ONLINE NON-MONOTONE DR-SUBMODULAR MAXIMIZATION BEYOND $1/e$

Anonymous authors
Paper under double-blind review

ABSTRACT

We give a 0.4013-approximate algorithm for adversarial online maximization of nonnegative, non-monotone DR-submodular functions over compact convex down-closed sets. To our knowledge, this is the first polynomial-oracle online guarantee strictly above $1/e$ for this problem class. Each action is committed before current-function feedback. The algorithm maintains one feasible state, uses one feasible gradient query and one Euclidean projection per round, and achieves $O(DB\sqrt{dT})$ expected approximate regret for diameter D and gradient bound B. The main result is a first-order inequality that holds simultaneously for every comparator. A strengthened delayed-trajectory comparison and an endpoint-retaining coordinate-chain comparison have complementary nonlinear residuals. A common-state substitution aligns their linear terms, and a fixed mixture of two feasible actions cancels the nonlinear terms exactly. This replaces objective-dependent offline auxiliary optimization with one online linear update. Continuous differentiability suffices; no positive anchor or interior-point condition is required. We prove extensions to non-anticipating adaptive adversaries and stochastic gradients. For every absolute tolerance $\varepsilon > 0$, the offline specialization attains expected coefficient $\alpha_* > 0.4013$ up to additive error ε using $O(1 + dD^2B^2/\varepsilon^2)$ feasible gradient queries and projections.

1 INTRODUCTION

Continuous diminishing returns (DR) provide a useful alternative to concavity in resource allocation, influence maximization, and probabilistic inference problems (Bian et al., 2017; 2019). A differentiable function is DR-submodular when increasing its argument can only decrease each partial derivative. Non-monotonicity allows additional allocation to reduce reward. Down-closed constraints ensure that reducing an allocation preserves feasibility. The online problem combines these ingredients with changing objectives, where a learner must choose an allocation before seeing the function that evaluates it (Chen et al., 2018).

The central approximation question has had different answers offline and online. Buchbinder & Feldman (2024) give the general offline coefficient 0.401, under smoothness and constraint-oracle assumptions. For adversarial online maximization over down-closed convex sets, the best prior coefficient is $1/e$ (Thang & Srivastav, 2021; Zhang et al., 2023; Pedramfar et al., 2024; Lu et al., 2026). In particular, Lu et al. (2026) attain $O(\sqrt{T})$ approximate regret with one feasible gradient query per round. We improve the coefficient while preserving that query budget and time exponent:

$$\mathbb{E}\sum_{t=1}^{T} f_t(p_t) \;\geq\; 0.4013\,\mathbb{E}\left[\max_{o\in\mathcal{K}}\sum_{t=1}^{T} f_t(o)\right] - O(DB\sqrt{dT}), \tag{1}$$

where D is the feasible-set diameter and B bounds the gradient norm on that set. To our knowledge, this is the first polynomial-oracle online guarantee strictly exceeding $1/e$ for this continuously differentiable function class over general down-closed convex sets. Feasibility of the queries is preserved, rather than introduced relative to Lu et al. (2026). The gain is in the leading approximation coefficient, although our estimator has a larger dimension-dependent regret constant, refined below using the feasible-set radius.

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.

Why online is harder than offline. An offline solver may inspect the same objective repeatedly before returning its solution. In online scenerio, however, the current action earns its reward before any current-function feedback, and the next objective may change adversarially. A solution optimized for today's function therefore cannot justify tomorrow's reward. This is a genuine extension of the optimization requirement: an online algorithm applied to a constant sequence gives an offline solver by returning a uniformly sampled played action, whereas an offline approximation guarantee alone supplies no regret bound for a changing sequence. Our result allows arbitrary oblivious sequences and also non-anticipating adaptive adversaries. There is no stationarity or independent-sampling assumption on the objectives.

Technical novelty: a comparison valid before seeing the objective. The required inequality must hold for every feasible state x and every comparator o, using an action distribution fixed independently of the current function. An offline guarantee at an objective-dependent local optimum is insufficient: the online state was learned from past rewards, and the hindsight comparator optimizes their sum. We combine two pointwise comparisons whose residuals can be controlled by a single online learner. In the box objective $J_x(a) = f(x \odot (\mathbf{1} - a))$, the substitutions $a = \mathbf{1} - x$ and $b = \mathbf{1} - o$ make the linear residual proportional to $o - x$, the same difference as in the trajectory comparison. All evaluations of J_x map into $[0, x] \subseteq \mathcal{K}$. Thus the auxiliary construction is feasible without an interior point or a separate box optimizer.

Technical novelty: the two retained terms that permit cancellation. The delayed exponential expansion of Buchbinder & Feldman (2024) contains a mixed early/late second-order term. We lower-bound that term by three quarters of the union value and retain its contribution. Separately, an ordered coordinate-chain argument retains the endpoint value $J_x(0) = f(x)$, in addition to $J_x(\mathbf{1} - o) = f(x \odot o)$. The strengthened trajectory bound can then assign equal coefficients to these two values. A fixed random mixture of the trajectory and box actions cancels them exactly. The resulting coefficient is $0.4013063665\ldots$. The essential advance is this comparator-uniform cancellation, not a new online ascent method or a round-by-round application of an offline solver.

Contributions. Figure 1 outlines the organization of the results.

1. We prove the first-order structural comparison with approximation coefficient above 0.4013 in Theorem 3.4. The main text explains the mixed-term bound and endpoint-retaining chain argument, and gives complete proofs of the common-state substitution and exact cancellation. The full structural verifications are in Appendices B.5 and C.1.

2. We implement the comparison with one feasible gradient and one projection per round. Theorem 5.1 gives $O(DB\sqrt{dT})$ expected approximate regret against oblivious and non-anticipating adaptive adversaries, using a gradient bound only on $\mathcal{K}$ and allowing empty-interior feasible sets.

3. We extend the guarantee to unbiased stochastic gradients, and give an online-to-offline specialization with an explicit inverse-square accuracy–oracle tradeoff in Corollary 6.1.

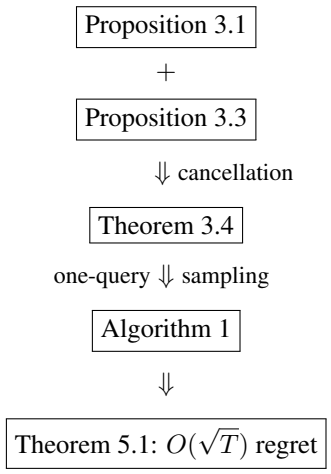


Figure 1: Organization of results.

Relation to prior techniques. The exponential expansion and delayed construction originate in Buchbinder & Feldman (2024). Coordinatewise hybrid telescoping is a standard ingredient of double-greedy analyses (Niazadeh et al., 2020; Roughgarden & Wang, 2018), and our ordered weighting and endpoint-retaining specialization are proved in full. The reduction to online linear optimization follows Pedramfar & Aggarwal (2024); Lu et al. (2026), and projected ascent is standard (Zinkevich, 2003). Detailed related work is provided in Appendix A.

Table 1: Approximation, accuracy, and cost. Online errors are cumulative approximate regret; offline errors are additive objective loss. Every offline row holds for every $\delta \in (0,1)$; $N_\delta = \lceil \delta^{-2} \rceil$ and $\Gamma_d = O(\sqrt{d})$. Online rows specialize the comparison fields to exact-gradient projected ascent. Buchbinder–Feldman instead require L-smoothness and their meta-solvability oracle. These are different oracle models and error scales, not a runtime dominance claim.

Setting	Reference	Coefficient	Error bound	Oracle budget / time
Online	Lu et al. (2026)	$1/e$	$O(DB\sqrt{T})$	One query + projection / round
	This work	$\alpha_* > \mathbf{0.4013}$	$O(DB\Gamma_d\sqrt{T})$	One query + projection / round
Offline	Buchbinder & Feldman (2024)	0.4010	$O(\delta D^2 L)$	$\mathrm{poly}(d, \delta^{-1})$ time
	This work	$\alpha_* > 0.4013$	$\delta DB\Gamma_d$	N_δ queries + N_δ projections

2 Problem and feedback model

Let $\mathcal{K} \subseteq [0,1]^d$ be nonempty, compact, convex, and down-closed. Then $\mathbf{0} \in \mathcal{K}$ and $D = \mathrm{diam}(\mathcal{K}) \le \sqrt{d}$. An exact Euclidean projection oracle $\Pi_{\mathcal{K}}$ is available; each projection is counted separately from an objective query. Products $x \odot o$, quotients, exponentials, and vector inequalities are coordinatewise. We write $x \oplus o = x + o - x \odot o$ and $[0, x] = \{z : 0 \le z \le x\}$.

Each reward $f_t : [0,1]^d \to \mathbb{R}_{\ge 0}$ is continuously differentiable and DR-submodular:

$$u \le v \implies \nabla f_t(u) \ge \nabla f_t(v), \qquad \sup_{z \in \mathcal{K}} \|\nabla f_t(z)\|_2 \le B. \tag{2}$$

DR-submodularity and nonnegativity hold on the cube $[0,1]^d$, and the quantitative gradient bound is needed only on $\mathcal{K}$. Boundary gradients are defined by continuous extension. No gradient-Lipschitz constant, positive anchor, or interior point is assumed. For example, $\{x \in [0,1]^d : \sum_i x_i \le 1,\ x_d = 0\}$ is covered despite having no all-positive point. The algorithm uses feasible boxes $[0, x]$, not neighborhoods requiring positive distance from the boundary.

At round t, the adversary fixes f_t before the learner's randomness, but does not reveal it. The learner commits to a reward-bearing action $p_t \in \mathcal{K}$ and then queries one full gradient at a chosen $z_t \in \mathcal{K}$. The query may differ from the action. This is post-decision first-order oracle feedback, not bandit feedback restricted to the played point. The received reward is $f_t(p_t)$, but neither that value nor any other value evaluation is needed in the update. An oblivious adversary fixes the full sequence in advance, while a non-anticipating adaptive adversary may choose f_t from the preceding history, but not from the current action, query coin, or oracle noise. For this latter model we assume $\mathbb{E} f_t(0) < \infty$, and the gradient bound then makes the regret below integrable. Write

$$\mathrm{Reg}_\alpha(T) = \mathbb{E}\left[\alpha \max_{o \in \mathcal{K}} \sum_{t=1}^{T} f_t(o) - \sum_{t=1}^{T} f_t(p_t)\right]. \tag{3}$$

Thus the regret guarantee uses the best fixed comparator for the realized sequence, inside the expectation. The constants D, B, T are used for tuning. If $D = 0$ or $B = 0$, the learner plays 0, queries at 0, and projects 0 on each round, and Appendix F.1 verifies that this gives nonpositive approximate regret. The step-size formulas below use $D, B > 0$.

We repeatedly use concavity along comparable segments and the resulting lattice inequality (Bian et al., 2017):

$$f(x \odot o) + f(x \oplus o) \le 2f(x) + \langle \nabla f(x), o - x \rangle. \tag{4}$$

To derive Eq. (4), apply the tangent upper bounds at x in the comparable directions $x \odot o - x \le 0$ and $x \oplus o - x \ge 0$. Adding the resulting inequalities, the two direction vectors sum to $o - x$. The union point $x \oplus o$ is an analysis-only value. Every actual query specified below lies in $[0, x]$, even when this union is outside $\mathcal{K}$.

3 A COMPARATOR-UNIFORM 0.4013 FIRST-ORDER COMPARISON

The structural goal is a comparison that holds simultaneously for every comparator $o \in \mathcal{K}$:

$$\alpha_* f(o) - \mathbb{E}[f(p(x)) \mid f, x] \le \langle G(f,x), o - x\rangle.$$

Here the action distribution depends only on the current feasible state x, and a single comparator-independent field $G(f,x)$ controls every comparator. Such a statement can be used online because the state and action distribution are fixed before the current function is observed. We first establish a strengthened trajectory comparison, then an endpoint-retaining box comparison with the same linear direction. Their nonlinear terms cancel to give Theorem 3.4. The main text records the constructions and proof ideas; Appendices B–D contain the complete structural proofs and coefficient certification.

