

# Operationalizing Stein’s Method for Online Linear Optimization: CLT-Based Optimal Tradeoffs

Zhiyu Zhang  
Carnegie Mellon University  
zhiyuzresearch@gmail.com

Aaditya Ramdas  
Carnegie Mellon University  
aramdas@cmu.edu

## Abstract

Adversarial online linear optimization (OLO) is essentially about making performance tradeoffs with respect to the unknown difficulty of the adversary. In the setting of one-dimensional fixed-time OLO on a bounded domain, it has been observed since Cover [Cov66] that achievable tradeoffs are governed by probabilistic inequalities, and these descriptive results can be converted into algorithms via dynamic programming, which, however, is not computationally efficient. We address this limitation by showing that Stein’s method, a classical framework underlying the proofs of probabilistic limit theorems, can be operationalized as computationally efficient OLO algorithms. The associated regret and total loss upper bounds are “additively sharp”, meaning that they surpass the conventional big-O optimality and match normal-approximation-based lower bounds by additive lower order terms. Our construction is inspired by the remarkably clean proof of a Wasserstein martingale central limit theorem (CLT) due to Röllin [Röl18].

Several concrete benefits can be obtained from this general technique. First, with the same computational complexity, the proposed algorithm improves upon the total loss upper bounds of online gradient descent (OGD) and multiplicative weight update (MWU). Second, our algorithm can realize a continuum of optimal two-point tradeoffs between the total loss and the maximum regret over comparators, improving upon prior works in parameter-free online learning. Third, by allowing the adversary to randomize on an unbounded support, we achieve sharp in-expectation performance guarantees for OLO with noisy feedback.

## 1 Introduction

We study a fundamental protocol in adversarial online learning: one-dimensional fixed-time *online linear optimization* (OLO) on a bounded domain. With a time horizon  $T \in \mathbb{N}_+$ , it is a repeated game between an algorithm we design and an adversary. The algorithm is defined by a sequence of functions  $(a_t)_{t \in [1:T]}$ , where each  $a_t : \mathbb{R}^{t-1} \rightarrow [-1, 1]$  can depend on  $T$ . The adversary is defined by another sequence of functions  $(l_t)_{t \in [1:T]}$ , where each  $l_t : \mathbb{R}^t \rightarrow \mathbb{R}$  can depend on  $T$  and the algorithm  $(a_t)_{t \in [1:T]}$ . In each (the  $t$ -th;  $t \in [1 : T]$ ) round,

- The algorithm picks the decision  $x_t = a_t(g_{1:t-1}) \in [-1, 1]$ .
- The adversary picks the loss gradient  $g_t = l_t(x_{1:t}) \in \mathbb{R}$ .
- The algorithm observes  $g_t$  and then suffers the linear loss  $g_t x_t$ .

At the end of the game, the performance of the algorithm is measured by its *total loss*,  $\text{Loss}_T := \sum_{t=1}^T g_t x_t$ , lower is better. A real world example is sequential betting:  $x_t$  is the amount of money a gambler bets on the negativity of  $g_t$ , and with the instantaneous payoff being  $-g_t x_t$ , the total amount of money they make is  $-\text{Loss}_T$ .

For intuition, let us briefly assume  $g_t \in [-1, 1]$  such that the worst case adversary would pick  $g_t = \text{sign}(x_t)$ . It is clear that optimizing only for this worst case would trivialize the problem, since the minimax-optimal algorithm is forced to always pick  $x_t = 0$ . The convention instead is to simultaneously consider adversaries with varied difficulty: although any nontrivial algorithm should suffer positive  $\text{Loss}_T$  in the worst case, one could indeed expect negative  $\text{Loss}_T$  if the adversary ends up being easy. Therefore every algorithm should realize a tradeoff, just like in betting where higher profits necessarily comes with higher risks. The objective is thus achieving suitable notions of “admissibility”, in the sense that no other algorithm can always guarantee lower  $\text{Loss}_T$  regardless of the adversary.

Rigorously studying this objective motivates a common surrogate performance metric called the *regret*: for all  $u \in [-1, 1]$  called a *comparator*, the regret of the algorithm with respect to  $u$  is defined as  $\text{Regret}_T(u) := \sum_{t=1}^T g_t(x_t - u)$ . If an algorithm guarantees  $\text{Regret}_T(u) \leq \psi_T(u)$  for some proper, closed and convex function  $\psi_T : [-1, 1] \rightarrow (-\infty, \infty]$ ,<sup>1</sup> then with the convex and 1-Lipschitz function  $\psi_T^* : \mathbb{R} \rightarrow \mathbb{R}$  being the convex conjugate of  $\psi_T$ ,  $\text{Loss}_T$  can be upper-bounded by the oracle inequality

$$\text{Loss}_T \leq \inf_{u \in [-1, 1]} \left[ \left( \sum_{t=1}^T g_t \right) u + \psi_T(u) \right] = -\sup_{u \in \mathbb{R}} \left[ \left( -\sum_{t=1}^T g_t \right) u - \psi_T(u) \right] = -\psi_T^* \left( -\sum_{t=1}^T g_t \right). \quad (1)$$

Since  $(\psi_T^*)^* = \psi_T$ , the total loss bound conversely implies the regret bound, and this two-way relation is called the *loss-regret duality* [MO14] (Lemma G.5). The idea is that upper-bounding the surrogate  $\text{Regret}_T(u)$  is equivalent to upper-bounding  $\text{Loss}_T$  itself in an adversary-dependent manner, where the cumulative bias  $\sum_{t=1}^T g_t$  can be used to quantify the difficulty of the adversary.<sup>2</sup> We will adopt this “loss-centric” viewpoint to characterize performance tradeoffs in OLO, whose relation to the conventional view of regret minimization is demonstrated through the following example.

**Example 1** (Uniform regret). *Consider the following definition which we call the uniform regret,*

$$\text{Regret}_T^{\text{unif}} := \sup_{u \in [-1, 1]} \text{Regret}_T(u) = \max \{ \text{Regret}_T(-1), \text{Regret}_T(1) \}.$$

*Any regret bound  $\text{Regret}_T(u) \leq \psi_T(u)$  can be converted into a uniform regret bound  $\text{Regret}_T^{\text{unif}} \leq \sup_{u \in [-1, 1]} \psi_T(u)$ , whose right hand side, if finite, can be regarded as a function  $\psi_T^{\text{unif}}$  invariant to its argument. Then, an upper bound on  $\text{Loss}_T$  is induced by the convex conjugate of  $\psi_T^{\text{unif}}$ ,  $(\psi_T^{\text{unif}})^*(s) = |s| - \sup_{u \in [-1, 1]} \psi_T(u)$ . Despite being weaker than  $\psi_T^*$ , i.e.,  $(\psi_T^{\text{unif}})^*(s) \leq \psi_T^*(s)$  for all  $s \in \mathbb{R}$ , this is a convenient relaxation adopted by most prior works, especially due to its connection to uniform convergence in statistical learning. See, e.g., [RS14].*

This work presents computationally efficient algorithms achieving general and sharp *upper bound functions* on  $\text{Regret}_T(u)$  and  $\text{Loss}_T$ . En route, we introduce the techniques of *Stein’s method* [Ste72, Ste86] to adversarial online learning and demonstrate intriguing algorithmic connections between sharp rates and *central limit theorems* (CLTs).

## 1.1 Motivation

To concretely motivate our results, we start from a central problem of the field: characterizing the conditions on  $\psi_T^*$  for the achievability of Eq.(1) against a large class of adversaries. In the special case of Boolean adversaries, the following landmark result is due to Cover [Cov66, Statement V].

**Theorem 1** (Cover’s characterization, adapted). *Assume that  $g_t \in \{-1, 1\}$  for all  $t \in [1 : T]$ , and define  $\text{RS}(n)$  as the distribution of the sum of  $n$  independent Rademacher random variables. Then, for all convex and 1-Lipschitz function  $\psi_T^* : \mathbb{R} \rightarrow (-\infty, \infty]$ , there exists an algorithm achieving the total loss bound Eq.(1) if and only if*

$$\mathbb{E}_{X \sim \text{RS}(T)} [\psi_T^*(X)] \leq 0. \quad (2)$$

*In particular, the corresponding algorithm outputs the expectation of a discrete derivative,*

$$x_t = -\frac{1}{2} \mathbb{E}_{X \sim \text{RS}(T-t)} \left[ \psi_T^* \left( \sum_{i=1}^{t-1} g_i + X + 1 \right) - \psi_T^* \left( \sum_{i=1}^{t-1} g_i + X - 1 \right) \right], \quad \forall t \in [1 : T]. \quad (3)$$

The necessity of Eq.(2) is straightforward, since any gambler cannot expect to make profit when the nature is uniformly random. Truly remarkable is the sufficiency of Eq.(2), proved constructively by Eq.(3). The construction reveals a fundamental dynamic programming principle which has inspired numerous subsequent

<sup>1</sup>Defined as  $\infty$  outside  $[-1, 1]$  following the convention.

<sup>2</sup>Intuitively, a larger  $|\sum_{t=1}^T g_t|$  means the adversary is more biased thus easier for the algorithm to exploit. Therefore in Eq.(1), the typical  $\psi_T^*$  function of interest would be an even function minimized at 0, meaning the loss bound is maximized at  $\sum_{t=1}^T g_t = 0$ .

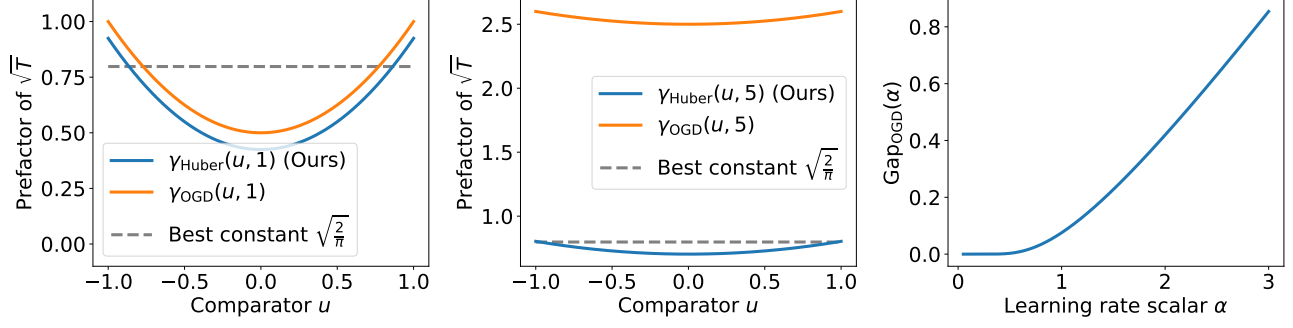


Figure 1: Comparison of the prefactors of  $\sqrt{T}$  in the regret bounds of our algorithm (denoted by  $\gamma_{\text{Huber}}(u, \alpha)$ ), OGD (denoted by  $\gamma_{\text{OGD}}(u, \alpha)$ ), and Cover’s algorithm (the optimal  $u$ -independent prefactor  $\sqrt{\frac{2}{\pi}}$ ; computationally efficient versions are provided by [KKW20, GHP22]).  $u$  is the comparator and  $\alpha$  is the scaling factor of the learning rate; see Section 4.2 for definitions. Left: with  $\alpha = 1$ ,  $\gamma_{\text{Huber}}(u, \alpha)$  and  $\gamma_{\text{OGD}}(u, \alpha)$  are compared as functions of  $u$ , lower is better. Our regret bound dominates that of OGD for all  $u \in [-1, 1]$ , while the optimal  $u$ -independent bound does not. Middle: with  $\alpha = 5$ , the improvement over OGD becomes more significant. Right: the margin of improvement  $\text{Gap}_{\text{OGD}}(\alpha) := \gamma_{\text{OGD}}(u, \alpha) - \gamma_{\text{Huber}}(u, \alpha)$  as a function of  $\alpha$ .

works connecting repeated games, stochastic analysis and partial differential equations (PDEs) [CBL06, RS14, KP17, DK20, HLPR23]. Based on that Cover proved the following in the same paper: still assuming  $g_t \in \{-1, 1\}$  for all  $t$ , an instantiation of Eq.(3) achieves the uniform regret bound  $\text{Regret}_T^{\text{unif}} \leq \sqrt{\frac{2}{\pi}}T + 1$ , where the leading constant  $\sqrt{\frac{2}{\pi}}$  cannot be improved further. This is better than the best known  $\text{Regret}_T^{\text{unif}} \leq \sqrt{T}$  bound of *online gradient descent* (OGD) [Zin03], and when adapted to the problem of *learning from expert advice* (LEA) with two experts, it is also better than the uniform regret bound of *multiplicative weight update* (MWU) [LW94, AHK12].

However, a crucial limitation of Cover’s algorithm is the computation. While OGD and MWU both take  $O(1)$  time per round, exactly evaluating Eq.(3) requires sweeping over the size- $(T - t + 1)$  support of  $\text{RS}(T - t)$  therefore is  $O(T)$  time. Such a gap becomes more evident if  $g_t \in \{-\varepsilon_t, \varepsilon_t\}$  for some distinct  $\varepsilon_1, \dots, \varepsilon_T \in \mathbb{R}_{>0}$ : the distribution underlying the accordingly generalized Eq.(3) has support size  $2^{T-t}$ , making the computation of  $x_t$  exponential-time. Furthermore, Cover’s algorithm requires the prior knowledge of  $|g_t|$  for all  $t$ , whereas OGD and MWU require only upper bounds on these  $|g_t|$ , at most.<sup>3</sup> The intuition is that some kind of continuous approximation is necessary, but a general, sharp and analytically scalable solution remains challenging.

- A natural attempt motivated by CLTs would be replacing the distribution  $\text{RS}(T - t)$  in Eq.(3) by the normal  $\mathcal{N}(0, T - t)$ , such that computing the Gaussian integral  $\mathbb{E}_{X \sim \mathcal{N}(0, T-t)}[\psi_T^*(s + X)]$  can be assumed to take  $O(1)$  time. The Lipschitzness of  $\psi_T^*$  gives the normal approximation error  $|\mathbb{E}_{X \sim \mathcal{N}(0, T-t)}[\psi_T^*(s + X)] - \mathbb{E}_{X \sim \text{RS}(T-t)}[\psi_T^*(s + X)]| = O(1), \forall s \in \mathbb{R}$  [Zhu14, Theorem 2], but simply combining it with Eq.(3) introduces  $O(T)$  cumulative error in the associated regret and total loss bounds.
- Existing works [KKW20, GHP22] refined this idea for the special case of Example 1. Building on the relation between  $(\psi_T^{\text{unif}})^*$  and the absolute value function, the proposed algorithms are based on querying the *error function*  $\text{erf} : \mathbb{R} \rightarrow \mathbb{R}$  which is computationally efficient. Via specialized Taylor-approximation-based proofs, they also match the leading  $\sqrt{\frac{2}{\pi}}T$  term in Cover’s uniform regret bound against the larger class of  $g_t \in [-1, 1]$  adversaries.

An immediate follow-up question is whether other *additively sharp* guarantees beyond Example 1 can be efficiently realized. Here “additively sharp” means that the obtained upper bound on  $\text{Loss}_T$  should be away from a probabilistic achievability boundary (e.g., Eq.(2)) by additive, rather than multiplicative, lower order factors. As motivated previously, this corresponds to the efficient and near-optimal realization of general performance tradeoffs, which is of central importance in OLO. Moreover, both OGD and MWU are by default equipped with

<sup>3</sup>Adaptive versions of OGD and MWU such as ADAGRAD [DHS11] do not require a priori upper bounds on  $|g_t|$ . The price to pay is a larger leading constant in the regret bound. These are orthogonal to the focus of our work.

certain comparator-dependent regret bound  $\text{Regret}_T(u) \leq \psi_T(u)$ , therefore instead of beating their respective relaxation  $\text{Regret}_T^{\text{unif}} \leq \sup_{u \in [-1, 1]} \psi_T(u)$  which is certainly a loose baseline, it is more sensible to compare the regret bound to such  $\psi_T$  functions and aim for pointwise dominance (see Figure 1, Left). By the loss-regret duality, this will ultimately ensure a dominance on the corresponding  $\text{Loss}_T$  upper bounds.

In this work we present a solution to this problem, among others. The connection to normal approximation motivates the use of Stein’s method, a classical analytical framework underlying various nonasymptotic, quantitative versions of CLTs.

## 1.2 Summary of Results

**Quantitative** This subsection assumes  $g_t \in [-1, 1]$  for all  $t$  unless stated otherwise. Our main result is a normal-approximation-based characterization of performance tradeoffs in OLO.

**Theorem 2** (Theorem 3 and 4, informal). *Let  $\psi_T^* : \mathbb{R} \rightarrow \mathbb{R}$  be an arbitrary convex and 1-Lipschitz function satisfying  $\mathbb{E}_{X \sim \mathcal{N}(0, T)}[\psi_T^*(X)] = 0$ . There exists an  $O(1)$ -time-per-round algorithm (Algorithm 1) guaranteeing*

$$\text{Loss}_T \leq -\psi_T^* \left( -\sum_{t=1}^T g_t \right) + O(\log T),$$

while it is impossible for any algorithm to guarantee  $\text{Loss}_T \leq -\psi_T^* \left( -\sum_{t=1}^T g_t \right) - \Omega(1)$ .

Compared to Cover’s characterization (Theorem 1), the above result overcomes the computational limitation of dynamic programming at the price of an  $O(\log T)$  gap between the upper and lower bounds. Since the  $\psi_T^*$  function typically grows with rate  $\Theta(\sqrt{T})$ , such a gap is on a lower order, meaning that the upper bound is “additively sharp”. Based on this master theorem, the following corollaries and extensions are obtained.

- *Dominating OGD and MWU.* For any hyperparameter  $\alpha \in \mathbb{R}_{>0}$ , a special case of our algorithm guarantees  $\text{Regret}_T(u) \leq \gamma_{\text{Huber}}(u, \alpha)\sqrt{T} + O(\log T)$ , where the prefactor of  $\sqrt{T}$  is given by

$$\gamma_{\text{Huber}}(u, \alpha) = \frac{u^2}{2\alpha} + \left( \alpha + \frac{1}{\alpha} \right) \Phi \left( \frac{1}{\alpha} \right) + \phi \left( \frac{1}{\alpha} \right) - \frac{1}{2}\alpha - \frac{1}{\alpha}.$$

This is compared to the OGD baseline with constant learning rate  $\alpha/\sqrt{T}$ , whose best known regret bound has the classical  $\sqrt{T}$ -prefactor  $\gamma_{\text{OGD}}(u, \alpha) = \frac{1}{2}(\alpha^{-1}u^2 + \alpha)$ . We show that  $\gamma_{\text{Huber}}(u, \alpha) < \gamma_{\text{OGD}}(u, \alpha)$  for all  $u \in [-1, 1]$  and  $\alpha \in \mathbb{R}_{>0}$ , and furthermore,  $\lim_{\alpha \rightarrow \infty} \gamma_{\text{Huber}}(u, \alpha) = \sqrt{\frac{2}{\pi}}$  while  $\lim_{\alpha \rightarrow \infty} \gamma_{\text{OGD}}(u, \alpha) = \infty$ . Combined with Eq.(1), it means the total loss upper bound of our algorithm dominates that of OGD, and their gap can be arbitrarily large. In contrast, improving the uniform regret bound of OGD as in [KKW20, GHP22] does not warrant such a dominance. Details are provided in Section 4, and see Figure 1 for visualizations.

The same argument can be applied to improve the MWU baseline. Notably, a quantitative separation is obtained by comparing the regret upper bound of our algorithm to an existing regret lower bound of MWU [GPS17]. See Appendix C.

- *Loss versus uniform regret.* Within performance tradeoffs in OLO, a notable special case first studied by Even-Dar et al. [EDKMW08] is the optimal “two-point” tradeoff between  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$ , i.e., the regret with respect to the average comparator  $u = 0$  versus the best comparator  $u \in \{-1, 1\}$ . With  $\Phi : \mathbb{R} \rightarrow (0, 1)$  being the standard normal CDF, we show that for any constant  $\varepsilon \in (0, \sqrt{\frac{\pi}{2}}]$ , an instantiation of our algorithm guarantees both  $\text{Loss}_T \leq \varepsilon\sqrt{T} + O(\log T)$  and  $\text{Regret}_T^{\text{unif}} \leq \gamma(\varepsilon)\sqrt{T} + O(\log T)$ , where  $\gamma(\varepsilon) \in [\sqrt{\frac{\pi}{2}}, \infty)$  is the unique solution of the equation  $\int_{-\infty}^{\varepsilon - \gamma(\varepsilon)} \Phi(x) dx = \frac{\varepsilon}{2}$  and cannot be improved further. This improves upon the state-of-the-art approach combining an unconstrained “parameter-free” OLO algorithm [ZCP22] with a constrained-to-unconstrained reduction [CO18]. See Appendix D.
- *Noisy feedback.* Existing additively sharp guarantees as well as our results above are based on the assumption of deterministically bounded adversaries. To build an even stronger parallel relation with CLTs,

we further allow the adversary to randomize on an unbounded domain, thereby giving sharp in-expectation performance guarantees for OLO with *noisy feedback*. Despite the differences in their settings, the obtained result can be viewed as the algorithmic analogue of a nonasymptotic Wasserstein martingale CLT [Röl18], which is of notable conceptual value. See Appendix E.

**Methodological** Our results are enabled by introducing Stein’s method [Ste72] to OLO. For context, Stein’s method refers to a collection of differential-operator-based techniques characterizing the distance of a distribution to a reference, such as the normal. Within probability theory, it is a canonical framework for proving quantitative limit theorems [CGS10, NP12], often being simpler than the Taylor-approximation-based alternatives. Our key observation is that the strengths of Stein’s method are highly congruent with the analytical difficulties in OLO, such that the proofs of certain probabilistic limit theorems, which are “descriptive” in nature, can be operationalized as “prescriptive” OLO algorithms with sharp rates. Specifically for this work, our construction parallels the remarkably clean proof of a Wasserstein martingale CLT due to Röllin [Röl18], while we believe the applicability of this idea extends much further. See Section 3.

### 1.3 Organization

Section 2 introduces the template of our algorithm, followed by the analysis in Section 3 which is the crux of this work. Section 4 presents an instantiation of the algorithm improving upon OGD. Section 5 concludes this work and discusses future directions. Other quantitative results and discussions of related works are deferred to the appendix; see the beginning of the appendix for its organization.

**Notation**  $Z$  represents the standard normal random variable with cumulative distribution function (CDF)  $\Phi$  and probability density function (PDF)  $\phi$ . More generally,  $\Phi_{\mu,\sigma}$  and  $\phi_{\mu,\sigma}$  represent the CDF and PDF of  $\mathcal{N}(\mu, \sigma^2)$ . For all Lebesgue-measurable  $f : \mathbb{R} \rightarrow \mathbb{R}$ , let  $\|f\|_\infty := \text{ess sup}_{x \in \mathbb{R}} |f(x)|$  be its essential supremum norm (with respect to the Lebesgue measure). Let  $\Pi_{[a,b]}(x) = \arg \min_{u \in [a,b]} |u - x|$  be the projection function to the interval  $[a, b]$ . For integers  $a \leq b$ ,  $[a : b]$  and  $x_{a:b}$  respectively denote the tuples  $[a, \dots, b]$  and  $[x_a, \dots, x_b]$ . Let  $\partial_1$  and  $\partial_2$  be the partial derivatives of a function with respect to the first and the second argument; similarly,  $\partial_{11}, \partial_{12}, \partial_{22}$  are the second order partial derivatives.

