

Sharp Minimax Regret for General Classes with Unbounded Weights

Wojciech Szpankowski

Abstract

In online learning, a learner sequentially predicts labels over T rounds, incurring loss at each step. The regret measures the learner's total loss over T rounds relative to the best expert from a given class. In this paper, we study the minimax regret with logarithmic loss for a general class of experts including the probit function, under mild regularity conditions, but with unbounded d -dimensional weights, a regime that has received comparatively little attention in the literature. The minimax regret with logarithmic loss is fully characterized by the Shtarkov sum, a fundamental complexity measure that we analyze precisely using tools from analytic combinatorics. Our main results provide a sharp characterization of the average Shtarkov sum and, consequently, of the average minimax regret with unbounded weights, up to second-order asymptotics. For weights bounded by $R \leq T$ the average minimax regret grows as $\frac{d}{2} \log T + \frac{d-2}{2} \log R + O(\log \log T)$. In particular, for $R = T^\alpha$ the leading term scales linearly with α : the regret grows as $(\frac{d}{2}(1+\alpha) - \alpha) \log T$ for $\alpha \leq 1$, and undergoes a phase transition at $\alpha = 1$, beyond which the regret approaches $d \log T$, independent of α . This solves the open problem of [Drmota et al. \(2025\)](#), who derived the regret only for logistic regression and $\alpha \leq 1/4$.

1 Introduction

Sequential prediction under logarithmic loss lies at the junction of online learning and universal coding: the minimax regret against a class of experts equals the logarithm of the Shtarkov sum ([Shtarkov, 1987](#)), and for smooth d -parameter classes over a bounded parameter set it grows as $\frac{d}{2} \log \frac{T}{2\pi} + \log \int_R \sqrt{\det I(w)} dw + o(1)$ ([Rissanen, 1996](#); [Clarke and Barron, 1990](#)) (see also [Jacquet and Szpankowski, 2004](#)). This classical picture breaks down when the parameters are unbounded. For generalized linear experts $p(\langle w, x \rangle)$ with $\|w\| \leq R$, the Fisher-information integral diverges as $R \rightarrow \infty$ (for $d \geq 2$). The regret is then known to move from $\frac{d}{2} \log T$ for bounded weights to $d \log T$ for $R = \infty$ ([Drmota et al., 2026](#)), yet how it moves between the two regimes, that is, its second-order dependence on R , has so far been characterized only for logistic regression with $R = O(T^{1/4})$ ([Drmota et al., 2025](#)). In this paper we describe the transition for all $R = O(T/(\log T)^{4.5})$ and for a broad class of experts, including Gaussian-type tails such as probit, whose analysis differs substantially from the logistic case. We show that the expected fixed-design regret equals $\frac{d}{2} \log T + \frac{d-2}{2} \log R + O(\log \log T)$. Hence for $R = T^\alpha$ the growth rate $\frac{d}{2}(1+\alpha) - \alpha$ increases linearly in α over the whole range $0 < \alpha < 1$, which solves the open problem of [Drmota et al. \(2025\)](#).

To set the stage for our discussion, we recall that the online learning problem can be described as a game between nature/environment and a learner/predictor. Broadly, the objective of the learner is to process past observations to predict the next realization of nature's labeling sequence. At each round $t \in \mathbb{N}$, the learner receives a d -dimensional data/feature vector $x_t \in \mathbb{R}^d$ to make a prediction $\hat{y}_t \in [0, 1]$ of the true label $y_t \in \{-1, 1\}$. Once a prediction is made, nature reveals the true label y_t , and the learner incurs some loss evaluated based on a predefined function $\ell : \hat{\mathcal{Y}} \times \mathcal{Y} \rightarrow \mathbb{R}^+$ where $\hat{\mathcal{Y}} = [0, 1]$ and $\mathcal{Y} = \{-1, 1\}$ are the prediction and label domains, respectively. In regret analysis,

we are interested in comparing the accumulated loss of the learner with that of the best strategy within a predefined class $\mathcal{H}$ of expert functions $h : \mathbb{R}^d \mapsto \hat{\mathcal{Y}}$. After T rounds, the pointwise regret is

$$R(\hat{y}^T, y^T, \mathcal{H} \mid x^T) = \sum_{t=1}^T \ell(\hat{y}_t, y_t) - \inf_{h \in \mathcal{H}} \sum_{t=1}^T \ell(h(x_t), y_t),$$

where $\hat{y}_t$ is the prediction based on prior observations y^{t-1} and x^t . Throughout, we write $y^t = (y_1, \dots, y_t) \in \{-1, 1\}^t$ and $x^t = (x_1, \dots, x_t)$ for $t \in [T]$.

In this paper, we focus specifically on the logarithmic loss and consider the general hypothesis class

$$\mathcal{H}_p = \{h_w : w \in \mathbb{R}^d\}, \quad h_w(x) := p(\langle w, x \rangle), \quad x \in \mathbb{R}^d, \quad (1)$$

where p is a function $\mathbb{R} \rightarrow [0, 1]$ and $\langle w, x \rangle$ is the scalar product of x and w . We do assume that p satisfies some mild conditions discussed in the next section. For example, logistic regression $p(w) = (1 + e^{-w})^{-1}$ or the probit function $p(\langle w, x \rangle) = \Phi(\langle w, x \rangle)$, where Φ is the CDF of the standard normal distribution, belong to $\mathcal{H}_p$. However, we do not bound the weights $w \in \mathbb{R}^d$, and this seems to have rarely been analyzed in depth, with the exception of [Drmotá et al. \(2026, 2025\)](#); [Qian et al. \(2024\)](#). Specifically, we assume that $\|w\| \leq R \leq \infty$, where R can grow with T . We are interested in the fixed design regret $r_T^R(\mathcal{H} \mid x^T)$ and its average $\bar{r}_T(\mathcal{H})$, where the feature vector x^T is known in advance. Specifically, for any given $\mathcal{H}$ and x^T we define

$$r_T^R(\mathcal{H} \mid x^T) := \inf_{\hat{y}^T} \sup_{y^T} R(\hat{y}^T, y^T, \mathcal{H} \mid x^T), \quad \bar{r}_T(\mathcal{H}) := \mathbb{E}_{x^T} [r_T(\mathcal{H} \mid x^T)]. \quad (2)$$

This notion of the fixed design regret is also known in the literature as transductive online learning ([Cesa-Bianchi and Shamir, 2011](#)). We often write $r_T(x^T) := r_T^R(\mathcal{H} \mid x^T)$.

It is well known ([Jacquet et al., 2021](#); [Shamir and Szpankowski, 2021](#)) that the minimax regret $r_T(x^T)$ can be studied through the so-called Shtarkov sum

$$S_R(x^T) = \sum_{y^T \in \{-1, 1\}^T} \sup_{\|w\| \leq R} P(y^T \mid x^T, w), \quad (3)$$

where $P(y^T \mid x^T, w) = \prod_{t=1}^T p(y_t \langle x_t, w \rangle)$, and then $r_T^R(x^T) = \log S_R(x^T)$. However, for technical reasons, to study second-order asymptotics we need to work with the regularized Shtarkov sum and its corresponding regularized regret defined in [Drmotá et al. \(2025\)](#) as

$$S_\varepsilon(x^T) = \sum_{y^T} \sup_w P(y^T \mid x^T, w) e^{-\varepsilon \|w\|^2}, \quad r_T^\varepsilon(x^T) = \log S_\varepsilon(x^T), \quad (4)$$

where $\varepsilon > 0$. We apply a multi-dimensional Fourier analysis to extract precise asymptotics of the regularized Shtarkov sum for any $\varepsilon \geq (\log T)^9/T^2$. Furthermore, we use Lemma 1 from [Drmotá et al. \(2025\)](#) that connects the regularized regret $r_T^\varepsilon(x^T)$ and the standard regret $r_T(x^T)$, that is,

$$r_T^{\varepsilon_1}(x^T) \leq r_T^R(x^T) \leq r_T^{\varepsilon_2}(x^T), \quad \varepsilon_1 = \frac{2d \log T}{R^2}, \quad \varepsilon_2 = \frac{1}{R^2}. \quad (5)$$

This demonstrates that the standard and regularized regrets grow at the same asymptotic rate for $R \leq T$. In short, precise asymptotics of the regularized regret allow us to extract second-order terms of the standard regret. That is the route we take in this paper.

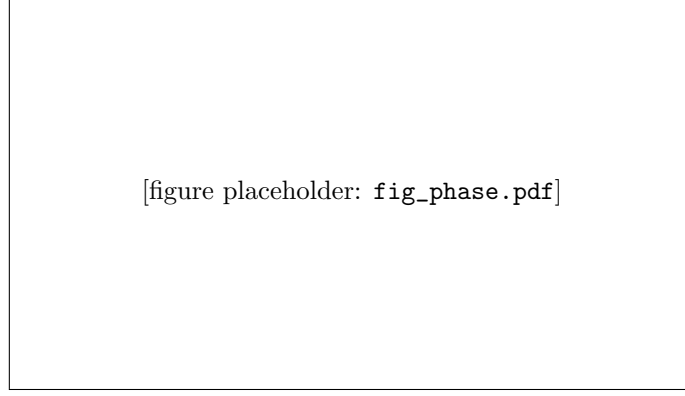


Figure 1: Phase transition of the regret growth constant β in $r_T \sim \beta d \log T$ for $R = T^\alpha$. In this paper we establish the red curve $\beta = (\alpha + 1)/2 - \alpha/d$ in the range $\alpha \in [0, 1]$ for a general class $\mathcal{H}$ with bounded Hessian. The lower thin green segment for $\alpha \in [0, 0.25]$ for logistic regression was proved in [Drmota et al. \(2025\)](#). The blue horizontal line for $\alpha \geq 1$ was derived in [Drmota et al. \(2026\)](#) for logistic regression.

Summary of contributions. In this paper, we present precise second-order asymptotics for the average Shtarkov sum and the average minimax regret for a large class of logistic-like and probit-like functions. Our contributions are threefold:

1. For a general class $\mathcal{H}$ of functions we derive precise second-order asymptotics of the average regularized Shtarkov sum for $\varepsilon = \Omega((\log T)^9/T^2)$ of the form

$$\mathbb{E}[S_\varepsilon(x^T)] = \left(\frac{T}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} \sqrt{\det(\overline{B}(w))} e^{-\varepsilon\|w\|^2} dw (1 + O(T^{-\beta(d)}))$$

for some $\beta(d) > 0$, where $\overline{B}(w)$ is a $d \times d$ matrix that depends on $p(\langle x, w \rangle)$ belonging to the class $\mathcal{H}$. We also obtain the corresponding expansion for the second moment $\mathbb{E}[S_\varepsilon(x^T)^2]$, which allows us to show that $\mathbb{E}[\log S_\varepsilon(x^T)] \sim \log \mathbb{E}[S_\varepsilon(x^T)]$.

2. By using a spectral decomposition of $\overline{B}(w)$ and corresponding asymptotics of the eigenvalues we obtain asymptotics of $\det(\overline{B}(w))$ of the form $\sim C\|w\|^{-d-2}$ (for some $C > 0$), which implies asymptotics of $\mathbb{E}[S_\varepsilon(x^T)]$ in terms of T and ε . Due to the relation $\mathbb{E}[\log S_\varepsilon(x^T)] \sim \log \mathbb{E}[S_\varepsilon(x^T)]$ this implies (for $\varepsilon = \Omega((\log T)^9/T^2)$)

$$r_T^\varepsilon = \mathbb{E}[\log S_\varepsilon(x^T)] = \frac{d}{2} \log T + \frac{d-2}{4} \log \frac{1}{\varepsilon} + \log C_d(p) + O(T^{-\beta(d)})$$

for some $\beta(d) > 0$ and some $C_d(p)$ that depends on $p \in \mathcal{H}$ and hence on the structure of the hypothesis class $\mathcal{H}$.

3. By applying Lemma 1 from [Drmota et al. \(2025\)](#) and the asymptotic formula for r_T^ε we obtain the corresponding second-order terms for the average regret as a function of R , namely

$$\frac{d}{2} \log T + \frac{d-2}{2} \log R - \frac{d-2}{2} \log(2d \log T) + \log C_d(p) \leq r_T^R \leq \frac{d}{2} \log T + \frac{d-2}{2} \log R + \log C_d(p)$$

for $R = O(T/(\log T)^{4.5})$. Observe that even for the logistic regression class, we extend its asymptotics from $R = O(T^{1/4})$ to $R = O(T/(\log T)^{4.5})$ (see also Figure 1). This solves the

open problem posed in [Drmota et al. \(2025\)](#). In particular, for $R = T^\alpha$ with $0 < \alpha < 1$ we obtain that the average regret grows linearly with α , that is, for $\alpha \leq 1$,

$$\bar{r}_T^R = \left(\frac{d}{2}(1 + \alpha) - \alpha \right) \log T + O(\log \log T).$$

Previously, in [Drmota et al. \(2025\)](#) this expression was known only for logistic regression and $\alpha \leq 1/4$, as graphically represented in [Figure 1](#). We recall that for $T \leq R \leq \infty$ the behaviour of the expected regret stays about $d \log T$ ([Drmota et al., 2026](#)), so we cover (almost) the whole range (just excluding the change of behaviour around $R \approx T$).