3.1 A STRENGTHENED DELAYED-TRAJECTORY COMPARISON

The first comparison uses a feasible state x both as the guiding point and as the fixed trajectory direction. For a delay $s \in (0,1)$ and time $v \in [0,1]$, define

$$Y_v(x) = (\mathbf{1} - x) \odot (\mathbf{1} - e^{-vx}) + x \odot (\mathbf{1} - e^{-(v-s)_+ x}), \tag{5}$$

$$\omega_v(x) = (\mathbf{1} - x) \odot e^{-vx} + \mathbf{1}_{\{v \ge s\}} x \odot e^{-(v-s)x}, \tag{6}$$

$$q_s(f,x) = \int_0^1 e^{v-1} \omega_v(x) \odot \nabla f(Y_v(x))\, dv. \tag{7}$$

This is a constant-direction specialization of the delayed construction of Buchbinder & Feldman (2024), not its objective-dependent direction-selection procedure. The inequality $1 - e^{-a} \le a$ gives $0 \le Y_v(x) \le x$; also $0 \le \omega_v(x) \le \mathbf{1}$ and $\dot{Y}_v = \omega_v \odot x$ almost everywhere. Thus the action $Y_1(x)$ and every location in the integral are feasible.

Define the scalar coefficients

$$\theta_s = \frac{(2-s)e^s - 1}{e}, \qquad \mu_s = \frac{e^s - 1}{e}, \qquad \zeta_s = \frac{(2-s)e^s - 2}{e} - \frac{(1-s)^2}{2e}\left(1 + \frac{3s}{4}\right). \tag{8}$$

Proposition 3.1. *For every $x, o \in \mathcal{K}$,*

$$f(Y_1(x)) \ge \theta_s f(o) - \mu_s f(x \odot o) - \zeta_s f(x \oplus o) - \langle q_s(f,x), o - x\rangle. \tag{9}$$

Consequently, for $\ell \ge \max\{0, \zeta_s\}$,

$$\begin{aligned} \theta_s f(o) - f(Y_1(x)) \le (\mu_s - \ell) f(x \odot o) + 2\ell f(x) \\ + \langle q_s(f,x) + \ell \nabla f(x), o - x\rangle. \end{aligned} \tag{10}$$

Proof sketch. The exponential expansion separates the trajectory into early and late phases. The strengthening comes from order two with one integration time in each phase. With $U = x \oplus o$, the associated point is

$$H = x \odot (o \oplus x) + (\mathbf{1} - x) \odot (o \oplus (x \oplus x)) = U + (\mathbf{1} - o) \odot x \odot (\mathbf{1} - x)^2.$$

Since $x_i(1 - x_i) \le 1/4$,

$$0 \le H - U \le \tfrac{1}{4}(\mathbf{1} - U).$$

Thus $W = U + 4(H - U)$ is in the cube, and directional concavity and nonnegativity give

$$f(H) \ge \tfrac{3}{4} f(U) + \tfrac{1}{4} f(W) \ge \tfrac{3}{4} f(U). \tag{11}$$

Retaining this mixed contribution, integrating the two phase bounds, and applying the lattice inequality with multiplier ℓ gives the two claims. Appendix B develops the phase comparisons, and Appendix B.5 gives the complete integration and free-ℓ step. □

The mixed contribution lowers ζ_s by $3s(1-s)^2/(8e)$ relative to dropping it. This slack will permit the choice $\ell = \mu_s/3$, which equalizes the two remaining nonlinear coefficients. The comparison holds for every comparator, rather than just an optimizer of the current function. That uniformity is necessary online: the hindsight comparator optimizes the sum of rewards. The proof uses only first derivatives and directional concavity; the full exponential inequality is an inherited result, while the mixed-term estimate and its use with a free lattice multiplier are the strengthening above.

3.2 A FEASIBLE BOX COMPARISON WITH THE SAME STATE

After choosing the lattice multiplier, the trajectory comparison will leave the sum $f(x \odot o) + f(x)$. We control this sum by a feasible action and a linear term in $o - x$. The endpoint value is essential: a box bound controlling only $f(x \odot o)$ would leave $f(x)$ uncancelled. The following lemma is an endpoint-retaining adaptation of coordinatewise hybrid telescoping (Niazadeh et al., 2020; Roughgarden & Wang, 2018), with a prescribed state rather than an objective-dependent optimizer.

Consider any nonnegative continuously differentiable DR-submodular F on the cube. For $a \in [0,1]^d$, set $r_i = a_i/(1+a_i)$ and $w_i = a_i(1-a_i)$. Order coordinates by nondecreasing w_i, resolving ties by index. For exposition relabel them in that order. Let $L^0 = 0$, $U^0 = \mathbf{1}$, and obtain L^i, U^i by setting coordinate i of L^{i-1}, U^{i-1} to r_i. Define

$$A_i = \int_0^1 \partial_i F(L^{i-1} + v r_i e_i)\, dv, \qquad C_i = \int_0^1 \partial_i F(U^{i-1} - v(1-r_i)e_i)\, dv, \tag{12}$$
$$g_i(F,a) = (1-r_i)A_i + r_i C_i.$$

The integrals are averages, so the definition also applies when $r_i = 0$.

Lemma 3.2. *For every $a, b \in [0,1]^d$,*

$$F(b) + F(0) - \tfrac{9}{4}F(r) \le \langle g(F,a), b - a\rangle. \tag{13}$$

Proof sketch. The lower chain raises one coordinate at a time from 0 to r, while the upper chain lowers it from $\mathbf{1}$ to r; A_i and C_i are their average partial derivatives. Replace the coordinates of b successively by those of r, forming $b^{(i)} = (r_1, \dots, r_i, b_{i+1}, \dots, b_d)$. DR-submodularity gives $A_i \ge C_i$. Comparing the ith hybrid segment with the two chain segments, and distinguishing $b_i \ge r_i$ from $b_i < r_i$, gives the common bound

$$F(b^{(i-1)}) - F(b^{(i)}) \le b_i(1-r_i)A_i - (1-b_i)r_i C_i.$$

The coefficient of b_i on the right is precisely $g_i(F,a)$. Centering at a_i and summing therefore yields

$$F(b) \le F(r) + \langle g(F,a), b-a\rangle + \sum_i \big[a_i(1-r_i)A_i - (1-a_i)r_i C_i\big].$$

The rational choice of r gives $a_i(1-r_i) = r_i$ and $(1-a_i)r_i = w_i(1-r_i)$. Since $r_i A_i = F(L^i) - F(L^{i-1})$ and $-(1-r_i)C_i = F(U^i) - F(U^{i-1})$, the residual sum becomes

$$\sum_i \big[a_i(1-r_i)A_i - (1-a_i)r_i C_i\big] = F(r) - F(0) + \sum_i w_i\big[F(U^i) - F(U^{i-1})\big].$$

The lower chain thus retains the endpoint $F(0)$ needed for cancellation. For the upper chain, summation by parts gives

$$\begin{aligned}\sum_i w_i\big[F(U^i) - F(U^{i-1})\big] &= w_d F(r) - w_1 F(\mathbf{1}) - \sum_{i<d}(w_{i+1} - w_i)F(U^i) \\ &\le \tfrac{1}{4}F(r).\end{aligned}$$

Here the weights are nondecreasing, $0 \le w_i \le 1/4$, and F is nonnegative. Individual upper-chain increments may have either sign, so their ordered telescoping is essential. Combining the bounds gives two copies of $F(r)$ and the additional $F(r)/4$, proving the stated $9/4$ coefficient while preserving $F(0)$. Appendix C.1 gives the complete proof, including the coordinate cases and endpoint conventions. □

We now specialize the inequality to a feasible box $[0, x]$. The complement substitution below turns the retained endpoint into $f(x)$ and expresses the linear residual in the same direction $o - x$ as the trajectory comparison.

Proposition 3.3. *For $x, o \in \mathcal{K}$, define*

$$J_x(a) = f(x \odot (\mathbf{1} - a)), \qquad P(x) = \frac{x}{2 - x}, \qquad k(f,x) = -g(J_x, \mathbf{1} - x). \tag{14}$$

Then

$$f(x \odot o) + f(x) - \tfrac{9}{4} f(P(x)) \le \langle k(f,x), o - x\rangle. \tag{15}$$

Both $P(x)$ and all original gradient locations defining $k(f,x)$ belong to $[0,x] \subseteq \mathcal{K}$.

Proof sketch. Apply Lemma 3.2 to J_x with the common-state substitution $a = \mathbf{1} - x$ and $b = \mathbf{1} - o$. Then $r = (\mathbf{1} - x)/(2 - x)$, $J_x(b) = f(x \odot o)$, $J_x(0) = f(x)$, and $J_x(r) = f(P(x))$. Most importantly,

$$b - a = x - o = -(o - x),$$

which aligns the linear term with the trajectory comparison after setting $k = -g$. Every auxiliary cube point maps into $[0, x]$, so the action and all original gradient locations are feasible. Appendix C.2 proves the substitution and its endpoint conventions in full. □

Although $a = \mathbf{1} - x$ need not belong to $\mathcal{K}$, its entire auxiliary cube maps into $[0, x]$. Feasibility concerns the original query, not the auxiliary coordinate. This is a substitution at a fixed state, not differentiation through J_x as x changes. It therefore introduces no derivatives of the coordinate order and no second derivatives of f. The box state is prescribed by x; its comparator is transformed by the same complement. This is what allows the box and trajectory residuals to be controlled by one online learner.

3.3 CANCELLATION AND THE 0.4013 FIRST-ORDER INEQUALITY

The two comparisons now have the same linear direction $o - x$. To cancel their nonlinear residuals, first match the coefficients of $f(x \odot o)$ and $f(x)$ in Eq. (10):

$$\mu_s - \ell = 2\ell \quad \Longrightarrow \quad \ell = \frac{\mu_s}{3}, \qquad b_s = \frac{2\mu_s}{3}.$$

Multiplying the box comparison by this common coefficient then leaves the box reward coefficient $9b_s/4$. Normalizing the two reward weights requires

$$\frac{9}{4}\lambda b_s = 1 - \lambda \quad \Longrightarrow \quad \lambda = \frac{1}{1 + 3\mu_s/2}.$$

These equations determine the coefficients, rather than an online tuning procedure. The one-dimensional delay choice is discussed in Appendix D.3; we fix

$$s = \frac{199}{500}, \qquad \ell = \frac{\mu_s}{3}, \qquad b_s = \frac{2\mu_s}{3}, \qquad \lambda = \frac{1}{1 + 3\mu_s/2}, \qquad \alpha_* = \lambda\theta_s. \tag{16}$$

Direct analytic bounds on the exponential (Appendix D.4) give

$$\ell - \zeta_s > 0.004823, \qquad \alpha_* > 0.40130636 > 0.4013. \tag{17}$$

In particular, Proposition 3.1 applies and $\mu_s - \ell = 2\ell = b_s$.

Theorem 3.4 (Feasible first-order comparison)**.** *Choose $p(x) = Y_1(x)$ with probability λ and $p(x) = P(x)$ otherwise. Define*

$$G(f, x) = \lambda\big[q_s(f, x) + \ell\nabla f(x) + b_s k(f, x)\big]. \tag{18}$$

For every $x, o \in \mathcal{K}$,

$$\alpha_* f(o) - \mathbb{E}[f(p(x)) \mid f, x] \le \langle G(f, x), o - x\rangle. \tag{19}$$

The action distribution depends only on x, and all gradients defining the field are feasible.

Proof sketch. Write $R_f(x, o) = f(x \odot o) + f(x)$. The two propositions 3.1 and 3.3 give

$$\begin{aligned} \theta_s f(o) - f(Y_1(x)) &\le b_s R_f(x, o) + \langle q_s(f, x) + \ell\nabla f(x), o - x\rangle, \\ R_f(x, o) - \tfrac{9}{4} f(P(x)) &\le \langle k(f, x), o - x\rangle. \end{aligned}$$

We multiply these inequalities by λ and λb_s, respectively, before adding together. The residual $R_f(x, o)$ cancels, while $(9/4)\lambda b_s = 1 - \lambda$ normalizes the two reward weights. The play coin realizes their weighted reward in expectation. This uses a mixture of feasible actions, rather than a concavity assertion about their average. Appendix D.2 gives the complete proof. □

Relation to upper-linearizability. Eq. (19) is the expectation-valued version of the comparison in Pedramfar & Aggarwal (2024); Lu et al. (2026): their deterministic reported action is replaced by a state-dependent distribution. Conditional expectation supplies the corresponding regret transfer, proved in Section 5. The field $G(f, x)$ is independent of the comparator and need not be the gradient of a scalar potential. The new work is establishing this field and its coefficient; the linear-learning reduction remains standard. Randomization is used for the reward mixture because DR-submodularity alone does not justify replacing a mixture of actions by their average.

