ONLINE CONVEX OPTIMIZATION WITH WARM STARTS: UPPER AND LOWER REGRET BOUNDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Online learners often begin from a retrieved, teacher-supplied, or previously trained parameter whose reliability is unknown. We study how this warm start affects static regret through its distance to the closest offline optimum. If that distance is revealed, online gradient descent achieves regret proportional to the distance times the square root of the horizon, and a matching lower bound holds at every distance up to half the domain diameter. Without the distance, a learning-rate ensemble and a translated parameter-free method adapt to its realized value, yet no algorithm can attain the oracle rate within one universal constant at every distance. Perfect trust in an exact warm start also entails linear sensitivity to small errors. With a known strong-convexity modulus, regularized follow-the-leader obtains a quadratic distance term without distance-dependent tuning, although its additive logarithmic term remains unmatched. A two-point bandit bound holds for oblivious losses in dimension at least two when a certified interior ball around the warm start contains a closest optimum. Logistic-loss experiments using the same online and offline objective illustrate the benefit of reliable retrieval and the cost of corruption.

1 INTRODUCTION

Machine learning systems rarely start a new task from scratch. A recommender may reuse the parameters of a similar user, and a fine-tuning pipeline may begin from a teacher or retrieved model. The downstream learner then faces a stream of losses from that initialization. A useful starting point can reduce the cost of adaptation; an unreliable one can increase it.

This paper studies that downstream layer. A single task consists of n rounds of online convex optimization over a compact convex set $\mathcal{K} \subset \mathbb{R}^d$ of diameter D. Before the first round the learner receives a warm start $h \in \mathcal{K}$: a feasible parameter, as returned by retrieval, distillation, or meta-initialization, rather than a forecast of the next loss or gradient. After the losses are revealed, write $F = \sum_{i=1}^{n} \ell_i$ for the cumulative objective, $\mathcal{W}^\star = \arg\min_{w \in \mathcal{K}} F(w)$ for the offline optimizer set, and

$$\varepsilon = \operatorname{dist}(h, \mathcal{W}^\star) .$$

We call ε the realized warm-start distance. It is an ex post property of (h, F) and is unobserved during the task unless supplied by an oracle. Static regret r_n is the excess cumulative loss relative to the best point of $\mathcal{K}$ in hindsight.

This downstream question complements the learning of a prior: once a parameter has been supplied, how does its distance to the eventual optimizer set affect attainable regret? It places warm-started optimization within the study of learning-augmented algorithms, with an initial solution as advice rather than a forecast of future losses.

The first difficulty is deciding how far to move from h. Small steps preserve a good initialization but can make recovery from a poor one slow. Large steps protect against a poor initialization while spending more regret when it was already accurate. If ε were revealed, the standard OGD potential would balance these effects at step size $\varepsilon/(L\sqrt{n})$ and give regret at most $\varepsilon L\sqrt{n}$, where L bounds subgradient norms (Zinkevich, 2003). This benchmark separates the value of a warm start from the difficulty of assessing its reliability. In applications the distance is unavailable, since it depends on losses the learner has not yet seen.

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.

Adapting from observed losses offers a way around this missing information. A grid of OGD experts maintains several levels of trust at once, while comparator-adaptive methods control regret directly as a function of distance from h. Both retain a cost for adaptation. We ask whether this cost is inherent by constructing hard tasks whose final optimizer remains at a prescribed distance.

Trust also raises a local question: how does requiring almost no regret at an exact warm start affect recovery from a small error? Following learning-augmented terminology, we call performance under exact advice consistency. Our lower bound studies its tradeoff with sensitivity to a nearby misspecification. We examine this question alongside strong convexity, which supplies stability through the losses themselves and allows a quadratic term in ε without estimating it.

In retrieval and meta-initialization pipelines, successive tasks receive different parameters. Taking these as inputs lets us sum one-task guarantees along the realized distance sequence and test, through a conditional lower bound, whether memory improves worst-case performance. Learning which parameter to supply next remains a complementary statistical problem.

Contributions.

- **Bounds at prescribed distances.** A barrier construction matches the standard OGD rate at every prescribed $\varepsilon \in [0, D/2]$. A supplied distance estimate q yields the bound $(L\sqrt{n}/2)(\varepsilon^2/q + q)$ and quantifies the effect of misspecification.
- **Limits of adaptation and trust.** An anchoring construction rules out a universal constant-factor oracle rate for algorithms that do not receive ε. A separate two-instance argument gives a consistency-sensitivity lower bound under local misspecification. The former is qualitative and does not establish a logarithmic lower bound.
- **Complementary guarantees.** Standard OCO tools give explicit grid and parameter-free guarantees without ε. With known modulus μ, RFTL centered at h has a term $\mu\varepsilon^2/2$, matched from below, plus an unmatched harmonic baseline. We extend the oracle bound to feasible two-point bandit queries and examine the dependence on distance in synthetic and retrieved-parameter tasks.

2 Related Works

Warm starts with predictions. Learned duals and discrete convex analysis connect the running time of an offline solver to the distance between a predicted solution and an optimum (Dinitz et al., 2021; Sakaue & Oki, 2022; Khodak et al., 2022). Measuring distance to the optimizer set avoids dependence on an arbitrary choice when optima are nonunique (Sakaue & Oki, 2023). We use the same distance-to-solution-set measure, but the learner here incurs losses while the objective is still being revealed. The quantity to control is therefore cumulative regret rather than the work needed to solve a known objective.

Optimistic and learning-augmented OCO. Predictable-sequence and optimistic methods exploit forecasts of gradients or regularity across successive losses (Chiang et al., 2012; Rakhlin & Sridharan, 2013; Dekel et al., 2017). Learning-augmented OCO also considers predictions of future costs and collections of imperfect hints (Chen et al., 2015; Li et al., 2023; Bhaskara et al., 2020a;b). A parameter warm start supplies different information: it suggests where a good fixed decision may lie, without predicting the next loss. Accurate initialization can consequently be useful even when successive gradients are difficult to forecast, while predictable gradients need not imply a small distance from h to the offline optimizer.

Meta-learning and bias learning. Lifelong learning and biased regularization estimate priors from collections of tasks (Baxter, 2000; Pentina & Lampert, 2014; Denevi et al., 2019; Balcan et al., 2019). MAML-style initialization and online meta-updates similarly use experience to improve future adaptation (Finn et al., 2017; 2019; Khodak et al., 2019; Cesa-Bianchi et al., 2022). Our guarantees apply after such a procedure supplies h, describing downstream regret conditional on its distance to the optimizer set. How small that distance becomes is the statistical question addressed by the prior-learning method.

Initialization- and comparator-adaptive OCO. The closest upper bounds are the initialization-dependent OGD potential (Zinkevich, 2003), Hedge aggregation (Freund & Schapire, 1997), and comparator-norm adaptive optimization (Streeter & McMahan, 2012; McMahan & Orabona, 2014; Cutkosky & Boahen, 2017; Mhammedi & Koolen, 2020). In particular, translating the origin to h makes the comparator norm equal to the warm-start distance. We combine the Hilbert-space Krichevsky-Trofimov guarantee of Orabona & Pál (2016) with the constrained reduction of Cutkosky & Orabona (2018); RFTL with a regularizer centered at h uses the standard regularization and stability argument (Hazan, 2016). The upper bounds thus specialize established algorithms. Our lower bounds ask a different question: how hard can a task be when the distance to its optimizer set is fixed exactly? The barrier construction enforces that distance, and the anchoring construction transfers the qualitative comparator-adaptive obstruction to a compact domain.

Consistency and robustness. Learning-augmented algorithms balance gains from accurate advice against protection from errors (Lykouris & Vassilvitskii, 2018; Purohit et al., 2018; Wei & Zhang, 2020; Mitzenmacher & Vassilvitskii, 2022; Eliáš et al., 2024). In this literature, consistency describes performance under perfect advice, while robustness describes worst-case protection under arbitrary advice error. We use consistency in that standard sense. Our two-instance lower bound addresses a local sensitivity question by coupling regret at an exact warm start with regret under a nearby misspecification through the learner's shared initial action.

3 One-Task Setup

Let $\mathcal{K} \subset \mathbb{R}^d$ be convex and compact with diameter

$$D := \sup_{u,v \in \mathcal{K}} \|u - v\|_2 .$$

A task has horizon n. The learner receives $h \in \mathcal{K}$ and then, for $i = 1, \dots, n$, plays $w_i \in \mathcal{K}$ and observes a convex L-Lipschitz loss $\ell_i : \mathcal{K} \to \mathbb{R}$ together with a subgradient $g_i \in \partial \ell_i(w_i)$ satisfying $\|g_i\|_2 \le L$. Write

$$F(w) := \sum_{i=1}^{n} \ell_i(w), \qquad \mathcal{W}^\star := \arg\min_{w \in \mathcal{K}} F(w), \qquad r_n := \sum_{i=1}^{n} \ell_i(w_i) - \min_{w \in \mathcal{K}} F(w).$$

Compactness and Lipschitzness imply that $\mathcal{W}^\star$ is nonempty and compact, so

$$\varepsilon := \operatorname{dist}(h, \mathcal{W}^\star) \in [0, D]$$

is attained. Fix any closest optimizer

$$w^\dagger \in \arg\min_{w \in \mathcal{W}^\star} \|h - w\|_2 ,$$

hence $\|h - w^\dagger\|_2 = \varepsilon$. Measuring distance to the set matters when the cumulative objective has several minimizers: an initialization that is already optimal has distance zero, regardless of which optimizer an offline solver happens to return.

The learner sees neither $w^\dagger$ nor ε during the task. An oracle may reveal ε before the first action in order to define a benchmark, but it does not reveal an optimizer. More generally, a positive scale q may be used as an estimate of the distance; its accuracy enters the bound explicitly. We contrast algorithms that receive ε with those that do not. In the latter case, the actions may use h and past feedback, but not ε or a closest optimizer.

The full-information upper bounds hold for each realized sequence satisfying the stated assumptions. They impose no distributional requirement on the source of h. OGD and the grid require only the subgradient at the played point; RFTL uses the revealed losses to solve its regularized subproblem. The two-point model in Section 7 replaces this feedback by function-value queries and introduces the geometry needed to keep those queries feasible.

Throughout, i indexes rounds. For repeated tasks, t indexes tasks, r_t denotes task regret, $R_T = \sum_t r_t$, and $\mathcal{F}_{t-1}$ is the history before task t. We write $H_n = \sum_{i=1}^{n} 1/i$. In the parameter-free method, a declared reference scale $\rho > 0$ controls regret relative to h; "parameter-free" refers specifically to the absence of ε-dependent tuning.

4 KNOWN-DISTANCE BENCHMARK

The initialization-dependent OGD bound provides a direct starting point. Its two terms charge for the initial distance to the comparator and for subsequent movement. Balancing them identifies the value of knowing the warm-start distance.

Theorem 1 (Warm-start OGD)**.** *Run*

$$w_1 = h, \qquad w_{i+1} = \Pi_{\mathcal{K}}(w_i - \eta g_i), \qquad g_i \in \partial \ell_i(w_i).$$

Then

$$r_n \le \frac{\varepsilon^2}{2\eta} + \frac{\eta}{2}\sum_{i=1}^{n} \|g_i\|_2^2 \le \frac{\varepsilon^2}{2\eta} + \frac{\eta n L^2}{2}.$$

The worst-case choice $\eta = D/(L\sqrt{n})$ gives $r_n \le (L\sqrt{n}/(2D))\varepsilon^2 + DL\sqrt{n}/2$. The infimum over $\eta > 0$ equals $\varepsilon L\sqrt{n}$.

For $\varepsilon > 0$, the oracle step $\eta = \varepsilon/(L\sqrt{n})$ makes the two terms equal. At $\varepsilon = 0$, staying at h gives zero regret. By contrast, the worst-case step size uses D in place of the unknown distance and always gives $r_n \le DL\sqrt{n}$, since $\varepsilon \le D$.

To see where the distance dependence enters, apply nonexpansiveness of projection to any $u \in \mathcal{K}$:

$$\|w_{i+1} - u\|_2^2 \le \|w_i - u\|_2^2 - 2\eta\langle g_i,\, w_i - u\rangle + \eta^2 \|g_i\|_2^2 .$$

Convexity bounds the instantaneous regret by the inner product. Summing telescopes the squared distances, and setting $u = w^\dagger$ leaves the initial term $\left\|h - w^\dagger\right\|_2^2 = \varepsilon^2$. This is precisely the standard OGD potential (Zinkevich, 2003). Replacing the oracle distance by an estimate gives a useful sensitivity calculation.

Theorem 2 (OGD with an estimated distance)**.** *Fix $q > 0$ and set $\eta = q/(L\sqrt{n})$. Then*

$$r_n \le \frac{L\sqrt{n}}{2}\Big(\frac{\varepsilon^2}{q} + q\Big).$$

If $\varepsilon > 0$ and $q = \gamma\varepsilon$ for $\gamma > 0$, the right-hand side equals $\varepsilon L\sqrt{n}\cdot(\gamma + \gamma^{-1})/2$. If $q \ge \varepsilon$, then $r_n \le qL\sqrt{n}$. If $q = \varepsilon$, the oracle bound $\varepsilon L\sqrt{n}$ is recovered.

A factor-two error in either direction increases this bound by only $5/4$. Larger errors have different interpretations: an underestimated q makes the distance penalty ε^2/q dominate, whereas an overestimated q spends more on movement. The worst-case choice $q = D$ is valid without a distance estimate, but even at $\varepsilon = 0$ its bound retains the baseline $DL\sqrt{n}/2$.

To determine whether the linear oracle rate can be improved, we need hard losses with the warm-start distance exactly equal to a prescribed value. A lower bound over a comparator ball would only locate a hard instance somewhere inside that ball. The next construction instead uses a barrier to force the offline optimizer onto its boundary while preserving uncertainty about its sign.

Theorem 3 (Lower bound at a prescribed distance)**.** *Fix any $\bar{\varepsilon} \in [0, D/2]$. For every possibly randomized algorithm there exists a convex L-Lipschitz sequence on $\mathcal{K} = [-D/2, D/2]$, with warm start $h = 0$ and distance $\varepsilon = \bar{\varepsilon}$, such that*

$$\mathbb{E}[r_n] \ge \frac{L\bar{\varepsilon}\sqrt{n}}{8}.$$

The bound holds even when $\bar{\varepsilon}$ is revealed before the task.

Proof idea. Take $m = n$ if n is odd and $m = n - 1$ otherwise, so m is odd and $m \ge n/2$. Sample Rademacher signs ξ_i and set

$$\ell_i(w) = \frac{L}{4}\xi_i w + \frac{3L}{4}\max\{0, |w| - \bar{\varepsilon}\}$$

for $i \le m$, with a possible final zero loss. Each loss is convex and L-Lipschitz. The cumulative objective is linear on $[-\bar{\varepsilon}, \bar{\varepsilon}]$ and increases away from that interval. Because m is odd, $S = \sum_{i=1}^{m} \xi_i$ is nonzero, so the unique optimizer is $w^\star = -\bar{\varepsilon}\,\mathrm{sign}(S)$. Its loss is $-L\bar{\varepsilon}\,|S|\,/4$. The learner's linear losses have conditional mean zero and its barrier losses are nonnegative. Hence $\mathbb{E}[r_n] \ge (L\bar{\varepsilon}/4)\mathbb{E}\,|S| \ge L\bar{\varepsilon}\sqrt{n}/8$ by Khintchine's inequality. Revealing $\bar{\varepsilon}$ leaves the fresh signs unpredictable, which is why the construction remains hard for the oracle learner.

