CONSTRAINED ONLINE LEARNING WITH NOISY CONSTRAINT VALUES

Anonymous authors
Paper under double-blind review

ABSTRACT

We study constrained online convex optimization with adversarial constraints and conditionally unbiased, finite-variance observations of constraint values and gradients. Under common feasibility, our LEDGER algorithm attains $O(\sqrt{T})$ expected regret and $O(\sqrt{T\log(eT)})$ expected budget violation, the largest cumulative overspend over any window. It uses a reflected exponential potential, clipped signed observations, and predictable adaptive regularization, with one feedback triple and one projection per round. Neither a Slater condition, independence between feedback channels, nor an absolute constraint-value bound is needed. A Gaussian testing lower bound proves that the budget rate has optimal horizon dependence under square-root regret at fixed positive noise, including the logarithm. The same obstruction holds for terminal violation, so the logarithm is not a cost of maximizing over windows; an $O(\sqrt{T})$ budget bound instead forces linear regret. In contrast, fixed positive Gaussian value noise yields a joint regret–hard-violation lower bound of $\Omega(\min\{\sigma,1\}T/\log^2 T)$, even with exact gradients in one dimension. The hard-violation construction matches arbitrarily many moments while preserving a feasible-endpoint gap and constant endpoint probabilities. Together, the bounds separate uncertainty about hard feasibility from learnable signed budgets. Deterministic restarts remove the horizon input without changing either upper rate.

1 INTRODUCTION

Constrained online convex optimization (COCO) models sequential resource allocation when both costs and constraints change over time. On round t, the learner chooses x_t from a known convex domain, incurs loss $\ell_t(x_t)$, and has net consumption $\gamma_t(x_t)$, where $\gamma_t(x) \le 0$ means that action x satisfies the round's constraint. For consumption $c_t(x)$ and allowance b_t, this is $\gamma_t(x) = c_t(x) - b_t$. With exact feedback, it is possible to achieve $O(\sqrt{T})$ regret against an action feasible on every round while keeping the sum of positive constraint violations at $O(\sqrt{T})$ (Sarkar and Sinha, 2026). In a noisy observation model, feasibility itself is uncertain: an unbiased estimate of a constraint value need not reveal whether the action satisfied the constraint. This distinction is not addressed by allowing noisy gradients while retaining exact values, or by assuming a fixed distribution of stochastic constraints (Yu et al., 2017; Wei et al., 2020; Yu et al., 2026).

We study adversarial convex losses and constraints observed through three local quantities: a loss direction, a constraint direction, and a scalar constraint value. The two directions guide the action update, but the scalar identifies the constraint level. Without it, the learner cannot distinguish $\gamma_t(x) \le 0$ from $\gamma_t(x) \le c$ for a constant c: the functions γ_t and $\gamma_t - c$ have identical gradients but different feasible sets. An additive loss offset cancels from regret; a constraint offset does not cancel from feasibility or budget overspend. We observe one noisy estimate of each quantity per round. The observation errors have conditional mean zero and bounded conditional second moments; they need not be independent or symmetric. We ask which joint regret–violation guarantees are possible with this information. The answer depends on the violation metric.

Hard violation, $\mathrm{ICV}_T = \sum_t [\gamma_t(x_t)]_+$, charges every infeasible round separately. We prove that Gaussian value noise of fixed standard deviation $\sigma > 0$ makes the worst-case maximum of expected regret and expected hard violation at least $\Omega(\min\{\sigma,1\}T/\log^2 T)$. This rules out every jointly

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.

Feedback / guarantee	Regret	Violation
Exact; potential methods (Sinha and Vaze, 2024; Yu et al., 2026)	$O(\sqrt{T})$	hard: $O(\sqrt{T}\log T)$
Exact; nested projections (Sarkar and Sinha, 2026)	$O(\sqrt{T})$	hard: $O(\sqrt{T})$
Noisy values and directions; LEDGER (Theorem 4.1)	$O(\sqrt{T})$	budget: $O(\sqrt{T\log(eT)})$
Noisy values; necessary budget (Proposition 3.2)	$O(\sqrt{T})$ required	budget: $\Omega(\sigma\sqrt{T\log T})$

Table 1: Joint guarantees under common feasibility. Rates are for fixed problem bounds; noisy-feedback bounds are in expectation. The last row is a lower bound conditioned on a uniform regret guarantee. Exact-feedback methods have different oracle access. The noisy upper bound matches horizon dependence for fixed positive noise.

polynomially sublinear pair, even for one-dimensional affine constraints with known gradients. We therefore consider *budget violation*,

$$\mathrm{WCV}_T = \max\left\{0, \max_{1\le s\le e\le T}\sum_{t=s}^{e}\gamma_t(x_t)\right\}.$$

It allows underspending to offset overspending within a window, but an early surplus cannot hide every later deficit. For example, alternating levels $-1, +1$ have zero terminal violation, hard violation $T/2$, and budget violation 1 for even T. This objective is appropriate when resources can be balanced across rounds, not when every infeasible action is unacceptable. Maximum-window accounting already appears in the exact-feedback constraint-satisfaction problem of Sinha and Vaze (2024, Sec. 3); the metric itself is not new.

Contributions. The paper gives both achievable guarantees and statistical obstructions for this noisy local-feedback model.

1. **Square-root regret with optimal-order budget control.** Under common feasibility, LEDGER achieves $O(\sqrt{T})$ expected regret and $O(\sqrt{T\log(eT)})$ expected maximum-window violation (Theorem 4.1). It uses one feedback triple and one projection onto the known decision set per round. Only conditional second moments are needed. In particular, the algorithm does not require a known bound on absolute constraint values. Deterministic restarts remove the horizon input without changing the rates (Corollary 4.2). The guarantee controls the true maximum overspend before expectation, rather than only a noisy balance or a terminal signed sum.
2. **A matching budget lower bound and an endpoint obstruction.** A two-instance Gaussian test gives a regret–budget tradeoff (Proposition 3.2). Uniform $O(\sqrt{T})$ expected regret forces $\Omega(\sigma\sqrt{T\log T})$ expected budget violation for fixed $\sigma > 0$. Moreover, exact $O(\sqrt{T})$ budget violation forces linear worst-case regret. A joint $\Omega(\min\{\sigma, 1\}\sqrt{T\log T})$ lower bound also matches the minimax order of the larger of the two criteria (Corollary 3.3). The obstruction already holds for terminal violation and known exact gradients: the logarithm reflects statistical uncertainty about the constraint level, not the number of windows or difficulty estimating a direction.
3. **Nearly linear hard-violation obstruction.** Moment-matched distributions with separated support endpoints give $\Omega(\sigma T^{1-1/(4n)})$ for each fixed matching order and $\Omega(\min\{\sigma, 1\}T/\log^2 T)$ when the order grows (Theorem 3.1). An explicit Chebyshev interpolation construction keeps the endpoint probabilities bounded away from zero. This is what prevents the testing reduction from losing additional powers of the matching order.

Technical approach. We couple signed reflection with adaptive regularization, building on established queue and potential methods (Lindley, 1952; Sinha and Vaze, 2024; Yu et al., 2026). A zero boundary derivative and clipping make the exponential drift compatible with finite variance. The regularizer cancels the weighted primal cost, while its unweighted sum remains small enough to charge back to true overspend. This yields $O(\log(eT))$ additional regret and $O(\sqrt{T\log(eT)})$ budget cost. Section 4.3 develops the coupled argument.

Related work. Long-term-constraint OCO began with Mahdavi et al. (2012). Adversarial time-varying constraints without Slater were studied by Yi et al. (2023); Guo et al. (2022), followed by near-optimal and optimal exact-feedback regret–hard-violation guarantees (Sinha and Vaze, 2024; Sarkar and Sinha, 2026). Stochastic-constraint methods instead use a fixed constraint distribution and a comparator feasible in expectation (Yu et al., 2017; Wei et al., 2020; Yu et al., 2026). Our comparator requirement is stronger, but the constraint sequence and conditional observation laws may vary adversarially. The fixed-length comparator windows of Liakopoulos et al. (2019) differ from our learner-side maximum overspend. For the lower bounds, Gaussian smoothing and moment matching have precedents in deconvolution and statistical estimation (Fan, 1991; Wu and Yang, 2019); the online regret–violation reduction and the feasible-endpoint construction are given explicitly here. Appendix A provides the detailed comparison.

2 Setting

Write $[a]_+ = \max\{a, 0\}$; logarithms are natural. Fix a compact convex set $\mathcal{K} \subseteq \mathbb{R}^d$ of Euclidean diameter at most $D > 0$ and a horizon $T \geq 1$. On round t, the learner plays $x_t \in \mathcal{K}$ and incurs a convex G-Lipschitz loss $\ell_t(x_t)$ and a convex G-Lipschitz constraint level $v_t = \gamma_t(x_t)$. The constraint is $\gamma_t(x) \leq 0$; a per-round allowance is absorbed into γ_t. The adversary may depend on the past and the current action, but not current feedback noise. The functions are not revealed. The learner has projection access to $\mathcal{K}$ and receives only the local feedback defined below.

Let $\mathcal{F}_t$ be the pre-feedback σ-field containing the history, x_t, the adversary's functions, and selected subgradients

$$p_t \in \partial \ell_t(x_t), \qquad q_t \in \partial \gamma_t(x_t), \qquad v_t = \gamma_t(x_t). \tag{1}$$

For known $\widehat{G} \geq G > 0$ and $\sigma \geq 0$, the oracle returns $(\tilde{p}_t, \tilde{q}_t, \tilde{v}_t)$ with

$$\mathbb{E}[\tilde{p}_t \mid \mathcal{F}_t] = p_t, \quad \mathbb{E}[\tilde{q}_t \mid \mathcal{F}_t] = q_t, \quad \tilde{v}_t = v_t + \varepsilon_t, \quad \mathbb{E}[\varepsilon_t \mid \mathcal{F}_t] = 0, \tag{2}$$

$$\mathbb{E}[\|\tilde{p}_t\|^2 \mid \mathcal{F}_t],\ \mathbb{E}[\|\tilde{q}_t\|^2 \mid \mathcal{F}_t] \leq \widehat{G}^2, \qquad \mathbb{E}[\varepsilon_t^2 \mid \mathcal{F}_t] \leq \sigma^2. \tag{3}$$

The three errors may be dependent on the same round, and their conditional distributions may vary with history. No symmetry, almost-sure observation bound, or sub-Gaussian tail is assumed. The value condition is stated for the error so that a bound on absolute constraint levels is unnecessary. Exact feedback makes all three channels exact; $\sigma = 0$ alone makes only the value channel exact.

Set $\mathcal{G}_t = \sigma(\mathcal{F}_t, \tilde{p}_t, \tilde{q}_t, \tilde{v}_t)$ and $\mathcal{G}_0 = \mathcal{F}_1$. The fields satisfy $\mathcal{G}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{G}_t$. Conditional mean-zero increments given $\mathcal{F}_t$ are therefore martingale differences for $(\mathcal{G}_t)$ by the tower property. All algorithmic weights and step sizes are chosen from the learner's past, hence are $\mathcal{G}_{t-1}$-measurable. We use *predictable* in this usual sense. The larger analysis field $\mathcal{F}_t$ also includes the adversary's current choices, and the new noise is centered conditional on these choices. The learner need not observe all the contents of $\mathcal{F}_t$.

Comparators and metrics. Let

$$\mathcal{U} = \{u \in \mathcal{K} : \mathbb{P}(\gamma_t(u) \leq 0 \text{ for all } t \leq T) = 1\}.$$

We assume $\mathcal{U} \neq \emptyset$: there is a deterministic action feasible on every round almost surely. No strict-feasibility margin is required. The expected-regret guarantee holds for each fixed $u \in \mathcal{U}$, not for a hindsight action selected using feedback noise. For deterministic oblivious functions this includes the best fixed feasible action, which is independent of observation noise. Define

$$\mathrm{Reg}_n(u) = \sum_{t=1}^{n} \big(\ell_t(x_t) - \ell_t(u)\big), \qquad \mathrm{WCV}_n = \max\left\{0, \max_{1 \leq s \leq e \leq n} \sum_{t=s}^{e} v_t\right\}, \tag{4}$$

$$\mathrm{ICV}_n = \sum_{t=1}^{n} [v_t]_+, \qquad \mathrm{CCV}_n = \left[\sum_{t=1}^{n} v_t\right]_+.$$

These are regret, budget violation, hard violation, and terminal violation, respectively. Always

$$\mathrm{CCV}_n \leq \mathrm{WCV}_n \leq \mathrm{ICV}_n. \tag{5}$$

All three violation metrics coincide when every $v_t \geq 0$. Our budget guarantee bounds the *expectation of the maximum* over windows, rather than a maximum of expectations or a deterministic safety limit. The classical Lindley recursion (Lindley, 1952), $\Lambda_0 = 0$, $\Lambda_t = [\Lambda_{t-1} + v_t]_+$, satisfies $\mathrm{WCV}_n = \max_{t \leq n} \Lambda_t$. Here Λ_t is the largest nonnegative sum over windows ending at t: reflection discards a negative accumulated balance so that a new window can start. The learner observes noisy values, so it maintains a clipped, regularized version of this balance.