4 THE ONE-QUERY ONLINE ALGORITHM

Section 3.3 establishes the structural inequality assuming access to the field $G(f, x)$. We now implement that field as an unbiased estimator using one feasible gradient query. The action mixture remains unchanged; only one of the field's three components is sampled on a round. We first state the estimator guarantee and the resulting online algorithm, then describe the sampler that realizes this guarantee.

Let $c_0 = 1 - e^{-1}$ and define

$$S_d = c_0 + \ell + b_s\sqrt{2d}, \qquad (\pi_0, \pi_1, \pi_2) = \frac{(c_0, \ell, b_s\sqrt{2d})}{S_d}, \qquad \Gamma_d = \lambda S_d. \tag{20}$$

The one-query field sampler defined below satisfies

$$\mathbb{E}[\widehat{G} \mid f, x] = G(f, x), \qquad \mathbb{E}[\|\widehat{G}\|_2^2 \mid f, x] \le B^2\Gamma_d^2. \tag{21}$$

Algorithm 1 combines the action mixture, this sampler, and a projected ascent update. The action coin and the query-branch coin are independent: the action is committed before the current-function query, whose answer updates the state for the next round. Appendix E proves the estimator guarantee, and Section 5 derives the regret guarantee from it.

Algorithm 1 Online maximization with one feasible gradient query

Require: Projection oracle $\Pi_{\mathcal{K}}$, horizon T, bounds $D, B > 0$.

1: Set the constants in Eq. (16), Eq. (20); $\eta = D/(B\Gamma_d\sqrt{T})$; $x_1 = 0$.
2: **for** $t = 1, \dots, T$ **do**
3: Without observing f_t, commit to $p_t = Y_1(x_t)$ with probability λ, or $p_t = P(x_t)$ otherwise.
4: Independently of the action coin, sample one branch J_t and its query point using Eq. (23).
5: Query the current-function oracle once, after committing p_t, and form $\widehat{G}_t$.
6: $x_{t+1} \leftarrow \Pi_{\mathcal{K}}(x_t + \eta\widehat{G}_t)$.
7: **end for**

4.1 SAMPLING THE TRAJECTORY INTEGRAL

A path time with density e^{v-1}/c_0 can be sampled as $V = \log(1+(e-1)U)$ for $U \sim \mathrm{Unif}[0, 1]$. The returned trajectory estimate is $c_0\omega_V(x) \odot \nabla f(Y_V(x))$. Sampling according to the weight in the definition of q_s makes this estimate unbiased. The query $Y_V(x)$ lies in $[0, x]$, and $\|\omega_V\|_\infty \le 1$ bounds this estimator's second moment by $c_0^2B^2$. Thus the integral specifies a one-sample expectation, not numerical integration or multiple oracle calls. The calculation is given in Appendix E.1.

4.2 SAMPLING THE BOX FIELD

Use the chains in Eq. (12) for $a = \mathbf{1} - x$, working in the sorted order and undoing the permutation on the returned vector. Draw I uniformly from $[d]$ and V uniformly from $[0, 1]$. With probability $1 - r_I$ set $Z = L^{I-1} + Vr_Ie_I$; otherwise set $Z = U^{I-1} - V(1 - r_I)e_I$. One original gradient gives

$$\widehat{k} = d\, x_I\, [\nabla f(x \odot (\mathbf{1} - Z))]_I e_I, \qquad \mathbb{E}\widehat{k} = k(f, x). \tag{22}$$

The positive sign comes from the chain rule for J_x and the definition $k = -g$. Although the chains specify $2d$ segments, only one random segment is queried. Their endpoints depend on x and the coordinate order, not on current gradient answers. Every such query lies in $[0, x]$. The useful

moment bound is $\mathbb{E}\|\widehat{k}\|_2^2 \le 2dB^2$, rather than the coarser d^2B^2: DR-submodularity bounds each sampled partial derivative by its values at the two endpoints of $[0, x]$. The endpoint gradients in this proof are not evaluated by the algorithm. The permutation-explicit construction of just the sampled segment, the sign calculation, and the moment bound are in Appendix E.2.

4.3 Sampling the complete field

After committing the action, independently select one branch J with these probabilities and evaluate only its gradient:

$$\widehat{G} = \begin{cases} \lambda c_0\, \omega_V(x) \odot \nabla f(Y_V(x))/\pi_0, & J = 0, \quad V \sim e^{v-1}/c_0, \\ \lambda \ell \nabla f(x)/\pi_1, & J = 1, \\ \lambda b_s \widehat{k}/\pi_2, & J = 2. \end{cases} \tag{23}$$

Only the selected branch is evaluated. Thus the single query is at $Y_V(x)$, x, or $x \odot (\mathbf{1} - Z)$, all in $[0, x]$; the integral notation specifies an expectation, not extra calls or numerical quadrature. The three branch expectations recover the three components of $G(f, x)$. The probabilities in Eq. (20) minimize the resulting second-moment bound among fixed positive branch probabilities by Cauchy–Schwarz; Appendix E.3 gives the full calculation. Importance sampling and projected ascent are standard; their role here is to implement the new comparison without increasing the query budget.

The coordinate sort and query construction take $O(d \log d)$ arithmetic operations and $O(d)$ memory, in addition to the one gradient and one projection in the online update. All state-dependent choices precede current oracle feedback.

5 Online regret from the first-order comparison

Theorem 3.4 gives the pointwise comparison

$$\alpha_* f_t(o) - \mathbb{E}[f_t(p_t) \mid \mathcal{H}_t] \le \langle G(f_t, x_t), o - x_t \rangle,$$

where $\mathcal{H}_t$ contains the history and f_t, x_t before the fresh draws. Section 4 supplies an unbiased estimator $\widehat{G}_t$ of this field with bounded second moment. Projected online ascent then controls the sum of these linear residuals. Algorithm 1 combines the action mixture, the one-query gradient estimator, and the update. We now state the guarantees obtained from these two inputs.

Theorem 5.1. *For every integer horizon $T \ge 1$, under Section 2, Algorithm 1 has*

$$\mathrm{Reg}_{\alpha_*}(T) \le DB\Gamma_d\sqrt{T} \tag{24}$$

for oblivious adversaries. For non-anticipating adaptive adversaries and the realized-sequence benchmark Eq. (3),

$$\mathrm{Reg}_{\alpha_*}(T) \le 2DB\Gamma_d\sqrt{T}. \tag{25}$$

The same bounds hold at coefficient 0.4013. Every round uses one feasible gradient and one projection. In particular, $\Gamma_d = O(\sqrt{d})$ with universal constants.

Proof sketch (oblivious adversaries). For an oblivious sequence, compactness and continuity give a fixed hindsight maximizer o^*. The pathwise projected-ascent inequality Eq. (44) gives

$$\sum_{t=1}^{T} \langle \widehat{G}_t, o^* - x_t \rangle \le \frac{D^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \|\widehat{G}_t\|_2^2.$$

Conditional unbiasedness lets us replace the field by its estimate against this fixed comparator, and the reward-mixture identity converts the structural comparison into approximate regret. The second-moment bound (Eq. (21)) and the step size $\eta = D/(B\Gamma_d\sqrt{T})$ give Eq. (24). Appendix F.1 gives the complete calculation. □

Proof sketch (non-anticipating adaptive adversaries). The hindsight maximizer o^* is now random and may depend on the realized estimator sequence. It is therefore invalid to discard the estimator error by conditioning against o^*. Instead, apply the comparator-uniform comparison and the

projection inequality pathwise. The cumulative estimator error is a martingale sum, whose second moment bounds the additional expected residual by $DB\Gamma_d\sqrt{T}$. Adding this correction to the projection term gives Eq. (25). The reward-mixture identity is used after taking expectation of the cumulative reward, not by conditioning against the random comparator. Appendix F.2 contains the complete proof, including the filtration, measurability, integrability, and orthogonality arguments, using the projection calculation in Appendix F.1. □

The adaptive-adversary statement is a static-comparator guarantee for a history-dependent reward sequence. It is distinct from adaptive regret, which measures performance on time intervals.

6 Extensions to Stochastic gradients and offline specialization

The structural comparison, its one-query implementation, and the online regret guarantees are now established. We now extend the guarantees to stochastic gradients as well as give the online-to-offline specialization.

Stochastic gradients. Appendix G uses a conditionally unbiased stochastic-gradient oracle whose conditional second moment is at most B_1^2. With the branch probabilities in Appendix G.1, the one-query field estimator remains conditionally unbiased and has second moment at most $B_1^2\widetilde{\Gamma}_d^2$, where $\widetilde{\Gamma}_d = \lambda(c_0+\ell+b_s d)$. Lemmas G.1 and G.2 give expected approximate-regret bounds $DB_1\widetilde{\Gamma}_d\sqrt{T}$ for oblivious adversaries and $2DB_1\widetilde{\Gamma}_d\sqrt{T}$ for non-anticipating adaptive adversaries, respectively. The stochastic noise may be temporally dependent, provided the conditional unbiasedness and second-moment conditions hold before every oracle call.

Offline specialization. A constant objective sequence turns the online guarantee into an offline solver without changing the algorithm. A uniformly sampled played action realizes the average reward, so selecting the output requires no value queries. We note that Corollary 6.1 provides slightly better offline approximation ratio than Buchbinder & Feldman (2024).

Corollary 6.1 (Offline specialization and oracle budget)**.** *For one fixed objective f, let* $\mathrm{OPT} = \max_{o\in\mathcal{K}} f(o)$. *For every integer $N \geq 1$, run Algorithm 1 with horizon N and $f_t = f$, and return p_τ for an independent uniform index $\tau \in [N]$. Then*

$$\mathbb{E}f(p_\tau) \geq \alpha_* \,\mathrm{OPT} - \frac{DB\Gamma_d}{\sqrt{N}}. \tag{26}$$

Consequently, the guarantee $\mathbb{E}f(p_\tau) \geq \alpha_* \,\mathrm{OPT} - \varepsilon$ is attained for every $\varepsilon > 0$ using

$$N_\varepsilon = \max\left\{1, \left\lceil \frac{D^2B^2\Gamma_d^2}{\varepsilon^2} \right\rceil\right\} \tag{27}$$

feasible gradient queries and the same number of projections. The statements also hold with Γ_R in place of Γ_d.

Proof. Apply Eq. (24) to the constant sequence and divide by N; uniform sampling realizes the average reward. □

7 Conclusion

We obtain coefficient 0.4013 for adversarial online non-monotone DR-submodular maximization using one feasible gradient and one projection per round. The key step is a comparator-uniform cancellation: a retained mixed trajectory term and an endpoint-retaining box comparison align at the same feasible state. This converts an offline-style approximation argument into an online first-order inequality without optimizing the current objective before acting. The guarantee requires only continuous differentiability and covers empty-interior down-closed sets. Explicit regret bounds, an offline specialization with a finite oracle budget, geometric refinements, stochastic feedback, and the realized-sequence adaptive benchmark accompany the construction.

AI USE STATEMENT

In this work, we used generative AI tools for [edit a research paper to improve readability, suggest a structure for a research paper]. We have not used generative AI tools for [help develop theoretical models or conceptual frameworks, formulate mathematical claims, assist in the writing of proofs, provide critical ingredients for proving mathematical claims], and [generate synthetic data sets, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis, interpret results] are not applicable to this work. We have reviewed all AI-assisted work. We have reviewed any statement that we use AI to improve readability to avoid any hallucination in claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

For our theoretical results, we have provided full proof of all our claims in Appendices. We have provided the references to the corresponding Appendix in the vicinity of the claim.

REFERENCES

An Bian, Kfir Y. Levy, Andreas Krause, and Joachim M. Buhmann. Continuous dr-submodular maximization: Structure and algorithms. In *Proceedings of Advances in Neural Information Processing Systems*, 2017.

Yatao A. Bian, Joachim M. Buhmann, and Andreas Krause. Optimal continuous dr-submodular maximization and applications to provable mean field inference. In *Proceedings of the International Conference on Machine Learning*, 2019.

Niv Buchbinder and Moran Feldman. Constrained submodular maximization via a nonsymmetric technique. *Mathematics of Operations Research*, 44(3):988–1005, 2019.

Niv Buchbinder and Moran Feldman. Constrained submodular maximization via new bounds for dr-submodular functions. In *Proceedings of the Annual ACM Symposium on Theory Of Computing*, 2024.

Lin Chen, Hamed Hassani, and Amin Karbasi. Online continuous submodular maximization. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2018.

Alina Ene and Huy L. Nguyen. Constrained submodular maximization: Beyond $1/e$. In *Proceedings of IEEE Annual Symposium on Foundations of Computer Scienc*, 2016.

