2} = Mt$, so the face assumptions hold on every realization. Use normal hint M and tangential hints twice those in the product model. On the face, transform the first alternative's probability p to $u = 2p - 1$. Then

$$-\log p - \log(1-p) = -\log(1-u^2) + \log 4, \qquad (2\delta Y_k)p = \delta Y_k u + \delta Y_k.$$

The extra cost constants cancel over each pair and in every policy comparison. The affine change of coordinates transforms the face controller exactly into the preceding tangential controller. Theorem 5 therefore gives, pathwise for either learner, an $O(\log(1 + K))$ discrepancy between full and face task losses. The constants are uniform over all realizations of (Y_k) and all learner hints, so the same estimate holds after expectation. Combining it with the product-model result proves $\mathbb{E}[L_{2K}^{\mathrm{mean}} - L_{2K}^{\mathrm{sat}}] = K\Delta_\theta + O(\log(1 + K))$ on the coupled simplex. Only the linear separation is transferred here; the product model's bounded excess bound is not asserted for the entire coupled controller.

I A BOUNDARY REGIME WITH NO LEADING TASK-LOSS SEPARATION

Consider $\mathcal{Z} = (0, \infty)$ with boundary comparator 0 and

$$\mathcal{R}(s) = \tfrac{1}{2}(s-1)^2 - \mu \log s, \qquad b_1 = 1, \qquad \mu, \eta > 0. \tag{40}$$

Fix $0 < a < b$, and suppose $c_t, m_t \in [aM, bM]$ for all t, where

$$\eta a M \ge 2\max\{|\mu - 1|, \sqrt{\mu}\}, \qquad k = (b-a)/a. \tag{41}$$

Costs need not be stochastic. Writing w_t for the played slack and $S_t = \sum_{i \le t} c_i$, the explicit state map is

$$s(H) = \frac{2\mu}{\eta H + \mu - 1 + \sqrt{(\eta H + \mu - 1)^2 + 4\mu}}, \quad w_t = s(S_{t-1} + m_t), \quad b_{t+1} = s(S_t). \tag{42}$$

Theorem 11 (Summable error for arbitrary hints)**.** *Under* (41)*, every allowed cost and hint sequence satisfies*

$$F_t,\ B_t \le \frac{C_k}{t^2}, \qquad B_t \le \mathcal{A}(e^{2k/t})F_t, \qquad C_k = 2\mu k^2(1 + e^{2k}). \tag{43}$$

Thus $\sum_{t=1}^{T} B_t$ and $\sum_{t=1}^{T} F_t$ are bounded independently of M, T. The divergence comparison factor is $1 + O(t^{-1})$, even though the Hessian grows as $\Theta(M^2t^2)$.

The proof uses $\mu/(2\eta H) \le s(H) \le 2\mu/(\eta H)$, $|(\log s)'(H)| \le 2/H$, and $|s'(H)| \le 4\mu/(\eta H^2)$. The two cumulative arguments in (42) differ by at most $(b-a)M$ and are both at least aMt. Their slack ratio is therefore within $e^{\pm 2k/t}$. Substitution in $D_{\mathcal{R}}(x, y) = \frac{1}{2}(x-y)^2 + \mu\psi(x/y)$ proves (43); Appendix K gives all constants.

If $c_t = MX_t$ are i.i.d. with $\mathrm{Var}(X_t) = v > 0$, every predictable hint has

$$\mathbb{E}[(c_t - m_t)^2 \mid \mathcal{G}_t] = M^2 v + (M\mathbb{E}X_t - m_t)^2$$

when $\mathcal{G}_t$ contains only past data and independent randomization. Hence even the mean predictor has cumulative squared error M^2vT. This is a difference in error measurements: Theorem 11 applies to the mean predictor as well as every other admissible predictor.

Theorem 12 (Uniform task asymptotics and policy comparison)**.** *Under* (41)*, uniformly over M, T and all allowed sequences,*

$$L_T := \sum_{t=1}^{T} c_t w_t = \frac{\mu}{\eta}\log T + O_{a,b,\mu,\eta}(1). \tag{44}$$

For any two hint policies evaluated on the same costs,

$$|L_T - L'_T| \le \frac{4\mu b(b-a)}{\eta a^2}\frac{\pi^2}{6}. \tag{45}$$

Moreover, $m_t = bM$ minimizes each round's loss, and therefore cumulative loss, among all hints in $[aM, bM]$.

Proof outline. The state map is decreasing and satisfies $|s'(H)| \le 4\mu/(\eta H^2)$. Two hints change its argument by at most $(b-a)M$, proving the summable policy bound. Uniformly for $H \ge aM$, $s(H) = \mu/(\eta H) + O(H^{-2})$. For $t \ge 2$, replacing $c_t/(S_{t-1}+m_t)$ by $\log(S_t/S_{t-1})$ incurs a summable $O((t-1)^{-2})$ error. The logarithms telescope to $\log T + O(1)$. Monotonicity and hint-independent future bases prove pointwise optimality of bM. Appendix K gives the complete uniform argument. □

The result permits bounded finite-horizon differences between policies. The common leading loss is a property of the specified controller: the boundary action 0 is already optimal for every positive cost. Changing the step size changes the controller and is a separate comparison from changing its prediction loss.

J Learning the mean-calibrated remainder

The reverse loss is not globally convex for a nonquadratic geometry, but it can become convex on the range encountered by the controller. Continue with (40)–(41), and write $q_t(m) = \eta^{-2} D_{\mathcal{R}}(s(S_{t-1}+m), s(S_t))$.

Proposition 15 proves that q_t is convex on the allowed interval after at most $\max\{1, \lceil 4k \rceil\}$ rounds, despite the global obstruction in Proposition 7.

Now let $c_t = MX_t$, where $X_t \in [a,b]$ are i.i.d. with mean $\bar{x}$ and variance $v > 0$. Define the history $\mathcal{F}_{t-1} = \sigma(X_1, \dots, X_{t-1})$. Use any $m_1 \in [aM, bM]$ and ordinary sample-mean prediction

$$m_t = \frac{S_{t-1}}{t-1} \quad (t \ge 2), \qquad \mathcal{Q}_t(m) = \mathbb{E}[q_t(m) \mid \mathcal{F}_{t-1}]. \tag{46}$$

Theorem 13 (Excess remainder of sample-mean prediction)**.** *For every $t \ge 2$,*

$$\frac{\mu v}{16\eta^2 b^2 t^2 (t-1)} \le \mathbb{E}[\mathcal{Q}_t(m_t) - \mathcal{Q}_t(M\bar{x})] \le \frac{2\mu v}{\eta^2 a^2 t^2 (t-1)}. \tag{47}$$

Consequently, the expected excess normalized remainder after round $n \ge 2$ has a tail of order $\Theta(n^{-2})$.

The proof combines a pathwise curvature bound with the sample-mean variance; Appendix K gives both calculations. The rate is also optimal over an unknown two-point family. For $p \in [1/4, 3/4]$, let $\mathbb{P}_p$ assign probabilities $p, 1-p$ to $X = b, a$, and let $\bar{x}_p = a + (b-a)p$. Write $\mathcal{Q}_{t,p}$ for the conditional risk under this law and $\mathbb{E}_p$ for expectation under it.

Theorem 14 (Minimax excess remainder)**.** *Under the same controller assumptions, for every $t \ge 2$,*

$$\inf_{\widehat{m}_t} \sup_{p \in [1/4, 3/4]} \mathbb{E}_p[\mathcal{Q}_{t,p}(\widehat{m}_t) - \mathcal{Q}_{t,p}(M\bar{x}_p)] \ge \frac{\mu(b-a)^2}{4096\eta^2 b^2 t^2 (t-1)}. \tag{48}$$

The infimum is over $[aM, bM]$-valued estimators based on the previous $t-1$ costs and independent randomization, without knowledge of p. Sample-mean prediction attains the same order uniformly over this family.

The proof uses two Bernoulli parameters separated by order $t^{-1/2}$ and the same uniform curvature lower bound. Thus t^{-3} is an optimal excess rate for this evolving loss, not a faster rate of estimating the mean: ordinary squared estimation error remains of order M^2/t. Theorems 12 and 3 distinguish this statistical guarantee from task optimality.

K Uniform half-line estimates and learning

Throughout this appendix assume (40) and (41). Constants may depend on a, b, μ, η but never on the cost or hint sequence, M, or T, unless explicitly stated.