Feasibility and local comparison. Convexity gives the affine minorant

$$\widehat{\gamma}_t(z) = v_t + \langle q_t, z - x_t \rangle \leq \gamma_t(z), \qquad A_t(z) = \tilde{v}_t + \langle \tilde{q}_t, z - x_t \rangle. \tag{6}$$

For every $\mathcal{F}_t$-measurable $z \in \mathcal{K}$, the estimation error $A_t(z) - \widehat{\gamma}_t(z)$ has conditional mean zero. At the played action, $A_t(x_t) = \tilde{v}_t$; differences satisfy $A_t(x_t) - A_t(z) = \langle \tilde{q}_t, x_t - z \rangle$. Thus value noise affects the balance update, whereas direction noise affects the primal comparison. This is the supporting-hyperplane inequality (Boyd and Vandenberghe, 2004, Sec. 3.1), applied before any nonlinear transformation of the observed value.

Lemma 2.1 (A bounded level for analysis only)**.** *Set $L = D\widehat{G}$ and $v_t^\circ = \max\{v_t, -L\}$. For every fixed $u \in \mathcal{U}$,*

$$-L \leq v_t^\circ \leq L, \qquad v_t \leq v_t^\circ \leq \langle q_t, x_t - u \rangle. \tag{7}$$

Proof. Feasibility and convexity give $v_t \leq \langle q_t, x_t - u \rangle$. Conditional Jensen gives $\|q_t\| \leq \widehat{G}$, so the inner product lies in $[-L, L]$. Taking the maximum of v_t and $-L$ proves the claim. □

In particular $v_t \leq L$, although v_t may be arbitrarily negative. The algorithm does not observe or use v_t°. Its role is to bound potential drift by monotonicity while retaining the same primal comparison. This removes any need for a known two-sided level bound.

3 Statistical limits with noisy values

All lower bounds use $\mathcal{K} = [0, 1]$, $\ell_t(x) = -x$, exact gradients, and independent $N(0, \sigma^2)$ value noise, with $D = G = \widehat{G} = 1$ and $|\gamma_t(x)| \leq 1$. The adversary is oblivious. Expectations in the hard-violation construction also include constraints drawn before play; Appendix B gives the deterministic-instance extraction and complete proofs.

3.1 Hard violation

For $\gamma_t(x) = x - \frac{1}{2} + \zeta_t$, the largest always-feasible action depends on the upper support endpoint of ζ_t, not its mean. Gaussian smoothing makes laws with different endpoints but many matching moments difficult to distinguish.

Theorem 3.1 (Hard-violation lower bound)**.** *For every $\sigma > 0$, integer $n \geq 1$, and horizon $T \geq 1$, set*

$$s = \min\left\{\frac{1}{2}, \frac{\sigma}{2}, \sigma\left(\frac{(2n)!}{12T}\right)^{1/(4n)}\right\}. \tag{8}$$

There are explicit symmetric laws P_n, P_n' supported in $[-s, s]$, with an action u_n feasible on every round under P_n, such that every learner satisfies

$$\max\{\mathbb{E}_{P_n}\mathrm{Reg}_T(u_n), \mathbb{E}_{P_n}\mathrm{ICV}_T, \mathbb{E}_{P_n'}\mathrm{ICV}_T\} \geq \frac{sT}{256n^2}. \tag{9}$$

For fixed n, σ and all sufficiently large T, this is at least $\sigma T^{1-1/(4n)}/(512n^2)$. Choosing the smallest n such that $(2n)!16^n \geq 12T$ gives

$$\Omega\left(\min\{\sigma, 1\} T \left(\frac{\log\log T}{\log T}\right)^2\right), \tag{10}$$

and hence $\Omega(\min\{\sigma, 1\}T/\log^2 T)$, with absolute implied constants.

Chebyshev interpolation matches $2n-1$ moments while keeping a support gap of order s/n^2 and constant endpoint masses. The weighted payoff inequality in Lemma B.3 accounts for negative regret without shrinking the testing tolerance with the gap. This excludes every joint $O(T^{1-\delta})$ regret–hard-violation guarantee for fixed $\delta > 0$, but does not exclude all sublinear rates. Appendix B.3 gives the construction.

3.2 Budget violation

For budget violation, consider constant affine constraints

$$\gamma_t^A(x) = \rho(x-1), \qquad \gamma_t^B(x) = \rho x, \qquad 0 < \rho \le 1. \tag{11}$$

The optimal feasible comparators are 1 in A and 0 in B, with identical constraint gradients. The transformed observations $\tilde{v}_t - \rho x_t$ have action-independent Gaussian laws with means $-\rho$ and 0.

Proposition 3.2 (Regret–budget tradeoff). *Fix $\sigma > 0$. Suppose a learner has uniform expected regret and budget bounds R_T and $\mathcal{B}_T$ over the normalized instance class above. For $0 < \mathcal{B}_T \le T/2$,*

$$R_T \ge \frac{T}{4}\exp\left(-\frac{4\mathcal{B}_T^2}{\sigma^2 T}\right). \tag{12}$$

When $0 < R_T < T/4$, this implies

$$\mathcal{B}_T \ge \frac{\sigma}{2}\sqrt{T\log\left(\frac{T}{4R_T}\right)}. \tag{13}$$

The same statements hold with terminal violation in place of budget violation.

Proof mechanism. Write $r_A = \mathbb{E}_A \mathrm{Reg}_T(1)$ and $b_B = \mathbb{E}_B \mathrm{WCV}_T$. The likelihood ratio $J = dP_B/dP_A$ of the transformed inputs, including learner randomization, satisfies $\mathbb{E}_A J^2 = e^{T\rho^2/\sigma^2}$. Set $Z = T^{-1}\sum_t (1-x_t) \in [0,1]$. Then $\mathbb{E}_A Z = r_A/T$, whereas $\mathbb{E}_B Z = 1 - b_B/(\rho T)$ because $\mathrm{WCV}_T = \rho \sum_t x_t$ in B. Changing measure gives $\mathbb{E}_B Z = \mathbb{E}_A[JZ]$; Cauchy–Schwarz and $\mathbb{E}_A Z^2 \le \mathbb{E}_A Z$ yield

$$1 - \frac{b_B}{\rho T} = \mathbb{E}_A[JZ] \le e^{T\rho^2/(2\sigma^2)}\sqrt{r_A/T}. \tag{14}$$

Set $\rho = 2\mathcal{B}_T/T$. The budget guarantee makes the left side at least $1/2$; squaring and rearranging proves (12). Appendix B.6 provides the likelihood calculation and the cases needed for the following corollary. □

Corollary 3.3 (Matching joint order and the square-root endpoint). *For each $T \ge 2$ and $\sigma > 0$, a pair of instances of the form (11) satisfies, for every learner,*

$$\max\{\mathbb{E}_A \mathrm{Reg}_T(1), \mathbb{E}_B \mathrm{WCV}_T\} \ge \frac{\min\{\sigma, 1\}}{4}\sqrt{T\log T}. \tag{15}$$

Under a uniform $R_T = O(\sqrt{T})$ guarantee, Proposition 3.2 requires $\mathcal{B}_T = \Omega(\sigma\sqrt{T\log T})$ for fixed $\sigma > 0$. Conversely, a uniform bound $\mathcal{B}_T \le C_B\sqrt{T}$, with fixed $C_B > 0$, forces $R_T \ge \frac{1}{4}e^{-4C_B^2/\sigma^2}T$ for all sufficiently large T.

These are uniform worst-case statements: ρ may depend on T, but the declared bounds remain fixed. Since all three violation metrics coincide in B, the logarithm is not caused by maximizing over windows. Theorem 4.1 matches this horizon dependence for the stronger window metric. Additional tradeoff and oracle consequences appear in Appendix B.7.

4 Achievable guarantees with signed balances

LEDGER achieves the optimal-order budget guarantee under square-root regret and finite-variance feedback. Algorithm 1 specifies the complete update, including all parameter choices, from the horizon and known problem bounds.

4.1 A REFLECTED EXPONENTIAL UPDATE

An unbiased value estimate need not remain unbiased after taking a positive part: at $v_t = 0$ with Gaussian noise, $\mathbb{E}\,[\tilde{v}_t]_+ = \sigma/\sqrt{2\pi}$. A constraint weight chosen using the current noisy value can also correlate with the current direction error. We instead accumulate signed values and choose all weights and step sizes before current feedback. Set

$$L = D\widehat{G}, \qquad B_v = (L^2 + \sigma^2)^{1/2} \geq L, \qquad \Phi_\alpha(q) = e^{\alpha q} - \alpha q - 1 \quad (q \geq 0). \tag{16}$$

Here L bounds the absolute one-round loss difference. Crucially, $\Phi'_\alpha(0) = 0$: reflection at zero has a quadratic, rather than first-order, potential cost. Clipping the value channel bounds positive queue increments without requiring exponential moments of the raw observations.

Algorithm 1: LEDGER: predictably regularized signed balance

Input: Horizon T, decision set $\mathcal{K}$, and known bounds $D, \widehat{G}, \sigma$.

$L \leftarrow D\widehat{G}$; $B_v \leftarrow \sqrt{L^2 + \sigma^2}$; $\alpha \leftarrow \sqrt{\log(eT)}/(64B_v\sqrt{T})$
$Q_0 \leftarrow 0$, $H_0 \leftarrow \alpha^2$; choose $x_1 \in \mathcal{K}$
for $t = 1, \ldots, T$ **do**
  $w_t \leftarrow \alpha(e^{\alpha Q_{t-1}} - 1)$; $H_t \leftarrow H_{t-1} + w_t^2$
  $\eta_t \leftarrow D/(2\widehat{G}\sqrt{t + H_t})$; $r_t \leftarrow 16\alpha B_v^2 + 4Lw_t/(\sqrt{H_t} + \sqrt{H_{t-1}})$
  play x_t; observe $(\tilde{p}_t, \tilde{q}_t, \tilde{v}_t)$
  $\bar{v}_t \leftarrow \max\{-1/\alpha, \min\{\tilde{v}_t, 1/\alpha\}\}$
  $x_{t+1} \leftarrow \Pi_{\mathcal{K}}\big(x_t - \eta_t(\tilde{p}_t + w_t\tilde{q}_t)\big)$
  $Q_t \leftarrow [Q_{t-1} + \bar{v}_t - r_t]_+$

The algorithm clips only the value channel; the directions remain unclipped. Its prescribed α satisfies $0 < \alpha \leq (64B_v)^{-1}$ because $\log(eT) \leq T$. All weights, step sizes, and regularizers are fixed before current feedback. Here Q_t tracks regularized overspend, $w_t = \Phi'_\alpha(Q_{t-1})$ converts that balance into a constraint penalty, and H_t accumulates squared penalties to scale the primal step. The subtraction of r_t stabilizes the queue; its total is charged back when bounding true overspend in (25). Each round uses one noisy feedback triple and one projection onto the known set $\mathcal{K}$, not a current feasible set. The adaptive part of r_t follows the regularization mechanism of Yu et al. (2026); exponential potentials in exact-feedback COCO were used by Sinha and Vaze (2024). Here reflection with zero boundary derivative, value clipping, and predictable primal step sizes make the method compatible with dependent finite-variance feedback.