Moran Feldman, Joseph Naor, and Roy Schwartz. A unified continuous greedy algorithm for submodular maximization. In *Proceedings of IEEE annual symposium on foundations of computer science*, 2011.

Yiyang Lu, Haresh Jadav, Mohammad Pedramfar, Ranveer Singh, and Vaneet Aggarwal. Upper-linearizability of online non-monotone dr-submodular maximization over down-closed convex sets. In *Proceedings of the International Conference on Machine Learning*, 2026.

Rad Niazadeh, Tim Roughgarden, and Joshua R. Wang. Optimal algorithms for continuous non-monotone submodular and DR-submodular maximization. *Journal of Machine Learning Research*, 21(125):1–31, 2020.

Mohammad Pedramfar and Vaneet Aggarwal. From linear to linearizable optimization: A novel framework with applications to stationary and non-stationary dr-submodular optimization. In *Proceedings of Advances in Neural Information Processing Systems*, 2024.

Mohammad Pedramfar, Yididiya Y. Nadew, Chris J. Quinn, and Vaneet Aggarwal. Unified projection-free algorithms for adversarial dr-submodular optimization. In *Proceedings of the International Conference on Learning Representations*, 2024.

Tim Roughgarden and Joshua R. Wang. An optimal learning algorithm for online unconstrained submodular maximization. In *Proceedings of the Annual Conference On Learning Theory*, 2018.

Nguyen Kim Thang and Abhinav Srivastav. Online non-monotone dr-submodular maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

Qixin Zhang, Zengde Deng, Zaiyi Chen, Kuangqi Zhou, Haoyuan Hu, and Yu Yang. Online learning for non-monotone dr-submodular maximization: From full information to bandit feedback. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2023.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of International Conference on Machine Learning*, 2003.

A Additional related work

Offline approximation and constraint oracles. Measured continuous greedy and its successors established approximation guarantees for constrained non-monotone submodular and DR-submodular maximization (Feldman et al., 2011; Bian et al., 2017; Ene & Nguyen, 2016; Buchbinder & Feldman, 2019). Buchbinder & Feldman (2024, Theorem 1.1) prove $0.401\ \mathrm{OPT} - O(\delta D^2 L)$ for every $\delta \in (0,1)$, in $\mathrm{poly}(d, \delta^{-1})$ time, for nonnegative L-smooth DR-submodular objectives over meta-solvable down-closed convex bodies. Meta-solvability provides efficient linear optimization after intersection with a polynomial number of explicit halfspaces. Their exponential expansion (Lemma 4.1) and nonnegative affine embedding (Lemma 4.2) are the prior structural inputs to our trajectory proof. Their offline construction also selects guiding points and auxiliary solutions using the current objective. Here the guiding point and direction are the same state, fixed by past online feedback; both candidate actions are explicit functions of that state.

Our offline corollary is an expected continuous-objective guarantee with feasible gradient and projection oracles. It gives a larger coefficient up to an arbitrary additive error without a gradient-Lipschitz assumption. The budget is $O(1 + dD^2B^2/\varepsilon^2)$ for every absolute error $\varepsilon > 0$, as specified in Eq. (27). The offline specialization and its numerical improvement over the coefficient 0.401 are secondary consequences of our online result. We therefore turn to the online literature, where the main advance is improving the approximation coefficient from $1/e$ to 0.4013 while retaining one feasible gradient query per round and $O(\sqrt{T})$ expected approximate regret.

Online approximation and query feasibility. Thang & Srivastav (2021); Zhang et al. (2023); Pedramfar et al. (2024) study non-monotone online DR-submodular maximization over down-closed sets. Lu et al. (2026) establish a $1/e$ first-order comparison and a single-gradient estimator, together with online regret consequences. Their query $1 - e^{-vx}$, $v \in [0,1]$, obeys $0 \le 1 - e^{-vx} \le x$, so it is already feasible on a down-closed set. Our advance is increasing the coefficient above $1/e$ while retaining feasible first-order access and one objective query per round.

To obtain the common projection-based comparison in Table 1, multiply the estimator of Lu et al. (2026, Theorem 1 and Lemma 2) by $c_0 = 1 - e^{-1}$. Its norm is at most $c_0 B$ under the exact-gradient bound used here. The standard projected-ascent proof therefore gives an oblivious-adversary bound $Dc_0B\sqrt{T}$. This explains the $O(DB\sqrt{T})$ entry without hiding our estimator's additional dimension dependence. Their framework also admits other linear learners, including the separation-oracle implementation presented in that paper. The first-order comparison, rather than that particular constraint-oracle implementation, is what is specialized in the table.

Coordinatewise comparisons and the reduction. The hybrid telescoping template is related to continuous double greedy (Niazadeh et al., 2020) and online unconstrained submodular maximization (Roughgarden & Wang, 2018). We prove the prescribed rational-point comparison with an ordered weighted sum, retain the endpoint, and use the complement substitution to express its residual at the same online state. Pedramfar & Aggarwal (2024) supply the general linear-to-linearizable viewpoint, and projected online ascent follows Zinkevich (2003). Our action distribution adds a conditional-expectation step to that reduction, not a new online optimization principle. The separate martingale argument handles the realized-sequence hindsight comparator for a non-anticipating adaptive adversary.

B COMPLETE PROOF OF THE STRENGTHENED TRAJECTORY COMPARISON

This appendix gives a complete proof of both inequalities in Proposition 3.1. The exponential expansion below is the attributed prior result; the embedding, comparator bounds, mixed-term estimate, and integration are verified here without importing steps from the main-text sketch.

Throughout, every function is nonnegative and continuously differentiable on its cube, with antitone gradient. In particular, f is defined on $[0,1]^d$, and $x, o \in \mathcal{K} \subseteq [0,1]^d$. Products and inequalities are coordinatewise, and $x \oplus o = x + o - x \odot o$. For $s \in (0,1)$, recall the local notation

$$\begin{aligned} Y_v(x) &= (\mathbf{1} - x) \odot (\mathbf{1} - e^{-vx}) + x \odot (\mathbf{1} - e^{-(v-s)_+ x}), \\ \omega_v(x) &= (\mathbf{1} - x) \odot e^{-vx} + \mathbf{1}_{\{v \geq s\}} x \odot e^{-(v-s)x}, \\ q_s(f,x) &= \int_0^1 e^{v-1} \omega_v(x) \odot \nabla f(Y_v(x))\, dv, \qquad C_v = Y_v + \omega_v \odot o. \end{aligned}$$

The scalar coefficients will be displayed again in the final integration.

Overview of the trajectory proof. Put $C_v = Y_v + \omega_v \odot o$. Since $\omega_v = \mathbf{1} - x - Y_v$ before s and $\omega_v = \mathbf{1} - Y_v$ after s, we have $Y_v \leq C_v \leq \mathbf{1}$. The forward tangent inequality gives $f(C_v) - f(Y_v) \leq \langle \omega_v \odot \nabla f(Y_v), o \rangle$. Using $\dot{Y}_v = \omega_v \odot x$ yields

$$\frac{d}{dv} f(Y_v) + f(Y_v) \geq f(C_v) - \langle \omega_v \odot \nabla f(Y_v), o - x \rangle. \tag{28}$$

The two-block exponential expansion gives the comparator-uniform phase bounds

$$f(C_v) \geq f(o) - f(x \odot o) - (1 - e^{-v}) f(x \oplus o), \qquad v < s, \tag{29}$$

$$f(C_v) \geq e^{-v} \left[e^s f(o) - (e^s - 1) f(x \oplus o) + \left(1 + \frac{3s}{4}\right)(v - s) f(x \oplus o) \right], \qquad v \geq s. \tag{30}$$

For the early phase, the block weights are x and $\mathbf{1} - x$, the initial points are 0 and o, and the directions are 0 and x. Diminishing increments bound the base value by $f(o) - f(x \odot o)$ and each positive-order integrand by $f(o) - f(x \odot o) - f(x \oplus o)$. Lemma B.4 gives the complete derivation, using the two-block specialization in Lemma B.2.

For the late phase, both blocks start at o; the first direction switches from 0 to x at s, and the second is always x. The strengthened term comes from order two with one integration time before s and one after it. With $U = x \oplus o$, the associated point H satisfies the geometric bound in Eq. (11), so its value is at least $3f(U)/4$. The mixed region has volume $2s(v - s)$ and factor $1/2!$, contributing $(3s/4)(v - s)f(U)$. The all-early regions and the first-order late region supply the other terms in Eq. (30). Lemma B.5 proves this bound in full, including the mixed-region calculation and the nonnegativity of every discarded integrand.

Multiplying Eq. (28) by e^{v-1} and integrating the two phase bounds produces θ_s, μ_s, and ζ_s in Eq. (8); the initial contribution is $e^{-1} f(0) \geq 0$. Finally, write $-\zeta_s f(U) = -\ell f(U) + (\ell - \zeta_s) f(U)$, use the lattice inequality for the first term, and discard the nonnegative residual union term. This gives Eq. (10). Appendix B.5 below contains the complete proof of Proposition 3.1, including the differential inequality, final integration, coefficient calculation, and free-ℓ step.

B.1 DR-SUBMODULAR DIRECTIONAL INEQUALITIES AND EXPONENTIAL INEQUALITY

For $u \leq v$, $\Delta \geq 0$, and $v + \Delta \leq \mathbf{1}$, integration of the antitone-gradient inequality gives

$$f(u + \Delta) - f(u) \geq f(v + \Delta) - f(v). \tag{31}$$

Concavity along a comparable segment follows because its directional derivative is nonincreasing. In particular,

$$\langle \nabla f(u), \Delta \rangle \geq f(u + \Delta) - f(u), \qquad \langle \nabla f(u), \Delta \rangle \leq f(u) - f(u - \Delta), \tag{32}$$

whenever the displayed points lie in the cube. These standard DR consequences are the only tangent inequalities needed (Bian et al., 2017).

Lemma B.1 (Buchbinder & Feldman (2024), Lemma 4.1)**.** *For $a \in [0,1]^n$, $t \geq 0$, and an integrable $\xi : [0,t] \to [0,1]^n$, let $y(t) = \int_0^t \xi(v)\, dv$. Then*

$$F(\mathbf{1} - a \odot e^{-y(t)}) \geq e^{-t}\left[F(\mathbf{1}-a) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{[0,t]^k} F\left((\mathbf{1}-a) \oplus \bigoplus_{j=1}^{k} \xi(v_j)\right) dv\right]. \tag{33}$$

This is the cited result itself; its proof is given in that reference. Continuity on a compact cube bounds the integrands, and the resulting exponential series is absolutely convergent.

B.2 Two-block specialization

If A has nonnegative entries and $c + Az$ remains in the original cube, then $F(z) = f(c + Az)$ is DR-submodular: $z \leq z'$ implies $c + Az \leq c + Az'$, and multiplication by $A^\top \geq 0$ preserves the resulting gradient inequality. It is also nonnegative. This proves the closure property needed for the following embedding.

Lemma B.2. *Let $b, a^{(1)}, a^{(2)} \in [0,1]^d$ and $\xi^{(i)} : [0,t] \to [0,1]^d$ be integrable, with $y^{(i)} = \int_0^t \xi^{(i)}(v)\, dv$. Then*

$$\begin{aligned} &f\big(\mathbf{1} - b \odot a^{(1)} \odot e^{-y^{(1)}} - (\mathbf{1}-b) \odot a^{(2)} \odot e^{-y^{(2)}}\big) \\ &\quad \geq e^{-t}\left[f(\mathbf{1} - b \odot a^{(1)} - (\mathbf{1}-b) \odot a^{(2)}) + \sum_{k \geq 1} \frac{1}{k!} \int_{[0,t]^k} \Xi_k(v)\, dv\right], \end{aligned} \tag{34}$$

where

$$\Xi_k(v) = f\left(b \odot \left[(\mathbf{1} - a^{(1)}) \oplus \bigoplus_{j=1}^{k} \xi^{(1)}(v_j)\right] + (\mathbf{1}-b) \odot \left[(\mathbf{1} - a^{(2)}) \oplus \bigoplus_{j=1}^{k} \xi^{(2)}(v_j)\right]\right). \tag{35}$$

Proof. Define $F(z^{(1)}, z^{(2)}) = f(b \odot z^{(1)} + (\mathbf{1}-b) \odot z^{(2)})$ on $[0,1]^{2d}$. The map is nonnegative linear and sends the larger cube into $[0,1]^d$, so the closure argument makes F nonnegative and DR-submodular. Apply Lemma B.1 with $a = (a^{(1)}, a^{(2)})$ and $\xi = (\xi^{(1)}, \xi^{(2)})$. Expanding the base term and each integrand proves the claim. This is the two-block specialization of the affine-embedding argument in Buchbinder & Feldman (2024, Lemma 4.2). □

B.3 Early-phase comparison

For $v < s$, $Y_{v,i} \leq (1 - x_i) v x_i \leq x_i$. For $v \geq s$,

$$Y_{v,i} \leq (1-x_i) v x_i + x_i (v - s) x_i = v x_i - s x_i^2 \leq x_i.$$

The expressions for ω_v give $0 \leq \omega_v \leq \mathbf{1}$, and differentiating within each phase gives $\dot{Y}_v = \omega_v \odot x$. In addition,

$$\omega_v = \mathbf{1} - Y_v - x \quad (v < s), \qquad \omega_v = \mathbf{1} - Y_v \quad (v \geq s). \tag{36}$$

The value at the single breakpoint has no effect on an integral or on the sampled estimator. It follows that $C_v = Y_v + \omega_v \odot o$ lies in the cube: in the early phase it is at most $\mathbf{1} - x$, and in the late phase it is $Y_v \oplus o$.

