Precise Regularized Minimax Regret with Unbounded Weights

Michael Drmota, Philippe Jacquet, Changlong Wu and Wojciech Szpankowski

Abstract

In online learning, a learner receives data in rounds and, at each round, predicts a label that is then compared to the true label, incurring a loss. The total loss over T rounds, when compared to the loss of the best expert from a class of experts or forecasters, is called the regret. In this paper, we focus on logarithmic loss for logistic-like experts with *unbounded* d-dimensional weights, a scenario that has been largely unexplored. To address the irregularities introduced by the unbounded weight norm, we introduce a *regularized* version of the average (fixed design) minimax regret by imposing a *soft constraint* on the weight norm. We demonstrate that the regularized minimax regret is fully characterized by a complexity measure we term the regularized Shtarkov sum. We also show how the behavior of the standard regret can be inferred from the regularized regret. Our main results provide a *precise* characterization of the regularized Shtarkov sum and, consequently, the regularized regret with unbounded weights up to second-order asymptotics. Notably, unlike the $d/2 \log T$ regret growth known for bounded weights, our results imply that the regularized regret grows as $(1/2 + \alpha/4)d \log T$ when the regularization parameter is of order $\Theta(T^{-\alpha})$ for $\alpha \leq 1/2$. We achieve this using tools from analytic combinatorics, including multidimensional Fourier analysis, the saddle point method, and the Mellin transform.

I. Introduction

The problem of online learning under logarithmic loss and its regret analysis has been intensively studied over the last decade [1], [2], [3], [4], [5]. However, even for logistic regression, there is a lack of precise second-order asymptotics (especially for unbounded weights), with a possible exception of [6] which is restricted to categorical data. In this paper, we initiate the study to fill this gap.

To set the stage of 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 $\mathbf{x}_t \in \mathbb{R}^d$ 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 $\ell : \hat{\mathcal{Y}} \times \mathcal{Y} \to \mathbb{R}^+$ 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 defined as

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

where $\hat{y}_t = g_t(y^{t-1}, \mathbf{x}^t)$ is the prediction based on prior observations y^{t-1} and $\mathbf{x}^t$. Throughout, we write $y^t = (y_1, \ldots, y_t)$ and $\mathbf{x}^t = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$ for $t \in [T]$. Note that the prediction rule g^T need not necessarily come from $\mathcal{H}$, which is also known as *improper* learning in the literature [3].

In this paper, we focus specifically on the *logarithmic loss* defined as:

$$\ell(\hat{y}_t, y_t) = -\frac{1 + y_t}{2} \log(\hat{y}_t) - \frac{1 - y_t}{2} \log(1 - \hat{y}_t).$$

We fix a function $p(w)$ from $\mathbb{R} \to [0,1]$ and restrict our study to the class of experts:

$$\mathcal{H}_p = \{h_{\mathbf{w}} : \mathbf{w} \in \mathbb{R}^d\}, \tag{1}$$

M. Drmota (email: drmota@tuwien.ac.at) is with TU Wien, Austria. P. Jacquet (email: philippe.jacquet@inria.fr) is with INRIA, France. C. Wu (email:wuchangl@hawaii.edu) and W. Szpankowski (email: szpan@purdue.edu) are with Purdue University, USA.

This work was partially supported by the NSF Center for Science of Information (CSoI) Grant CCF-0939370, by NSF Grants CCF-2006440, CCF- 2211423, and in addition by Google Research Grant.

where $h_{\mathbf{w}}$ (as a function from $\mathbb{R}^d \to \mathbb{R}$) is defined by

$$h_{\mathbf{w}}(\mathbf{x}) := p(\langle \mathbf{w}, \mathbf{x} \rangle), \text{ for } \mathbf{x} \in \mathbb{R}^d$$

and $\langle \mathbf{w}, \mathbf{x} \rangle$ is the scalar product of $\mathbf{x}$ and $\mathbf{w}$. Hereafter, we focus on the logistic regression with $p(w) = (1 + \exp(-w))^{-1}$ [7], [8], since its precise analysis is the key step to analyze a larger class of functions and other losses. In the last Section V, we discuss the quantum tomography problem with a different loss function departing from the classical logistic regression, however, our methodology still applies.