Why the regularizer has two parts. The constant $16\alpha B_v^2$ pays for the clipping bias and the weight-dependent curvature cost in the potential drift. It cannot by itself cancel the adaptive primal cost, which grows with $\sqrt{H_t}$. The second part is chosen so that multiplication by $w_t/2$ gives exactly $2L(\sqrt{H_t} - \sqrt{H_{t-1}})$; this is the cancellation in (24). At $w_t = 0$ it vanishes, and $H_0 = \alpha^2 > 0$ keeps the denominator nonzero. Neither term is a relaxation of the actual constraint: the full sum of r_t is charged back in the true-window certificate. Thus the proof must control the regularizer twice, once weighted for regret and once unweighted for budget violation.

4.2 SQUARE-ROOT REGRET AND OPTIMAL-ORDER BUDGET CONTROL

Theorem 4.1 (LEDGER: finite-variance regret and budget guarantee)**.** *Assume conditional unbiasedness, (3), and $\mathcal{U} \neq \emptyset$. Set L, B_v as in Algorithm 1; the inequalities below hold for any fixed $0 < \alpha \leq (64B_v)^{-1}$. For every fixed $u \in \mathcal{U}$ and every $n \leq T$,*

$$\mathbb{E}\mathrm{Reg}_n(u) + \mathbb{E}\Phi_\alpha(Q_n) + 8\alpha B_v^2 \mathbb{E}\sum_{t=1}^{n} w_t \leq 2L\sqrt{n} + 2L\alpha + 60\alpha^2 B_v^2 n. \tag{17}$$

Let $K_0 = 4 + 6L$ and $\Lambda_T = \log((T+1)K_0T)$. Then

$$\mathbb{E}\mathrm{WCV}_T \leq \frac{\Lambda_T}{\alpha} + 17\alpha B_v^2 T + 4L\sqrt{2T\Lambda_T} + 4\sigma\sqrt{T}. \tag{18}$$

For the algorithm's prescribed parameters,

$$\mathbb{E}\mathrm{Reg}_T(u) \leq 2L\sqrt{T} + \frac{1}{32} + \frac{15}{1024}\log(eT) = O(\sqrt{T}), \qquad \mathbb{E}\mathrm{WCV}_T = O\Big(\sqrt{T\log(eT)}\Big). \tag{19}$$

The constants depend only on $(\widehat{G}, D, \sigma)$, not on an absolute level bound. The bounds allow an adaptive adversary and dependent feedback channels, without a Slater condition.

The regret rate has no multiplicative logarithm. Together with Proposition 3.2, the budget rate has optimal T-dependence, including the logarithm, among algorithms with uniform $O(\sqrt{T})$ regret for fixed positive Gaussian value noise and fixed problem bounds.

4.3 Proof design: what changes under noisy values

Technical novelty. The technical contribution is a coupling that turns finite-variance signed feedback into an exponential certificate for true window overspend while preserving $O(\sqrt{T})$ regret. Two features make this possible. First, subtracting the tangent at zero from the exponential makes reflection compatible with conditional centering: the linear drift is the predictable comparison term $w_t v_t^\circ$, while boundary crossings enter only through a quadratic remainder. Clipping at $1/\alpha$ converts that remainder and the clipping bias into terms of order $\alpha B_v^2 w_t + \alpha^2 B_v^2$. The constant part of r_t absorbs the weight-dependent term, leaving an $O(\alpha^2 B_v^2 T)$ contribution to regret, which is only logarithmic at our choice of α.

Second, the adaptive part of r_t links primal cancellation to the cost of regularizing the balance. Multiplication by w_t turns it into an exact increment of $\sqrt{H_t}$; its unweighted sum is controlled by $\sqrt{T \log(H_T/\alpha^2)}$. The same exponential-moment estimate controls both this logarithmic cost and the running queue maximum. This closes the argument: the regularization that makes the regret proof work also costs only $O(\sqrt{T \log(eT)})$ in the true budget certificate. Both weighted cancellation and the logarithmic-accumulator argument adapt Yu et al. (2026, Alg. 1 and Lemmas 10–11). Our contribution is to make them compatible with noisy signed reflection and to transfer the resulting control to true windows. Appendix C gives the complete proof.

Primal comparison and negative levels. Lemma 2.1 gives $v_t^\circ = \max\{v_t, -L\} \le \langle q_t, x_t - u\rangle$. For analysis, set $\tilde{v}_t^\circ = v_t^\circ + \varepsilon_t$ and $\bar{v}_t^\circ = \mathrm{clip}(\tilde{v}_t^\circ, -1/\alpha, 1/\alpha)$. Retaining the same error gives $\bar{v}_t \le \bar{v}_t^\circ$ and $\mathbb{E}[(\tilde{v}_t^\circ)^2 \mid \mathcal{F}_t] \le B_v^2$; the learner still uses the actual observation. Raising levels below $-L$ to $-L$ preserves the primal comparison because $\langle q_t, x_t - u\rangle \ge -L$, and monotonicity makes the reflected drift larger. Thus the known geometry, direction bound, and noise variance control the second moment needed in the potential proof. For $d_t = \tilde{p}_t + w_t \tilde{q}_t$, predictability gives $\mathbb{E}[d_t \mid \mathcal{F}_t] = p_t + w_t q_t$ and $\mathbb{E}[\|d_t\|^2 \mid \mathcal{F}_t] \le 2\widehat{G}^2(1 + w_t^2)$. The latter bound follows from the squared triangle inequality even when the channels are dependent. To see why the primal update has the required adaptive cost, put $S_t = t + H_t$, so $S_t - S_{t-1} = 1 + w_t^2$. The nonincreasing step sizes and projection onto $\mathcal{K}$ give the pathwise inequality

$$\sum_{t=1}^{n} \langle d_t, x_t - u\rangle \le \frac{D^2}{2\eta_n} + \frac{1}{2}\sum_{t=1}^{n} \eta_t \|d_t\|^2. \tag{20}$$

Predictability permits conditioning before multiplying the direction moment bound by η_t. Since $\sum_{t\le n}(S_t - S_{t-1})/\sqrt{S_t} \le 2(\sqrt{S_n} - \alpha)$, the expected right side is at most $2L\mathbb{E}\sqrt{S_n}$. Convexity and conditional unbiasedness bound the expected left side below by $\mathbb{E}\mathrm{Reg}_n(u) + \mathbb{E}\sum_{t\le n} w_t v_t^\circ$. Using $\sqrt{n + H_n} \le \sqrt{n} + \sqrt{H_n}$ therefore yields

$$\mathbb{E}\mathrm{Reg}_n(u) + \mathbb{E}\sum_{t\le n} w_t v_t^\circ \le 2L\sqrt{n} + 2L\mathbb{E}\sqrt{H_n}. \tag{21}$$

Reflection, clipping, and cancellation. Extend the potential by $F(q) = \Phi_\alpha([q]_+)$ on $\mathbb{R}$. Because $\Phi'_\alpha(0) = 0$, this extension is continuously differentiable at zero. Its almost-everywhere second derivative is zero for $q < 0$ and $\alpha^2 e^{\alpha q}$ for $q > 0$. Thus, for $q \ge 0$ and $\alpha z \le 1$, Taylor's integral remainder gives

$$F(q + z) - F(q) \le \alpha(e^{\alpha q} - 1)z + \tfrac{3}{2}\alpha^2 e^{\alpha q} z^2. \tag{22}$$

Indeed, the second derivative along the segment is at most $e\alpha^2 e^{\alpha q}$, and $e/2 < 3/2$. This also covers steps crossing zero: reflection contributes through the quadratic remainder.

Monotonicity of F allows us to use $q = Q_{t-1}$ and $z = \bar{v}_t^\circ - r_t$ in (22) to bound the actual drift. Clipping and $r_t \ge 0$ ensure $\alpha z \le 1$. Moreover, $|\mathbb{E}[\bar{v}_t^\circ \mid \mathcal{F}_t] - v_t^\circ| \le \alpha B_v^2$ and $\mathbb{E}[(\bar{v}_t^\circ)^2 \mid \mathcal{F}_t] \le B_v^2$.

Using $\alpha^2 e^{\alpha Q_{t-1}} = \alpha(w_t + \alpha)$ gives

$$\begin{aligned}\mathbb{E}[\Phi_\alpha(Q_t) - \Phi_\alpha(Q_{t-1}) \mid \mathcal{F}_t] &\le w_t v_t^\circ - w_t r_t + 4\alpha B_v^2 w_t + 3\alpha w_t r_t^2 + 3\alpha^2(B_v^2 + r_t^2) \\ &\le w_t v_t^\circ - \tfrac{1}{2} w_t r_t + 60\alpha^2 B_v^2.\end{aligned} \tag{23}$$

The second line uses $4\alpha B_v^2 \le r_t/4$, $3\alpha r_t \le 51/256 < 1/4$, and $r_t \le 17B_v/4$. Thus half of the regularization absorbs clipping bias and the weight-dependent remainder. The remaining half obeys the exact identity

$$\tfrac{1}{2} w_t r_t = 8\alpha B_v^2 w_t + 2L(\sqrt{H_t} - \sqrt{H_{t-1}}). \tag{24}$$

Summing (23) and combining with (21) cancels $2L\mathbb{E}\sqrt{H_n}$, leaving the initial term $2L\sqrt{H_0} = 2L\alpha$ and proving (17).

True maximum-window violation. The lower bound $\mathrm{Reg}_n(u) \ge -Ln$ and (17) imply $\mathbb{E}e^{\alpha Q_n} \le K_0 T$. For $M_T = \max_{0\le n\le T} Q_n$, the pathwise inequality $e^{\alpha M_T} \le \sum_{n=0}^T e^{\alpha Q_n}$ and Jensen's inequality give $\mathbb{E}M_T \le \Lambda_T/\alpha$. For every window $[s, e]$, the reflected update gives $\sum_{t=s}^e \bar{v}_t \le Q_e - Q_{s-1} + \sum_{t=s}^e r_t$. Substituting $v_t = \bar{v}_t + (\tilde{v}_t - \bar{v}_t) - \varepsilon_t$ yields the simultaneous pathwise certificate

$$\mathrm{WCV}_T \le M_T + \sum_{t=1}^T r_t + \sum_{t=1}^T [\tilde{v}_t - \bar{v}_t]_+ + 2 \max_{0\le n\le T} \left| \sum_{t=1}^n \varepsilon_t \right|. \tag{25}$$

Negative clipping residuals reduce overspend and need not be charged. The positive residual is bounded by $[\tilde{v}_t - \bar{v}_t]_+ \le \alpha(\tilde{v}_t^\circ)^2$, so its expected total is at most $\alpha B_v^2 T$. Doob's inequality bounds the noise term by $4\sigma\sqrt{T}$.

Paying for the regularizer. The adaptive part of r_t is controlled by the logarithmic growth of H_t:

$$\sum_{t=1}^T \frac{w_t}{\sqrt{H_t} + \sqrt{H_{t-1}}} \le \sqrt{T \sum_{t=1}^T \frac{H_t - H_{t-1}}{H_t}} \le \sqrt{T \log \frac{H_T}{\alpha^2}}. \tag{26}$$

The first step uses Cauchy–Schwarz and $H_t - H_{t-1} = w_t^2$; the second uses $(b-a)/b \le \log(b/a)$. Since $H_T/\alpha^2 = 1 + \sum_{t=1}^T (w_t/\alpha)^2 \le (1 + \sum_{t=1}^T w_t/\alpha)^2$, the exponential-moment bound and Jensen's inequality give $\mathbb{E}\log(H_T/\alpha^2) \le 2\Lambda_T$. Consequently, $\mathbb{E}\sum_{t=1}^T r_t \le 16\alpha B_v^2 T + 4L\sqrt{2T\Lambda_T}$. Combining these costs in (25) proves (18).

Error accounting and the logarithm. The certificate separates three costs besides the running balance: clipping contributes at most $\alpha B_v^2 T$, regularization contributes $16\alpha B_v^2 T + 4L\sqrt{2T\Lambda_T}$, and the value-noise martingale contributes $4\sigma\sqrt{T}$. Thus the raw noise partial sums need no union bound over windows. Choosing α of order $\sqrt{\log(eT)}/(B_v\sqrt{T})$ balances the queue cost Λ_T/α with the terms linear in αT; the master inequality then charges only $O(\log(eT))$ additional regret. Although the queue-maximum argument also introduces a logarithm, eliminating that step alone cannot improve the worst-case budget rate: Proposition 3.2 already forces the same logarithmic order for terminal violation.