Lemma B.3. *For $x, o \in [0,1]^d$, $0 \leq \Delta \leq \mathbf{1} - o$, and $0 \leq o' \leq o$,*

$$f(o' + (\mathbf{1} - x) \odot \Delta) \geq f(o) + f(0) - f(x \oplus o) - f(o - o'). \tag{37}$$

Proof. In Eq. (31), use bases $0 \leq o - o'$ and increment $o' + (\mathbf{1} - x) \odot \Delta$. The increment fits because $o + (\mathbf{1} - x) \odot \Delta \leq \mathbf{1}$. This gives

$$f(o' + (\mathbf{1} - x) \odot \Delta) - f(0) \geq f(o + (\mathbf{1} - x) \odot \Delta) - f(o - o').$$

Next use bases $o \le x \oplus o$ and increment $(\mathbf{1} - x) \odot \Delta$. Its upper endpoint is in the cube since $\mathbf{1} - (x \oplus o) = (\mathbf{1} - x) \odot (\mathbf{1} - o)$. Therefore

$$f(o + (\mathbf{1} - x) \odot \Delta) - f(o) \ge f(x \oplus o + (\mathbf{1} - x) \odot \Delta) - f(x \oplus o) \ge -f(x \oplus o).$$

Adding the two displays proves the result. This arbitrary-comparator specialization follows the delayed-comparison reasoning of Buchbinder & Feldman (2024); all feasibility and sign steps are given here. □

Lemma B.4. *For $0 \le v < s$,*

$$f(C_v) \ge f(o) - f(x \odot o) - (1 - e^{-v}) f(x \oplus o).$$

Proof. Here

$$C_v = (\mathbf{1} - x) - (\mathbf{1} - x) \odot (\mathbf{1} - o) \odot e^{-vx}.$$

Apply Lemma B.2 with $b = x$, $a^{(1)} = \mathbf{1}$, $\xi^{(1)} = 0$, $a^{(2)} = \mathbf{1} - o$, and $\xi^{(2)} = x$ on $[0, v]$. The base term is $f((\mathbf{1} - x) \odot o)$. Diminishing increments give

$$f((\mathbf{1} - x) \odot o) \ge f(o) - f(x \odot o) + f(0) \ge f(o) - f(x \odot o).$$

For expansion order $k \ge 1$ put $W_k = \bigoplus_{j=1}^{k} x$. The integrand is $f((\mathbf{1} - x) \odot (o \oplus W_k))$. Lemma B.3, with $o' = (\mathbf{1} - x) \odot o$ and $\Delta = (\mathbf{1} - o) \odot W_k$, bounds it below by $f(o) - f(x \oplus o) - f(x \odot o)$. It is independent of the sampled times. The series therefore gives

$$\begin{aligned} f(C_v) &\ge e^{-v} \big[f(o) - f(x \odot o) + (e^v - 1)(f(o) - f(x \oplus o) - f(x \odot o)) \big] \\ &= f(o) - f(x \odot o) - (1 - e^{-v}) f(x \oplus o). \end{aligned}$$

A lower bound on an individual retained integrand may be negative; replacing that integrand by its lower bound remains valid. □

B.4 Late-phase comparison and mixed second-order term

Lemma B.5. *For $s \le v \le 1$,*

$$f(C_v) \ge e^{-v} \left[e^s f(o) - (e^s - 1) f(x \oplus o) + \left(1 + \frac{3s}{4}\right)(v - s) f(x \oplus o) \right].$$

Proof. Now $C_v = Y_v \oplus o$. Set $\xi'(u) = 0$ for $u < s$ and $\xi'(u) = x$ for $u \ge s$. Use Lemma B.2 with $b = x$, $a^{(1)} = a^{(2)} = \mathbf{1} - o$, $\xi^{(1)} = \xi'$, and $\xi^{(2)} = x$. Its base term is $f(o)$, since

$$\mathbf{1} - C_v = x \odot (\mathbf{1} - o) \odot e^{-(v-s)x} + (\mathbf{1} - x) \odot (\mathbf{1} - o) \odot e^{-vx}.$$

The order-k integrand is

$$\Xi_k(u) = f\left(x \odot \left[o \oplus \bigoplus_{j=1}^{k} \xi'(u_j) \right] + (\mathbf{1} - x) \odot \left[o \oplus \bigoplus_{j=1}^{k} x \right] \right). \tag{38}$$

For all times in $[0, s)$, Lemma B.3 with $o' = o$ bounds the integrand $f(o + (\mathbf{1} - x) \odot (\mathbf{1} - o) \odot W_k)$ by $f(o) - f(x \oplus o)$. Summing these all-early regions contributes $(e^s - 1)(f(o) - f(x \oplus o))$.

Put $U = x \oplus o$. For $k = 1$ and $u_1 \in [s, v]$, both blocks equal $o \oplus x$, contributing $(v - s) f(x \oplus o)$. For $k = 2$ with exactly one early time, the integrand is

$$\begin{aligned} &f\big(x \odot (o \oplus x) + (\mathbf{1} - x) \odot (o \oplus (x \oplus x))\big) \\ &\quad = f\big(o + (\mathbf{1} - o) \odot [x^2 + (\mathbf{1} - x) \odot (2x - x^2)]\big) \\ &\quad = f\big(o + (\mathbf{1} - o) \odot [x + x \odot (\mathbf{1} - x)^2]\big) = f(H). \end{aligned}$$

Here

$$H = U + (\mathbf{1} - o) \odot x \odot (\mathbf{1} - x)^2, \qquad \mathbf{1} - U = (\mathbf{1} - o) \odot (\mathbf{1} - x).$$

For each coordinate, $x_i(1 - x_i) \le 1/4$ gives

$$0 \le H - U \le \tfrac{1}{4}(\mathbf{1} - U).$$

Consequently $W = U + 4(H - U)$ belongs to the cube, $U \le W$, and $H = \frac{3}{4}U + \frac{1}{4}W$. Directional concavity along this comparable segment and nonnegativity prove locally that

$$f(H) \ge \tfrac{3}{4}f(U) + \tfrac{1}{4}f(W) \ge \tfrac{3}{4}f(U).$$

The region has volume $2s(v - s)$ and factor $1/2!$, adding $(3s/4)(v - s)f(x \oplus o)$. These retained regions are disjoint. Every other integrand is nonnegative. Combining their contributions with the base term and multiplying by e^{-v} proves Eq. (30). □

The factor $3/4$ in the mixed-term bound is attained already by $f(z) = 1 - z$ in one dimension with $o = 0$, $x = 1/2$: then $U = 1/2$, $H = 5/8$. This observation concerns precisely the mixed integrand used here. The equal guiding point and direction x are essential to the displayed formula for H.

B.5 Integration and complete proof of Proposition 3.1

Complete proof of Proposition 3.1. The feasibility identities in Appendix B.3 give $Y_v \le C_v \le \mathbf{1}$ and $\dot{Y}_v = \omega_v \odot x$ almost everywhere. The forward tangent inequality gives

$$f(C_v) - f(Y_v) \le \langle \nabla f(Y_v), \omega_v \odot o \rangle.$$

Using the derivative of $f(Y_v)$ and rearranging yields, within this appendix,

$$\frac{d}{dv} f(Y_v) + f(Y_v) \ge f(C_v) - \langle \omega_v \odot \nabla f(Y_v), o - x \rangle.$$

The path is continuous and piecewise continuously differentiable, so this almost-everywhere inequality can be integrated across the breakpoint. After multiplication by e^{v-1}, its integrated left side is $f(Y_1) - e^{-1}f(0)$. By Lemma B.4, the early-phase value contribution is

$$\frac{e^s - 1}{e}\big[f(o) - f(x \odot o)\big] - \frac{e^s - 1 - s}{e} f(x \oplus o).$$

By Lemma B.5, the late contribution is

$$\frac{(1-s)e^s}{e} f(o) - \frac{(1-s)(e^s - 1)}{e} f(x \oplus o) + \frac{(1-s)^2}{2e}\left(1 + \frac{3s}{4}\right) f(x \oplus o).$$

Adding gives coefficients $\theta_s, -\mu_s, -\zeta_s$, where

$$\theta_s = \frac{(2-s)e^s - 1}{e}, \qquad \mu_s = \frac{e^s - 1}{e}, \qquad \zeta_s = \frac{(2-s)e^s - 2}{e} - \frac{(1-s)^2}{2e}\left(1 + \frac{3s}{4}\right).$$

The field integral is exactly $-\langle q_s(f, x), o - x \rangle$. Thus a slightly stronger intermediate expression is

$$f(Y_1) \ge \theta_s f(o) - \mu_s f(x \odot o) - \zeta_s f(x \oplus o) - \langle q_s(f, x), o - x \rangle + e^{-1} f(0).$$

Dropping the nonnegative last term proves the first inequality of Proposition 3.1.

For the second inequality, the tangent bounds from x to $x \odot o \le x$ and to $x \oplus o \ge x$ have directions summing to $o - x$. Adding them gives

$$f(x \odot o) + f(x \oplus o) \le 2f(x) + \langle \nabla f(x), o - x \rangle.$$

Write $-\zeta_s f(x \oplus o) = -\ell f(x \oplus o) + (\ell - \zeta_s) f(x \oplus o)$. Applying the last inequality with multiplier $\ell \ge 0$ gives

$$\begin{aligned} f(Y_1) \ge \theta_s f(o) - (\mu_s - \ell) f(x \odot o) - 2\ell f(x) \\ + (\ell - \zeta_s) f(x \oplus o) - \langle q_s(f, x) + \ell \nabla f(x), o - x \rangle. \end{aligned} \tag{39}$$

When $\ell \ge \max\{0, \zeta_s\}$, the remaining union term is nonnegative and may be discarded. Rearranging proves the second inequality in full:

$$\theta_s f(o) - f(Y_1(x)) \le (\mu_s - \ell) f(x \odot o) + 2\ell f(x) + \langle q_s(f, x) + \ell \nabla f(x), o - x \rangle.$$

□

Choosing $\ell = \mu_s/3$ equalizes the two remaining nonlinear coefficients, explaining the parameter used in the main theorem.

Every step uses DR inequalities, first-derivative integration, continuity, and nonnegativity. In particular, the structural statement is valid for the continuously differentiable class without a quantitative smoothness bound. The evaluations of comparator-dependent values in this appendix are statements in the analysis. The query algorithm accesses only the feasible path and box locations listed in Section 4.

C COMPLETE PROOF OF THE ENDPOINT-RETAINING BOX COMPARISON

This appendix gives the complete proof of Lemma 3.2, then verifies the common-state substitution in Proposition 3.3. The sampling implementation of the resulting field is developed separately in Appendix E.2.