Related work. In universal coding, the minimax regret under logarithmic loss is given by the Shtarkov sum ([Shtarkov, 1987](#)). For smooth parametric classes over a bounded parameter set it admits the precise expansion of [Rissanen \(1996\)](#); [Clarke and Barron \(1990\)](#), and analytic-combinatorics refinements go back to [Drmota and Szpankowski \(2004\)](#). Logistic-type experts have been studied extensively under various formulations of regret, mostly with bounded weights ([Kakade and Ng, 2005](#); [McMahan and Streeter, 2012](#); [Hazan et al., 2014](#); [Rakhlin and Sridharan, 2015](#); [Foster et al., 2018](#); [Shamir, 2020](#); [Bilodeau et al., 2020](#); [Jézéquel et al., 2020](#); [Agarwal et al., 2022](#); [Wu et al., 2022, 2026](#)); see [Cesa-Bianchi and Lugosi \(2006\)](#) for background. For the sequential regret, where both x^T and y^T are chosen adaptively, the minimax regret is at most $\frac{d}{2} \log(TR^2/d)$ for any weight range R ([Foster et al., 2018](#); [Shamir, 2020](#); [Wu et al., 2022](#)). This bound becomes vacuous as $R \rightarrow \infty$. Unbounded comparators are also central in parameter-free and scale-free online learning ([Orabona and Pál, 2016](#); [Mayo et al., 2022](#)), but not at the precision considered here. For the fixed design (transductive) regret ([Kakade and Kalai, 2005](#)), lower bounds of order $\frac{d}{2} \log(T/d^2)$, independent of R , are known ([Shamir, 2020](#); [Wu et al., 2022](#); [Mayo et al., 2022](#)); related expected-regret notions are studied in [Wu et al. \(2023\)](#). Recently, [Drmota et al. \(2026\)](#) showed that the fixed design regret is at most $2d \log T$ for a general monotone class even with $R = \infty$, and equals $(1 + o(1))d \log(T/d)$ for logistic regression. The transition from $\frac{d}{2} \log T$ to $d \log T$ as R grows has been characterized precisely only in [Drmota et al. \(2025\)](#), for logistic regression and $R = O(T^{1/4})$. Unbounded weights were also treated in [Qian et al. \(2024\)](#), though without precise asymptotics. To the best of our knowledge, [Drmota et al. \(2025, 2026\)](#) are the only works giving precise fixed design regret with unbounded weights. We extend these results to a general class of functions, including probit, and to the full range $R = O(T/(\log T)^{4.5})$. This extension requires substantially sharper tail estimates for the Fourier representation of the Shtarkov sum.

2 Main result

The main goal of this paper is to estimate the average regret $\bar{r}_T^R$ for $\|w\| < R$. Throughout we assume that the vectors $x_1, \dots, x_T$ are i.i.d. random vectors $X_1, \dots, X_T$ that follow a probability distribution over bounded support. In what follows we assume that X is uniformly distributed in the unit ball. This simplifies several of our computations.

We focus specifically on the logarithmic loss defined over a hypothesis class of experts $\mathcal{H}_p = \{h_w : w \in \mathbb{R}^d\}$, $h_w(x) := p(\langle w, x \rangle)$, $x \in \mathbb{R}^d$. We shall study a general class of functions p with bounded first and second derivatives; for example, logistic regression $p(w) = (1 + e^{-w})^{-1}$ or the probit function $p(w) = \Phi(w)$ (where Φ is the CDF of the standard normal distribution). The logistic case was discussed in [Drmota et al. \(2025\)](#). All necessary calculations for logistic-like functions can also be done in a slightly more general framework:

(H) The function p satisfies $1 - p(w) = p(-w)$ and $-\log p(w)$ is convex with bounded first and second derivatives. Furthermore, $((\log p(u))'u)' \sim -ce^{-\alpha u}u^\beta$ for $u \rightarrow \infty$ for some $c, \alpha > 0$, $\beta \geq 1$, $((\log p(u))'u)' \sim \gamma$ for $u \rightarrow -\infty$ for some $\gamma > 0$, and $((\log p(u))'u)'' \neq 0$ at all zeros of $((\log p(u))'u)'$, which are only finitely many.

Clearly, the logistic functions $p(w) = (1 + e^{-w})^{-1}$ satisfy all assumptions (with $c = \alpha = \beta = \gamma = 1$). Note that the probit case $p(w) = \Phi(w)$ does not satisfy (H). It only satisfies the conditions $1 - p(w) = p(-w)$ and that $-\log p(w)$ is convex. However, the derivatives are not bounded and the other conditions are not satisfied either. Actually the tail behaviour of Φ is completely different from the exponential tail behaviour for p that satisfy (H). This makes the analysis even more challenging and interesting.

The main result of this paper is formulated next. We delay the proof to Section 3 for p that satisfies (H), while the probit case is discussed in Appendix C.

Theorem 1. *Let $w \in \mathbb{R}^d$ and $\varepsilon > 0$. Assume that the features x_t are i.i.d., generated by a uniform distribution over the d -dimensional unit ball B_d . Suppose further that p satisfies (H) or that $p(w) = \Phi(w)$. Then for $1 \leq R \leq O(T/\log^{4.5} T)$ we have for the expected minimax regret*

$$\bar{r}_T^R = \frac{d}{2} \log T + \frac{d-2}{2} \log R + O(\log \log T). \quad (6)$$

The proof follows from two propositions that we state now and prove below and in the appendices. The roadmap of the proof was partially outlined in the introduction: we first analyze the generalized Shtarkov sum $S_\varepsilon(x^T)$ defined in (4) using a multidimensional Fourier analysis, resulting in the following proposition.

Proposition 2. *Under the same assumptions as in Theorem 1, for $\varepsilon = \Omega((\log T)^9/T^2)$ there exists $\beta(d) > 0$ such that*

$$\mathbb{E}[S_\varepsilon(x^T)] = \left(\frac{T}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} \sqrt{\det(\bar{B}(w))} e^{-\varepsilon\|w\|^2} dw (1 + O(T^{-\beta(d)})) \quad (7)$$

with

$$\bar{B}(w) = \frac{1}{\text{Vol}(B_d)} \int_{B_d} \frac{[p'(\langle w, x \rangle)]^2}{p(\langle w, x \rangle)p(-\langle w, x \rangle)} x \otimes x dx, \quad (8)$$

where $x \otimes x = xx^\tau$ is the tensor product of x with itself, τ denoting the transpose. Furthermore (see Appendix B),

$$\mathbb{E}[S_\varepsilon(x^T)^2] = \mathbb{E}[S_\varepsilon(x^T)]^2 (1 + O(T^{-\beta(d)})). \quad (9)$$

Before we proceed, we briefly indicate how Proposition 2 establishes

$$\mathbb{E}[\log S_\varepsilon(X^T)] = \log \mathbb{E}[S_\varepsilon(X^T)] (1 + O(T^{-\beta(d)})). \quad (10)$$

To simplify the notation we set $S = S_\varepsilon(x^T)$. By Jensen's inequality we directly have $\mathbb{E}[\log S] \leq \log \mathbb{E}[S]$. Thus, it is sufficient to obtain a lower bound. Observe that (9), proved in Appendix B, implies $\text{Var}[S] = O((\mathbb{E}[S])^2 T^{-\beta})$. Then Chebyshev's inequality

$$\mathbb{P}\left(\left|\frac{S}{\mathbb{E}[S]} - 1\right| \geq \frac{1}{2}\right) \leq 4 \frac{\text{Var}[S]}{(\mathbb{E}[S])^2} = O(T^{-\beta})$$

can be used to obtain the lower bound $\mathbb{E}[\log S] \geq \log \mathbb{E}[S] (1 + O(T^{-\beta(d)}))$.

The second main result needed to prove Theorem 1 is presented next.

Proposition 3. Suppose that the matrix $\overline{B}(w)$ is given by (8). Then, as $\|w\| \rightarrow \infty$, we have

$$\det(\overline{B}(w)) = A_d(p) \|w\|^{-d-2} (1 + O(\|w\|^{-2})), \quad (11)$$

where

$$A_d(p) = \frac{v_{d-1}}{v_d} \frac{2^d}{(d+2)^{d-1}} q^*(1)^{d-1} q^*(3).$$

Here $v_d = \pi^{d/2} / \Gamma(\frac{d}{2} + 1) = \text{Vol}(B_d)$ denotes the volume of the unit ball B_d and $q^*(s) = \int_0^\infty q(w) w^{s-1} dw$ the Mellin transform of the function $q(w) = p'(w)^2 / (p(w)p(-w))$. For example, for the logistic function we have $A_d(p) = \frac{v_{d-1}}{v_d} \frac{\pi^2}{3(d+2)^{d-1}}$.

Given Propositions 2 and 3 we show below how to establish the following asymptotic properties (which then also lead to Theorem 1).

Proposition 4. Under the assumptions of Theorem 1 we have, as $\varepsilon \rightarrow 0$ but $\varepsilon \gg (\log T)^9 T^{-2}$,

$$\mathbb{E}[S_\varepsilon(x^T)] = \begin{cases} T^{d/2} (C_d(p) \varepsilon^{-\frac{d}{4} + \frac{1}{2}} + O(1)) (1 + O(T^{-\beta(d)})), & d \geq 3, \\ T^{1/2} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty B(w)^{1/2} dw + O(\varepsilon^{1/4}) \right) (1 + O(T^{-\beta(1)})), & d = 1, \\ T \left(\frac{\sqrt{A_2(p)}}{2} \log \varepsilon^{-1} + O(1) \right) (1 + O(T^{-\beta(2)})), & d = 2, \end{cases} \quad (12)$$

where

$$C_d(p) = (2\pi)^{-\frac{d}{2}} \sqrt{A_d(p)} \frac{dv_d}{2} \Gamma\left(\frac{d-2}{4}\right).$$

In particular, for $d \geq 3$ and $\varepsilon = \Theta(T^{-2\alpha})$ with $0 < \alpha < 1$, we have

$$\bar{r}_T^\varepsilon = \mathbb{E}[\log S_\varepsilon(X^T)] = \left(\left(\frac{1}{2} + \frac{\alpha}{2} \right) d - \alpha \right) \log T + O(1). \quad (13)$$

Proof. Proposition 4 follows from Propositions 2 and 3. We only do the calculations for $d \geq 3$; the cases $d = 1$ and $d = 2$ are simple variants. By Proposition 3 we have for $\|w\| \geq 1$ $\det(\overline{B}(w)) = A_d(p) \|w\|^{-d-2} (1 + O(\|w\|^{-2}))$. Consequently we obtain (with $s_d = dv_d$ denoting the surface measure of the d -dimensional unit sphere)

$$\begin{aligned} \int_{\mathbb{R}^d} \sqrt{\det \overline{B}(w)} e^{-\varepsilon \|w\|^2} dw &= \int_{\|w\| \geq 1} \sqrt{\det \overline{B}(w)} e^{-\varepsilon \|w\|^2} dw + O(1) \\ &= \sqrt{A_d(p)} \int_{\|w\| \geq 1} \|w\|^{-\frac{d+2}{2}} (1 + O(\|w\|^{-2})) e^{-\varepsilon \|w\|^2} dw + O(1) \\ &= s_d \sqrt{A_d(p)} \int_1^\infty r^{d-1-\frac{d+2}{2}} (1 + O(r^{-2})) e^{-\varepsilon r^2} dr + O(1) \\ &= \frac{s_d}{2} \sqrt{A_d(p)} \varepsilon^{-\frac{d}{4} + \frac{1}{2}} \int_\varepsilon^\infty v^{\frac{d}{4} - \frac{3}{2}} (1 + O(v^{-1}\varepsilon)) e^{-v} dv + O(1) \\ &= \frac{s_d}{2} \sqrt{A_d(p)} \varepsilon^{-\frac{d}{4} + \frac{1}{2}} \int_0^\infty v^{\frac{d}{4} - \frac{3}{2}} e^{-v} dv + O(1) \\ &= \frac{s_d}{2} \sqrt{A_d(p)} \Gamma\left(\frac{d-2}{4}\right) \varepsilon^{-\frac{d}{4} + \frac{1}{2}} + O(1), \end{aligned}$$

where we have used the substitution $\varepsilon r^2 = v$ and the property that $\int_0^\varepsilon v^{\frac{d}{4} - \frac{3}{2}} e^{-v} dv = O(\varepsilon^{\frac{d}{4} - \frac{1}{2}})$. This establishes (12), and with the help of (10) we directly get (13). $\square$

The proof of Theorem 1 is now immediate. We just use (13), and after setting $\varepsilon = 1/R^2$ and invoking (5) we are done.

3 Analysis and preliminary results

In order to prove Theorem 1 we just have to prove Propositions 2 and 3. Whereas Proposition 3 is a direct generalization of a corresponding property from Drmota et al. (2025), the proof of Proposition 2 requires a lot of preliminaries. In what follows we comment on the main steps that are used to obtain this result in the case where hypothesis (H) is satisfied (here we refer to Drmota et al. (2025) for more details). The probit case ($p(w) = \Phi(w)$) requires an even more technical analysis since the tail of Φ is much smaller than in the case of hypothesis (H). This is detailed in Appendix C.

Note that we significantly extend the results of Drmota et al. (2025). First we extend the logistic case $p(w) = 1/(1 + e^{-w})$ to hypothesis (H). More importantly, we extend the range from $\varepsilon = \Omega(T^{-1/2})$ to $\varepsilon = \Omega((\log T)^9/T^2)$, which fills the gap that was left in Drmota et al. (2025). Finally, we can also handle the probit case.

3.1 Fourier representation of the Shtarkov sum

Our objective is to find precise asymptotics for the generalized Shtarkov sum $S_\varepsilon(x^T)$ defined in (4). Note that for any y^T and x^T we have for any $\varepsilon > 0$

$$P(y^T | x^T, w) = \prod_{t=1}^T p(y_t \langle x_t, w \rangle) \quad \text{and} \quad P_\varepsilon(y^T | x^T, w) = \prod_{t=1}^T p(y_t \langle x_t, w \rangle) e^{-\varepsilon \|w\|^2}. \quad (14)$$

We define $L(y^T | x^T, w) = -\log P(y^T | x^T, w)$ and $L_\varepsilon(y^T | x^T, w) = L(y^T | x^T, w) + \varepsilon \|w\|^2$. Then

$$\nabla L_\varepsilon(y^T | x^T, w) = - \sum_{t=1}^T \frac{p'(y_t \langle x_t, w \rangle)}{p(y_t \langle x_t, w \rangle)} y_t x_t + 2\varepsilon w$$

and

$$\nabla^2 L_\varepsilon(y^T | x^T, w) = \sum_{t=1}^T \frac{[p'(y_t \langle x_t, w \rangle)]^2 - p''(y_t \langle x_t, w \rangle) p(y_t \langle x_t, w \rangle)}{p^2(y_t \langle x_t, w \rangle)} x_t \otimes x_t + 2\varepsilon I, \quad (15)$$

where $x \otimes x$ denotes the matrix $(x_i x_j)_{1 \leq i, j \leq d}$ and I the identity matrix.

To study the Shtarkov sum, we need a better understanding of the optimal w_ε^* defined as $w_\varepsilon^* = \arg \min_{w \in \mathbb{R}^d} L_\varepsilon(y^T | x^T, w)$, which is the (unique) solution of the equation $\nabla L_\varepsilon(y^T | x^T, w) = 0$. Similarly, for every $a \in \mathbb{R}^d$ the equation

$$\nabla L_\varepsilon(y^T | x^T, w) = a \quad (16)$$

has a unique solution $w_\varepsilon^*(a)$. This follows from hypothesis (H), namely from the convexity of $-\log p$ and from the tail estimates. Furthermore, if we denote $G_{y^T|x^T,\varepsilon}(w) := \nabla L_\varepsilon(y^T | x^T, w)$, then $w_\varepsilon^* = G_{y^T|x^T,\varepsilon}^{-1}(0)$.