While we assume that $\mathbf{x}_t$ lies on a compact manifold $\mathcal{M}_{\mathbf{x}}$ (e.g., $\mathcal{M}_{\mathbf{x}} = [-1, 1]^d$, the unit ball $\mathcal{B}_d$, or the sphere $\mathcal{S}_d$), we do *not* bound the weights $\mathbf{w} \in \mathbb{R}^d$, and this seems to have never been analyzed in depth, to the best of our knowledge. Specifically, we assume that $\|\mathbf{w}\| \leq R \leq \infty$, where R can grow with T.

We are interested in the *fixed design* regret where the feature vector $\mathbf{x}^T$ is known in advance. Specifically, for any given $\mathcal{H}$ and $\mathbf{x}^T$, the *fixed design* minimax regret is defined as

$$r_T(\mathcal{H}|\mathbf{x}^T) := \inf_{g^T} \sup_{y^T} \mathcal{R}(g^T, y^T, \mathcal{H}|\mathbf{x}^T). \tag{2}$$

When the class $\mathcal{H}$ is clear, as in our case, we simply write $r_T^R(\mathbf{x}^T) := r_T(\mathcal{H}|\mathbf{x}^T)$. This notion was also known in the literature as *transductive online learning* [9]. To decouple it from the feature vector $\mathbf{x}^T$, one either maximizes over all possible $\mathbf{x}^T$ or takes the average over the features. We study here the *averaged* fixed design minimax regret as

$$\bar{r}_T(\mathcal{H}) := \mathbb{E}_{\mathbf{x}^T}[r_T(\mathcal{H}|\mathbf{x}^T)],$$

where the feature vector $\mathbf{x}^T$ is generated by an i.i.d. process. The importance of fixed-design regret lies in the fact that it is a universal lower bound for various regrets discussed in the literature [4], [10].

As discussed in [6], [11] the minimax regret $r_T(\mathbf{x}^T)$ can be studied through the so called Shtarkov sum which for bounded $\|\mathbf{w}\| \leq R$ becomes

$$S_R(\mathbf{x}^T) = \sum_{y^T} \sup_{||\mathbf{w}|| \leq R} P(y^T|\mathbf{x}^T, \mathbf{w}) \tag{3}$$

where $P(y^T|\mathbf{x}^T, \mathbf{w}) = \prod_{t=1}^T p(y_t\langle \mathbf{x}_t, \mathbf{w} \rangle)$, and the regret is then $r_T^R(\mathbf{x}^T) = \log S_R(\mathbf{x}^T)$. While the Shtarkov sum approach provides an exact solution, there are two main issues: computational and analytical. The optimization problem $\sup_{\|\mathbf{w}\| \leq R} P(y^T|\mathbf{x}^T, \mathbf{w})$ is non-convex and, more problematically, most of the optimal solutions

$$\mathbf{w}^* = \arg \sup_{||\mathbf{w}|| \leq R} P(y^T|\mathbf{x}^T, \mathbf{w}) \tag{4}$$

lie on the boundary $\|\mathbf{w}\| = R$. To address these issues, one often resorts to regularization (see [12], [13]).

In view of these challenges, we introduce and study a *regularized* version of the minimax regret. We first notice that for the logarithmic loss function we can write $\ell(\hat{y}, y) = -\log P(y|\hat{y})$ and $\ell(h_{\mathbf{w}}(\mathbf{x}), y) = -\log P(y|\mathbf{x}, \mathbf{w})$, leading to the *regularized* pointwise regret

$$\mathcal{R}_T^{\varepsilon}(\hat{y}^T, y^T|\mathbf{x}^T) = -\sum_{t=1}^T \log P(y_t|\hat{y}_t) + \sup_{\mathbf{w} \in \mathbb{R}^d} \sum_{t=1}^T \log P(y_t|\mathbf{x}_t, \mathbf{w}) e^{-\varepsilon ||\mathbf{w}||^2} \tag{5}$$