## 2 Algorithm

**Stein equation** We begin by introducing the key element of Stein’s method, *Stein equation* and its solution. Appendix B provides the minimal technical background to keep the present work self-contained, while interested readers are referred to several canonical resources for comprehensive treatments [CGS10, Ros11, NP12].

**Definition 2.1** (Solution of Stein equation). *Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a convex and 1-Lipschitz function. For all  $\mu \in \mathbb{R}$  and  $\sigma \in \mathbb{R}_{>0}$ , consider the linear ordinary differential equation*

$$\sigma^2 f'(x) - (x - \mu)f(x) = h(x) - \mathbb{E}_{Z \sim \mathcal{N}(0,1)}[h(\mu + \sigma Z)], \quad \forall x \in \mathbb{R}, \quad (4)$$

*called the Stein equation associated with the distribution  $\mathcal{N}(\mu, \sigma^2)$  and the target function  $h$ . It has a unique bounded solution, and we denote it by  $f_{\mu,\sigma,h} : \mathbb{R} \rightarrow \mathbb{R}$ .*

The function  $f_{\mu,\sigma,h}$  can be expressed as integrals using  $h$ , but due to the modularity of Stein’s method, our analysis is performed at an abstract level without using  $h$ -specific structures. This is a major benefit of Stein’s method over more granular alternatives.

We also note the following properties of  $f_{\mu,\sigma,h}$ . Due to the Lipschitzness of  $h$ , the derivative  $f'_{\mu,\sigma,h}$  is Lipschitz, therefore the second derivative  $f''_{\mu,\sigma,h}$  exists almost everywhere [NP12, Proposition 3.5.1]. Likewise,  $h'$  exists almost everywhere and  $\|h'\|_\infty$  equals the minimum Lipschitz constant of  $h$ . A concept called *Stein factors* [Gau25] refers to upper bounds on  $\|f_{\mu,\sigma,h}\|_\infty$ ,  $\|f'_{\mu,\sigma,h}\|_\infty$  and  $\|f''_{\mu,\sigma,h}\|_\infty$  expressed using  $\|h'\|_\infty$ ; see Lemma B.2. We will immediately use  $\|f_{\mu,\sigma,h}\|_\infty \leq \|h'\|_\infty \leq 1$ .

---

**Algorithm 1** 1D fixed-time bounded-domain OLO via Stein’s method.

---

**Require:** A convex and 1-Lipschitz function  $h : \mathbb{R} \rightarrow \mathbb{R}$ .

- 1: Define  $s_0 := 0$ , and pick an arbitrary  $\rho_0 \in \mathbb{R}_{>0}$ .
- 2: **for**  $t = 1, \dots, T$  **do**
- 3:   Pick an arbitrary  $\rho_t \in [0, \rho_{t-1}]$ , and we further require  $\rho_t > 0$  if  $t < T$ . Define  $c_t := \rho_{t-1}^2 - \rho_t^2 \in \mathbb{R}_{\geq 0}$ .
- 4:   Based on the function  $f_{\mu, \sigma, h}$  from Definition 2.1, output

$$x_t = \mathbb{E}_{Z \sim \mathcal{N}(0,1)} [f_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z)]. \quad (5)$$

This is a valid output on the domain  $[-1, 1]$  since  $\|f_{s_{t-1}, \rho_{t-1}, h}\|_{\infty} \leq 1$ .

- 5:   Observe the loss gradient  $g_t \in \mathbb{R}$  and define  $s_t := s_{t-1} + g_t \in \mathbb{R}$ .
  - 6: **end for**
- 

**Algorithm** With the above, we present Algorithm 1. A number of remarks are in order.

- The function  $h$  serves as a proxy of  $\psi_T^*$  from the targeted Eq.(1), specifying the desirable tradeoff on the total loss. The algorithm is invariant to vertical shifts on  $h$ , i.e.,  $h(x) \leftarrow h(x) + \text{constant}$  for all  $x \in \mathbb{R}$ .
- The parameter  $\rho_t$  serves as a guess on the adversary’s “standard deviation to go”,  $\sqrt{\sum_{i=t+1}^T g_i^2}$ , and it is allowed to depend on  $g_{1:t-1}$ . Requiring  $\rho_t > 0$  for all  $t < T$  is to ensure the well-posedness of Eq.(5). Analogously, the induced  $c_t = \rho_{t-1}^2 - \rho_t^2$  is a guess on the adversary’s instantaneous variance,  $g_t^2$ .

For intuition, if it is known that  $|g_t| \leq 1$  for all  $t$ , then one could simply set  $\rho_t = \sqrt{T-t}$  thus  $c_t = 1$ . We remark that in its full generality, hard constraints on the adversary are not required, as the associated performance guarantee (Theorem 3) will hold against arbitrary unconstrained adversaries, taking the misspecification error of  $\rho_{0:T}$  into account.

While the output Eq.(5) itself is sufficient for our analysis, its reliance on  $f_{\mu, \sigma, h}$  may obfuscate the intuition. The following lemma spells out the explicit dependence of  $x_t$  on the function  $h$ , and based on that, we discuss the intuition by analyzing its small- $c_t$  approximation.

**Lemma 2.1** (Lemma G.1, simplified). *Let  $\text{Exp}(1)$  be the exponential distribution with parameter 1. The output Eq.(5) of Algorithm 1 is equivalent to*

$$x_t = -\mathbb{E}_{Z \sim \mathcal{N}(0,1); \tau \sim \text{Exp}(1)} \left[ h' \left( s_{t-1} + \sqrt{\rho_{t-1}^2 - e^{-2\tau} c_t} Z \right) \right].$$

Suppose for some  $t \in [1 : T]$  we have  $\rho_{t-1} = \sqrt{T-t+1}$  and  $c_t \approx 0$ . By Lemma 2.1,  $x_t \approx -\mathbb{E}_Z[h'(s_{t-1} + \sqrt{T-t+1}Z)]$ , which we define as the output  $\tilde{x}_t$  of a simpler,  $c_t$ -independent analogue of Algorithm 1. This can be justified from the following two perspectives.

**Universality and regularization** The fundamental difficulty in OLO originates from the uncertainty in the adversary. Consider the  $t$ -th round where  $g_{1:t-1}$  is known but  $g_{t:T}$  is not. If the algorithm is given the oracle knowledge that  $g_{t:T}$  is sampled iid from a known distribution  $\mathcal{D}$ , then the distribution of  $\sum_{t=1}^T g_t$  is also known.  $x_t$  can then be assigned as the expectation of the best-in-hindsight decision,  $x_t \leftarrow \mathbb{E}[\arg \min_{u \in [-1,1]} \sum_{t=1}^T g_t u]$ . Tie-breaking of  $\arg \min$  is arbitrary assuming the density exists.

The intuition of  $\tilde{x}_t$  follows from this idea, but with two more ingredients. First, if CLT can be applied, then the distribution of  $\sum_{t=1}^T g_t$  is insensitive to  $\mathcal{D}$ , as it can be approximated by  $\mathcal{N}(s_{t-1}, T-t+1)$  as long as  $T \gg t$ . This embodies the universality principle in high-dimensional probability [O’D14, VH14]. Second, the function  $h$  introduces certain inductive bias, therefore the algorithm should minimize the regularized objective,

$$\tilde{x}_t = \mathbb{E}_{X \sim \mathcal{N}(s_{t-1}, T-t+1)} \left[ \arg \min_{u \in [-1,1]} Xu + h^*(-u) \right] = -\mathbb{E}_Z[h'(s_{t-1} + \sqrt{T-t+1}Z)]. \quad (6)$$

Picking  $h(s) = |s|$  gives a constant convex conjugate  $h^*$ , corresponding to the case without regularization.



**Continuous time dynamics** Another interpretation of  $\tilde{x}_t$  is via the *potential method* in online learning [CBL06, Ora25]. Let  $\varphi_T : (-\infty, T] \times \mathbb{R} \rightarrow \mathbb{R}$  be the *heat potential*

$$\varphi_T(t, s) := \mathbb{E}_Z[h(s + \sqrt{T-t}Z)],$$

which is the unique solution of the *backward heat equation* (BHE)  $(\partial_1 + \frac{1}{2}\partial_{22})\varphi_T = 0$  under the terminal condition  $\varphi_T(T, s) = h(s)$ . It is straightforward to verify  $\tilde{x}_t = -\partial_2\varphi_T(t-1, s_{t-1})$ . This is the continuous time limit of the potential method, whose idea underlies a growing stream of recent works on better-than-big-O rates in OLO [DK20, KKW20, GHP22, ZCP22, HLP23, HLP24, ZYCP24]. It is also connected to Eq.(6) via the equivalence between the potential method and the *Follow the Regularized Leader* (FTRL) algorithm on linear losses [Ora25, Section 7.3].

Going from  $\tilde{x}_t$  to  $x_t$ , the key challenge is rigorously accounting for nonzero  $c_t$ . In particular, the bottleneck has been the *discretization argument*: showing that a suitable modification of  $\tilde{x}_t$  (equivalently, a numerical scheme on the BHE) works well against discrete time adversaries. Stein’s method offers substantial advantages for this purpose.

**Computational complexity** By Lemma 2.1,  $x_t$  can be implemented by a two-dimensional Gaussian quadrature whose time complexity is independent of  $t$  and  $T$ . Assuming this computational subroutine is exact (which is the convention of the field), the conclusion is that Algorithm 1 takes  $O(1)$  time per round. In addition, the  $h$  function of interest is sometimes simple enough such that  $x_t$  can be explicitly reduced to common Gaussian-integral-type special functions. Examples include the error function and the *Owen’s T function*, for which fast and stable specialized implementations exist. This is demonstrated through concrete special cases in Appendix H.

### 3 Analysis

The main result of this work is Theorem 3, the total loss upper bound of Algorithm 1. An equivalent regret bound follows from the loss-regret duality (Corollary 15 in Appendix G). Applications and extensions are presented in Section 4, Appendix C, Appendix D and Appendix E.

**Theorem 3** (Main result; upper bound on  $\text{Loss}_T$ ). *Against all adversaries (potentially unbounded a priori), Algorithm 1 given any convex and 1-Lipschitz function  $h$  guarantees*

$$\text{Loss}_T \leq \underbrace{-\mathbb{E}_Z \left[ h \left( \sum_{t=1}^T g_t + \rho_T Z \right) \right]}_{=:-\bar{\psi}_T^*(-\sum_{t=1}^T g_t)} + \underbrace{\mathbb{E}_Z[h(\rho_0 Z)] + \sum_{t=1}^T \left( \sqrt{\frac{2}{\pi}} \frac{\max\{g_t^2 - c_t, 0\}}{\rho_{t-1}} + \frac{2c_t |g_t| + |g_t^3|}{\rho_{t-1}^2} \right)}_{=:\text{err}_T}.$$

To parse this result, we divide the RHS into two parts, respectively defined as  $-\bar{\psi}_T^*(-\sum_{t=1}^T g_t)$  and  $\text{err}_T$ .

- $-\bar{\psi}_T^*(-\sum_{t=1}^T g_t)$  is the primary component of the bound due to the continuous time dynamics of the OLO game. The most interpretable case is when  $\rho_T = 0$  and  $h$  is even, meaning that the function  $\bar{\psi}_T^*$  is a vertical shift of  $h$ . Generally,  $\bar{\psi}_T^*$  mirrors  $\psi_T^*$  from the targeted total loss bound Eq.(1): it is convex, 1-Lipschitz, and analogous to Cover’s achievability condition Eq.(2) we have  $\mathbb{E}_{X \sim \mathcal{N}(0, \rho_0^2 - \rho_T^2)}[\bar{\psi}_T^*(X)] = 0$ . Achieving  $\text{Loss}_T \leq -\bar{\psi}_T^*(-\sum_{t=1}^T g_t)$  would be ideal, but as discussed previously, additional discretization error is necessary. Controlling this error has been the bottleneck in prior works.
- The key value of Theorem 3 is showing that the discretization error  $\text{err}_T$  of Algorithm 1 can be controlled uniformly over all convex and Lipschitz  $h$  functions. Moreover,  $\text{err}_T$  is a lower order term against many constrained adversary classes, thereby addressing the aforementioned bottleneck. For example, against the simple adversary class satisfying  $|g_t| \leq 1$  for all  $t$ , setting  $h(x) = |x|$  and  $\rho_t = \sqrt{T-t}$  in Algorithm 1 gives  $\bar{\psi}_T^*(s) = |s| - \sqrt{\frac{2}{\pi}}T$  for all  $s \in \mathbb{R}$  while  $\text{err}_T = O(\log T)$ . This corresponds to the particular performance tradeoff in Example 1; also see Section 4.1 for the equivalent argument on the regret.

The following result shows that against adversaries satisfying  $|g_t| \leq 1$ , Theorem 3 with  $\rho_t = \sqrt{T-t}$  is optimal modulo the additive  $O(\log T)$  factor. The proof is based on the standard nonasymptotic Wasserstein CLT for iid sums (e.g., [Ros11, Theorem 3.2]), presented in Appendix G.

**Theorem 4** (Lower bound on  $\text{Loss}_T$ ). *There exists an absolute constant  $c > 0$  such that the following holds. For any OLO algorithm and any 1-Lipschitz function  $h : \mathbb{R} \rightarrow \mathbb{R}$ , there exists a Boolean adversary ( $g_t \in \{-1, 1\}$  for all  $t$ ) inducing*

$$\text{Loss}_T > -h\left(\sum_{t=1}^T g_t\right) + \mathbb{E}_Z[h(\sqrt{T}Z)] - c.$$

**Technical sketch** The crux of this work is the proof of Theorem 3, which is short and considerably different from all known analyses in OLO. It is inspired by the work of Röllin [Röl18] combining Lindeberg’s method and Stein’s method to prove a quantitative (i.e., nonasymptotic) Wasserstein martingale CLT; that is, bounding the 1-Wasserstein distance between a martingale and an appropriately scaled normal distribution.

By its dual representation, the 1-Wasserstein distance is an *integral probability metric* over the class of 1-Lipschitz functions. Fixing any function  $h$  in this class, our key observation is that it is still valid to apply Röllin’s argument on arbitrary data sequences generated by the OLO adversary. While such sequences are not martingales, the difference in the analysis is an additional bias term which is precisely the  $\text{Loss}_T$  induced by Algorithm 1. In other words, Theorem 3 (and its generalization allowing randomized adversaries) may be viewed as an algorithmic analogue of Röllin’s CLT, where the function  $h$  associated with the algorithm serves as the *witness function* measuring the statistical distance. This will be further discussed in Appendix E.

While the 1-Wasserstein distance is defined over all 1-Lipschitz functions, in Theorem 3 we also require  $h$  to be convex, differing from Röllin’s argument in the probabilistic context. This is motivated by the duality argument from Section 1, and technically, it ensures that the function  $f_{\mu, \sigma, h}$  is nonincreasing, leading to a low  $\text{err}_T$  through Eq.(8) in the proof. Nonconvex performance tradeoffs in OLO is generally a more delicate subject, which we further discuss in Appendix F.

*Proof of Theorem 3.* The proof is divided into the following steps. It uses several standard technical lemmas of Stein’s method, provided in Appendix B.

**Step 1** The first step is the telescoping argument from Lindeberg’s method. Let  $Z_0, \dots, Z_T$  be a sequence of iid standard normal random variables. For all  $t \in [0 : T]$ , define

$$m_t := \rho_T Z_0 + \sum_{i=1}^T (g_i \mathbf{1}[i \leq t] + \sqrt{c_i} Z_i \mathbf{1}[i > t]) \stackrel{d}{=} s_t + \rho_t Z_t,$$

where  $\stackrel{d}{=}$  denotes the equality of marginal distributions (with respect to the randomness of  $Z_{0:T}$ ). Starting from the telescopic sum  $h(m_T) - h(m_0) = \sum_{t=1}^T (h(m_t) - h(m_{t-1}))$ , we take the expectation with respect to  $Z_{0:T}$  to obtain

$$\mathbb{E}_{Z_T} \left[ h \left( \sum_{t=1}^T g_t + \rho_T Z_T \right) \right] = \mathbb{E}_{Z_0} [h(\rho_0 Z_0)] + \sum_{t=1}^T (\mathbb{E}_{Z_t} [h(s_t + \rho_t Z_t)] - \mathbb{E}_{Z_{t-1}} [h(s_{t-1} + \rho_{t-1} Z_{t-1})]). \quad (7)$$

From this point the subscript of  $Z$  is dropped.

**Step 2** The next step is using the Stein equation and *Stein’s lemma* (Lemma B.1) to simplify the expression under the sum for each  $t \in [1 : T]$ . Since  $f_{\mu, \sigma, h}$  satisfies the Stein equation Eq.(4),

$$\rho_{t-1}^2 f'_{s_{t-1}, \rho_{t-1}, h}(x) - (x - s_{t-1}) f_{s_{t-1}, \rho_{t-1}, h}(x) = h(x) - \mathbb{E}_Z [h(s_{t-1} + \rho_{t-1} Z)], \quad \forall x \in \mathbb{R}.$$

Plugging in  $x \leftarrow s_t + \rho_t Z$  and taking the expectation with respect to  $Z$ ,

$$\begin{aligned} & \mathbb{E}_Z [h(s_t + \rho_t Z)] - \mathbb{E}_Z [h(s_{t-1} + \rho_{t-1} Z)] \\ &= \mathbb{E}_Z \left[ \rho_{t-1}^2 f'_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) - (s_t + \rho_t Z - s_{t-1}) f_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) \right] \\ &= \mathbb{E}_Z \left[ c_t f'_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) - g_t f_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) \right] \end{aligned}$$



$$+ \underbrace{\mathbb{E}_Z \left[ \rho_t^2 f'_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) - \rho_t Z f_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) \right]}_{=0}. \quad (\text{Lemma B.1})$$

On the last line, the integrability assumption required by Lemma B.1 follows from Lemma B.2 (Stein factors).

**Step 3** The final step is plugging in Stein factors and exploiting the convexity of  $h$ . Recall that  $f'_{s_{t-1}, \rho_{t-1}, h}$  is absolutely continuous. By Taylor's theorem with integral remainder,

$$\begin{aligned} f'_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) &= f'_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z) + g_t \int_0^1 f''_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z + \lambda g_t) d\lambda, \\ f_{s_{t-1}, \rho_{t-1}, h}(s_t + \rho_t Z) &= f_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z) + g_t f'_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z) \\ &\quad + g_t^2 \int_0^1 (1 - \lambda) f''_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z + \lambda g_t) d\lambda. \end{aligned}$$

Therefore, combining the above,

$$\begin{aligned} &\mathbb{E}_Z [h(s_t + \rho_t Z)] - \mathbb{E}_Z [h(s_{t-1} + \rho_{t-1} Z)] \\ &= -g_t \underbrace{\mathbb{E}_Z [f_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z)]}_{=x_t} + \underbrace{(c_t - g_t^2) \mathbb{E}_Z [f'_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z)]}_{=: \triangle} \\ &\quad + \underbrace{\mathbb{E}_Z \left[ c_t g_t \int_0^1 f''_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z + \lambda g_t) d\lambda - g_t^3 \int_0^1 (1 - \lambda) f''_{s_{t-1}, \rho_{t-1}, h}(s_{t-1} + \rho_t Z + \lambda g_t) d\lambda \right]}_{=: \diamond}. \end{aligned}$$

For the terms  $\triangle$  and  $\diamond$  on the RHS, we use Lemma B.2 (Stein factors) and Lemma B.3 ( $f'_{s_{t-1}, \rho_{t-1}, h}$  is non-positive),

$$\begin{aligned} \triangle &\leq \sqrt{\frac{2}{\pi}} \rho_{t-1}^{-1} \cdot \max\{g_t^2 - c_t, 0\}, \\ \diamond &\leq \frac{2c_t |g_t|}{\rho_{t-1}^2} + \frac{2|g_t^3|}{\rho_{t-1}^2} \int_0^1 (1 - \lambda) d\lambda = \frac{2c_t |g_t| + |g_t^3|}{\rho_{t-1}^2}. \end{aligned} \quad (8)$$

Combining all three steps completes the proof.  $\square$

## 4 Application: Dominating OGD

This section instantiates our results with specific  $h$  functions. We will assume  $|g_t| \leq 1$  for simplicity, therefore setting  $\rho_t = \sqrt{T-t}$  gives  $\text{err}_T = O(\log T)$ . Proofs are deferred to Appendix G, and the closed forms of the output  $x_t$  are presented in Appendix H.

### 4.1 Absolute Value Function

As a warm-up, picking  $h(x) = |x|$  corresponds to minimizing the uniform regret in Example 1. The following regret bound matches existing ones [KKW20, GHP22] in the optimal leading order term  $\sqrt{\frac{2}{\pi}T}$ .

**Corollary 5** (Regret: absolute value). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = |x|$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \text{Regret}_T^{\text{unif}} \leq \sqrt{\frac{2}{\pi}T} + O(\log T), \quad \forall u \in [-1, 1].$$

## 4.2 Huber Function

A highlight of this work is picking  $h$  as the *Huber function*, whose inductive bias matches that of the OGD baseline. The following regret bound generalizes Corollary 5 in the limit of  $\eta \rightarrow \infty$ . By the regret version of Theorem 4 (Corollary 16 in Appendix G), it cannot be improved by  $\Omega(\log T)$  uniformly over  $u \in [-1, 1]$ .

**Corollary 6** (Regret: Huber). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = \frac{\eta}{2}x^2 \cdot \mathbf{1}[|x| \leq \eta^{-1}] + (|x| - \frac{1}{2\eta}) \cdot \mathbf{1}[|x| > \eta^{-1}]$  for  $\eta = \frac{\alpha}{\sqrt{T}}$  and  $\alpha \in \mathbb{R}_{>0}$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{u^2}{2\alpha} + \left( \alpha + \frac{1}{\alpha} \right) \Phi\left(\frac{1}{\alpha}\right) + \phi\left(\frac{1}{\alpha}\right) - \frac{1}{2}\alpha - \frac{1}{\alpha} \right]}_{=: \gamma_{\text{Huber}}(u, \alpha)} \sqrt{T} + O(\log T), \quad \forall u \in [-1, 1].$$

Since the algorithm takes  $O(1)$  time per round, we compare it to the OGD baseline with constant learning rate  $\eta = \frac{\alpha}{\sqrt{T}}$  which outputs  $x_t = \Pi_{[-1,1]}(x_{t-1} - \eta g_{t-1})$ . Against the same adversary class, the OGD baseline guarantees  $\text{Regret}_T(u) \leq \frac{1}{2}(\alpha^{-1}u + \alpha)\sqrt{T}$  [Ora25, Theorem 2.13], where the prefactor of  $\sqrt{T}$  will be denoted as  $\gamma_{\text{OGD}}(u, \alpha)$ . To our knowledge, this is the sharpest known regret bound of OGD that depends on  $u$ ,  $\alpha$  and  $T$ . As  $O(\log T)$  in Corollary 6 is a lower order term, we focus on the prefactors of  $\sqrt{T}$  and define the *margin of improvement* as

$$\text{Gap}_{\text{OGD}}(\alpha) := \gamma_{\text{OGD}}(u, \alpha) - \gamma_{\text{Huber}}(u, \alpha) = \left( \alpha + \frac{1}{\alpha} \right) \left[ 1 - \Phi\left(\frac{1}{\alpha}\right) \right] - \phi\left(\frac{1}{\alpha}\right).$$

The following observations can be made; also see Figure 1 in Section 1 for visualizations.