We analyze the Shtarkov sum via a multidimensional Fourier transform, which we can prove to exist as in Drmota et al. (2025). For convenience we set

$$h_{y^T|x^T}(a) = P(y^T | x^T, G_{y^T|x^T}^{-1}(a)) e^{-\varepsilon \|G_{y^T|x^T}^{-1}(a)\|^2}.$$

By assumption (H) it follows that (for given y^T , x^T , and $\varepsilon > 0$)

$$\limsup_{\|a\| \rightarrow \infty} \frac{\|G_{y^T|x^T}^{-1}(a)\|}{\|a\|} > 0. \quad (17)$$

Actually this also holds in the probit case. Hence, the Fourier transform

$$\tilde{h}_{y^T|x^T}(z) = \int_{\mathbb{R}^d} h_{y^T|x^T}(a) e^{-i\langle a, z \rangle} da = \int_{\mathbb{R}^d} \exp(-L_\varepsilon(y^T | x^T, G_{y^T|x^T, \varepsilon}^{-1}(a))) e^{-i\langle a, z \rangle} da$$

of $h_{y^T|x^T}(a)$ exists and the inverse Fourier transform has an absolutely convergent integral $h_{y^T|x^T}(a) = (2\pi)^{-d} \int_{\mathbb{R}^d} \tilde{h}_{y^T|x^T}(z) e^{i\langle a, z \rangle} dz$. Actually we only need $h_{y^T|x^T}(0)$ since $\sup_{w \in \mathbb{R}^d} P_\varepsilon(y^T | x^T, w) = P_\varepsilon(y^T | x^T, G_{y^T|x^T, \varepsilon}^{-1}(0)) = h_{y^T|x^T}(0)$. Consequently we can evaluate the Shtarkov sum as follows:

$$\begin{aligned} \sup_{w \in \mathbb{R}^d} P_\varepsilon(y^T | x^T, w) &= P_\varepsilon(y^T | x^T, G_{y^T|x^T, \varepsilon}^{-1}(0)) = \exp(-L(y^T | x^T, G_{y^T|x^T, \varepsilon}^{-1}(0))) \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} \tilde{h}_{y^T|x^T}(z) dz = (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h_{y^T|x^T}(a) e^{-i\langle a, z \rangle} da dz \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \exp(-L_\varepsilon(y^T | x^T, G_{y^T|x^T, \varepsilon}^{-1}(a))) e^{-i\langle a, z \rangle} da dz \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} P(y^T | x^T, w) e^{-i\langle \nabla L(y^T | x^T, w), z \rangle} \\ &\quad \times e^{-\varepsilon \|w\|^2 - 2i\varepsilon \langle w, z \rangle} \det(\nabla^2 L_\varepsilon(y^T | x^T, w)) dw dz, \end{aligned}$$

where we have used the substitution $a = G_{y^T|x^T, \varepsilon}(w) = \nabla L_\varepsilon(y^T | x^T, w)$. To complete our derivation, we observe that the determinant $\det(\nabla^2 L_\varepsilon(y^T | x^T, w))$ contains in each of its d^2 entries a sum of the form $\sum_{t=1}^T [\dots]$. In order to proceed further we use the explicit formula for the determinant and expand all these d^2 sums (where we have to use d^2 different indices $t = 1, \dots, T$). Nevertheless each summand of the expanded determinant depends on at most d different y_t 's. Thus, $T - d$ of the y_t 's do not appear in this summand. Suppose that these are the first $T - d$ components. Then, by summing over y^{T-d} we can use the following simplification:

$$\sum_{y^{T-d} \in \{-1, 1\}^{T-d}} P(y^{T-d} | x^{T-d}, w) \exp(-i\langle \nabla L(y^{T-d} | x^{T-d}, w), z \rangle) = \prod_{t=1}^{T-d} f(w, x_t, z),$$

where

$$f(w, x, z) = p(\langle x, w \rangle) \exp\left(-i \frac{p'(\langle x, w \rangle)}{p(\langle x, w \rangle)} \langle x, z \rangle\right) + p(-\langle x, w \rangle) \exp\left(i \frac{p'(-\langle x, w \rangle)}{p(-\langle x, w \rangle)} \langle x, z \rangle\right). \quad (18)$$

For example, for $d = 1$ this leads to the representation

$$\begin{aligned}
S_\varepsilon(x^T) &= \sum_{y^T} \sup_w P(y^T | x^T, w) e^{-\varepsilon w^2} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{y^T} P(y^T | x^T, w) e^{-\varepsilon w^2} e^{-i \frac{\partial}{\partial w} L_\varepsilon(y^T | x^T, w) z} \frac{\partial^2}{\partial w^2} L_\varepsilon(y^T | x^T, w) dw dz \\
&= \sum_{s=1}^T \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{y^T} P(y^T | x^T, w) e^{-\varepsilon w^2} e^{-i \frac{\partial}{\partial w} L_\varepsilon(y^T | x^T, w) z} \left(\frac{p'(y_s x_s w)^2}{p(y_s x_s w)^2} - \frac{p''(y_s x_s w)}{p(y_s x_s w)} \right) x_s^2 dw dz \\
&\quad + 2\varepsilon \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{y^T} P(y^T | x^T, w) e^{-\varepsilon w^2} e^{-i \frac{\partial}{\partial w} L_\varepsilon(y^T | x^T, w) z} dw dz \\
&= \sum_{s=1}^T \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{t \neq s} f(x_t, z, w) \\
&\quad \times p(y_s x_s w) \exp \left(i \frac{p'(y_s x_s w)}{p(y_s x_s w)} x_s z \right) \\
&\quad \times \left(\frac{p'(y_s x_s w)^2}{p(y_s x_s w)^2} - \frac{p''(y_s x_s w)}{p(y_s x_s w)} \right) x_s^2 e^{-\varepsilon w^2} e^{-2i\varepsilon w z} dw dz \\
&\quad + 2\varepsilon \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{1 \leq t \leq T} f(x_t, z, w) e^{-\varepsilon w^2} e^{-2i\varepsilon w z} dw dz.
\end{aligned} \tag{20}$$

Assuming that $X^T = (X_1, \dots, X_T)$ are i.i.d., we find for $d = 1$ the expected Shtarkov sum

$$\begin{aligned}
\mathbb{E}[S_\varepsilon(X^T)] &= \frac{T}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(w, z)^{T-1} B(w, z) e^{-\varepsilon w^2} e^{-2i\varepsilon w z} dw dz \\
&\quad + \frac{2\varepsilon}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(w, z)^T e^{-\varepsilon w^2} e^{-2i\varepsilon w z} dw dz,
\end{aligned} \tag{21}$$

where $f(w, z)$ abbreviates

$$f(w, z) = \mathbb{E}[f(X, w, z)] = \mathbb{E} \left[p(Xw) \exp \left(-i \frac{p'(Xw)}{p(Xw)} Xz \right) + p(-Xw) \exp \left(i \frac{p'(-Xw)}{p(-Xw)} Xz \right) \right]$$

and $B(w, z)$ is given by

$$\begin{aligned}
B(w, z) &= \mathbb{E} \left[p(Xw) \exp \left(-i \frac{p'(Xw)}{p(Xw)} Xz \right) \left(\frac{p'(Xw)^2}{p(Xw)^2} - \frac{p''(Xw)}{p(Xw)} \right) X^2 \right] \\
&\quad + \mathbb{E} \left[p(-Xw) \exp \left(i \frac{p'(-Xw)}{p(-Xw)} Xz \right) \left(\frac{p'(-Xw)^2}{p(-Xw)^2} - \frac{p''(-Xw)}{p(-Xw)} \right) X^2 \right].
\end{aligned}$$

Note that

$$B(w, 0) = B(w) = \mathbb{E} \left[\frac{p'(Xw)^2}{p(Xw)p(-Xw)} X^2 \right]. \tag{22}$$

If $d > 1$, then a more complicated expansion holds. However, the first term is still of the form

$$\frac{T^d}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(w, z)^{T-d} \det(B(w, z)) e^{-\varepsilon \|w\|^2} e^{-2i\varepsilon \langle w, z \rangle} dw dz,$$

where $f(w, z) = \mathbb{E}[f(w, X, z)]$ and $B(w, z)$ is a matrix-valued function with $B(w, 0) = B(w)$. This expression is the main tool that we use to study asymptotically the average minimax regret.

3.2 Roadmap to the proof of Proposition 2

In order to keep the presentation as simple as possible we discuss here only the case $d = 1$ and concentrate on functions p satisfying (H). We also refer to [Drmota et al. \(2025\)](#) for several technical computations. However, as mentioned before, we generalize the results of [Drmota et al. \(2025\)](#) significantly (we cover not only the range $\varepsilon = \Omega(1/\sqrt{T})$ but the range $\varepsilon = \Omega((\log T)^9/T^2)$), including the probit case ($p(w) = \Phi(w)$), which needs even more technical refinements and is discussed in Appendix C.

3.2.1 Convergence issues of the Fourier integral

We want to emphasize that the method described above has some convergence issues. The integral in (20) has to be interpreted as an iterated integral. It is not an absolutely convergent double integral on $\mathbb{R} \times \mathbb{R}$. This comes from the direct use of the integral formula for the inverse Fourier transform. Nevertheless, by taking expectation with respect to X^T we find the integral representation (21), which turns out to be an absolutely convergent integral on $\mathbb{R} \times \mathbb{R}$ —this is actually a kind of miracle.

In order to transfer the asymptotics from $\mathbb{E}[S_\varepsilon(X^T)]$ to $\mathbb{E}[\log S_\varepsilon(X^T)]$ we need the corresponding asymptotic expansion for the second moment $\mathbb{E}[S_\varepsilon(X^T)^2]$ (compare with the comments following Proposition 2). We can proceed formally very similarly to obtain an integral representation for $\mathbb{E}[S_\varepsilon(X^T)^2]$; however, at this stage this miracle does not happen. As discussed in Appendix B, we need another regularizing parameter to cover this case.

3.2.2 Gaussian approximation

The integral (21) contains the factor f^{T-1} . It is therefore natural to consider a local expansion for f and apply a Laplace-like approximation in order to evaluate the integral (21) asymptotically. Actually we have

$$f(w, z) = e^{-\frac{1}{2}z^2 B(w)} \left(1 + O(z^3 \min(1, |w|^{-4})) + O(z^4 \min(1, |w|^{-6})) \right)$$

with $B(w)$ from (22). Furthermore we have the property $B(w) = \Theta(\min(1, |w|^{-3}))$. For example, if we restrict ourselves to the range

$$\mathcal{R} = \{(w, z) \in \mathbb{R} \times \mathbb{R} : |z| \leq \max(1, |w|)^{3/2} T^{-1/2} (\log T)^{1/2}\}, \quad (23)$$

then we directly obtain

$$\begin{aligned} & \frac{T}{2\pi} \iint_{\mathcal{R}} f(w, z)^{T-1} B(w, z) e^{-\varepsilon w^2} e^{-2i\varepsilon w z} dw dz \\ & \sim \frac{T}{2\pi} \int_{-\infty}^{\infty} B(w, 0) e^{-\varepsilon w^2} \int_{|z| \leq \max(1, |w|)^{3/2} T^{-1/2} (\log T)^{1/2}} e^{-\frac{T-1}{2} B(w) z^2} dz dw \\ & \sim \frac{\sqrt{T}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} B(w)^{1/2} e^{-\varepsilon w^2} dw. \end{aligned} \quad (24)$$

This is exactly the asymptotic leading term (from Proposition 2). However, it is a non-trivial task to show that the remaining parts of the integral (21) outside the range $\mathcal{R}$ of (23) are negligible.

3.2.3 Asymptotic properties of $f(z, w)$ and $B(w, z)$

The following lemmas collect some asymptotic properties of $f(z, w)$ and $B(w, z)$ under hypothesis (H). They are used to upper bound the tails in the integral representation (20). Most of the properties have already been stated in Drmota et al. (2025), but only for the logistic case. Proofs of Lemmas 5–8 can be found in Appendix A.

Lemma 5. *For $|w| \leq 1$ we uniformly have $|f(z, w)| \leq \min(1, C/|z|)$ for some constant $C > 0$. Furthermore, for $|w| \geq 1$ we have*

$$|f(z, w)| \leq \min\left(1, \frac{C_1}{\sqrt{|wz|}} + \frac{C_2 e^{|w|}}{|wz|}\right).$$

Lemma 6. *Suppose that $c_1 > 0$ is a given constant. Then there exists $c_2 > 0$ such that*

$$|f(z, w)| \leq 1 - \frac{c_2 \max\{1, \log(|z|/|w|)\}}{|w|}$$

uniformly for $|z| \geq c_1|w|$.

Lemma 7. *Suppose that $c_1 > 0$ is a given constant. Then*

$$f(z, w) = 1 - \frac{k(z/w)}{w} + O\left(\frac{z}{w} e^{-w}\right) \quad (25)$$

uniformly for $|w| \geq 1$ and $|z|/|w| \geq c_1$, where

$$k(v) = \int_{-\infty}^{\infty} p(u) \left(1 - \exp\left(i \frac{p'(u)}{p(u)} uv\right)\right) du$$

satisfies $k(v) = \log v + \log \log v + O(1)$ and $k'(v) = O(|v|^{-1})$ as $|v| \rightarrow \infty$. Furthermore we have the upper bound $\frac{\partial}{\partial z} f(z, w) = O\left(\frac{1}{|zw|}\right)$.