where $\varepsilon \geq 0$ and $\sup_{\mathbf{w} \in \mathbb{R}^d}$ is *unconstrained*. Then, the regularized minimax regrets are defined as

$$r_T^{\varepsilon}(\mathbf{x}^T) = \inf_{\hat{y}^T} \max_{y^T} \mathcal{R}_T^{\varepsilon}(\hat{y}^T, y^T|\mathbf{x}^T), \qquad \bar{r}_T^{\varepsilon}(\mathcal{H}) := \mathbb{E}_{\mathbf{x}^T}[r_T^{\varepsilon}(\mathcal{H}|\mathbf{x}^T)] \tag{6}$$

and the *generalized* Shtarkov sum is

$$S_{\varepsilon}(\mathbf{x}^T) = \sum_{y^T} \sup_{\mathbf{w}} P(y^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon ||\mathbf{w}||^2}. \tag{7}$$

We show in Section II that

$$r_T^{\varepsilon}(\mathbf{x}^T) = \log S_{\varepsilon}(\mathbf{x}^T)$$

holds as well. Notice that the optimization

$$\mathbf{w}_{\varepsilon}^* := \arg_{\mathbf{w} \in \mathbb{R}^d} P(y^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon ||\mathbf{w}||^2}$$

is much easier to compute since it is *log-concave* and $||\mathbf{w}_\varepsilon^*|| < \infty$ holds always. However, unlike the regularization introduced in [12], [13], here we are not aiming to control the computational complexity. Instead, the introduction of regularization is primarily for the sake of analytical considerations (i.e., allowing the existence of the Fourier transform, as detailed in Section II).

Furthermore, our regularized regret can be interpreted as a *soft-constraint* on $R = ||\mathbf{w}||$ with $R \sim 1/\sqrt{\varepsilon}$. In fact, in Lemma 1 we prove for all $\mathbf{x}^T$ the following relations between standard (classical) regret $r_T^R(\mathbf{x}^T)$ for $\|\mathbf{w}\| \le R$ and the regularized regret $r_T^\varepsilon(\mathbf{x}^T)$

$$r_T^R(\mathbf{x}^T) \le r_T^\varepsilon(\mathbf{x}^T) \le r_T^{R\sqrt{d\log T}}(\mathbf{x}^T) + O(1) \tag{8}$$

for $\varepsilon = 1/R^2$. This demonstrates that the standard regret and the regularized regret grow asymptotically the same order for polynomially growing R, giving us another justification to study precisely the regularized regret and the associated Shtarkov sum.

A. Related Work

Online learning under logarithmic loss can be viewed as universal compression (source coding) with side information, as discussed in [14], [15] and [16], [17], [18], [19], [20]. The logistic-type class of experts, as in (1), was studied extensively in [3], [7], [2], [8], [4] under various formulations of regret. In particular, it is known that for any range R of the weight $\mathbf{w}$, the minimax regret can be upper bounded by $(d/2)\log\big(TR^2/d\big)$ for the *sequential* regret, i.e., where both $\mathbf{x}^T$ and y^T are selected sequentially [3], [8], [4]. For the *fixed design* regret we study here, the precise dependency on the weight norm R is largely unexplored. Several prior results, such as [8], [4], [21], have demonstrated that the regret lower bound grows as $(d/2)\log\big(T/d^2\big)$ (with no dependency on R), which can deviate arbitrarily from the generic $(d/2)\log\big(TR^2/d\big)$ upper bound for $R \to \infty$. Recently, [22] showed that for *fixed design* regret, the upper bound can be improved to $2d\log T$ for a general *monotone class* even with $R = \infty$. The authors of [22] also demonstrated that for $R = \infty$, the $(1 + o(1))d\log(T/d)$ regret holds for logistic regression. This leaves open the question of *precisely* characterizing the fixed design regret in the transition region of the regret from $(d/2)\log T$ to $d\log T$ as $R \to \infty$. In this paper we provide an answer when $R = o(T^{1/4})$ for the regularized and classical regret. To the best of our knowledge, [22] is the only work that studies the precise characterization of fixed design regret with unbounded weights (although recently [23] studied unbounded weights but not its precise behavior). We should emphasize that studying the transition region poses substantial technical challenges if a precise characterization is desired (i.e., precise up to the *second* order asymptotic). Our findings are most closely related to [11], [6]. In [6], a precise maximal minimax regret is analyzed, but only for a *finite* number of feature values (see also [15]).