Corollary 1 (Minimax bounds with known distance)**.** *On $\varepsilon \in [0, D/2]$, the minimax expected regret of algorithms that are given ε satisfies*

$$\frac{L\varepsilon\sqrt{n}}{8} \le \mathfrak{R}^\star(\varepsilon) \le \varepsilon L\sqrt{n}.$$

5 Unknown Distance

The estimated-distance bound suggests maintaining several candidate scales rather than committing to one estimate. We first aggregate a finite grid of OGD step sizes, then replace that grid by a comparator-adaptive method. Both use observed gradients to decide how much to trust h.

Theorem 4 (Grid aggregation)**.** *Let $\eta_j = D/(L\sqrt{n}\,2^j)$ for $j = 0, \dots, J$ with $J = \lceil \log_2 n \rceil$, and let a Hedge master with learning rate $\sqrt{8\ln(J+1)/(nD^2L^2)}$ play the convex combination of the corresponding warm-start OGD experts. All experts reuse the single subgradient observed at the master point. The deterministic master satisfies*

$$r_n \le \min_{0\le j\le J}\Big\{\frac{\varepsilon^2}{2\eta_j} + \frac{\eta_j n L^2}{2}\Big\} + DL\sqrt{2n\ln(J+1)},$$

and therefore

$$r_n \le \frac{3}{2}\varepsilon L\sqrt{n} + \frac{3DL}{2\sqrt{n}} + DL\sqrt{2n\ln(1+\lceil\log_2 n\rceil)}.$$

Proof idea. Let $p_{i,j}$ be the master weight and $w_{i,j}$ expert j's action. The learner plays $w_i = \sum_j p_{i,j} w_{i,j}$ and observes one subgradient g_i there. Each expert then takes a projected step along this same g_i, while Hedge uses the linearized losses $\langle g_i,\, w_{i,j}\rangle$. Subtracting their minimum leaves the weights unchanged and puts every loss in $[0, LD]$, because $|\langle g_i,\, w_{i,j} - w_{i,k}\rangle| \le LD$. Convexity at the played point converts the resulting linearized bound into regret. This construction needs one gradient query and $J+1$ projected updates per round.

For $\varepsilon > D/n$, the grid contains a step within a factor of two of the oracle step, yielding the $3/2$ coefficient. Below that scale, the smallest expert contributes at most $3DL/(2\sqrt{n})$. The remaining term is the Hedge aggregation cost. Since it does not shrink with ε, the grid bound can be dominated by aggregation when the warm start is very accurate. A comparator-adaptive method offers a different balance: an explicit cost at h together with a distance-dependent logarithmic term.

Theorem 5 (Translated parameter-free reduction)**.** *Shift the domain to $\mathcal{K}' := \mathcal{K} - h$, which contains the origin. Run the Hilbert-space Krichevsky-Trofimov method of Orabona & Pál (2016, Cor. 5) on L-normalized linearized losses, then apply the constrained-set reduction of Cutkosky & Orabona (2018, Thm. 3). Let $\rho > 0$ be a reference scale fixed before the task (the KT initial wealth on the normalized problem). Playing $w_i = u_i + h$ yields*

$$r_n \le 2L\varepsilon\sqrt{n\ln\Big(1 + \frac{24n^2\varepsilon^2}{\rho^2}\Big)} + 2L\rho\Big(1 - \frac{1}{e\sqrt{\pi n}}\Big).$$

The method uses ρ, L, and n, but not ε.

Proof idea. Translation preserves regret and sends $w^\dagger$ to a comparator of norm ε. Corollary 5 of Orabona & Pál (2016) controls unconstrained linearized regret for unit-ball gradients. The constrained reduction modifies the gradients without increasing their norms and incurs a factor of two; rescaling by L gives the displayed constants. At $\varepsilon = 0$, the bound is at most $2L\rho$. Decreasing ρ therefore improves the guarantee at an exact warm start, but increases the logarithmic term at positive distances. The reference scale sets this balance before the task rather than estimating ε online.

Could another method that does not receive ε remove both costs and retain a constant multiple of the oracle rate everywhere? The following result rules this out, even in one dimension. It leaves open the magnitude of the necessary gap, including whether the logarithm in Theorem 5 is optimal.

Theorem 6 (Impossibility of uniform adaptation)**.** *Fix $d = 1$ and $\mathcal{K} = [-D/2, D/2]$. There do not exist an algorithm A whose actions use neither ε nor a closest optimizer and a finite constant C such that, for every horizon n, every convex L-Lipschitz sequence with warm start $h = 0$, and every distance $\varepsilon \in (0, D/4]$,*

$$\mathbb{E}[r_n] \le C\varepsilon L\sqrt{n}.$$

Proof idea. For arbitrary m, nonzero $u \in [-D/4, D/4]$, and gradients $g_i \in [-L/4, L/4]$, play $g_i w$ for m rounds and $L\,|w-u|$ for m further rounds. Writing $G = \sum_i g_i$, the cumulative objective is $Gw + mL\,|w-u|$. Since $|G| \le mL/4$, it is strictly decreasing to the left of u and strictly increasing to the right. Its unique optimizer is therefore u, so the warm-start distance is exactly $|u|$.

The first m actions cannot depend on the value of u, which appears only in future losses. Define $A_m(g) = \mathbb{E}_A[\sum_{i=1}^m g_i w_i]$. A universal factor-$C$ guarantee would imply

$$A_m(g) - uG \le C\,|u|\,L\sqrt{2m}$$

for every admissible u. Choose the sign of u so that $uG \le 0$, then send $|u|$ to zero while holding g fixed. This forces $A_m(g) \le 0$. If the g_i are independent signs of magnitude $L/4$, predictability gives $\mathbb{E}_g A_m(g) = 0$; because the average is over finitely many nonpositive terms, $A_m(g) = 0$ for every sign vector. Taking $g_i = L/4$ and $u = -D/4$ then gives regret at least $mLD/16$, contradicting the proposed square-root bound for large m. The argument is qualitative: it rules out a uniform constant, not a specific logarithmic rate. The same obstruction is classical for unconstrained comparator-norm adaptive linear optimization (Streeter & McMahan, 2012); here the anchoring losses pin the comparator at a prescribed distance inside a compact domain.

6 Consistency, Sensitivity, and Strong Convexity

The reference-scale tradeoff becomes especially clear at $\varepsilon = 0$. Staying at h is then optimal, but the learner must choose its first action before it can distinguish this case from a nearby misspecification. We measure consistency by worst-case regret at an exact warm start and local sensitivity by regret on tasks at distance $\delta > 0$. This second quantity is not the global robustness measure used in competitive analysis. The same initial action couples the two regrets.

Theorem 7 (Consistency and local sensitivity)**.** *Let $\mathcal{K} = [-D/2, D/2]$ and let A be any possibly randomized algorithm whose actions use neither ε nor a closest optimizer. For every $\delta \in (0, D/2]$ there exist one-task environments $\mathcal{E}_0$ and $\mathcal{E}_\delta$, both using the same warm start h, such that their distances are 0 and δ, respectively, and*

$$\mathbb{E}[r(A, \mathcal{E}_0)] + \mathbb{E}[r(A, \mathcal{E}_\delta)] \ge L\delta.$$

The first action W is a function of h and internal randomness only. Choose the sign σ so that $h + \sigma\delta \in \mathcal{K}$ and consider the two one-round losses $L\,|w-h|$ and $L\,|w-(h+\sigma\delta)|$. Pointwise,

$$L\,|W-h| + L\,|W-(h+\sigma\delta)| \ge L\delta$$

by the triangle inequality. Expectation preserves the inequality, and padding with zero losses gives the same construction for any horizon. An algorithm with zero consistency cost must therefore incur at least $L\delta$ on one of these nearby tasks. Combining this obstruction with the lower bound at a prescribed distance gives the following result.

Corollary 2 (Combined lower bound)**.** *For an algorithm A that does not receive ε or a closest optimizer, let $C_0(A)$ be its worst-case expected regret on n-round tasks with $\varepsilon = 0$. For every $\delta \in (0, D/2]$ there exists a task with distance δ such that*

$$\mathbb{E}[r(A, \mathcal{E}_\delta)] \ge \max\Big\{\frac{L\sqrt{n}}{8}\delta,\ L\delta - C_0(A)\Big\}.$$

Allowing a positive consistency budget permits movement away from h. Substituting that budget into the OGD potential gives a corresponding upper bound.

Proposition 1 (Tuned warm-start OGD)**.** *For every consistency budget $C_0 \in (0, DL\sqrt{n}/2]$, warm-start OGD with $\eta = 2C_0/(nL^2)$ attains*

$$r_n \le \frac{nL^2\varepsilon^2}{4C_0} + C_0.$$

At $\varepsilon = 0$ this is $r_n \le C_0$. Separately, playing h gives $r_n \le nL\varepsilon$. Selecting the better policy with known ε gives

$$\min\Big\{nL\varepsilon,\ \frac{nL^2\varepsilon^2}{4C_0} + C_0\Big\}.$$

Selecting its smaller branch requires ε; tuned OGD alone need not attain it.

At $C_0 = \Theta(L\sqrt{n}\,\varepsilon)$, tuned OGD matches Corollary 2 up to constants. This is the scale at which the bounds have the same order; away from it, the available upper and lower bounds need not coincide.

Strong convexity provides another source of stability. When each loss has known modulus $\mu > 0$, the accumulated objective becomes more strongly curved as observations arrive. Regularized follow-the-leader can use that curvature while keeping its initial regularizer centered at h.

Theorem 8 (Warm-start RFTL)**.** *Assume every ℓ_i is μ-strongly convex and μ is known. Set $w_1 = h$ and*

$$w_{i+1} \in \arg\min_{w \in \mathcal{K}} \Big\{ \frac{\mu}{2} \|w - h\|_2^2 + \sum_{s=1}^{i} \ell_s(w) \Big\}.$$

Then

$$r_n \le \frac{\mu}{2}\varepsilon^2 + \frac{2L^2}{\mu} H_n, \qquad H_n := \sum_{i=1}^{n} \frac{1}{i}.$$

The update does not use ε.

The standard RFTL decomposition separates the initial regularization cost from the stability of successive leaders (Hazan, 2016). Evaluating the regularizer at $w^\dagger$ gives $\mu\varepsilon^2/2$. Before the ith loss, the regularized objective is $i\mu$-strongly convex, which bounds the next leader's movement by $2L/(i\mu)$. Lipschitzness converts this into a charge of $2L^2/(i\mu)$, and summing gives the harmonic term. Thus μ determines the update, while the unknown distance appears only when evaluating the comparator in the proof.

The quadratic dependence is also necessary on a simple family where all uncertainty is concentrated in the first round.

Theorem 9 (Strongly convex lower bound at a prescribed distance)**.** *Assume $L \ge \mu D$. For every $\varepsilon \in [0, D/2]$ and every possibly randomized algorithm there exists a one-task sequence of μ-strongly convex L-Lipschitz losses on $[-D/2, D/2]$, with $h = 0$ and distance ε, such that $\mathbb{E}[r_n] \ge \mu\varepsilon^2/2$.*

Proof idea. Repeat $\ell_i(w) = \frac{\mu}{2}(w - \sigma\varepsilon)^2$ after sampling $\sigma = \pm 1$. The side condition $L \ge \mu D$ keeps the family L-Lipschitz. The first action is independent of σ, and its expected excess is at least $\mu\varepsilon^2/2$. Later rounds contribute a nonnegative amount because the same comparator minimizes every loss.

This family matches the distance-dependent term but does not force the harmonic baseline: after the first loss, its optimizer is known. The bounds therefore leave an additive gap. The RFTL upper bound improves on the convex oracle bound when $4LH_n/(\mu\sqrt{n}) \le \varepsilon \le L\sqrt{n}/\mu$. At very small distances, staying at h can instead give a smaller bound, $nL\varepsilon$. Corollary 4 and Theorem 10 derive these comparisons and a convex combination of the two policies. Choosing the smaller bound uses ε, whereas any fixed combination can be run without it.

7 Extensions: Repeated Tasks and Bandit Feedback

For T tasks with warm starts h_t and distances ε_t, restarting a one-task method at each h_t bounds $R_T = \sum_t r_t$ by summation. These upper bounds hold task by task, without independence assumptions or updates to the warm-start source.

Whether memory can improve this worst-case dependence requires a separate argument. For every target profile $(\bar\varepsilon_1, \dots, \bar\varepsilon_T) \in [0, D/2]^T$ and every algorithm, including one that retains arbitrary state across tasks, there are deterministic convex L-Lipschitz task sequences on $[-D/2, D/2]$ with distances exactly equal to those values such that

$$\mathbb{E}[R_T] \ge \frac{L\sqrt{n}}{8} \sum_{t=1}^{T} \bar\varepsilon_t.$$

Apply Theorem 3 independently on each task. The fresh signs remain unpredictable conditional on the past, giving $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}] \ge L\bar\varepsilon_t\sqrt{n}/8$. Summing and averaging fixes a deterministic collection with the displayed bound. Memory cannot help on this family because earlier tasks reveal nothing about future signs; useful transfer requires additional relationships between tasks.

Past performance can nevertheless guide a choice among supplied sources. If each task arrives with K candidate warm starts and full losses are revealed after every round, a Hedge master over their OGD trajectories, each run with the worst-case step size, attains

$$R_T \le \min_{k\in[K]} R_T^{(k)} + nLD\sqrt{2T\ln K}$$

(Proposition 8). Here $R_T^{(k)}$ is the regret of using source k throughout. Any two feasible trajectories differ by at most nLD in task loss, so subtracting the minimum puts the source-loss vector in $[0, nLD]^K$ even when static regrets are negative. The master learns source weights, rather than new parameter values. Propositions 6 and 10 give complementary concentration bounds for i.i.d. tasks and the independent lower-bound family.

Two-point bandit feedback with known distance. With function-value feedback, even staying close to h requires room for exploration. We use a two-point protocol: the learner incurs $\ell_i(w_i)$ at a center w_i and observes the same loss at $w_i \pm \delta u_i$. For a uniformly sampled unit direction u_i, the estimator is

$$\widehat{g}_i = \frac{d}{2\delta}\big(\ell_i(w_i+\delta u_i) - \ell_i(w_i-\delta u_i)\big)u_i, \qquad u_i \sim \mathrm{Unif}(\mathbb{S}^{d-1}).$$

A certified ball around h makes these symmetric queries feasible after shrinking the decision set. The next result sets both the query radius and the step size using the known distance.

Proposition 2 (Two-point bandit bound with known distance)**.** *Assume $d \ge 2$, $B(h,r) \subseteq \mathcal{K}$ with $r > 0$, and $\varepsilon \le r$. The losses are oblivious. Run projected OGD on the shrunken domain $\{w : B(w,\delta) \subseteq \mathcal{K}\}$ with $\delta = \varepsilon/\sqrt{n}$ and $\eta = \varepsilon/(L\sqrt{2dn})$, using the symmetric two-point estimator on the unit sphere. All queries lie in $\mathcal{K}$, and*

$$\mathbb{E}[r_n] \le (\sqrt{2d}+3)L\varepsilon\sqrt{n}.$$

If $\varepsilon = 0$, playing h gives $r_n = 0$.

The estimator is unbiased for the gradient of the ball-smoothed loss. Its second moment is at most $2dL^2$: the symmetric loss difference is an odd, $2L\delta$-Lipschitz function on the sphere, so the sphere Poincaré inequality controls its variance (Ledoux, 2001). Applying the OGD potential to these estimators gives the $\sqrt{2d}$ term. The comparator $h + (1-\delta/r)(w^\dagger - h)$ lies in the shrunken domain and is at most $\delta\varepsilon/r \le \delta$ from $w^\dagger$. Smoothing and this displacement together contribute at most $3Ln\delta = 3L\varepsilon\sqrt{n}$. Oblivious losses keep the hindsight comparator independent of the sampled directions.