Lemma 8. *We uniformly have*

$$|B(z, w)| = O(\min(1, |w|^{-3})), \quad B(w) = \Theta(\min(1, |w|^{-3})),$$

$$B(z, w) = B(w) + O(z^2 \min(1, |w|^{-5})).$$

More precisely,

$$B(w) = \frac{p'(0)^2}{3p(0)^2} + O(w^2) \quad (w \rightarrow 0), \quad B(w) = \frac{1}{2|w|^3} \int_{-\infty}^{\infty} \frac{v^2 p'(v)^2}{p(v)(1-p(v))} dv + O(w^{-4}) \quad (w \rightarrow \infty).$$

Furthermore, for some constant $c_1 > 0$,

$$\begin{aligned} f(w, z) &= 1 - \Theta(z^2 \min(1, |w|^{-3})) \quad (\text{for } |z| \leq \max(1, c_1|w|)) \\ &= 1 - \frac{z^2}{2} B(w) + O(z^3 \min(1, |w|^{-4})) \\ &= e^{-\frac{1}{2} z^2 B(w)} \left(1 + O(z^3 \min(1, |w|^{-4})) + O(z^4 \min(1, |w|^{-6}))\right). \end{aligned}$$

Finally, suppose that $|z| \geq c_1|w|$ for some $c_1 > 0$. Then we have uniformly $\frac{\partial}{\partial z} B(w, z) = O\left(\frac{1}{w^3 z}\right)$.

3.2.4 Some key estimates

We recall that $\mathbb{E}[S_\varepsilon(X^T)]$ can be represented as $TJ_0 + 2\varepsilon J_1$, where J_0 and J_1 are double integrals that look very similar (see (21)). We have already seen that the range $|z| \leq \max(1, |w|)^{3/2} T^{-1/2} (\log T)^{1/2}$ in the double integral J_0 provides the (proposed) asymptotic leading term for $\mathbb{E}[S_\varepsilon(X^T)]$. The main issue is to provide proper upper bounds for the tails. More precisely, we split up the integrals in (21) into three parts (for the sake of brevity, we only deal with J_0), where we focus on the second and third parts, in which we have to use the upper bounds for $B(w, z)$ and $f(w, z)$ in a proper way.

A: The case $|w| \leq 1$. Here we proceed as in [Drmotá et al. \(2025\)](#).

B: The case $1 \leq |w| \leq \eta T$, where $\eta = \eta(T) = c_6 (\log T)^{-3} \rightarrow 0$ and c_6 is sufficiently large. The range $|z| \leq c_1 |w|$ can again be handled as in [Drmotá et al. \(2025\)](#), so we do not repeat it. However, if $|z| \geq c_1 |w|$ we argue in a more precise way (in order to cover also the range $\varepsilon = \Omega((\log T)^9 / T^2)$ and not only the range $\varepsilon = \Omega(1/\sqrt{T})$).

First suppose that $c_1 |w| \leq z \leq e^{2w^2}$. Here we use the formula (a direct consequence of Lemma 10)

$$\int_I H(z) e^{-2i\varepsilon zw} dz = O\left(\frac{1}{\varepsilon w} \left(\max_{z \in I} |H(z)| + \int_I |H'(z)| dz \right)\right) \quad (26)$$

for the function $H(z) = B(w, z) f(w, z)^{T-1}$ and the interval $I = [c_1 |w|, e^{2w^2}]$. By Lemmas 8 and 6 we find $\max_{z \in I} |B(w, z) f(w, z)^{T-1}| = O(w^{-3} e^{-cT/w})$ for some constant $c > 0$. For the derivative we thus have the upper bound

$$\begin{aligned} |H'(z)| &\leq |B_z| |f|^{T-1} + |B| (T-1) |f|^{T-2} |f_z| \\ &= O\left(\frac{e^{-cT/w}}{w^3 z}\right) + O\left(\frac{T}{w^3} e^{-c'T \frac{\log(z/w)}{w}} \frac{1}{zw}\right) \end{aligned}$$

(for some constants $c, c' > 0$), which leads to

$$\int_{c_1 |w|}^{e^{2w^2}} |H'(z)| dz = O\left(\frac{e^{-cT/|w|}}{w^3} \log(e^{2w^2})\right) + O\left(\frac{1}{|w|^3} e^{-c'T \frac{\log c_1}{|w|}}\right) = O(e^{-cT/|w|}).$$

Consequently, we obtain

$$\begin{aligned} J_{21} &= \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{c_1 |w|}^{e^{2w^2}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon wz} dz dw \\ &= O\left(\int_1^{\eta T} e^{-cT/w - \varepsilon w^2} dw\right) \\ &= O\left(\int_1^{(T/\varepsilon)^{1/3}} e^{-cT^{2/3} \varepsilon^{1/3}} dw + \int_{(T/\varepsilon)^{1/3}}^{\eta T} e^{-\varepsilon w^2} dw\right) \\ &= O(e^{-(c/2)T^{2/3} \varepsilon^{1/3}}) = O\left(\frac{1}{T}\right) \end{aligned}$$

as long as $\varepsilon \geq c'' T^{-2} (\log T)^9$ for a proper constant $c'' > 0$.

In the range $|z| \geq e^{2w^2}$ we use the upper bound $|f(z, w)| \leq C_2 \exp(e^{|w|})/z$ and find

$$\int_{e^{2w^2}}^\infty |B(w, z) f(w, z)^{T-1}| dz = O\left(\frac{w^{-3}}{T} e^{(T-1)w^2 - 2(T-2)w^2}\right) = O\left(\frac{1}{T} e^{-Tw^2/2}\right),$$

and consequently

$$J_{22} = \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{|z| \geq e^{2w^2}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw = O\left(\frac{1}{T}\right).$$

We note that it was essential to use inequality (26). Had we upper bounded the corresponding integral by $\int_I |H(z)| dz$, it would have been impossible to cover the range $\varepsilon = \Omega((\log T)^9/T^2)$. The same comment applies to the third range C.

C: The case $|w| \geq \eta T$. We partition the integration with respect to z into three ranges: $|z| \leq c_1|w|$, $c_1|w| \leq |z| \leq e^{2w^2}$, and $|z| \geq e^{2w^2}$. In the first case we use the trivial estimate $Bf^{T-1} = O(w^{-3})$ and obtain

$$\begin{aligned} J_{31} &= \frac{1}{2\pi} \int_{|w| \geq \eta T} \int_{|z| \leq c_1|w|} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{\eta T}^{\infty} w^{-2} e^{-\varepsilon w^2} dw\right) \\ &= O\left(\frac{1}{\eta^3 T^3} e^{-\varepsilon \eta^2 T^2}\right) = O(e^{-c(\log T)^3}). \end{aligned}$$

For the second case we set $H(z) = B(w, z) f(w, z)^{T-1}$ and use the property that $\max_{c_1|w| \leq z \leq e^{2w^2}} |H(z)| + \int_{c_1|w|}^{e^{2w^2}} |H'(z)| dz = O(e^{-cT/|w|})$ for some $c > 0$, which leads to $\int_{c_1|w|}^{e^{2w^2}} B(w, z) f(w, z)^{T-1} e^{-2i\varepsilon w z} dz = O\left(\frac{1}{\varepsilon|w|} e^{-cT/|w|}\right)$. Thus,

$$\begin{aligned} J_{32} &= \frac{1}{2\pi} \int_{|w| \geq \eta T} \int_{c_1|w| \leq |z| \leq e^{2w^2}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{\eta T}^{\infty} \frac{1}{\varepsilon|w|} e^{-cT/|w|} e^{-\varepsilon w^2} dw\right) = O\left(\frac{1}{\varepsilon \eta T} e^{-\varepsilon \eta^2 T^2}\right) = O(e^{-c(\log T)^3}). \end{aligned}$$

Finally, for $|z| \geq e^{2w^2}$ we use (as above) the upper bound $|f(w, z)| \leq C_2 e^{e|w|}/|z|$ and the resulting bound $\int_{e^{2w^2}}^{\infty} |B(w, z) f(w, z)^{T-1}| dz = O(\frac{1}{T} e^{-2Tw^2})$ to obtain

$$\begin{aligned} J_{33} &= \frac{1}{2\pi} \int_{|w| \geq \eta T} \int_{|z| \geq e^{2w^2}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{\eta T}^{\infty} \frac{1}{T} e^{-Tw^2} e^{-\varepsilon w^2} dw\right) = O(e^{-\eta^3 T^3}), \end{aligned}$$

which is certainly negligible.

3.2.5 The case $d > 1$

In principle, the case $d > 1$ can be handled very similarly to the case $d = 1$. For example, there is the same kind of Gaussian approximation

$$f(w, z) = \exp\left(-\frac{1}{2} z^T B(w) z\right) \times \begin{cases} (1 + O(\|z\|^3 + \|z\|^4)) & \text{for } \|w\| \leq 1, \\ \left(1 + O\left(\frac{\|z\|^3}{\|w\|^4} + \frac{\|z\|^2 |\sin \varphi|^2}{\|w\|}\right)\right) & \text{for } \|w\| > 1, \end{cases}$$

that leads to the asymptotic leading term for $\mathbb{E}[S_\varepsilon(X^T)]$. However, most parts of the $2d$ -dimensional integral do not contribute to the asymptotic leading term and have to be upper bounded. For example, instead of Lemma 5 we have an upper bound of the form (compare with [Drmotá et al., 2025](#))

$$|f(z, w)| \leq \min \left(1, C_1 \frac{\log \|z\|}{\|z\|} \right) \quad (27)$$

if $\|w\| \leq 1$ (for some constant $C_1 > 0$), and

$$|f(z, w)| \leq \min \left(1, C_2 \frac{\log(\|z\| \|w\|) + e^{\|w\|}}{\sqrt{\|z\| \|w\|}} \right) \quad (28)$$

if $\|w\| \geq 1$ (for some constant $C_2 > 0$). We do not repeat the technical details but refer (again) to [Drmotá et al. \(2025\)](#), where the d -dimensional case was worked out for the logistic case $p(w) = (1 + e^{-w})^{-1}$.

3.3 Proof of Proposition 3

We recall the notation B_d for the d -dimensional unit ball. Its d -dimensional volume is $v_d = \pi^{d/2}/\Gamma(\frac{d}{2} + 1)$. We also note that S_d denotes the corresponding sphere, which has $((d-1)$ -dimensional) surface measure $s_d = dv_d = 2\pi^{d/2}/\Gamma(\frac{d}{2})$. We now provide the spectral decomposition of $\overline{B}(w)$.

Lemma 9. (i) Suppose that $d \geq 2$, let $u = w/\|w\|$ and set $q(w) = \frac{p'(w)^2}{p(w)p(-w)}$. Then

$$\overline{B}(w) = \phi(w)(I_d - u \otimes u) + \lambda(w) u \otimes u,$$

where I_d is the identity operator on $\mathbb{R}^d$ (thus $I_d - u \otimes u$ is the identity operator on the hyperplane orthogonal to u), with

$$\lambda(w) = \frac{v_{d-1}}{v_d} \int_0^\pi \cos(\theta)^2 \sin(\theta)^d q(\cos(\theta)\|w\|) d\theta \quad (29)$$

and

$$\phi(w) = \frac{1}{d+2} \int_0^\pi \sin(\theta)^{d+2} q(\cos(\theta)\|w\|) d\theta \quad (30)$$

the eigenvalues of $\overline{B}(w)$, with multiplicity 1 and $d-1$ respectively.

(ii) The eigenvalue $\lambda(w)$ is asymptotically $\lambda(w) = \frac{2v_{d-1}}{v_d} q^*(3) \|w\|^{-3} (1 + O(\|w\|^{-2}))$, where $q^*(s)$ denotes the Mellin transform of q . Furthermore, the eigenvalue $\phi(w)$ is asymptotically $\phi(w) = \frac{2}{d+2} \frac{q^*(1)}{\|w\|} (1 + O(\|w\|^{-2}))$. Consequently

$$\det \overline{B}(w) = \lambda(w) \phi^{d-1}(w) = \frac{v_{d-1}}{v_d} \frac{2^d}{(d+2)^{d-1}} q^*(1)^{d-1} q^*(3) \|w\|^{-d-2} (1 + O(\|w\|^{-2})).$$

The proof of the spectral decomposition is the same as in [Drmotá et al. \(2025\)](#). Only the proof of the asymptotics for $\lambda(w)$ and $\phi(w)$ needs a small modification. Since $\lambda(w)$ and $\phi(w)$ depend only on the norm $w = \|w\|$ we also use the shorthand notation $\lambda(w)$ and $\phi(w)$.

We now apply the machinery of analytic combinatorics, especially Mellin transform asymptotics, which can be learned from [Flajolet and Sedgewick \(2008\)](#) and [Szpankowski \(2001\)](#). As in [Drmotá et al. \(2025\)](#) one uses the Mellin transforms

$$\lambda^*(s) = \int_0^\infty \lambda(w) w^{s-1} dw, \quad \phi^*(s) = \int_0^\infty \phi(w) w^{s-1} dw.$$

As in [Drmotá et al. \(2025\)](#) it follows that

$$\begin{aligned}\lambda^*(s) &= \frac{2v_{d-1}}{v_d} \int_0^1 (1-y^2)^{(d-1)/2} y^2 \int_0^\infty q(yw) w^{s-1} dw dy \\ &= \frac{2v_{d-1}}{v_d} q^*(s) \int_0^1 (1-y^2)^{(d-1)/2} y^{2-s} dy = \frac{2v_{d-1}}{v_d} q^*(s) \beta_1^*(3-s),\end{aligned}$$

where $q^*(s) = \int_0^\infty q(w) w^{s-1} dw$ is the Mellin transform of q and

$$\beta_1^*(s) = \frac{\Gamma(\frac{d+1}{2})\Gamma(\frac{s}{2})}{2\Gamma(\frac{d+1+s}{2})}$$

is the Mellin transform of the function $(1-y^2)^{(d-1)/2}$ defined over $[0,1]$. As usual Γ denotes the Euler gamma function. Note that $q^*(s)$ and $\beta_1^*(s)$ are analytic for $\text{Re}(s) > 0$. Thus, the dominating singularity of $\lambda^*(s)$ is a simple pole at $s = 3$ (coming from $\Gamma(\frac{3-s}{2})$) with residue $\text{res}(\lambda^*, s=3) = -\frac{2v_{d-1}}{v_d} q^*(3)$. This implies that $\lambda(w) = \lambda(\|w\|) = \frac{2v_{d-1}}{v_d} q^*(3) \|w\|^{-3} (1 + O(\|w\|^{-2}))$ as $w \rightarrow \infty$, since $s = 5$ is the next pole of $\lambda^*(s)$.