K.1 State map and derivative bounds

The derivatives are

$$\mathcal{R}'(s) = s - 1 - \mu/s, \qquad \mathcal{R}''(s) = 1 + \mu/s^2, \qquad \mathcal{R}'''(s) = -2\mu/s^3.$$

Since $\mathcal{R}'(1) = -\mu$, the state equation $\mathcal{R}'(s(H)) = -\mu - \eta H$ becomes

$$s(H)^2 + (\eta H + \mu - 1)s(H) - \mu = 0.$$

The unique positive root is (42). For $H \geq aM$, put $L = \eta H$ and $A = L + \mu - 1$. The threshold implies

$$\tfrac{1}{2}L \leq A \leq \tfrac{3}{2}L, \qquad L \geq 2\sqrt{\mu}.$$

Since $\sqrt{A^2 + 4\mu} \leq \sqrt{13}\, L/2$, the denominator in (42) is less than $4L$. It is at least $2A \geq L$. Therefore

$$\frac{\mu}{2\eta H} \leq s(H) \leq \frac{2\mu}{\eta H} \leq \sqrt{\mu}. \tag{49}$$

Differentiating the implicit equation gives

$$s'(H) = -\frac{\eta s(H)}{\sqrt{A^2 + 4\mu}}, \qquad (\log s)'(H) = -\frac{\eta}{\sqrt{A^2 + 4\mu}}.$$

Thus

$$|(\log s)'(H)| \leq \frac{2}{H}, \qquad |s'(H)| \leq \frac{4\mu}{\eta H^2}. \tag{50}$$

The representation $s(H) = \mu/(A + s(H))$ also gives a useful uniform expansion:

$$\begin{aligned} \left| s(H) - \frac{\mu}{\eta H} \right| &= \frac{\mu|\mu - 1 + s(H)|}{\eta H\,[A + s(H)]} \\ &\leq \frac{2\mu(|\mu - 1| + \sqrt{\mu})}{(\eta H)^2}. \end{aligned} \tag{51}$$

This bound is valid on the entire range $H \geq aM$, not just asymptotically.

K.2 Proof of Theorem 11

Set $H_m = S_{t-1} + m_t$ and $H_c = S_t$. Both lie in $[aMt, bMt]$ and $|H_m - H_c| \leq (b-a)M$. Integrating (50) on the interval between them gives

$$\left| \log \frac{w_t}{b_{t+1}} \right| \leq \frac{2k}{t}, \qquad |w_t - b_{t+1}| \leq \frac{4\mu(b-a)}{\eta a^2 M t^2}. \tag{52}$$

For either orientation, (32) in one dimension gives

$$D_{\mathcal{R}}(x, y) = \tfrac{1}{2}(x - y)^2 + \mu\psi(x/y).$$

For all real u, Taylor's formula bounds $0 \leq e^u - 1 - u \leq \frac{1}{2}e^{|u|}u^2$. Using (52), either log-divergence term is at most $2\mu e^{2k}k^2/t^2$. The quadratic term is at most

$$\frac{8\mu^2(b-a)^2}{\eta^2 a^4 M^2 t^4} = \frac{8\mu^2 k^2}{\eta^2 a^2 M^2 t^4} \leq \frac{2\mu k^2}{t^4} \leq \frac{2\mu k^2}{t^2}.$$

This proves the first part of (43). Summing uses $\sum_{t=1}^{\infty} t^{-2} = \pi^2/6$. The slack-ratio bound and Theorem 10 prove $B_t \leq \mathcal{A}(e^{2k/t})F_t$, including the case $F_t = B_t = 0$. Finally, (49) on $H \in [aMt, bMt]$ shows that $1 + \mu/s(H)^2 = \Theta(M^2t^2)$ with the permitted fixed-parameter constants.

K.3 Task asymptotics and policy comparison

Proof. The cumulative arguments for two policies differ by $m_t - m'_t$ and are at least aMt. The derivative bound therefore gives $c_t|w_t - w'_t| \le 4\mu b(b-a)/(\eta a^2 t^2)$. Summing proves (45).

For the leading coefficient, write $A = \eta H + \mu - 1$. The state equation gives $s(H) = \mu/(A + s(H))$. Since $A \ge \eta H/2$ and $s(H) \le \sqrt{\mu}$ under (41),

$$\left| s(H) - \frac{\mu}{\eta H} \right| \le \frac{2\mu(|\mu - 1| + \sqrt{\mu})}{(\eta H)^2} \qquad (H \ge aM).$$

Consequently,

$$\left| c_t w_t - \frac{\mu}{\eta} \frac{c_t}{S_{t-1} + m_t} \right| \le \frac{2\mu(|\mu - 1| + \sqrt{\mu})b}{\eta^2 a^2 M t^2}. \tag{53}$$

The threshold bounds M away from zero in terms of the fixed parameters, so this error has a uniformly bounded sum. For $t \ge 2$, put $S = S_{t-1} \ge aM(t-1)$. Using $0 \le x - \log(1+x) \le x^2/2$ for $x \ge 0$ gives

$$\left| \frac{c_t}{S + m_t} - \frac{c_t}{S} \right| = \frac{c_t m_t}{S(S + m_t)} \le \frac{b^2}{a^2(t-1)^2},$$
$$\left| \frac{c_t}{S} - \log \frac{S + c_t}{S} \right| \le \frac{c_t^2}{2S^2} \le \frac{b^2}{2a^2(t-1)^2}.$$

Both errors are summable independently of the sequence and scale. The logarithms telescope:

$$\sum_{t=2}^{T} \log \frac{S_t}{S_{t-1}} = \log \frac{S_T}{c_1} = \log T + O_{a,b}(1), \qquad \frac{aT}{b} \le \frac{S_T}{c_1} \le \frac{bT}{a}.$$

The first ratio $c_1/m_1 \le b/a$ and (53) prove (44), including $T = 1$. Finally, s is decreasing and the base trajectory is hint-independent. Increasing the current hint lowers the positive-cost loss without changing any later base state, proving the last claim. □

K.4 The boundary comparator and the first stability term

Equation (69) gives the explicit comparator potential and initial stability divergence, including their cancelling $\mu \log M$ terms. The remaining stability terms can be bounded directly. For $t \ge 2$, the base $b_t = s(S_{t-1})$ and played state $w_t = s(S_{t-1} + m_t)$ have cumulative arguments at least $aM(t-1)$ and separation at most bM. The same estimates as above, with b/a in place of k, give $V_t \le C/(t-1)^2$. This proves a uniformly summable stability tail. The first term is finite for every M, but is not uniformly bounded as $M \to \infty$. It cannot be discarded when explaining the M-independent remainder in (44).

K.5 Convexity after burn-in

Proposition 15 (Convexity after a finite burn-in)**.** *For every cost history, q_t is convex on $[aM, bM]$ whenever $t \ge \max\{1, \lceil 4k \rceil\}$.*

Indeed, with $w = s(S_{t-1} + m)$,

$$q''_t(m) = \frac{1}{\mathcal{R}''(w)} \left[1 + \frac{\eta(m - c_t)\mathcal{R}'''(w)}{\mathcal{R}''(w)^2} \right]. \tag{54}$$

Here $\mathcal{R}''(w) = 1 + \mu/w^2$ and $\mathcal{R}'''(w) = -2\mu/w^3$. The bracket is at least one if $m \le c_t$ and at least $1 - 4k/t$ otherwise.

Equation (9) gives $q'_t(m) = (m - c_t)/\mathcal{R}''(w)$, where $w = s(S_{t-1} + m)$. Since $w' = -\eta/\mathcal{R}''(w)$, differentiating yields (54). If $m \le c_t$, the product $(m - c_t)\mathcal{R}'''(w)$ is nonnegative. If $m > c_t$, the bracket in that equation is

$$1 - \frac{2\eta\mu(m - c_t)w}{(w^2 + \mu)^2} \ge 1 - \frac{2\eta(m - c_t)w}{\mu} \ge 1 - \frac{4(b-a)}{at} = 1 - \frac{4k}{t}.$$

The last inequality uses $m - c_t \le (b-a)M$ and $w \le 2\mu/(\eta a M t)$. At and after the stated integer threshold, $q''_t(m) \ge 0$ throughout the closed hint interval, proving convexity.

K.6 Normalization and the learning tail

Proof. For every realized history and $m \in [aM, bM]$, put $s = s(S_{t-1} + m)$. The state bounds imply $\mu/(2\eta bMt) \le s \le 2\mu/(\eta aMt) \le \sqrt{\mu}$. Since $\Psi_t''(m) = s^2/(s^2+\mu)$, we obtain the pathwise curvature sandwich

$$\frac{\mu}{8\eta^2 b^2 M^2 t^2} \le \Psi_t''(m) \le \frac{4\mu}{\eta^2 a^2 M^2 t^2}. \tag{55}$$

Independence of X_t from the past makes $M\bar{x}$ its conditional cost mean. Theorem 2 thus identifies the excess as $D_{\Psi_t}(M\bar{x}, m_t)$. Both arguments lie in $[aM, bM]$. Integral Taylor expansion gives

$$\frac{\mu(m_t - M\bar{x})^2}{16\eta^2 b^2 M^2 t^2} \le \mathcal{Q}_t(m_t) - \mathcal{Q}_t(M\bar{x}) \le \frac{2\mu(m_t - M\bar{x})^2}{\eta^2 a^2 M^2 t^2}.$$

The estimator is unbiased, and independence of the preceding observations gives

$$\mathbb{E}(m_t - M\bar{x})^2 = \frac{M^2}{(t-1)^2} \sum_{i=1}^{t-1} \mathrm{Var}(X_i) = \frac{M^2 v}{t-1}.$$

Taking expectations proves (47); no independence between m_t and the random potential Ψ_t is needed. Finally, $t^{-3} \le [t^2(t-1)]^{-1} \le 2t^{-3}$ for $t \ge 2$. Integral comparison yields an order-n^{-2} tail, and nonnegativity permits exchanging expectation and the infinite sum. □

The preceding proof establishes the pathwise curvature sandwich (55) and the sample-mean variance calculation. For an explicit tail comparison, $t^{-3} \le [t^2(t-1)]^{-1} \le 2t^{-3}$ implies

$$\frac{1}{2(n+1)^2} \le \sum_{t=n+1}^{\infty} \frac{1}{t^2(t-1)} \le \frac{1}{n^2} \qquad (n \ge 2)$$

by integration of x^{-3}. Multiplying by the constants in (47) bounds the expected excess normalized remainder. The corresponding excess in B_t is multiplied by η^2, and its contribution to the regret bound by η. This calculation concerns excess relative to the conditional Bayes risk; it does not remove the irreducible conditional variance in the raw squared error or assert optimality for task loss.