C.1 PROOF OF LEMMA 3.2: ENDPOINT-RETAINING COORDINATE-CHAIN INEQUALITY

We prove Lemma 3.2, recalling its notation so that the argument is self-contained. Let F be nonnegative and continuously differentiable on $[0,1]^d$, with antitone gradient, and let $a, b \in [0,1]^d$. Set $r_i = a_i/(1+a_i)$ and $w_i = a_i(1-a_i)$. Sort by nondecreasing w_i, resolving ties by index, and temporarily relabel the coordinates in that order. Starting from $L^0 = 0$ and $U^0 = \mathbf{1}$, obtain L^i, U^i by setting coordinate i to r_i. The averages and field are

$$A_i = \int_0^1 \partial_i F(L^{i-1} + v r_i e_i)\, dv, \qquad C_i = \int_0^1 \partial_i F(U^{i-1} - v(1-r_i)e_i)\, dv,$$
$$g_i(F,a) = (1-r_i)A_i + r_i C_i.$$

The assertion of Lemma 3.2 is

$$F(b) + F(0) - \tfrac{9}{4}F(r) \le \langle g(F,a), b-a\rangle.$$

Proof overview. The lower chain raises one coordinate at a time from 0 to r, while the upper chain lowers it from 1 to r. The quantities A_i and C_i are the corresponding average partial derivatives. Replace the coordinates of b successively by those of r, forming $b^{(i)} = (r_1, \dots, r_i, b_{i+1}, \dots, b_d)$. The two coordinate cases and $A_i \ge C_i$ give the hybrid bound

$$F(b) \le F(r) + \langle g(F,a), b-a\rangle + \sum_i \big[a_i(1-r_i)A_i - (1-a_i)r_i C_i\big].$$

The rational choice of r makes $a_i(1-r_i) = r_i$ and $(1-a_i)r_i = w_i(1-r_i)$. Consequently the last sum becomes

$$F(r) - F(0) + \sum_i w_i\big[F(U^i) - F(U^{i-1})\big]. \tag{40}$$

Sorting the w_i makes the summation-by-parts remainder nonpositive: its interior coefficients are $-(w_{i+1} - w_i)$, and the function is nonnegative. The remaining endpoint term is at most $w_d F(r) \le \frac{1}{4}F(r)$. The two copies of $F(r)$ and this endpoint term give the coefficient $9/4$. Individual upper-chain increments may have either sign, so replacing each weight by its maximum before summing would not be valid. The complete proof below includes both coordinate cases, the chord bound, $A_i \ge C_i$, all endpoint cases, and tie conventions.

Complete proof of Lemma 3.2. In this coordinate order, define $b^{(i)} = (r_1, \dots, r_i, b_{i+1}, \dots, b_d)$. Thus $b^{(0)} = b$ and $b^{(d)} = r$. If $b_i \ge r_i$, then

$$F(b^{(i-1)}) - F(b^{(i)}) = \int_{r_i}^{b_i} \partial_i F(r_1, \dots, r_{i-1}, u, b_{i+1}, \dots, b_d)\, du.$$

Every point in this integral dominates every point $L^{i-1} + v r_i e_i$ for $0 \le v \le 1$. Antitonicity bounds its partial derivative above by A_i, so the loss is at most $(b_i - r_i)A_i$.

If $b_i < r_i$, the same loss is the negative of an integral from b_i to r_i. Every integration point is dominated by every $U^{i-1} - v(1-r_i)e_i$. Its partial derivative is therefore at least C_i. Taking the negative integral bounds the loss by $(b_i - r_i)C_i$. Every point on the lower ith segment is coordinatewise no larger than every point on the upper ith segment: earlier coordinates equal r, the current coordinate is respectively at most or at least r_i, and later coordinates are 0 or 1. Integrating the antitone-gradient inequality therefore gives $A_i \ge C_i$. The difference between the chord expression $b_i(1-r_i)A_i - (1-b_i)r_i C_i$ and the first bound is $r_i(1-b_i)(A_i - C_i) \ge 0$; the difference from the second is $b_i(1-r_i)(A_i - C_i) \ge 0$. This proves the claimed coordinatewise loss bound for every $b_i \in [0,1]$.

Summing gives

$$F(b) \le F(r) + \langle g(F,a), b-a\rangle + \sum_i \big[a_i(1-r_i)A_i - (1-a_i)r_iC_i\big].$$

The choice of r gives $a_i(1-r_i) = r_i$ and $(1-a_i)r_i = w_i(1-r_i)$. The last sum is consequently

$$F(r) - F(0) + \sum_i w_i\big[F(U^i) - F(U^{i-1})\big].$$

Ordering the weights makes their summation-by-parts remainder have a useful sign:

$$\sum_i w_i\big[F(U^i) - F(U^{i-1})\big] = w_dF(r) - w_1F(\mathbf{1}) - \sum_{i<d}(w_{i+1}-w_i)F(U^i) \le \tfrac{1}{4}F(r).$$

This uses nonnegativity and $0 \le w_i \le 1/4$, and proves the stated inequality. The increments along the upper chain may have either sign, so replacing each weight by its maximum before summing would be invalid. Ordering the weights is what makes the telescoped remainder nonpositive.

When $a_i = 0$, $r_i = 0$ and $A_i = \partial_i F(L^{i-1})$ by its integral definition. The identity $r_iA_i = F(L^i) - F(L^{i-1})$ remains valid. When $a_i = 1$, $r_i = 1/2$ and $w_i = 0$. Equal weights make the corresponding summation-by-parts coefficient vanish, so any deterministic tie order works. Sorting by (w_i, i) is a measurable choice with $O(d \log d)$ comparison cost. The field may change discontinuously when coordinates exchange order; the regret proof needs only its pointwise comparison and moment bound.

□

The coordinate telescoping template is related to the continuous and online double-greedy analyses of Niazadeh et al. (2020); Roughgarden & Wang (2018). The prescribed rational point and the ordered weighted sum in Eq. (40) give the endpoint-retaining inequality required for the present common-state construction.

C.2 Proof of Proposition 3.3 and complement feasibility

For $x, o \in \mathcal{K}$, set

$$J_x(a) = f(x \odot (\mathbf{1} - a)), \qquad P(x) = \frac{x}{2-x}, \qquad k(f,x) = -g(J_x, \mathbf{1} - x).$$

We prove

$$f(x \odot o) + f(x) - \tfrac{9}{4}f(P(x)) \le \langle k(f,x), o - x\rangle,$$

and the feasibility of the action and every original-domain gradient location.

Complete proof of Proposition 3.3. The gradient $\nabla J_x(a) = -x \odot \nabla f(x \odot (\mathbf{1} - a))$ is antitone in a: the argument order is reversed once by the complement, and the gradient order once more by the negative multiplier. Thus J_x is DR-submodular. In Lemma 3.2 set $a = \mathbf{1} - x$, $b = \mathbf{1} - o$. Then $r = (\mathbf{1} - x)/(2 - x)$, $J_x(b) = f(x \odot o)$, $J_x(0) = f(x)$, and $J_x(r) = f(P(x))$. Most importantly,

$$b - a = x - o = -(o - x),$$

which gives Eq. (15) with the stated sign. The denominator $2 - x$ lies coordinatewise in $[1, 2]$, so $P(x)$ is defined even at $x_i = 0$ or 1 and satisfies $0 \le P(x) \le x$. Every auxiliary cube point z, including every chain point, corresponds to $0 \le x \odot (\mathbf{1} - z) \le x$, which is feasible by down-closedness.

The auxiliary state $a = \mathbf{1} - x$ need not lie in $\mathcal{K}$: it is an argument of J_x, whose entire cube maps into $[0, x] \subseteq \mathcal{K}$. In particular,

$$x \odot \left(\mathbf{1} - \frac{\mathbf{1} - x}{2 - x}\right) = \frac{x}{2 - x} = P(x).$$

The construction is defined at both coordinate endpoints. The matching comparator difference follows from $b = \mathbf{1} - o$, not from any coordinate ordering between o and x. At fixed x, the chain rule differentiates only the auxiliary argument, so it introduces neither derivatives of the coordinate order nor second derivatives of f. □

D CANCELLATION AND CERTIFICATION OF THE APPROXIMATION COEFFICIENT

This appendix derives the cancellation weights, proves Theorem 3.4, and certifies the delay and approximation coefficient. The online learner uses these fixed constants; no parameter optimization is performed on a round.

D.1 DERIVATION OF THE CANCELLATION PARAMETERS

For any delay $s \in (0,1)$ satisfying $\zeta_s \le \mu_s/3$, the coefficients of the two nonlinear values in the trajectory comparison are $\mu_s - \ell$ and 2ℓ. Matching them to a single copy of the box comparison requires

$$\mu_s - \ell = 2\ell = b_s, \qquad \text{hence} \qquad \ell = \mu_s/3, \qquad b_s = 2\mu_s/3.$$

Normalizing the two reward coefficients into probabilities requires

$$\frac{9}{4}\lambda b_s = 1 - \lambda, \qquad \text{hence} \qquad \lambda = \frac{1}{1 + 3\mu_s/2}.$$

The remaining requirement $\zeta_s \le \mu_s/3$ is exactly the condition permitting the nonnegative residual union term to be discarded. Thus the coefficient is derived from the two comparisons, with only the delay left to choose.

D.2 COMPLETE PROOF OF THEOREM 3.4

Complete proof of Theorem 3.4. Because $\mu_s - \ell = 2\ell = b_s$, multiplying the trajectory inequality Eq. (10) by λ and moving its nonlinear terms to the left gives

$$\alpha_* f(o) - \lambda f(Y_1(x)) - \lambda b_s \big[f(x \odot o) + f(x)\big] \le \lambda \langle q_s(f,x) + \ell \nabla f(x), o - x\rangle.$$

Multiplying the box inequality Eq. (15) by λb_s gives

$$\lambda b_s \big[f(x \odot o) + f(x)\big] - \tfrac{9}{4}\lambda b_s f(P(x)) \le \lambda b_s \langle k(f,x), o - x\rangle.$$

Adding cancels both nonlinear values exactly. Using $(9/4)\lambda b_s = 1 - \lambda$ and the definition of G leaves

$$\alpha_* f(o) - \lambda f(Y_1(x)) - (1-\lambda) f(P(x)) \le \langle G(f,x), o - x\rangle.$$

The play coin realizes the weighted reward in expectation. Its distribution depends only on x, and Propositions 3.1 and 3.3 give the feasibility of both actions and all gradient locations. This uses a mixture of feasible actions, rather than a concavity assertion about their average. The argument holds for every o with the same field $G(f,x)$. □

D.3 CHOICE OF THE DELAY

The resulting one-dimensional coefficient family is

$$\alpha(s) = \lambda \theta_s = \frac{(2-s)e^s - 1}{e + \frac{3}{2}(e^s - 1)}.$$

For perspective, the sign of $\alpha'(s)$ is the sign of

$$\psi(s) = (e - \tfrac{3}{2})(1 - s) + \tfrac{3}{2} - \tfrac{3}{2}e^s.$$

Indeed, differentiating the displayed quotient gives $\alpha'(s) = e^s \psi(s)/(e + \frac{3}{2}(e^s - 1))^2$. Since $\psi'(s) = -(e - \frac{3}{2}) - \frac{3}{2}e^s < 0$, $\psi(0) > 0$, and $\psi(1) < 0$, the unrestricted scalar expression has a unique maximizer. The rational choice $s = 199/500$ gives the required coefficient with positive feasibility slack, as certified in Appendix D.4. No maximization routine is part of the online algorithm.

D.4 Rigorous certification of the 0.4013 coefficient

The algorithm fixes the rational delay $s = 199/500$; no parameter optimization is required. The following elementary bounds justify both its approximation coefficient and the sign used in discarding the union value.

For $0 \le x \le 1$, let $S_n(x) = \sum_{j=0}^{n} x^j/j!$. The first omitted term is $x^{n+1}/(n+1)!$, and every subsequent ratio is at most $x/(n+2)$. Therefore

$$S_n(x) \le e^x \le S_n(x) + \frac{x^{n+1}}{(n+1)!}\frac{1}{1 - x/(n+2)}. \tag{41}$$

All endpoints are rational when x is rational. With $n = 18$, applying this at $x = 1$ and $x = 199/500$ and substituting into the rational expressions in Eq. (8) and Eq. (16) yields

$$0.40130636651 < \alpha_* < 0.40130636653, \qquad 0.004823 < \ell - \zeta_s < 0.004825. \tag{42}$$

For clarity, the two expressions whose signs and values are bounded are

$$\alpha_* = \frac{(2-s)e^s - 1}{e + \frac{3}{2}(e^s - 1)}, \qquad \ell - \zeta_s = \frac{(s - \frac{5}{3})e^s + \frac{5}{3} + \frac{1}{2}(1-s)^2(1 + 3s/4)}{e}.$$

Using the positive numerator and denominator intervals in Eq. (41) gives Eq. (42) by rational arithmetic. The decimal endpoints in Eq. (42) denote exact rationals. The first interval verifies the claimed coefficient, while the second verifies $\ell \ge \max\{0, \zeta_s\}$ because $\ell > 0$. Also

$$\mu_s - \ell = 2\ell = b_s, \qquad \frac{9}{4}\lambda b_s = \frac{3}{2}\lambda\mu_s = 1 - \lambda,$$

which are the two exact cancellation identities used in Theorem 3.4.