We can make a similar analysis for $\phi^*(s)$ and arrive at $\phi^*(s) = \frac{2}{d+2} q^*(s) \beta_2^*(1-s)$, where $\beta_2^*(s)$ is the Mellin transform of the function $(1-y^2)^{(d+1)/2}$. Hence,

$$\phi^*(s) = \frac{1}{d+2} q^*(s) \frac{\Gamma(\frac{d+3}{2})\Gamma(\frac{1-s}{2})}{\Gamma(\frac{d+4-s}{2})}.$$

Here the dominating singularity is a simple pole at $s = 1$ with residue $\text{res}(\phi^*, s=1) = -\frac{2}{d+2} q^*(1)$. The next pole is $s = 3$. Thus, we have $\phi(w) = \phi(\|w\|) = \frac{2}{d+2} \frac{q^*(1)}{\|w\|} (1 + O(\|w\|^{-2}))$, and consequently

$$\det \bar{B}(w) = \lambda(w) \phi(w)^{d-1} = \frac{v_{d-1}}{v_d} \frac{2^d}{(d+2)^{d-1}} q^*(1)^{d-1} q^*(3) \|w\|^{-d-2} (1 + O(\|w\|^{-2})),$$

as proposed.

AI disclosure

I used Claude (Anthropic), a large language model, to check the manuscript for mathematical, typographical, and consistency errors. All results, derivations, and proofs were developed by the author, who verified every suggested correction independently and takes full responsibility for the content of the paper.

References

- Naman Agarwal, Satyen Kale, and Julian Zimmert. Efficient methods for online multiclass logistic regression. In *Proceedings of the 33rd International Conference on Algorithmic Learning Theory (ALT)*, volume 167 of *PMLR*, pages 3–33, 2022.
- Blair Bilodeau, Dylan J. Foster, and Daniel M. Roy. Tight bounds on minimax regret under logarithmic loss via self-concordance. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *PMLR*, pages 919–929, 2020.

- Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- Nicolò Cesa-Bianchi and Ohad Shamir. Efficient transductive online learning via randomized rounding. arXiv preprint arXiv:1106.2429, 2011.
- Bertrand S. Clarke and Andrew R. Barron. Information-theoretic asymptotics of Bayes methods. *IEEE Transactions on Information Theory*, 36(3):453–471, 1990.
- Michael Drmota and Wojciech Szpankowski. Precise minimax redundancy and regret. *IEEE Transactions on Information Theory*, 50(11):2686–2707, 2004.
- Michael Drmota, Philippe Jacquet, Changlong Wu, and Wojciech Szpankowski. Precise regularized minimax regret with unbounded weights. *IEEE Transactions on Information Theory*, 71:8753–8776, 2025.
- Michael Drmota, Philippe Jacquet, Changlong Wu, and Wojciech Szpankowski. Phase transition of regret for logistic regression with large weights. In *Algorithmic Learning Theory (ALT)*, PMLR, 2026.
- Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2008.
- Dylan J. Foster, Satyen Kale, Haipeng Luo, Mehryar Mohri, and Karthik Sridharan. Logistic regression: The importance of being improper. In *Proceedings of the 31st Conference on Learning Theory (COLT)*, 2018.
- Elad Hazan, Tomer Koren, and Kfir Y. Levy. Logistic regression: Tight bounds for stochastic and online optimization. In *Proceedings of the 27th Conference on Learning Theory (COLT)*, pages 197–209, 2014.
- M. N. Huxley. *Area, Lattice Points, and Exponential Sums*, volume 13 of *London Mathematical Society Monographs, New Series*. Clarendon Press, Oxford, 1996.
- Philippe Jacquet and Wojciech Szpankowski. Markov types and minimax redundancy for Markov sources. *IEEE Transactions on Information Theory*, 50:1393–1402, 2004.
- Philippe Jacquet, Gil I. Shamir, and Wojciech Szpankowski. Precise minimax regret for logistic regression with categorical feature values. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory (ALT)*, volume 132 of *PMLR*, pages 755–771, 2021.
- Rémi Jézéquel, Pierre Gaillard, and Alessandro Rudi. Efficient improper learning for online logistic regression. In *Proceedings of the 33rd Conference on Learning Theory (COLT)*, volume 125 of *PMLR*, pages 2085–2108, 2020.
- Sham M. Kakade and Adam Tauman Kalai. From batch to transductive online learning. In *Advances in Neural Information Processing Systems 18*, pages 611–618, 2005.
- Sham M. Kakade and Andrew Y. Ng. Online bounds for Bayesian algorithms. In *Advances in Neural Information Processing Systems 17*, pages 641–648, 2005.
- Jack J. Mayo, Hédi Hadiji, and Tim van Erven. Scale-free unconstrained online learning for curved losses. In *Proceedings of the 35th Conference on Learning Theory (COLT)*, volume 178 of *PMLR*, pages 1–34, 2022.

- H. Brendan McMahan and Matthew Streeter. Open problem: Better bounds for online logistic regression. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, volume 23 of *PMLR*, pages 44.1–44.3, 2012.
- Francesco Orabona and Dávid Pál. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems 29*, pages 577–585, 2016.
- Jian Qian, Alexander Rakhlin, and Nikita Zhivotovskiy. Refined risk bounds for unbounded losses via transductive priors. arXiv preprint arXiv:2410.21621, 2024.
- Alexander Rakhlin and Karthik Sridharan. Sequential probability assignment with binary alphabets and large classes of experts. arXiv preprint arXiv:1501.07340, 2015.
- Jorma Rissanen. Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42(1):40–47, 1996.
- Gil I. Shamir. Logistic regression regret: What’s the catch? In *Proceedings of the 33rd Conference on Learning Theory (COLT)*, PMLR, pages 3296–3319, 2020.
- Gil I. Shamir and Wojciech Szpankowski. Low complexity approximate Bayesian logistic regression for sparse online learning. arXiv preprint arXiv:2101.12113; see also ISIT 2024, 2021.
- Yu. M. Shtarkov. Universal sequential coding of single messages. *Problems of Information Transmission*, 23(3):175–186, 1987.
- Wojciech Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.
- Changlong Wu, Mohsen Heidari, Ananth Grama, and Wojciech Szpankowski. Precise regret bounds for log-loss via a truncated Bayesian algorithm. In *Advances in Neural Information Processing Systems*, 2022.
- Changlong Wu, Mohsen Heidari, Ananth Grama, and Wojciech Szpankowski. Expected worst case regret via stochastic sequential covering. *Transactions on Machine Learning Research*, 2023.
- Changlong Wu, Ananth Grama, and Wojciech Szpankowski. Online universal learning from information-theoretic perspective. *Foundations and Trends in Communications and Information Theory*, 2026. URL <https://www.cs.purdue.edu/homes/spa/papers/fnt25.pdf>.

A Proofs of Lemmas 5–8

The proofs of Lemmas 5–8 rely mainly on the following two properties (which can be found in [Huxley, 1996](#)). Lemma 10 is used more frequently; Lemma 11 is only used when Lemma 10 does not apply.

Lemma 10. *Let $\beta_1 < \beta_2$ be real numbers. Assume that h is continuously differentiable on $[\beta_1, \beta_2]$ and has a monotone nonvanishing derivative. Then for each continuous function H of bounded variation we have*

$$\left| \int_{\beta_1}^{\beta_2} H(x) e^{ih(x)} dx \right| \leq 2 \frac{\max_{[\beta_1, \beta_2]} |H| + V_{\beta_1}^{\beta_2}(H)}{\min_{[\beta_1, \beta_2]} |h'|}, \quad (31)$$

where $V_{\beta_1}^{\beta_2}(H)$ denotes the total variation of H on $[\beta_1, \beta_2]$.

Lemma 11. *Let $\beta_1 < \beta_2$ be real numbers. Assume that h is twice continuously differentiable on $[\beta_1, \beta_2]$ such that the second derivative is non-zero. Then for each continuous function H of bounded variation we have*

$$\left| \int_{\beta_1}^{\beta_2} H(x) e^{ih(x)} dx \right| \leq 8 \frac{\max_{[\beta_1, \beta_2]} |H| + V_{\beta_1}^{\beta_2}(H)}{\min_{[\beta_1, \beta_2]} \sqrt{|h''|}}. \quad (32)$$

The proof of Lemma 5 is actually the same as that given in [Drmota et al. \(2025\)](#). Instead of the explicit logistic function $p(w) = (1 + e^{-w})^{-1}$ we use the properties stated in (H). More precisely, we set $h(x) = xz \frac{p'(xw)}{p(xw)}$, which gives $h'(x) = z((\log p(u))'u)'$ (with $u = xw$). If $((\log p(u))'u)' \neq 0$ then we apply Lemma 10, and if we are close to the zeros of the function $((\log p(u))'u)'$, where Lemma 10 does not apply, we use the property that $((\log p(u))'u)'' \neq 0$ in a small neighborhood of the zero, so that Lemma 11 applies.

Lemma 6 is an extension of a corresponding property in [Drmota et al. \(2025\)](#), where it was only proved that $|f(w, z)| \leq 1 - \frac{c_2}{|w|}$ for some constant $c_2 > 0$ if $|z| \geq c_1|w|$. Suppose that $w > 0$. Then we first have $\left| \int_0^1 p(-xw) e^{ih(x)} dx \right| = O\left(\frac{1}{|z|}\right)$. Similarly to the proof of Lemma 10 we have

$$\begin{aligned} \left| \int_0^1 p(xw) e^{ih(-x)} dx \right| &= \left| \left(\int_0^{1/w} + \int_{1/w}^{(|u_0|+\eta)/w} + \int_{(|u_0|+\eta)/w}^{c/w} + \int_{c/w}^1 \right) p(xw) e^{ih(-x)} dx \right| \\ &\leq 1 - \frac{c}{w} + O\left(\frac{1}{|z|} + \frac{1}{\sqrt{|zw|}} + \frac{e^c}{|z|c}\right), \end{aligned}$$

where $c > |u_0| + \eta$ and u_0 denotes the zero of $((\log p(u))'u)'$. By setting $c = \max\{1, \log(|z|/|w|)\}$ we thus obtain

$$|f(z, w)| \leq 1 - \frac{\max\{1, \log(|z|/|w|)\}}{w} + O\left(\frac{\max\{1, \log(|z|/|w|)\}}{|z|} + \frac{1}{\sqrt{|zw|}}\right).$$

Thus, if $|z| \geq c'_1 w$ for a sufficiently large constant c'_1 , we have $|f(z, w)| \leq 1 - \frac{c_2 \max\{1, \log(|z|/|w|)\}}{w}$ for some constant $c_2 > 0$. Finally, the case $c_1 w \leq |z| \leq c'_1 w$ follows as in [Drmota et al. \(2025\)](#).

Lemma 7 is a more precise form of Lemma 6 and relies first on the computation

$$\begin{aligned} f(w, z) &= \frac{1}{2} \int_{-1}^1 \left(p(xw) e^{ixz \frac{p'(xw)}{p(xw)}} + p(-xw) e^{-ixz \frac{p'(-xw)}{p(-xw)}} \right) dx \\ &= 1 - \int_{-1}^1 p(xw) \left(1 - e^{ixz \frac{p'(xw)}{p(xw)}} \right) dx \\ &= 1 - \frac{k(z/w)}{w} + \int_{|x| \geq 1} p(xw) \left(1 - e^{ixz \frac{p'(xw)}{p(xw)}} \right) dx. \end{aligned}$$

It remains to upper bound the remaining integral. This can be done with the help of the tail estimates provided by hypothesis (H), for which we need the asymptotic property $k(v) = \log v + \log \log v + O(1)$ and the upper bound for $\frac{\partial}{\partial z} f(w, z)$. For both properties one uses again Lemma 10 (we omit the easy details).

Finally, Lemma 8 is a direct analogue of corresponding properties from Drmota et al. (2025) for the logistic case. We only added an upper bound for $\frac{\partial}{\partial z} B(w, z)$, which is immediate from the integral representation of $B(w, z)$ and a direct application of Lemma 10.

B The second moment

The main difference from the first moment is that the second moment $\mathbb{E}[S_\varepsilon(x^T)^2]$ cannot be explicitly represented as a convergent multi-dimensional integral. In order to handle this problem we use a regularized version of $S_\varepsilon(x^T)$ of the form

$$S_\varepsilon^\nu(x^T) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \prod_{t=1}^T f(w, x_t, z) e^{-\varepsilon w^2 - 2i\varepsilon w z} \nabla^2 L_\varepsilon(y^T | x^T, w) e^{-\nu z^2} dw dz, \quad (33)$$

which is absolutely convergent as a double integral if $\nu > 0$ and has the property that $S_\varepsilon(x^T) = \lim_{\nu \rightarrow 0} S_\varepsilon^\nu(x^T)$. As in Drmota et al. (2025) it follows that

$$|S_\varepsilon(x^T) - S_\varepsilon^\nu(x^T)| = O\left(\left(\frac{\nu}{\varepsilon^2} + \frac{\nu^{1/2}}{\varepsilon^{3/2}}\right) e^{O(T^2/\varepsilon)}\right),$$

and consequently (since $\mathbb{E}[S_\varepsilon(x^T)] = O(\sqrt{T})$)

$$|\mathbb{E}[S_\varepsilon(x^T)^2] - \mathbb{E}[S_\varepsilon^\nu(x^T)^2]| = O\left(\left(\frac{\nu}{\varepsilon^2} + \frac{\nu^{1/2}}{\varepsilon^{3/2}}\right) e^{O(T^2/\varepsilon)}\right).$$

In particular, if we choose

$$\nu = e^{-CT^2/\varepsilon} > 0 \quad (34)$$

for a sufficiently large constant C , we find

$$|\mathbb{E}[S_\varepsilon(x^T)^2] - \mathbb{E}[S_\varepsilon^\nu(x^T)^2]| = O(\varepsilon^{-2} e^{-C'T^2/\varepsilon}) = O(T e^{-C'(\log T)^9})$$

if $\varepsilon \gg (\log T)^9 T^{-2}$.