4.4 Removing the horizon input

Corollary 4.2 (Unknown horizon)**.** *Run Algorithm 1 in consecutive epochs of planned lengths* $1, 2, 4, \ldots$*, restarting its state and using each epoch length in place of* T*. For every deterministic horizon* T *and every fixed comparator feasible on its first* T *rounds, the resulting algorithm satisfies* $\mathbb{E}\mathrm{Reg}_T(u) = O(\sqrt{T})$ *and* $\mathbb{E}\mathrm{WCV}_T = O(\sqrt{T\log(eT)})$ *under the same assumptions.*

Every window intersects each epoch in at most one window. Summing the geometric epoch bounds, including the unfinished final epoch, preserves both rates (Appendix D). This removes the horizon input.

4.5 Consequences for the violation metric

As a direct consequence of the preceding theorems, terminal and maximum-window violation have the same joint minimax order. Fix $\sigma > 0$ and let $\mathfrak{C}_\sigma$ be the instances on $[0,1]$ satisfying Section 2 with declared bounds $D = G = \widehat{G} = 1$ and $|\gamma_t(x)| \le 1$. For $V_T \in \{\mathrm{CCV}_T, \mathrm{WCV}_T\}$, define

$$\mathfrak{M}_T(V) = \inf_{\mathcal{A}} \sup_{I \in \mathfrak{C}_\sigma} \sup_{u \in \mathcal{U}(I)} \max\{\mathbb{E}_I \mathrm{Reg}_T^{\mathcal{A}}(u), \mathbb{E}_I V_T^{\mathcal{A}}\},$$

where the infimum is over learners with the stated local oracle access. Then

$$\mathfrak{M}_T(\mathrm{CCV}) = \Theta_\sigma(\sqrt{T \log T}), \qquad \mathfrak{M}_T(\mathrm{WCV}) = \Theta_\sigma(\sqrt{T \log T}). \tag{27}$$

Indeed, $\mathrm{CCV}_T \le \mathrm{WCV}_T$ and Theorem 4.1 give both upper bounds. Corollary 3.3 gives both lower bounds because the two metrics coincide in its violating instance. The subscript records that the comparison is for fixed positive noise, not a claim of optimal dependence on σ.

This corollary is not an additional algorithmic result: it records that LEDGER controls the stronger metric at the same joint minimax order. A small terminal sum alone can still conceal a large intervening deficit for a particular learner.

Scope of the guarantee. The resource interpretation is temporary overspend, not per-round safety. The bound is in expectation and scales with the full horizon; it does not assert a separate square-root bound in each window's length. The comparator must be fixed and almost surely feasible on every round, not a noise-dependent hindsight optimum. Finally, the upper bound uses known moment envelopes and depends on $B_v = \sqrt{L^2 + \sigma^2}$, whereas the Gaussian budget obstruction scales with σ. Matching the horizon exponent and logarithm therefore leaves open a sharp interpolation to vanishing or zero value noise. The prescribed constants are conservative; Appendix C.6 explains why the asymptotic guarantee is not a claim of validated finite-horizon performance.

5 Conclusion

Noisy values separate hard feasibility from budget control. A nearly linear lower bound rules out jointly polynomially sublinear regret and hard violation. For maximum-window budget violation, LEDGER achieves $O(\sqrt{T})$ expected regret and $O(\sqrt{T \log(eT)})$ expected violation under conditional finite variance and common feasibility, without a Slater margin or absolute level bound. Gaussian testing proves the optimal budget horizon dependence, including its logarithm, even for terminal violation. Thus changing from terminal accounting to maximum-window accounting does not increase the joint minimax order at fixed positive noise, whereas charging each positive violation creates a nearly linear obstruction. All bounds concern true played values, and the window maximum is taken before expectation.

Reproducibility statement

Algorithm 1 specifies the complete update and parameter choices. Appendices B–D give the explicit lower-bound constructions, finite-variance analysis, and unknown-horizon proof.

AI use statement

Generative AI tools were used in manuscript revision. The authors are responsible for the correctness, originality, and final presentation of the work.

REFERENCES

S. Agrawal and N. R. Devanur. Bandits with concave rewards and convex knapsacks. In *Proceedings of the fifteenth ACM conference on Economics and computation*, 2014.

A. Badanidiyuru, R. Kleinberg, and A. Slivkins. Bandits with knapsacks. *Journal of the ACM (JACM)*, 2018.

H. Balasundaram, K. K. Mahendran, and R. Vaze. Breaking the $O(\sqrt{T})$ cumulative constraint violation barrier while achieving $O(\sqrt{T})$ static regret in constrained online convex optimization. *arXiv preprint arXiv:2603.20671*, 2026.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

M. Castiglioni, A. Celli, A. Marchesi, G. Romano, and N. Gatti. A unifying framework for online optimization with long-term constraints. In *Advances in Neural Information Processing Systems*, 2022.

S. Chaudhary and D. Kalathil. Safe online convex optimization with unknown linear safety constraints. In *Proceedings of the AAAI conference on artificial intelligence*, 2022.

J. L. Doob. *Stochastic Processes*. Wiley, 1953.

J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 1991.

H. Guo, X. Liu, H. Wei, and L. Ying. Online convex optimization with hard constraints: Towards the best of two worlds and beyond. In *Advances in Neural Information Processing Systems*, 2022.

N. Immorlica, K. Sankararaman, R. Schapire, and A. Slivkins. Adversarial bandits with knapsacks. *Journal of the ACM*, 2022.

R. Jenatton, J. Huang, and C. Archambeau. Adaptive algorithms for online convex optimization with long-term constraints. In *International Conference on Machine Learning*, 2016.

J. Lekeufack and M. I. Jordan. An optimistic algorithm for online convex optimization with adversarial constraints. *arXiv preprint arXiv:2412.08060*, 2024.

N. Liakopoulos, A. Destounis, G. Paschos, T. Spyropoulos, and P. Mertikopoulos. Cautious regret minimization: Online optimization with long-term budget constraints. In *International Conference on Machine Learning*, 2019.

D. V. Lindley. The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48:277–289, 1952.

Y. Lu, M. Pedramfar, M. Wang, and V. Aggarwal. BAGEL: Adversarially constrained online convex optimization under separation oracle access. *arXiv preprint arXiv:2502.16744v3*, 2026.

M. Mahdavi, R. Jin, and T. Yang. Trading regret for efficiency: online convex optimization with long term constraints. *The Journal of Machine Learning Research*, 2012.

S. Mannor, J. N. Tsitsiklis, and J. Y. Yu. Online learning with sample path constraints. *Journal of Machine Learning Research*, 2009.

M. J. Neely and H. Yu. Online convex optimization with time-varying constraints. *arXiv preprint arXiv:1702.04783*, 2017.

H. Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 1955.

D. Sarkar and A. Sinha. A geometric approach to constrained online learning. *arXiv preprint arXiv:2605.21107*, 2026.

A. Sinha. BanditQ: Fair bandits with guaranteed rewards. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.

A. Sinha and R. Vaze. Optimal algorithms for online convex optimization with adversarial constraints. In *Advances in Neural Information Processing Systems*, 2024.

A. Sinha and R. Vaze. Beyond $\widetilde{O}(\sqrt{T})$ constraint violation for online convex optimization with adversarial constraints. In *Advances in Neural Information Processing Systems*, 2025.

F. E. Stradi, M. Castiglioni, A. Marchesi, N. Gatti, and C. Kroer. No-regret learning under adversarial resource constraints: A spending plan is all you need! In *Advances in Neural Information Processing Systems*, 2025.

L. N. Trefethen. *Approximation Theory and Approximation Practice*. Society for Industrial and Applied Mathematics, 2013.

A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, 2009.

I. Usmanova, A. Krause, and M. Kamgarpour. Safe convex learning under uncertain constraints. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 2019.

R. Vaze and A. Sinha. $O(\sqrt{T})$ static regret and instance-dependent constraint violation for constrained online convex optimization. In *Advances in Neural Information Processing Systems*, 2025.

X. Wei, H. Yu, and M. J. Neely. Online primal-dual mirror descent under stochastic constraints. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2020.

Y. Wu and P. Yang. Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *The Annals of Statistics*, 2019.

X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and K. H. Johansson. Regret and cumulative constraint violation analysis for distributed online constrained convex optimization. *IEEE Transactions on Automatic Control*, 2023.

H. Yu and M. J. Neely. A low complexity algorithm with $O(\sqrt{T})$ regret and $O(1)$ constraint violations for online convex optimization with long-term constraints. *Journal of Machine Learning Research*, 2020.

H. Yu, M. Neely, and X. Wei. Online convex optimization with stochastic constraints. In *Advances in Neural Information Processing Systems*, 2017.

K. Yu, J. Lee, and D. Lee. Constrained online convex optimization without Slater's condition. *arXiv preprint arXiv:2606.31480*, 2026.

J. Yuan and A. Lamperski. Online convex optimization for cumulative constraints. In *Advances in Neural Information Processing Systems*, 2018.

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning*, 2003.

Exact-value step	Difficulty under noisy values	Construction used here
Positive-part increment	At $v_t = 0$, Gaussian noise gives $\mathbb{E}\,[\tilde{v}_t]_+ = \sigma/\sqrt{2\pi}$.	Signed increment, with a floor at zero.
Unreflected signed balance	Earlier negative values can hide a later positive window.	Reflect the accumulated balance, rather than each observation.
Exponential drift with a nonzero boundary derivative	Reflection creates a first-order noise cost, and raw finite-variance noise need not have exponential moments.	Use $\Phi'_\alpha(0) = 0$ and clip only the value channel.
Thresholded direction	A noisy gate is not the exact positive-part subgradient and can be correlated with direction noise.	Raw direction and a predictable weight.
Comparator feasibility	Conditional mean feasibility need not survive a positive part; direction noise also enters $A_t(u)$.	Compare through the affine constraint estimator.
Bounding earlier weights by the terminal value	A running balance need not be monotone, so its terminal value need not dominate earlier weights.	Bound exponential moments and then the running maximum.

Table 2: Proof obstacles when adapting exact-feedback queue methods to noisy values. The rows concern different constructions. Section 4.3 explains the changes.

A Related work in full

COCO with exact constraint observation. Online convex optimization with long-term constraints was introduced by Mahdavi et al. (2012), who traded regret against the cost of projecting onto a fixed constraint set and obtained $O(\sqrt{T})$ regret with $O(T^{3/4})$ violation of the *signed* long-term constraint; refinements for fixed constraints appear in Jenatton et al. (2016); Yuan and Lamperski (2018); Yu and Neely (2020). For time-varying constraints, Neely and Yu (2017) use Slater's condition. In contrast, the distributed time-varying-constraint guarantees of Yi et al. (2023) do not require Slater: their tradeoff gives $O(\sqrt{T})$ static network regret and $O(T^{3/4})$ network cumulative violation at parameter $\kappa = 1/2$. In the centralized adversarial setting, Guo et al. (2022) obtain $O(\sqrt{T})$ regret and $O(T^{3/4})$ hard violation under common feasibility, also without Slater. Sinha and Vaze (2024) gave the first near-minimax-optimal algorithm, $O(\sqrt{T})$ regret and $\widetilde{O}(\sqrt{T})$ hard violation, by running a Lipschitz-adaptive online subroutine on surrogate losses that combine the loss with the constraint through an exponential Lyapunov potential. Its dependence on exact constraint values motivates the use of signed constraint observations proposed here. Dimension- and instance-dependent refinements, optimistic algorithms, and projection-free interfaces have also been studied (Sinha and Vaze, 2025; Vaze and Sinha, 2025; Balasundaram et al., 2026; Lekeufack and Jordan, 2024; Lu et al., 2026). Using nested feasible-set projections and a self-contraction argument, Sarkar and Sinha (2026) obtain the exact $O(\sqrt{T})$ pair for general convex losses. These methods use exact values or feasible-set information not available in our model. For the positive-part queue constructions, Table 2 isolates why direct substitution of a noisy value does not preserve the proof. Theorem 3.1 rules out their joint polynomially sublinear regret–hard-violation rates under Gaussian value noise.