- Consistent with its notation,  $\text{Gap}_{\text{OGD}}(\alpha)$  is independent of  $u$ , and it is always positive due to a standard *Mills ratio* estimate (Lemma G.7). Therefore Corollary 6 dominates the regret bound of OGD for all  $u \in [-1, 1]$ , while they are both incomparable to sharp uniform regret bounds, including our Corollary 5 and the results of [KKW20, GHP22]. By Eq.(1), this ensures that the  $\text{Loss}_T$  upper bound of our algorithm is lower than that of OGD regardless of  $\sum_{t=1}^T g_t$  – that is, regardless of the difficulty of the adversary.
- $\text{Gap}_{\text{OGD}}(\alpha)$  grows unboundedly with  $\alpha$ . For all  $u \in [-1, 1]$ ,  $\gamma_{\text{OGD}}(u, \alpha)$  diverges as  $\alpha \rightarrow \infty$ , while

$$\lim_{\alpha \rightarrow \infty} \gamma_{\text{Huber}}(u, \alpha) = \lim_{\alpha \rightarrow \infty} \alpha \left[ \Phi\left(\frac{1}{\alpha}\right) - \Phi(0) \right] + \phi\left(\frac{1}{\alpha}\right) = 2\phi(0) = \sqrt{\frac{2}{\pi}},$$

recovering Corollary 5.

- To reason about the “effective learning rate” of our algorithm, consider  $\tilde{x}_t$  from Eq.(6) which is the small- $c_t$  approximation of the output  $x_t$ . Suppose for simplicity that  $s_{t-1} = 0$  for some  $t \in [1 : T]$ . By symmetry, we have  $\tilde{x}_t = 0$ , meaning that  $\tilde{x}_{t+1}$  only depends on  $g_t$ ,  $\eta$  and  $\rho_t$ . Linearizing it with respect to  $g_t$  gives (see Remark H.1 for derivation)

$$\tilde{x}_{t+1} - \tilde{x}_t = -\eta \cdot \text{erf}\left(\frac{1}{\sqrt{2}\eta\rho_t}\right) g_t + o(g_t).$$

Treating the absolute value of the linear coefficient as the effective learning rate, we see that it is increasing with respect to  $t$  and eventually reaches the nominal value  $\eta$ , the learning rate of the OGD baseline. This is different from typical OLO algorithms where the learning rate is constant (if  $T$  is known) or decreasing (if  $T$  is unknown).

Other interesting cases of  $h$  are analyzed in the appendix.

## 5 Conclusion

Utilizing Stein’s method, this paper presents a general, sharp and analytically scalable framework to design and analyze OLO algorithms. Inspired by Röllin’s proof of a Wasserstein martingale CLT [Röl18], our main result

is a computationally efficient improvement to a seminal dynamic programming algorithm by Cover [Cov66]. Starting from there, we demonstrate a number of more specific quantitative benefits, including (i) dominating the total loss upper bounds of OGD and MWU; (ii) achieving optimal tradeoffs between the total loss and the uniform regret; and (iii) handling OLO with noisy feedback. Together, they imply an intriguing algorithmic relation between adversarial online learning and probabilistic limit theorems, whose applicability might extend much further.

This work only studies the basic setting of OLO with a one-dimensional bounded domain and a known time horizon. Generalizing such conditions (that is, achieving additively sharp rates while allowing general dimension, unknown time horizon or unbounded domain) would be a highly valuable direction for future works. Other than OLO, the idea of universality may also benefit online learning problems with curved losses or bandit feedback, where little is known about better-than-big-O optimality. Finally, besides giving improvements to existing problems, we hope the connection to probabilistic limit theorems could help motivate and answer new problems in adversarial online learning, such as operationalizing the Kolmogorov distance versions of the martingale CLT.

## Acknowledgment

This work is assisted by AI tools for literature search, preliminary calculation and writing. The authors take full responsibility for the originality and correctness of the results.

## References

- [ABRT08] Jacob Abernethy, Peter L Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. In *Conference on Learning Theory*, pages 415–423, 2008.
- [AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of computing*, 8(1):121–164, 2012.
- [AR25] Shubhada Agrawal and Aaditya Ramdas. Eventually LIL regret: Almost sure  $\ln \ln t$  regret for a sub-Gaussian mixture on unbounded data. *arXiv preprint arXiv:2512.12325*, 2025.
- [BEZ20] Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Finite-time 4-expert prediction problem. *Communications in Partial Differential Equations*, 45(7):714–757, 2020.
- [BK25] Alankrita Bhatt and Victoria Kostina. Prediction with expert advice under additive noise. In *Advances in Neural Information Processing Systems*, 2025.
- [CBL06] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- [CGS10] Louis HY Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein’s method*. Springer Science & Business Media, 2010.
- [CO18] Ashok Cutkosky and Francesco Orabona. Black-box reductions for parameter-free online learning in banach spaces. In *Conference On Learning Theory*, pages 1493–1529. PMLR, 2018.
- [Cov66] Thomas M Cover. Behavior of sequential predictors of binary sequences. Technical report, Stanford University, 1966. <https://isl.stanford.edu/~cover/papers/paper3.pdf>.
- [DHS11] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(7), 2011.
- [DK20] Nadejda Drenska and Robert V Kohn. Prediction with expert advice: A PDE perspective. *Journal of Nonlinear Science*, 30(1):137–173, 2020.
- [EDKMW08] Eyal Even-Dar, Michael Kearns, Yishay Mansour, and Jennifer Wortman. Regret to the best vs. regret to the average. *Machine Learning*, 72(1):21–37, 2008.

- [FRS18] Dylan J Foster, Alexander Rakhlin, and Karthik Sridharan. Online learning: Sufficient statistics and the Burkholder method. In *Conference On Learning Theory*, pages 3028–3064. PMLR, 2018.
- [Gau25] Robert E Gaunt. On Stein factors in Stein’s method for normal approximation. *Statistics & Probability Letters*, 219:110339, 2025.
- [GHP22] Laura Greenstreet, Nicholas JA Harvey, and Victor Sanches Portella. Efficient and optimal fixed-time regret with two experts. In *International Conference on Algorithmic Learning Theory*, pages 436–464. PMLR, 2022.
- [Gor41] Robert D Gordon. Values of Mills’ ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *The Annals of Mathematical Statistics*, 12(3):364–366, 1941.
- [GPS16] Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Towards optimal algorithms for prediction with expert advice. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 528–547. SIAM, 2016.
- [GPS17] Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Tight lower bounds for multiplicative weights algorithmic families. In *44th International Colloquium on Automata, Languages, and Programming*, pages 48–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2017.
- [HLP24] Nicholas JA Harvey, Christopher Liaw, and Victor S Portella. Continuous prediction with experts’ advice. *Journal of Machine Learning Research*, 25(228):1–32, 2024.
- [HLPR23] Nicholas JA Harvey, Christopher Liaw, Edwin Perkins, and Sikander Randhawa. Optimal anytime regret with two experts. *Mathematical Statistics and Learning*, 6(1):87–142, 2023.
- [JO19] Kwang-Sung Jun and Francesco Orabona. Parameter-free online convex optimization with sub-exponential noise. In *Conference on learning theory*, pages 1802–1823. PMLR, 2019.
- [KKW20] Vladimir A Kobzar, Robert V Kohn, and Zhilei Wang. New potential-based bounds for prediction with expert advice. In *Conference on Learning Theory*, pages 2370–2405. PMLR, 2020.
- [Koo13] Wouter M Koolen. The Pareto regret frontier. *Advances in Neural Information Processing Systems*, 26, 2013.
- [KP11] Michael Kapralov and Rina Panigrahy. Prediction strategies without loss. *Advances in Neural Information Processing Systems*, 24, 2011.
- [KP17] Anna R Karlin and Yuval Peres. *Game theory, alive*, volume 101. American Mathematical Soc., 2017.
- [Liu25] Zijian Liu. Online convex optimization with heavy tails: Old algorithms, new regrets, and applications. *arXiv preprint arXiv:2508.07473*, 2025.
- [LW94] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.
- [MO14] H Brendan McMahan and Francesco Orabona. Unconstrained online linear learning in Hilbert spaces: Minimax algorithms and normal approximations. In *Conference on Learning Theory*, pages 1020–1039. PMLR, 2014.
- [NP12] Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus: from Stein’s method to universality*, volume 192. Cambridge University Press, 2012.
- [O’D14] Ryan O’Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.
- [OJ23] Francesco Orabona and Kwang-Sung Jun. Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*, 70(1):436–455, 2023.

- [Ora25] Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213v7*, 2025.
- [Owe80] Donald B Owen. A table of normal integrals: A table. *Communications in Statistics-Simulation and Computation*, 9(4):389–419, 1980.
- [Röl18] Adrian Röllin. On quantitative bounds in the mean martingale central limit theorem. *Statistics & Probability Letters*, 138:171–176, 2018.
- [Ros11] Nathan Ross. Fundamentals of Stein’s method. *Probability Surveys*, 8:210–293, 2011.
- [RS14] Alexander Rakhlin and Karthik Sridharan. Statistical learning and sequential prediction. *Unpublished lecture notes*, 2014. [http://www.mit.edu/~rakhlin/courses/stat928/stat928\\_notes.pdf](http://www.mit.edu/~rakhlin/courses/stat928/stat928_notes.pdf).
- [RS17] Alexander Rakhlin and Karthik Sridharan. On equivalence of martingale tail bounds and deterministic regret inequalities. In *Conference on Learning Theory*, pages 1704–1722. PMLR, 2017.
- [Ste72] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory*, volume 6, pages 583–603. University of California Press, 1972.
- [Ste86] Charles Stein. Approximate computation of expectations. *Lecture Notes-Monograph Series*, 7:i–164, 1986.
- [SV19] Glenn Shafer and Vladimir Vovk. *Game-theoretic foundations for probability and finance*. John Wiley & Sons, 2019.
- [vdH19] Dirk van der Hoeven. User-specified local differential privacy in unconstrained adaptive online learning. *Advances in Neural Information Processing Systems*, 32, 2019.
- [VH14] Ramon Van Handel. Probability in high dimension. *Unpublished lecture notes*, 2014. <https://web.math.princeton.edu/~rvan/APC550.pdf>.
- [WSR24] Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27, 2024.
- [ZC22] Jiujia Zhang and Ashok Cutkosky. Parameter-free regret in high probability with heavy tails. *Advances in Neural Information Processing Systems*, 35:8000–8012, 2022.
- [ZCP22] Zhiyu Zhang, Ashok Cutkosky, and Ioannis Paschalidis. PDE-based optimal strategy for unconstrained online learning. In *International Conference on Machine Learning*, pages 26085–26115. PMLR, 2022.
- [Zhu14] Kangping Zhu. *Two problems in applications of PDE*. PhD thesis, New York University, 2014.
- [Zin03] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936, 2003.
- [ZYCP24] Zhiyu Zhang, Heng Yang, Ashok Cutkosky, and Ioannis C Paschalidis. Improving adaptive online learning using refined discretization. In *International Conference on Algorithmic Learning Theory*, pages 1208–1233. PMLR, 2024.

# Appendix

**Organization** Appendix A discusses related works, followed by the preliminaries of Stein’s method in Appendix B. Appendix C continues the argument in Section 4, showing that a special case of Algorithm 1 improves upon MWU. Appendix D shows that Algorithm 1 can attain a continuum of optimal two-point tradeoffs between  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$ . Appendix E extends our results to OLO with noisy feedback and further compares our result to Röllin’s martingale CLT. Appendix F discusses the generalization to nonconvex performance tradeoffs in OLO. Appendix G contains omitted proofs. Finally, Appendix H presents the closed-form expressions of  $f_{\mu,\sigma,h}$  and  $x_t$  in special cases, thus improving the concreteness of our algorithm.

## A Related Works

From a quantitative perspective, our work is related to two general research themes in online learning: (i) achieving leading-constant-optimal upper bounds on the uniform regret; and (ii) achieving comparator-dependent regret bounds. Besides those, from a methodological perspective, our work touches upon the deep connection between adversarial online learning and martingale inequalities. The latter has been studied by a line of works whose objectives and techniques are very different from ours.

**Sharp uniform regret bounds** Consider the uniform regret  $\text{Regret}_T^{\text{unif}}$  from Example 1 which is the maximum of the regret over all comparators. While most works focus on the conventional big-O optimality when bounding the  $\text{Regret}_T^{\text{unif}}$ , here we note several exceptional settings where sharper bounds have been achieved. Specifically for  $\text{Regret}_T^{\text{unif}}$ , this precisely means achieving the optimal constant multiplier on the leading order term of the bound, without being subsumed by big O.

As discussed in Section 1.1, for the known-horizon version of OLO on a one-dimensional bounded domain (i.e., the setting of this work), [Cov66] gave the first algorithm achieving the optimal leading constant, and computational efficient improvements were later given by [KKW20, GHP22]. Generalizing the domain to higher dimensions while still achieving the leading constant optimality is however a delicate subject requiring a careful treatment of the shape of the domain.

- The simple case is the  $d$ -dimensional Euclidean norm ball: for any  $d \geq 2$ , [ABRT08] showed that OGD with constant learning rate achieves the optimal constant multiplier in front of the leading order term  $\sqrt{T}$ .
- The difficult case is the  $d$ -dimensional probability simplex (that is, the LEA problem with  $d$  experts): MWU with constant learning rate is leading-constant-optimal in the regime of both  $d \rightarrow \infty$  and  $T \rightarrow \infty$  [CBL06, Theorem 3.7]. Only taking  $T \rightarrow \infty$ , precisely leading-constant-optimal algorithms for  $d = 2, 3, 4$  are given by [Cov66], [GPS16] and [BEZ20] respectively. The regime in between remains open.

In a different direction, [HLPR23] achieved leading constant optimality for the *anytime* version of OLO (i.e.,  $T$  is unknown by the algorithm) on a one-dimensional bounded domain. A common technical theme shared by these recent progresses is a continuous time perspective on the OLO problem, pioneered by [Zhu14, DK20, HLPR23] and realized by tools from PDE and stochastic analysis. The bottleneck is the discretization argument on the PDE, discussed in Section 2, for which Stein’s method offers analytical advantages.

**Parameter-free OLO** Going beyond the uniform regret, there is a subfield within OLO studying the realization of  $u$ -dependent upper bounds on  $\text{Regret}_T(u)$ . Depending on the context, this is often called *parameter-free online learning* or *comparator-adaptive online learning*. See [Ora25, Section 10] for an exposition.

A key motivation of this topic originates from the practical need to handle unbounded domains. In one dimension, it means the domain of the algorithm is  $\mathbb{R}$  as opposed to  $[-1, 1]$ . The definition of  $\text{Regret}_T^{\text{unif}}$  becomes vacuous, therefore the typical quantitative objective is to achieve a  $u$ -dependent upper bound on  $\text{Regret}_T(u)$  whose dependence on both  $|u| \rightarrow \infty$  and  $T \rightarrow \infty$  is asymptotically optimal. Based on a similar technical backbone as [HLPR23], an algorithm with the optimal leading constant is given by [ZCP22]. Such results are also meaningful for the settings with a bounded domain, due to a projection technique of [CO18] showing that if a particular  $u$ -dependent regret bound is achievable in the unconstrained setting for all  $u \in \mathbb{R}$ , then it is also achievable in the constrained setting for all  $u$  within the domain of the algorithm.



Compared to the state of the art in parameter-free OLO, our work has a subtle but important difference: rather than considering both  $|u| \rightarrow \infty$  and  $T \rightarrow \infty$ , we keep  $u$  fixed on a bounded domain and only consider the asymptotic regime of  $T \rightarrow \infty$ . Via the loss-regret duality (Lemma G.5), this is closely related to the objectives of [Cov66], which we use to motivate our techniques. Appendix D shows that an application of such techniques improves upon the combination of [ZCP22] and [CO18].

**Connection to martingale concentration** Conceptually, our work can be viewed as another evidence of the deep connection between adversarial online learning and martingale inequalities. In a broader context, such a connection has been studied by many others, including [RS17, FRS18, JO19, OJ23, WSR24, AR25], just to name a few. There is also an emerging field called game-theoretic probability [SV19] dedicated to the systematic study of this connection.

Our work differs from the mentioned works in several major aspects. The most notable one is that while these works have characterized various intriguing relations between game-theoretic regret bounds and martingale *concentration inequalities* (e.g., exponential tail probability bounds), we focus on such a relation with martingale CLTs on the probabilistic side. This allows us to obtain OLO algorithms with additively sharp rates which is beyond the reach of existing concentration-based arguments; for example, the seminal work of [RS17] proposed a conversion from martingale concentration inequalities to the achievability of certain regret bounds, which, as noted by the authors, pays a price through the multiplying constants. Other differences include: (i) using Stein’s method in adversarial online learning has not been considered by prior works; and (ii) we emphasize the somewhat unusual perspective of total loss minimization in OLO which includes uniform regret minimization as a special case, while the latter is often the starting point of the works mentioned above.

## B Preliminaries of Stein’s Method

This section presents the preliminaries of Stein’s method, based on [CGS10, NP12].

**Stein’s lemma** The motivation of Stein’s method comes from the celebrated Stein’s lemma. We only need its special case for one-dimensional normal distributions [CGS10, Lemma 2.1].

**Lemma B.1** (Stein’s lemma). *Let  $X \sim \mathcal{N}(\mu, \sigma^2)$ . For all absolutely continuous function  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\mathbb{E}[|f'(X)|] < \infty$ , we have*

$$\sigma^2 \mathbb{E}[f'(X)] = \mathbb{E}[(X - \mu)f(X)].$$

**Solution of Stein equation** Consider the Stein equation Eq.(4) with a convex 1-Lipschitz function  $h$ ,

$$\sigma^2 f'(x) - (x - \mu)f(x) = h(x) - \mathbb{E}_Z[h(\mu + \sigma Z)], \quad \forall x \in \mathbb{R}.$$

With a change of variable from the special case of  $\mu = 0$  and  $\sigma = 1$  [NP12, Proposition 3.2.2 and Remark 3.2.4], it is a standard fact that the unique bounded solution  $f_{\mu, \sigma, h}$  is given by the two equivalent expressions,

$$\begin{aligned} f_{\mu, \sigma, h}(x) &= \frac{1}{\sigma^2 \phi_{\mu, \sigma}(x)} \int_{-\infty}^x (h(z) - \mathbb{E}_Z[h(\mu + \sigma Z)]) \phi_{\mu, \sigma}(z) dz \\ &= -\frac{1}{\sigma^2 \phi_{\mu, \sigma}(x)} \int_x^{\infty} (h(z) - \mathbb{E}_Z[h(\mu + \sigma Z)]) \phi_{\mu, \sigma}(z) dz. \end{aligned} \tag{9}$$

Besides, the solution  $f_{\mu, \sigma, h}$  admits another equivalent representation through the Ornstein–Uhlenbeck (OU) semigroup  $P_t$ ,

$$(P_t h)(x) := \mathbb{E}_Z \left[ h \left( \mu + e^{-t}(x - \mu) + \sigma \sqrt{1 - e^{-2t}} Z \right) \right],$$

and since  $h$  is differentiable almost everywhere,

$$\begin{aligned} f_{\mu, \sigma, h}(x) &= -\frac{\partial}{\partial x} \int_0^{\infty} ((P_{\tau} h)(x) - \mathbb{E}_Z[h(\mu + \sigma Z)]) d\tau \\ &= -\int_0^{\infty} \frac{\partial}{\partial x} (P_{\tau} h)(x) d\tau \end{aligned}$$

$$= - \int_0^\infty e^{-\tau} \mathbb{E}_Z \left[ h' \left( \mu + e^{-\tau}(x - \mu) + \sigma \sqrt{1 - e^{-2\tau}} Z \right) \right] d\tau. \quad (10)$$

This is a change of variable from [NP12, Proposition 3.5.1].

**Stein factors** When restricted to the class of Lipschitz  $h$  functions, Stein factors refer to bounds on the essential supremum norm of  $f_{\mu,\sigma,h}$ ,  $f'_{\mu,\sigma,h}$  and  $f''_{\mu,\sigma,h}$  via the Lipschitz constant of  $h$ . Here we present the version from [Gau25] with a change of variable. Besides being useful for error control, it also certifies the integrability requirement of Lemma B.1 in our analysis.

**Lemma B.2** (Stein factors). *For all 1-Lipschitz function  $h$ ,*

$$\|f_{\mu,\sigma,h}\|_\infty \leq 1, \quad \|f'_{\mu,\sigma,h}\|_\infty \leq \sqrt{\frac{2}{\pi}} \sigma^{-1}, \quad \|f''_{\mu,\sigma,h}\|_\infty \leq 2\sigma^{-2}.$$

**Monotonicity** Finally, the following result states the monotonicity of  $f_{\mu,\sigma,h}$  when the function  $h$  is convex and Lipschitz. This is a straightforward observation from the semigroup representation Eq.(10) as well as the fact that  $f_{\mu,\sigma,h} \in C^1$ .

**Lemma B.3** (Monotonicity). *For all convex 1-Lipschitz function  $h$ ,  $f'_{\mu,\sigma,h}(x) \leq 0$  for all  $x \in \mathbb{R}$ .*

## C Application: Dominating MWU

This section further develops the argument from Section 4, showing that another special case of our algorithm improves upon the MWU baseline. In particular, it requires setting the function  $h$  in Algorithm 1 as the *log-sum-exp function*. The following  $\text{Regret}_T(u)$  upper bound is a corollary of Theorem 3, and by Corollary 16, it cannot be improved by  $\Omega(\log T)$  uniformly over  $u \in [-1, 1]$ .

**Corollary 7** (Regret: log-sum-exp). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = \frac{1}{\eta} \ln(\cosh(\eta x))$  for  $\eta = \frac{\alpha}{\sqrt{T}}$  and  $\alpha \in \mathbb{R}_{>0}$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{1}{2\alpha} \ln((1+u)^{1+u}(1-u)^{1-u}) + \mathbb{E}_Z [\alpha^{-1} \ln(\cosh(\alpha Z))] \right]}_{=: \gamma_{\text{LSE}}(u, \alpha)} \sqrt{T} + O(\log T), \quad \forall u \in [-1, 1].$$

*In particular, the RHS follows the convention  $0^0 = 1$  when  $u = \pm 1$ .*

While the MWU baseline has been mostly studied in the so-called “expert problem”, its special case with two experts can be equivalently adapted to our setting. This is a well-known fact, and we provide the details in Appendix C.1 for concreteness. With the constant learning rate  $\eta = \frac{\alpha}{\sqrt{T}}$ , the obtained OLO algorithm guarantees [Ora25, Section 7.5]

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{1}{2\alpha} \ln((1+u)^{1+u}(1-u)^{1-u}) + \frac{1}{2}\alpha \right]}_{=: \gamma_{\text{MWU}}(u, \alpha)} \sqrt{T}. \quad (11)$$

Relaxing this bound gives  $\text{Regret}_T^{\text{unif}} \leq [(\ln 2)\frac{1}{\alpha} + \frac{1}{2}\alpha]\sqrt{T}$  which is tight for the MWU baseline, as [GPS17] showed that MWU must suffer  $\text{Regret}_T^{\text{unif}} \geq \sqrt{2(\ln 2)T}$  even when non-increasing (as opposed to constant) learning rates are allowed.