Thus, it suffices to compute $\mathbb{E}[S_\varepsilon^\nu(x^T)^2]$. Since $\nu > 0$, $S_\varepsilon^\nu(x^T)$ is represented as an absolutely convergent double integral (33), and by dominated convergence we obtain, writing $\Xi = e^{-\varepsilon(w_1^2 + w_2^2) - 2i\varepsilon(w_1 z_1 - w_2 z_2)} e^{-\nu(z_1^2 + z_2^2)}$ and $d\mathbf{v} = dw_1 dw_2 dz_1 dz_2$,

$$\begin{aligned} \mathbb{E}[S_\varepsilon^\nu(x^T)^2] &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^4} \mathbb{E}\left[\prod_{t=1}^T (f(w_1, x_t, z_1) f(w_2, x_t, z_2)) \nabla^2 L_\varepsilon(y^T | x^T, w_1) \nabla^2 L_\varepsilon(y^T | x^T, w_2)\right] \Xi d\mathbf{v} \\ &= \frac{T^2 - T}{(2\pi)^2} \int_{\mathbb{R}^4} f(w_1, w_2, z_1, z_2)^{T-1} B(z_1, w_1) B(z_2, w_2) \Xi d\mathbf{v} \\ &\quad + \frac{T}{(2\pi)^2} \int_{\mathbb{R}^4} f(w_1, w_2, z_1, z_2)^{T-1} B_2(z_1, z_2, w_1, w_2) \Xi d\mathbf{v} \\ &\quad + 2 \frac{2\varepsilon T}{(2\pi)^2} \int_{\mathbb{R}^4} f(w_1, w_2, z_1, z_2)^{T-1} f(w_1, z_1) B(z_2, w_2) \Xi d\mathbf{v} \\ &\quad + \frac{(2\varepsilon)^2}{(2\pi)^2} \int_{\mathbb{R}^4} f(w_1, w_2, z_1, z_2)^T \Xi d\mathbf{v} \\ &= (T^2 - T) \cdot J_0^\nu + T \cdot J_{0,2}^\nu + 2\varepsilon T \cdot J_1^\nu + (2\varepsilon)^2 J_2^\nu, \end{aligned}$$

where

$$\begin{aligned} f(w_1, w_2, z_1, z_2) &= \mathbb{E}[f(w_1, X, z_1)f(w_2, X, z_2)] \\ &= \frac{1}{2} \int_{-1}^1 \left(\frac{e^{-ixz_1/(1+e^{xw_1})}}{1+e^{-xw_1}} + \frac{e^{ixz_1/(1+e^{-xw_1})}}{1+e^{xw_1}} \right) \left(\frac{e^{ixz_2/(1+e^{xw_2})}}{1+e^{-xw_2}} + \frac{e^{-ixz_2/(1+e^{-xw_2})}}{1+e^{xw_2}} \right) dx \end{aligned}$$

(written here for the logistic case) and

$$B_2(z_1, z_2, w_1, w_2) = \mathbb{E}[f(w_1, X, z_1)f(w_2, X, z_2)p(Xw_1)p(-Xw_1)p(Xw_2)p(-Xw_2)X^4].$$

By using corresponding upper bounds for $f(w_1, w_2, z_1, z_2)$ (compare with [Drmota et al., 2025](#)) one can handle the integrals J_0^ν , $J_{0,2}^\nu$, J_1^ν , J_2^ν . As in the case of the expected value it follows, for example, that $J_0^\nu = J_0^2(1 + O(T^{-\beta})) + O((1 - \kappa)^T \log \nu^{-1})$ for some $\beta > 0$ and some $\kappa > 0$. Putting all these estimates together we finally obtain (by setting $\nu = \exp(-CT^2/\varepsilon)$, see (34)) $\mathbb{E}[S_\varepsilon^\nu(x^T)^2] = \mathbb{E}[S_\varepsilon(x^T)]^2(1 + O(T^{-\beta}))$ and consequently

$$\mathbb{E}[S_\varepsilon(x^T)^2] = \mathbb{E}[S_\varepsilon(x^T)]^2(1 + O(T^{-\beta})). \quad (35)$$

C The probit class

We assume that $d = 1$, so that the X_t are i.i.d. and uniformly distributed on $[-1, 1]$. We also use as $p(w)$ the Gaussian distribution function (the probit case)

$$p(w) = \Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-v^2/2} dv.$$

Note that p satisfies $1 - p(w) = p(-w)$. Since $\Phi'(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}$ and $\Phi(w) > 1 - e^{-w^2/2}/(\sqrt{2\pi}w)$ (for $w > 0$), and more precisely

$$\Phi(w) = 1 - \frac{e^{-w^2/2}}{\sqrt{2\pi}w} + \frac{e^{-w^2/2}}{\sqrt{2\pi}w^3}(1 + O(w^{-2})) \quad (w \rightarrow +\infty),$$

$p = \Phi$ satisfies the properties $(\log p(w))'' < 0$ and $(\log(1 - p(w)))'' < 0$ and

$$p'(w) = O(1), \quad p''(w) = O(1), \quad p'''(w) = O(1),$$

$$(\log p)' = O(w), \quad (\log p)'' = O(w^2), \quad (\log p)''' = O(w^3),$$

and similarly $(\log(1 - p))' = O(w)$, $(\log(1 - p))'' = O(w^2)$, $(\log(1 - p))''' = O(w^3)$. It is also easy to check that the property

$$\limsup_{|a| \rightarrow \infty} \frac{|G_{y^T|x^T}^{-1}(a)|}{|a|} > 0 \quad (36)$$

holds. Suppose that y^T and x^T are given and without loss of generality assume that $w > 0$. Then $G_{y^T|x^T}(w) = w \sum_{1 \leq t \leq T, y_t x_t < 0} x_t^2 + 2\varepsilon w + O(1/w)$. Thus, $a = G_{y^T|x^T}(w) \geq (2\varepsilon + o(1))w$, which implies (36).

With the help of these properties of $p = \Phi$ it follows that the expected regularized Shtarkov sum is given by

$$\begin{aligned} \mathbb{E}[S_\varepsilon(X^T)] &= \frac{T}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(w, z)^{T-1} B(w, z) e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz \\ &\quad + \frac{2\varepsilon}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz =: T \cdot J_0 + 2\varepsilon \cdot J_1, \end{aligned}$$

where

$$\begin{aligned} f(w, z) &= \mathbb{E}[f(w, X, z)] \\ &= \frac{1}{2} \int_{-1}^1 \left(\Phi(xw) \exp\left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz\right) + \Phi(-xw) \exp\left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz\right) \right) dx, \end{aligned} \quad (37)$$

and where $B(w, z)$ abbreviates

$$\begin{aligned} B(w, z) &= \mathbb{E} \left[\Phi(Xw) \exp\left(-i \frac{e^{-X^2 w^2/2}}{\sqrt{2\pi}\Phi(Xw)} Xz\right) \left(\frac{e^{-X^2 w^2}}{2\pi\Phi(Xw)^2} + \frac{Xwe^{-X^2 w^2/2}}{\sqrt{2\pi}\Phi(Xw)} \right) X^2 \right. \\ &\quad \left. + \Phi(-Xw) \exp\left(i \frac{e^{-X^2 w^2/2}}{\sqrt{2\pi}\Phi(-Xw)} Xz\right) \left(\frac{e^{-X^2 w^2}}{2\pi\Phi(-Xw)^2} - \frac{Xwe^{-X^2 w^2/2}}{\sqrt{2\pi}\Phi(-Xw)} \right) X^2 \right] \\ &= \frac{1}{2} \int_{-1}^1 \left(\exp\left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz\right) \left(\frac{e^{-x^2 w^2}}{2\pi\Phi(xw)} + \frac{xwe^{-x^2 w^2/2}}{\sqrt{2\pi}} \right) x^2 \right. \\ &\quad \left. + \exp\left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz\right) \left(\frac{e^{-x^2 w^2}}{2\pi\Phi(-xw)} - \frac{xwe^{-x^2 w^2/2}}{\sqrt{2\pi}} \right) x^2 \right) dx. \end{aligned}$$

We also set

$$B(w) = B(w, 0) = \mathbb{E} \left[\frac{X^2 e^{-X^2 w^2}}{2\pi\Phi(Xw)\Phi(-Xw)} \right] = \frac{1}{2} \int_{-1}^1 \frac{x^2 e^{-x^2 w^2}}{2\pi\Phi(xw)\Phi(-xw)} dx.$$

Our goal is to give a full proof of the following asymptotic property that is the basis of all other results, see Theorem 1 (we skip the corresponding computations for the second moment and for $d > 1$; they are more involved but follow precisely the lines given here combined with the methods of [Drmota et al., 2025](#)).

Proposition 12. *Suppose that $p = \Phi$ and that the X_t are i.i.d. and uniformly distributed on $[-1, 1]$. Then we have uniformly for $\varepsilon \geq (\log T)^9/T^2$*

$$\mathbb{E}[S_\varepsilon(X^T)] = \frac{\sqrt{T}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} B(w)^{1/2} e^{-\varepsilon w^2} dw + O(T^{-\frac{1}{2}-\beta})$$

for some $\beta > 0$.

We start with some technical lemmas.

Lemma 13. *We uniformly have*

$$\begin{aligned} |B(z, w)| &= O(\min(1, |w|^{-3})), \quad B(w) = \Theta(\min(1, |w|^{-3})), \\ B(z, w) &= B(w) + O(|z| \min(1, |w|^{-4})). \end{aligned}$$

More precisely,

$$B(w) = \frac{2}{3\pi} + O(w^2) \quad (w \rightarrow 0), \quad B(w) = \frac{1}{4\pi w^3} \int_{-\infty}^{\infty} \frac{v^2 e^{-v^2}}{\Phi(v)\Phi(-v)} dv + O(w^{-1} e^{-w^2/2}) \quad (w \rightarrow \infty).$$

Furthermore, for some constant $c_1 > 0$,

$$\begin{aligned} f(w, z) &= 1 - \Theta(z^2 \min(1, |w|^{-3})) \quad (\text{for } |z| \leq \max(1, c_1 |w|)) \\ &= 1 - \frac{z^2}{2} B(w) + O(z^3 \min(1, |w|^{-4})) \\ &= e^{-\frac{1}{2} z^2 B(w)} \left(1 + O(z^3 \min(1, |w|^{-4})) + O(z^4 \min(1, |w|^{-6})) \right). \end{aligned}$$

Proof. By symmetry we can rewrite the definition of $f(w, z)$:

$$f(w, z) = \int_0^1 \left(\Phi(xw) \exp \left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz \right) + \Phi(-xw) \exp \left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz \right) \right) dx.$$

By differentiation we directly obtain $f(w, 0) = \int_0^1 (\Phi(xw) + \Phi(-xw)) dx = 1$,

$$\begin{aligned} \frac{\partial f}{\partial z}(w, z) &= \frac{-i}{\sqrt{2\pi}} \int_0^1 x e^{-x^2 w^2/2} \left(\exp \left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz \right) \right. \\ &\quad \left. - \exp \left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz \right) \right) dx, \\ \frac{\partial f}{\partial z}(w, 0) &= 0, \\ \frac{\partial^2 f}{\partial z^2}(w, z) &= -\frac{1}{2\pi} \int_0^1 x^2 e^{-x^2 w^2} \left(\frac{1}{\Phi(xw)} \exp \left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz \right) \right. \\ &\quad \left. + \frac{1}{\Phi(-xw)} \exp \left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz \right) \right) dx, \\ \frac{\partial^2 f}{\partial z^2}(w, 0) &= -\frac{1}{2\pi} \int_0^1 \frac{x^2 e^{-x^2 w^2}}{\Phi(xw)\Phi(-xw)} dx = -B(w), \\ \frac{\partial^3 f}{\partial z^3}(w, z) &= \frac{i}{2\sqrt{2\pi}^3} \int_0^1 x^3 e^{-3x^2 w^2/2} \left(\frac{1}{\Phi(xw)^2} \exp \left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz \right) \right. \\ &\quad \left. - \frac{1}{\Phi(-xw)^2} \exp \left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz \right) \right) dx. \end{aligned}$$

Since $\Phi(xw)\Phi(-xw) = \Phi(0)^2 + O(w^2 x^2)$ and $e^{-x^2 w^2} = 1 + O(x^2 w^2)$ we have (as $w \rightarrow 0$)

$$B(w) = \int_{-1}^1 \frac{x^2}{4\pi\Phi(0)^2} dx + O\left(\int_{-1}^1 w^2 x^4 dx\right) = \frac{2}{3\pi} + O(w^2).$$

For $w \rightarrow \infty$ we obtain

$$\begin{aligned} B(w) &= \frac{1}{2} \int_{-1}^1 \frac{x^2 e^{-x^2 w^2}}{2\pi\Phi(xw)\Phi(-xw)} dx = \frac{1}{4\pi w^3} \int_{-w}^w \frac{v^2 e^{-v^2}}{\Phi(v)\Phi(-v)} dv \\ &= \frac{1}{4\pi w^3} \int_{-\infty}^{\infty} \frac{v^2 e^{-v^2}}{\Phi(v)\Phi(-v)} dv + O\left(w^{-3} \int_w^{\infty} v^3 e^{-v^2/2} dv\right) \\ &= \frac{1}{4\pi w^3} \int_{-\infty}^{\infty} \frac{v^2 e^{-v^2}}{\Phi(v)\Phi(-v)} dv + O(w^{-1} e^{-w^2/2}). \end{aligned}$$

Putting these two properties together we also get $B(w) = \Theta(\min(1, |w|^{-3}))$. Similarly it follows that $\frac{\partial^3 f}{\partial z^3}(w, z) = O(1)$ as $w \rightarrow 0$ (uniformly in z) and, as $w \rightarrow \infty$,

$$\begin{aligned} \frac{\partial^3 f}{\partial z^3}(w, z) &= O\left(\int_0^1 x^3 e^{-3x^2 w^2/2} \left(\frac{1}{\Phi(xw)^2} + \frac{1}{\Phi(-xw)^2} \right) dx\right) \\ &= O\left(\int_0^{\infty} x^3 (xw)^2 e^{x^2 w^2/2} e^{-3x^2 w^2/2} dx\right) = O(w^{-4}). \end{aligned}$$

Thus it immediately follows that $f(z, w) = 1 - \frac{z^2}{2}B(w) + O(z^3 \min(1, |w|^{-4}))$, and by expanding $f(z, w) = e^{\log f(z, w)}$ we obtain the third representation for $f(z, w)$ (where we use $z^4 \min(1, |w|^{-6})$ as the order of $z^4 B(w)^2$). Next, if $|z| \leq \max(1, c_1 |w|)$ (for some sufficiently small constant $c_1 > 0$) it follows that $z^2 B(w) \geq z^3 \min(1, |w|^{-4})$. Thus we also get $f(w, z) = 1 - \Theta(z^2 \min(1, |w|^{-3}))$.