Stochastic and i.i.d. constraints. Mannor et al. (2009) showed that sublinear regret and violation need not be jointly attainable against adversarial constraints when the fixed comparator is required to be feasible only in the long run. Yu et al. (2017) studied i.i.d. constraints observed after the action and obtained $O(\sqrt{T})$ regret and violation via a virtual-queue primal–dual method under Slater's condition; Wei et al. (2020) weakened Slater to a bounded-Lagrange-multiplier assumption. More recently, Yu et al. (2026) obtain $O(\sqrt{T})$ expected regret and $O(\sqrt{T}\log T)$ expected cumulative violation for stochastic constraints without Slater's condition using adaptive dual regularization. These stochastic-constraint results use a fixed constraint distribution and a comparator feasible for its mean. Our constraints may vary adversarially, but the same comparator must be feasible on every round. The observation models also differ: we receive local noisy values and directions of the current constraint, rather than exact feedback on a sampled constraint. Yu et al. (2026, Thm. 5) also treat adversarial constraints, obtaining $O(\sqrt{T})$ regret and $O(\sqrt{T}\log T)$ hard violation with exact constraint feedback. That extension uses positive parts of exact values, rather than

noisy signed observations. Signed queue updates are already standard in this literature. LEDGER builds on the adaptive-regularization mechanism of Yu et al. (2026, Alg. 1 and Lemmas 10–11), replacing exact-value positive-part updates by a clipped signed reflected balance. Its zero-boundary-derivative potential and predictable primal step sizes permit dependent finite-variance feedback and yield $O(\sqrt{T\log(eT)})$ maximum-window violation with $O(\sqrt{T})$ regret. Liakopoulos et al. (2019) instead use windows of a fixed length for a comparator benchmark, not our learner-side metric. Sinha (2024) studies stochastic reward constraints with bandit feedback in a fair-allocation model.

Window accounting and its precedent. The maximum signed sum over contiguous windows and its reflected-queue representation appear in Sinha and Vaze (2024, Sec. 3), for online constraint satisfaction with exact feedback and no loss objective. We use this metric with adversarial losses and noisy local observations. Building on that accounting, the result here is a joint regret–violation guarantee under finite variance, with a matching noise-dependent obstruction. The expected maximum, rather than just an expected terminal balance, is the object controlled in (25).

Budgeted and knapsack formulations. Castiglioni et al. (2022) give a black-box framework for long-term-constraint problems covering stochastic and adversarial feedback, with a global comparator whose adversarial reward fraction depends on a strict-feasibility parameter. Stradi et al. (2025) show that a prescribed spending plan suffices in a setting where the plan prescribes expected use under time-varying reward–cost distributions; our allowance is absorbed into each round's adversarial constraint, with a roundwise-feasible comparator. Adversarial Bandits with Knapsacks (Badanidiyuru et al., 2018; Immorlica et al., 2022; Agrawal and Devanur, 2014) uses a different stopping-time and competitive-ratio formulation; our additive regret objective against a per-round-feasible comparator is not directly comparable.

Noise, deconvolution and lower bounds. Our hard-violation lower bounds are Le Cam two-point arguments (Tsybakov, 2009) in which the two worlds are matched to high order. In Proposition B.8, the first unmatched moment is the fourth, producing an eighth-order KL bound. The extension to arbitrary order in Theorem 3.1 is related to nonparametric deconvolution, where matching moments against a supersmooth (Gaussian) kernel produces logarithmic rather than polynomial rates (Fan, 1991). Chebyshev approximation and moment matching have also been used to construct statistically indistinguishable laws in other estimation problems (Wu and Yang, 2019). Here we interpolate at Chebyshev extrema and evaluate just outside their interval. The interpolation coefficients yield two explicit symmetric probability laws with matching moments, a support gap of order s/n^2, and endpoint masses bounded below by an absolute constant. A weighted payoff inequality converts that gap into an online regret–hard-violation obstruction. The interpolation identities are classical (Trefethen, 2013); the resulting distributions and online reduction are proved in Appendix B. For the budget metric, Proposition 3.2 instead uses a Gaussian likelihood-ratio second moment with constant affine constraints. Under uniform square-root regret, it forces $\Omega(\sigma\sqrt{T\log T})$ budget violation, even for the terminal metric, matching the T-dependence of Theorem 4.1 at fixed positive noise.

Safe learning. A related literature keeps a stochastic constraint satisfied at every round with high probability, using confidence bounds and a pessimistic feasible region: see Usmanova et al. (2019) for safe convex optimization with an unknown constraint and Chaudhary and Kalathil (2022) for online convex optimization with unknown linear safety constraints. Those results are for a *fixed* constraint function, where a pessimistic estimate can be built and refined; the adversarial case has no fixed function to estimate, and Theorem 3.1 excludes joint polynomially sublinear regret and hard violation in our adversarial noisy-value model.

What is, and is not, the contribution. Signed queues, exponential potentials, adaptive dual regularization, and Gaussian moment matching are established tools. In particular, Yu et al. (2026, Lemmas 10–11) provide both weighted regularizer cancellation and the logarithmic-accumulator bound on its unweighted sum. Our upper-bound contribution is their coupling under conditional finite variance: reflection has only a quadratic boundary cost, predictable coefficients avoid same-round noise correlations, and the regularizer can be paid for both in the weighted primal comparison and in the unweighted true-window certificate. The lower bounds distinguish two statistical obstructions within the same oracle model. Gaussian shift testing forces the budget logarithm, while explicit

moment-matched support endpoints force nearly linear hard violation. Neither result is a claim of hardness caused by projection complexity or missing derivative information.

B CONSTRUCTIONS AND PROOFS OF THE LOWER BOUNDS

The hard-violation argument combines Gaussian smoothing with a weighted payoff inequality. The smoothing calculation is a standard moment-matching method (Fan, 1991; Wu and Yang, 2019); the reduction to testing uses total variation and Pinsker's inequality (Tsybakov, 2009, Ch. 2). We give the distribution construction and every constant explicitly. Throughout, ϕ_a is the density of $N(a, \sigma^2)$, $\phi = \phi_0$, and $P * \phi$ is the density of $\zeta + \varepsilon$ for independent $\zeta \sim P$ and $\varepsilon \sim N(0, \sigma^2)$.

B.1 GAUSSIAN-MIXTURE BOUNDS

Lemma B.1 (Gaussian identity and mixture comparison)**.** *For all real a, b,*

$$\int_{\mathbb{R}} \frac{\phi_a(x)\phi_b(x)}{\phi_0(x)}\,dx = e^{ab/\sigma^2}. \tag{28}$$

Let $P = \sum_i w_i \delta_{sa_i}$ and $P' = \sum_i w_i' \delta_{sa_i}$ be probability laws on a common finite node set, write $c_i = w_i - w_i'$ and $M_k = \sum_i c_i a_i^k$, and suppose $P' * \phi \geq c\phi_0$ pointwise for $c > 0$. Then

$$\mathrm{KL}(P * \phi \| P' * \phi) \leq \frac{1}{c} \sum_{k=0}^{\infty} \frac{(s^2/\sigma^2)^k}{k!} M_k^2. \tag{29}$$

Proof. Completing the square gives $\phi_a \phi_b / \phi_0 = e^{ab/\sigma^2} \phi_{a+b}$, proving (28). If $p = P * \phi$ and $q = P' * \phi$, then

$$\mathrm{KL}(p\|q) \leq \chi^2(p\|q) \leq c^{-1} \int \frac{(p-q)^2}{\phi_0} = c^{-1} \sum_{i,j} c_i c_j e^{s^2 a_i a_j/\sigma^2}.$$

Expanding each exponential and summing the finite node indices proves (29). □

Lemma B.2 (Smoothing of symmetric moment-matched laws)**.** *Let P, P' be finite symmetric probability laws on $[-s, s]$, with equal moments of orders $0, 1, \ldots, 2n-1$. If $s \leq \sigma/2$, then*

$$\mathrm{KL}(P * \phi \| P' * \phi) \leq \frac{6}{(2n)!} (s/\sigma)^{4n}. \tag{30}$$

Proof. For $|a| \leq s$, pairing opposite Gaussian shifts gives

$$\tfrac{1}{2}(\phi_a(x) + \phi_{-a}(x)) = \phi_0(x) e^{-a^2/(2\sigma^2)} \cosh(ax/\sigma^2) \geq e^{-1/8} \phi_0(x).$$

A central atom also satisfies this lower bound, so $P' * \phi \geq e^{-1/8}\phi_0$. In the notation of Lemma B.1, $M_k = 0$ for $k < 2n$, and $|M_k| \leq 2$ because normalized nodes lie in $[-1, 1]$. With $u = s^2/\sigma^2 \leq 1/4$,

$$\mathrm{KL}(P * \phi \| P' * \phi) \leq 4e^{1/8} \sum_{k \geq 2n} \frac{u^k}{k!} \leq 4e^{1/8+u} \frac{u^{2n}}{(2n)!} \leq 6 \frac{u^{2n}}{(2n)!}.$$

□

B.2 A WEIGHTED PAYOFF INEQUALITY AND THE ONLINE REDUCTION

Lemma B.3 (Payoff with endpoint probabilities)**.** *Let P be supported in $[-a, a]$ and P' in $[-b, b]$, with $0 \leq a < b$, $w_a = P(\{a\}) > 0$, and $w_b = P'(\{b\}) > 0$. Define $V_P(z) = \mathbb{E}_{\zeta \sim P}[z + \zeta]_+$, $V_{P'}$ analogously, and $\Delta = b - a$. Then, for every $z \in \mathbb{R}$,*

$$-a - z + \frac{V_P(z)}{w_a} + \min\left\{ \frac{V_{P'}(z)}{w_b}, \Delta \right\} \geq \Delta. \tag{31}$$

No symmetry or absence of mass between a and b is needed for this inequality.

Proof. The endpoint atoms give $V_P(z) \geq w_a[z+a]_+$ and $V_{P'}(z) \geq w_b[z+b]_+$. Hence the left side of (31) is at least

$$[-a-z]_+ + \min\{[z+b]_+, \Delta\}.$$

This is at least Δ for $z \leq -b$, equals Δ for $-b \leq z \leq -a$, and equals Δ for $z \geq -a$. □

Lemma B.4 (Online observations and bounded testing statistics)**.** *Let $\gamma_t(x) = x - \frac{1}{2} + \zeta_t$ and $\ell_t(x) = -x$ on $[0,1]$, with $\zeta_{1:T}$ drawn independently from P before play, and with exact gradients and independent Gaussian value noise. Suppose P is supported in $[-a,a]$, $a \leq 1/2$. Writing $y_t = x_t - 1/2$, the following hold.*

1. *The action $u_P = 1/2 - a$ is feasible for every realization, and $\mathrm{Reg}_T(u_P) = \sum_t(-a - y_t)$.*
2. *The transformed inputs $z_t = \tilde{v}_t - y_t = \zeta_t + \varepsilon_t$ are i.i.d. with law $P * \phi$. For the learner's independent random seed U, the action sequence is the same measurable function of $(z_{1:T}, U)$ in both worlds, with y_t depending only on $(z_{1:t-1}, U)$.*
3. *$\mathbb{E}_P \mathrm{ICV}_T = \mathbb{E}_P \sum_t V_P(y_t)$.*
4. *For any statistic $\Xi(z_{1:T}, U) \in [0, K]$ and another law P',*

$$|\mathbb{E}_P \Xi - \mathbb{E}_{P'} \Xi| \leq K\sqrt{T\mathrm{KL}(P * \phi \| P' * \phi)/2}.$$

Proof. The first claim follows from $\gamma_t(u_P) = \zeta_t - a \leq 0$ and the linear loss. Given the strategy and its seed, the original feedback is recovered recursively from $\tilde{v}_t = z_t + y_t$. Thus the action-dependent shift adds no information, proving the second claim. Independence of the current level from the past action gives the third. Finally, the joint input laws are $(P * \phi)^{\otimes T} \otimes P_U$ and $(P' * \phi)^{\otimes T} \otimes P_U$. Their relative entropy is $T\mathrm{KL}(P * \phi \| P' * \phi)$; Pinsker's inequality and the range of Ξ prove the last claim. □

Proposition B.5 (Master hard-violation lower bound)**.** *Under the endpoint conditions of Lemma B.3, suppose $b \leq 1/2$ and $T\mathrm{KL}(P * \phi \| P' * \phi) \leq 1/2$. Every learner then satisfies*

$$\max\{\mathbb{E}_P \mathrm{Reg}_T(u_P), \mathbb{E}_P \mathrm{ICV}_T, \mathbb{E}_{P'} \mathrm{ICV}_T\} \geq \frac{(b-a)T}{2(1 + w_a^{-1} + w_b^{-1})}. \tag{32}$$

Proof. Set $R = \mathbb{E}_P \mathrm{Reg}_T(u_P)$, $V = \mathbb{E}_P \mathrm{ICV}_T$, $V' = \mathbb{E}_{P'} \mathrm{ICV}_T$, and

$$\Xi = \sum_{t=1}^{T} \min\{V_{P'}(y_t)/w_b, b-a\} \in [0, T(b-a)].$$

Lemma B.3 and the first three parts of Lemma B.4 give $R + V/w_a + \mathbb{E}_P \Xi \geq (b-a)T$. The input total variation is at most $1/2$, so $\mathbb{E}_P \Xi \leq \mathbb{E}_{P'} \Xi + (b-a)T/2 \leq V'/w_b + (b-a)T/2$. Therefore $R + V/w_a + V'/w_b \geq (b-a)T/2$. A weighted sum with these positive coefficients is at most $(1 + w_a^{-1} + w_b^{-1})$ times the largest term. This remains true if R is negative. □

B.3 Explicit moment-matched laws with substantial endpoint mass

We use the elementary Chebyshev identities $\mathrm{T}_m(\cos\theta) = \cos(m\theta)$ and $\mathrm{T}_m(\cosh t) = \cosh(mt)$, together with Lagrange interpolation (Trefethen, 2013). The following construction is explicit and does not require solving a moment problem.