The insight from Section 4 carries over to the comparison between Corollary 7 and Eq.(11). Specifically, we define the  $u$ -independent  $\text{Gap}_{\text{MWU}}(\alpha) := \gamma_{\text{MWU}}(u, \alpha) - \gamma_{\text{LSE}}(u, \alpha)$  as the difference between the two prefactors of  $\sqrt{T}$ . Lemma G.2 shows that  $\text{Gap}_{\text{MWU}}(\alpha) > 0$  for all  $\alpha > 0$ , and  $\lim_{\alpha \rightarrow \infty} \gamma_{\text{LSE}}(u, \alpha) = \sqrt{\frac{2}{\pi}}$  while  $\gamma_{\text{MWU}}(u, \alpha)$  diverges. Analogous to Figure 1 in the main paper, results from this regime are visualized in Figure 2. Combining it with the lower bound of [GPS17] gives a quantitative separation between our algorithm and MWU.

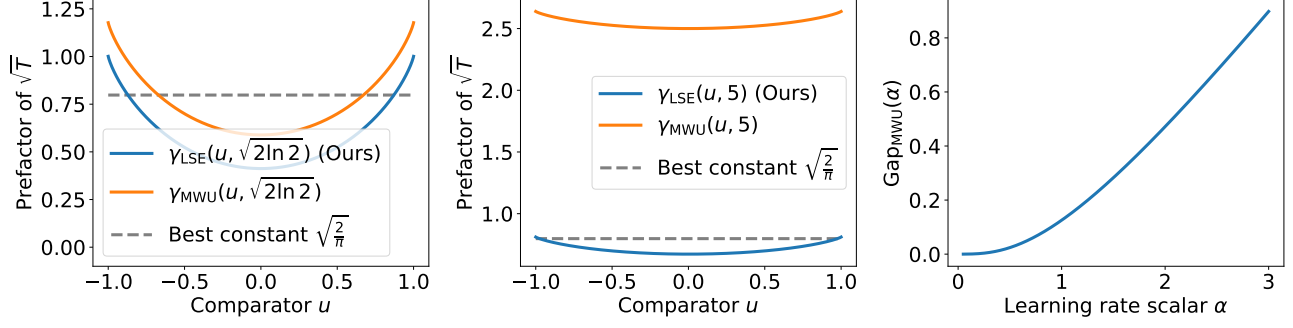


Figure 2: Comparison of the prefactors of  $\sqrt{T}$  in the regret bounds of our algorithm (represented by  $\gamma_{\text{LSE}}(u, \alpha)$ ), OGD (represented by  $\gamma_{\text{MWU}}(u, \alpha)$ ), and Cover’s algorithm (the optimal  $u$ -independent prefactor  $\sqrt{\frac{2}{\pi}}$ ). Analogous to Figure 1, but based on the log-sum-exp regime (Corollary 7). Left: with  $\alpha = \sqrt{2\ln 2}$  which minimizes  $\sup_{u \in [-1, 1]} \gamma_{\text{MWU}}(1, \alpha)$ ,  $\gamma_{\text{LSE}}(u, \alpha)$  and  $\gamma_{\text{MWU}}(u, \alpha)$  are compared as functions of  $u$ , lower is better. Middle: with  $\alpha = 5$ , the gap between  $\gamma_{\text{MWU}}(u, \alpha)$  and  $\gamma_{\text{LSE}}(u, \alpha)$  widens. Right: the margin of improvement  $\text{Gap}_{\text{MWU}}(\alpha) := \gamma_{\text{MWU}}(u, \alpha) - \gamma_{\text{LSE}}(u, \alpha)$  as a function of  $\alpha$ .

### C.1 Details of the Baseline

**LEA and the MWU algorithm** The MWU baseline was originally designed to solve the problem of *distributional LEA*, with the following definition. Consider a repeated game lasting  $T$  rounds. In each round,

- The LEA algorithm picks the decision  $w_t \in \Delta_d$ , where  $\Delta_d = \{w \in \mathbb{R}^d : \|w\|_1 = 1, w_i \geq 0, \forall i \in [1 : d]\}$  represents the probability simplex embedded in  $\mathbb{R}^d$ .
- The adversary picks the feedback  $l_t \in \mathbb{R}^d$  satisfying  $\|l_t\|_\infty \leq 1$ .
- The LEA algorithm suffers the loss  $\langle l_t, w_t \rangle$ .

At the end of the game, the performance of the algorithm is measured by the *LEA regret*: for any comparator  $\tilde{u} \in \Delta_d$ ,

$$\text{Regret}_T^{\text{LEA}}(\tilde{u}) := \sum_{t=1}^T \langle l_t, w_t - \tilde{u} \rangle.$$

Our work is only relevant to the case of  $d = 2$ . In this case, with the learning rate  $\eta = \frac{\alpha}{\sqrt{T}}$  where  $\alpha \in \mathbb{R}_{>0}$ , the MWU algorithm from [Ora25, Section 7.5] outputs

$$x_{t,j} = \frac{\exp\left(-\eta \sum_{\tau=1}^{t-1} l_{\tau,j}\right)}{\sum_{i=1}^d \exp\left(-\eta \sum_{\tau=1}^{\tau-1} l_{k,i}\right)}, \quad \forall j \in [1 : 2],$$

and guarantees the LEA regret bound (typically stated using the KL divergence)

$$\text{Regret}_T^{\text{LEA}}(\tilde{u}) \leq \left[ \frac{1}{\alpha} (\tilde{u}_1 \ln \tilde{u}_1 + \tilde{u}_2 \ln \tilde{u}_2 + \ln 2) + \frac{1}{2} \alpha \right] \sqrt{T}, \quad \forall \tilde{u} \in \Delta_2. \quad (12)$$

**Converting LEA to 1D OLO** With the above MWU algorithm denoted as  $\mathcal{A}$ , we use it to define the following OLO algorithm for the setting we consider. In each round,

- Query  $\mathcal{A}$  for its output  $w_t \in \Delta_2$ , and then output  $x_t = w_{t,1} - w_{t,2} \in [-1, 1]$  in one-dimensional OLO.
- After receiving the loss gradient  $g_t \in \mathbb{R}$  in OLO, define  $l_t = [g_t, -g_t]$  and send it to  $\mathcal{A}$  as its  $t$ -th feedback in LEA. As long as  $|g_t| \leq 1$ , it ensures  $\|l_t\|_\infty \leq 1$ .

Furthermore, for any OLO comparator  $u \in [-1, 1]$ , define an LEA comparator  $\tilde{u}$  by  $\tilde{u}_1 = \frac{1}{2}(1 + u)$  and  $\tilde{u}_2 = \frac{1}{2}(1 - u)$ . These procedures ensure that for all  $t$  we have  $g_t x_t = \langle l_t, w_t \rangle$ , and the total loss of the comparators satisfies  $\sum_{t=1}^T g_t u = \sum_{t=1}^T \langle l_t, \tilde{u} \rangle$ . The point is that any achievable upper bound on  $\text{Regret}_T^{\text{LEA}}(\tilde{u})$  can also be realized in the OLO problem we consider, and plugging in Eq.(12) gives Eq.(11).

## D Application: Trading off Total Loss and Uniform Regret

This section presents the second of the three applications summarized in Section 1.2: Algorithm 1 can achieve a continuum of optimal two-point tradeoffs between the total loss and the uniform regret.

The motivation is the following. Our main result (Theorem 3) shows that given any convex and 1-Lipschitz  $h$  function, Algorithm 1 guarantees additively sharp regret and total loss upper bounds, each realizing a performance tradeoff specified by  $h$ . While different  $h$  functions are generally incomparable, a natural follow-up question is whether one could characterize the optimal  $h$  given a reasonable preference over all possible tradeoffs. The two-point tradeoff we consider corresponds to such a preference.

**Setting** Starting from the necessary notations, let  $\text{alg}$  represent the OLO algorithm and  $\text{adv}$  represent the adversary. Throughout this section the adversary  $\text{adv}$  is assumed to satisfy  $g_t \in [-1, 1]$  for all  $t$ . Fixing the algorithm and the adversary, both the total loss  $\text{Loss}_T$  and the uniform regret  $\text{Regret}_T^{\text{unif}}$  are determined, therefore we overload such notations by  $\text{Loss}_T(\text{alg}, \text{adv})$  and  $\text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv})$  to clarify their dependencies. Similarly, the notation  $\text{Regret}_T(u)$  is overloaded by  $\text{Regret}_T(u, \text{alg}, \text{adv})$ .

Given any constant  $\varepsilon \in (0, \sqrt{\frac{\pi}{2}}]$  (independent of  $T$ ), we are interested in characterizing

$$\gamma(\varepsilon) := \limsup_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \inf_{\text{alg}} \left\{ \sup_{\text{adv}} \text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv}); \sup_{\text{adv}} \text{Loss}_T(\text{alg}, \text{adv}) \leq \varepsilon \sqrt{T} \right\}. \quad (13)$$

Here we emphasize that the supremum over  $\text{adv}$  needs to respect the boundedness assumption  $g_t \in [-1, 1]$ . The interpretation is that up to an additive  $o(\sqrt{T})$  error, the optimal uniform regret bound subject to the constraint  $\text{Loss}_T \leq \varepsilon \sqrt{T}$  is  $\text{Regret}_T^{\text{unif}} \leq \gamma(\varepsilon) \sqrt{T}$ . Furthermore, we aim to design a computationally efficient algorithm achieving these performance guarantees.

**Background** Two-point tradeoffs of this flavor are ubiquitous in statistics and machine learning, such as the tradeoff between Type-I and Type-II errors in hypothesis testing. The pioneering works of [EDKMW08, KP11] initiated this study in OLO, although the emphasis is on the big-O optimality rather than the leading constant optimality in a different quantitative regime ( $\varepsilon \propto T^{-1/2}$  as opposed to constant  $\varepsilon$ ; also see Appendix A). Interestingly, the algorithm of [KP11] is based on a ‘‘Stein-like’’ ODE, while it appears that the authors might not be aware of this connection at the time therefore the analysis was by brute force and suboptimal. Without considering computational efficiency, [Koo13] characterized a similar optimality frontier in the two-expert LEA problem, by passing a Cover-style achievability condition to the  $T \rightarrow \infty$  limit. A general strategy for trading off  $\text{Loss}_T$  with  $\text{Regret}_T^{\text{unif}}$  is via the combination of an unconstrained parameter-free OLO algorithm on the domain  $\mathbb{R}$  and a projection technique of [CO18]. This gives the state-of-the-art baseline to be introduced shortly.

**Results** Our first result is a lemma connecting the simultaneous achievability of upper bounds on  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$  to the achievability of a particular  $u$ -dependent upper bound on  $\text{Regret}_T(u)$ , possibly of independent interest. It holds with and without the assumption of  $g_t \in [-1, 1]$ .

**Lemma D.1** (Optimal two-point tradeoff via  $u$ -dependent regret bound). *For any  $a, b \in \mathbb{R}$  and any OLO algorithm  $\text{alg}$ , the following two conditions are equivalent:*

1.  $\sup_{\text{adv}} \text{Loss}_T(\text{alg}, \text{adv}) \leq a$  and  $\sup_{\text{adv}} \text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv}) \leq b$ ;
2.  $\sup_{\text{adv}, u \in [-1, 1]} \text{Regret}_T(u, \text{alg}, \text{adv}) \leq a + (b - a)|u|$ .

By Lemma D.1, upper-bounding  $\gamma(\varepsilon)$  from Eq.(13) can be reduced to finding the lowest  $\bar{\gamma}(\varepsilon)$  such that the regret bound  $\text{Regret}_T(u) \leq [\varepsilon + (\bar{\gamma}(\varepsilon) - \varepsilon)|u|]\sqrt{T}$  can be guaranteed up to an  $o(\sqrt{T})$  additive error. This fits into the scope of our general results: in particular, Corollary 15 (the regret version of our main theorem) shows

that it suffices to pick the  $h$  function in Algorithm 1 as the convex conjugate of the targeted  $u$ -dependent regret bound (viewed as a function of  $u$ ). It can be verified that for any  $b \in \mathbb{R}_{>0}$ , the convex conjugate of the function

$$f(x) = \begin{cases} a + b|x|, & x \in [-1, 1], \\ \infty, & \text{else} \end{cases}$$

is the (vertically shifted) *soft-thresholded absolute value*  $f^*(y) = -a + \max\{|y| - b, 0\}$ . Therefore we are motivated to instantiate Corollary 15 with the soft-thresholded absolute value function, resulting in the following constructive regret upper bound parameterized by a learning rate scalar  $\alpha$ .

**Corollary 8** (Regret: soft-thresholded). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = \max\{|x| - \eta^{-1}, 0\}$  for  $\eta = \frac{\alpha}{\sqrt{T}}$  and  $\alpha \in \mathbb{R}_{>0}$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{|u|}{\alpha} + \frac{2}{\alpha} \Phi\left(\frac{1}{\alpha}\right) + 2\phi\left(\frac{1}{\alpha}\right) - \frac{2}{\alpha} \right]}_{=: \gamma_{\text{STh}}(u, \alpha)} \sqrt{T} + O(\log T), \quad \forall u \in [-1, 1].$$

The obtained Corollary 8 is analogous to Corollary 6 (the case dominating OGD) and Corollary 7 (the case dominating MWU). In particular, they all recover the sharp uniform regret upper bound (Corollary 5, as well as the results of [KKW20, GHP22]) when the learning rate scalar  $\alpha$  goes to infinity. Corollary 8 is also additively sharp due to the general regret lower bound, Corollary 16.

By reparameterizing the bound using  $\varepsilon$  and taking the  $T \rightarrow \infty$  limit, we obtain the following exact characterization of  $\gamma(\varepsilon)$  from Eq.(13), realized by a computationally efficient algorithm.

**Theorem 9** (Characterization of  $\gamma(\varepsilon)$ ). *For all constant  $\varepsilon \in (0, \sqrt{\frac{2}{\pi}}]$ ,  $\gamma(\varepsilon)$  from Eq.(13) is the unique solution of the equation*

$$\int_{-\infty}^{\varepsilon - \gamma(\varepsilon)} \Phi(x) dx = \frac{\varepsilon}{2}.$$

Furthermore, for all  $T \in \mathbb{N}_+$ , there exists a special case of Algorithm 1 that simultaneously guarantees  $\text{Loss}_T \leq \varepsilon\sqrt{T} + O(\log T)$  and  $\text{Regret}_T^{\text{unif}} \leq \gamma(\varepsilon)\sqrt{T} + O(\log T)$ .

In retrospect, the crucial step in the above analysis is using Lemma D.1 to convert the two-point tradeoff between  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$  to a particular “continuous” tradeoff on  $\text{Regret}_T(u)$ , such that our general regret upper bound (Corollary 15) and lower bound (Corollary 16) can be applied. Alternatively, one may view this conversion from the dual perspective (i.e., total loss upper bounds), which gives another derivation of the right  $h$  function to consider by solving an infinite-dimensional linear program. We provide a sketch of this argument in Appendix D.1.

**Comparison to baseline** Prior to this work, the state-of-the-art algorithm trading off  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$  is a combination of [ZCP22] and [CO18], concretely introduced in Appendix D.1. With  $\text{erfi} : \mathbb{R} \rightarrow \mathbb{R}$  being the *imaginary error function* (i.e.,  $\text{erfi}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(z^2) dz$ ), such an algorithm runs in  $O(1)$  time per round while simultaneously guaranteeing  $\text{Loss}_T \leq \varepsilon\sqrt{T}$  and

$$\text{Regret}_T^{\text{unif}} \leq \varepsilon\sqrt{T} + \sqrt{2T} \cdot \text{erfi}^{-1}\left(\frac{\sqrt{2}}{\sqrt{\pi}\varepsilon}\right) \quad (14)$$

against all adversaries satisfying  $g_t \in [-1, 1]$  for all  $t$ .

Comparing our Theorem 9 to this baseline, we see that up to additive logarithmic terms, both results guarantee the same  $\text{Loss}_T$  upper bounds, but their  $\text{Regret}_T^{\text{unif}}$  bounds differ in the prefactors of  $\sqrt{T}$ . Demonstrating the advantage of Theorem 9 amounts to showing

$$\gamma(\varepsilon) < \varepsilon + \sqrt{2} \cdot \text{erfi}^{-1}\left(\frac{\sqrt{2}}{\sqrt{\pi}\varepsilon}\right) \quad (15)$$

for all  $\varepsilon \in (0, \sqrt{\frac{2}{\pi}}]$ , which is proved in Lemma G.4. In particular, when  $\varepsilon = \sqrt{\frac{2}{\pi}}$ , we have LHS =  $\sqrt{\frac{2}{\pi}}$  while RHS =  $\sqrt{\frac{2}{\pi}} + \sqrt{2} \cdot \text{erfi}^{-1}(1) > \text{LHS}$ . Also see Figure 3 for visualizations.

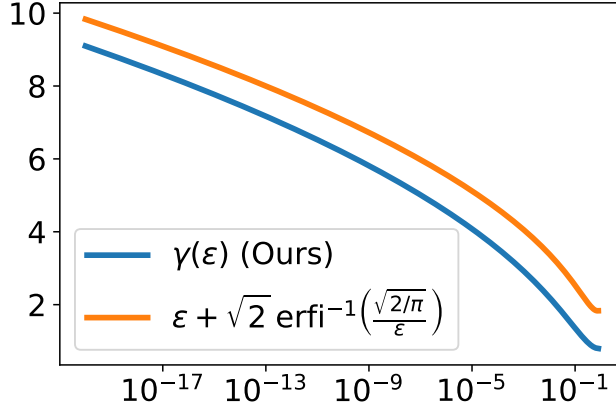


Figure 3: Comparison of the prefactors of  $\sqrt{T}$  in the  $\text{Regret}_T^{\text{unif}}$  upper bound of our algorithm (LHS of Eq.(15), represented by blue) and the baseline [CO18, ZCP22] (RHS of Eq.(15), represented by orange); lower is better. Both are functions of  $\varepsilon \in (0, \sqrt{\frac{2}{\pi}}]$  which represents the budget on  $\text{Loss}_T$ . Our result dominates that of the baseline.

## D.1 Details of the Baseline

The baseline we consider starts from a constrained-to-unconstrained reduction first proposed by [CO18] and systematically summarized in [Ora25, Section 9]. To date it has become a key element in the toolbox of online learning. The version for our setting is the following [Ora25, Remark 9.6].

**Lemma D.2** (Reduction to unconstrained OLO). *Let  $\mathcal{A}$  be an arbitrary algorithm for one-dimensional unconstrained OLO (i.e.,  $x_t \in \mathbb{R}$  as opposed to  $x_t \in [-1, 1]$ ) achieving  $\text{Regret}_T(u) \leq \psi_T(u)$  for all  $u \in \mathbb{R}$ , against all adversaries satisfying  $g_t \in [-1, 1]$  for all  $t$ . Then, there exists an algorithm (which is a post-processed version of  $\mathcal{A}$ ) satisfying  $x_t \in [-1, 1]$  for all  $t$  while achieving*

$$\text{Regret}_T(u) \leq \psi_T(u), \quad \forall u \in [-1, 1],$$

against the same adversary class.

Motivated by this lemma, the main objective within the field of parameter-free online learning is to achieve sharp comparator-dependent regret bounds on unconstrained domains. For our setting the state of the art is [ZCP22], building on the techniques of [HLPR23]. Given a potential function defined as

$$\varphi(t, s) = \varepsilon \sqrt{t} \left( \sqrt{\pi} \int_0^{\frac{s}{\sqrt{2t}}} \text{erfi}(z) \, dz - 1 \right),$$

the unconstrained algorithm outputs a similar discrete derivative as Cover’s algorithm Eq.(3),

$$x_t = -\frac{1}{2} \left[ \varphi \left( t, \sum_{i=1}^{t-1} g_i + 1 \right) - \varphi \left( t, \sum_{i=1}^{t-1} g_i - 1 \right) \right] \in \mathbb{R}, \quad \forall t \in [1 : T].$$

It guarantees [ZCP22, intermediate result within Theorem 4]

$$\text{Regret}_T(u) \leq \varepsilon \sqrt{T} + |u| \sqrt{2T} \cdot \text{erfi}^{-1} \left( \frac{\sqrt{2}|u|}{\sqrt{\pi}\varepsilon} \right), \quad \forall u \in \mathbb{R}. \quad (16)$$

Setting  $|u| = 1$  and combining it with Lemma D.2 results in the Eq.(14) referenced in our work.

A quantitative strength of this baseline is that Eq.(16) has the optimal constant multiplier  $\sqrt{2}$  on the leading order factor  $|u| \sqrt{T \log |u|}$ , as  $|u|$  and  $T$  both tend to infinity. Broadly speaking, this is related to the “additive sharpness” targeted by our work, while we study optimality in a more detailed scaling regime where  $u$  is fixed (on a bounded domain) and only  $T \rightarrow \infty$ .



## D.2 Dual Argument of Lemma D.1 via Linear Programming

In this subsection, we sketch the dual argument of Lemma D.1, converting the two-point tradeoff between  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$  to a particular upper bound function on  $\text{Loss}_T$  to aim for. Technically, this relies on solving an infinite-dimensional linear program which could be of independent interest. Since our formal results are already derived from Lemma D.1, the following sketch will omit many technical subtleties, prioritizing simplicity over mathematical rigor.

Recall from Theorem 3 that for any convex and 1-Lipschitz function  $\psi_T^* : \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $\mathbb{E}_{X \sim \mathcal{N}(0,T)}[\psi_T^*(X)] = 0$ , there exists a special case of Algorithm 1 guaranteeing

$$\text{Loss}_T \leq -\psi_T^* \left( -\sum_{t=1}^T g_t \right) + O(\log T)$$

against all adversaries satisfying  $g_t \in [-1, 1]$  for all  $t$ . This is additively sharp due to Theorem 4. We are now interested in finding the optimal tradeoff between  $\text{Loss}_T$  and  $\text{Regret}_T^{\text{unif}}$  induced by such  $\psi_T^*$  functions.

For the simplicity of this sketch, we restrict the search space to  $\psi_T^*$  functions that are also even. Given this restriction, notice that the above  $\text{Loss}_T$  upper bound leads to the adversary-independent upper bound

$$\text{Loss}_T \leq \sup_{s \in \mathbb{R}} [-\psi_T^*(s)] + O(\log T) = -\psi_T^*(0) + O(\log T),$$

as well as the uniform regret bound (via the loss-regret duality, Lemma G.5)

$$\begin{aligned} \text{Regret}_T^{\text{unif}} &\leq \max_{u \in \{-1, 1\}} (\psi_T^*)^*(u) + O(\log T) \\ &= \sup_{s \in \mathbb{R}} [s - \psi_T^*(s)] + O(\log T) && (\psi_T^* \text{ is even}) \\ &= \lim_{s \rightarrow \infty} [s - \psi_T^*(s)] + O(\log T). && (\psi_T^* \text{ is 1-Lipschitz}) \end{aligned}$$

Omitting the additive  $O(\log T)$  terms from the RHS, the task of minimizing  $\text{Regret}_T^{\text{unif}}$  under the constraint  $\text{Loss}_T \leq \varepsilon\sqrt{T}$  can be modeled as the following optimization problem,

$$\text{Problem One} \leftarrow \begin{cases} \min_{f: \mathbb{R} \rightarrow \mathbb{R}} & \lim_{x \rightarrow \infty} [x - f(x)] \\ \text{s.t.} & \mathbb{E}_{X \sim \mathcal{N}(0,T)}[f(X)] = 0, \\ & -f(0) = \varepsilon\sqrt{T}, \\ & f \text{ is even, convex and 1-Lipschitz.} \end{cases}$$

The idea is that the optimal value of Problem One normalized by  $\frac{1}{\sqrt{T}}$  gives  $\gamma(\varepsilon)$ . While additional effort is required to formally establish their equivalence, here we do not treat such subtleties. The minimizing argument  $f$  is the total loss upper bound function to aim for, thus also the  $h$  function to apply in Algorithm 1. The subsequent sketch is divided into four steps.