Using only the almost-sure estimate $\|\widehat{g}_i\|_2 \le dL$ instead gives a factor $d+3$. That estimate also supports the reduction in Section B.10, which chooses its radius without ε. Both constructions still need the certified ball: at a boundary warm start, the first symmetric query may already leave $\mathcal{K}$.

8 Experiments

We examine how regret varies with realized distance and how that relationship changes when retrieval is corrupted. The 45 one-vs-one binary tasks from UCI Digits each contain 100 online rounds. Before each task, 10 labeled examples per class form a support set, disjoint from its online examples. Features are scaled to $[0,1]$ and normalized to unit norm; decisions lie in the radius-5 Euclidean ball. Each online loss is logistic loss plus $(\lambda/2)\|w\|_2^2$, with $\lambda = 10^{-2}$. The offline comparator minimizes the sum of these same regularized losses, so the reported distance and regret use one common objective. We approximate that optimum from three initializations, requiring a projected-gradient residual at most 10^{-8}; the distances are numerical estimates.

Retrieval averages up to five previous comparators, ranked by cosine similarity of class-mean difference descriptors computed only from the support sets; the first task starts at the origin. Thus the current task's descriptor is available before online adaptation, while its comparator is never used to choose its own warm start. To isolate the effect of initialization, no-warm-start OGD and warm-start OGD use the same worst-case step size $D/(L\sqrt{n})$ from the origin and h, respectively. Oracle-tuned

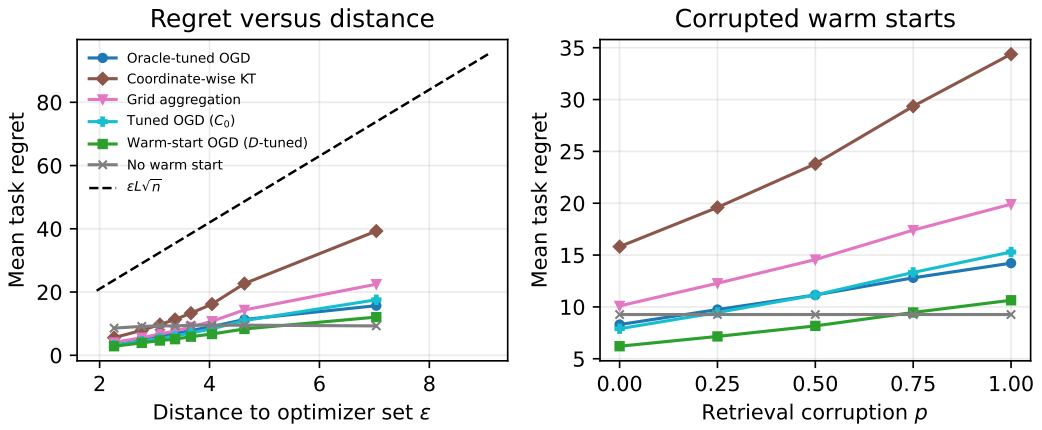


Figure 1: UCI Digits, 45 one-vs-one tasks and 30 seeds. A disjoint labeled support set determines each task descriptor before its online rounds. Left: mean task regret against realized distance; the dashed line is the OGD bound with known distance. Right: mean regret under retrieval corruption. Grid aggregation and oracle-tuned OGD track distance closely, while warm-start OGD with the worst-case step size is best at intact retrieval but falls behind the no-warm-start baseline at full corruption.

OGD uses $\varepsilon/(L\sqrt{n})$, tuned OGD uses $C_0 = DL\sqrt{n}/4$, and the grid follows Theorem 4. Coordinate-wise KT provides an empirical comparator-adaptive baseline, rather than an implementation of the Hilbert-space reduction. Each selected retrieval source is replaced by a random previous comparator with probability $p \in \{0, 0.25, 0.5, 0.75, 1\}$. We repeat the protocol with 30 fixed seeds and report seed-level bootstrap intervals.

The realized distance ranges from 1.95 to 9.07, with retrieval score and distance having Spearman correlation -0.848. In Figure 1, regret for grid aggregation and oracle-tuned OGD varies closely with distance. Yet warm-start OGD with the worst-case step size has the lowest mean regret at intact retrieval, 6.21 compared with 8.30 for oracle-tuned OGD and 10.10 for grid aggregation. Its larger step works better on these tasks even though the oracle step minimizes the worst-case potential. Distance-dependent guarantees therefore need not rank algorithms by their observed mean performance.

Corruption reverses the benefit of warm-start initialization: its mean regret rises from 6.21 at $p = 0$ to 10.64 at $p = 1$, with a 95% seed-level bootstrap interval of $[10.51, 10.78]$, while the no-warm-start mean remains 9.26. Because these two methods share a step size, this comparison isolates the cost of starting from an unreliable parameter. The experiment illustrates downstream sensitivity under a fixed retrieval rule; it does not estimate performance of learned retrieval on new task populations. Section B.21 gives shuffled and adversarial diagnostics, numerical results, and synthetic checks of the proof constructions.

9 Limitations and Conclusion

The bounds concern convex adaptation from an externally supplied parameter. They do not predict distance from retrieval scores or cover nonconvex losses. The Digits study uses labeled descriptors and the same regularized objective online and offline, while the bandit guarantees additionally require oblivious losses and certified interior geometry.

Open questions include the sharp price of unknown distance, a lower bound for the strongly convex harmonic term, and the consistency-sensitivity tradeoff under strong convexity. The parameter-free and strongly convex updates avoid ε but require ρ and μ. Overall, our results separate the value of a warm start from the cost of exploiting it without the distance.

AI USE STATEMENT

Generative AI tools assisted with drafting and copyediting, preparation of numerical verification code, and checks of cross-references and bibliographic metadata. The authors formulated the research questions, model, and results; independently checked every theorem statement and proof; reproduced every numerical constant reported in the paper; and verified each bibliographic entry against the publisher record. The authors take full responsibility for the manuscript and its claims.

ETHICS STATEMENT

This work involves no human-subject study or private data. The empirical study uses the public UCI Digits dataset. In applications such as recommendation, an unreliable warm start may worsen performance or inherit biases from its source. Our regret guarantees neither establish fairness nor justify deployment without checking the loss, feedback, and geometric assumptions in the intended setting.

REPRODUCIBILITY STATEMENT

The appendix provides complete proofs, explicit estimators and assumptions, and full experimental settings. The supplementary material specifies fixed seeds, preprocessing, solver tolerances, package versions, and the procedures used to regenerate every diagnostic.

REFERENCES

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the 23rd Annual Conference on Learning Theory*, pp. 28–40, 2010.

Maria-Florina Balcan, Mikhail Khodak, and Ameet Talwalkar. Provable guarantees for gradient-based meta-learning. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 424–433, 2019.

Jonathan Baxter. A model of inductive bias learning. *Journal of Artificial Intelligence Research*, 12: 149–198, 2000.

Aditya Bhaskara, Ashok Cutkosky, Ravi Kumar, and Manish Purohit. Online learning with imperfect hints. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 822–831, 2020a.

Aditya Bhaskara, Ashok Cutkosky, Ravi Kumar, and Manish Purohit. Online linear optimization with many hints. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9530–9539, 2020b.

Nicolò Cesa-Bianchi, Pierre Laforgue, Andrea Paudice, and Massimiliano Pontil. Multitask online mirror descent. *Transactions on Machine Learning Research*, 2022.

Niangjun Chen, Anish Agarwal, Adam Wierman, Siddharth Barman, and Lachlan L. H. Andrew. Online convex optimization using predictions. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 191–204, 2015.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Proceedings of the 25th Annual Conference on Learning Theory*, pp. 6.1–6.20, 2012.

Ashok Cutkosky and Kwabena Boahen. Online learning without prior information. In *Proceedings of the 2017 Conference on Learning Theory*, pp. 643–677, 2017.

Ashok Cutkosky and Francesco Orabona. Black-box reductions for parameter-free online learning in Banach spaces. In *Proceedings of the 31st Conference on Learning Theory*, pp. 1493–1529, 2018.

Ofer Dekel, Arthur Flajolet, Nika Haghtalab, and Patrick Jaillet. Online learning with a hint. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5299–5308, 2017.

Giulia Denevi, Carlo Ciliberto, Riccardo Grazzi, and Massimiliano Pontil. Learning-to-learn stochastic gradient descent with biased regularization. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 1566–1575, 2019.

Michael Dinitz, Sungjin Im, Thomas Lavastida, Benjamin Moseley, and Sergei Vassilvitskii. Faster matchings via learned duals. In *Advances in Neural Information Processing Systems*, volume 34, pp. 10393–10406, 2021.

Marek Eliáš, Haim Kaplan, Yishay Mansour, and Shay Moran. Learning-augmented algorithms with explicit predictors. In *Advances in Neural Information Processing Systems*, volume 37, pp. 97972–98008, 2024.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 1126–1135, 2017.

Chelsea Finn, Aravind Rajeswaran, Sham Kakade, and Sergey Levine. Online meta-learning. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 1920–1930, 2019.

Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 385–394, 2005.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2 (3–4):157–325, 2016.

Mikhail Khodak, Maria-Florina Balcan, and Ameet Talwalkar. Adaptive gradient-based meta-learning methods. In *Advances in Neural Information Processing Systems*, volume 32, pp. 5917–5928, 2019.

Mikhail Khodak, Maria-Florina Balcan, Ameet Talwalkar, and Sergei Vassilvitskii. Learning predictions for algorithms with predictions. In *Advances in Neural Information Processing Systems*, volume 35, pp. 3542–3555, 2022.

Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

Pengfei Li, Jianyi Yang, Adam Wierman, and Shaolei Ren. Robust learning for smoothed online convex optimization with feedback delay. In *Advances in Neural Information Processing Systems*, volume 36, pp. 16889–16916, 2023.

Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 3296–3305, 2018.

H. Brendan McMahan and Francesco Orabona. Unconstrained online linear learning in Hilbert spaces: Minimax algorithms and normal approximations. In *Proceedings of the 27th Conference on Learning Theory*, pp. 1020–1039, 2014.

Zakaria Mhammedi and Wouter M. Koolen. Lipschitz and comparator-norm adaptivity in online learning. In *Proceedings of the 33rd Conference on Learning Theory*, pp. 2858–2887, 2020.

Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. *Communications of the ACM*, 65(7):33–35, 2022.

Francesco Orabona and Dávid Pál. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems*, volume 29, pp. 577–585, 2016.

Anastasia Pentina and Christoph H. Lampert. A PAC-Bayesian bound for lifelong learning. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 991–999, 2014.

Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In *Advances in Neural Information Processing Systems*, volume 31, pp. 9661–9670, 2018.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, pp. 993–1019, 2013.

Shinsaku Sakaue and Taihei Oki. Discrete-convex-analysis-based framework for warm-starting algorithms with predictions. In *Advances in Neural Information Processing Systems*, volume 35, pp. 20988–21000, 2022.

Shinsaku Sakaue and Taihei Oki. Rethinking warm-starts with predictions: Learning predictions close to sets of optimal solutions for faster L-/L$^\natural$-convex function minimization. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 29760–29776, 2023.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

Matthew Streeter and H. Brendan McMahan. No-regret algorithms for unconstrained online convex optimization. In *Advances in Neural Information Processing Systems*, volume 25, pp. 2402–2410, 2012.

Alexander Wei and Fred Zhang. Optimal robustness-consistency trade-offs for learning-augmented online algorithms. In *Advances in Neural Information Processing Systems*, volume 33, pp. 8042–8053, 2020.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003.

A Notation and Reusable Proof Tools

Throughout the appendix we use the notation of the main text. The decision set is the compact convex set $\mathcal{K} \subset \mathbb{R}^d$ of diameter D. A one-task instance has horizon n, warm start $h \in \mathcal{K}$, convex L-Lipschitz losses ℓ_i, cumulative objective F, nonempty compact optimizer set $\mathcal{W}^\star$, closest optimizer $w^\dagger$, distance $\varepsilon = \operatorname{dist}(h, \mathcal{W}^\star)$, and static regret r_n. When T tasks are present we restore indices $t = 1, \dots, T$, write h_t, ε_t, r_t, and $R_T = \sum_t r_t$, and write $\mathcal{W}_t^\star$ and $w_t^\dagger$ for the corresponding optimizer objects. Euclidean projection onto a closed convex set C is denoted Π_C. All norms are Euclidean unless stated otherwise.

We repeatedly use three standard facts. (i) Convexity yields $\ell_i(w_i) - \ell_i(u) \le \langle g_i, w_i - u \rangle$ for $g_i \in \partial \ell_i(w_i)$. (ii) Euclidean projection is nonexpansive. (iii) Hedge over K experts with losses in an interval of length B has regret at most $B\sqrt{2n \ln K}$ (Freund & Schapire, 1997). The same upper bound holds under Hoeffding's lemma with learning rate $\sqrt{8 \ln K/(nB^2)}$, which in fact yields the stronger $B\sqrt{n \ln K/2}$. Khintchine's inequality is used in the form $\mathbb{E}[|S|] \ge \sqrt{m/2}$ for a sum of m independent Rademacher random variables.

The proofs below establish the benchmark with known distance, guarantees that do not use ε, and the lower bounds. Subsequent extensions treat feasible bandit queries, repeated tasks, and the experimental constructions.

B Proofs and Additional Results

B.1 Lower Bound at a Prescribed Distance

Full proof of Theorem 3. We prove the repeated-task extension at the same time; setting $T = 1$ gives the stated one-task result. Fix a target profile $\bar{\varepsilon}_1, \dots, \bar{\varepsilon}_T \in [0, D/2]$. Let

$$m := \begin{cases} n, & \text{if } n \text{ is odd,} \\ n-1, & \text{if } n \text{ is even.} \end{cases}$$

Then m is odd and $m \ge n/2$. For each task t, sample independent Rademacher variables $\xi_{t,1}, \dots, \xi_{t,m}$. Set the warm start to $h_t = 0$ and define the first m losses by

$$\ell_t^{(i)}(w) := \frac{L}{4}\,\xi_{t,i} w + \frac{3L}{4} \max\{0, |w| - \bar{\varepsilon}_t\}, \qquad i = 1, \dots, m.$$

If $m < n$, set the remaining $n - m$ losses to zero.

Step 1: Each loss is convex and L-Lipschitz. The map $w \mapsto \max\{0, |w| - \bar{\varepsilon}_t\}$ is convex and 1-Lipschitz. Hence each $\ell_t^{(i)}$ is convex. Its subgradients belong to

$$\frac{L}{4}\xi_{t,i} + \frac{3L}{4}\,\partial\big(\max\{0, |w| - \bar{\varepsilon}_t\}\big),$$

so every subgradient has magnitude at most $L/4 + 3L/4 = L$.