Finally we consider $B(w, z)$. Clearly we have $B(w, z) = O(1)$ for $w \rightarrow 0$ (uniformly in z). For $w \rightarrow \infty$ we argue similarly as above and obtain

$$|B(w, z)| \leq 4 \int_0^\infty \left(\frac{e^{-x^2 w^2}}{2\pi\Phi(-|xw|)} + \frac{x|w|e^{-x^2 w^2/2}}{\sqrt{2\pi}} \right) x^2 dx$$

$$= O\left(\int_0^\infty w x^3 e^{-x^2 w^2/2} dx \right) = O(w^{-3}).$$

Finally we use the expansion $e^{iu} = 1 + O(u)$ to obtain $B(z, w) = B(w) + O(|z| \min(1, |w|^{-4}))$, which completes the proof of the lemma. $\square$

Lemma 14. *There exists $w_0 > 0$ with the following property. For $|w| \leq w_0$ we uniformly have $|f(z, w)| \leq \min(1, C/|z|)$ for some constant $C > 0$. Furthermore, for $|w| \geq w_0$ we have*

$$|f(z, w)| \leq \min\left(1, \frac{C_1}{\sqrt{|wz|}} + \frac{C_2 e^{|w|^2/2}}{|w^2 z|}\right)$$

for some constants $C_1, C_2 > 0$.

Proof. We consider the function $h(x) = \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz$, which satisfies $h'(x) = z \bar{h}(xw)$, where

$$\bar{h}(u) = -\frac{u^2 e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} - \frac{u e^{-u^2}}{2\pi\Phi(u)^2} + \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)}.$$

We have the following three different asymptotic regimes:

$$\bar{h}(u) \sim -\frac{u^2 e^{-u^2/2}}{\sqrt{2\pi}} \quad (u \rightarrow \infty), \quad \bar{h}(u) \sim \sqrt{\frac{2}{\pi}} \quad (u \rightarrow 0), \quad \bar{h}(u) \sim -2u \quad (u \rightarrow -\infty).$$

Furthermore there is a unique zero $u_0 > 0$ of $\bar{h}$, which satisfies $\bar{h}'(u_0) < 0$. Now, $f(w, z)$ can be represented as

$$f(w, z) = \frac{1}{2} \int_{-1}^1 \Phi(xw) e^{ih(x)} dx + \frac{1}{2} \int_{-1}^1 \Phi(-xw) e^{ih(-x)} dx.$$

Note that $\Phi(xw)$ is uniformly bounded and that its variation is uniformly bounded, too. Next we set $w_0 = u_0 - \eta$ for a sufficiently small $\eta > 0$ such that $\bar{h}'(u) \neq 0$ on the interval $[u_0 - \eta, u_0 + \eta]$. If $|w| \leq w_0$ then we have $|xw| \leq u_0 - \eta$ for $|x| \leq 1$, so that $|h'(x)| = |z| |\bar{h}(xw)| \geq C'|z|$ for $|x| \leq 1$, for some constant $C' > 0$. Hence, a direct application of Lemma 10 implies $|f(w, z)| \leq C/|z|$ for some constant $C > 0$.

Next, if $|w| \geq w_0$ then we cut out the interval $I_0(w) := [\frac{u_0 - \eta}{w}, \frac{u_0 + \eta}{w}]$ from the interval $[-1, 1]$ if there is a non-empty intersection. For $x \in I_0(w)$ (which is equivalent to $xw \in [u_0 - \eta, u_0 + \eta]$) we have $h''(x) = zw \bar{h}'(xw)$ and consequently $|h''(x)| \geq C''|wz|$ for $x \in I_0(w)$, for some constant $C'' > 0$. Hence, by applying Lemma 11 we obtain an upper bound for the corresponding integral of the form $O(1/\sqrt{|wz|})$. Finally, in the remaining part of the interval $[-1, 1]$ we certainly have $h'(x) \neq 0$ and $|h'(x)| \geq C'''|z|w^2 e^{-w^2/2}$ for some constant $C''' > 0$. Thus, we can apply Lemma 10 again and obtain (for $|w| \geq w_0$) $|f(z, w)| \leq \frac{C_1}{\sqrt{|wz|}} + \frac{C_2 e^{|w|^2/2}}{|w^2 z|}$ for proper constants $C_1, C_2 > 0$.

Finally, we note that we always have the trivial upper bound $|f(z, w)| \leq 1$. $\square$

Lemma 15. Suppose that $c_1 > 0$ is a given constant. Then there exists $c_2 > 0$ such that

$$|f(z, w)| \leq 1 - \frac{c_2 \max\{1, \sqrt{\log(|z|/|w|)}\}}{|w|}$$

uniformly for $|z| \geq c_1|w|$.

Proof. We consider first the case $|z| \geq c'_1|w|$, where c'_1 will be chosen sufficiently large. Without loss of generality we assume that $w > 0$. As in the proof of Lemma 13 it follows that $|2 \int_0^1 \Phi(xw)e^{ih(-x)} dx| = O(1/|z|)$ and that

$$\begin{aligned} \left| \int_0^1 \Phi(-xw)e^{ih(x)} dx \right| &= \left| \left(\int_0^{(u_0-\eta)/w} + \int_{(u_0-\eta)/w}^{(u_0+\eta)/w} + \int_{(u_0+\eta)/w}^{c/w} + \int_{c/w}^1 \right) \Phi(-xw)e^{ih(x)} dx \right| \\ &\leq 1 - \frac{\kappa c}{w} + O\left(\frac{1}{|z|} + \frac{1}{\sqrt{|zw|}} + \frac{e^{c^2/2}}{|z|c^2}\right), \end{aligned}$$

where $c > u_0 + \eta$ and $\kappa > 0$ is a constant. By setting $c = \max\{1, 2\sqrt{\log(|z|/|w|)}\}$ we thus obtain

$$|f(w, z)| \leq 1 - \frac{\kappa \max\{1, \sqrt{2\log(|z|/|w|)}\}}{w} + O\left(\frac{\max\{1, \sqrt{\log(|z|/|w|)}\}}{|z|} + \frac{1}{\sqrt{|zw|}}\right) + O\left(\frac{1}{w \log(|z|/|w|)}\right).$$

Thus, if $|z| \geq c'_1 w$ for a sufficiently large constant c'_1 , we have $|f(z, w)| \leq 1 - \frac{c_2 \max\{1, \sqrt{\log(|z|/|w|)}\}}{w}$ for some constant $c_2 > 0$.

Finally we consider the interval $c_1 w \leq |z| \leq c'_1 w$. With $c = |z|/w$ we have

$$\begin{aligned} f(z, w) &= \int_0^1 \left(\Phi(xw) \exp\left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz\right) + \Phi(-xw) \exp\left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz\right) \right) dx \\ &= \frac{1}{w} \int_0^w \left(\Phi(v) \exp\left(-i \frac{e^{-v^2/2}}{\sqrt{2\pi}\Phi(v)} vc\right) + \Phi(-v) \exp\left(i \frac{e^{-v^2/2}}{\sqrt{2\pi}\Phi(-v)} vc\right) \right) dv. \end{aligned}$$

By continuity it follows that, uniformly for $c_1 \leq c \leq c'_1$ and $w \geq w_0$, $\left| \int_0^1 \Phi(v) \exp\left(-i \frac{e^{-v^2/2}}{\sqrt{2\pi}\Phi(v)} vc\right) dv \right| \leq \int_0^1 \Phi(v) dv - \frac{c_2}{2}$ for some constant $c_2 > 0$. Hence $|f(z, w)| \leq \frac{1}{w} \int_0^w (\Phi(v) + \Phi(-v)) dv - \frac{c_2}{w} = 1 - \frac{c_2}{w}$, as proposed. (The case $w < 0$ is completely the same.) $\square$

Lemma 16. Suppose that $c_1 > 0$ is a given constant. Then

$$f(z, w) = 1 - \frac{k(z/w)}{w} + O\left(\frac{z}{w} e^{-w^2/2}\right) \quad (38)$$

uniformly for $|w| \geq 1$ and $|z|/|w| \geq c_1$, where

$$k(v) = \int_{-\infty}^{\infty} \Phi(u) \left(1 - \exp\left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v\right)\right) du$$

satisfies $k(v) = \sqrt{2\log|v|} + O(1)$ and $k'(v) = O(|v|^{-1})$ as $|v| \rightarrow \infty$. Furthermore we have the upper bound $\frac{\partial}{\partial z} f(z, w) = O\left(\frac{1}{|zw|}\right)$.

Proof. Due to symmetry we have

$$\begin{aligned}
f(z, w) &= \int_0^1 \left(\Phi(-xw) \exp \left(i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(-xw)} xz \right) + \Phi(xw) \exp \left(-i \frac{e^{-x^2 w^2/2}}{\sqrt{2\pi}\Phi(xw)} xz \right) \right) dx \\
&= 1 - \frac{1}{w} \int_0^w \Phi(-u) \left(1 - \exp \left(i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(-u)} u \frac{z}{w} \right) \right) du \\
&\quad - \frac{1}{w} \int_0^w \Phi(u) \left(1 - \exp \left(-i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} u \frac{z}{w} \right) \right) du.
\end{aligned}$$

Furthermore we have

$$\int_w^\infty \Phi(-u) \left| 1 - \exp \left(i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(-u)} u \frac{z}{w} \right) \right| du = O \left(\int_w^\infty \frac{e^{-u^2/2}}{u} du \right) = O(e^{-w^2/2})$$

and

$$\int_w^\infty \Phi(u) \left| 1 - \exp \left(-i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} u \frac{z}{w} \right) \right| du = O \left(\frac{z}{w} \int_w^\infty e^{-u^2/2} u du \right) = O \left(\frac{z}{w} e^{-w^2/2} \right).$$

Thus, (38) holds. Next we observe that (for $v \geq 1$)

$$\begin{aligned}
&\int_{-\infty}^1 \Phi(u) \left(1 - \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) \right) du = O(1), \\
&\int_1^{\sqrt{2\log v}} \Phi(u) \left(1 - \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) \right) du = \int_1^{\sqrt{2\log v}} \Phi(u) du - \int_1^{\sqrt{2\log v}} \Phi(u) \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) du \\
&\quad = \sqrt{2\log v} + O(1), \\
&\int_{\sqrt{2\log v}}^\infty \Phi(u) \left(1 - \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) \right) du = O \left(v \int_{\sqrt{2\log v}}^\infty e^{-u^2/2} u du \right) = O(1).
\end{aligned}$$

Note that $\left| \frac{\partial}{\partial u} \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right| \geq \frac{1}{20}$ for $1 \leq u \leq \sqrt{2\log v}$. Hence, we can apply Lemma 10 to get the upper bound for the corresponding integral, $\int_1^{\sqrt{2\log v}} \Phi(u) \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) du = O(1)$. This shows that $k(v) = \sqrt{2\log v} + O(1)$ for $v \geq 1$. The same arguments apply for $v \leq -1$.

Finally, we have to deal with

$$k'(v) = \frac{-i}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) e^{-u^2/2} u du,$$

where we again assume that $v \geq 1$. Since $\int_{\sqrt{2\log v}}^\infty e^{-u^2/2} u du = O(1/v)$ and, by Lemma 10, $\int_{-1}^1 \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) e^{-u^2/2} u du = O(1/v)$, it remains to consider the range $1 \leq |u| \leq \sqrt{2\log v}$. By another application of Lemma 10 it follows immediately that $\int_{-\sqrt{2\log v}}^{-1} \exp \left(i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} v \right) e^{-u^2/2} du = O(1/v)$. Here we use the property that $\Phi(u) \sim e^{-u^2/2}/(|u|\sqrt{2\pi})$ as $u \rightarrow -\infty$, so that $\left| \frac{\partial}{\partial u} \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right| \geq \frac{v}{20}$ for $-\sqrt{2\log v} \leq u \leq -1$. For the last integral we use the substitution $e^{-u^2/2} = U$ to obtain

$$\begin{aligned}
&\int_1^{\sqrt{2\log v}} \exp \left(i \frac{e^{-u^2/2} u}{\sqrt{2\pi}\Phi(u)} v \right) e^{-u^2/2} u du \\
&\quad = \int_{1/v}^{1/\sqrt{e}} \exp \left(i \frac{U \sqrt{2\log(1/U)}}{\sqrt{2\pi}\Phi(\sqrt{2\log(1/U)})} v \right) dU.
\end{aligned}$$

Since $|\frac{\partial}{\partial U} \frac{U\sqrt{2\log(1/U)}}{\sqrt{2\pi}\Phi(\sqrt{2\log(1/U)})}v| \geq \frac{v}{20}$ for $1/v \leq U \leq 1/\sqrt{e}$, we can apply Lemma 10 another time and obtain that this integral is $O(1/v)$, which completes the proof of the upper bound for $k'(v)$. The upper bound for the derivative $\frac{\partial}{\partial z}f(z, w)$ follows in the same way (we omit the details). $\square$

Lemma 17. *Suppose that $|z| \geq c_1|w|$ for some $c_1 > 0$. Then we have uniformly $\frac{\partial}{\partial z}B(w, z) = O(\frac{1}{w^3z})$.*

Proof. By using the substitution $u = xw$ we obtain the representation

$$B(w, z) = w^{-3} \int_{-w}^w \exp\left(-i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} u \frac{z}{w}\right) \left(\frac{e^{-u^2}}{2\pi\Phi(u)} + \frac{ue^{-u^2/2}}{\sqrt{2\pi}}\right) u^2 du$$

and consequently

$$\frac{\partial B(w, z)}{\partial z} = -iw^{-4} \int_{-w}^w \exp\left(-i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} u \frac{z}{w}\right) \left(\frac{ue^{-3u^2/2}}{(2\pi)^{3/2}\Phi(u)^2} + \frac{u^2e^{-u^2}}{2\pi\Phi(u)}\right) u^2 du.$$

With $v = z/w$, by Lemma 10 we certainly have

$$\int_{-1}^1 \exp\left(-i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} uv\right) \left(\frac{ue^{-3u^2/2}}{(2\pi)^{3/2}\Phi(u)^2} + \frac{u^2e^{-u^2}}{2\pi\Phi(u)}\right) u^2 du = O\left(\frac{1}{v}\right).$$

Next, by using the substitution $U = e^{-u^2/2}$ and again Lemma 10 we obtain

$$\begin{aligned} & \int_1^w \exp\left(-i \frac{e^{-u^2/2}}{\sqrt{2\pi}\Phi(u)} uv\right) \left(\frac{ue^{-3u^2/2}}{(2\pi)^{3/2}\Phi(u)^2} + \frac{u^2e^{-u^2}}{2\pi\Phi(u)}\right) u^2 du \\ &= \int_{e^{-w^2/2}}^{1/\sqrt{e}} \exp\left(-i \frac{U\sqrt{2\log(1/U)}}{\sqrt{2\pi}\Phi(\sqrt{2\log(1/U)})} v\right) \\ & \quad \times \left(\frac{2\log(1/U)U^2}{(2\pi)^{3/2}\Phi(\sqrt{2\log(1/U)})^2} + \frac{(2\log(1/U))^{3/2}U}{2\pi\Phi(\sqrt{2\log(1/U)})}\right) dU \\ &= O\left(\frac{1}{v}\right). \end{aligned}$$

Finally, a direct application of Lemma 10 (and the property $\Phi(u) \sim e^{-u^2/2}/(\sqrt{2\pi}|u|)$ as $u \rightarrow -\infty$) provides the same bound $O(1/v)$ for the integral over $[-w, -1]$ and completes the proof of the lemma. $\square$

We recall the representation $\mathbb{E}S_\varepsilon(X^T) = T \cdot J_0 + 2\varepsilon \cdot J_1$, where

$$J_1 = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz, \quad J_0 = \frac{1}{2\pi} \int_{\mathbb{R}^2} B(z, w) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz.$$

The main challenge in the computation of the integrals J_0 and J_1 are the parts that correspond to large w . We start with the integral J_0 .