**Step 1: reformulation using derivative** To simplify Problem One, notice that for all feasible function  $f$ , the derivative  $f'$  exists almost everywhere. Since

$$f(x) = f(0) + \int_0^x f'(s) \, ds, \quad \forall x \in \mathbb{R},$$

we have

$$\lim_{x \rightarrow \infty} [x - f(x)] = -f(0) + \int_0^\infty [1 - f'(x)] \, dx,$$

$$\mathbb{E}_{X \sim \mathcal{N}(0,T)}[f(X)] = 2 \int_0^\infty f(\sqrt{T}z) \phi(z) \, dz$$

$$\begin{aligned}
&= 2 \int_0^\infty \left( f(0) + \int_0^{\sqrt{T}z} f'(s) ds \right) \phi(z) dz \\
&= f(0) + 2 \int_0^\infty \left( \int_{\frac{s}{\sqrt{T}}}^\infty \phi(z) dz \right) f'(s) ds \quad (\text{swap integrals}) \\
&= f(0) + 2 \int_0^\infty \left[ 1 - \Phi \left( \frac{s}{\sqrt{T}} \right) \right] f'(s) ds.
\end{aligned}$$

Combining these properties, we can change the argument of Problem One from  $f$  to its derivative  $f'$ , thereby converting the problem to the following infinite dimensional linear program that shares the same optimal value.

$$\text{Problem Two} \leftarrow \begin{cases} \min_{g: \mathbb{R}_{\geq 0} \rightarrow [0,1]} & \varepsilon\sqrt{T} + \int_0^\infty [1 - g(x)] dx \\ \text{s.t.} & -\varepsilon\sqrt{T} + 2 \int_0^\infty \left[ 1 - \Phi \left( \frac{x}{\sqrt{T}} \right) \right] g(x) dx = 0. \end{cases}$$

Notably, compared to Problem One, Problem Two drops the monotonicity requirement on the function  $g$ . This is a tight relaxation which can be verified.

**Step 2: the dual problem** The second step is to apply the standard duality analysis from convex optimization. With the multiplier  $\lambda \in \mathbb{R}$ , the Lagrangian of Problem Two is

$$\begin{aligned}
L(g, \lambda) &:= \varepsilon\sqrt{T} + \int_0^\infty [1 - g(x)] dx + \lambda \left( 2 \int_0^\infty \left[ 1 - \Phi \left( \frac{x}{\sqrt{T}} \right) \right] g(x) dx - \varepsilon\sqrt{T} \right) \\
&= (1 - \lambda)\varepsilon\sqrt{T} + \int_0^\infty \left[ 1 + \left( 2\lambda - 2\lambda\Phi \left( \frac{x}{\sqrt{T}} \right) - 1 \right) g(x) \right] dx.
\end{aligned}$$

Minimizing it with respect to  $g : \mathbb{R} \rightarrow [0, 1]$  is equivalent to assigning  $g(x)$  pointwise to 0 or 1, giving the Lagrange dual function

$$q(\lambda) := \min_{g: \mathbb{R}_{\geq 0} \rightarrow [0,1]} L(g, \lambda) = (1 - \lambda)\varepsilon\sqrt{T} + \int_0^\infty \min \left\{ 1, 2\lambda \left[ 1 - \Phi \left( \frac{x}{\sqrt{T}} \right) \right] \right\} dx.$$

Accordingly, the dual of Problem Two is maximizing  $q(\lambda)$  over  $\lambda \in \mathbb{R}$ .

**Step 3: solving the dual problem** Regarding this dual problem, it can be verified that the maximizing argument satisfies  $\lambda \geq 1$ . For any such  $\lambda$ , define the notation

$$x_\lambda := \sqrt{T} \cdot \Phi^{-1} \left( 1 - \frac{1}{2\lambda} \right) \geq 0,$$

such that

$$q(\lambda) = (1 - \lambda)\varepsilon\sqrt{T} + x_\lambda + 2\lambda \int_{x_\lambda}^\infty \left[ 1 - \Phi \left( \frac{x}{\sqrt{T}} \right) \right] dx.$$

Differentiating it gives

$$\begin{aligned}
q'(\lambda) &= -\varepsilon\sqrt{T} + \frac{\partial x_\lambda}{\partial \lambda} + 2 \int_{x_\lambda}^\infty \left[ 1 - \Phi \left( \frac{x}{\sqrt{T}} \right) \right] dx - 2\lambda \left[ 1 - \Phi \left( \frac{x_\lambda}{\sqrt{T}} \right) \right] \frac{\partial x_\lambda}{\partial \lambda} \\
&= -\varepsilon\sqrt{T} + 2 \int_{x_\lambda}^\infty \left[ 1 - \Phi \left( \frac{x}{\sqrt{T}} \right) \right] dx.
\end{aligned}$$

The maximizing argument  $\lambda^*$  satisfies  $q(\lambda^*) = 0$ , which is equivalent to the solution of

$$\int_{-\infty}^{-\Phi^{-1}(1 - \frac{1}{2\lambda^*})} \Phi(x) dx = \frac{\varepsilon}{2}.$$

**Step 4: final result** Summarizing the above and observing that strong duality holds, we have

$$\gamma(\varepsilon) = \frac{q(\lambda^*)}{\sqrt{T}} = \varepsilon + \Phi^{-1} \left( 1 - \frac{1}{2\lambda^*} \right),$$

therefore  $\int_{-\infty}^{\varepsilon - \gamma(\varepsilon)} \Phi(x) dx = \frac{\varepsilon}{2}$ , recovering the result from Theorem 9. Furthermore, the minimizing argument  $f$  of Problem One is given by the soft-thresholded absolute value,

$$f(x) = -\varepsilon\sqrt{T} + \max \left\{ |x| - \sqrt{T} \cdot \Phi^{-1} \left( 1 - \frac{1}{2\lambda^*} \right), 0 \right\}, \quad \forall x \in \mathbb{R}.$$

This is the  $h$  function used to instantiate our algorithm in Corollary 8.

## E Extension: OLO with Noisy Feedback

This section extends the algorithmic results so far to the setting with stochastic adversaries, also known as OLO with noisy feedback. By doing this, we further develop the parallel relation between performance guarantees in OLO and Wasserstein martingale CLTs [Röl18]. Motivations are two-fold.

- The intuition underlying the proof of our main theorem (Theorem 3) is to view the adversary’s cumulative decision  $\sum_{t=1}^T g_t$  as a stochastic process indexed by  $T$ . After a suitable notion of bias is properly taken care of, this can be essentially analyzed as a martingale whose normal approximation properties follow from Stein’s method. While so far we have only considered deterministic  $\sum_{t=1}^T g_t$ , it is natural to expect that the same idea still works when  $\sum_{t=1}^T g_t$  is by default a stochastic process. This section performs this extension.
- In OLO, achievable performance guarantees are governed by the complexity of the targeted adversary class. While typically the adversaries are constrained in a deterministic manner (e.g., by assuming  $g_t \in [-1, 1]$ ), an alternative setting is to impose soft, distributional constraints such that  $|g_t|$  is allowed to be large (albeit with low probability). There have been plenty of works on this extension [JO19, vdH19, ZC22, BK25, Liu25], but to our knowledge, achieving better-than-big-O optimality remains open.

After introducing the setting, we present the extensions of our main results in Appendix E.1. The relation to Röllin’s martingale CLT is discussed in Appendix E.2.

**Setting** With a time horizon  $T \in \mathbb{N}_+$ , let  $\Omega = \mathbb{R}^T$  and  $\mathcal{F} = \mathcal{B}(\mathbb{R}^T)$  represent the sequence space and its Borel  $\sigma$ -algebra. Let  $G_t(\omega) = \omega_t, \omega \in \Omega, t \in [1 : T]$  be the canonical coordinate process (intuitively, the *loss process* in OLO), and let  $\mathcal{F}_t = \sigma(G_1, \dots, G_t), t \in [0, T]$  be the canonical filtration. Consider the following interaction between an OLO algorithm and an adversary, analogous to the deterministic setting in Section 1.

- The algorithm is defined by a sequence of Borel functions  $(a_t)_{t \in [1:T]}$  where each  $a_t : \mathbb{R}^{t-1} \rightarrow [-1, 1]$  can depend on  $T$ . This defines a *decision process*  $(X_t)_{t \in [1:T]}$  where

$$X_t(\omega) = a_t(G_1(\omega), \dots, G_{t-1}(\omega)),$$

which is predictable with respect to  $(\mathcal{F}_t)_{t \in [0:T]}$ .

- The adversary is defined by a probability measure  $\mathbb{P}$  on  $(\Omega, \mathcal{F})$ , which can arbitrarily depend on  $T$  and the algorithm  $(a_t)_{t \in [1:T]}$ . This specifies the law of the loss process  $(G_t)_{t \in [1:T]}$ , and by fixing the algorithm, the law of the decision process  $(X_t)_{t \in [1:T]}$  as well.

On the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0:T]}, \mathbb{P})$ , the performance of the algorithm is measured by the total loss

$$\text{Loss}_T = \sum_{t=1}^T G_t X_t$$

which is a random variable, and the goal of the algorithm is to minimize its expectation in the same sense as in Section 1. We use the plain  $\mathbb{E}$  to denote the expectation with respect to  $\mathbb{P}$ . This is not to be confused with  $\mathbb{E}_Z$  in the rest of our analysis, which takes the expectation with respect to the external randomness in  $Z$ . Picking  $\mathbb{P}$  as a Dirac measure essentially recovers the deterministic setting.

For notational convenience, we also write  $x_t = X_t(\omega)$  and  $g_t = G_t(\omega)$  such that all the pathwise notations are consistent with the deterministic setting.

**OLO with noise** The above is equivalent to the classical problem of OLO with noisy feedback: after a standard deterministic adversary picks  $g_t$ , the algorithm observes  $\hat{g}_t = g_t + \sigma_t$  where the noise  $\sigma_t$  is conditionally zero mean, and the performance of the algorithm is evaluated by the expectation of its total loss  $\sum_{t=1}^T \hat{g}_t x_t$ . We note that our analysis does not place explicit assumptions on the tail of the noise  $\sigma_t$  (e.g., sub-Gaussianity) which are required by typical martingale concentration inequalities. Instead, assumptions will be imposed on the conditional moments of  $\sigma_t$ . This will be made clear in Corollary 11.

## E.1 Main Result

**Algorithm** Consider applying Algorithm 1 to every sample path of the above stochastic setting. To prepare for its population-level guarantees, we now reformulate the intermediate quantities of Algorithm 1 as stochastic processes.

- On each sample path, Algorithm 1 requires picking a parameter  $\rho_t$  based on  $g_1, \dots, g_{t-1}$ . On the population level, with  $\rho_t$  represented by some  $P_t(\omega)$ , this amounts to the algorithm picking a predictable process  $(P_t)_{t \in [0:T]}$  with respect to  $(\mathcal{F}_t)_{t \in [0:T]}$ , such that (i) for all  $\omega \in \Omega$  and  $t \in [1:T]$ ,  $P_t(\omega) \in [0, P_{t-1}(\omega)]$ ; and (ii) for all  $\omega \in \Omega$  and  $t < T$ ,  $P_t(\omega) > 0$ . We note that  $P_0$  is simply a positive real number.
- Accordingly, based on the pathwise quantity  $c_t = \rho_{t-1}^2 - \rho_t^2$ , we define the process  $(C_t)_{t \in [1:T]}$  by  $C_t = P_{t-1}^2 - P_t^2$ , which is also predictable with respect to  $(\mathcal{F}_t)_{t \in [0:T]}$ .
- Finally, define the process  $(S_t)_{t \in [1:T]}$  by  $S_t = \sum_{t=1}^T G_t$ .

Using these notations, applying Algorithm 1 to the stochastic setting is equivalent to the update rule

$$X_t = \mathbb{E}_Z [f_{S_{t-1}, P_{t-1}, h}(S_{t-1} + P_t Z)] . \quad (17)$$

**Total loss bound** A key strength of our main result (Theorem 3) is that the stated total loss bound holds against arbitrary adversaries, including those that are unconstrained beforehand. Therefore taking the expectation of Theorem 3 (more precisely, a partial result in its proof) leads to the following theorem.

**Theorem 10** (Expected total loss bound). *In the stochastic setting, against all adversaries, Algorithm 1 (i.e., Eq.(17)) given any convex and 1-Lipschitz function  $h$  guarantees*

$$\begin{aligned} \mathbb{E}[\text{Loss}_T] &\leq -\mathbb{E} \left[ \mathbb{E}_Z \left[ h \left( \sum_{t=1}^T G_t + P_T Z \right) \right] \right] + \mathbb{E}_Z[h(P_0 Z)] \\ &\quad + \mathbb{E} \sum_{t=1}^T \left( \sqrt{\frac{2}{\pi}} \frac{\max\{\mathbb{E}[G_t^2 | \mathcal{F}_{t-1}] - C_t, 0\}}{P_{t-1}} + \frac{2C_t |G_t| + |G_t^3|}{P_{t-1}^2} \right) . \end{aligned}$$

*Proof of Theorem 10.* We start from the following partial result in the proof of Theorem 3 which holds pathwise on the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0:T]}, \mathbb{P})$ ,

$$\begin{aligned} \mathbb{E}_Z[h(S_t + P_t Z)] - \mathbb{E}_Z[h(S_{t-1} + P_{t-1} Z)] \\ \leq -G_t X_t + (C_t - G_t^2) \mathbb{E}_Z [f'_{S_{t-1}, P_{t-1}, h}(S_{t-1} + P_t Z)] + \frac{2C_t |G_t| + |G_t^3|}{P_{t-1}^2} . \end{aligned}$$

Taking the expectation conditioned on  $\mathcal{F}_{t-1}$  before applying Lemma B.2 (Stein factors) and Lemma B.3 ( $f'_{s_{t-1}, \rho_{t-1}, h}$  is non-positive) gives

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{E}_Z[h(S_t + P_t Z)] - \mathbb{E}_Z[h(S_{t-1} + P_{t-1} Z)] \middle| \mathcal{F}_{t-1} \right] \\ & \leq -\mathbb{E}[G_t | \mathcal{F}_{t-1}] X_t + \sqrt{\frac{2}{\pi}} \frac{\max\{\mathbb{E}[G_t^2 | \mathcal{F}_{t-1}] - C_t, 0\}}{P_{t-1}} + \frac{2C_t \mathbb{E}[|G_t| | \mathcal{F}_{t-1}] + \mathbb{E}[|G_t^3| | \mathcal{F}_{t-1}]}{P_{t-1}^2}. \end{aligned}$$

Plugging it back into the telescopic sum Eq.(7) from the proof of Theorem 3 and taking the expectation completes the proof.  $\square$

Despite the generality of Theorem 10, the most interpretable case is when the second moment of  $G_t$  conditioned on  $\mathcal{F}_{t-1}$  is bounded by a known constant, e.g.,  $\mathbb{E}[G_t^2 | \mathcal{F}_{t-1}] \leq 1$ . In this case we obtain the following simplification by setting  $P_t$  and  $C_t$  deterministically.

**Corollary 11** (Bounded second moment). *In the stochastic setting, if the adversary satisfies the moment condition  $\mathbb{E}[G_t^2 | \mathcal{F}_{t-1}] \leq 1$  for all  $t \in [1 : T]$ , then by setting  $P_t = \sqrt{T-t}$  deterministically, Algorithm 1 (i.e., Eq.(17)) given any convex and 1-Lipschitz function  $h$  guarantees*

$$\mathbb{E}[\text{Loss}_T] \leq -\mathbb{E} \left[ h \left( \sum_{t=1}^T G_t \right) \right] + \mathbb{E}_Z[h(\sqrt{T}Z)] + \sum_{t=1}^T \frac{2\mathbb{E}[|G_t|] + \mathbb{E}[|G_t^3|]}{T-t+1}.$$

The value of this total loss bound is its dependence on the first three moments of  $G_t$  as opposed to more specific tail conditions. If  $G_t$  for all  $t \in [1 : T]$  is supported on the interval  $[-1, 1]$ , then the sum on the RHS is  $O(\log T)$ , meaning that Corollary 11 essentially recovers its deterministic counterpart (Theorem 3) against  $g_t \in [-1, 1]$  adversaries which is additively sharp. We also note that just like Theorem 3, Corollary 11 holds for all convex and 1-Lipschitz  $h$  functions, therefore results from other sections of this work can be applied in an orthogonal manner.

**Without convexity** Next we provide a variant of Theorem 10 to be used shortly. Notice that when proving Theorem 10, the convexity of the  $h$  function is only used through Lemma B.3. Removing this step gives the following analogous result without convexity, whose proof is nearly identical to that of Theorem 10. Implications will be further discussed in Appendix F.

**Theorem 12** (Variant of Theorem 10 without convexity). *In the stochastic setting, against all adversaries, Algorithm 1 (i.e., Eq.(17)) given any 1-Lipschitz function  $h$  guarantees*

$$\begin{aligned} \mathbb{E}[\text{Loss}_T] & \leq -\mathbb{E} \left[ \mathbb{E}_Z \left[ h \left( \sum_{t=1}^T G_t + P_T Z \right) \right] \right] + \mathbb{E}_Z[h(P_0 Z)] \\ & \quad + \mathbb{E} \sum_{t=1}^T \left( \sqrt{\frac{2}{\pi}} \frac{|\mathbb{E}[G_t^2 | \mathcal{F}_{t-1}] - C_t|}{P_{t-1}} + \frac{2C_t |G_t| + |G_t^3|}{P_{t-1}^2} \right). \end{aligned}$$

## E.2 From Total Loss Bound to Martingale CLT

The construction of our OLO algorithm is based on the remarkably clean proof of a quantitative Wasserstein martingale CLT due to [Röll18]. We now further elucidate their connection from the opposite direction, more specifically showing that Röllin's martingale CLT can be proved via the above Theorem 12. In this way, we justify viewing the obtained total loss upper bounds as algorithmic analogues of the martingale CLT.

Concretely, we first restate the result of Röllin [Röll18, Theorem 2.1]. For two random variables  $X$  and  $Y$  taking values in  $\mathbb{R}$ , define their Wasserstein distance as

$$d_W(X, Y) := \sup_{h \in \text{Lip}_1(\mathbb{R})} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|,$$

where  $\text{Lip}_1(\mathbb{R})$  denotes the collection of all 1-Lipschitz functions from  $\mathbb{R}$  to  $\mathbb{R}$ .

**Theorem 13** (Röllin's martingale CLT). *For a martingale  $(L_t)_{t \in [1:T]}$  adapted to the filtration  $(\mathcal{F}_t)_{t \in [0:T]}$ , define the notations  $V_T = \sum_{t=1}^T \mathbb{E}[L_t^2 | \mathcal{F}_{t-1}]$  and  $v_T = \sum_{t=1}^T \mathbb{E}[L_t^2]$ . Assume that  $V_T = v_T$  almost surely. Then, for any  $a \geq 0$ ,*

$$d_W \left( \frac{1}{\sqrt{v_T}} \sum_{t=1}^T L_t, Z \right) \leq \frac{1}{\sqrt{v_T}} \left[ 3 \sum_{t=1}^T \mathbb{E} \frac{|L_t|^3}{v_T - \sum_{i=1}^{t-1} \mathbb{E}[L_i^2 | \mathcal{F}_{i-1}] + a^2} + 2a \right].$$

Next, we give a concise rephrasing of Röllin's original proof starting from our Theorem 12.

*Proof of Theorem 13.* Consider applying Theorem 12 with the loss process  $(G_t)_{t \in [1:T]}$  being the martingale  $(L_t)_{t \in [1:T]}$  from Theorem 13. The algorithm from Theorem 12 requires choosing the process  $(P_t)_{t \in [1:T]}$ , and here we consider equipping it with oracle knowledge: let  $C_t = \mathbb{E}[L_t^2 | \mathcal{F}_{t-1}]$ ,  $P_0 = \sqrt{\sum_{t=1}^T \mathbb{E}[L_t^2] + a^2}$ , and thus

$$\begin{aligned} P_t &= \sqrt{P_0^2 - \sum_{t=1}^t C_t} = \sqrt{v_T - \sum_{i=1}^t \mathbb{E}[L_i^2 | \mathcal{F}_{i-1}] + a^2}, \\ P_T &= \sqrt{P_0^2 - \sum_{t=1}^T C_t} = \sqrt{v_T - V_T + a^2} = a, \quad a.s. \end{aligned}$$

The last equality is due to the assumption that  $V_T = v_T$  almost surely. We note that such choices of  $(C_t)_{t \in [1:T]}$  and  $(P_t)_{t \in [1:T]}$  satisfy their respective measurability requirements. Then, Theorem 12 yields

$$\mathbb{E} \left[ \sum_{t=1}^T L_t X_t \right] \leq -\mathbb{E} \left[ \mathbb{E}_Z \left[ h \left( \sum_{t=1}^T L_t + aZ \right) \right] \right] + \mathbb{E}_Z [h(\sqrt{v_T + a^2} Z)] + \mathbb{E} \sum_{t=1}^T \frac{2\mathbb{E}[L_t^2 | \mathcal{F}_{t-1}] |L_t| + |L_t^3|}{v_T - \sum_{i=1}^{t-1} \mathbb{E}[L_i^2 | \mathcal{F}_{i-1}] + a^2},$$

where by the Lipschitzness of  $h$ ,

$$\begin{aligned} -\mathbb{E} \left[ \mathbb{E}_Z \left[ h \left( \sum_{t=1}^T L_t + aZ \right) \right] \right] &\leq -\mathbb{E} \left[ h \left( \sum_{t=1}^T L_t \right) \right] + \mathbb{E}_Z [|aZ|] \leq -\mathbb{E} \left[ h \left( \sum_{t=1}^T L_t \right) \right] + a, \\ \mathbb{E}_Z [h(\sqrt{v_T + a^2} Z)] &\leq \mathbb{E}_Z [h(\sqrt{v_T} Z)] + a. \end{aligned}$$

Furthermore,

$$\mathbb{E} \sum_{t=1}^T \frac{2\mathbb{E}[L_t^2 | \mathcal{F}_{t-1}] |L_t| + |L_t^3|}{v_T - \sum_{i=1}^{t-1} \mathbb{E}[L_i^2 | \mathcal{F}_{i-1}] + a^2} \leq \mathbb{E} \sum_{t=1}^T \frac{3|L_t^3|}{v_T - \sum_{i=1}^{t-1} \mathbb{E}[L_i^2 | \mathcal{F}_{i-1}] + a^2}.$$

Since  $(X_t)_{t \in [1:T]}$  is predictable with respect to the natural filtration  $(\mathcal{F}_t)_{t \in [1:T]}$  of  $(L_t)_{t \in [1:T]}$  and the latter is a martingale, we have  $\mathbb{E}[\sum_{t=1}^T L_t X_t] = 0$ . Therefore combining the above, for any 1-Lipschitz  $h$  function we have

$$\mathbb{E} \left[ h \left( \sum_{t=1}^T L_t \right) \right] - \mathbb{E}_Z [h(\sqrt{v_T} Z)] \leq 3\mathbb{E} \sum_{t=1}^T \frac{|L_t^3|}{v_T - \sum_{i=1}^{t-1} \mathbb{E}[L_i^2 | \mathcal{F}_{i-1}] + a^2} + 2a.$$

Notice that the inequality still holds with  $h \leftarrow -h$ . Finally scaling all  $L_t$  and  $a$  by  $\frac{1}{\sqrt{v_T}}$  completes the proof.  $\square$

We remark that the major technical elements of this proof are already presented in [Röl18], with some of them being classical techniques of Stein's method (Appendix B). Nonetheless, we find it conceptually very interesting to rephrase the proof through performance guarantees in OLO, indicating that many of their technical difficulties are in common. A notable observation is that the instrumental OLO algorithm constructed within the proof is allowed to have prior knowledge on the quadratic variation of the adversary, whereas in the actual OLO problem the algorithm does not have such prior knowledge therefore the convexity of the  $h$  function becomes crucial. Formally connecting this observation to game-theoretic probability [SV19] would be a valuable direction for future works.