K.7 Proof of Theorem 14

Put $n = t - 1$ and choose $p_\pm = 1/2 \pm 1/(8\sqrt{n})$. Both parameters belong to $[1/4, 3/4]$. Their cost means $\theta_\pm = M[a + (b-a)p_\pm]$ have separation $\Delta = \theta_+ - \theta_- = M(b-a)/(4\sqrt{n})$. Let $P_\pm$ denote the laws of the n past observations, let $A = \sum_\omega \sqrt{P_+(\omega)P_-(\omega)}$ be their affinity, and let $\mathrm{TV}(P_+, P_-) = \frac{1}{2}\sum_\omega |P_+(\omega) - P_-(\omega)|$ be their total variation distance. Independence gives $A = (1 - 1/(16n))^{n/2}$. Applying Cauchy–Schwarz after factoring a difference of squares gives

$$\mathrm{TV}(P_+, P_-) \le \sqrt{1 - A^2} = \sqrt{1 - (1 - 1/(16n))^n} \le \frac{1}{4}.$$

The last step uses $(1-x)^n \ge 1 - nx$ for $x \in [0, 1]$. Independent estimator randomization leaves this bound unchanged.

Any estimator induces a test by choosing the nearer of θ_+, θ_-. On a testing error its squared estimation error is at least $\Delta^2/4$. The sum of the two testing error probabilities is at least $1 - \mathrm{TV}(P_+, P_-)$. Consequently

$$\max_{\sigma \in \{-,+\}} \mathbb{E}_{p_\sigma}(\widehat{m}_t - \theta_\sigma)^2 \ge \frac{\Delta^2}{8}(1 - \mathrm{TV}(P_+, P_-)) \ge \frac{M^2(b-a)^2}{256n}.$$

The pathwise curvature sandwich (55) holds under either parameter, despite the dependence of the geometry on the sample. The conditional excess-risk identity therefore gives

$$\mathcal{Q}_{t,p}(\widehat{m}_t) - \mathcal{Q}_{t,p}(M\bar{x}_p) \ge \frac{\mu(\widehat{m}_t - M\bar{x}_p)^2}{16\eta^2 b^2 M^2 t^2}.$$

Taking expectations, the maximum over the two parameters, and the infimum over estimators proves (48). The sample-mean upper bound follows from Theorem 13 and $\mathrm{Var}_p(X) \le (b-a)^2/4$. This is a minimax statement about the roundwise excess loss, with fixed a, b, μ, η and admissible M; it is not a task-regret lower bound.

Table 1: Uniform boundary rates for fixed η, μ and sufficiently large M. Write $\gamma = (q-2)/(q-1) = 1-r$. The remainder column is an upper bound: it can vanish for perfect hints.

Curvature exponent	$\sum_{t=1}^{T} B_t$	Actual loss $\sum_{t=1}^{T} c_t w_t$
$1 < q < 2$	$O(M^\gamma)$	$\Theta(M^\gamma)$
$q = 2$	$O(1)$	$\Theta(\log(1+T))$
$q > 2$	$O(M^\gamma)$	$\Theta(M^\gamma T^\gamma)$

L General boundary curvature: where the transition occurs

Replace the log barrier by

$$\mathcal{R}(s) = \tfrac{1}{2}(s-1)^2 + \mu\mathcal{B}(s), \qquad \mathcal{B}''(s) \sim c_0 s^{-q} \quad (s \downarrow 0), \qquad r = \frac{1}{q-1}, \tag{56}$$

where $\mathcal{B} \in C^2(0,\infty)$ is convex, $c_0 > 0$, and $q > 1$. Then $\mathcal{B}'(s) \to -\infty$ at zero. The quadratic term makes $\mathcal{R}'$ tend to $+\infty$ at infinity, so the mirror update is well defined. For $1 < q < 2$, $\mathcal{B}$ has a finite boundary value but a divergent gradient; the result does not require a value-divergent barrier. Keep $b_1 = 1$, $c_t, m_t \in [aM, bM]$, and fixed $\mu, \eta > 0$.

Theorem 16 (Summable remainder and task-loss transition)**.** *For all sufficiently large M, uniformly over allowed sequences and $T \geq 1$,*

$$F_t,\ B_t \leq C M^{1-r} t^{-(r+1)}, \qquad \sum_{t=1}^{T} c_t w_t = \Theta\left(M^{1-r} \sum_{t=1}^{T} t^{-r} \right). \tag{57}$$

Here constants and the threshold may depend on $\mathcal{B}, a, b, \mu, \eta$, but not on M, T or the sequences. Thus the positive prediction-error remainder is summable for every $q > 1$, whereas task loss has the three regimes in Table 1.

To see the mechanism, define $s(H) = (\mathcal{R}')^{-1}(\mathcal{R}'(1) - \eta H)$. Integration and inversion of (56) give

$$s(H) \sim \left(\frac{\mu c_0}{(q-1)\eta H} \right)^r, \qquad [\mathcal{R}''(s(H))]^{-1} = \Theta(H^{-(r+1)}).$$

The conjugate divergence has displacement $\eta(m_t - c_t) = O(M)$, and every intermediate cumulative argument is $\Theta(Mt)$. Its integral Taylor formula gives the first bound in (57). Separately, $c_t w_t = \Theta(M^{1-r} t^{-r})$, giving the second.

Theorem 17 (Uniform integral approximation and policy tails)**.** *Under the assumptions of Theorem 16, for $T > n \geq 1$,*

$$\left| \sum_{t=n+1}^{T} c_t w_t - \int_{S_n}^{S_T} s(h)\, dh \right| \leq C M^{1-r} n^{-r}. \tag{58}$$

For any two allowed policies on the same costs and $T > n \geq 0$,

$$\sum_{t=n+1}^{T} c_t |w_t - w_t'| \leq C' M^{1-r} (1+n)^{-r}. \tag{59}$$

The constants and scale threshold are uniform as in Theorem 16.

Proof outline. Implicit differentiation gives $|s'(H)| = \eta / \mathcal{R}''(s(H)) = O(H^{-r-1})$. For $t \geq 2$, replacing the hint-shifted evaluation by any point of $[S_{t-1}, S_t]$ changes the cost-weighted action by at most $C M^{1-r}(t-1)^{-r-1}$. Integration and summation give (58). Applying the same derivative bound between two hints also covers $t = 1$ and proves (59). Appendix M supplies the details. □

Every pair of policies therefore has an absolutely convergent loss difference for every $q > 1$. The transition at $q = 2$ concerns the common integral term, rather than a transition in the cumulative value of changing hints.

The exact identity locates the source of growing task loss. Set $Q(s) = \mathcal{R}(s) - s\mathcal{R}'(s)$. For comparator 0, (6) becomes

$$\eta \sum_{t=1}^{T} c_t w_t = Q(b_{T+1}) - Q(1) + \sum_{t=1}^{T} (B_t - V_t), \qquad Q'(s) = -s\mathcal{R}''(s). \tag{60}$$

For $t \geq 2$, V_t obeys the same summable order as B_t. Meanwhile $Q'(s) \sim -\mu c_0 s^{1-q}$, so Q has a finite boundary limit if $q < 2$, logarithmic divergence if $q = 2$, and power divergence if $q > 2$. For the exact log potential, the comparator contribution and the first stability term both contain $\mu \log M$; these terms cancel. The stability tail is summable, recovering the scale-uniform logarithmic law. Appendix M gives this cancellation and the general estimates. Under (56) alone, the $q = 2$ rate is logarithmic in order; a bounded additive error requires more than curvature equivalence.