Step 2: The task optimum lies exactly at distance $\bar{\varepsilon}_t$ from the warm start. Let $S_t := \sum_{i=1}^m \xi_{t,i}$. Since m is odd, $S_t \neq 0$ deterministically. The cumulative loss on the nonzero rounds is

$$F_t(w) = \frac{L}{4} S_t w + \frac{3mL}{4} \max\{0, |w| - \bar{\varepsilon}_t\}.$$

Inside the interval $|w| \le \bar{\varepsilon}_t$, $F_t(w) = \frac{L}{4} S_t w$ is linear, so it is minimized at $w_t^\star = -\bar{\varepsilon}_t \operatorname{sign}(S_t)$. It remains to rule out minimizers outside that interval. If $w > \bar{\varepsilon}_t$, then

$$\frac{d}{dw} F_t(w) = \frac{L}{4} S_t + \frac{3mL}{4} \ge \frac{L}{4}(-m + 3m) = \frac{mL}{2} > 0.$$

Hence F_t is strictly increasing on $(\bar{\varepsilon}_t, \infty)$. Similarly, if $w < -\bar{\varepsilon}_t$, then

$$\frac{d}{dw} F_t(w) = \frac{L}{4} S_t - \frac{3mL}{4} \le \frac{L}{4}(m - 3m) = -\frac{mL}{2} < 0,$$

so F_t is strictly decreasing on $(-\infty, -\bar{\varepsilon}_t)$. Therefore the unique minimizer over $\mathcal{K} = [-D/2, D/2]$ is $w_t^\star$, and $\varepsilon_t = \operatorname{dist}(0, \mathcal{W}_t^\star) = |w_t^\star| = \bar{\varepsilon}_t$.

Step 3: Lower-bound the learner's regret. Let $w_{t,i}$ denote the learner's action on round i of task t. Conditioned on the task history up to round $i-1$, the action $w_{t,i}$ is measurable while $\xi_{t,i}$ is independent and mean zero, so $\mathbb{E}[\xi_{t,i} w_{t,i} \mid \mathcal{F}_{t,i-1}] = 0$. Taking expectations and summing, $\mathbb{E}\big[\sum_{i=1}^{m} \frac{L}{4} \xi_{t,i} w_{t,i}\big] = 0$. The barrier term is always nonnegative, hence $\mathbb{E}\Big[\sum_{i=1}^{n} \ell_t^{(i)}(w_{t,i})\Big] \geq 0$. The comparator $w_t^\star$ lies in $[-\bar{\varepsilon}_t, \bar{\varepsilon}_t]$, so the barrier term vanishes at $w_t^\star$. Therefore

$$\sum_{i=1}^{n} \ell_t^{(i)}(w_t^\star) = \frac{L}{4} S_t w_t^\star = -\frac{L\bar{\varepsilon}_t}{4} |S_t| .$$

Thus, conditioned on the past tasks, $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}] \geq \frac{L\bar{\varepsilon}_t}{4} \mathbb{E}[|S_t|]$. By Khintchine's inequality, $\mathbb{E}[|S_t|] \geq \sqrt{m/2}$. Since $m \geq n/2$, we have $m/2 \geq n/4$ and therefore $\sqrt{m/2} \geq \sqrt{n/4} = \sqrt{n}/2$. Hence $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}] \geq \frac{L\bar{\varepsilon}_t\sqrt{n}}{8}$. Taking expectations and summing over tasks gives $\mathbb{E}[R_T] \geq \frac{L\sqrt{n}}{8} \sum_{t=1}^{T} \bar{\varepsilon}_t$.

Step 4: Fix a deterministic task sequence. The displayed quantity is the expectation over the sampled Rademacher signs $\{\xi_{t,i}\}$. Hence some deterministic realization of all those signs yields at least the same expected regret against the learner's internal randomness. This proves the existential form of the theorem. □

B.2 Warm-Start OGD Proof

Proof of Theorem 1. Let $w^\dagger$ be a closest optimizer, so $\|h - w^\dagger\|_2 = \varepsilon$. By convexity of each loss,

$$\ell_i(w_i) - \ell_i(w^\dagger) \leq \langle g_i, w_i - w^\dagger \rangle.$$

The projected update is

$$w_{i+1} = \Pi_{\mathcal{K}}(w_i - \eta g_i).$$

Because Euclidean projection is nonexpansive and $w^\dagger \in \mathcal{K}$,

$$\begin{aligned} \|w_{i+1} - w^\dagger\|_2^2 &\leq \|w_i - \eta g_i - w^\dagger\|_2^2 \\ &= \|w_i - w^\dagger\|_2^2 - 2\eta \langle g_i, w_i - w^\dagger \rangle + \eta^2 \|g_i\|_2^2 . \end{aligned}$$

Rearranging gives

$$\langle g_i, w_i - w^\dagger \rangle \leq \frac{\|w_i - w^\dagger\|_2^2 - \|w_{i+1} - w^\dagger\|_2^2}{2\eta} + \frac{\eta}{2} \|g_i\|_2^2 .$$

Summing over $i = 1, \ldots, n$ telescopes:

$$\begin{aligned} r_n &= \sum_{i=1}^{n} \big(\ell_i(w_i) - \ell_i(w^\dagger)\big) \\ &\leq \sum_{i=1}^{n} \langle g_i, w_i - w^\dagger \rangle \\ &\leq \frac{\|w_1 - w^\dagger\|_2^2}{2\eta} + \frac{\eta}{2} \sum_{i=1}^{n} \|g_i\|_2^2 \\ &= \frac{\varepsilon^2}{2\eta} + \frac{\eta}{2} \sum_{i=1}^{n} \|g_i\|_2^2 . \end{aligned}$$

Using $\|g_i\|_2 \leq L$ yields

$$r_n \leq \frac{\varepsilon^2}{2\eta} + \frac{\eta n L^2}{2}.$$

For the robust tuning $\eta = D/(L\sqrt{n})$,

$$r_n \leq \frac{\varepsilon^2 L\sqrt{n}}{2D} + \frac{DL\sqrt{n}}{2}.$$

The right-hand side

$$\phi(\eta) \coloneqq \frac{\varepsilon^2}{2\eta} + \frac{\eta n L^2}{2}$$

is minimized over $\eta > 0$ at $\eta^\star = \varepsilon/(L\sqrt{n})$ when $\varepsilon > 0$, and has infimum 0 when $\varepsilon = 0$. In both cases,

$$\inf_{\eta>0} \phi(\eta) = \varepsilon L\sqrt{n},$$

which gives the oracle rate. □

B.3 OGD WITH AN ESTIMATED DISTANCE

Proof of Theorem 2. Substitute $\eta = q/(L\sqrt{n})$ into Theorem 1:

$$r_n \le \frac{\varepsilon^2}{2\eta} + \frac{\eta n L^2}{2} = \frac{\varepsilon^2 L\sqrt{n}}{2q} + \frac{qL\sqrt{n}}{2} = \frac{L\sqrt{n}}{2}\Big(\frac{\varepsilon^2}{q} + q\Big).$$

If $\varepsilon > 0$ and $q = \gamma\varepsilon$, then $\varepsilon^2/q + q = \varepsilon(\gamma + \gamma^{-1})$, so the bound equals $\varepsilon L\sqrt{n} \cdot (\gamma + \gamma^{-1})/2$. If $q \ge \varepsilon$, then $\varepsilon^2/q + q \le \varepsilon + q \le 2q$, hence $r_n \le qL\sqrt{n}$. If $q = \varepsilon$, the oracle value $\varepsilon L\sqrt{n}$ is recovered. If $\varepsilon = 0$, the bound is $qL\sqrt{n}/2 \ge 0$ and playing h yields $r_n = 0$. □

B.4 ORACLE MINIMAX CHARACTERIZATION

Proof of Corollary 1. The lower bound is exactly the one-task case of Theorem 3. The upper bound is the oracle tuning of Theorem 1: if $\varepsilon > 0$, the choice $\eta = \varepsilon/(L\sqrt{n})$ yields $r_n \le \varepsilon L\sqrt{n}$, and if $\varepsilon = 0$ playing h yields $r_n = 0$. □

B.5 EXPLICIT MULTISCALE ADAPTATION PROOF

Proof of Theorem 4. Fix a task and suppress the task index. Choose

$$w^\dagger \in \mathcal{W}^\star \qquad \text{with} \qquad \big\|h - w^\dagger\big\|_2 = \varepsilon.$$

For each $j = 0, \dots, J$, let expert j use the step size

$$\eta_j = \frac{D}{L\sqrt{n}\,2^j},$$

start from $w_1^{(j)} = h$, and update by

$$w_{i+1}^{(j)} \coloneqq \Pi_{\mathcal{K}}\Big(w_i^{(j)} - \eta_j g_i\Big),$$

where

$$g_i \in \partial \ell^{(i)}(\bar{w}_i)$$

is any subgradient observed at the master's prediction. The master maintains Hedge weights $p_i \in \Delta_{J+1}$ and plays

$$\bar{w}_i \coloneqq \sum_{j=0}^{J} p_i^{(j)} w_i^{(j)}.$$

Because $\mathcal{K}$ is convex, $\bar{w}_i \in \mathcal{K}$.

Step 1: Compare the master to the best surrogate expert. Define the shifted surrogate losses

$$z_i^{(j)} \coloneqq \langle g_i,\, w_i^{(j)} \rangle - \min_{0\le k\le J} \langle g_i,\, w_i^{(k)} \rangle.$$

Since all expert actions lie in $\mathcal{K}$ and $\|g_i\|_2 \le L$,

$$0 \le z_i^{(j)} \le LD.$$

Indeed, $z_i^{(j)}$ is nonnegative by definition, and

$$z_i^{(j)} \le \max_{0\le j'\le J} \langle g_i,\, w_i^{(j')} \rangle - \min_{0\le k\le J} \langle g_i,\, w_i^{(k)} \rangle \le L \max_{j',k} \Big\| w_i^{(j')} - w_i^{(k)} \Big\|_2 \le LD.$$

Applying Hedge with range $B = LD$ and the learning rate $\sqrt{8\ln(J+1)/(nD^2L^2)}$ of Theorem 4 (the Hoeffding choice) gives

$$\sum_{i=1}^{n}\sum_{j=0}^{J} p_i^{(j)} z_i^{(j)} \le \min_{0\le j\le J} \sum_{i=1}^{n} z_i^{(j)} + LD\sqrt{2n\ln(J+1)}.$$

Hoeffding's lemma yields the stronger overhead $LD\sqrt{n\ln(J+1)/2}$, which we record as the coarser $LD\sqrt{2n\ln(J+1)}$.

Expanding the shifted surrogate losses and canceling the common minima yields

$$\sum_{i=1}^{n}\sum_{j=0}^{J} p_i^{(j)} \langle g_i,\, w_i^{(j)}\rangle \le \min_{0\le j\le J} \sum_{i=1}^{n} \langle g_i,\, w_i^{(j)}\rangle + LD\sqrt{2n\ln(J+1)}.$$

By convexity of each loss at the point $\bar{w}_i$,

$$\ell^{(i)}(\bar{w}_i) - \ell^{(i)}(w^\dagger) \le \langle g_i,\, \bar{w}_i - w^\dagger\rangle = \sum_{j=0}^{J} p_i^{(j)} \langle g_i,\, w_i^{(j)}\rangle - \langle g_i,\, w^\dagger\rangle.$$

Combining the last two displays and summing over rounds gives

$$r \le \min_{0\le j\le J} \sum_{i=1}^{n} \langle g_i,\, w_i^{(j)} - w^\dagger\rangle + LD\sqrt{2n\ln(J+1)}.$$

Step 2: Plug in the expert guarantees. Each expert runs projected gradient descent on the linear losses $w \mapsto \langle g_i,\, w\rangle$. The standard OGD potential argument therefore gives

$$\sum_{i=1}^{n} \langle g_i,\, w_i^{(j)} - w^\dagger\rangle \le \frac{\varepsilon^2}{2\eta_j} + \frac{\eta_j n L^2}{2}.$$

Hence

$$r \le \min_{0\le j\le J} \left\{ \frac{\varepsilon^2}{2\eta_j} + \frac{\eta_j n L^2}{2} \right\} + LD\sqrt{2n\ln(J+1)}.$$

This proves the first display.

Step 3: Simplify the grid optimum. If $\varepsilon > D/n$, choose j so that

$$\eta_j \le \frac{\varepsilon}{L\sqrt{n}} < 2\eta_j.$$

Then

$$\frac{\varepsilon^2}{2\eta_j} \le \varepsilon L\sqrt{n}, \qquad \frac{\eta_j n L^2}{2} < \frac{\varepsilon L\sqrt{n}}{2},$$

so the expert term is at most $\frac{3}{2}\varepsilon L\sqrt{n}$.

If $\varepsilon \le D/n$, choose the smallest step size η_J. Since $J = \lceil \log_2 n \rceil$, we have

$$\frac{D}{2Ln^{3/2}} \le \eta_J \le \frac{D}{Ln^{3/2}}.$$

Thus

$$\frac{\varepsilon^2}{2\eta_J} \le \frac{(D/n)^2}{2} \cdot \frac{2Ln^{3/2}}{D} = \frac{DL}{\sqrt{n}}, \qquad \frac{\eta_J n L^2}{2} \le \frac{DL}{2\sqrt{n}}.$$

So in this case the expert term is at most $3DL/(2\sqrt{n})$.

Combining the two cases yields the uniform simplification

$$\min_{0\le j\le J} \left\{ \frac{\varepsilon^2}{2\eta_j} + \frac{\eta_j n L^2}{2} \right\} \le \frac{3}{2}\varepsilon L\sqrt{n} + \frac{3DL}{2\sqrt{n}}.$$

Substituting into the first display proves the second bound. □

B.6 Lower Bound for Pure Expert-Advice Aggregation

Proposition 3 (Black-box aggregation lower bound)**.** *Let $K \geq 2$ and $B > 0$, and consider the standard expert-advice protocol with K experts and losses in $[0, B]$. Then for every possibly randomized master algorithm and every horizon $n \geq \lfloor \log_2 K \rfloor$, there exists a deterministic loss sequence such that the master suffers expected regret at least*

$$\frac{B}{4}\sqrt{n\lfloor \log_2 K \rfloor}$$

against the best expert in hindsight.

Proof. Let

$$m \coloneqq \lfloor \log_2 K \rfloor, \qquad s \coloneqq \left\lfloor \frac{n}{m} \right\rfloor.$$

Since $n \geq m$, we have $s \geq 1$. Ignore all but the first 2^m experts and index them by sign vectors $b = (b_1, \ldots, b_m) \in \{-1, +1\}^m$. For any remaining experts, simply duplicate the losses of the expert $b = (+1, \ldots, +1)$; this can only decrease the benchmark "best expert" loss.

Partition the first ms rounds into m blocks of length s. For each block $r \in [m]$ and each round $q \in [s]$ in that block, sample an independent Rademacher random variable $\xi_{r,q}$ and define the loss of expert b on that round by

$$z_{r,q}^{(b)} \coloneqq \frac{B}{2}\big(1 - b_r \xi_{r,q}\big).$$

For any remaining rounds, set all expert losses to zero.

Step 1: The master's expected loss. Fix a round (r, q) and let $p_{r,q}$ be the master's distribution over experts before observing $\xi_{r,q}$. Then

$$\begin{aligned}\mathbb{E}_{\xi_{r,q}}\left[\sum_b p_{r,q}(b) z_{r,q}^{(b)} \,\middle|\, p_{r,q}\right] &= \frac{B}{2}\left(1 - \sum_b p_{r,q}(b) b_r \,\mathbb{E}[\xi_{r,q}]\right) \\ &= \frac{B}{2}.\end{aligned}$$

Summing over the ms nonzero rounds shows that the master's expected loss is $Bms/2$.