For $n = 1$, use $P_1 = \delta_0$ and $P_1' = \frac{1}{2}\delta_{-s} + \frac{1}{2}\delta_s$. For $n \geq 2$, set

$$m = 2n-2, \quad c = \cosh(1/m), \quad a = s/c, \quad b = s, \quad \xi_j = \cos(j\pi/m) \quad (0 \leq j \leq m). \tag{33}$$

Define the Lagrange coefficients at the point $c > 1$ by

$$\lambda_j = \prod_{\substack{0 \leq k \leq m \\ k \neq j}} \frac{c - \xi_k}{\xi_j - \xi_k}, \qquad \beta_j = \frac{\lambda_j + \lambda_{m-j}}{2}, \qquad A = \frac{1 + \cosh 1}{2}. \tag{34}$$

The probability laws are

$$P_n = \frac{1}{A} \sum_{\substack{0 \le j \le m \\ j \text{ even}}} \beta_j \delta_{a\xi_j}, \tag{35}$$

$$P'_n = \frac{1}{A} \left(\frac{\delta_s + \delta_{-s}}{2} + \sum_{\substack{0 \le j \le m \\ j \text{ odd}}} (-\beta_j) \delta_{a\xi_j} \right). \tag{36}$$

Lemma B.6 (Validity, moments, endpoints, and gap)**.** *The laws P_n, P'_n above are symmetric probabilities with identical moments through order $2n-1$. Their maximum support points a_n, b_n satisfy*

$$b_n - a_n \ge \frac{s}{16n^2}, \qquad P_n(\{a_n\}) \ge \frac{1}{3}, \quad P'_n(\{b_n\}) \ge \frac{1}{3}. \tag{37}$$

For $n \ge 2$, $a_n = a$ and $b_n = s$; for $n = 1$, $a_1 = 0$ and $b_1 = s$.

Proof. The case $n = 1$ is immediate. Assume $n \ge 2$, so m is positive and even. The nodes ξ_j are strictly decreasing. In (34), every numerator is positive and exactly j denominator factors are negative, hence $\operatorname{sign}(\lambda_j) = (-1)^j$. Because m is even, λ_j and λ_{m-j} have the same sign, so $\operatorname{sign}(\beta_j) = (-1)^j$ and $\beta_j = \beta_{m-j}$.

Interpolation of the constant polynomial and of T_m gives

$$\sum_{j=0}^{m} \lambda_j = 1, \qquad \sum_{j=0}^{m} |\lambda_j| = \sum_{j=0}^{m} (-1)^j \lambda_j = \mathrm{T}_m(c) = \cosh 1.$$

Symmetrization preserves these two sums because paired coefficients have the same sign. Thus $\sum_j \beta_j = 1$, $\sum_j |\beta_j| = \cosh 1$, and their positive mass is A. The negative mass is $A - 1$. Equations (35)–(36) therefore specify nonnegative masses summing to one; symmetry follows from $\xi_{m-j} = -\xi_j$ and $\beta_{m-j} = \beta_j$.

For every polynomial p of degree at most m, interpolation and symmetrization give

$$\sum_{j=0}^{m} \beta_j p(a\xi_j) = \frac{p(ac) + p(-ac)}{2} = \frac{p(s) + p(-s)}{2}.$$

Rearranging this identity proves that P_n and P'_n have the same moments through degree $m = 2n-2$. Both are symmetric, so their moment of the next, odd degree $2n-1$ is zero as well.

The atom at a in P_n has mass $w_a = \beta_0/A$. Each factor in $\lambda_0 = \prod_{k=1}^{m} (c - \xi_k)/(1 - \xi_k)$ is greater than one. Also $\lambda_m > 0$, whence $\beta_0 \ge 1/2$ and $w_a \ge 1/(1 + \cosh 1) > 1/3$. The atom at s in P'_n has mass exactly $w_b = 1/(2A) = 1/(1 + \cosh 1) > 1/3$. These atoms establish the claimed support endpoints. Finally, $m \le 2n$, $\cosh(1/m) < 2$, and $\cosh t - 1 \ge t^2/2$ imply

$$s - a = s \frac{\cosh(1/m) - 1}{\cosh(1/m)} \ge \frac{s}{4m^2} \ge \frac{s}{16n^2}.$$

□

B.4 Proof of the hard-violation theorem

Proof of Theorem 3.1. Use (8) and the laws in (35)–(36), or the stated two-point construction for $n = 1$. Let ζ_t have the selected law and set $\gamma_t(x) = x - 1/2 + \zeta_t$. The comparator $u_n = 1/2 - a_n$ is feasible under P_n on every realization; the comparator $1/2 - s$ is feasible under P'_n. The levels lie in $[-1, 1]$ on $[0, 1]$ because $s \le 1/2$. Since $s \le \sigma/2$, Lemma B.2 and the chosen scale imply

$$T\mathrm{KL}(P_n * \phi \| P'_n * \phi) \le \frac{6T}{(2n)!} (s/\sigma)^{4n} \le \frac{1}{2}.$$

By Lemma B.6, $1 + w_a^{-1} + w_b^{-1} \le 7$ and $b_n - a_n \ge s/(16n^2)$. Proposition B.5 therefore gives the stronger finite-horizon bound $sT/(224n^2)$, and hence the stated $sT/(256n^2)$.

For fixed n, σ, the third term in (8) eventually attains the minimum. Moreover, $((2n)!/12)^{1/(4n)} \geq (1/6)^{1/(4n)} \geq 1/2$. Thus the lower bound is at least $\sigma T^{1-1/(4n)}/(512n^2)$.

For the growing-order bound, let $n(T)$ be the smallest positive integer satisfying $(2n)!16^n \geq 12T$. The third term in (8) is then at least $\sigma/2$, so $s = \min\{\sigma, 1\}/2$ and the lower bound is

$$\frac{\min\{\sigma, 1\}T}{512n(T)^2}. \tag{38}$$

For completeness, $n(T) \leq \lceil \log T \rceil + 1$: the inequality $(2r)!16^r \geq 12e^{r-1}$ holds at $r = 1$ and its left side grows by a factor of at least 192 when r increases by one, while the right side grows by e. This proves the $T/\log^2 T$ statement directly. Stirling's formula (Robbins, 1955) gives $\log((2n)!16^n/12) = 2n \log n + O(n)$. Taking $n = \lceil 2 \log T/\log\log T \rceil$ for sufficiently large T satisfies the required inequality. Hence $n(T) = O(\log T/\log\log T)$, proving the sharper (10). □

Remark B.7 (Randomized instances and deterministic extraction)**.** The instance distribution is oblivious: the entire level sequence is drawn before play. Only the current value is observed through noise. The comparator u_n is fixed and feasible for every realization in its world, so the expected-regret quantifier is the same as in the upper bound. Whichever expected quantity attains the lower-bound maximum has a realization of the level sequence whose conditional expectation is at least as large. This supplies a deterministic oblivious instance for each learner. The lower bound already grants the full loss and the exact constraint slope, since both are identical in the two worlds. It does not rely on concealing gradient information.

B.5 An illustrative mean- and variance-matched pair

Proposition B.8 (Fourth-moment separation)**.** *For $\sigma > 0$, $T \geq 32$, and $s = \min\{\sigma(8T)^{-1/8}, 1/4\}$, take $P = \frac{1}{2}\delta_{-s} + \frac{1}{2}\delta_s$ and $P' = \frac{1}{8}\delta_{-2s} + \frac{3}{4}\delta_0 + \frac{1}{8}\delta_{2s}$. Under the preceding linear-loss and affine-constraint model,*

$$\max\{\mathbb{E}_P \mathrm{Reg}_T(1/2 - s), \mathbb{E}_P \mathrm{ICV}_T, \mathbb{E}_{P'} \mathrm{ICV}_T\} \geq \frac{sT}{36} \geq \min\left\{\frac{\sigma T^{7/8}}{48}, \frac{T}{144}\right\}.$$

Proof. The laws agree through the third moment. On normalized nodes $-2, -1, 0, 1, 2$, their signed weights are $-1/8, 1/2, -3/4, 1/2, -1/8$. The fourth-moment difference is -3, all odd differences vanish, and for even $k \geq 6$ the absolute difference is at most $2^k/4$. Also $P' * \phi \geq (3/4)\phi_0$. Since $s \leq \sigma/2$ under the stated choice, $u = s^2/\sigma^2 \leq 1/4$, and Lemma B.1 gives

$$\begin{aligned}\mathrm{KL}(P * \phi \| P' * \phi) &\leq \frac{4}{3}\left(\frac{9}{24}u^4 + \frac{1}{16}\sum_{k \geq 6}\frac{(4u)^k}{k!}\right) \\ &\leq \frac{4}{3}\left(\frac{9}{24} + \frac{4096e}{16 \cdot 720 \cdot 16}\right)u^4 \leq 0.6u^4 \leq (s/\sigma)^8.\end{aligned}$$

Thus $T\mathrm{KL} \leq 1/8$. The endpoint parameters are $a = s$, $b = 2s$, $w_a = 1/2$, and $w_b = 1/8$. Proposition B.5 yields $sT/22$, which implies $sT/36$. Finally $36\,8^{1/8} < 48$, proving the displayed constants. For the bound $s \leq \sigma/2$, either $s = \sigma(8T)^{-1/8}$ and $T \geq 32$, or $s = 1/4 \leq \sigma(8T)^{-1/8}$, which gives $\sigma \geq 1/2$. □

B.6 The Gaussian budget tradeoff

Proof of Proposition 3.2. Use (11) with exact gradients $p_t = -1$, $q_t = \rho$ and independent Gaussian value noise. In world A, $u_A = 1$ is optimal and $\mathrm{Reg}_T(1) = \sum_t (1 - x_t) \geq 0$; in world B, $u_B = 0$ is the only feasible action and

$$\mathrm{WCV}_T = \mathrm{CCV}_T = \mathrm{ICV}_T = \rho \sum_t x_t.$$

The transformed inputs $Y_t = \tilde{v}_t - \rho x_t$ have laws $N(-\rho, \sigma^2)$ in A and $N(0, \sigma^2)$ in B, independently over rounds. The learner's actions are the same measurable function of the transformed history and

its seed in both worlds, because $\tilde{v}_t = Y_t + \rho x_t$ reconstructs the feedback. For their joint input laws, including the common seed distribution, the likelihood ratio is

$$J = \frac{dP_B}{dP_A} = \exp\left(\frac{\rho}{\sigma^2}\sum_{t=1}^{T} Y_t + \frac{T\rho^2}{2\sigma^2}\right), \qquad \mathbb{E}_A J^2 = e^{T\rho^2/\sigma^2}.$$

The second identity follows by the Gaussian moment-generating function. With $Z = T^{-1}\sum_t (1 - x_t) \in [0,1]$, $r_A = \mathbb{E}_A \mathrm{Reg}_T(1)$, and $b_B = \mathbb{E}_B \mathrm{WCV}_T$, Cauchy–Schwarz and $Z^2 \le Z$ give

$$1 - \frac{b_B}{\rho T} = \mathbb{E}_B Z = \mathbb{E}_A[JZ] \le e^{T\rho^2/(2\sigma^2)}\sqrt{r_A/T}.$$

For $\rho = 2\mathcal{B}_T/T \in (0,1]$, the uniform budget guarantee implies $b_B \le \mathcal{B}_T = \rho T/2$, so

$$R_T \ge r_A \ge \frac{T}{4} e^{-T\rho^2/\sigma^2} = \frac{T}{4}\exp\left(-\frac{4\mathcal{B}_T^2}{\sigma^2 T}\right).$$

Taking logarithms proves (13) when $0 < R_T < T/4$. All quantities are bounds for the normalized class; the actual slope ρ may be smaller than one without changing the declared constants. □

Proof of Corollary 3.3. Set $a_\sigma = \min\{\sigma, 1\}$, $\rho = (a_\sigma/2)\sqrt{\log T/T} \in (0,1]$, and $h = (a_\sigma/4)\sqrt{T\log T} = \rho T/2$. If $r_A \ge h$, (15) follows. Otherwise, (14) yields

$$1 - \frac{b_B}{\rho T} \le T^{1/8}\sqrt{h/T} = \frac{\sqrt{a_\sigma}}{2}\frac{(\log T)^{1/4}}{T^{1/8}} \le \frac{1}{2},$$

using $a_\sigma \le \sigma$, $a_\sigma \le 1$, and $\log T \le \sqrt{T}$. Hence $b_B \ge h$, proving the joint lower bound for every $T \ge 2$.