Evolving forward targets. The convex forward loss also changes with history. For a frozen power action map proportional to $(M\beta + m)^{-r}$, where $\beta \geq 0$ is normalized prior cumulative cost and $c = MX$ with $X \in [a, b]$, its normalized Bayes hint is $[\mathbb{E}(\beta + X)^{-r}]^{-1/r} - \beta$. It increases from a negative-order power mean toward $\mathbb{E}X$ as history grows. Proposition 18 gives the precise expansion and the limiting connection to the full barrier map.

M General curvature and evolving forward targets

M.1 Gradient and state asymptotics

Let $\mathcal{R}$ satisfy (56). For any fixed $s_0 > 0$ sufficiently small,

$$\mathcal{B}'(s) = \mathcal{B}'(s_0) - \int_s^{s_0} \mathcal{B}''(u)\, du.$$

Bounding $\mathcal{B}''(u)$ above and below by $(1 \pm \epsilon) c_0 u^{-q}$ for sufficiently small u and integrating shows

$$\mathcal{B}'(s) \sim -\frac{c_0}{q-1} s^{-(q-1)} \qquad (s \downarrow 0). \tag{61}$$

The finite integration constants are negligible because $q > 1$. The same asymptotic holds for $\mathcal{R}'(s)$ with c_0 replaced by μc_0. Since $\mathcal{R}''(s) = 1 + \mu \mathcal{B}''(s) \geq 1$, $\mathcal{R}'$ is strictly increasing. Its limit at zero is $-\infty$, and for $s \geq 1$ convexity of $\mathcal{B}$ gives $\mathcal{R}'(s) \geq s - 1 + \mu \mathcal{B}'(1) \to +\infty$. Thus the gradient is onto $\mathbb{R}$.

Define

$$s(H) = (\mathcal{R}')^{-1}(\mathcal{R}'(1) - \eta H).$$

As $H \to \infty$, $s(H) \downarrow 0$, and (61) gives

$$s(H) \sim C_* H^{-r}, \qquad C_* = \left(\frac{\mu c_0}{(q-1)\eta} \right)^r, \qquad r = \frac{1}{q-1}. \tag{62}$$

For clarity, this inversion can be justified directly: for any $\epsilon > 0$, the state equation is eventually trapped between constants arbitrarily close to $\mu c_0 s^{-(q-1)}/(q-1) = \eta H$. Taking the power $-r$ squeezes $s(H)H^r$ to C_*.

It follows that there are $H_0 > 0$ and positive constants C_1, C_2, C_3, C_4 such that for all $H \geq H_0$,

$$C_1 H^{-r} \leq s(H) \leq C_2 H^{-r}, \qquad C_3 H^{-(r+1)} \leq \frac{1}{\mathcal{R}''(s(H))} \leq C_4 H^{-(r+1)}. \tag{63}$$

Indeed, $\mathcal{R}''(s) \sim \mu c_0 s^{-q}$ and $rq = r + 1$. All constants may depend on the full potential and fixed parameters. An asymptotic assumption alone cannot give a uniform entry threshold over all potentials sharing the same q, c_0.

M.2 Proof of Theorem 16

Choose M_0 so that $aM_0 \geq H_0$. For $M \geq M_0$ and any $t \geq 1$,

$$aMt \leq S_{t-1} + m_t,\ S_t \leq bMt.$$

The played and updated states have the representations $w_t = s(S_{t-1} + m_t)$ and $b_{t+1} = s(S_t)$. Consider the dual endpoints

$$u_m = \mathcal{R}'(1) - \eta(S_{t-1} + m_t), \qquad u_c = \mathcal{R}'(1) - \eta S_t.$$

Their separation has magnitude $\eta|m_t - c_t| \leq \eta(b-a)M$. Every dual point between them is $\mathcal{R}'(1) - \eta H$ with $H \in [aMt, bMt]$. By (63), the conjugate Hessian along that interval is at most $C_4(aMt)^{-(r+1)}$. The integral Taylor formula in either orientation therefore gives

$$D_{\mathcal{R}^*}(u_m, u_c),\ D_{\mathcal{R}^*}(u_c, u_m) \leq \frac{1}{2}\eta^2(b-a)^2M^2C_4(aMt)^{-(r+1)} \leq CM^{1-r}t^{-(r+1)}.$$

Bregman duality proves the remainder bound in (57). Since $r > 0$,

$$\sum_{t=1}^{\infty} t^{-(r+1)} \leq 1 + \frac{1}{r} < \infty.$$

This proves summability for every $q > 1$. It is an upper bound rather than an unconditional lower bound, since $m_t = c_t$ makes both divergences zero.

For task loss, (63) gives directly

$$aC_1b^{-r}M^{1-r}t^{-r} \leq c_tw_t \leq bC_2a^{-r}M^{1-r}t^{-r}.$$

Summing yields the second part of (57), with uniform positive constants. For $r > 1$, $\sum_{t=1}^{T} t^{-r}$ lies between 1 and $1 + 1/(r-1)$. For $r = 1$, it is comparable to $\log(1+T)$. For $0 < r < 1$, it is comparable to T^{1-r}. Since $r > 1$, $r = 1$, and $r < 1$ correspond respectively to $q < 2$, $q = 2$, and $q > 2$, Table 1 follows.

M.3 Proof of Theorem 17

The uniform state estimates (63) imply, after increasing the same fixed threshold if necessary, $|s'(H)| \leq CH^{-r-1}$ for every $H \geq aM$. For $t \geq 2$, every point $h \in [S_{t-1}, S_t]$ and the argument $S_{t-1} + m_t$ are at least $aM(t-1)$, and their distance is at most bM. The mean value theorem yields

$$\begin{aligned}\left| c_ts(S_{t-1} + m_t) - \int_{S_{t-1}}^{S_t} s(h)\,dh \right| &\leq \int_{S_{t-1}}^{S_t} CbM[aM(t-1)]^{-r-1}\,dh \\ &\leq Cb^2a^{-r-1}M^{1-r}(t-1)^{-r-1}.\end{aligned}$$

For $n \geq 1$, the sum of $(t-1)^{-r-1}$ over $t \geq n+1$ is at most $(1+1/r)n^{-r}$. Summation proves (58). For two policies, the intermediate cumulative arguments are at least aMt, and their separation is at most $(b-a)M$ even at $t = 1$. Hence

$$c_t|w_t - w_t'| \leq Cb(b-a)a^{-r-1}M^{1-r}t^{-r-1}.$$

The full sum is at most a fixed multiple of M^{1-r}, and the sum after round $n \geq 1$ is at most a fixed multiple of $M^{1-r}n^{-r}$. Adjusting the constant gives (59) for every $n \geq 0$. The absolute convergence assertion follows directly.

For fixed M, this approximation also gives leading coefficients under the curvature equivalence alone. Define $C_* = (\mu c_0/((q-1)\eta))^r$, so that $s(H) \sim C_*H^{-r}$. For $0 < r < 1$, the integral term diverges and $L_T \sim C_*S_T^{1-r}/(1-r)$. For $r = 1$, $L_T \sim C_* \log S_T \sim C_* \log T$. For $r > 1$, write $L_\infty = \sum_{t \geq 1} c_tw_t < \infty$; letting the upper endpoint tend to infinity in (58) gives $L_\infty - L_T \sim C_*S_T^{1-r}/(r-1)$. In the last case the approximation error is $O(M^{1-r}T^{-r})$, smaller than the tail by a factor of order T^{-1}. These assertions hold uniformly over the allowed sequences at fixed M, because the state equivalence is uniform on sufficiently large tails and $aMT \leq S_T \leq bMT$.

M.4 Stability tails and the comparator potential

For $t \geq 2$, the base and played dual arguments correspond to $H = S_{t-1}$ and $H = S_{t-1} + m_t$. They differ by at most bM in cumulative cost and are both at least $aM(t-1)$. Repeating the conjugate-Hessian estimate gives

$$V_t \leq C' M^{1-r}(t-1)^{-(r+1)} \qquad (t \geq 2). \tag{64}$$

Thus $\sum_{t\geq 2} V_t$ is finite and has the same uniform scale as the positive remainder. The initial term V_1 is finite for any fixed M, but its dependence on M must be analyzed separately.

For any $s > 0$, conjugacy gives $\mathcal{R}^*(\mathcal{R}'(s)) = s\mathcal{R}'(s) - \mathcal{R}(s) = -Q(s)$. Therefore (6) with $v = 0$ is exactly (60). Differentiation requires only C^2 regularity:

$$Q'(s) = -s\mathcal{R}''(s) \sim -\mu c_0 s^{1-q}.$$

If $1 < q < 2$, this derivative is integrable at zero, and for a finite limit $Q(0+)$,

$$Q(0+) - Q(s) \sim \frac{\mu c_0}{2-q} s^{2-q}.$$

If $q = 2$,

$$Q(s) = \mu c_0 \log(1/s) + o(\log(1/s)).$$

If $q > 2$,

$$Q(s) \sim \frac{\mu c_0}{q-2} s^{2-q}.$$

These statements follow by the same two-sided integration argument used in (61). Together with the convergent sums of B_t and V_t for fixed M, they locate any divergent growth in T in the comparator term. They do not replace the direct uniform-in-M task-loss proof above: initial contributions can cancel, as the exact log example demonstrates.