## F Discussion of Nonconvex Tradeoffs

Deviating from the rest of this work, this section discusses nonconvex performance tradeoffs in OLO, i.e., achieving

$$\text{Loss}_T \leq -\text{Bound} \left( -\sum_{t=1}^T g_t \right) \quad (18)$$

analogous to Eq.(1) but the function  $\text{Bound} : \mathbb{R} \rightarrow \mathbb{R}$  here is nonconvex. Since the convex conjugate of any function is automatically convex, the function  $\text{Bound}$  cannot be expressed as a convex conjugate  $\psi_T^*$ , therefore total loss bounds of this type cannot be recovered by regret bounds through the loss-regret duality. For such a nonconvex tradeoff regime, existing results are scarce.

First, we note that Cover’s characterization of performance tradeoffs in OLO (Theorem 1) also works in this nonconvex regime since the adversary is assumed to be Boolean. Concretely, the following generalization of Theorem 1 is true: for all 1-Lipschitz function  $\text{Bound} : \mathbb{R} \rightarrow (-\infty, \infty]$ , there exists an algorithm achieving the total loss bound Eq.(18) against Boolean adversaries if and only if

$$\mathbb{E}_{X \sim \text{RS}(T)}[\text{Bound}(X)] \leq 0.$$

The associated algorithm is still given by Eq.(3), and our goal is to give a computationally efficient improvement.

By inspection, the proof of our Theorem 3 only uses the convexity of the  $h$  function through Lemma B.3. Therefore when the  $h$  function given to Algorithm 1 is 1-Lipschitz but not necessarily convex, the following variant of Theorem 3 is immediate. Also see Theorem 12 for the analogue of this result in the stochastic setting of OLO.

**Theorem 14** (Variant of Theorem 3 without convexity). *Against all adversaries (potentially unbounded a priori), Algorithm 1 given any 1-Lipschitz function  $h$  guarantees*

$\text{Loss}_T$

$$\leq -\mathbb{E}_Z \left[ h \left( \sum_{t=1}^T g_t + \rho_T Z \right) \right] + \mathbb{E}_Z[h(\rho_0 Z)] + \sum_{t=1}^T \left( \sqrt{\frac{2}{\pi}} \frac{|g_t^2 - c_t|}{\rho_{t-1}} + \frac{2c_t |g_t| + |g_t^3|}{\rho_{t-1}^2} \right).$$

Similar to the idea of Theorem 3, the first two terms on the RHS essentially match Cover’s achievability condition therefore constitute the ideal bound to aim for, whereas the remaining sum should be understood as the discretization error. To maximize the interpretability, consider again the case where  $|g_t| \leq 1$  for all  $t$  and accordingly Algorithm 1 sets  $\rho_t = \sqrt{T-t}$  and  $c_t = 1$ . We have

$$\sum_{t=1}^T \frac{2c_t |g_t| + |g_t^3|}{\rho_{t-1}^2} = O(\log T),$$

but the other part of the discretization error can only be generally bounded by

$$\sum_{t=1}^T \sqrt{\frac{2}{\pi}} \frac{|g_t^2 - c_t|}{\rho_{t-1}} = \max_{t \in [1:T]} |g_t^2 - c_t| \cdot O(\sqrt{T}).$$

While this is still a meaningful result, unless  $c_t = g_t^2$  for all  $t$  we can no longer argue about the additive sharpness of the obtained total loss bound since the  $O(\sqrt{T})$  discretization error is often significant.

To summarize, in this nonconvex tradeoff regime:

- If the adversary is Boolean, then by setting  $\rho_t = \sqrt{T-t}$  thus  $g_t^2 = c_t$  for all  $t$ , we obtain a computationally efficient improvement of Cover’s algorithm, and the associated total loss bound is additively sharp (akin to the convex tradeoff regime, Theorem 3).
- If the adversary satisfies  $|g_t| \leq 1$  but is not Boolean, then Cover’s total loss bound does not apply, whereas our bound is meaningful but not additively sharp. This is due to an inflated discretization error term related to the sequence of misspecification error  $|g_t^2 - c_t|$ .

## G Omitted Proofs

This section presents the omitted proofs thus far. Appendix G.1 proves the complete version of Lemma 2.1, giving a more direct representation of Algorithm 1's output. Appendix G.2 proves Theorem 4 (our general lower bound on  $\text{Loss}_T$ ), and then uses the loss-regret duality to convert both Theorem 3 and Theorem 4 to matching upper and lower bounds on  $\text{Regret}_T(u)$ . Appendix G.3 proves the instantiations of these results on iconic  $h$  functions. Appendix G.4 presents the omitted proofs for Appendix D. Appendix G.5 proves a number of technical lemmas. Appendix G.6 summarizes several classical lemmas we use, whose proofs are omitted.

### G.1 Rewriting the Algorithm's Output

The following lemma is the complete version of Lemma 2.1 in Section 2.

**Lemma G.1** (Rewriting the output). *The output Eq.(5) of Algorithm 1 is equivalent to*

$$\begin{aligned} x_t &= -\frac{1}{2} \int_0^1 \frac{1}{\sqrt{\tau}} \mathbb{E}_Z \left[ h' \left( s_{t-1} + \sqrt{\rho_{t-1}^2 - \tau c_t} Z \right) \right] d\tau \\ &= -\mathbb{E}_{Z \sim \mathcal{N}(0,1); \tau \sim \text{Beta}(\frac{1}{2}, 1)} \left[ h' \left( s_{t-1} + \sqrt{\rho_{t-1}^2 - \tau c_t} Z \right) \right], \end{aligned}$$

as well as

$$\begin{aligned} x_t &= -\int_0^\infty e^{-\tau} \mathbb{E}_Z \left[ h' \left( s_{t-1} + \sqrt{\rho_{t-1}^2 - e^{-2\tau} c_t} Z \right) \right] d\tau \\ &= -\mathbb{E}_{Z \sim \mathcal{N}(0,1); \tau \sim \text{Exp}(1)} \left[ h' \left( s_{t-1} + \sqrt{\rho_{t-1}^2 - e^{-2\tau} c_t} Z \right) \right]. \end{aligned}$$

An immediate remark is that the integrals are well-posed: due to the Lipschitzness of  $h'$ ,  $h'$  exists almost everywhere, and whenever  $h'$  exists it is bounded. Therefore the expectation of  $h'$  exists on distributions with density.

*Proof of Lemma G.1.* Observe that the two equations in the lemma are equivalent by a change of variable. To connect them to Eq.(5), the proof is divided into two parts to avoid the technical nuances of taking the  $\rho_t \rightarrow 0$  limit. The first part tackles the degenerate case of  $\rho_t = 0$  by brute force, which is only possible at  $t = T$  by the requirement of Algorithm 1. The second part tackles the nondegenerate case of  $\rho_t > 0$  where a more structured continuous argument is applicable.

**Part 1** If  $\rho_t = 0$ , i.e.,  $t = T$  and  $\rho_{t-1}^2 = c_t$ , then Eq.(5) becomes  $x_t = f_{s_{t-1}, \rho_{t-1}, h}(s_{t-1})$ . Using the integral representation of the function  $f_{\mu, \sigma, h}$  through the OU semigroup (Eq.(10) in Appendix B), we have

$$f_{\mu, \sigma, h}(\mu) = -\int_0^\infty e^{-\tau} \mathbb{E}_Z \left[ h' \left( \mu + \sigma \sqrt{1 - e^{-2\tau}} Z \right) \right] d\tau.$$

Plugging in  $\mu \leftarrow s_{t-1}$  and  $\sigma \leftarrow \rho_{t-1}$  results in the equivalence of Eq.(5) and the second equation of the lemma.

**Part 2** Next consider  $\rho_t \neq 0$ , i.e.,  $\rho_{t-1}^2 > c_t \geq 0$ . Define the notations  $Y \sim \mathcal{N}(\mu, \sigma^2 - c)$  for general  $\sigma^2 > c \geq 0$ , and  $u(c) := \mathbb{E}_Y[f_{\mu, \sigma, h}(Y)] = \mathbb{E}_Z[f_{\mu, \sigma, h}(\mu + \sqrt{\sigma^2 - c} Z)]$ . The goal of Part 2 is to prove  $u(0) = -\mathbb{E}_Z[h'(\mu + \sigma Z)]$  and

$$u(c) = -\frac{1}{2\sqrt{c}} \int_0^c \frac{1}{\sqrt{\tau}} \mathbb{E}_Z \left[ h' \left( \mu + \sqrt{\sigma^2 - \tau} Z \right) \right] d\tau, \quad \forall c \in (0, \sigma^2). \quad (19)$$

After that, by plugging in  $\mu \leftarrow s_{t-1}$ ,  $\sigma \leftarrow \rho_{t-1}$  and  $c \leftarrow c_t$ , the LHS  $u(c)$  recovers the definition of  $x_t$  in Eq.(5) while the RHS recovers the first equation of the objective lemma. Starting from this point, the subscripts of  $f_{\mu, \sigma, h}$  will be dropped for conciseness.

*Step 1.* Our first step is to differentiate the Stein equation Eq.(4) with respect to  $x$ , where all derivatives exist almost everywhere.

$$\sigma^2 f''(x) - (x - \mu) f'(x) - f(x) = h'(x), \quad a.e.$$

Taking the expectation with respect to  $Y \sim \mathcal{N}(\mu, \sigma^2 - c)$  which is absolutely continuous w.r.t. the Lebesgue measure,

$$\mathbb{E}_Y[\sigma^2 f''(Y) - (Y - \mu)f'(Y) - f(Y)] = \mathbb{E}_Y[h'(Y)].$$

Meanwhile, applying Stein's lemma (Lemma B.1) to the random variable  $Y$  and the function  $f'(x)$  gives (the integrability requirement of Lemma B.1 follows from bounds on the Stein factors, Lemma B.2)

$$(\sigma^2 - c)\mathbb{E}_Y[f''(Y)] = \mathbb{E}_Y[(Y - \mu)f'(Y)].$$

Combining it with the above and with  $u(c) = \mathbb{E}_Y[f(Y)]$ ,

$$\mathbb{E}_Y[cf''(Y)] - u(c) = \mathbb{E}_Y[h'(Y)].$$

Plugging in  $c = 0$  gives  $u(0) = -\mathbb{E}_Z[h'(\mu + \sigma Z)]$ .

*Step 2.* The second step is to consider the derivative  $u'(c)$ . Differentiating the definition of  $u(c) = \mathbb{E}_Z[f(\mu + \sqrt{\sigma^2 - c}Z)]$  gives

$$u'(c) = -\frac{1}{2\sqrt{\sigma^2 - c}}\mathbb{E}_Z\left[Zf'\left(\mu + \sqrt{\sigma^2 - c}Z\right)\right].$$

Applying Stein's lemma (Lemma B.1) again to the random variable  $Z$  and the function  $f'(\mu + \sqrt{\sigma^2 - c}x)$ , where the integrability requirement again follows from Lemma B.2,

$$\sqrt{\sigma^2 - c} \cdot \mathbb{E}_Z\left[f''\left(\mu + \sqrt{\sigma^2 - c}Z\right)\right] = \mathbb{E}_Z\left[Zf'\left(\mu + \sqrt{\sigma^2 - c}Z\right)\right],$$

therefore  $u'(c) = -\frac{1}{2}\mathbb{E}_Y[f''(Y)]$  and thus we obtain the ODE

$$2cu'(c) + u(c) = -\mathbb{E}_Y[h'(Y)].$$

*Step 3.* To solve this ODE, we reformulate it into

$$(\sqrt{cu(c)})' = -\frac{1}{2\sqrt{c}}\mathbb{E}_Y[h'(Y)] = -\frac{1}{2\sqrt{c}}\mathbb{E}_Z\left[h'\left(\mu + \sqrt{\sigma^2 - c}Z\right)\right],$$

and integrating it results in Eq.(19). □

## G.2 General Regret and Total Loss Bounds

We first restate Theorem 4 and prove it as follows.

**Theorem 4** (Lower bound on  $\text{Loss}_T$ ). *There exists an absolute constant  $c > 0$  such that the following holds. For any OLO algorithm and any 1-Lipschitz function  $h : \mathbb{R} \rightarrow \mathbb{R}$ , there exists a Boolean adversary  $(g_t \in \{-1, 1\})$  for all  $t$ ) inducing*

$$\text{Loss}_T > -h\left(\sum_{t=1}^T g_t\right) + \mathbb{E}_Z[h(\sqrt{T}Z)] - c.$$

*Proof of Theorem 4.* The proof is based on a nonasymptotic Wasserstein CLT for iid sums. Recall from Section 1 that  $\text{RS}(n)$  denotes the distribution of the sum of  $n$  independent Rademacher random variables.

First, for any function  $h : \mathbb{R} \rightarrow \mathbb{R}$ , no OLO algorithm can guarantee

$$\text{Loss}_T < -h\left(\sum_{t=1}^T g_t\right) + \mathbb{E}_{X \sim \text{RS}(T)}[h(X)]$$

against all Boolean adversaries. To see this, notice that against iid Rademacher  $g_{1:T}$ , taking the expectation of this bound would lead to the  $0 < 0$  contradiction.

Next, by Lemma G.6 which is due to [Ros11], there exists an absolute constant  $c > 0$  such that for any 1-Lipschitz function  $h : \mathbb{R} \rightarrow \mathbb{R}$  we have

$$\left|\mathbb{E}_{X \sim \text{RS}(T)}[h(X)] - \mathbb{E}_Z[h(\sqrt{T}Z)]\right| < c.$$

Combining it with the above impossibility argument completes the proof. □

The following two corollaries are the regret analogues of Theorem 3 and Theorem 4. The proofs are due to the loss-regret duality (Lemma G.5) therefore omitted.

**Corollary 15** (Upper bound on  $\text{Regret}_T(u)$ ). *Algorithm 1 with  $\rho_T = 0$  guarantees the regret bound*

$$\text{Regret}_T(u) \leq h^*(-u) + \mathbb{E}_Z[h(\rho_0 Z)] + \sum_{t=1}^T \left( \sqrt{\frac{2}{\pi}} \frac{\max\{g_t^2 - c_t, 0\}}{\rho_{t-1}} + \frac{2c_t |g_t| + |g_t^3|}{\rho_{t-1}^2} \right), \quad \forall u \in [-1, 1].$$

**Corollary 16** (Lower bound on  $\text{Regret}_T(u)$ ). *There exists an absolute constant  $c > 0$  such that the following statement holds. For any OLO algorithm and any convex and 1-Lipschitz function  $h : \mathbb{R} \rightarrow \mathbb{R}$ , there exists a Boolean adversary ( $g_t \in \{-1, 1\}$  for all  $t$ ) and a comparator  $u \in [-1, 1]$  such that*

$$\text{Regret}_T(u) > h^*(-u) + \mathbb{E}_Z[h(\sqrt{T}Z)] - c.$$

### G.3 Special Cases

Below we restate and prove the specializations of Corollary 15 in Section 4, Appendix C and Appendix D.

**Corollary 5** (Regret: absolute value). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = |x|$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \text{Regret}_T^{\text{unif}} \leq \sqrt{\frac{2}{\pi}} T + O(\log T), \quad \forall u \in [-1, 1].$$

*Proof of Corollary 5.* Starting from Corollary 15, it can be verified that  $h^*(-u) = -h(0) = 0$  for all  $u \in [-1, 1]$ , and  $\mathbb{E}_Z[h(\sqrt{T}Z)] = \sqrt{T} \mathbb{E}_Z[|Z|] = \sqrt{\frac{2}{\pi}} T$ .  $\square$

**Corollary 6** (Regret: Huber). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = \frac{\eta}{2} x^2 \cdot \mathbf{1}[|x| \leq \eta^{-1}] + (|x| - \frac{1}{2\eta}) \cdot \mathbf{1}[|x| > \eta^{-1}]$  for  $\eta = \frac{\alpha}{\sqrt{T}}$  and  $\alpha \in \mathbb{R}_{>0}$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{u^2}{2\alpha} + \left( \alpha + \frac{1}{\alpha} \right) \Phi\left(\frac{1}{\alpha}\right) + \phi\left(\frac{1}{\alpha}\right) - \frac{1}{2}\alpha - \frac{1}{\alpha} \right]}_{=:\gamma_{\text{Huber}}(u, \alpha)} \sqrt{T} + O(\log T), \quad \forall u \in [-1, 1].$$

*Proof of Corollary 6.* Starting from Corollary 15,

$$h^*(-u) = -ux - h(x) \Big|_{x=-\frac{u}{\eta}} = \frac{u^2}{2\eta}.$$

For all  $v > 0$ ,

$$\begin{aligned} \mathbb{E}_Z[h(vZ)] &= \int_{-\infty}^{-(v\eta)^{-1}} \left( -vz - \frac{1}{2\eta} \right) \phi(z) \, dz + \frac{\eta}{2} \int_{-(v\eta)^{-1}}^{(v\eta)^{-1}} v^2 z^2 \phi(z) \, dz + \int_{(v\eta)^{-1}}^{\infty} \left( vz - \frac{1}{2\eta} \right) \phi(z) \, dz \\ &= 2v\phi\left(\frac{1}{v\eta}\right) - \frac{1}{\eta} \Phi\left(-\frac{1}{v\eta}\right) + \frac{\eta}{2} v^2 [\Phi(z) - z\phi(z)] \Big|_{z=-(v\eta)^{-1}}^{(v\eta)^{-1}} \\ &= v^2 \eta \left[ \Phi\left(\frac{1}{v\eta}\right) - \frac{1}{2} \right] + v\phi\left(\frac{1}{v\eta}\right) - \frac{1}{\eta} \left[ 1 - \Phi\left(\frac{1}{v\eta}\right) \right]. \end{aligned}$$

Therefore with  $v \leftarrow \sqrt{T}$  and  $\eta = \frac{\alpha}{\sqrt{T}}$ ,

$$\begin{aligned} \text{Regret}_T(u) &\leq \frac{u^2}{2\eta} + \eta T \left[ \Phi\left(\frac{1}{\eta\sqrt{T}}\right) - \frac{1}{2} \right] + \sqrt{T} \phi\left(\frac{1}{\eta\sqrt{T}}\right) - \frac{1}{\eta} \left[ 1 - \Phi\left(\frac{1}{\eta\sqrt{T}}\right) \right] + O(\log T) \\ &= \left[ \frac{u^2}{2\alpha} + \left( \alpha + \frac{1}{\alpha} \right) \Phi\left(\frac{1}{\alpha}\right) + \phi\left(\frac{1}{\alpha}\right) - \frac{1}{2}\alpha - \frac{1}{\alpha} \right] \sqrt{T} + O(\log T). \end{aligned} \quad \square$$

**Corollary 7** (Regret: log-sum-exp). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = \frac{1}{\eta} \ln(\cosh(\eta x))$  for  $\eta = \frac{\alpha}{\sqrt{T}}$  and  $\alpha \in \mathbb{R}_{>0}$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{1}{2\alpha} \ln((1+u)^{1+u}(1-u)^{1-u}) + \mathbb{E}_Z [\alpha^{-1} \ln(\cosh(\alpha Z))] \right]}_{=: \gamma_{\text{LSE}}(u, \alpha)} \sqrt{T} + O(\log T), \quad \forall u \in [-1, 1].$$

In particular, the RHS follows the convention  $0^0 = 1$  when  $u = \pm 1$ .

*Proof of Corollary 7.* Starting from Corollary 15, we first consider  $u \in (-1, 1)$ .

$$\begin{aligned} h^*(-u) &= -ux - \frac{1}{\eta} \ln(\cosh(\eta x)) \Big|_{x=\frac{1}{\eta} \tanh^{-1}(-u)} \\ &= \frac{1}{\eta} \left[ -u \tanh^{-1}(-u) - \ln \frac{1}{\sqrt{1-u^2}} \right] & (\cosh(\tanh^{-1}(-u)) = \frac{1}{\sqrt{1-u^2}}) \\ &= \frac{1}{2\eta} \left[ -u \ln \frac{1-u}{1+u} + \ln(1-u^2) \right] & (\tanh^{-1}(-u) = \frac{1}{2} \ln \frac{1-u}{1+u}) \\ &= \frac{1}{2\eta} [(1+u) \ln(1+u) + (1-u) \ln(1-u)]. \end{aligned}$$

The case of  $u = \pm 1$  follows from  $\lim_{x \rightarrow 0^+} x \ln x = 0$ . Further plugging the definition of  $h$  into  $\mathbb{E}_Z[h(\sqrt{T}Z)]$  completes the proof.  $\square$

**Corollary 8** (Regret: soft-thresholded). *Assume  $|g_t| \leq 1$  for all  $t$ . If  $h(x) = \max\{|x| - \eta^{-1}, 0\}$  for  $\eta = \frac{\alpha}{\sqrt{T}}$  and  $\alpha \in \mathbb{R}_{>0}$ , then Algorithm 1 with  $\rho_t = \sqrt{T-t}$  guarantees*

$$\text{Regret}_T(u) \leq \underbrace{\left[ \frac{|u|}{\alpha} + \frac{2}{\alpha} \Phi\left(\frac{1}{\alpha}\right) + 2\phi\left(\frac{1}{\alpha}\right) - \frac{2}{\alpha} \right]}_{=: \gamma_{\text{STh}}(u, \alpha)} \sqrt{T} + O(\log T), \quad \forall u \in [-1, 1].$$

*Proof of Corollary 8.* Starting from Corollary 15,

$$h^*(-u) = \sup_{x \in \mathbb{R}} -ux - \max\{|x| - \eta^{-1}, 0\} = -ux \Big|_{x=-\eta^{-1} \text{sign}(u)} = \eta^{-1} |u|.$$

For all  $v > 0$ ,

$$\begin{aligned} \mathbb{E}_Z[h(vZ)] &= \int_{-\infty}^{-(v\eta)^{-1}} \left( -vz - \frac{1}{\eta} \right) \phi(z) dz + \int_{(v\eta)^{-1}}^{\infty} \left( vz - \frac{1}{\eta} \right) \phi(z) dz \\ &= 2v\phi\left(\frac{1}{v\eta}\right) - \frac{2}{\eta} \left[ 1 - \Phi\left(\frac{1}{v\eta}\right) \right]. \end{aligned}$$

Substituting  $v \leftarrow \sqrt{T}$  and  $\eta = \frac{\alpha}{\sqrt{T}}$  completes the proof.  $\square$

## G.4 Proofs for the Two-Point Tradeoff

Below we restate and prove the results from Appendix D.