If $R_T \le C_R\sqrt{T}$ and $0 < \mathcal{B}_T \le T/2$, then for all sufficiently large T, (13) gives

$$\mathcal{B}_T \ge \frac{\sigma}{2}\sqrt{T\left(\tfrac{1}{2}\log T - \log(4C_R)\right)}.$$

When $\mathcal{B}_T > T/2$, the same asymptotic lower order is automatic for fixed σ. A zero uniform budget bound forces $x_t = 0$ almost surely in world B; absolute continuity of the Gaussian input laws then forces the same actions in A, giving $r_A = T$. Thus that case cannot coexist with square-root regret for large T. Finally, insert $\mathcal{B}_T = C_B\sqrt{T}$ in (12) for $T \ge 4C_B^2$ to prove the endpoint statement. □

Scope of the lower bounds. The hard-violation bound uses fresh random levels and Gaussian smoothing. It does not extend to every bounded-noise model merely by truncating the Gaussian; support effects can change the testing problem. The budget construction instead has constant constraints and varies their slope with the horizon, which is permitted in a uniform worst-case bound. It still applies if the learner knows both candidate instances and the common feasible action 0. Neither lower bound applies when the value channel is exact.

B.7 Further tradeoff and oracle consequences

The logarithm persists beyond square-root regret. If a uniform regret bound satisfies $R_T \le CT^\beta$ for fixed $C > 0$ and $0 \le \beta < 1$, then (13) gives

$$\mathcal{B}_T \ge \frac{\sigma}{2}\sqrt{T\big((1-\beta)\log T - \log(4C)\big)} \tag{39}$$

whenever $0 < \mathcal{B}_T \le T/2$ and the radicand is positive. For fixed σ, the case $\mathcal{B}_T > T/2$ already exceeds this order for large T; a zero budget bound forces linear regret by absolute continuity. Thus every fixed polynomial improvement over linear regret requires $\Omega(\sigma\sqrt{T\log T})$ budget violation. Conversely, a uniform $\mathcal{B}_T = o(\sigma\sqrt{T\log T})$ implies the necessary regret lower bound $R_T \ge T^{1-o(1)}$ by (12). These are direct consequences of the finite-horizon test, not additional assumptions on the adversary.

Exact derivatives do not reveal the missing level. Both lower bounds remain valid if the learner is given the common feasible action 0 and may query exact derivatives at arbitrary points. In the hard-violation construction, $\ell_t'(x) = -1$ and $\gamma_t'(x) = 1$ everywhere in both worlds; in the budget construction, $\gamma_t'(x) = \rho$ in both worlds. All higher derivatives vanish. Such answers, and the disclosed feasible action, are identical across the alternatives and add no distinguishing information to the transformed noisy values. These obstructions therefore hold for smooth affine functions. They do not cover an oracle that reveals exact constraint values or permits additional value samples per round.

C LEDGER: FINITE-VARIANCE ANALYSIS

We prove Theorem 4.1. Throughout, $L = D\widehat{G}$, $B_v = (L^2 + \sigma^2)^{1/2} \geq L$, and $0 < \alpha \leq (64B_v)^{-1}$ is fixed within the epoch. The prescribed choice is $\alpha = \sqrt{\log(eT)}/(64B_v\sqrt{T})$. All quantities refer to that algorithm. Comparators are fixed deterministic $u \in \mathcal{U}$; conditional expectations use the pre-noise field $\mathcal{F}_t$ and martingales use the post-feedback filtration $(\mathcal{G}_t)$.

Proof map and attribution. The projection comparison in Lemma C.2 is a predictable variable-step version of standard OGD (Zinkevich, 2003). The weighted regularizer cancellation in (49) and the logarithmic-accumulator argument in Lemma C.5 adapt Yu et al. (2026, Alg. 1 and Lemmas 10–11). The proof-specific work is to make these components compatible with the reflected potential and dependent finite-variance feedback (Lemma C.3), and to control the true expected maximum-window violation, paying for both clipping and regularization ((52)). Exponential moments, Jensen's inequality, and Doob's inequality provide the remaining classical steps. Section 4.3 explains why each modification is needed.

C.1 CLIPPING, DOMINATION, AND INTEGRABILITY

For analysis only, define

$$v_t^\circ = \max\{v_t, -L\}, \qquad \tilde{v}_t^\circ = v_t^\circ + \varepsilon_t, \qquad \bar{v}_t^\circ = \mathrm{clip}(\tilde{v}_t^\circ, -1/\alpha, 1/\alpha).$$

These quantities are not queried or computed by the algorithm.

Lemma C.1 (Dominating observations and predictable regularizers)**.** *For every round,*

$$\bar{v}_t \leq \bar{v}_t^\circ, \qquad \mathbb{E}[(\bar{v}_t^\circ)^2 \mid \mathcal{F}_t] \leq B_v^2, \tag{40}$$

$$|\mathbb{E}[\bar{v}_t^\circ \mid \mathcal{F}_t] - v_t^\circ| \leq \mathbb{E}[|\tilde{v}_t^\circ - \bar{v}_t^\circ| \mid \mathcal{F}_t] \leq \alpha B_v^2, \tag{41}$$

$$\mathbb{E}[[\tilde{v}_t - \bar{v}_t]_+ \mid \mathcal{F}_t] \leq \alpha B_v^2, \tag{42}$$

$$16\alpha B_v^2 \leq r_t \leq 16\alpha B_v^2 + 4L \leq \tfrac{17}{4} B_v. \tag{43}$$

All terms whose expectations occur in the bounds below are integrable; no unconditional moment of a negative raw level or raw value observation is needed.

Proof. Lemma 2.1 gives $|v_t^\circ| \leq L$ and $v_t \leq v_t^\circ$. Monotonicity of clipping proves $\bar{v}_t \leq \bar{v}_t^\circ$. The oracle error conditions give $\mathbb{E}[(\tilde{v}_t^\circ)^2 \mid \mathcal{F}_t] = (v_t^\circ)^2 + \mathbb{E}[\varepsilon_t^2 \mid \mathcal{F}_t] \leq B_v^2$. Clipping does not increase absolute value. For every real z,

$$|z - \mathrm{clip}(z, -1/\alpha, 1/\alpha)| = (|z| - 1/\alpha)_+ \leq \alpha z^2,$$

which proves (41). Moreover,

$$[\tilde{v}_t - \bar{v}_t]_+ = (\tilde{v}_t - 1/\alpha)_+ \leq (\tilde{v}_t^\circ - 1/\alpha)_+ \leq \alpha(\tilde{v}_t^\circ)^2,$$

so (42) holds even for very negative v_t. Since $\sqrt{H_t} + \sqrt{H_{t-1}} \geq w_t \geq 0$, the adaptive part of r_t lies in $[0, 4L]$. Use $\alpha B_v \leq 1/64$ and $L \leq B_v$ for the final upper bound. Finally, $r_t \geq 0$ and $\bar{v}_t \leq 1/\alpha$ imply $0 \leq Q_t \leq t/\alpha$. Thus the queue, weights, accumulated weights, and potential have deterministic bounds at each finite horizon. Conditional direction second moments imply integrability of all primal terms. Loss differences and violation metrics are bounded by Ln; the positive clipping residual has the preceding integrable bound. No exponential moment of the raw observations, or absolute moment of negative levels, is used. □

C.2 PREDICTABLE PRIMAL COMPARISON

Lemma C.2 (Variable-step primal bound)**.** *For every fixed $u \in \mathcal{U}$ and $n \le T$,*

$$\mathbb{E}\mathrm{Reg}_n(u) + \mathbb{E}\sum_{t=1}^{n} w_t v_t^\circ \le 2L\sqrt{n} + 2L\mathbb{E}\sqrt{H_n}. \tag{44}$$

Proof. Write $d_t = \tilde{p}_t + w_t\tilde{q}_t$, $S_t = t + H_t$, and $S_0 = \alpha^2$. The step sizes $\eta_t = D/(2\widehat{G}\sqrt{S_t})$ are nonincreasing. Expanding the one-step projection inequality and telescoping the squared distances to u yields

$$\sum_{t=1}^{n}\langle d_t, x_t - u\rangle \le \frac{D^2}{2\eta_n} + \frac{1}{2}\sum_{t=1}^{n}\eta_t\|d_t\|^2. \tag{45}$$

Both w_t and η_t are $\mathcal{F}_t$-measurable, and arbitrary dependence between the two direction channels is allowed by

$$\mathbb{E}[\|d_t\|^2 \mid \mathcal{F}_t] \le 2\widehat{G}^2(1 + w_t^2) = 2\widehat{G}^2(S_t - S_{t-1}).$$

Since

$$\sum_{t=1}^{n}\frac{S_t - S_{t-1}}{\sqrt{S_t}} \le 2(\sqrt{S_n} - \sqrt{S_0}),$$

the expectation of the right side of (45) is at most $L\mathbb{E}\sqrt{S_n} + L\mathbb{E}(\sqrt{S_n} - \alpha) \le 2L\mathbb{E}\sqrt{S_n}$. Convexity and feasibility give

$$\langle p_t, x_t - u\rangle \ge \ell_t(x_t) - \ell_t(u), \qquad \langle q_t, x_t - u\rangle \ge v_t^\circ.$$

Conditional unbiasedness and $w_t \ge 0$ therefore lower-bound the expected left side of (45) by the left side of (44). Finally, $\sqrt{n + H_n} \le \sqrt{n} + \sqrt{H_n}$. □

C.3 REFLECTED-POTENTIAL DRIFT AND CANCELLATION

Lemma C.3 (Drift with a zero boundary derivative)**.** *For every round,*

$$\mathbb{E}[\Phi_\alpha(Q_t) - \Phi_\alpha(Q_{t-1}) \mid \mathcal{F}_t] \le w_t v_t^\circ - \tfrac{1}{2}w_t r_t + 60\alpha^2 B_v^2. \tag{46}$$

Proof. Extend the potential to $\mathbb{R}$ by $F(q) = e^{\alpha[q]_+} - \alpha\,[q]_+ - 1$. The extension is continuously differentiable at zero. For $q \ge 0$ and any z with $\alpha z \le 1$, Taylor's integral remainder gives

$$F(q + z) - F(q) \le \alpha(e^{\alpha q} - 1)z + \tfrac{3}{2}\alpha^2 e^{\alpha q} z^2. \tag{47}$$

Indeed, the almost-everywhere second derivative is zero on the negative half-line and $\alpha^2 e^{\alpha y}$ on the positive half-line. Along the segment from q to $q + z$ it is at most $e\alpha^2 e^{\alpha q}$, and $e/2 < 3/2$. The function F is nondecreasing. Since $\bar{v}_t \le \bar{v}_t^\circ$,

$$\Phi_\alpha(Q_t) - \Phi_\alpha(Q_{t-1}) \le F(Q_{t-1} + \bar{v}_t^\circ - r_t) - F(Q_{t-1}).$$