For $q = 2$, the curvature equivalence $\mathcal{B}''(s) \sim c_0 s^{-2}$ alone allows slowly varying lower-order corrections. It gives no bounded additive error after replacing $Q(s)$ by $\mu c_0 \log(1/s)$. This is why Theorem 16 asserts an order bound, while Theorem 12 makes the stronger bounded-error claim only for the specified quadratic-plus-log potential.

M.5 From the power-law target to the full potential

Proposition 18 (History-dependent forward Bayes target)**.** *Let $X \in [a, b]$, $r > 0$, and $\beta \geq 0$. For a frozen action map $z(m) = C(M\beta + m)^{-r}$ on $m \in [aM, bM]$, with $C, M > 0$, the forward Bayes hint for $c = MX$ is $M\alpha_r(\beta)$, where*

$$\alpha_r(\beta) = \left(\mathbb{E}(\beta + X)^{-r}\right)^{-1/r} - \beta \in [a, b]. \tag{65}$$

For nonconstant X, α_r is strictly increasing and strictly below $\mathbb{E}X$, with

$$\alpha_r(\beta) = \mathbb{E}X - \frac{(r+1)\operatorname{Var}(X)}{2\beta} + O_{a,b,r}(\beta^{-2}) \qquad (\beta \to \infty). \tag{66}$$

For the full potential (56)*, its exact target satisfies $m^\star_{F,M}(\beta)/M \to \alpha_r(\beta)$ as $M \to \infty$ at each fixed β, with prior sum $S_{\mathrm{prev}} = M\beta$. Separately, for the exact log potential, fixed M, and prior sum $S \to \infty$,*

$$m^\star_F(S) = \mathbb{E}c - \frac{\operatorname{Var}(c)}{S} + O_{M,a,b,\mu,\eta}(S^{-2}). \tag{67}$$

M.6 Exact derivatives for the growing-history target

The proof of Proposition 18 below gives the Taylor argument for (67). Here are exact derivative formulas that justify its remainder estimates. Write $A(H) = \eta H + \mu - 1$ and $D(H) = \sqrt{A(H)^2 + 4\mu}$. The state map can also be written as $s(H) = (D(H) - A(H))/2$, so

$$s'(H) = \frac{\eta}{2}\left(\frac{A(H)}{D(H)} - 1\right), \qquad s''(H) = \frac{2\mu\eta^2}{D(H)^3}, \qquad s'''(H) = -\frac{6\mu\eta^3 A(H)}{D(H)^5}.$$

Expanding these expressions at $H = S + u$ gives the derivative estimates used in the proof, uniformly on each fixed bounded interval of u. For fixed M, both the support of $c - \mathbb{E}c$ and the possible target deviations lie in such an interval. The third-derivative bound therefore controls the Taylor remainders uniformly over all possible costs and target locations. This uniformity is at fixed cost scale; the interval grows with M. Neither the fixed-history limit nor (67) justifies interchanging the large-scale and growing-history limits without further estimates.

M.7 Complete proof of Proposition 18

Proof. The decreasing map defines a strictly convex local prediction potential through $\Psi'(m) = -z(m)/\eta$. State matching $z(m_F^\star) = \mathbb{E}z(MX)$ gives (65) and a unique solution in the hint interval. Put $A_p = \mathbb{E}(\beta + X)^{-p}$. Differentiation and Jensen's inequality give

$$\alpha_r'(\beta) = \frac{A_{r+1}}{A_r^{1+1/r}} - 1 \geq 0,$$

strictly for nonconstant X. Jensen applied to $x \mapsto (\beta + x)^{-r}$ also gives $\alpha_r(\beta) \leq \mathbb{E}X$, with the same strictness. Boundedness of X yields

$$A_r = \beta^{-r}\left(1 - \frac{r\mathbb{E}X}{\beta} + \frac{r(r+1)\mathbb{E}X^2}{2\beta^2} + O(\beta^{-3})\right).$$

Taking the power $-1/r$ and subtracting β proves (66). For the general potential, write $s(H) \sim C_* H^{-r}$ as above. For each fixed β, the relative error is uniformly small for $H \geq M(\beta + a)$ as $M \to \infty$. Hence $s(M(\beta + x))/(C_* M^{-r})$ converges uniformly to $(\beta + x)^{-r}$ on $x \in [a, b]$. The exact normalized target lies in $[a, b]$. Every convergent subsequence must therefore solve the limiting state-matching equation, whose unique solution is $\alpha_r(\beta)$. Compactness then gives convergence of the full sequence:

$$\frac{m_{F,M}^\star(\beta)}{M} \longrightarrow \alpha_r(\beta) \qquad (M \to \infty,\ \beta \text{ fixed}). \tag{68}$$

At $\beta = 0$ this is the negative-order power mean, including the harmonic mean when $r = 1$. The power action map is a local model; the stated limiting connection is what relates it to a globally admissible mirror map.

For the exact log map and fixed M, differentiation of (42) gives, uniformly for bounded u,

$$s'(S+u) = -\frac{\mu}{\eta S^2} + O(S^{-3}), \qquad s''(S+u) = \frac{2\mu}{\eta S^3} + O(S^{-4}), \qquad s'''(S+u) = O(S^{-4}).$$

Put $H = S + \mathbb{E}c$ and $\Delta = m_F^\star(S) - \mathbb{E}c$. The target lies in $[aM, bM]$, so both Δ and $c - \mathbb{E}c$ are bounded for fixed M. Expanding $s(H + \Delta) = \mathbb{E}s(H + c - \mathbb{E}c)$ to second order yields

$$s'(H)\Delta = \tfrac{1}{2}s''(H)(\mathrm{Var}(c) - \Delta^2) + O(S^{-4}).$$

Because $|s'(H)|$ is bounded below by a positive multiple of S^{-2}, this first implies $\Delta = O(S^{-1})$. Substitution gives $\Delta = \frac{1}{2}\mathrm{Var}(c)s''(H)/s'(H) + O(S^{-2})$. The derivative ratio is $-2/S + O(S^{-2})$, proving (67). □

M.8 Initial scale cancellation for the log barrier

For the exact log potential, $Q(s) = \frac{1}{2} - \frac{1}{2}s^2 + \mu - \mu \log s$, so

$$\begin{aligned} P_T(0) &= \tfrac{1}{2}(1 - b_{T+1}^2) - \mu \log b_{T+1} = \mu \log M + \mu \log T + O(1), \\ V_1 &= \tfrac{1}{2}(w_1 - 1)^2 + \mu(w_1 - 1 - \log w_1) = \mu \log M + O(1). \end{aligned} \tag{69}$$

Indeed, the uniform state bounds give $b_{T+1} = \Theta((MT)^{-1})$ and $w_1 = \Theta(M^{-1})$; both states are bounded above under (41). The errors in (69) are therefore uniform over M, T and the allowed sequences. Subtracting V_1 cancels the initial $\mu \log M$ contribution. The remaining sums of B_t and V_t for $t \geq 2$ are uniformly bounded, recovering $\eta L_T = \mu \log T + O(1)$ from the exact identity. Discarding V_1 would lose this uniform conclusion. Under (56) alone, the $q = 2$ task rate is logarithmic in order; the bounded-error formula (44) uses the exact log potential.

N A STABLE VERTEX IN MULTIPLE DIMENSIONS

The disappearance of leading policy differences extends beyond separable half-lines. Let $d \geq 1$, write $\mathcal{Z} = \{z \in \mathbb{R}^d : z_i > 0,\ \sum_{i=1}^d z_i < 1\}$, and set $z_0 = 1 - \sum_{i=1}^d z_i$. Choose weights $\mu_0, \dots, \mu_d > 0$ and the coupled simplex potential

$$R(z) = -\sum_{i=0}^{d} \mu_i \log z_i. \tag{70}$$

Reduced costs $c_{t,i}$ represent the cost of alternative i minus that of alternative 0. The comparator $z = 0$ puts all mass on alternative 0. Fix an interior initialization and $\eta > 0$.

Theorem 19 (Common logarithmic loss under a cumulative margin)**.** *Suppose $|c_{t,i}| \leq G$ and $|m_{t,i}| \leq H$ for finite $G, H > 0$. Assume there are $\delta > 0$ and an integer $t_0 \geq 1$ such that*

$$S_{t,i} \geq \delta t \qquad (i = 1, \dots, d,\ t \geq t_0). \tag{71}$$

Uniformly over these cost sequences and all allowed predictable hints,

$$F_t + B_t = O(t^{-2}), \qquad L_T = \frac{\sum_{i=1}^d \mu_i}{\eta} \log T + O(1). \tag{72}$$

For any two such policies and $T > n \geq 1$,

$$\sum_{t=n+1}^{T} |\langle c_t, z_t - z'_t \rangle| \leq \frac{C}{n}. \tag{73}$$

Constants depend only on $d, G, H, \delta, t_0, \eta$, the weights, and initialization. Individual costs and hints may have either sign.