Step 2: The best expert in hindsight. For each block r, define

$$S_r \coloneqq \sum_{q=1}^{s} \xi_{r,q}.$$

The cumulative loss of expert b over the nonzero rounds is

$$\frac{Bms}{2} - \frac{B}{2}\sum_{r=1}^{m} b_r S_r.$$

Because the experts include all sign vectors in $\{-1, +1\}^m$, the best expert chooses $b_r = \operatorname{sign}(S_r)$ for each block and achieves

$$\frac{Bms}{2} - \frac{B}{2}\sum_{r=1}^{m} |S_r|.$$

Hence the master's expected regret under this random construction is

$$\frac{B}{2}\sum_{r=1}^{m} \mathbb{E}[|S_r|].$$

Step 3: Lower-bound the block fluctuations. Each S_r is a sum of s i.i.d. Rademacher variables, so Khintchine's inequality gives

$$\mathbb{E}[|S_r|] \geq \sqrt{\frac{s}{2}}.$$

Therefore

$$\mathbb{E}[\text{Regret}] \geq \frac{Bm}{2}\sqrt{\frac{s}{2}}.$$

Since $s = \lfloor n/m \rfloor \geq n/(2m)$,

$$\sqrt{\frac{s}{2}} \geq \sqrt{\frac{n}{4m}}.$$

Substituting,

$$\mathbb{E}[\text{Regret}] \geq \frac{B}{4}\sqrt{nm}.$$

Step 4: Fix a deterministic sequence. The displayed quantity is the expectation over the random signs $\{\xi_{r,q}\}$. Hence some deterministic realization of those signs yields at least the same expected regret against the master's internal randomness. This proves the proposition. □

Corollary 3 (Unrestricted black-box aggregation)**.** *Suppose an expert-advice master must handle every sequence of loss vectors in $[0, B]^K$, without restrictions on how expert losses are generated. For every horizon $n \geq \lfloor \log_2 K \rfloor$, the worst-case additive overhead over the best expert is $\Omega(B\sqrt{n \log K})$.*

Proof. This is exactly Proposition 3. □

The unrestricted loss-vector assumption is essential. OGD experts driven by a shared gradient generate structured loss vectors, and the construction above need not be realizable by those experts. It therefore does not establish a matching lower bound for Theorem 4 or for adaptation to the unknown distance.

B.7 Translated Parameter-Free Adaptation Proof

Lemma 1 (Imported unconstrained KT bound)**.** *Let $\{g_i\}_{i=1}^n$ be vectors in a Hilbert space with $\|g_i\|_2 \leq 1$, and let $\rho > 0$ be the Krichevsky-Trofimov initial wealth. The Hilbert-space algorithm of Orabona & Pál (2016, Cor. 5) produces iterates z_i satisfying, for every comparator u,*

$$\sum_{i=1}^{n} \langle g_i,\, z_i - u \rangle \leq \|u\|_2 \sqrt{n \ln\Big(1 + \frac{24n^2\,\|u\|_2^2}{\rho^2}\Big)} + \rho\Big(1 - \frac{1}{e\sqrt{\pi n}}\Big).$$

Proof. This is Orabona & Pál (2016, Cor. 5) with initial endowment ρ on unit-ball rewards (equivalently, losses). The displayed constants 24 and $e\sqrt{\pi n}$ are those of that corollary. □

Lemma 2 (Constrained doubling)**.** *Let $W \subset V$ be closed and convex in a reflexive Banach space. If an unconstrained algorithm A on V has linearized regret $R_A(u)$, the reduction of Cutkosky & Orabona (2018, Thm. 3) produces feasible iterates in W whose linearized regret is at most $2R_A(u)$ for every $u \in W$. The reduction feeds modified subgradients of the Cutkosky surrogate rather than the original subgradients; their dual norms are no larger than those of the original subgradients, so they remain admissible for Lemma 1, whose bound holds for every unit-ball sequence.*

Proof of Theorem 5. Translate the domain to $\mathcal{K}' := \mathcal{K} - h$, which is convex, compact, and contains the origin. Centered losses $\tilde{\ell}_i(u) = \ell_i(u+h)$ have the same subgradient norms. The closest optimizer maps to $u^\dagger = w^\dagger - h$ with $\|u^\dagger\|_2 = \varepsilon$, and regret is preserved.

Let a_i be the modified subgradient supplied by the constrained reduction from the observed g_i. By Lemma 2, $\|a_i\|_2 \leq \|g_i\|_2 \leq L$. Applying Lemma 1 to a_i/L and rescaling gives

$$\sum_{i=1}^{n} \langle a_i,\, z_i - u^\dagger \rangle \leq L\varepsilon \sqrt{n \ln\Big(1 + \frac{24n^2\varepsilon^2}{\rho^2}\Big)} + L\rho\Big(1 - \frac{1}{e\sqrt{\pi n}}\Big).$$

Lemma 2 bounds the feasible linearized regret by twice this upper bound and keeps iterates in $\mathcal{K}'$. Convexity then bounds convex-loss regret by linearized regret, proving Theorem 5. The construction uses ρ, L, and n, but not ε. □

B.8 Impossibility of Uniform Adaptation

Proof of Theorem 6. Fix $d = 1$ and $\mathcal{K} = [-D/2, D/2]$. Suppose an ε-blind A and finite C satisfy the proposed bound on every admissible task and horizon. Let $m \geq 1$, let $u \in [-D/4, D/4] \setminus \{0\}$, and let $g_1, \ldots, g_m \in [-L/4, L/4]$. Define a $2m$-round task by $\ell_i(w) = g_i w$ for $i \leq m$ and $\ell_{m+j}(w) = L\,|w - u|$ for $j = 1, \ldots, m$. Each loss is convex and L-Lipschitz. Writing $S = \sum_{i=1}^{m} g_i$, the cumulative objective is $F(w) = Sw + mL\,|w - u|$. Because $|S| \leq mL/4$, one has $S - mL \leq -3mL/4 < 0$ on $(-\infty, u)$ and $S + mL \geq 3mL/4 > 0$ on (u, ∞), so u is the unique minimizer and $\varepsilon = |u|$.

The first m losses do not depend on u. Since A is ε-blind, its first m actions are also independent of u and define an online linear algorithm. Task regret is at least $\sum_{i=1}^{m} g_i(w_i - u)$, because the remaining losses vanish at u and are nonnegative elsewhere. The assumed bound with horizon $n = 2m$ therefore gives

$$\mathbb{E}\Big[\sum_{i=1}^{m} g_i(w_i - u)\Big] \leq C\,|u|\,L\sqrt{2m}.$$

Write $A_m(g) := \mathbb{E}_{\mathrm{alg}}[\sum_{i=1}^{m} g_i w_i]$ for a deterministic gradient sequence g. Then $A_m(g) - Su \leq C\,|u|\,L\sqrt{2m}$. For each fixed g, choose the sign of u so that $Su \leq 0$ and send $|u| \to 0$ through $(0, D/4]$. The right-hand side vanishes and A_m does not depend on u, so $A_m(g) \leq 0$.

Now let $g_i = \xi_i L/4$ with independent Rademacher signs ξ_i. Predictability gives $\mathbb{E}_g A_m(g) = 0$. This expectation is a finite average of nonpositive terms, hence $A_m(g) = 0$ for every sign sequence. Taking the constant sequence $g_i = L/4$ for all i and $u = -D/4$ yields

$$\mathbb{E}\Big[\sum_{i=1}^{m} g_i(w_i - u)\Big] = -u \cdot m \cdot \frac{L}{4} = \frac{mLD}{16}.$$

The assumed bound requires this to be at most $C(D/4)L\sqrt{2m}$, which fails for large m. This contradiction proves Theorem 6. The argument does not produce a logarithmic lower bound and is not claimed to match Theorem 5. □

B.9 Bandit Feedback Extension

The linear dependence on the known distance extends to bandit convex optimization with two-point feedback at the standard $\sqrt{n}$ rate and a $\sqrt{d}$ dimensional factor. The key extra issue is query feasibility: both $w_{t,i} + \delta u_{t,i}$ and $w_{t,i} - \delta u_{t,i}$ must remain inside $\mathcal{K}$.

Lemma 3 (Warm-start interior shrinkage)**.** *Let $\mathcal{K} \subset \mathbb{R}^d$ be convex, let $h \in \mathcal{K}$, and assume $B(h, r) \subseteq \mathcal{K}$ for some $r > 0$. For any $\theta \in \mathcal{K}$ and any $\delta \in [0, r]$, define*

$$\theta_\delta := h + \left(1 - \frac{\delta}{r}\right)(\theta - h).$$

Then

$$\theta_\delta \in \mathcal{K}_\delta := \{w \in \mathcal{K} : B(w, \delta) \subseteq \mathcal{K}\}.$$

Moreover,

$$\|\theta_\delta - h\|_2 \leq \|\theta - h\|_2, \qquad \|\theta_\delta - \theta\|_2 = \frac{\delta}{r}\|\theta - h\|_2.$$

Proof. If $\delta = 0$, the claim is immediate. Assume $\delta > 0$. Let $z \in \mathbb{R}^d$ satisfy $\|z\|_2 \leq \delta$. Then

$$\theta_\delta + z = \left(1 - \frac{\delta}{r}\right)\theta + \frac{\delta}{r}\left(h + \frac{r}{\delta}z\right).$$

Because $\|rz/\delta\|_2 \leq r$, the point $h + (r/\delta)z$ belongs to $B(h, r) \subseteq \mathcal{K}$. Also $\theta \in \mathcal{K}$. By convexity of $\mathcal{K}$, the convex combination $\theta_\delta + z$ belongs to $\mathcal{K}$. Since this holds for every $\|z\|_2 \leq \delta$, we have $B(\theta_\delta, \delta) \subseteq \mathcal{K}$, so $\theta_\delta \in \mathcal{K}_\delta$. The norm identities follow directly from the definition of θ_δ. □

Proposition 4 (Two-point bandit warm-start OGD with feasible queries). *Assume $\mathcal{K} \subset \mathbb{R}^d$, $d \geq 2$, is convex and compact, and that the losses are oblivious. For task t, assume $B(h_t, r) \subseteq \mathcal{K}$ with $r > 0$ and $\varepsilon_t \leq r$. Let $w_t^\dagger \in \mathcal{W}_t^\star$ be a closest offline optimizer, so $\|h_t - w_t^\dagger\|_2 = \varepsilon_t$. For $\varepsilon_t > 0$, set*

$$\delta_t = \frac{\varepsilon_t}{\sqrt{n}}, \qquad \eta_t = \frac{\varepsilon_t}{dL\sqrt{n}}.$$

Run projected OGD on

$$\mathcal{K}_{\delta_t} := \{w \in \mathcal{K} : B(w, \delta_t) \subseteq \mathcal{K}\},$$

initialized at h_t. On round i, sample $u_{t,i} \sim \mathrm{Unif}(S^{d-1})$, query

$$\ell_t^{(i)}(w_{t,i} + \delta_t u_{t,i}) \quad \textit{and} \quad \ell_t^{(i)}(w_{t,i} - \delta_t u_{t,i}),$$

and use the estimator

$$\widehat{g}_{t,i} = \frac{d}{2\delta_t}\left[\ell_t^{(i)}(w_{t,i} + \delta_t u_{t,i}) - \ell_t^{(i)}(w_{t,i} - \delta_t u_{t,i})\right] u_{t,i}.$$

Then all queries lie in $\mathcal{K}$, and

$$\mathbb{E}[r_t] \leq (d+3)L\varepsilon_t\sqrt{n} = O(dL\varepsilon_t\sqrt{n}).$$

If $\varepsilon_t = 0$, playing h_t deterministically gives $r_t = 0$.

Proof. Because $\varepsilon_t \leq r$, we have $\delta_t = \varepsilon_t/\sqrt{n} \leq r$. Since $B(h_t, r) \subseteq \mathcal{K}$, we have $h_t \in \mathcal{K}_{\delta_t}$. The algorithm projects every iterate onto $\mathcal{K}_{\delta_t}$, so

$$B(w_{t,i}, \delta_t) \subseteq \mathcal{K}.$$

Therefore both query points $w_{t,i} \pm \delta_t u_{t,i}$ lie in $\mathcal{K}$.

Define the shrunken comparator

$$\theta_{t,\delta}^\dagger = h_t + \left(1 - \frac{\delta_t}{r}\right)(w_t^\dagger - h_t).$$

By Lemma 3,

$$\theta_{t,\delta}^\dagger \in \mathcal{K}_{\delta_t}, \qquad \|h_t - \theta_{t,\delta}^\dagger\|_2 \leq \varepsilon_t,$$

and

$$\|\theta_{t,\delta}^\dagger - w_t^\dagger\|_2 = \frac{\delta_t}{r}\varepsilon_t \leq \delta_t,$$

where the last inequality uses $\varepsilon_t \leq r$.

For each loss, define the ball-smoothed loss

$$\widehat{\ell}_t^{(i)}(w) := \mathbb{E}_{v \sim \mathrm{Unif}(B^d)}\left[\ell_t^{(i)}(w + \delta_t v)\right].$$

The standard two-point estimator of Flaxman et al. (2005); Shamir (2017) is unbiased for the gradient of this ball-smoothed loss:

$$\mathbb{E}[\widehat{g}_{t,i} \mid \mathcal{F}_{t,i-1}] = \nabla\widehat{\ell}_t^{(i)}(w_{t,i}),$$

and, since each $\ell_t^{(i)}$ is L-Lipschitz,

$$\|\widehat{g}_{t,i}\|_2 \leq dL.$$

The OGD potential inequality on $\mathcal{K}_{\delta_t}$, with comparator $\theta_{t,\delta}^\dagger$, gives

$$\sum_{i=1}^n \langle \widehat{g}_{t,i}, w_{t,i} - \theta_{t,\delta}^\dagger\rangle \leq \frac{\|h_t - \theta_{t,\delta}^\dagger\|_2^2}{2\eta_t} + \frac{\eta_t}{2}\sum_{i=1}^n \|\widehat{g}_{t,i}\|_2^2.$$

Because the losses are oblivious, $\theta_{t,\delta}^\dagger$ is a function of $(\ell_t^{(1)}, \ldots, \ell_t^{(n)})$ and is independent of the sphere samples. Conditioning on $w_{t,i}$ and the loss sequence therefore gives

$$\mathbb{E}\left[\langle \widehat{g}_{t,i}, w_{t,i} - \theta_{t,\delta}^\dagger\rangle \;\middle|\; w_{t,i}, \ell_t^{(\cdot)}\right] = \langle \nabla\widehat{\ell}_t^{(i)}(w_{t,i}), w_{t,i} - \theta_{t,\delta}^\dagger\rangle.$$

Convexity of each smoothed loss upgrades the inner product to a function-value gap. Taking expectations and using $\|\widehat{g}_{t,i}\|_2 \le dL$ yields

$$\mathbb{E}\left[\sum_{i=1}^{n} \widehat{\ell}_t^{(i)}(w_{t,i}) - \widehat{\ell}_t^{(i)}(\theta_{t,\delta}^\dagger)\right] \le \frac{\varepsilon_t^2}{2\eta_t} + \frac{\eta_t n d^2 L^2}{2} = dL\varepsilon_t\sqrt{n}.$$

It remains to convert smoothed regret against $\theta_{t,\delta}^\dagger$ into original regret against $w_t^\dagger$. Lipschitzness gives

$$|\widehat{\ell}_t^{(i)}(w) - \ell_t^{(i)}(w)| \le L\delta_t$$

for every feasible w. Therefore

$$\begin{aligned}\mathbb{E}[r_t] &= \mathbb{E}\left[\sum_{i=1}^{n} \ell_t^{(i)}(w_{t,i}) - \ell_t^{(i)}(w_t^\dagger)\right] \\ &\le dL\varepsilon_t\sqrt{n} + 2L\delta_t n + Ln\|\theta_{t,\delta}^\dagger - w_t^\dagger\|_2 \\ &\le dL\varepsilon_t\sqrt{n} + 2L\varepsilon_t\sqrt{n} + L\varepsilon_t\sqrt{n} \\ &= (d+3)L\varepsilon_t\sqrt{n}.\end{aligned}$$

This proves the claim. □

Discussion. Proposition 4 preserves the linear dependence on ε_t from Theorem 1 at the same $\sqrt{n}$ rate, with a dimension penalty d from the almost-sure bound on the two-point estimator. We sharpen this to $\sqrt{d}$ in Proposition 2 below using a second-moment bound. The smoothing and estimation template is standard in bandit convex optimization (Flaxman et al., 2005; Agarwal et al., 2010; Shamir, 2017). A one-point analogue is not claimed here; one-point estimators do not cancel the function-value offset of a two-point difference and require a separate analysis.