**Lemma D.1** (Optimal two-point tradeoff via  $u$ -dependent regret bound). *For any  $a, b \in \mathbb{R}$  and any OLO algorithm  $\text{alg}$ , the following two conditions are equivalent:*

1.  $\sup_{\text{adv}} \text{Loss}_T(\text{alg}, \text{adv}) \leq a$  and  $\sup_{\text{adv}} \text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv}) \leq b$ ;
2.  $\sup_{\text{adv}, u \in [-1, 1]} \text{Regret}_T(u, \text{alg}, \text{adv}) \leq a + (b - a)|u|$ .

*Proof of Lemma D.1.* Let us drop  $\text{alg}$  and  $\text{adv}$  from the notations. By definition,  $\text{Loss}_T = \text{Regret}_T(0)$  and  $\text{Regret}_T^{\text{unif}} = \max_{u \in \{-1, 1\}} \text{Regret}_T(u)$ , therefore the second condition in the lemma trivially implies the first one. Now consider going from the first to the second:

$$\begin{aligned} \text{Regret}_T(u) &= \sum_{t=1}^T g_t(x_t - u) \\ &= (1 - |u|) \sum_{t=1}^T g_t x_t + |u| \sum_{t=1}^T g_t(x_t - \text{sign}(u)) \\ &\leq (1 - |u|)a + |u|b. \end{aligned} \quad \square$$

**Theorem 9** (Characterization of  $\gamma(\varepsilon)$ ). *For all constant  $\varepsilon \in (0, \sqrt{\frac{2}{\pi}}]$ ,  $\gamma(\varepsilon)$  from Eq.(13) is the unique solution of the equation*

$$\int_{-\infty}^{\varepsilon - \gamma(\varepsilon)} \Phi(x) dx = \frac{\varepsilon}{2}.$$

Furthermore, for all  $T \in \mathbb{N}_+$ , there exists a special case of Algorithm 1 that simultaneously guarantees  $\text{Loss}_T \leq \varepsilon\sqrt{T} + O(\log T)$  and  $\text{Regret}_T^{\text{unif}} \leq \gamma(\varepsilon)\sqrt{T} + O(\log T)$ .

*Proof of Theorem 9.* Throughout this proof, let  $c > 0$  be a sufficiently large absolute constant. We only prove matching upper and lower bounds on  $\gamma(\varepsilon)$ , since the corresponding algorithm can be simply obtained from Corollary 8.

An upper bound on  $\gamma(\varepsilon)$  is immediate: with the hyperparameter  $\alpha$  satisfying  $\gamma_{\text{STh}}(0, \alpha)\sqrt{T} + c \log T = \varepsilon\sqrt{T}$ , applying Corollary 8 yields

$$\begin{aligned} \gamma(\varepsilon) &= \limsup_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \inf_{\text{alg}} \left\{ \sup_{\text{adv}} \text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv}); \sup_{\text{adv}} \text{Loss}_T(\text{alg}, \text{adv}) \leq \varepsilon\sqrt{T} \right\} \\ &\leq \lim_{T \rightarrow \infty} \left( \gamma_{\text{STh}}(1, \alpha) + c \frac{\log T}{\sqrt{T}} \right) \Big|_{\gamma_{\text{STh}}(0, \alpha) = \varepsilon - cT^{-1/2} \log T} \\ &= \gamma_{\text{STh}}(1, \alpha) \Big|_{\gamma_{\text{STh}}(0, \alpha) = \varepsilon}. \end{aligned}$$

As for the lower bound, applying Corollary 16 on the  $h$  function from Corollary 8 shows that for any algorithm  $\text{alg}$ , there exists a Boolean adversary  $\text{adv}$  and a comparator  $u \in [-1, 1]$  such that

$$\text{Regret}_T(u, \text{alg}, \text{adv}) > \gamma_{\text{STh}}(u, \alpha)\sqrt{T} - c.$$

By Lemma D.1, this is equivalent to the violation of either

$$\sup_{\text{adv}} \text{Loss}_T(\text{alg}, \text{adv}) \leq \gamma_{\text{STh}}(0, \alpha)\sqrt{T} - c$$

or

$$\sup_{\text{adv}} \text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv}) \leq \gamma_{\text{STh}}(1, \alpha)\sqrt{T} - c.$$

Therefore,

$$\begin{aligned} \gamma(\varepsilon) &= \limsup_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \inf_{\text{alg}} \left\{ \sup_{\text{adv}} \text{Regret}_T^{\text{unif}}(\text{alg}, \text{adv}); \sup_{\text{adv}} \text{Loss}_T(\text{alg}, \text{adv}) \leq \varepsilon\sqrt{T} \right\} \\ &\geq \lim_{T \rightarrow \infty} \left( \gamma_{\text{STh}}(1, \alpha) - \frac{c}{\sqrt{T}} \right) \Big|_{\gamma_{\text{STh}}(0, \alpha) = \varepsilon + cT^{-1/2}} \\ &= \gamma_{\text{STh}}(1, \alpha) \Big|_{\gamma_{\text{STh}}(0, \alpha) = \varepsilon}. \end{aligned}$$

Combining the upper and lower bounds, we have  $\gamma(\varepsilon) = \gamma_{\text{STh}}(1, \alpha) \Big|_{\gamma_{\text{STh}}(0, \alpha) = \varepsilon}$ . Finally applying Lemma G.3 completes the proof.  $\square$



## G.5 Proofs of Technical Lemmas

Our analysis uses several technical lemmas proved below.

**Lemma G.2.** *For all  $\alpha \in \mathbb{R}_{>0}$ , let  $f(\alpha) := \mathbb{E}_Z [\alpha^{-1} \ln(\cosh(\alpha Z))]$ . Then,  $f(\alpha) < \frac{1}{2}\alpha$  and  $\lim_{\alpha \rightarrow \infty} f(\alpha) = \sqrt{\frac{2}{\pi}}$ .*

*Proof of Lemma G.2.* Starting from the first part of the lemma, applying the fact of  $\cosh(x) < \exp\left(\frac{x^2}{2}\right)$  for all  $x \in \mathbb{R} \setminus \{0\}$  gives

$$f(\alpha) < \mathbb{E}_Z \left[ \alpha^{-1} \ln \left( \exp \left( \frac{1}{2} \alpha^2 Z^2 \right) \right) \right] = \frac{1}{2} \alpha \mathbb{E}_Z [Z^2] = \frac{1}{2} \alpha.$$

Next, for all  $x \in \mathbb{R}$ ,

$$\alpha^{-1} \ln(\cosh(\alpha x)) = |x| + \alpha^{-1} \ln \frac{1 + \exp(-2\alpha |x|)}{2}.$$

Denote the second term on the RHS as  $g(\alpha, x)$ , which satisfies the pointwise convergence  $\lim_{\alpha \rightarrow \infty} g(\alpha, x) = 0$ . Furthermore, for all  $\alpha \geq 1$  we have  $|g(\alpha, x)| \leq \alpha^{-1} \ln 2 \leq \ln 2$ , therefore by the dominated convergence theorem,  $\lim_{\alpha \rightarrow \infty} \mathbb{E}_Z [g(\alpha, Z)] = 0$ . This leads to

$$\lim_{\alpha \rightarrow \infty} f(\alpha) = \mathbb{E}_Z [|Z|] = \sqrt{\frac{2}{\pi}}. \quad \square$$

**Lemma G.3.** *For the quantity  $\gamma_{\text{STh}}(u, \alpha)$  defined in Corollary 8, we have*

$$\int_{-\infty}^{\gamma_{\text{STh}}(0, \alpha) - \gamma_{\text{STh}}(1, \alpha)} \Phi(x) dx = \frac{1}{2} \gamma_{\text{STh}}(0, \alpha).$$

*Proof of Lemma G.3.* Recall that

$$\gamma_{\text{STh}}(u, \alpha) = \frac{|u|}{\alpha} + \frac{2}{\alpha} \Phi\left(\frac{1}{\alpha}\right) + 2\phi\left(\frac{1}{\alpha}\right) - \frac{2}{\alpha},$$

therefore  $\gamma_{\text{STh}}(0, \alpha) - \gamma_{\text{STh}}(1, \alpha) = -\alpha^{-1}$ . It can be simply verified that for all  $x \in \mathbb{R}$ ,  $\int_{-\infty}^x \Phi(z) dz = x\Phi(x) + \phi(x)$ . Therefore

$$\int_{-\infty}^{-\alpha^{-1}} \Phi(x) dx = -\frac{1}{\alpha} \Phi\left(-\frac{1}{\alpha}\right) + \phi\left(-\frac{1}{\alpha}\right) = -\frac{1}{\alpha} \left[ 1 - \Phi\left(\frac{1}{\alpha}\right) \right] + \phi\left(\frac{1}{\alpha}\right) = \frac{1}{2} \gamma_{\text{STh}}(0, \alpha). \quad \square$$

**Lemma G.4.** *For all  $\varepsilon \in (0, \sqrt{\frac{2}{\pi}}]$ ,  $\gamma(\varepsilon)$  defined in Eq.(13) satisfies*

$$\gamma(\varepsilon) - \varepsilon < \sqrt{2} \cdot \text{erfi}^{-1} \left( \frac{\sqrt{2}}{\sqrt{\pi\varepsilon}} \right).$$

*Proof of Lemma G.4.* It suffices to prove

$$\int_{-\infty}^{\varepsilon - \gamma(\varepsilon)} \Phi(x) dx > \int_{-\infty}^{-\sqrt{2} \cdot \text{erfi}^{-1} \left( \frac{\sqrt{2}}{\sqrt{\pi\varepsilon}} \right)} \Phi(x) dx,$$

where LHS =  $\frac{\varepsilon}{2}$  by Theorem 9. By a change of variable  $z = \text{erfi}^{-1} \left( \frac{\sqrt{2}}{\sqrt{\pi\varepsilon}} \right)$ , this is further equivalent to proving

$$\text{erfi}(z) \int_{-\infty}^{-\sqrt{2}z} \Phi(x) dx < \sqrt{\frac{2}{\pi}}, \quad \forall z \geq \text{erfi}^{-1}(1). \quad (20)$$

To this end, notice that for all  $x < 0$  we have

$$\int_{-\infty}^x \Phi(z) dz = x\Phi(x) + \phi(x)$$

$$\begin{aligned}
&= x[1 - \Phi(-x)] + \phi(-x) \\
&< 2\phi(-x).
\end{aligned}
\tag{Mills ratio bound; Lemma G.7}$$

Therefore the LHS of Eq.(20) is upper-bounded by

$$\operatorname{erfi}(z) \int_{-\infty}^{-\sqrt{2}z} \Phi(x) dx < 2\operatorname{erfi}(z)\phi(\sqrt{2}z) = \sqrt{\frac{2}{\pi}} \exp(-z^2) \int_0^z \exp(x^2) dx \leq \sqrt{\frac{2}{\pi}}. \quad \square$$

## G.6 Summary of Existing Lemmas

This subsection summarizes several known results from the literature, whose proofs are omitted.

The first one is a formal version of the loss-regret duality [Ora25, Theorem 10.6] used throughout our analysis. Notice that if the function  $\psi$  has a convex and compact domain, then the following lemma can still be applied by assigning  $\psi$  as  $\infty$  outside its domain, which is standard.

**Lemma G.5** (Loss-regret duality). *Let  $\psi : \mathbb{R} \rightarrow (-\infty, \infty]$  be a proper, closed and convex function, and let  $\psi^*$  be its convex conjugate. For any two sequences  $x_1, \dots, x_T \in \mathbb{R}$  and  $g_1, \dots, g_T \in \mathbb{R}$ , we have*

$$\sum_{t=1}^T g_t x_t \leq -\psi^* \left( -\sum_{t=1}^T g_t \right)$$

if and only if

$$\sum_{t=1}^T g_t (x_t - u) \leq \psi(u), \quad \forall u \in \mathbb{R}.$$

The next lemma is a Wasserstein normal approximation bound for iid Rademacher sums, simplified from [Ros11, Theorem 3.2]. The constant is not optimized.

**Lemma G.6** (Nonasymptotic Wasserstein CLT for iid Rademacher sums). *Let  $X_1, \dots, X_n$  be iid Rademacher random variables. If  $S_n = \sum_{i=1}^n X_i$  and  $Z$  has the standard normal distribution, then for any 1-Lipschitz function  $f : \mathbb{R} \rightarrow \mathbb{R}$  we have*

$$|\mathbb{E}[f(S_n)] - \mathbb{E}_Z[f(\sqrt{n}Z)]| \leq 1 + \sqrt{\frac{2}{\pi}}.$$

The final lemma is a standard estimate of the Mills ratio, due to [Gor41].

**Lemma G.7** (Mills ratio bound). *Consider the Mills ratio defined as  $m(x) := \frac{1-\Phi(x)}{\phi(x)}$ . For all  $x > 0$ ,*

$$\frac{x}{x^2 + 1} < m(x) < \frac{1}{x}.$$

## H Closed Forms of the Output

To provide additional intuition and ease the implementation of our algorithm, this section derives the closed forms of Algorithm 1's output (as well as the  $f_{\mu,\sigma,h}$  function), focusing on the two special cases of  $h$  from Section 4. The goal is to convert the computation of  $x_t$  to querying common Gaussian-integral-type special functions. Auxiliary lemmas for the derivation are provided in Appendix H.3.

### H.1 Absolute Value Function

Consider the setting of Section 4.1.

**Proposition 17** (Algorithm: absolute value). *If  $h(x) = |x|$ , then in Algorithm 1,*

$$f_{\mu,\sigma,h}(x) = \begin{cases} 1 - \frac{\mu + \mathbb{E}_Z[|\mu + \sigma Z|]}{\sigma^2} \frac{\Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}, & x \leq 0, \\ -1 - \frac{\mu - \mathbb{E}_Z[|\mu + \sigma Z|]}{\sigma^2} \frac{1 - \Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}, & x > 0. \end{cases}$$

Furthermore, the output Eq.(5) is equivalent to

$$x_t = 1 - \int_0^1 \frac{1}{\sqrt{\tau}} \Phi \left( \frac{s_{t-1}}{\sqrt{\rho_{t-1}^2 - \tau c_t}} \right) d\tau, \quad (21)$$

which can be further expressed as one of the following three cases:

- If  $\rho_{t-1}^2 > c_t = 0$ , then

$$x_t = 1 - 2\Phi \left( \frac{s_{t-1}}{\rho_{t-1}} \right).$$

- If  $\rho_{t-1}^2 = c_t > 0$  which is only possible at  $t = T$  by the requirement of Algorithm 1, then

$$x_t = -\text{sign}(s_{t-1}) \left[ 1 + \sqrt{\frac{\pi}{2}} \frac{|s_{t-1}|}{\rho_{t-1}} - \sqrt{\frac{\pi}{2}} \frac{s_{t-1}}{\rho_{t-1}} \left( 2\Phi \left( \frac{s_{t-1}}{\rho_{t-1}} \right) - 1 \right) - \sqrt{2\pi} \phi \left( \frac{s_{t-1}}{\rho_{t-1}} \right) \right].$$

- If  $\rho_{t-1}^2 > c_t > 0$ , then with Owen's  $T$  function  $T(x, y) := \frac{1}{2\pi} \int_0^y \frac{e^{-\frac{1}{2}x^2(1+z^2)}}{1+z^2} dz$ ,

$$x_t = 1 - 2\Phi \left( \frac{s_{t-1}}{\sqrt{\rho_{t-1}^2 - c_t}} \right) - \frac{\sqrt{2\pi}\rho_{t-1}}{\sqrt{c_t}} \phi \left( \frac{s_{t-1}}{\rho_{t-1}} \right) \left[ 1 - 2\Phi \left( \frac{s_{t-1}}{\rho_{t-1}} \sqrt{\frac{c_t}{\rho_{t-1}^2 - c_t}} \right) \right] - \frac{2\sqrt{2\pi}s_{t-1}}{\sqrt{c_t}} T \left( \frac{s_{t-1}}{\rho_{t-1}}, \sqrt{\frac{c_t}{\rho_{t-1}^2 - c_t}} \right). \quad (22)$$

For implementation, both the normal CDF and the Owen's  $T$  function can be queried from standard software packages, such as SciPy.

Several remarks are in order.

- To make sense of these somewhat complicated expressions, first notice that all the expressions are consistent with the standard dimensional analysis, where  $\mu$ ,  $\sigma$  and  $\sqrt{c}$  sharing the same “unit”.
- Next, the case of  $\rho_{t-1}^2 > c_t = 0$  is the most interpretable one. It can be recovered from the general case of  $\rho_{t-1}^2 > c_t > 0$ : as  $c_t \rightarrow 0$ , the second line of Eq.(22) vanishes, therefore  $x_t \rightarrow 1 - 2\Phi(s_{t-1}/\rho_{t-1})$ . This equals the probability of a standard Brownian motion starting at  $s_{t-1}$  ending up being negative after  $\rho_{t-1}^2$  time. Also see Section 2 for discussions of this continuous time approximation.
- As shown below, the “last round special case”  $\rho_{t-1}^2 = c_t > 0$  and the general case  $\rho_{t-1}^2 > c_t > 0$  are proved using different strategies: the former is based on the closed form of  $f_{\mu,\sigma,h}$  while the latter is based on the integral Eq.(21). Therefore a simple sanity check is to verify that the latter recovers the former in the limit of  $c_t \rightarrow \rho_{t-1}^2$ , which follows from  $T(x, \infty) = \frac{1}{2}[1 - \Phi(|x|)]$  [Owe80, Eq.(2.4) in Table II].

*Proof of Proposition 17.* The proof is divided into three parts: the first part analyzes the function  $f_{\mu,\sigma,h}$ , the second part derives the general integral representation Eq.(21) of  $x_t$ , and the third part relates  $x_t$  to common special functions.

**Part 1** Starting from the representation Eq.(9) of the solution of Stein equation, for all  $x \leq 0$ ,

$$\begin{aligned} f_{\mu,\sigma,h}(x) &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \int_{-\infty}^x (-z - \mathbb{E}_Z[|\mu + \sigma Z|]) \phi_{\mu,\sigma}(z) dz \\ &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left( \int_{-\infty}^x (-z + \mu) \phi_{\mu,\sigma}(z) dz - (\mu + \mathbb{E}_Z[|\mu + \sigma Z|]) \int_{-\infty}^x \phi_{\mu,\sigma}(z) dz \right) \\ &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left( \sigma^2 \int_{-\infty}^x \phi'_{\mu,\sigma}(z) dz - (\mu + \mathbb{E}_Z[|\mu + \sigma Z|]) \int_{-\infty}^x \phi_{\mu,\sigma}(z) dz \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} (\sigma^2 \phi_{\mu,\sigma}(x) - (\mu + \mathbb{E}_Z[|\mu + \sigma Z|]) \Phi_{\mu,\sigma}(x)) \\
&= 1 - \frac{\mu + \mathbb{E}_Z[|\mu + \sigma Z|]}{\sigma^2} \frac{\Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}.
\end{aligned}$$

The other case of  $x > 0$  is similar,

$$\begin{aligned}
f_{\mu,\sigma,h}(x) &= -\frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \int_x^\infty (z - \mathbb{E}_Z[|\mu + \sigma Z|]) \phi_{\mu,\sigma}(z) dz \\
&= -\frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left( -\int_x^\infty (-z + \mu) \phi_{\mu,\sigma}(z) dz + (\mu - \mathbb{E}_Z[|\mu + \sigma Z|]) \int_x^\infty \phi_{\mu,\sigma}(z) dz \right) \\
&= -\frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left( -\sigma^2 \int_x^\infty \phi'_{\mu,\sigma}(z) dz + (\mu - \mathbb{E}_Z[|\mu + \sigma Z|]) \int_x^\infty \phi_{\mu,\sigma}(z) dz \right) \\
&= -\frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} (\sigma^2 \phi_{\mu,\sigma}(x) + (\mu - \mathbb{E}_Z[|\mu + \sigma Z|]) (1 - \Phi_{\mu,\sigma}(x))) \\
&= -1 - \frac{\mu - \mathbb{E}_Z[|\mu + \sigma Z|]}{\sigma^2} \frac{1 - \Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}.
\end{aligned}$$

**Part 2** Next we prove Eq.(21). By Lemma G.1,  $x_t$  can be represented as the following integral with  $\mu \leftarrow s_{t-1}$ ,  $\sigma \leftarrow \rho_{t-1} > 0$  and  $c \leftarrow c_t > 0$ ,

$$x_t = -\frac{1}{2} \int_0^1 \frac{1}{\sqrt{\tau}} \mathbb{E}_Z \left[ \text{sign} \left( \mu + \sqrt{\sigma^2 - \tau c} Z \right) \right] d\tau.$$

For the integrand, it is valid to only consider  $\tau < 1$  and obtain

$$\mathbb{E}_Z \left[ \text{sign} \left( \mu + \sqrt{\sigma^2 - \tau c} Z \right) \right] = 1 - 2\mathbb{P} \left( \mu + \sqrt{\sigma^2 - \tau c} Z \leq 0 \right) = 2\Phi \left( \frac{\mu}{\sqrt{\sigma^2 - \tau c}} \right) - 1.$$

Therefore combining them gives Eq.(21) stated in the proposition.

**Part 3** The final step is to relate Eq.(21) to common special functions. The case of  $\rho_{t-1}^2 > c_t = 0$  simply follows from Eq.(21). The general case of  $\rho_{t-1}^2 > c_t \geq 0$  follows from Lemma H.1, which is separated for reusability. Below we only consider the last round special case,  $\rho_{t-1}^2 = c_t > 0$ .

Let  $\mu \leftarrow s_{t-1}$  and  $\sigma \leftarrow \rho_{t-1} > 0$ . The definition of  $x_t$  in Eq.(5) and the above results on  $f_{\mu,\sigma,h}$  give

$$x_t = f_{\mu,\sigma,h}(\mu) = \begin{cases} 1 - \frac{\mu + \mathbb{E}_Z[|\mu + \sigma Z|]}{\sigma^2} \frac{\Phi_{\mu,\sigma}(\mu)}{\phi_{\mu,\sigma}(\mu)}, & \mu \leq 0, \\ -1 - \frac{\mu - \mathbb{E}_Z[|\mu + \sigma Z|]}{\sigma^2} \frac{1 - \Phi_{\mu,\sigma}(\mu)}{\phi_{\mu,\sigma}(\mu)}, & \mu > 0, \end{cases}$$

where  $\Phi_{\mu,\sigma}(\mu) = \frac{1}{2}$  and  $\phi_{\mu,\sigma}(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$ . These two cases on  $\mu$  can be combined into

$$x_t = -\text{sign}(\mu) \left[ 1 + \frac{\sqrt{2\pi}}{2\sigma} (|\mu| - \mathbb{E}_Z[|\mu + \sigma Z|]) \right],$$

where

$$\mathbb{E}_Z[|\mu + \sigma Z|] = \int_{-\infty}^{-\frac{\mu}{\sigma}} (-\mu - \sigma x) \phi(x) dx + \int_{-\frac{\mu}{\sigma}}^{\infty} (\mu + \sigma x) \phi(x) dx = \mu \left[ 2\Phi \left( \frac{\mu}{\sigma} \right) - 1 \right] + 2\sigma \phi \left( \frac{\mu}{\sigma} \right). \quad \square$$

## H.2 Huber Function

Consider the setting of Section 4.2.