Proof outline. Put $\theta_{t,i} = \eta(S_{t-1,i} + m_{t,i}) - y_{1,i}$. The first-order equations are $\mu_i / z_{t,i} = \theta_{t,i} + \mu_0 / z_{t,0}$. The margin makes every $\theta_{t,i}$ of order t, hence $z_{t,i} = O(t^{-1})$ and $z_{t,0} = 1 - O(t^{-1})$. It follows uniformly that

$$z_{t,i} = \frac{\mu_i}{\eta S_{t-1,i}} + O(t^{-2}), \qquad [\nabla^2 R(z_t)]^{-1} \preceq \operatorname{diag}(z_{t,i}^2/\mu_i) = O(t^{-2})I.$$

The same bounds hold along bounded hint segments. Theorem 3 proves the policy tail bound, and conjugate Taylor expansion bounds both divergences. Finally, $c_{t,i}/S_{t-1,i} = \log(S_{t,i}/S_{t-1,i}) + O(t^{-2})$ for all sufficiently large t. Summing telescopes, and $\delta T \leq S_{T,i} \leq GT$ gives (72). Appendix O handles initialization and all uniform constants. □

The cumulative margin permits temporary changes in the best alternative. It requires that accumulated costs eventually favor a fixed vertex by a linear margin. The coefficient counts the barrier weights on the vanishing coordinates; the weight on the winning coordinate affects only the bounded term. Together with (33), this identifies the geometric distinction: approaching a vertex suppresses all reduced directions here, whereas approaching a face can leave useful tangential sensitivity.

A simple vertex has exactly d active inequalities with independent normals. Their slacks give invertible coordinates centered at the vertex.

Corollary 20 (Simple polytope vertices)**.** *The conclusions of Theorem 19 also hold at a simple vertex of a bounded full-dimensional polytope with a convex quadratic-plus-log-barrier potential, when costs, hints, and the cumulative margin are expressed in its active-slack coordinates. The logarithmic coefficient is the sum of the active barrier weights divided by η.*

Appendix O proves localization and the resulting expansion; constants may additionally depend on the polytope and potential.

O PROOF OF THE MULTIDIMENSIONAL VERTEX THEOREM

The domain in Section N is bounded, and its log potential tends to infinity at every boundary point. Consequently $R(z) - \langle y, z\rangle$ has an interior minimizer for every $y \in \mathbb{R}^d$. Its Hessian is

$$\nabla^2 R(z) = \operatorname{diag}_{1\le i\le d}(\mu_i/z_i^2) + \frac{\mu_0}{z_0^2}\mathbf{1}\mathbf{1}^\top \succ 0, \tag{74}$$

where $\mathbf{1}$ is the vector of d ones. Strict convexity and the inverse function theorem verify all mirror assumptions.

O.1 UNIFORM STATE AND INVERSE-HESSIAN ESTIMATES

Put $K = \max\{G, H\}$. The following estimates hold for every counterfactual hint $u \in [-K, K]^d$, so they cover played actions, updated states, and the segments joining their dual coordinates. Let $z_t(u)$ denote that action and put $\theta_i = \eta(S_{t-1,i} + u_i) - y_{1,i}$ and $\nu = \mu_0/z_{t,0}(u)$. The gradient equation reads

$$z_{t,i}(u) = \frac{\mu_i}{\theta_i + \nu} \quad (i \ge 1). \tag{75}$$

For all t larger than a fixed integer N, the cumulative margin and bounded costs give

$$\frac{\eta\delta}{2}t \le \theta_i \le C_0 t.$$

Here and below N and constants depend only on the parameters listed in Theorem 19. Since $\nu > 0$, (75) implies $z_{t,i}(u) \le 2\mu_i/(\eta\delta t)$. Increasing N makes their sum at most $1/2$. Therefore $z_{t,0}(u) \ge 1/2$ and $\mu_0 \le \nu \le 2\mu_0$. The difference between the denominator in (75) and $\eta S_{t-1,i}$ is uniformly bounded. Both denominators have lower bounds proportional to t, proving

$$z_{t,i}(u) = \frac{\mu_i}{\eta S_{t-1,i}} + O(t^{-2}). \tag{76}$$

Inverting the positive-semidefinite inequality from (74) yields

$$0 \prec J_t(u) \preceq \operatorname{diag}_{i\ge1}(z_{t,i}(u)^2/\mu_i) \preceq C_1 t^{-2} I.$$

For the finitely many $t < N$, all dual states belong to a common compact set: costs and counterfactual hints are bounded, and initialization is fixed. The inverse mirror map sends it to a compact subset of $\mathcal{Z}$. Its continuous inverse Hessian is uniformly bounded there. Increasing C_1 establishes $J_t(u) \preceq C_1 t^{-2} I$ for every $t \ge 1$.

O.2 DIVERGENCES AND POLICY DIFFERENCES

The prediction potential has Hessian J_t. The hint and cost endpoints lie in $[-K, K]^d$, and their distance is at most $\sqrt{d}(G + H)$. The integral Taylor formula therefore bounds each of F_t, B_t by $\eta^2 C_1 d(G + H)^2/(2t^2)$. For policy differences use $\|c_t\| \le \sqrt{d}G$ and hint diameter at most $2\sqrt{d}H$ in (13). The absolute sum after round n is at most

$$2\eta C_1 dGH \sum_{t=n+1}^{\infty} t^{-2} \le \frac{2\eta C_1 dGH}{n}.$$

In particular every policy difference converges absolutely as $T \to \infty$.

O.3 THE LEADING TASK COEFFICIENT

Increase $N \ge 2$ so that $S_{t-1,i} > 0$ and $|c_{t,i}/S_{t-1,i}| \le 1/2$ for $t \ge N$. Equation (76) and bounded costs give

$$\langle c_t, z_t\rangle = \sum_{i=1}^{d} \frac{\mu_i}{\eta}\frac{c_{t,i}}{S_{t-1,i}} + O(t^{-2}).$$

For $|x| \le 1/2$, the integral identity $x - \log(1+x) = \int_0^x u/(1+u)\,du$ implies $|x - \log(1+x)| \le x^2$. This holds for signed x and yields

$$\sum_{t=N}^{T} \frac{c_{t,i}}{S_{t-1,i}} = \log \frac{S_{T,i}}{S_{N-1,i}} + O(1) = \log T + O(1).$$

The final equality uses $\delta T \le S_{T,i} \le GT$ and $\delta(N-1) \le S_{N-1,i} \le G(N-1)$. Each earlier task loss has absolute value at most G, because $\sum_{i \ge 1} z_{t,i} < 1$. Those finitely many terms are uniformly bounded; the same argument covers $T < N$ after enlarging the constant. This proves (72) without requiring any individual cost gap to be positive.

O.4 Proof of Corollary 20

Let the original polytope lie in $\mathbb{R}^d$, and let its simple vertex be v. Enumerate the active affine slacks as $s_1, \dots, s_d$ and define $z_i = s_i(w)$. Independence of their normals gives an invertible matrix B such that $w = v + Bz$. An original cost $\widetilde{c}_t$ becomes $c_t = B^\top \widetilde{c}_t$, and hints transform in the same way. Then $\langle \widetilde{c}_t, w_t - v \rangle = \langle c_t, z_t \rangle$. The Bregman divergence and controller equations transform by composition of the potential with this affine map. All bounds and the margin in the corollary refer to these transformed coordinates.

The reduced open domain is bounded, is contained in $(0,\infty)^d$, and contains $(0,\epsilon)^d$ for some $\epsilon > 0$: every inactive inequality has strictly positive slack at v. The transformed potential can be written

$$R(z) = -\sum_{i=1}^{d} \mu_i \log z_i + \phi(z),$$

where ϕ consists of the quadratic term, any affine term, and the inactive log barriers. It is convex on the domain and has a C^2 extension to a neighborhood of 0. The complete potential has positive-definite Hessian and diverges at the boundary of a bounded domain, so it satisfies the standing mirror assumptions.

We first justify that the controller reaches this neighborhood, uniformly over bounded hints. For a counterfactual hint $u \in [-K, K]^d$ with $K = \max\{G, H\}$, put $\theta_t = \eta(S_{t-1} + u) - y_1$. The action minimizes $\langle \theta_t, z \rangle + R(z)$. The margin implies $c_1 t \le \theta_{t,i} \le c_2 t$ for large t, for fixed constants $c_1, c_2 > 0$. The potential is bounded below on its bounded domain: its quadratic and affine terms are bounded below, and every slack is bounded above, so each negative logarithm is bounded below. As a feasible comparison use $z = t^{-1}\mathbf{1}$ for large t. The objective there is at most $(\sum_i \mu_i) \log t + C$ because ϕ is bounded near zero. Optimality consequently gives

$$c_1 t \sum_i z_{t,i}(u) \le \Big(\sum_i \mu_i\Big) \log t + C',$$

so $z_t(u) \to 0$ uniformly.