E One-query field estimation

This appendix gives the complete implementation analysis for the field in Theorem 3.4. We first prove the baseline estimator guarantee Eq. (21), then give the geometry-dependent and stochastic-gradient refinements. These are estimation results: Appendix F derives the corresponding regret bounds from conditional unbiasedness and the second-moment guarantees.

E.1 Sampling the trajectory integral

Let $c_0 = 1 - e^{-1}$. A path time with density e^{v-1}/c_0 can be sampled as $V = \log(1 + (e-1)U)$ for $U \sim \mathrm{Unif}[0,1]$. The definition of q_s gives

$$\mathbb{E}\big[c_0\omega_V(x) \odot \nabla f(Y_V(x)) \mid f, x\big] = q_s(f, x).$$

The query $Y_V(x)$ lies in $[0, x]$, and $\|\omega_V\|_\infty \le 1$ bounds this estimator's second moment by $c_0^2 B^2$. Thus the integral specifies a one-sample expectation, not numerical integration or multiple oracle calls.

E.2 Constructing the sampled box-field segment

Let $\sigma = (\sigma_1, \dots, \sigma_d)$ be the permutation sorting $w_i = (1 - x_i)x_i$. Set $a = \mathbf{1} - x$, $r = a/(\mathbf{1} + a)$. Draw a position j uniformly from $[d]$ and write $i = \sigma_j$. The endpoints immediately before processing this coordinate are

$$L_h = \begin{cases} r_h, & h \in \{\sigma_1, \dots, \sigma_{j-1}\}, \\ 0, & \text{otherwise}, \end{cases} \qquad U_h = \begin{cases} r_h, & h \in \{\sigma_1, \dots, \sigma_{j-1}\}, \\ 1, & \text{otherwise}. \end{cases}$$

With probability $1 - r_i$, put $Z = L + V r_i e_i$; otherwise put $Z = U - V(1 - r_i)e_i$, using an independent uniform V. Then query at $z = x \odot (\mathbf{1} - Z)$ and return $dx_i[\nabla f(z)]_i e_i$ in the original coordinate system. No other segment is evaluated. This is the permutation-explicit version of Eq. (22).

Unbiasedness and second moment. The chain rule gives $-x_i[\nabla f(x \odot (\mathbf{1} - Z))]_i = \partial_i J_x(Z)$. Conditional on $I = i$, the two segment averages therefore have mean $g_i(J_x, \mathbf{1} - x)$ before negation; averaging over the uniform coordinate gives $\mathbb{E}\widehat{k}_i = -g_i(J_x, \mathbf{1} - x) = k_i(f, x)$. This explains the positive sign. Equivalently, for fixed i, the conditional mean of $-x_i[\nabla f(z)]_i$ is $(1-r_i)A_i + r_i C_i$ for $F = J_x$. The minus sign in $k = -g$ gives the positive multiplier in the returned vector. Averaging the uniformly selected coordinate proves unbiasedness. For $x_i = 0$, the returned coordinate is zero and all original query locations still belong to $[0, x]$. Thus lower-dimensional feasible sets need no special extension of the oracle.

The useful moment bound is $\mathbb{E}\|\widehat{k}\|_2^2 \le 2dB^2$, rather than the coarser d^2B^2. For completeness, define the analysis-only endpoint bounds $m_i = \max\{|\partial_i f(0)|, |\partial_i f(x)|\}$. Since $0 \le z \le x$, antitonicity places $\partial_i f(z)$ between these endpoint partials. The uniform-coordinate estimator therefore satisfies

$$\mathbb{E}\|\widehat{k}\|_2^2 \le d \sum_i x_i^2 m_i^2 \le d\big(\|\nabla f(0)\|_2^2 + \|\nabla f(x)\|_2^2\big) \le 2dB^2. \tag{43}$$

The endpoint gradients are not queried to construct this estimator. □

E.3 Unbiasedness and second moment of the complete field

After committing the action, independently select one branch J with the probabilities in Eq. (20) and evaluate only its gradient, as specified in Eq. (23). Only the selected branch is evaluated. Thus the single query is at $Y_V(x)$, x, or $x \odot (\mathbf{1} - Z)$, all in $[0, x]$; the integral notation specifies an expectation, not extra calls or numerical quadrature.

The trajectory expectation from Appendix E.1, the endpoint branch, and $\mathbb{E}\widehat{k} = k$ give

$$\mathbb{E}[\widehat{G} \mid f, x] = \lambda\big[q_s(f, x) + \ell \nabla f(x) + b_s k(f, x)\big] = G(f, x).$$

For the moment bound, $\|\omega_V\|_\infty \le 1$ and Eq. (43) give

$$\mathbb{E}[\|\widehat{G}\|_2^2 \mid f, x] \le \lambda^2 B^2 \left[\frac{c_0^2}{\pi_0} + \frac{\ell^2}{\pi_1} + \frac{2db_s^2}{\pi_2}\right] = B^2 \Gamma_d^2.$$

This proves Eq. (21). These probabilities minimize this bound among fixed positive branch probabilities by Cauchy–Schwarz. Importance sampling and projected ascent are standard; their role here is to implement the new comparison without increasing the query budget.

F Online Regret and Learning Extensions

This appendix completes the online guarantees in Theorem 5.1. The inputs are the comparator-uniform inequality of Theorem 3.4 and the unbiased, second-moment-controlled estimator of Section 4, whose complete analysis is in Appendix E.

F.1 Projection inequality and conditioning

For any $o \in \mathcal{K}$, nonexpansiveness of projection gives

$$\|x_{t+1} - o\|_2^2 \le \|x_t + \eta \widehat{G}_t - o\|_2^2 = \|x_t - o\|_2^2 - 2\eta \langle \widehat{G}_t, o - x_t \rangle + \eta^2 \|\widehat{G}_t\|_2^2.$$

Rearrange, sum, discard the nonnegative final squared distance, and use $\|x_1 - o\|_2 \le D$. This proves the pathwise inequality

$$\sum_{t=1}^{T} \langle \widehat{G}_t, o - x_t \rangle \le \frac{D^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \|\widehat{G}_t\|_2^2. \tag{44}$$

It holds simultaneously for all comparators on each realized update sequence. This is the standard projected online ascent argument (Zinkevich, 2003).

Let $\mathcal{F}_{t-1}$ contain all past actions, queries, responses, and adversary randomness, and let $\mathcal{H}_t$ additionally contain the current function f_t and state x_t, before the fresh play and query draws. Let $\mathcal{F}_t$ also contain the current draws and response, so $\mathcal{F}_{t-1} \subseteq \mathcal{H}_t \subseteq \mathcal{F}_t \subseteq \mathcal{H}_{t+1}$. The state is $\mathcal{F}_{t-1}$-measurable; the sorted-coordinate convention is measurable, and projection preserves measurability

of the next state. For stochastic feedback, the oracle conditions are imposed given every quantity fixed before the call, including $\mathcal{H}_t$, the play, the chosen branch, coordinate, time, and query location. This ensures $\mathbb{E}[\widehat{G}_t \mid \mathcal{H}_t] = G(f_t, x_t)$ by iterated expectation. The query draws are independent of the play coin conditional on $\mathcal{H}_t$. The adversary can use the complete preceding history but fixes f_t before these current draws.

Write $G_t = G(f_t, x_t)$ and $\bar{f}_t(x_t) = \lambda f_t(Y_1(x_t)) + (1-\lambda) f_t(P(x_t))$. The structural and sampling inputs say, respectively,

$$\alpha_* f_t(o) - \bar{f}_t(x_t) \le \langle G_t, o - x_t \rangle \quad \text{for every } o \in \mathcal{K},$$

$$\mathbb{E}[f_t(p_t) \mid \mathcal{H}_t] = \bar{f}_t(x_t), \qquad \mathbb{E}[\widehat{G}_t \mid \mathcal{H}_t] = G_t, \qquad \mathbb{E}[\|\widehat{G}_t\|_2^2 \mid \mathcal{H}_t] \le B^2 \Gamma_d^2.$$

These identities separate action randomization from field estimation.

Complete proof of the oblivious part of Theorem 5.1. First take $D, B > 0$. For an oblivious sequence, compactness and continuity give a fixed hindsight maximizer o^*. Before the fresh draws, o^* and x_t are fixed under conditioning on $\mathcal{H}_t$, so the displayed conditional reward and estimator identities apply against this comparator. Sum Eq. (19) with $o = o^*$, take expectations, and use these identities, Eq. (44), and Eq. (21). The result is

$$\mathrm{Reg}_{\alpha_*}(T) \le \mathbb{E} \sum_{t=1}^{T} \langle \widehat{G}_t, o^* - x_t \rangle \le \frac{D^2}{2\eta} + \frac{\eta}{2} \sum_t \mathbb{E}\|\widehat{G}_t\|_2^2 \le \frac{D^2}{2\eta} + \frac{\eta T B^2 \Gamma_d^2}{2} = DB\Gamma_d \sqrt{T}$$

with $\eta = D/(B\Gamma_d\sqrt{T})$. The same calculation uses Γ_R when the geometry-dependent estimator is selected. Nonnegative rewards and $\alpha_* > 0.4013$ permit rounding down the coefficient. Feasibility and the per-round query count follow from Section 4 and the projection update. □

F.2 Non-anticipating adaptive adversary

Complete proof of the adaptive part of Theorem 5.1. Use the nested information sets from Appendix F.1. Compactness and continuity give a hindsight maximizer of the realized cumulative function; select a measurable maximizer o^*. This comparator can depend on the entire realized estimate sequence and is not assumed to be $\mathcal{H}_t$-measurable. By contrast, x_t and G_t are $\mathcal{H}_t$-measurable, and $\widehat{G}_t$ is $\mathcal{F}_t$-measurable.

The integrability condition is exactly the one in Section 2. Integrating the gradient along feasible segments gives $|f_t(x) - f_t(0)| \le DB$ for $x \in \mathcal{K}$. Thus $\mathbb{E} f_t(0) < \infty$ makes the cumulative reward and its maximum integrable. The second-moment bound above makes all the linear and martingale terms below integrable as well.

Set $e_t = G_t - \widehat{G}_t$ and $M_T = \sum_{t=1}^{T} e_t$. Conditional unbiasedness gives $\mathbb{E}[e_t \mid \mathcal{H}_t] = 0$, and hence $\mathbb{E}[e_t \mid \mathcal{F}_{t-1}] = 0$. In particular,

$$\mathbb{E}\langle e_t, x_t \rangle = 0.$$

For $j < t$, e_j is $\mathcal{H}_t$-measurable, so

$$\mathbb{E}\langle e_j, e_t \rangle = \mathbb{E}\langle e_j, \mathbb{E}[e_t \mid \mathcal{H}_t] \rangle = 0.$$

The conditional mean of $\widehat{G}_t$ is G_t. Expanding the square, then using these zero cross terms, gives

$$\begin{aligned} \mathbb{E}\|M_T\|_2^2 &= \sum_t \mathbb{E}\|e_t\|_2^2 \\ &= \sum_t \mathbb{E}\big[\|\widehat{G}_t\|_2^2 - \|G_t\|_2^2\big] \le \sum_t \mathbb{E}\|\widehat{G}_t\|_2^2 \le T B^2 \Gamma_d^2. \end{aligned}$$

The structural inequality and the projection bound hold simultaneously for every comparator on each realized sequence. Apply them pathwise to o^* and substitute $G_t = \widehat{G}_t + e_t$:

$$\alpha_* \max_{o \in \mathcal{K}} \sum_t f_t(o) - \sum_t \bar{f}_t(x_t) \le \frac{D^2}{2\eta} + \frac{\eta}{2} \sum_t \|\widehat{G}_t\|_2^2 + \sum_t \langle e_t, o^* - x_t \rangle.$$

Since $0 \in \mathcal{K}$, $\|o^*\|_2 \le D$. Therefore

$$\mathbb{E}\sum_t \langle e_t, o^* - x_t\rangle = \mathbb{E}\langle M_T, o^*\rangle - \mathbb{E}\sum_t \langle e_t, x_t\rangle$$
$$\le D\,\mathbb{E}\|M_T\|_2 \le D\sqrt{\mathbb{E}\|M_T\|_2^2} \le DB\Gamma_d\sqrt{T}.$$

The reward-mixture identity is used only after taking expectation of the cumulative reward:

$$\mathbb{E}\sum_t f_t(p_t) = \mathbb{E}\sum_t \bar{f}_t(x_t).$$

It is not paired conditionally with o^*, and remains valid when future functions depend on the realized play. Taking expectations in the pathwise bound and setting $\eta = D/(B\Gamma_d\sqrt{T})$ yields

$$\mathrm{Reg}_{\alpha_*}(T) \le \frac{D^2}{2\eta} + \frac{\eta T B^2\Gamma_d^2}{2} + DB\Gamma_d\sqrt{T} = 2DB\Gamma_d\sqrt{T}.$$

Nonnegativity permits rounding α_* down to 0.4013. Feasibility and the oracle budget are unchanged because the adversary fixes f_t before the fresh draws. □

G Stochastic gradient extensions

This appendix extends the online guarantees to a single stochastic-oracle model: conditionally unbiased gradients with uniformly bounded conditional second moment. Appendix G.1 gives the estimator and its moment bound. Appendices G.2 and G.3 then prove the regret guarantees for oblivious and non-anticipating adaptive adversaries, respectively.