Apply (47) to this upper bound with $q = Q_{t-1}$ and $z = \bar{v}_t^\circ - r_t$. The increment condition holds since $r_t \ge 0$ and $\alpha\bar{v}_t^\circ \le 1$. Using $\alpha^2 e^{\alpha q} = \alpha(w_t + \alpha)$ and $(\bar{v}_t^\circ - r_t)^2 \le 2(\bar{v}_t^\circ)^2 + 2r_t^2$ gives

$$\Phi_\alpha(Q_t) - \Phi_\alpha(Q_{t-1}) \le w_t(\bar{v}_t^\circ - r_t) + 3\alpha(w_t + \alpha)((\bar{v}_t^\circ)^2 + r_t^2).$$

Lemma C.1 bounds its conditional expectation by

$$w_t v_t^\circ - w_t r_t + 4\alpha B_v^2 w_t + 3\alpha w_t r_t^2 + 3\alpha^2(B_v^2 + r_t^2). \tag{48}$$

The regularizer absorbs the two positive terms proportional to w_t, because

$$4\alpha B_v^2 \le r_t/4, \qquad 3\alpha r_t \le 3\alpha(16\alpha B_v^2 + 4L) \le \frac{51}{256} < \frac{1}{4}.$$

Also $3\alpha^2(B_v^2 + r_t^2) \le (915/16)\alpha^2 B_v^2 \le 60\alpha^2 B_v^2$. Substitution in (48) proves the claim. The zero derivative at the reflection boundary is what avoids a first-order reflection cost. □

Proof of (17). The adaptive regularizer obeys the exact identity

$$\tfrac{1}{2} w_t r_t = 8\alpha B_v^2 w_t + 2L \frac{w_t^2}{\sqrt{H_t} + \sqrt{H_{t-1}}} = 8\alpha B_v^2 w_t + 2L(\sqrt{H_t} - \sqrt{H_{t-1}}). \tag{49}$$

There is no division by zero since $H_{t-1} \geq \alpha^2 > 0$. Sum (46), take expectations, and use (44) to obtain

$$\begin{aligned} \mathbb{E}\mathrm{Reg}_n(u) + \mathbb{E}\Phi_\alpha(Q_n) &\leq 2L\sqrt{n} + 2L\mathbb{E}\sqrt{H_n} - \tfrac{1}{2}\mathbb{E}\sum_{t=1}^{n} w_t r_t + 60\alpha^2 B_v^2 n \\ &= 2L\sqrt{n} + 2L\alpha - 8\alpha B_v^2 \mathbb{E}\sum_{t=1}^{n} w_t + 60\alpha^2 B_v^2 n. \end{aligned}$$

Here $\Phi_\alpha(Q_0) = 0$ and $\sqrt{H_0} = \alpha$. □

C.4 Running maximum and total regularization

Lemma C.4 (Exponential moments and the running maximum)**.** *With* $K_0 = 4 + 6L$ *and* $\Lambda_T = \log((T+1)K_0 T)$,

$$\mathbb{E}e^{\alpha Q_n} \leq K_0 T \quad (0 \leq n \leq T), \qquad \mathbb{E}\max_{0 \leq n \leq T} Q_n \leq \Lambda_T/\alpha. \tag{50}$$

Proof. Since $\mathrm{Reg}_n(u) \geq -Ln$, (17) gives $\mathbb{E}\Phi_\alpha(Q_n) \leq Ln + 2L\sqrt{n} + 2L\alpha + 60\alpha^2 B_v^2 n$. For $z \geq 0$, $2z \leq e^z$, hence $e^{\alpha q} \leq 2(1 + \Phi_\alpha(q))$ for $q \geq 0$. Consequently,

$$\mathbb{E}e^{\alpha Q_n} \leq 2 + 2Ln + 4L\sqrt{n} + 4L\alpha + 120\alpha^2 B_v^2 n.$$

For $1 \leq n \leq T$, use $\sqrt{n} \leq n$, $4L\alpha \leq 1/16$, and $120\alpha^2 B_v^2 \leq 120/4096$ to bound this by $(4+6L)T$. The case $n = 0$ is immediate. Write $M_T = \max_{0 \leq n \leq T} Q_n$. Pathwise $e^{\alpha M_T} \leq \sum_{n=0}^{T} e^{\alpha Q_n}$, so Jensen's inequality yields $\alpha \mathbb{E}M_T \leq \log \mathbb{E}e^{\alpha M_T} \leq \Lambda_T$. □

Lemma C.5 (Total regularization cost)**.**

$$\mathbb{E}\sum_{t=1}^{T} r_t \leq 16\alpha B_v^2 T + 4L\sqrt{2T\Lambda_T}. \tag{51}$$

Proof. Following the regularization-cost argument of Yu et al. (2026, Lem. 11), Cauchy–Schwarz and $H_t - H_{t-1} = w_t^2$ imply

$$\sum_{t=1}^{T} \frac{w_t}{\sqrt{H_t} + \sqrt{H_{t-1}}} \leq \sum_{t=1}^{T} \frac{w_t}{\sqrt{H_t}} \leq \sqrt{T \sum_{t=1}^{T} \frac{H_t - H_{t-1}}{H_t}} \leq \sqrt{T \log \frac{H_T}{\alpha^2}},$$

where the last step uses $(b-a)/b \leq \log(b/a)$ for $0 < a \leq b$. Set $a_t = w_t/\alpha = e^{\alpha Q_{t-1}} - 1 \geq 0$. Then

$$\frac{H_T}{\alpha^2} = 1 + \sum_{t=1}^{T} a_t^2 \leq \left(1 + \sum_{t=1}^{T} a_t\right)^2.$$

By Jensen's inequality and Lemma C.4,

$$\mathbb{E}\log \frac{H_T}{\alpha^2} \leq 2\log\left(1 + \sum_{t=1}^{T} \mathbb{E}a_t\right) \leq 2\log(1 + K_0 T^2) \leq 2\Lambda_T.$$

Apply Jensen's inequality once more to the square root and add $16\alpha B_v^2 T$. □

C.5 Transfer to true windows and parameter choice

Proof of (18) *and* (19). The reflected update gives $Q_t \geq Q_{t-1} + \bar{v}_t - r_t$. Thus, for every window $[s, e] \subseteq [1, T]$,

$$\sum_{t=s}^{e} \bar{v}_t \leq Q_e - Q_{s-1} + \sum_{t=s}^{e} r_t \leq M_T + \sum_{t=1}^{T} r_t.$$

Write $S_n^\varepsilon = \sum_{t=1}^{n} \varepsilon_t$, with $S_0^\varepsilon = 0$. Since $v_t = \bar{v}_t + (\tilde{v}_t - \bar{v}_t) - \varepsilon_t$, pathwise

$$\mathrm{WCV}_T \leq M_T + \sum_{t=1}^{T} r_t + \sum_{t=1}^{T} [\tilde{v}_t - \bar{v}_t]_+ + 2 \max_{0 \leq n \leq T} |S_n^\varepsilon|. \tag{52}$$

The expected positive clipping residual is at most $\alpha B_v^2 T$ by (42). Negative clipping residuals only reduce the window sum. The tower property transfers $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_t] = 0$ to $\mathcal{G}_{t-1}$, so S_n^ε is a square-integrable martingale. Orthogonality of its increments and Doob's L^2 maximal inequality (Doob, 1953) give

$$\mathbb{E} \max_{0 \leq n \leq T} |S_n^\varepsilon| \leq 2\sqrt{\mathbb{E}(S_T^\varepsilon)^2} \leq 2\sigma\sqrt{T}.$$

Take expectations in (52) and apply Lemmas C.4 and C.5 to prove (18).

For $\alpha = \sqrt{\log(eT)}/(64 B_v \sqrt{T})$, the restriction $\alpha B_v \leq 1/64$ follows from $\log(eT) \leq T$. Dropping the nonnegative potential and weighted-sum terms in (17) yields

$$\mathbb{E}\mathrm{Reg}_T(u) \leq 2L\sqrt{T} + 2L\alpha + \frac{15}{1024}\log(eT) \leq 2L\sqrt{T} + \frac{1}{32} + \frac{15}{1024}\log(eT).$$

Substitution in (18) gives the explicit budget bound

$$\mathbb{E}\mathrm{WCV}_T \leq 64 B_v \Lambda_T \sqrt{\frac{T}{\log(eT)}} + \frac{17}{64} B_v \sqrt{T \log(eT)} + 4L\sqrt{2T\Lambda_T} + 4\sigma\sqrt{T}. \tag{53}$$

For fixed problem bounds, $\Lambda_T = O(\log(eT))$ and $\log(eT) = O(\sqrt{T})$, proving both rates in (19). □

C.6 Finite-horizon interpretation of the constants

The displayed constants prioritize a transparent finite-variance proof rather than optimized finite-horizon performance. Lemma 2.1 gives $v_t \leq L$ pathwise, so $\mathrm{WCV}_T \leq LT$ independently of the algorithm. Consequently, the right side of (53) can always be capped at LT. For example, with $D = \widehat{G} = \sigma = 1$ and $T = 10^4$, we have $B_v = \sqrt{2}$ and $\Lambda_T = \log(10(T+1)T)$. Substitution gives approximately 61,795 in (53), exceeding the trivial bound $LT = 10{,}000$. At this horizon the explicit certificate therefore does not establish an improvement over the trivial bound. This does not contradict the asymptotic rate or demonstrate poor realized performance. The numerical checks validate implementation identities, not empirical regret or budget scaling; tighter constants and practical parameter selection are not established by the present analysis.

D Unknown horizon and interpretation of the guarantee

Proof of Corollary 4.2. Use epochs $I_j = \{2^j, \ldots, 2^{j+1} - 1\}$ of planned length $h_j = 2^j$, $j = 0, 1, \ldots$. At the first round of each epoch reset $Q_0 = 0$ and $H_0 = \alpha_j^2$, choose an action in $\mathcal{K}$, and set $\alpha_j = \sqrt{\log(eh_j)}/(64 B_v \sqrt{h_j})$. Use local round numbers in all other updates.

For any prefix of $n \leq h$ rounds of an epoch with planned length h, Theorem 4.1 gives

$$\mathbb{E}\mathrm{Reg}_n(u) \leq 2L\sqrt{n} + 2L\alpha_h + 60\alpha_h^2 B_v^2 n \leq C_R\sqrt{h},$$

where C_R depends only on L, B_v. Indeed, $\log(eh)/\sqrt{h}$ is uniformly bounded for $h \geq 1$. Repeating the maximum-queue and regularization bounds through n, with the same α_h, gives

$$\mathbb{E}\mathrm{WCV}_n \leq \frac{\Lambda_n}{\alpha_h} + 17\alpha_h B_v^2 n + 4L\sqrt{2n\Lambda_n} + 4\sigma\sqrt{n} \leq C_W\sqrt{h\log(eh)},$$

where $\Lambda_n = \log((n+1)K_0 n)$. The bounds apply after any past history because conditional unbiasedness and second-moment envelopes are unchanged by deterministic restarts. The same fixed comparator is feasible in every epoch being considered.

For $J = \lfloor \log_2 T \rfloor$, epoch J is the last nonempty epoch; it may be unfinished. Regret adds across the epochs. Every global window meets each epoch in an empty set or a contiguous window, hence, pathwise,

$$\mathrm{WCV}_T \le \sum_{j=0}^{J} \mathrm{WCV}^{(j)}_{|I_j \cap [1,T]|}.$$

Since $2^J \le T$,

$$\sum_{j=0}^{J} \sqrt{2^j} \le \frac{\sqrt{T}}{1-2^{-1/2}}, \qquad \sum_{j=0}^{J} \sqrt{2^j \log(e2^j)} \le \frac{\sqrt{T\log(eT)}}{1-2^{-1/2}}.$$

Summing the expectation bounds proves both claims. □

Feedback and implementation. The reference implementation fixes all update coefficients before receiving the feedback triple. Exponential weights can be stored in logarithmic form: maintain $\log H_t$, compute $\log w_t$ using a stable log-exponential-minus-one evaluation, and form the bounded ratios $w_t/(\sqrt{H_t}+\sqrt{H_{t-1}})$ and $w_t/\sqrt{t+H_t}$ directly. This avoids numerical overflow in the weight while leaving the mathematical update unchanged. Numerical tests of identities and certificates are provided as implementation checks, not experimental evidence for a worst-case theorem.