Lemma 4 (Sphere second-moment bound)**.** *Let $d \ge 2$. For every convex L-Lipschitz loss ℓ, every $\delta > 0$, every feasible center w, and $u \sim \mathrm{Unif}(S^{d-1})$, the two-point estimator*

$$\widehat{g} = \frac{d}{2\delta}\left[\ell(w+\delta u) - \ell(w-\delta u)\right] u$$

satisfies

$$\mathbb{E}\big[\|\widehat{g}\|_2^2 \mid w\big] \le 2dL^2.$$

Proof. Define

$$F(u) := \ell(w+\delta u) - \ell(w-\delta u), \qquad u \in S^{d-1}.$$

Then $F(-u) = -F(u)$, so $\mathbb{E}_{u\sim S^{d-1}} F(u) = 0$. Moreover, because ℓ is L-Lipschitz on the feasible query region,

$$|F(u) - F(v)| \le L\delta\|u-v\|_2 + L\delta\|u-v\|_2 = 2L\delta\|u-v\|_2.$$

Thus F is $2L\delta$-Lipschitz on the sphere. By the Poincaré inequality on S^{d-1} (see Ledoux, 2001, Sec. 2.3),

$$\mathbb{E}F(u)^2 = \mathrm{Var}(F(u)) \le \frac{4L^2\delta^2}{d-1}.$$

Therefore

$$\mathbb{E}\|\widehat{g}\|_2^2 = \frac{d^2}{4\delta^2}\mathbb{E}F(u)^2 \le \frac{d^2}{d-1}L^2 \le 2dL^2,$$

where the last inequality uses $d \ge 2$. □

Proof of Proposition 2. The main-text algorithm uses the feasible set and smoothing radius of Proposition 4 with

$$\delta_t = \frac{\varepsilon_t}{\sqrt{n}}, \qquad \eta_t = \frac{\varepsilon_t}{L\sqrt{2dn}}.$$

The feasibility and shrunken-comparator arguments are identical to the proof of Proposition 4. The only change is the use of Lemma 4. Thus

$$\mathbb{E}\left[\sum_{i=1}^{n} \widehat{\ell}_t^{(i)}(w_{t,i}) - \widehat{\ell}_t^{(i)}(\theta_{t,\delta}^{\dagger})\right] \le \frac{\varepsilon_t^2}{2\eta_t} + \frac{\eta_t}{2}\sum_{i=1}^{n}\mathbb{E}\|\widehat{g}_{t,i}\|_2^2$$

$$\le \frac{\varepsilon_t^2}{2\eta_t} + \eta_t dnL^2 = \sqrt{2d}\, L\varepsilon_t\sqrt{n}.$$

The smoothing and comparator-shrinkage conversion adds at most $3L\varepsilon_t\sqrt{n}$, exactly as in Proposition 4. Therefore

$$\mathbb{E}[r_t] \le (\sqrt{2d}+3)L\varepsilon_t\sqrt{n}.$$

□

B.10 Parameter-Free Bandit Adaptation

The methods of Propositions 2 and 4 use ε to set both the step size and the smoothing radius. Choosing the radius from the certified ball instead allows us to apply Lemmas 1 and 2 without distance information. This reduction uses the almost-sure estimator bound dL, so its leading term has a factor d rather than the second-moment factor $\sqrt{d}$.

Proposition 5 (Parameter-free two-point bandit, explicit)**.** *Assume $d \ge 2$, $\mathcal{K} \subset \mathbb{R}^d$ convex and compact, and $B(h,r) \subseteq \mathcal{K}$ with $r > 0$. The losses are oblivious. Fix $\rho > 0$ before the task and set $\delta = r/n$. Run the translated Krichevsky-Trofimov method of Lemmas 1 and 2 on the centered shrunken domain $\mathcal{K}_\delta - h$, feeding the two-point estimator*

$$\widehat{g}_i = \frac{d}{2\delta}\big(\ell_i(w_i+\delta s_i) - \ell_i(w_i-\delta s_i)\big)s_i, \qquad s_i \sim \mathrm{Unif}(S^{d-1}).$$

All queries lie in $\mathcal{K}$. With $G := dL$,

$$\mathbb{E}[r_n] \le 2G\varepsilon\sqrt{n\ln\Big(1+\frac{24n^2\varepsilon^2}{\rho^2}\Big)} + 2G\rho\Big(1-\frac{1}{e\sqrt{\pi n}}\Big) + 2Lr + L\varepsilon.$$

Equivalently, the leading term is $2dL\varepsilon\sqrt{n\ln(1+24n^2\varepsilon^2/\rho^2)}$.

Proof. Because $\delta \le r$ and $B(h,r) \subseteq \mathcal{K}$, one has $h \in \mathcal{K}_\delta$. The algorithm projects onto $\mathcal{K}_\delta$, so both query points lie in $\mathcal{K}$. Write $G := dL$. Let $\widehat{\ell}_i$ denote the δ-smoothed loss. The estimator satisfies $\|\widehat{g}_i\|_2 \le G$ almost surely and is unbiased for $\nabla\widehat{\ell}_i(w_i)$. Because the losses are oblivious, the hindsight shrunken comparator $u_\delta := h + (1-\delta/r)(w^\dagger - h)$ is independent of the sphere samples, so

$$\mathbb{E}\big[\langle \widehat{g}_i, w_i - u_\delta\rangle \;\big|\; w_i, \ell_{1:n}\big] = \langle \nabla\widehat{\ell}_i(w_i), w_i - u_\delta\rangle.$$

Convexity of each smoothed loss converts this identity into an upper bound on smoothed regret against u_δ. The KT guarantee applies pathwise using the almost-sure bound G.

Translate to $\mathcal{K}_\delta - h$. Lemma 2 sends modified subgradients $\tilde{g}$ of the Cutkosky surrogate to the unconstrained method; $\|\tilde{g}\|_* \le \|\widehat{g}_i\|_2 \le G$. Apply Lemma 1 pathwise to $\tilde{g}/G$, which is valid because that bound holds for every unit-ball sequence, then scale by G and double. This produces the first two terms against u_δ, whose distance from h is at most ε. The algorithm's Euclidean projection onto $\mathcal{K}_\delta$ is the reduction's feasible map Π_W. Lipschitz conversion from smoothed losses to original losses costs at most $2L\delta n = 2Lr$. Comparator shrinkage costs at most $Ln\cdot(\delta/r)\varepsilon = L\varepsilon$. Adding these terms yields the claim. The construction uses d, L, n, r, ρ and does not use ε in the algorithm. □

B.11 Exact-Consistency Lower Bound Proof

Proof of Theorem 7. Let ξ denote the internal randomness of the algorithm. Before any loss is observed, the algorithm's first action is a measurable function $W(\xi, h) \in \mathcal{K}$ of the warm start and its random seed.

Fix $\delta \in (0, D/2]$. Since $\mathcal{K} = [-D/2, D/2]$ has diameter D, at least one of $h+\delta$ or $h-\delta$ lies in $\mathcal{K}$. Choose a sign $\sigma \in \{-1,+1\}$ such that $h + \sigma\delta \in \mathcal{K}$.

Consider the following one-round environments:

$$\mathcal{E}_0 : \ell_0(w) = L\,|w - h|\,, \qquad \mathcal{E}_\delta : \ell_\delta(w) = L\,|w - (h + \sigma\delta)|\,.$$

Both losses are convex and L-Lipschitz on $\mathcal{K}$. The unique minimizers are h and $h + \sigma\delta$, respectively, so the first environment has warm-start distance 0 and the second has warm-start distance δ. If one wants $n > 1$ rounds, simply set all remaining losses to zero.

For a fixed realization of ξ, the regrets on the first round are

$$r(\mathcal{A}, \mathcal{E}_0; \xi) = L\,|W(\xi, h) - h|\,, \qquad r(\mathcal{A}, \mathcal{E}_\delta; \xi) = L\,|W(\xi, h) - h - \sigma\delta|\,.$$

By the triangle inequality,

$$r(\mathcal{A}, \mathcal{E}_0; \xi) + r(\mathcal{A}, \mathcal{E}_\delta; \xi) \geq L\,|\sigma\delta| = L\delta.$$

Taking expectation over ξ proves

$$\mathbb{E}[r(\mathcal{A}, \mathcal{E}_0)] + \mathbb{E}[r(\mathcal{A}, \mathcal{E}_\delta)] \geq L\delta.$$

□

B.12 Constructive Frontier Algorithm

Proof of Proposition 1. Substitute $\eta = 2C_0/(nL^2)$ into Theorem 1's bound $r_n \leq \varepsilon^2/(2\eta) + \eta n L^2/2$:

$$\frac{\varepsilon^2}{2\eta} = \frac{\varepsilon^2 \cdot nL^2}{4C_0}, \qquad \frac{\eta n L^2}{2} = C_0.$$

Playing h on every round gives $r_n = F(h) - F(w^\dagger) \leq nL\varepsilon$ since F is nL-Lipschitz on $\mathcal{K}$. These are guarantees for two different policies. An oracle can choose the policy with the smaller displayed bound using ε, but their minimum is not a simultaneous guarantee for tuned OGD. □

Comparison to the lower bound. At $C_0 = \Theta(L\sqrt{n}\,\varepsilon)$, the tuned OGD upper bound has order $L\varepsilon\sqrt{n}$, matching Corollary 2. This corresponds to $\eta = \Theta(\varepsilon/(L\sqrt{n}))$. For smaller consistency budgets, the term $nL^2\varepsilon^2/(4C_0)$ can substantially exceed the combined lower bound. A sharp characterization across all budgets remains open.

B.13 Consistency-Sensitivity Tradeoff

Proof of Corollary 2. Fix $\delta \in (0, D/2]$.

First, apply Theorem 3 with target distance $\bar{\varepsilon} = \delta$. This yields a one-task environment $\mathcal{E}_\delta^{(1)}$ with warm-start distance δ such that

$$\mathbb{E}[r(\mathcal{A}, \mathcal{E}_\delta^{(1)})] \geq \frac{L\sqrt{n}}{8}\delta.$$

Next, apply Theorem 7. There exist an exact-warm-start environment $\mathcal{E}_0$ and an environment at distance δ, $\mathcal{E}_\delta^{(2)}$ such that

$$\mathbb{E}[r(\mathcal{A}, \mathcal{E}_0)] + \mathbb{E}[r(\mathcal{A}, \mathcal{E}_\delta^{(2)})] \geq L\delta.$$

By definition of $C_0(\mathcal{A})$,

$$\mathbb{E}[r(\mathcal{A}, \mathcal{E}_0)] \leq C_0(\mathcal{A}),$$

so

$$\mathbb{E}[r(\mathcal{A}, \mathcal{E}_\delta^{(2)})] \geq L\delta - C_0(\mathcal{A}).$$

Now choose whichever of the two distance-δ environments, $\mathcal{E}_\delta^{(1)}$ or $\mathcal{E}_\delta^{(2)}$, gives larger regret. That environment satisfies

$$\mathbb{E}[r(\mathcal{A}, \mathcal{E}_\delta)] \geq \max\left\{\frac{L\sqrt{n}}{8}\delta, L\delta - C_0(\mathcal{A})\right\}.$$

□

B.14 Strongly Convex Faster-Rate Proof

Proof of Theorem 8. Fix a task t and choose $w_t^\dagger \in \mathcal{W}_t^\star$ with

$$\left\| h_t - w_t^\dagger \right\|_2 = \varepsilon_t.$$

Write $h = h_t$ and $w^\dagger = w_t^\dagger$. For $i = 0, 1, \ldots, n$, define

$$F_i(w) \coloneqq \frac{\mu}{2} \|w - h\|_2^2 + \sum_{s=1}^{i} \ell^{(s)}(w), \qquad w \in \mathcal{K}.$$

By definition, w_{i+1} minimizes F_i over $\mathcal{K}$, and $w_1 = h$.

Step 1: Regret decomposition. We write

$$\begin{aligned} r_t &= \sum_{i=1}^{n} \big(\ell^{(i)}(w_i) - \ell^{(i)}(w^\dagger)\big) \\ &= \sum_{i=1}^{n} \big(\ell^{(i)}(w_i) - \ell^{(i)}(w_{i+1})\big) + \sum_{i=1}^{n} \big(\ell^{(i)}(w_{i+1}) - \ell^{(i)}(w^\dagger)\big). \end{aligned}$$

The second sum is the standard be-the-regularized-leader (BTRL) overhead and admits the textbook bound

$$\sum_{i=1}^{n} \big(\ell^{(i)}(w_{i+1}) - \ell^{(i)}(w^\dagger)\big) \le R(w^\dagger) - R(w_1) = \frac{\mu}{2}\varepsilon_t^2, \tag{1}$$

where $R(w) \coloneqq \frac{\mu}{2} \|w - h\|_2^2$ and $w_1 = h$ gives $R(w_1) = 0$. We prove (1) by induction on n. The base case $n = 0$ is the empty inequality. Assume the bound holds for $n - 1$ rounds with any comparator $u \in \mathcal{K}$. Applied with $u = w_{n+1}$, it gives

$$\sum_{i=1}^{n-1} \ell^{(i)}(w_{i+1}) \le \sum_{i=1}^{n-1} \ell^{(i)}(w_{n+1}) + R(w_{n+1}) - R(w_1).$$

Adding $\ell^{(n)}(w_{n+1})$ to both sides and using the definition of F_n,

$$\sum_{i=1}^{n} \ell^{(i)}(w_{i+1}) \le F_n(w_{n+1}) - R(w_1).$$

Optimality $F_n(w_{n+1}) \le F_n(w^\dagger)$ then yields

$$\sum_{i=1}^{n} \ell^{(i)}(w_{i+1}) \le \sum_{i=1}^{n} \ell^{(i)}(w^\dagger) + R(w^\dagger) - R(w_1),$$

which is (1) for n rounds. Hence

$$r_t \le \frac{\mu}{2}\varepsilon_t^2 + \sum_{i=1}^{n} \big(\ell^{(i)}(w_i) - \ell^{(i)}(w_{i+1})\big).$$

Step 2: One-step stability. The function F_{i-1} is $i\mu$-strongly convex, because it is the sum of the μ-strongly convex regularizer and $i - 1$ losses, each also μ-strongly convex. Since w_i minimizes F_{i-1} over $\mathcal{K}$, the strong-convexity inequality at a constrained minimizer gives

$$F_{i-1}(w_{i+1}) - F_{i-1}(w_i) \ge \frac{i\mu}{2} \|w_{i+1} - w_i\|_2^2.$$

Because w_{i+1} minimizes $F_i = F_{i-1} + \ell^{(i)}$,

$$F_i(w_{i+1}) \le F_i(w_i),$$

which rearranges to

$$F_{i-1}(w_{i+1}) - F_{i-1}(w_i) \le \ell^{(i)}(w_i) - \ell^{(i)}(w_{i+1}).$$

Combining the last two displays gives

$$\frac{i\mu}{2}\left\|w_{i+1}-w_i\right\|_2^2 \le \ell^{(i)}(w_i)-\ell^{(i)}(w_{i+1}).$$

Since each loss is L-Lipschitz,

$$\ell^{(i)}(w_i)-\ell^{(i)}(w_{i+1}) \le L\left\|w_i-w_{i+1}\right\|_2.$$

Therefore

$$\left\|w_i-w_{i+1}\right\|_2 \le \frac{2L}{i\mu},$$

and so

$$\ell^{(i)}(w_i)-\ell^{(i)}(w_{i+1}) \le \frac{2L^2}{i\mu}.$$

Step 3: Sum the stability terms. Summing over $i=1,\dots,n$ yields

$$\sum_{i=1}^{n}\left(\ell^{(i)}(w_i)-\ell^{(i)}(w_{i+1})\right) \le \frac{2L^2}{\mu}\sum_{i=1}^{n}\frac{1}{i} = \frac{2L^2}{\mu}H_n.$$

Substituting into the decomposition from Step 1 gives

$$r_t \le \frac{\mu}{2}\varepsilon_t^2 + \frac{2L^2}{\mu}H_n.$$

Finally, $H_n \le 1+\log n$ yields the theorem. □

B.15 Strongly Convex Lower Bound at a Prescribed Distance

Proof of Theorem 9. Fix $\varepsilon \in [0, D/2]$ and let the warm start be $h=0$. Sample a random sign $\sigma \in \{-1,+1\}$ uniformly, and define for every round $i=1,\dots,n$ the loss

$$\ell^{(i)}(w) := \frac{\mu}{2}(w-\sigma\varepsilon)^2, \qquad w \in [-D/2, D/2].$$

Step 1: Verify the assumptions. Each loss is μ-strongly convex. Its derivative is

$$\frac{d}{dw}\ell^{(i)}(w) = \mu(w-\sigma\varepsilon),$$

so on $\mathcal{K}=[-D/2, D/2]$,

$$\left|\ell^{(i)\prime}(w)\right| \le \mu\left(\frac{D}{2}+\varepsilon\right) \le \mu D \le L.$$

Hence every loss is also L-Lipschitz.