**Proposition 18** (Algorithm: Huber). *If  $h(x) = \frac{k}{2}x^2 \cdot \mathbf{1}[|x| \leq k^{-1}] + (|x| - \frac{1}{2k}) \cdot \mathbf{1}[|x| > k^{-1}]$  for some  $k \in \mathbb{R}_{>0}$ , then in Algorithm 1,*

$$f_{\mu,\sigma,h}(x) = \begin{cases} 1 - \frac{\mu + \mathbb{E}_Z[h(\mu + \sigma Z)] + (2k)^{-1} \frac{\Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}}{\sigma^2}, & x < -k^{-1}, \\ \frac{1}{2\sigma^2 \phi_{\mu,\sigma}(x)} \left[ (k\mu^2 + k\sigma^2 - 2\mathbb{E}_Z[h(\mu + \sigma Z)])\Phi_{\mu,\sigma}(x) - k\sigma^2(x + \mu)\phi_{\mu,\sigma}(x) \right. \\ \quad \left. - (k\mu^2 + k\sigma^2 + 2\mu + k^{-1})\Phi_{\mu,\sigma}(-k^{-1}) + \sigma^2(1 + k\mu)\phi_{\mu,\sigma}(-k^{-1}) \right], & -k^{-1} \leq x \leq k^{-1}, \\ -1 - \frac{\mu - \mathbb{E}_Z[h(\mu + \sigma Z)] - (2k)^{-1} \frac{1 - \Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}}{\sigma^2}, & x > k^{-1}. \end{cases}$$

Furthermore, the output Eq.(5) is equivalent to

$$x_t = 1 - \frac{1}{2} \int_0^1 \frac{1}{\sqrt{\tau}} \left[ (1 - ks_{t-1})\Phi\left(\frac{s_{t-1} - k^{-1}}{\sqrt{\rho_{t-1}^2 - \tau c_t}}\right) + (1 + ks_{t-1})\Phi\left(\frac{s_{t-1} + k^{-1}}{\sqrt{\rho_{t-1}^2 - \tau c_t}}\right) \right. \\ \left. - k\sqrt{\rho_{t-1}^2 - \tau c_t} \left( \phi\left(\frac{s_{t-1} - k^{-1}}{\sqrt{\rho_{t-1}^2 - \tau c_t}}\right) - \phi\left(\frac{s_{t-1} + k^{-1}}{\sqrt{\rho_{t-1}^2 - \tau c_t}}\right) \right) \right] d\tau. \quad (23)$$

Slightly different from Proposition 17, in Proposition 18 we leave the output  $x_t$  as a general integral Eq.(23) for brevity. Nonetheless, the same idea still holds: Eq.(23) can be further converted to a closed form using Gaussian-integral-type special functions. Concretely,

- If  $\rho_{t-1}^2 > c_t = 0$ , then

$$x_t = 1 - \Phi\left(\frac{s_{t-1} - k^{-1}}{\rho_{t-1}}\right) - \Phi\left(\frac{s_{t-1} + k^{-1}}{\rho_{t-1}}\right) \\ + k \left[ s_{t-1} \left( \Phi\left(\frac{s_{t-1} - k^{-1}}{\rho_{t-1}}\right) - \Phi\left(\frac{s_{t-1} + k^{-1}}{\rho_{t-1}}\right) \right) + \rho_{t-1} \left( \phi\left(\frac{s_{t-1} - k^{-1}}{\rho_{t-1}}\right) - \phi\left(\frac{s_{t-1} + k^{-1}}{\rho_{t-1}}\right) \right) \right].$$

- If  $\rho_{t-1}^2 = c_t > 0$  which is only possible at  $t = T$  by the requirement of Algorithm 1, then  $x_t = f_{s_{t-1}, \rho_{t-1}, h}(s_{t-1})$ . The closed form of  $f_{\mu,\sigma,h}$  is derived in Proposition 18.
- If  $\rho_{t-1}^2 > c_t > 0$ , then it suffices to apply the Gaussian integral formulas from Lemma H.1 and H.2. The final result is the closed form of  $x_t$  based on Owen's T function.

**Remark H.1** (Effective learning rate). *Intuition can be obtained by reasoning about the effective learning rate of Algorithm 1. Suppose  $\rho_t^2 > c_{t+1} = 0$  for some  $t \in [0 : T - 1]$ , giving*

$$x_{t+1} = 1 - \Phi\left(\frac{s_t - k^{-1}}{\rho_t}\right) - \Phi\left(\frac{s_t + k^{-1}}{\rho_t}\right) \\ + k \left[ s_t \left( \Phi\left(\frac{s_t - k^{-1}}{\rho_t}\right) - \Phi\left(\frac{s_t + k^{-1}}{\rho_t}\right) \right) + \rho_t \left( \phi\left(\frac{s_t - k^{-1}}{\rho_t}\right) - \phi\left(\frac{s_t + k^{-1}}{\rho_t}\right) \right) \right].$$

Denote the RHS as  $\tilde{x}_{t+1}$ , the output of the  $c_t$ -independent analogue of Algorithm 1. Taking the derivative with respect to  $s_t$  gives

$$\frac{\partial \tilde{x}_{t+1}}{\partial s_t} = -\frac{2}{\rho_t} \phi\left(\frac{s_t - k^{-1}}{\rho_t}\right) + k \left[ \Phi\left(\frac{s_t - k^{-1}}{\rho_t}\right) - \Phi\left(\frac{s_t + k^{-1}}{\rho_t}\right) \right]$$

$$+ k \left[ \frac{s_t}{\rho_t} \left( \phi \left( \frac{s_t - k^{-1}}{\rho_t} \right) - \phi \left( \frac{s_t + k^{-1}}{\rho_t} \right) \right) - \frac{s_t - k^{-1}}{\rho_t} \phi \left( \frac{s_t - k^{-1}}{\rho_t} \right) + \frac{s_t + k^{-1}}{\rho_t} \phi \left( \frac{s_t + k^{-1}}{\rho_t} \right) \right].$$

Furthermore,  $\tilde{x}_{t+1} = 0$  when  $s_t = 0$ .

Now suppose  $s_{t-1} = 0$  therefore  $s_t = g_t$  and  $\tilde{x}_t = 0$ . Based on the above, we linearize  $\tilde{x}_{t+1}$  with respect to  $g_t$  (also  $s_t$ , as  $s_t = g_t$ ) and obtain

$$\begin{aligned} \tilde{x}_{t+1} - \tilde{x}_t &= \tilde{x}_{t+1} \\ &= \tilde{x}_{t+1} \Big|_{s_t=0} + \frac{\partial \tilde{x}_{t+1}}{\partial s_t} \Big|_{s_t=0} \cdot g_t + o(g_t) \\ &= k \left[ \Phi \left( -\frac{1}{k\rho_t} \right) - \Phi \left( \frac{1}{k\rho_t} \right) \right] + o(g_t) \\ &= -k \cdot \operatorname{erf} \left( \frac{1}{\sqrt{2}k\rho_t} \right) g_t + o(g_t). \end{aligned}$$

The absolute value of  $g_t$ 's coefficient can be regarded as the effective learning rate.

*Proof of Proposition 18.* The proof generalizes that of Proposition 17. Again it is divided into two parts analyzing  $f_{\mu,\sigma,h}$  and  $x_t$  separately.

**Part 1** Consider  $f_{\mu,\sigma,h}$ . Starting from the representation Eq.(9) of the solution of Stein equation, for all  $x < -k^{-1}$ ,

$$\begin{aligned} f_{\mu,\sigma,h}(x) &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left[ \int_{-\infty}^x (-z + \mu) \phi_{\mu,\sigma}(z) dz - \left( \frac{1}{2k} + \mu + \mathbb{E}_Z[h(\mu + \sigma Z)] \right) \Phi_{\mu,\sigma}(x) \right] \\ &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left[ \sigma^2 \phi_{\mu,\sigma}(x) - \left( \frac{1}{2k} + \mu + \mathbb{E}_Z[h(\mu + \sigma Z)] \right) \Phi_{\mu,\sigma}(x) \right] \\ &= 1 - \frac{\mu + \mathbb{E}_Z[h(\mu + \sigma Z)] + (2k)^{-1}}{\sigma^2} \frac{\Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}. \end{aligned}$$

Similarly, for all  $x > k^{-1}$ ,

$$\begin{aligned} f_{\mu,\sigma,h}(x) &= -\frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left[ \int_x^\infty (z - \mu) \phi_{\mu,\sigma}(z) dz - \left( \frac{1}{2k} - \mu + \mathbb{E}_Z[h(\mu + \sigma Z)] \right) (1 - \Phi_{\mu,\sigma}(x)) \right] \\ &= -\frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left[ \sigma^2 \phi_{\mu,\sigma}(x) - \left( \frac{1}{2k} - \mu + \mathbb{E}_Z[h(\mu + \sigma Z)] \right) (1 - \Phi_{\mu,\sigma}(x)) \right] \\ &= -1 - \frac{\mu - \mathbb{E}_Z[h(\mu + \sigma Z)] - (2k)^{-1}}{\sigma^2} \frac{1 - \Phi_{\mu,\sigma}(x)}{\phi_{\mu,\sigma}(x)}. \end{aligned}$$

As for the regime in between: for all  $x \in [-k^{-1}, k^{-1}]$ ,

$$\begin{aligned} f_{\mu,\sigma,h}(x) &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left[ \int_{-\infty}^{-k^{-1}} \left( -z - \frac{1}{2k} \right) \phi_{\mu,\sigma}(z) dz + \int_{-k^{-1}}^x \frac{k}{2} z^2 \phi_{\mu,\sigma}(z) dz - \mathbb{E}_Z[h(\mu + \sigma Z)] \Phi_{\mu,\sigma}(x) \right] \\ &= \frac{1}{\sigma^2 \phi_{\mu,\sigma}(x)} \left[ \sigma^2 \phi_{\mu,\sigma}(-k^{-1}) - \left( \mu + \frac{1}{2k} \right) \Phi_{\mu,\sigma}(-k^{-1}) + \frac{k}{2} \int_{-k^{-1}}^x z^2 \phi_{\mu,\sigma}(z) dz - \mathbb{E}_Z[h(\mu + \sigma Z)] \Phi_{\mu,\sigma}(x) \right], \end{aligned}$$

where

$$\int_{-k^{-1}}^x z^2 \phi_{\mu,\sigma}(z) dz = [(\mu^2 + \sigma^2) \Phi_{\mu,\sigma}(z) - \sigma^2 (z + \mu) \phi_{\mu,\sigma}(z)] \Big|_{z=-k^{-1}}^x.$$

Plugging it back,

$$\begin{aligned} f_{\mu,\sigma,h}(x) &= \frac{1}{2\sigma^2 \phi_{\mu,\sigma}(x)} \left[ (k\mu^2 + k\sigma^2 - 2\mathbb{E}_Z[h(\mu + \sigma Z)]) \Phi_{\mu,\sigma}(x) - k\sigma^2 (x + \mu) \phi_{\mu,\sigma}(x) \right. \\ &\quad \left. - (k\mu^2 + k\sigma^2 + 2\mu + k^{-1}) \Phi_{\mu,\sigma}(-k^{-1}) + \sigma^2 (1 + k\mu) \phi_{\mu,\sigma}(-k^{-1}) \right]. \end{aligned}$$

**Part 2** Consider  $x_t$ . By the same substitution  $\mu \leftarrow s_{t-1}$ ,  $\sigma \leftarrow \rho_{t-1} > 0$  and  $c \leftarrow c_t > 0$  as in Proposition 17,  $x_t$  can be reformulated as

$$x_t = -\frac{1}{2} \int_0^1 \frac{1}{\sqrt{\tau}} \mathbb{E}_Z \left[ h' \left( \mu + \sqrt{\sigma^2 - \tau c} Z \right) \right] d\tau,$$

where

$$\begin{aligned} \mathbb{E}_Z \left[ h' \left( \mu + \sqrt{\sigma^2 - \tau c} Z \right) \right] &= - \int_{-\infty}^{\frac{-k^{-1}-\mu}{\sqrt{\sigma^2-\tau c}}} \phi(z) dz + \int_{\frac{k^{-1}-\mu}{\sqrt{\sigma^2-\tau c}}}^{\infty} \phi(z) dz + k \int_{\frac{-k^{-1}-\mu}{\sqrt{\sigma^2-\tau c}}}^{\frac{k^{-1}-\mu}{\sqrt{\sigma^2-\tau c}}} \left( \mu + \sqrt{\sigma^2 - \tau c} z \right) \phi(z) dz \\ &= -1 + (1 - k\mu) \Phi \left( \frac{-k^{-1} + \mu}{\sqrt{\sigma^2 - \tau c}} \right) + (1 + k\mu) \Phi \left( \frac{k^{-1} + \mu}{\sqrt{\sigma^2 - \tau c}} \right) \\ &\quad - k\sqrt{\sigma^2 - \tau c} \left[ \phi \left( \frac{-k^{-1} + \mu}{\sqrt{\sigma^2 - \tau c}} \right) - \phi \left( \frac{k^{-1} + \mu}{\sqrt{\sigma^2 - \tau c}} \right) \right]. \end{aligned} \quad \square$$

### H.3 Auxiliary Lemmas

The above derivation uses the following two Gaussian integral lemmas, derived by somewhat complicated but mechanical calculations.

**Lemma H.1.** For all  $\mu \in \mathbb{R}$ ,  $\sigma \in \mathbb{R}_{>0}$  and  $c \in (0, \sigma^2)$ ,

$$\begin{aligned} &1 - \int_0^1 \frac{1}{\sqrt{\tau}} \Phi \left( \frac{\mu}{\sqrt{\sigma^2 - \tau c}} \right) d\tau \\ &= 1 - 2\Phi \left( \frac{\mu}{\sqrt{\sigma^2 - c}} \right) - \frac{\sqrt{2\pi}\sigma}{\sqrt{c}} \phi \left( \frac{\mu}{\sigma} \right) \left[ 1 - 2\Phi \left( \frac{\mu}{\sigma} \sqrt{\frac{c}{\sigma^2 - c}} \right) \right] - \frac{2\sqrt{2\pi}\mu}{\sqrt{c}} T \left( \frac{\mu}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}} \right). \end{aligned}$$

*Proof of Lemma H.1.* Denote the LHS of the objective equality as  $f(\mu)$ . Taking the derivative gives

$$\begin{aligned} f'(u) &= -\frac{1}{\sqrt{2\pi}} \int_0^1 \frac{1}{\sqrt{\tau(\sigma^2 - \tau c)}} \exp \left( -\frac{\mu^2}{2(\sigma^2 - \tau c)} \right) d\tau \\ &= -\frac{2}{\sqrt{2\pi c}} \int_0^{\sqrt{\frac{c}{\sigma^2 - c}}} \frac{1}{1 + \tau^2} \exp \left( -\frac{\mu^2(1 + \tau^2)}{2\sigma^2} \right) d\tau \quad (\tau \leftarrow \frac{\tau^2 \sigma^2}{(1 + \tau^2)c}) \\ &= -\frac{2\sqrt{2\pi}}{\sqrt{c}} T \left( \frac{\mu}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}} \right), \quad (T(x, y) := \frac{1}{2\pi} \int_0^y \frac{e^{-\frac{1}{2}x^2(1+z^2)}}{1+z^2} dz) \end{aligned}$$

where the last line uses the definition of Owen's T function. It is known that [Owe80, Eq.(c00,000.1) in Table I]

$$\int T(x, y) dx = xT(x, y) + \frac{y}{\sqrt{2\pi(1+y^2)}} \Phi \left( x\sqrt{1+y^2} \right) - \phi(x)\Phi(yx) + \frac{1}{2}\phi(x).$$

Therefore, since  $f(0) = 0$ ,

$$\begin{aligned} f(\mu) &= -\frac{2\sqrt{2\pi}}{\sqrt{c}} \int_0^\mu T \left( \frac{x}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}} \right) dx \\ &= -\frac{2\sqrt{2\pi}\sigma}{\sqrt{c}} \left[ xT \left( x, \sqrt{\frac{c}{\sigma^2 - c}} \right) + \frac{\sqrt{c}}{\sqrt{2\pi}\sigma} \Phi \left( \frac{\sigma x}{\sqrt{\sigma^2 - c}} \right) - \phi(x)\Phi \left( \frac{\sqrt{c}x}{\sqrt{\sigma^2 - c}} \right) + \frac{1}{2}\phi(x) \right] \Big|_{x=0}^{\frac{\mu}{\sigma}} \\ &= 1 - 2\Phi \left( \frac{\mu}{\sqrt{\sigma^2 - c}} \right) - \frac{\sqrt{2\pi}\sigma}{\sqrt{c}} \phi \left( \frac{\mu}{\sigma} \right) \left[ 1 - 2\Phi \left( \frac{\mu}{\sigma} \sqrt{\frac{c}{\sigma^2 - c}} \right) \right] - \frac{2\sqrt{2\pi}\mu}{\sqrt{c}} T \left( \frac{\mu}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}} \right). \end{aligned} \quad \square$$

**Lemma H.2.** For all  $\mu \in \mathbb{R}$ ,  $\sigma \in \mathbb{R}_{>0}$  and  $c \in (0, \sigma^2)$ ,

$$\begin{aligned} &\int_0^1 \frac{\sqrt{\sigma^2 - \tau c}}{\sqrt{\tau}} \phi \left( \frac{\mu}{\sqrt{\sigma^2 - \tau c}} \right) d\tau \\ &= \sqrt{\sigma^2 - c} \cdot \phi \left( \frac{\mu}{\sqrt{\sigma^2 - c}} \right) + \frac{\sqrt{2\pi}\mu\sigma}{\sqrt{c}} \phi \left( \frac{\mu}{\sigma} \right) \left[ \Phi \left( \frac{\mu}{\sigma} \sqrt{\frac{c}{\sigma^2 - c}} \right) - \frac{1}{2} \right] + \frac{\sqrt{2\pi}}{\sqrt{c}} (\sigma^2 - \mu^2) T \left( \frac{\mu}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}} \right). \end{aligned}$$



*Proof of Lemma H.2.* Similar to the first step of Lemma H.1, with the change of variable  $\tau \leftarrow \frac{\tau^2 \sigma^2}{(1+\tau^2)c}$ , the LHS of the objective equality can be converted as

$$\int_0^1 \frac{\sqrt{\sigma^2 - \tau c}}{\sqrt{\tau}} \phi\left(\frac{\mu}{\sqrt{\sigma^2 - \tau c}}\right) d\tau = \frac{2\sigma^2}{\sqrt{2\pi c}} \int_0^{\sqrt{\frac{c}{\sigma^2 - c}}} \frac{1}{(1+\tau^2)^2} \exp\left(-\frac{\mu^2(1+\tau^2)}{2\sigma^2}\right) d\tau.$$

For brevity, denote the obtained RHS as  $\diamond$ . The objective now is to further simplify this term, specifically relating it to Owen's T function,  $T(x, y) := \frac{1}{2\pi} \int_0^y \frac{e^{-\frac{1}{2}x^2(1+z^2)}}{1+z^2} dz$ .

For notational convenience, define  $v = \frac{\mu^2}{2\sigma^2} \geq 0$ . Taking the following derivative yields

$$\begin{aligned} \frac{d}{d\tau} \left[ \frac{\tau}{1+\tau^2} \exp(-v(1+\tau^2)) \right] &= \left[ \frac{1-2v\tau^2}{1+\tau^2} - \frac{2\tau^2}{(1+\tau^2)^2} \right] \exp(-v(1+\tau^2)) \\ &= \left[ -\frac{1+2v\tau^2}{1+\tau^2} + \frac{2}{(1+\tau^2)^2} \right] \exp(-v(1+\tau^2)) \\ &= \left[ -2v + \frac{2v-1}{1+\tau^2} + \frac{2}{(1+\tau^2)^2} \right] \exp(-v(1+\tau^2)), \end{aligned}$$

therefore

$$\frac{\exp(-v(1+\tau^2))}{(1+\tau^2)^2} = \frac{1}{2} \frac{d}{d\tau} \left[ \frac{\tau}{1+\tau^2} \exp(-v(1+\tau^2)) \right] + v \exp(-v(1+\tau^2)) + \left( \frac{1}{2} - v \right) \frac{\exp(-v(1+\tau^2))}{1+\tau^2}.$$

Integrating it with respect to  $\tau$  gives

$$\begin{aligned} &\int_0^y \frac{\exp(-v(1+\tau^2))}{(1+\tau^2)^2} d\tau \\ &= \frac{1}{2} \frac{\tau}{1+\tau^2} \exp(-v(1+\tau^2)) \Big|_{\tau=0}^y + v e^{-v} \int_0^y \exp(-v\tau^2) d\tau + \left( \frac{1}{2} - v \right) \int_0^y \frac{\exp(-v(1+\tau^2))}{1+\tau^2} d\tau \\ &= \frac{1}{2} \frac{y}{1+y^2} \exp(-v(1+y^2)) + \sqrt{\pi v} e^{-v} \left[ \Phi(\sqrt{2v}y) - \frac{1}{2} \right] + 2\pi \left( \frac{1}{2} - v \right) T(\sqrt{2v}, y). \end{aligned}$$

Specializing it to  $\diamond$ ,

$$\begin{aligned} &\diamond \\ &= \frac{2\sigma^2}{\sqrt{2\pi c}} \left[ \frac{1}{2} \frac{y}{1+y^2} \exp(-v(1+y^2)) + \sqrt{\pi v} e^{-v} \left[ \Phi(\sqrt{2v}y) - \frac{1}{2} \right] + 2\pi \left( \frac{1}{2} - v \right) T(\sqrt{2v}, y) \right] \Big|_{y=\sqrt{\frac{c}{\sigma^2 - c}}, v=\frac{\mu^2}{2\sigma^2}} \\ &= \frac{\sqrt{\sigma^2 - c}}{\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2(\sigma^2 - c)}\right) + \frac{\mu\sigma}{\sqrt{c}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \left[ \Phi\left(\frac{\mu}{\sigma} \sqrt{\frac{c}{\sigma^2 - c}}\right) - \frac{1}{2} \right] \\ &\quad + \frac{\sqrt{2\pi}}{\sqrt{c}} (\sigma^2 - \mu^2) T\left(\frac{\mu}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}}\right) \\ &= \sqrt{\sigma^2 - c} \cdot \phi\left(\frac{\mu}{\sqrt{\sigma^2 - c}}\right) + \frac{\sqrt{2\pi}\mu\sigma}{\sqrt{c}} \phi\left(\frac{\mu}{\sigma}\right) \left[ \Phi\left(\frac{\mu}{\sigma} \sqrt{\frac{c}{\sigma^2 - c}}\right) - \frac{1}{2} \right] + \frac{\sqrt{2\pi}}{\sqrt{c}} (\sigma^2 - \mu^2) T\left(\frac{\mu}{\sigma}, \sqrt{\frac{c}{\sigma^2 - c}}\right). \quad \square \end{aligned}$$