G.1 Stochastic gradient estimation

Let $\mathcal{H}_t$ contain the complete preceding history together with the current function f_t and state x_t, before the fresh play and query draws. For this subsection, pre-call information means every quantity fixed before the current oracle call: $\mathcal{H}_t$, the play, the chosen branch, coordinate, time, and query location. The adversary fixes f_t before the fresh play and query draws, and the query draws are independent of the play coin conditional on $\mathcal{H}_t$. The oracle conditions below are imposed after conditioning on all this pre-call information; iterated expectation then preserves the unbiasedness of the sampled field.

Suppose $\xi_t(z)$ satisfies

$$\mathbb{E}[\xi_t(z) \mid \text{pre-call information}] = \nabla f_t(z), \qquad \mathbb{E}[\|\xi_t(z)\|_2^2 \mid \text{pre-call information}] \le B_1^2. \tag{45}$$

The path and endpoint estimators have second moments at most $c_0^2 B_1^2$ and $\ell^2 B_1^2$, respectively, before branch importance weighting. The stochastic box estimator obeys

$$\mathbb{E}\|dx_I[\xi_t(z)]_I e_I\|_2^2 \le d^2 B_1^2$$

because $x_I \le 1$ and a coordinate's squared magnitude is bounded by the vector's squared norm at that queried point. Define

$$\widetilde{S}_d = c_0 + \ell + b_s d, \qquad (\widetilde{\pi}_0, \widetilde{\pi}_1, \widetilde{\pi}_2) = (c_0, \ell, b_s d)/\widetilde{S}_d,$$

and

$$\widetilde{\Gamma}_d = \lambda\widetilde{S}_d = \lambda(c_0 + \ell + b_s d). \tag{46}$$

Use these probabilities in the three-branch sampler Eq. (23), replacing the selected exact gradient by the single stochastic response ξ_t. Conditional on the selected query point, this replacement preserves the mean of every branch. Iterated expectation over the path-time, coordinate, segment, and branch draws therefore recovers $G(f_t, x_t)$. The three unscaled branch second-moment bounds are $c_0^2 B_1^2$, $\ell^2 B_1^2$, and $b_s^2 d^2 B_1^2$, so importance sampling with probabilities proportional to their square roots gives

$$\mathbb{E}[\widehat{G}_t \mid \mathcal{H}_t] = G(f_t, x_t), \qquad \mathbb{E}[\|\widehat{G}_t\|_2^2 \mid \mathcal{H}_t] \le B_1^2\widetilde{\Gamma}_d^2. \tag{47}$$

Each round still uses one feasible stochastic-gradient query and one projection. The noise may be temporally dependent; only the conditional requirements in Eq. (45) are used.

G.2 OBLIVIOUS ADVERSARY

Run Algorithm 1 with the stochastic response $\xi_t(z_t)$ in place of $\nabla f_t(z_t)$, branch probabilities $(\widetilde{\pi}_0, \widetilde{\pi}_1, \widetilde{\pi}_2)$, and step size

$$\eta = \frac{D}{B_1 \widetilde{\Gamma}_d \sqrt{T}}. \tag{48}$$

The degenerate cases $D = 0$ or $B_1 = 0$ are handled by always playing, querying, and projecting 0.

Lemma G.1 (Stochastic gradients against an oblivious adversary)**.** *Under Section 2 and the stochastic-oracle assumption Eq. (45), the stochastic version of Algorithm 1 satisfies*

$$\mathrm{Reg}_{\alpha_*}(T) \le D B_1 \widetilde{\Gamma}_d \sqrt{T} \tag{49}$$

for every oblivious adversary. The same bound holds at coefficient 0.4013*, and every round uses one feasible stochastic-gradient query and one projection.*

Proof. Let $\mathcal{F}_{t-1}$ contain all past actions, queries, oracle responses, and algorithmic randomness. Let $\mathcal{H}_t$ additionally contain the current function f_t and state x_t, before the fresh play and query draws, and let $\mathcal{F}_t$ additionally contain those draws and the current stochastic-oracle response. Thus

$$\mathcal{F}_{t-1} \subseteq \mathcal{H}_t \subseteq \mathcal{F}_t \subseteq \mathcal{H}_{t+1}.$$

The state x_t is $\mathcal{F}_{t-1}$-measurable. The pre-call conditioning in Eq. (45), followed by iterated expectation over all current query draws, gives Eq. (47). The action coin is conditionally independent of the query draws, while the oracle conditions hold after conditioning on the realized play. Thus, writing

$$G_t = G(f_t, x_t), \qquad \bar{f}_t(x_t) = \lambda f_t(Y_1(x_t)) + (1-\lambda) f_t(P(x_t)),$$

we have

$$\mathbb{E}[f_t(p_t) \mid \mathcal{H}_t] = \bar{f}_t(x_t), \qquad \mathbb{E}[\widehat{G}_t \mid \mathcal{H}_t] = G_t, \qquad \mathbb{E}[\|\widehat{G}_t\|_2^2 \mid \mathcal{H}_t] \le B_1^2 \widetilde{\Gamma}_d^2. \tag{50}$$

For every realized update sequence and every $o \in \mathcal{K}$, nonexpansiveness of Euclidean projection gives

$$\|x_{t+1} - o\|_2^2 \le \|x_t + \eta \widehat{G}_t - o\|_2^2 = \|x_t - o\|_2^2 - 2\eta \langle \widehat{G}_t, o - x_t \rangle + \eta^2 \|\widehat{G}_t\|_2^2.$$

After rearranging and summing,

$$\sum_{t=1}^{T} \langle \widehat{G}_t, o - x_t \rangle \le \frac{D^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \|\widehat{G}_t\|_2^2. \tag{51}$$

This inequality holds pathwise and simultaneously for all comparators.

For an oblivious sequence, compactness and continuity give a fixed hindsight maximizer $o^* \in \arg\max_{o \in \mathcal{K}} \sum_t f_t(o)$. Theorem 3.4 gives

$$\alpha_* f_t(o^*) - \bar{f}_t(x_t) \le \langle G_t, o^* - x_t \rangle.$$

We sum this inequality, take expectations, and use Eq. (50). The fixed comparator o^* and the state x_t are measurable before the current oracle response. Applying Eq. (51) and the conditional moment bound yields

$$\begin{aligned}
\mathrm{Reg}_{\alpha_*}(T) &\le \mathbb{E} \sum_{t=1}^{T} \langle \widehat{G}_t, o^* - x_t \rangle \\
&\le \frac{D^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \mathbb{E}\|\widehat{G}_t\|_2^2 \\
&\le \frac{D^2}{2\eta} + \frac{\eta T B_1^2 \widetilde{\Gamma}_d^2}{2} = D B_1 \widetilde{\Gamma}_d \sqrt{T},
\end{aligned}$$

where the last equality uses Eq. (48). Nonnegativity and $\alpha_* > 0.4013$ permit rounding down the coefficient. If $D = 0$, then $\mathcal{K} = \{0\}$. If $B_1 = 0$, Eq. (45) and Jensen's inequality imply $\nabla f_t(z) = 0$ at every feasible query point, so the stated zero-regret convention applies. □

G.3 NON-ANTICIPATING ADAPTIVE ADVERSARY

Lemma G.2 (Stochastic gradients against a non-anticipating adaptive adversary)**.** *Under Section 2 and the stochastic-oracle assumption Eq. (45), the stochastic version of Algorithm 1 satisfies*

$$\mathrm{Reg}_{\alpha_*}(T) \le 2DB_1\widetilde{\Gamma}_d\sqrt{T} \tag{52}$$

for every non-anticipating adaptive adversary and the realized-sequence benchmark Eq. (3). The same bound holds at coefficient 0.4013*, and every round uses one feasible stochastic-gradient query and one projection.*

Proof. Use the information sets from Appendix G.2. The adversary may use $\mathcal{F}_{t-1}$ when selecting f_t, but fixes f_t before the current play, query, and oracle draws. Hence x_t, f_t, and G_t are $\mathcal{H}_t$-measurable, while Eq. (50) remains valid. Compactness and continuity give a measurable hindsight maximizer o^* of the realized cumulative reward. Unlike in the oblivious case, o^* may depend on the entire realized sequence and need not be $\mathcal{H}_t$-measurable.

The integrability assumption in Section 2 is sufficient here. Indeed, the exact gradient bound gives $|f_t(x) - f_t(0)| \le DB$ on $\mathcal{K}$, while Eq. (50) makes all stochastic linear terms square-integrable.

Set

$$e_t = G_t - \widehat{G}_t, \qquad M_T = \sum_{t=1}^{T} e_t.$$

Conditional unbiasedness gives $\mathbb{E}[e_t \mid \mathcal{H}_t] = 0$ and therefore $\mathbb{E}[e_t \mid \mathcal{F}_{t-1}] = 0$. Since x_t is $\mathcal{F}_{t-1}$-measurable,

$$\mathbb{E}\langle e_t, x_t\rangle = 0.$$

For $j < t$, the error e_j is $\mathcal{H}_t$-measurable, so martingale orthogonality gives

$$\mathbb{E}\langle e_j, e_t\rangle = \mathbb{E}\langle e_j, \mathbb{E}[e_t \mid \mathcal{H}_t]\rangle = 0.$$

Moreover, because $\mathbb{E}[\widehat{G}_t \mid \mathcal{H}_t] = G_t$,

$$\begin{aligned}\mathbb{E}\|M_T\|_2^2 &= \sum_{t=1}^{T} \mathbb{E}\|e_t\|_2^2 \\ &= \sum_{t=1}^{T} \mathbb{E}\big[\|\widehat{G}_t\|_2^2 - \|G_t\|_2^2\big] \\ &\le \sum_{t=1}^{T} \mathbb{E}\|\widehat{G}_t\|_2^2 \le TB_1^2\widetilde{\Gamma}_d^2.\end{aligned}$$

The structural comparison and Eq. (51) hold pathwise for every comparator. Apply them to the realized o^* and substitute $G_t = \widehat{G}_t + e_t$:

$$\alpha_* \max_{o\in\mathcal{K}} \sum_{t=1}^{T} f_t(o) - \sum_{t=1}^{T} \bar{f}_t(x_t) \le \frac{D^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\|\widehat{G}_t\|_2^2 + \sum_{t=1}^{T}\langle e_t, o^* - x_t\rangle.$$

Since $0 \in \mathcal{K}$, we have $\|o^*\|_2 \le D$. Consequently,

$$\begin{aligned}\mathbb{E}\sum_{t=1}^{T}\langle e_t, o^* - x_t\rangle &= \mathbb{E}\langle M_T, o^*\rangle - \mathbb{E}\sum_{t=1}^{T}\langle e_t, x_t\rangle \\ &\le D\,\mathbb{E}\|M_T\|_2 \le D\sqrt{\mathbb{E}\|M_T\|_2^2} \le DB_1\widetilde{\Gamma}_d\sqrt{T}.\end{aligned}$$

The reward identity is used after taking expectation:

$$\mathbb{E}\sum_{t=1}^{T} f_t(p_t) = \mathbb{E}\sum_{t=1}^{T} \bar{f}_t(x_t).$$

Taking expectations in the pathwise comparison, applying Eq. (50), and using Eq. (48) now gives

$$\begin{aligned}\mathrm{Reg}_{\alpha_*}(T) &\le \frac{D^2}{2\eta} + \frac{\eta T B_1^2 \widetilde{\Gamma}_d^2}{2} + D B_1 \widetilde{\Gamma}_d \sqrt{T} \\ &= 2 D B_1 \widetilde{\Gamma}_d \sqrt{T}.\end{aligned}$$

Nonnegativity permits rounding α_* down to 0.4013. Feasibility and the one-query budget are unchanged because each stochastic oracle call is made at the same feasible location as in the exact-gradient algorithm. The degenerate cases follow as in Lemma G.1. □