Step 2: The warm-start distance is exactly ε. The cumulative loss is

$$\sum_{i=1}^{n}\ell^{(i)}(w) = \frac{n\mu}{2}(w-\sigma\varepsilon)^2.$$

Its unique minimizer over $\mathcal{K}$ is

$$\theta^\star = \sigma\varepsilon.$$

Therefore

$$\operatorname{dist}(h, \mathcal{W}^\star) = |\sigma\varepsilon| = \varepsilon.$$

Step 3: Lower-bound the regret. Let W be the learner's first action. Since no loss has yet been observed, W may depend on the warm start and the learner's internal randomness, but not on the fresh random sign σ.

The comparator loss is zero at $\theta^\star = \sigma\varepsilon$. Thus the learner's first-round regret equals

$$\ell^{(1)}(W) - \ell^{(1)}(\theta^\star) = \frac{\mu}{2}(W - \sigma\varepsilon)^2.$$

Taking expectation over σ conditional on W gives

$$\begin{aligned}\mathbb{E}_\sigma\Big[\frac{\mu}{2}(W-\sigma\varepsilon)^2 \,\Big|\, W\Big] &= \frac{\mu}{4}\big((W-\varepsilon)^2 + (W+\varepsilon)^2\big) \\ &= \frac{\mu}{2}\big(W^2 + \varepsilon^2\big) \\ &\geq \frac{\mu}{2}\varepsilon^2.\end{aligned}$$

Since the total regret is at least the first-round regret, taking expectation over the learner's randomness yields

$$\mathbb{E}[r_t] \geq \frac{\mu}{2}\varepsilon^2.$$

Step 4: Fix one sign. The displayed lower bound is the average of the two choices $\sigma \in \{-1, +1\}$. Therefore at least one deterministic sign yields a one-task loss sequence with warm-start distance ε and expected regret at least $\mu\varepsilon^2/2$. □

B.16 Comparing Strongly Convex Upper Bounds

The unmatched harmonic term in Theorem 8 matters when comparing it with the convex oracle bound. The following sufficient condition makes both terms of the RFTL upper bound small enough to improve on that benchmark.

Corollary 4 (When curvature helps)**.** *The oracle comparison of Theorems 1 and 8 gives*

$$\min\Big\{\varepsilon L\sqrt{n},\ \frac{\mu}{2}\varepsilon^2 + \frac{2L^2}{\mu}H_n\Big\}.$$

Whenever $4LH_n/(\mu\sqrt{n}) \leq \varepsilon \leq L\sqrt{n}/\mu$, the strongly convex bound is no larger than the linear bound.

Proof of Corollary 4. The displayed inequality is simply the minimum of the two upper bounds proved in Theorems 1 and 8.

For the second claim, assume

$$\frac{4LH_n}{\mu\sqrt{n}} \leq \varepsilon_t \leq \frac{L\sqrt{n}}{\mu}.$$

Then

$$\frac{2L^2}{\mu}H_n \leq \frac{\varepsilon_t L\sqrt{n}}{2}$$

by the lower bound on ε_t, and

$$\frac{\mu}{2}\varepsilon_t^2 \leq \frac{\varepsilon_t L\sqrt{n}}{2}$$

by the upper bound on ε_t. Adding the two displays gives

$$\frac{\mu}{2}\varepsilon_t^2 + \frac{2L^2}{\mu}H_n \leq \varepsilon_t L\sqrt{n}.$$

□

B.17 High-Probability Cumulative-Regret Lower Bound

Independent copies of the lower-bound construction also control cumulative regret with high probability. The result retains the sum of prescribed distances and subtracts a deviation determined by the bounded range of task regrets.

Proposition 6 (High-probability lower bound)**.** *Adopt the construction of Theorem 3: T independent tasks, each with warm start $h_t = 0$ and distance $\bar{\varepsilon}_t \in [0, D/2]$, generated by independent Rademacher sign sequences $(\xi_{t,i})_{i=1}^n$. Let $R_T = \sum_{t=1}^T r_t$ be the cumulative per-task regret of any (possibly randomized) algorithm. Then for every $\delta \in (0,1)$, with probability at least $1-\delta$ over the joint randomness,*

$$R_T \;\geq\; \frac{L\sqrt{n}}{8} \sum_{t=1}^{T} \bar{\varepsilon}_t \;-\; DLn\,\sqrt{2T\,\log(1/\delta)}.$$

The bound is positive when the average distance satisfies $\frac{1}{T}\sum_t \bar{\varepsilon}_t > 8D\sqrt{2n\log(1/\delta)/T}$.

Proof. By the proof of Theorem 3, the conditional expectation of the per-task regret satisfies $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}] \geq L\bar{\varepsilon}_t\sqrt{n}/8$ for each t, where $\mathcal{F}_{t-1}$ is the sigma-algebra generated by the first $t-1$ tasks. Hence

$$\mathbb{E}[R_T] \;\geq\; \frac{L\sqrt{n}}{8} \sum_{t=1}^{T} \bar{\varepsilon}_t.$$

Each per-task regret satisfies $r_t \in [-nLD, nLD]$ almost surely: the algorithm and the comparator both lie in $\mathcal{K} = [-D/2, D/2]$ and each loss is L-Lipschitz, so a single-round excess lies in $[-LD, LD]$ and n rounds lie in $[-nLD, nLD]$. Negative static regret is allowed.

Define the martingale-difference sequence $X_t := \mathbb{E}[r_t \mid \mathcal{F}_{t-1}] - r_t$. Then each X_t takes values in an interval of length $W = 2nLD$. Hoeffding's lemma therefore yields

$$\Pr\left(\sum_{t=1}^{T} X_t \geq s\right) \;\leq\; \exp\left(-\frac{2s^2}{TW^2}\right) = \exp\left(-\frac{s^2}{2T(nLD)^2}\right).$$

Taking $s = nLD\sqrt{2T\log(1/\delta)}$ gives, with probability at least $1-\delta$,

$$R_T \;=\; \sum_{t=1}^{T} \mathbb{E}[r_t \mid \mathcal{F}_{t-1}] - \sum_{t=1}^{T} X_t \;\geq\; \sum_{t=1}^{T} \mathbb{E}[r_t \mid \mathcal{F}_{t-1}] - DLn\sqrt{2T\log(1/\delta)}.$$

Using the conditional lower bound $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}] \geq L\bar{\varepsilon}_t\sqrt{n}/8$ for every t gives

$$R_T \;\geq\; \frac{L\sqrt{n}}{8} \sum_{t=1}^{T} \bar{\varepsilon}_t - DLn\sqrt{2T\log(1/\delta)},$$

which is the displayed bound. □

The probability is over both the independent hard tasks and the learner's randomness. This statement concerns the specified random family; fixing a deterministic sequence by averaging gives the expectation guarantee of Theorem 3, not automatically the same high-probability guarantee.

B.18 Convex Combination of the Warm Start and RFTL

Theorem 10 (Convex combination with RFTL)**.** *Let v_i be the RFTL iterate of Theorem 8. For every $\alpha \in [0,1]$, the combination $w_i = \alpha h + (1-\alpha)v_i$ satisfies*

$$r_n \leq \alpha nL\varepsilon + (1-\alpha)\Big(\frac{\mu}{2}\varepsilon^2 + \frac{2L^2}{\mu}H_n\Big).$$

Selecting between playing h and RFTL with known ε gives $\min\{nL\varepsilon,\, \mu\varepsilon^2/2 + (2L^2/\mu)H_n\}$, which improves on Theorem 8 whenever $\varepsilon < 2LH_n/(\mu n)$.

Proof of Theorem 10. Let v_i be the RFTL iterate of Theorem 8 and write $w_i = \alpha h + (1-\alpha)v_i$. Choose $w^\dagger \in \mathcal{W}^\star$ with $\|h - w^\dagger\|_2 = \varepsilon$. Theorem 8 gives

$$\sum_{i=1}^{n} \ell_i(v_i) - F(w^\dagger) \le \frac{\mu}{2}\varepsilon^2 + \frac{2L^2}{\mu} H_n.$$

Convexity of $\mathcal{K}$ keeps w_i feasible. Convexity of each ℓ_i yields $\ell_i(w_i) \le \alpha\ell_i(h) + (1-\alpha)\ell_i(v_i)$. Summing and subtracting $F(w^\dagger)$ produces

$$r_n \le \alpha\big[F(h) - F(w^\dagger)\big] + (1-\alpha)\Big(\frac{\mu}{2}\varepsilon^2 + \frac{2L^2}{\mu} H_n\Big).$$

Lipschitzness of F gives $F(h) - F(w^\dagger) \le nL\varepsilon$. Dropping the nonnegative $\mu\varepsilon^2/2$ term shows that playing h has a strictly smaller upper bound than RFTL whenever $\varepsilon < 2LH_n/(\mu n)$. Selecting the better endpoint from these bounds uses ε; the guarantee for any fixed α does not. □

Figure 2 compares upper bounds rather than observed regrets. In the left panel, RFTL has the smaller bound at larger horizons, while oracle OGD has the smaller bound when the harmonic baseline dominates. In the right panel, staying at h has the smaller bound for sufficiently accurate warm starts. The red dotted curve marks the sufficient condition $\varepsilon = 2LH_n/(\mu n)$. Selecting a policy from either diagram uses the plotted distance.

B.19 A Standard No-Warm-Start Lower Bound

Proposition 7 (Standard no-warm-start lower bound)**.** *Consider a single task with no warm start information, $d = 1$, $\mathcal{K} = [-D/2, D/2]$, and n rounds of L-Lipschitz convex losses. Every online algorithm has worst-case expected regret at least*

$$\frac{DL}{4}\sqrt{n}.$$

Proof. Let $\xi_1, \dots, \xi_n$ be i.i.d. Rademacher random variables and define the linear losses

$$\ell^{(i)}(w) = L\xi_i w, \qquad w \in [-D/2, D/2].$$

These losses are convex and L-Lipschitz.

Let w_i be the learner's action at round i. Because w_i is measurable with respect to the past and ξ_i is independent of the past,

$$\mathbb{E}[\xi_i w_i] = 0.$$

Therefore

$$\mathbb{E}\left[\sum_{i=1}^{n} \ell^{(i)}(w_i)\right] = 0.$$

On the other hand, for the offline comparator,

$$\min_{w\in[-D/2,D/2]} \sum_{i=1}^{n} \ell^{(i)}(w) = \min_{w\in[-D/2,D/2]} Lw \sum_{i=1}^{n} \xi_i = -\frac{LD}{2}\left|\sum_{i=1}^{n} \xi_i\right|.$$

Thus the expected regret is

$$\mathbb{E}[r] = \frac{LD}{2}\,\mathbb{E}\left[\left|\sum_{i=1}^{n} \xi_i\right|\right].$$

Khintchine's inequality gives

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} \xi_i\right|\right] \ge \sqrt{\frac{n}{2}},$$

and hence

$$\mathbb{E}[r] \ge \frac{LD}{2}\sqrt{\frac{n}{2}} \ge \frac{DL}{4}\sqrt{n}.$$

□

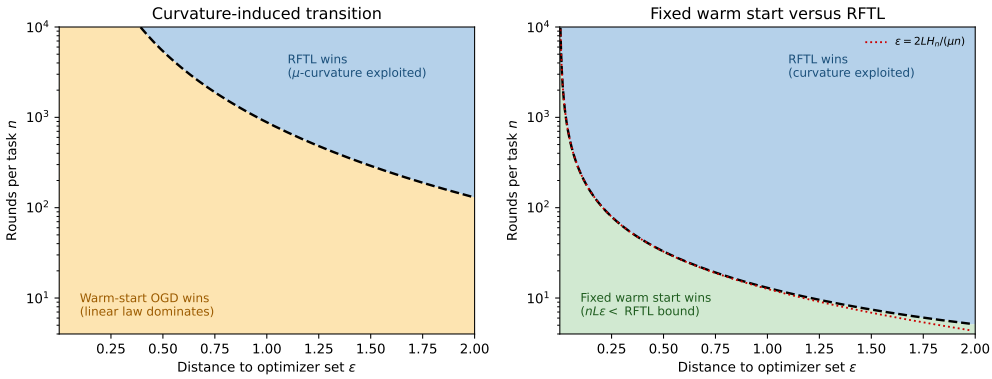


Figure 2: Phase boundaries in (ε, n) for the strongly convex upper bounds with $L = 1$ and $\mu = 1/2$. Left: comparison of oracle OGD and RFTL from Corollary 4. Right: comparison of staying at h and RFTL from Theorem 10. Each point is colored by the smaller upper bound; dashed curves mark equality, and the red dotted curve in the right panel is $\varepsilon = 2LH_n/(\mu n)$.