A: The case $|w| \leq 1$. If $|w| \leq 1$ then we have the uniform bound $|f(z, w)| \leq C/|z|$ by Lemma 14. We first look at the case $|z| \leq 2C$. Here we certainly have the uniform representation (see Lemma 13) $f(z, w) = e^{-\frac{z^2}{2}TB(w)}(1 + O(z^3))$ and, by continuity, $|f(z, w)| \leq e^{-c_1 z^2}$ for some constant $c_1 > 0$. If $|z| \geq 2C$ then we have the trivial estimate $|f(z, w)| \leq C/|z| \leq 1/2$. Consequently,

$$\begin{aligned}
J_{0,1} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-1}^1 B(z, w) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz \\
&= \frac{1}{2\pi} \int_{-1}^1 \int_{|z| \leq T^{-1/3}} e^{-TB(w)\frac{z^2}{2}} (1 + O(z^2 + Tz^3)) e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\
&\quad + \frac{1}{2\pi} \int_{-1}^1 \int_{T^{-1/3} \leq |z| \leq 2C} O(e^{-c_1 T z^2}) e^{-\varepsilon w^2} dz dw + \frac{1}{2\pi} \int_{-1}^1 \int_{|z| \geq 2C} (C/|z|)^T e^{-\varepsilon w^2} dz dw \\
&= \int_{-1}^1 \frac{B(w)^{1/2}}{\sqrt{2\pi T}} e^{-2\varepsilon^2 w^2 / (TB(w))} e^{-\varepsilon w^2} dw + O(T^{-1}) + O(e^{-c_1 T^{1/3}}) + O(2^{-T}) \\
&= \frac{1}{\sqrt{2\pi T}} \int_{-1}^1 B(w)^{1/2} e^{-\varepsilon w^2} dw + O(T^{-1}).
\end{aligned}$$

B: The case $1 \leq |w| \leq \eta T$, where $\eta = \eta(T) = c_6(\log T)^{-3} \rightarrow 0$ and c_6 is sufficiently large. We divide the integral over z into several parts. The first part is the interval $|z| \leq z_1 = |w|^{3/2} T^{-1/2} \eta^{-1/6}$, where we use Lemma 13:

$$\begin{aligned}
f(z, w)^{T-1} &= e^{-TB(w)\frac{z^2}{2} + O(|z^2/w^3| + T|z^3/w^4|) + O(T|z^4/w^6|)} \\
&= e^{-TB(w)\frac{z^2}{2}} (1 + O(|z^2/w^3| + T|z^3/w^4|) + O(T|z^4/w^6|)).
\end{aligned}$$

By using the substitution $v = z(TB(w))^{1/2} = \Theta(zT^{1/2}w^{-3/2})$ we have (with $v_1 = z_1(TB(w))^{1/2} = \Theta(\eta^{-1/6})$)

$$\int_{|z| \leq z_1} e^{-TB(w)\frac{z^2}{2} - 2i\varepsilon w z} dz = \frac{\sqrt{2\pi}}{\sqrt{TB(w)}} e^{-2\varepsilon^2 w^2 / (TB(w))} + O\left(\frac{|w|^{3/2} \eta^{1/6}}{T^{1/2}} e^{-c\eta^{-1/3}}\right)$$

for some $c > 0$. Note that $c\eta^{-1/3} = cc_6 \log T$, so that $e^{-c\eta^{-1/3}} = T^{-cc_6}$ is an arbitrarily large negative power of T (when c_6 is sufficiently large). Similarly we obtain

$$\begin{aligned}
\int_{|z| \leq z_1} e^{-TB(w)\frac{z^2}{2}} |z^2/w^3| dz &= O\left(\frac{|w|^{3/2}}{T^{3/2}}\right), \quad \int_{|z| \leq z_1} e^{-TB(w)\frac{z^2}{2}} T|z^3/w^4| dz = O\left(\frac{|w|^2}{T}\right), \\
\int_{|z| \leq z_1} e^{-TB(w)\frac{z^2}{2}} T|z^4/w^6| dz &= O\left(\frac{|w|^{3/2}}{T^{3/2}}\right).
\end{aligned}$$

Summing up we find

$$\begin{aligned}
J_{0,21} &= \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{|z| \leq z_1} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz \\
&= \frac{1}{\sqrt{2\pi T}} \int_{|w| \geq 1} B(w)^{1/2} e^{-2\varepsilon^2 w^2 / (TB(w))} e^{-\varepsilon w^2} dw + O\left(T^{-1/2} \int_{\eta T}^{\infty} w^{-3/2} e^{-\varepsilon w^2} dw\right) \\
&\quad + O\left(T^{-3/2} \int_1^{\eta T} w^{-3/2} e^{-\varepsilon w^2} dw\right) + O\left(T^{-1} \int_1^{\eta T} w^{-1} e^{-\varepsilon w^2} dw\right) \\
&= \frac{1}{\sqrt{2\pi T}} \int_{|w| \geq 1} B(w)^{1/2} e^{-2\varepsilon^2 w^2 / (TB(w))} e^{-\varepsilon w^2} dw + O\left(\frac{\log T}{T}\right).
\end{aligned}$$

Next suppose that $z_1 = |w|^{3/2}T^{-1/2}\eta^{-1/6} \leq |z| \leq c_1|w|$, where c_1 is sufficiently small, which ensures that (see Lemma 13) $B(w) \frac{z^2}{2} \geq C\left(\frac{|z|^3}{|w|^4} + \frac{|z|^4}{|w|^6}\right)$. Recall that $B(w) = \Theta(|w|^{-3})$ for $|w| \geq 1$. Hence, there exists a constant $c > 0$ such that $|f(z, w)| \leq e^{-cz^2/|w|^3}$ uniformly for $z_1 \leq |z| \leq c_1|w|$. This implies that the corresponding integral is upper bounded by

$$\begin{aligned} J_{0,22} &= \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{z_1 \leq |z| \leq c_1|w|} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz \\ &= O\left(\frac{\eta^{1/6} e^{-c\eta^{-1/3}}}{T^{1/2}} \int_{1 \leq w \leq \eta T} w^{-3/2} e^{-\varepsilon w^2} dw\right) = O\left(\frac{\eta^{1/6} e^{-c\eta^{-1/3}}}{T^{1/2}}\right). \end{aligned}$$

Next suppose that $c_1|w| \leq |z| \leq c'_1|w|$ for an arbitrary constant $c'_1 > 1$. By Lemma 15 we have the upper bound $|f(z, w)| \leq 1 - c_2/|w| \leq e^{-c_2/|w|}$, and so we have (provided that $(T/\varepsilon)^{1/3} \leq \eta T$, or equivalently $\varepsilon \geq c_6^{-3}(\log T)^9/T^2$)

$$\begin{aligned} J_{0,23} &= \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{c_1|w| \leq |z| \leq c'_1|w|} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{1 \leq w \leq \eta T} w^{-2} e^{-c_2 T/|w| - \varepsilon w^2} dw\right) \\ &= O\left(\int_{1 \leq w \leq (T/\varepsilon)^{1/3}} w^{-2} e^{-c_2 T/w} dw\right) + O\left(\int_{(T/\varepsilon)^{1/3} \leq w \leq \eta T} e^{-\frac{\varepsilon}{2} w^2} dw\right) \\ &= O(e^{-c_2 T^{2/3} \varepsilon^{1/3}}) = O\left(\frac{1}{T}\right) \end{aligned}$$

if c_6 is sufficiently large.

Finally suppose that $|z| \geq c_1|w|$ (for some $c_1 > 1$). This range is split up into $c_1|w| \leq |z| \leq e^{|w|^3}$ and $|z| \geq e^{|w|^3}$. In the first case we use (26) for the function $H(z) = B(w, z) f(w, z)^{T-1}$ and the interval $I = [c_1|w|, e^{|w|^3}]$. Clearly we have $\max_{z \in I} |B(w, z) f(w, z)^{T-1}| = O(w^{-3} e^{-cT/|w|})$ for some constant $c > 0$. For the derivative we have the upper bound

$$\begin{aligned} |H'(z)| &\leq |B_z| |f|^{T-1} + |B| (T-1) |f|^{T-2} |f_z| \\ &= O\left(\frac{e^{-cT/|w|}}{w^3 z}\right) + O\left(\frac{T}{w^3} e^{-c'T \frac{\sqrt{2 \log |z/w|}}{|w|}} \frac{\sqrt{2 \log |z/w|}}{\sqrt{2 \log |z/w|} zw}\right) \end{aligned}$$

(for some constants $c, c' > 0$), which leads to

$$\int_{c_1|w|}^{e^{|w|^3}} |H'(z)| dz = O\left(\frac{e^{-cT/|w|}}{w^3} \log(e^{|w|^3})\right) + O\left(\frac{T}{w^3} e^{-c'T \frac{\sqrt{2 \log c_1}}{|w|}} \sqrt{2 \log e^{|w|^3}}\right) = O(e^{-cT/|w|}).$$

Thus, we obtain

$$\begin{aligned} J_{0,24} &= \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{c_1|w|}^{e^{|w|^3}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_1^{\eta T} e^{-cT/w - \varepsilon w^2} dw\right) \\ &= O\left(\int_1^{(T/\varepsilon)^{1/3}} e^{-cT^{2/3} \varepsilon^{1/3}} dw + \int_{(T/\varepsilon)^{1/3}}^{\eta T} e^{-\varepsilon w^2} dw\right) \\ &= O(e^{-(c/2)T^{2/3} \varepsilon^{1/3}}) = O\left(\frac{1}{T}\right) \end{aligned}$$

as long as $\varepsilon \geq c''(\log T)^3 T^{-2}$ for a proper constant $c'' > 0$. In the second case we get $|f(z, w)| \leq C_2 e^{w^2/2}/|z|$ and

$$\int_{e^{|w|^3}}^{\infty} |B(w, z) f(w, z)^{T-1}| dz = O\left(\frac{|w|^{-3}}{T} e^{(T-1)w^2/2 - (T-2)|w|^3}\right) = O\left(\frac{1}{T} e^{-T|w|^3/2}\right),$$

and consequently

$$J_{0,25} = \frac{1}{2\pi} \int_{1 \leq |w| \leq \eta T} \int_{|z| \geq e^{|w|^3}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw = O\left(\frac{1}{T}\right).$$

C: The case $|w| \geq \eta T$. We partition the integration with respect to z into three ranges: $|z| \leq c_1|w|$, $c_1|w| \leq |z| \leq e^{|w|^3}$, and $|z| \geq e^{|w|^3}$. In the first case we use the trivial estimate $Bf^{T-1} = O(w^{-3})$ and obtain

$$\begin{aligned} J_{0,31} &= \frac{1}{2\pi} \int_{|w| \geq \eta T} \int_{|z| \leq c_1|w|} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{\eta T}^{\infty} w^{-2} e^{-\varepsilon w^2} dw\right) \\ &= O\left(\frac{1}{\eta^3 T^3} e^{-\varepsilon \eta^2 T^2}\right) = O(e^{-c(\log T)^3}). \end{aligned}$$

For the second case we set $H(z) = B(w, z) f(w, z)^{T-1}$ and use the property that $\max_{c_1|w| \leq z \leq e^{|w|^3}} |H(z)| + \int_{c_1|w|}^{e^{|w|^3}} |H'(z)| dz = O(e^{-cT/|w|})$ for some $c > 0$, which leads to $\int_{c_1|w|}^{e^{|w|^3}} B(w, z) f(w, z)^{T-1} e^{-2i\varepsilon w z} dz = O\left(\frac{1}{\varepsilon|w|} e^{-cT/|w|}\right)$. Thus,

$$\begin{aligned} J_{0,32} &= \frac{1}{2\pi} \int_{|w| \geq \eta T} \int_{c_1|w| \leq |z| \leq e^{|w|^3}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{\eta T}^{\infty} \frac{1}{\varepsilon|w|} e^{-cT/|w|} e^{-\varepsilon w^2} dw\right) = O\left(\frac{1}{\varepsilon \eta T} e^{-\varepsilon \eta^2 T^2}\right) = O(e^{-c(\log T)^3}). \end{aligned}$$

Finally, for $|z| \geq e^{|w|^3}$ we use (as above) the upper bound $|f(w, z)| \leq C_2 e^{w^2/2}/|z|$ and the resulting bound $\int_{e^{|w|^3}}^{\infty} |B(w, z) f(w, z)^{T-1}| dz = O\left(\frac{1}{T} e^{-T|w|^3/2}\right)$ to obtain

$$\begin{aligned} J_{0,33} &= \frac{1}{2\pi} \int_{|w| \geq \eta T} \int_{|z| \geq e^{|w|^3}} B(w, z) f(w, z)^{T-1} e^{-\varepsilon w^2 - 2i\varepsilon w z} dz dw \\ &= O\left(\int_{\eta T}^{\infty} \frac{1}{T} e^{-T|w|^3/2} e^{-\varepsilon w^2} dw\right) = O(e^{-\eta^3 T^4/2}), \end{aligned}$$

which is certainly negligible. This completes the proof. $\square$