In this neighborhood the gradient of ϕ is uniformly bounded. The first-order equations improve the preliminary localization to

$$\frac{\mu_i}{z_{t,i}(u)} = \eta S_{t-1,i} + \eta u_i - y_{1,i} + \partial_i \phi(z_t(u)), \qquad z_{t,i}(u) = \frac{\mu_i}{\eta S_{t-1,i}} + O(t^{-2}).$$

Convexity of ϕ gives $\nabla^2 R(z) \succeq \operatorname{diag}(\mu_i / z_i^2)$ and hence the same inverse-Hessian estimate as for the simplex. The proof of Theorem 19 now applies: conjugate Taylor expansion bounds both orientations, sensitivity gives the policy tail, and logarithmic telescoping gives the coefficient $\sum_i \mu_i/\eta$. For finitely many early rounds, bounded costs, hints, and domain again give uniform constants. The argument requires a simple vertex with no additional displayed active inequalities; redundant active facets cannot be absorbed into the smooth function ϕ.

P Deterministic numerical illustrations and checks

This paper's conclusions are mathematical. The calculations below provide reproducible illustrations of the stated model and checks on algebraic identities. They use no application data and do not evaluate a deployed scheduler. The supplementary script `numerical-checks.py` in `experiments/` generates Figure 2, its numerical data, and a machine-readable validation report.

P.1 ILLUSTRATION PROTOCOL

For panels (a)–(c), use the exact quadratic-plus-log potential with $\mu = 1$, $\eta = 1/2$, $a = 1$, $b = 2$, and $M = 20$. For trajectory plots, use the deterministic alternating costs $20, 40, 20, 40, \dots$ and $T = 4096$.

Panel (a) uses the fixed arithmetic-mean hint $m_t = 30$. It compares cumulative half squared error with $\sum_t F_t/\eta^2$ and $\sum_t B_t/\eta^2$. The geometric curves approach finite limits while squared error grows linearly. This comparison concerns error terms, not complete regret bounds.

Panel (b) plots $L_T - (\mu/\eta)\log T$ for four predictable policies: constant arithmetic-mean hint 30, constant reset harmonic-mean hint $80/3$, maximal hint 40, and the sample mean of past costs (initialized at 30). The harmonic label denotes a fixed reset target, not the changing forward Bayes predictor. All policies use the same costs and the same η, μ. The bounded residuals illustrate the common leading coefficient in Theorem 12.

Panel (c) uses a two-point distribution with equal probabilities at $X = 1$ and $X = 2$. For each frozen normalized history β, a scalar root solver computes the exact forward Bayes hint from

$$s(M(\beta + \alpha)) = \tfrac{1}{2}s(M(\beta + 1)) + \tfrac{1}{2}s(M(\beta + 2)), \qquad \alpha \in [1, 2].$$

It is compared with the shifted harmonic expression $\alpha_1(\beta) = [\frac{1}{2}(\beta + 1)^{-1} + \frac{1}{2}(\beta + 2)^{-1}]^{-1} - \beta$ and the reverse target $3/2$. This separates a reset asymptotic target from the target along growing histories.

Panel (d) uses the same costs and the constant hint 30, with potentials

$$\mathcal{B}_q(s) = \begin{cases} s^{2-q}/[(q-1)(q-2)], & q \neq 2, \\ -\log s, & q = 2, \end{cases} \qquad \mathcal{B}_q''(s) = s^{-q},$$

for $q \in \{3/2, 2, 3\}$. The corresponding state equation is solved in $\log s$ coordinates. It plots $L_T/M^{(q-2)/(q-1)}$, illustrating the bounded, logarithmic, and power-growth regimes. All plotted data are generated by the included script; none are inherited from a different controller or a different experiment.

P.2 INDEPENDENT NUMERICAL CHECKS

The supplementary numerical report records the following checks.

- It evaluates the regret identity for a two-dimensional quadratic-plus-log potential on a simplex, with signed costs and interior and boundary comparators. It computes divergences from their defining potentials and compares the two sides using 60-decimal-digit arithmetic.
- It checks primal–dual divergence orientation, the reverse-loss gradient, and the conditional excess-risk identity on a finite cost distribution in the same nonseparable geometry.
- It integrates smooth approximations to the extremal directional-curvature profiles to check approach to the sharp factor in Theorem 9.
- It checks the explicit half-line remainder and policy-difference bounds over several cost scales and parameters, including a case where the reverse loss is nonconvex before the stated burn-in and convex afterward.
- For two-point i.i.d. costs, it enumerates all possible counts of past observations with their binomial probabilities to compute sample-mean excess risk. It compares these finite sums with both constants in (47), avoiding a Monte Carlo estimate.

These finite calculations can detect algebraic and implementation errors; the uniform claims and asymptotic rates rely on the proofs.

P.3 TASK SENSITIVITY, SIGNED COSTS, AND BOUNDARY FACES

The script `policy-checks.py` produces the additional checks and Figure 3. It saves trajectory data and a validation report alongside its source in the supplementary material. On 30 simplex

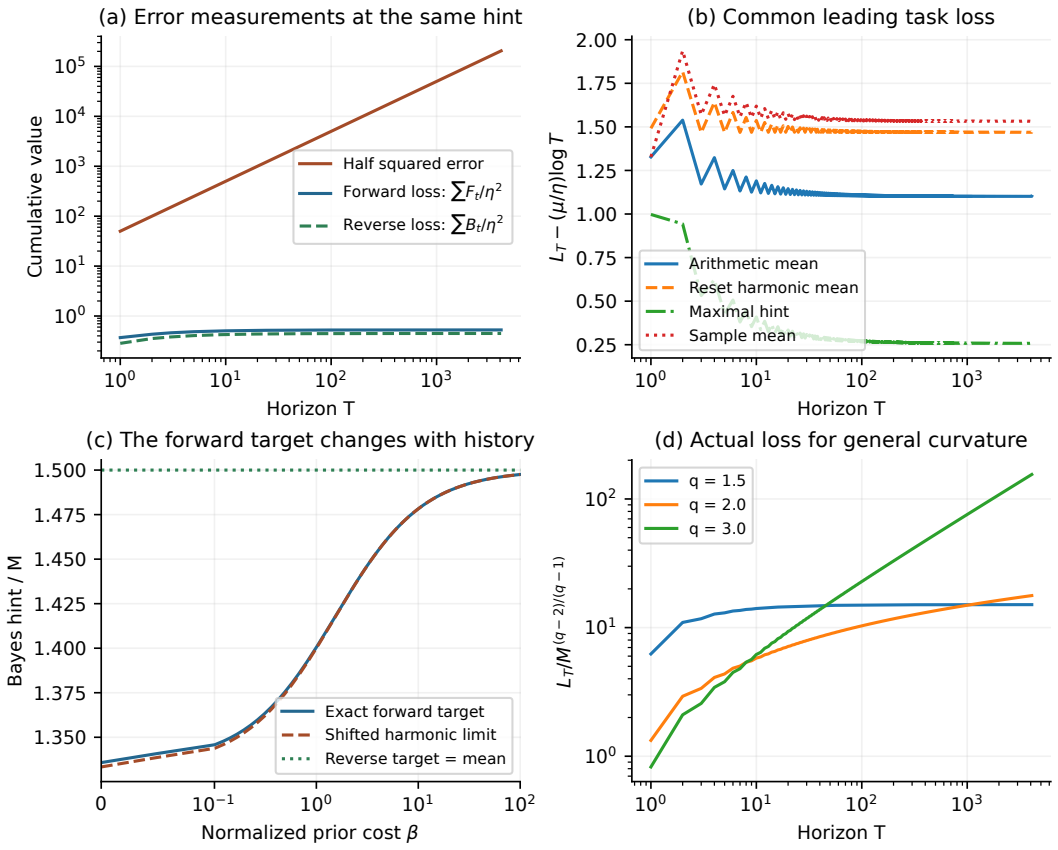

Figure 2: Deterministic illustrations of the specified mirror controller. (a) Cumulative prediction-error terms at the same mean hint. (b) Task-loss residual after subtracting the common logarithmic leading term, for four policies at fixed controller parameters. (c) Exact and limiting forward targets as normalized history grows, compared with the reverse target. (d) Task loss for three boundary-curvature exponents, normalized by its predicted cost-magnitude scale. The horizontal variable is the horizon in (a), (b), and (d), and normalized prior cumulative cost in (c). The calculations illustrate the theorems rather than providing additional empirical claims.