B.20 Source Aggregation and Stochastic Tasks

Proposition 8 (Multiple warm-start sources)**.** *Suppose each task t comes with K candidate warm starts $h_t^{(1)}, \dots, h_t^{(K)} \in \mathcal{K}$. For each source k, let $R_T^{(k)}$ be the total regret of warm-start OGD with the worst-case step size from Theorem 1, run with source k on every task. Assume that after each round the full loss function is revealed, so that all K source-specific trajectories can be replayed exactly offline.*

Then there is a deterministic master algorithm such that

$$R_T \le \min_{k\in[K]} R_T^{(k)} + nLD\sqrt{2T\ln K}.$$

After shifting each task's source-regret vector by its minimum, the shifted coordinates lie in $[0, nLD]$ because any two feasible trajectories differ by at most nLD in cumulative loss. Hedge with this range pays $nLD\sqrt{2T\ln K}$. Applying the worst-case step size to each warm-start OGD instance gives

$$R_T \le \frac{L\sqrt{n}}{2D} \min_{k\in[K]} \sum_{t=1}^{T} \big(\varepsilon_t^{(k)}\big)^2 + \frac{TDL\sqrt{n}}{2} + nLD\sqrt{2T\ln K}.$$

Proof. Let $p_t \in \Delta_K$ be the Hedge distribution at task t. For every round i of task t, let source k propose its own action $w_{t,i}^{(k)}$, and let the master play the convex combination

$$\bar{w}_{t,i} := \sum_{k=1}^{K} p_t^{(k)} w_{t,i}^{(k)}.$$

For each source k, define its cumulative task loss by

$$C_t^{(k)} := \sum_{i=1}^{n} \ell_t^{(i)}\Big(w_{t,i}^{(k)}\Big).$$

Because $\mathcal{K}$ is convex, $\bar{w}_{t,i} \in \mathcal{K}$. By convexity of the losses,

$$\sum_{i=1}^{n} \ell_t^{(i)}(\bar{w}_{t,i}) \le \sum_{k=1}^{K} p_t^{(k)} C_t^{(k)}.$$

Subtracting

$$m_t := \min_{\theta\in\mathcal{K}} F_t(\theta)$$

from both sides gives the task-wise regret bound

$$r_t^{\text{master}} \le \sum_{k=1}^{K} p_t^{(k)} r_t^{(k)}.$$

Shift the source-regret vector at task t by $\min_k r_t^{(k)}$. The shifted coordinates lie in $[0, nLD]$. The standard Hedge guarantee (Freund & Schapire, 1997) on the shifted task-level loss vectors $(r_t^{(1)}, \dots, r_t^{(K)})$ gives the deterministic bound

$$\sum_{t=1}^{T}\sum_{k=1}^{K} p_t^{(k)} r_t^{(k)} \le \min_{k\in[K]} \sum_{t=1}^{T} r_t^{(k)} + nLD\sqrt{2T\ln K}.$$

Summing the task-wise bound

$$r_t^{\text{master}} \le \sum_{k=1}^{K} p_t^{(k)} r_t^{(k)}$$

over tasks and combining with Hedge yields

$$R_T \le \min_{k\in[K]} \sum_{t=1}^{T} r_t^{(k)} + nLD\sqrt{2T\ln K}.$$

This is exactly the first display.

The second display follows from the worst-case step-size guarantee in Theorem 1, namely

$$R_T^{(k)} \le \frac{L\sqrt{n}}{2D} \sum_{t=1}^{T} \big(\varepsilon_t^{(k)}\big)^2 + \frac{TDL\sqrt{n}}{2},$$

and from the shifted range nLD. □

Proposition 9 (Multiple warm-start sources, strongly convex)**.** *Suppose each task t comes with K candidate warm starts $h_t^{(1)}, \dots, h_t^{(K)} \in \mathcal{K}$ and that every loss $\ell_t^{(i)}$ is also μ-strongly convex on $\mathcal{K}$. For each source k, let $R_T^{(k)}$ be the total regret of warm-start RFTL from Theorem 8 run with source k on every task. Assume that after each round the full loss function is revealed, so that all K source-specific trajectories can be replayed exactly offline.*

Then there is a deterministic master algorithm such that

$$R_T \le \min_{k\in[K]} R_T^{(k)} + nLD\sqrt{2T\ln K}.$$

With the same source-loss normalization as in Proposition 8, warm-start RFTL gives

$$R_T \;\le\; \frac{\mu}{2} \min_{k\in[K]} \sum_{t=1}^{T} \big(\varepsilon_t^{(k)}\big)^2 \;+\; \frac{2TL^2}{\mu} H_n \;+\; nLD\sqrt{2T\ln K}.$$

Proof. Replace warm-start OGD by warm-start RFTL in the proof of Proposition 8. Convexity of each $\ell_t^{(i)}$ still gives

$$r_t^{\text{master}} \;\le\; \sum_{k=1}^{K} p_t^{(k)} r_t^{(k)}.$$

Static source regrets may be negative, so the same minimum shift as in Proposition 8 is required: the shifted coordinates lie in $[0, nLD]$ because any two feasible trajectories differ by at most nLD in cumulative loss. Hedge (Freund & Schapire, 1997) on those shifted vectors yields

$$\sum_{t=1}^{T}\sum_{k=1}^{K} p_t^{(k)} r_t^{(k)} \;\le\; \min_{k\in[K]} \sum_{t=1}^{T} r_t^{(k)} \;+\; nLD\sqrt{2T\ln K}.$$

Combining the per-task and Hedge bounds gives the first display. The second display follows by substituting the per-source RFTL guarantee from Theorem 8. □

For $K = 1$, the aggregation cost vanishes and both results reduce to their one-task bounds summed over tasks. With an i.i.d. model for tasks and their supplied warm starts, the same deterministic bounds can instead be summarized through the distribution of the warm-start distance.

Proposition 10 (I.i.d. task sequences)**.** *Assume the tasks are drawn i.i.d. and warm-start OGD with the worst-case step size from Theorem 1 is run independently on each task. If $\bar{\varepsilon}^2 := \mathbb{E}[\varepsilon_t^2] < \infty$, then*

$$\mathbb{E}\left[\frac{R_T}{T}\right] \le \frac{\bar{\varepsilon}^2 L\sqrt{n}}{2D} + \frac{DL\sqrt{n}}{2}.$$

Moreover, for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\frac{R_T}{T} \le \frac{\bar{\varepsilon}^2 L\sqrt{n}}{2D} + \frac{DL\sqrt{n}}{2} + \frac{DL\sqrt{n}}{2}\sqrt{\frac{\ln(1/\delta)}{2T}}.$$

Proof. By the worst-case tuning in Theorem 1,

$$r_t \le \frac{L\sqrt{n}}{2D}\varepsilon_t^2 + \frac{DL\sqrt{n}}{2}.$$

Taking expectations and using identical distribution across tasks gives

$$\mathbb{E}\left[\frac{R_T}{T}\right] = \mathbb{E}[r_1] \le \frac{\bar{\varepsilon}^2 L\sqrt{n}}{2D} + \frac{DL\sqrt{n}}{2}.$$

The bound $r_t \le (L\sqrt{n}/(2D))\varepsilon_t^2 + DL\sqrt{n}/2$ is deterministic. Consequently it is enough to concentrate the bounded variables $\varepsilon_t^2 \in [0, D^2]$. Hoeffding's inequality yields, with probability at least $1-\delta$,

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \le \bar{\varepsilon}^2 + D^2\sqrt{\frac{\ln(1/\delta)}{2T}}.$$

Substituting into the preceding bound gives

$$\frac{R_T}{T} \le \frac{\bar{\varepsilon}^2 L\sqrt{n}}{2D} + \frac{DL\sqrt{n}}{2} + \frac{DL\sqrt{n}}{2}\sqrt{\frac{\ln(1/\delta)}{2T}}.$$

Only the squared distances need to be concentrated, because the per-task regret inequality holds pathwise. □

B.21 Experiment Details

This section documents the synthetic distance-dependent regret curves (Figure 3), the analytical comparisons of upper bounds (Figure 2), and the UCI Digits study in the main text (Figure 1).

Synthetic regret experiments (Figure 3). Both panels use the common choices

$$d = 1, \qquad D = 2, \qquad L = 1, \qquad n = 201.$$

We choose n odd so that the lower-bound construction of Theorem 3 uses all n rounds without zero-padding. We evaluate the distance grid

$$\varepsilon \in \{0, 0.05, 0.10, \ldots, 1.00\}$$

and average over 300 independently sampled tasks for each value, using random seed 0.

Left panel: convex lower-bound family. The left panel uses exactly the task family from the proof of Theorem 3. Each task samples i.i.d. signs $\xi_1, \ldots, \xi_n \in \{-1, 1\}$ and uses the piecewise-linear losses from that construction.

We compare the following four methods.

1. **Oracle-tuned OGD**: warm-start OGD with $\eta = \varepsilon/(L\sqrt{n})$.
2. **Warm-start OGD with worst-case tuning**: warm-start OGD with $\eta = D/(L\sqrt{n})$.

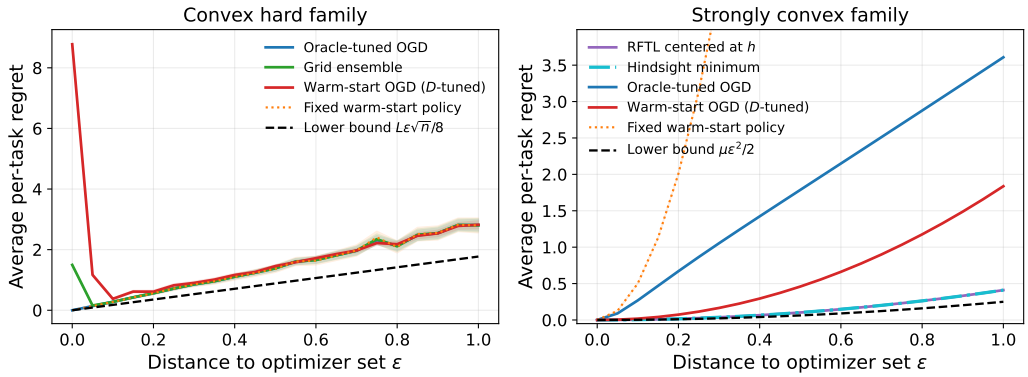


Figure 3: Witness families ($D = 2$, $L = 1$, $n = 201$). Left: convex construction of Theorem 3. Right: repeated quadratic family of Theorem 9 with $\mu = 1/2$. Curves are means over 300 tasks; bands are ± 2 standard errors.

3. **Grid ensemble**: a Hedge master over warm-start OGD experts with step sizes

$$\eta_j = \frac{D}{L\sqrt{n}\,2^j}, \qquad j = 0, \dots, \lceil \log_2 n \rceil.$$

   The master learning rate is

$$\eta_{\text{meta}} = \sqrt{\frac{8 \ln(\lceil \log_2 n \rceil + 1)}{n D^2 L^2}}.$$

   On each round, the master plays the weighted average of the expert actions and updates the expert weights from the shifted linearized losses induced by the subgradient observed at the master point, exactly as in Theorem 4.

4. **Fixed warm-start policy**: $w_{t,i} = h_t = 0$ for every round, providing the fixed-action baseline used by Theorem 10.

The barrier vanishes at $h = 0$, so staying there avoids the additional losses incurred outside $[-\varepsilon, \varepsilon]$. This makes the fixed policy $w_i = h$ competitive on this particular family, even though the lower bound applies to every algorithm. The quadratic family below gives a contrasting case in which moving away from h is much more effective.

Right panel: strongly convex quadratic family. The right panel uses a companion strongly convex witness family inspired by the proof of Theorem 9. Fix

$$\mu = \frac{1}{2}.$$

Each task samples a sign $\sigma \in \{-1, 1\}$ uniformly at random and then uses the same quadratic loss on every round,

$$\ell^{(i)}(w) = \frac{\mu}{2}(w - \sigma\varepsilon)^2.$$

For $\mathcal{K} = [-1, 1]$ and $\varepsilon \in [0, 1]$, these losses are μ-strongly convex and have gradient norm at most

$$\mu\,|w - \sigma\varepsilon| \le \mu(1 + \varepsilon) \le 1 = L.$$

On this panel we compare five methods:

1. **RFTL centered at** h: the strongly convex algorithm from Theorem 8.
2. **Oracle-tuned OGD**: warm-start OGD with $\eta = \varepsilon/(L\sqrt{n})$.
3. **Warm-start OGD with worst-case tuning**: warm-start OGD with $\eta = D/(L\sqrt{n})$.
4. **Fixed warm-start policy**: $w_{t,i} = h_t = 0$ for every round.

5. **Hindsight minimum** (Theorem 10): the minimum of the fixed warm-start policy and RFTL centered at h, corresponding to the hindsight choice $\alpha \in \{0, 1\}$ in their convex combination.

On this strongly convex quadratic family, RFTL centered at h achieves regret $\sum_{i=1}^{n} \frac{\mu}{2}(\varepsilon/i)^2 = \frac{\mu\varepsilon^2}{2}\sum_{i=1}^{n} i^{-2} \le \frac{\pi^2\mu\varepsilon^2}{12}$ because its closed-form iterate is $w_{i+1} = \frac{i}{i+1}\sigma\varepsilon$. This is smaller than the regret $\frac{n\mu\varepsilon^2}{2}$ of the fixed policy $w_i = h$ for all $\varepsilon > 0$ and $n \ge 2$, so the hindsight minimum coincides with RFTL on this family. This is a benchmark, not an implementable selection rule. The fixed policy shows that never leaving the warm start can cost a factor proportional to n relative to RFTL on this witness family.

Real-data validation on UCI Digits (Figure 1). We construct $T = 45$ one-vs-one tasks from UCI Digits. Each task uses $n = 100$ online rounds and a separate support set of 10 labeled examples per class. Support and online examples never overlap within a task. Features are scaled to $[0, 1]$ and projected onto the unit sphere. The feasible set is the Euclidean ball of radius 5. The per-round loss is logistic loss plus $(\lambda/2)\|w\|_2^2$ with $\lambda = 10^{-2}$, so the cumulative online loss equals the constrained offline objective. The comparator is a high-accuracy constrained minimizer checked by the projected-gradient mapping from three initializations. The implied Lipschitz constant is $L = 1 + \lambda \cdot 5 = 1.05$. The warm start on task t is the mean of up to $k = 5$ nearest previous comparators. We compare no-warm-start OGD, warm-start OGD with the worst-case step size, oracle-tuned OGD, tuned OGD with $C_0 = DL\sqrt{n}/4$, grid aggregation, and a coordinate-wise Krichevsky-Trofimov surrogate of Theorem 5, plus retrieval corruption $p \in \{0, 0.25, 0.5, 0.75, 1\}$. All quantities use 30 seeds $\{0, \dots, 29\}$. The supplementary material records optimizer diagnostics and numerical summaries.

The projected-gradient mapping residual has median 9.66×10^{-9} and maximum 1.00×10^{-8} across tasks and seeds. At intact retrieval, the no-warm-start mean regret is 9.26 (95% seed-level bootstrap interval $[9.23, 9.29]$), compared with 6.21 ($[6.16, 6.25]$) for warm-start OGD with the worst-case step size. At $p = 1$, the latter has mean 10.64 ($[10.51, 10.78]$). Shuffled retrieval gives mean 10.40 for the same method. The hindsight adversarial warm start is a diagnostic only: this method has mean 31.21, while the parameter-free baseline has mean 126.71. Full numerical summaries, including medians and rank correlations, are in the supplementary material.

The retrieval descriptor is the normalized difference of the two labeled class means, computed only from the disjoint support set before adaptation. The first task has no previous comparator and starts at zero; later tasks average the closest $\min\{5, t-1\}$ previous comparators. Thus the study evaluates adaptation with small labeled task descriptors, not label-free retrieval. Their mean cosine similarity has Spearman correlation -0.848 with realized distance. Bootstrap intervals resample the 30 seed-level means and describe this experimental protocol; they are not confidence guarantees for transfer to an unseen task population.

Reproducibility. The supplementary material includes all procedures needed to regenerate Figures 1 to 3. Synthetic curves average 300 tasks per distance value with ± 2 standard errors, and Figure 2 evaluates the closed-form upper bounds on a 240×240 grid. It also specifies package versions, random seeds, preprocessing, solver tolerances, and the commands used for every experiment.