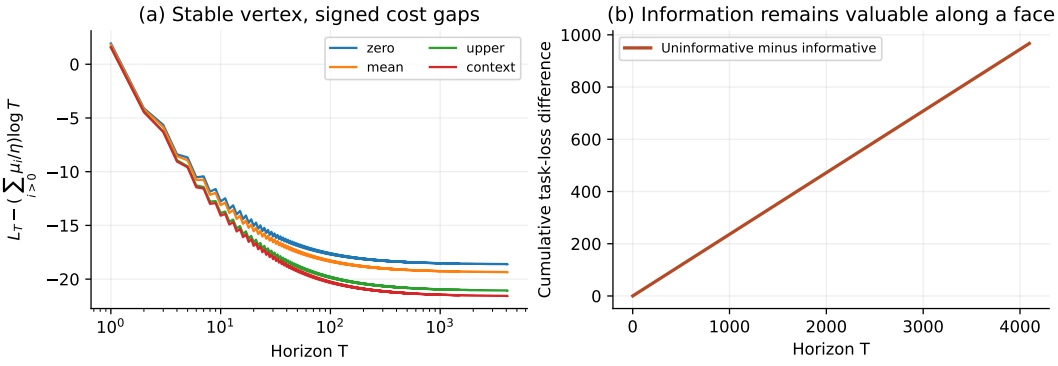

Figure 3: Task comparisons at fixed controller parameters. (a) On the coupled simplex, signed per-round cost gaps maintain a cumulative vertex margin; four hint policies have a common logarithmic leading loss. (b) Along a boundary face, informative hints yield an exact linear reduction in task loss. The second panel plots an analytic identity; the first is a deterministic illustration of Theorem 19.

instances generated with NumPy seed 7129, it checks the sensitivity formula by numerical integration of the inverse Hessian. It also checks the stability correction against independently evaluated conjugates. These are finite double-precision checks with an absolute residual tolerance of 10^{-9}.

For the simplex panel, use three reduced coordinates, step size $\eta = 0.3$, barrier weights $(\mu_0, \mu_1, \mu_2, \mu_3) = (0.7, 0.4, 0.9, 1.1)$, and initial reduced state $(0.20, 0.25, 0.15)$. With $\epsilon_t = (-1)^{t+1}$, the deterministic reduced costs are

$$c_t = (1, 1.4, 0.8) + \epsilon_t(2, 2.5, 1.6).$$

All even-round gaps are negative, while each cumulative coordinate is at least $0.8t$. The horizon is 4096. The four policies use zero hints, the constant mean vector $(1, 1.4, 0.8)$, the constant upper vector $(4, 4, 4)$, and the exact context-dependent costs. All hints are predictable; the last policy observes ϵ_t as its context. The plot subtracts the theoretical coefficient $8 \log T$ from cumulative loss, without fitting a slope or offset. The program additionally checks that the last doubling-block slope is within one percent of 8. Finite maxima of scaled Hessians, divergences, and policy differences are reported as numerical diagnostics, not as proofs of uniform bounds.

The face panel uses the product-domain construction (33) with $\eta = 1/2$ and $\delta = 1$. Its common positive coordinate cancels from the policy difference. The plotted line is the exact formula $\delta h T$; the code separately checks it from the roundwise actions for nine pairs of step sizes and tangential cost magnitudes. For $q \in \{3/2, 2, 3\}$, the code also compares direct integration of the scalar state map with the identity

$$\int_{S_n}^{S_T} s(H)\, dH = \frac{Q(s(S_T)) - Q(s(S_n))}{\eta},$$

and reports the normalized errors in (58) and the policy tails on four doubling intervals. Finally, exact binomial sums check the total-variation bound used in the minimax proof at seven sample sizes. For a nonsimplex check of Corollary 20, use the square $(0, 1)^2$, $\eta = 0.4$, initial state $(0.3, 0.4)$, active weights $(0.8, 1.2)$, and opposite-facet weights $(1.3, 0.7)$. Add the quadratic term with Hessian $\left(\begin{smallmatrix} 2 & 0.3 \\ 0.3 & 1 \end{smallmatrix}\right)$. At prior sums $(t-1)(1, 1.5)$ for $t \in \{16, 64, 256, 1024, 4096\}$, a damped Newton solve checks the state expansion for hints $(-2, 3)$ and $(3, -2)$. The code verifies stationarity and, at the largest horizon, relative error below 0.5% in the leading state expression.

Q Protocols for boundary reduction and learned policies

The script `boundary_learning.py` in the supplementary `experiments/` folder produces Figure 1. It uses NumPy, SciPy, and Matplotlib, and saves all plotted arrays and a verification report. The entire supplement is reproduced by `python run_all.py` from its extracted directory. The supplied `requirements.txt` records the tested versions; no network access is needed during execution.

Q.1 A NONSIMPLE VERTEX WITH SIGNED COSTS

Use the three-dimensional pyramid

$$P = \{(x, y, z) : z + x \geq 0,\ z - x \geq 0,\ z + y \geq 0,\ z - y \geq 0,\ 1 - z \geq 0\}.$$

Four facets meet at the apex $(0, 0, 0)$, so the vertex is nonsimple. In the displayed slack order, the barrier weights are $(0.4, 0.7, 0.6, 0.9, 0.8)$. Add the quadratic term with Hessian

$$Q = \begin{pmatrix} 0.3 & 0.05 & 0 \\ 0.05 & 0.4 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

and use $\eta = 0.4$ and initial state $(0, 0, 0.45)$. For $\epsilon_t = (-1)^{t+1}$, let $c_t = (0, 0, 1) + \epsilon_t(1.6, -1.2, 1.5)$, with horizon $T = 4096$. The four predictable policies use hints 0, $(0, 0, 1)$, $(0, 0, 3)$, and c_t, respectively; the last observes the parity context before acting. These costs satisfy (15) with $\delta = 1/4$ from round 3 onward. Indeed $|x|, |y| \leq z$ on P and $\|(x, y, z)\| \leq \sqrt{3}z$; the cumulative coefficients satisfy $S_{t,3} - |S_{t,1}| - |S_{t,2}| \geq (\sqrt{3}/4)t$ for $t \geq 3$. Individual even-round vertical costs are negative.

A damped Newton method solves the full stationarity equation, using the preceding solution only as a numerical initialization. It checks strict feasibility and relative stationarity residual at most $2 \cdot 10^{-12}$. The plotted residual subtracts the predicted coefficient $\nu_v/\eta = 2.6/0.4 = 6.5$ without fitting. The four final doubling-block slopes are 6.482, 6.480, 6.476, and 6.458. These finite slopes agree with the asymptotic coefficient within one percent. The code also records scaled inverse Hessians and absolute differences between policies on the last half of the trajectory.

Q.2 FACE REDUCTION AND BLOCK SENSITIVITY

Use the two-dimensional reduced simplex with weights $(1, 1, 0.7)$, $\eta = 0.5$, and initial reduced state $(0.4, 0.2)$, so the initial probabilities of alternatives 0 and 1 are equal. For $t \in \{8, 16, \ldots, 4096\}$, take prior cumulative reduced costs $(2, 2(t-1))$ and current hint $(-0.7, 2)$. The corresponding face has weights $(1, 1)$, initial probability $1/2$, and first reduced cumulative cost 2. An independent scalar multiplier solver computes full and face actions. The code compares their states and inverse Hessians, checks the tangential action derivative by a centered finite difference with step 10^{-3}, and changes only the normal hint from 2 to -2. At the final doubling, state error decreases by a factor approximately 0.500, whereas the cross-block inverse Hessian and the normal-hint action difference decrease by factors approximately 0.250. The middle panel plots the state and tangent-block errors multiplied by t.

Q.3 TWO LEARNED POLICIES ON A COUPLED SIMPLEX

Use the paired-cost construction of Theorem 6 with $\theta = 0.35$, $\delta = 1$, $M = 2$, $\eta = 0.5$, weights $(1, 1, 0.7)$, and the same initial reduced state $(0.4, 0.2)$. There are $K = 2048$ pairs, or 4096 rounds, in each of 64 independent streams. The streams use separate NumPy generators with consecutive integer seeds beginning at 20260926. All policies within one stream share its costs. On odd rounds the mean and saturated learners use only the previous pairs; on even rounds they use the already observed cost with its sign reversed. An oracle using the unknown sign of θ is recorded only as a reference and is not supplied to either learner. No hyperparameter is selected using the observed results.

The predicted leading difference per pair is $\Delta_\theta = 0.05222974$. At $K = 2048$, the coupled-simplex experiment gives a mean task-loss difference of 105.673 with standard error 0.540 across streams; the plotted band is the mean plus or minus two standard errors at each horizon. It is a descriptive pointwise band, not a simultaneous confidence region. The leading asymptotic line is $K\Delta_\theta$; it omits finite-horizon corrections.

For an independent check, enumerate the $k - 1$ previous positive outcomes under their binomial distribution at each odd round of the product model. This gives an exact expected final policy gap of 106.592 and saturated-policy task excess of 0.4384 relative to its task oracle, below the proved bound 2.7814. Applying the same cost streams to the face controller agrees with that exact enumeration within 1.23 standard errors at the final horizon. All raw outcomes, cumulative losses, standard errors, and enumeration results are included in the supplement. Finite diagnostics complement the proofs and do not verify their uniform quantifiers.