B. Summary of Contributions

In this paper, we present for the first time precise second-order asymptotics for the regularized Shtarkov sum and consequently the regularized minimax regret as in (6) with $\varepsilon > 0$. While our derivations do not directly work for $\varepsilon = 0$, we extend our findings to $\varepsilon \to 0$ as long as $\varepsilon \gg T^{-1/2}$ (which means that there exists a sufficiently large positive constant c such that $\varepsilon \ge cT^{-1/2}$) showing a phase transition of the leading term of of the (average) Shtarkov sum and the corresponding one for the regret. We should emphasize that to analyze such a phase transition, we need precise expression for the second-order terms. This result also shed lights on the classical minimax regret as shown in (27) when the weight norm R grows as $O(T^{1/4})$.

More precisely, we represent first in (9) the regularized minimax regret as the logarithm of the generalized Shtarkov sum. Then in Theorem 1 we present precise second-order asymptotic expansion of the average Shtarkov sum and hence the regularized minimax regret for logistic regression with *unbounded* weights. We prove that for $\varepsilon \gg T^{-1/2}$ the regularized minimax regret grows as $(d/2)\log(2T/\pi) + \log C_d(\varepsilon)$ where $C_d(\varepsilon)$ has a complicated multidimensional integral which we explicitly evaluate for $\varepsilon \to 0$ in Theorem 2. Furthermore, for $\varepsilon = \Theta(T^{-\alpha})$ with $\alpha \le 1/2$, we show in Theorem 2 that the leading term grows as $((1/2 + \alpha/4)d - \alpha/2)\log T$ for regularized and classical regret. We also conjecture it reaches $d\log T$ for $\varepsilon \sim 1/T^2$. We accomplish it using powerful analytic techniques[1] such as saddle point method, Mellin transform, and multidimensional Fourier transform (see [24], [25]),

[1]A. Odlyzko argued: "Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision."

hopefully initiating an analytic learning theory (see [16]) in which problems of machine learning are solved by tools of complex analysis.[2]

We present here analysis for the simplest model for the logical regression and uniform distribution for $\mathbf{x}$. However, we must emphasize that the hardest challenge is to find the right approach for the simplest case and then bring more technical approaches to generalize. This is particularly true for analytic techniques as witnessed by regret analysis in information theory: the Shtarkov sum was first precisely analyzed for a memoryless source [26], then extended to Markov sources [17] and finally to renewal processes [27].

II. Main Results

In this section we present our main results with most proofs delegated to the next two sections and the Appendix. Before we start our discussion, we derive the connection between the regularized regret (6) and generalized Shtarkov sum (7). Note that, for any given $\mathbf{x}^T$, the predictor $\hat{y}_t$ can be compactly represented as a distribution Q over $\{-1,+1\}^T$ such that $\hat{y}_t = Q(+1|y^{t-1}, \mathbf{x}^T)$ and $\ell(\hat{y}_t, y_t) = -\log Q(y_t|y^{t-1}, \mathbf{x}^T)$. Then the regularized regret can be written as

$$\begin{aligned} r_T^{\varepsilon}(\mathcal{H}|\mathbf{x}^T) &= \min_Q \max_{y^T}[-\log Q(y^T|\mathbf{x}^T) + \sup_{\mathbf{w}} \log P(y^T|\mathbf{x}^T) e^{-\varepsilon||\mathbf{w}||^2}] \\ &= \min_Q \max_{y^T}[-\log Q(y^T|\mathbf{x}^T) + \log P_{\varepsilon}^*(y^T|\mathbf{x}^T)] + \log \sum_{v^T} \sup_{\mathbf{w}} P(v^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} \\ &\overset{(a)}{=} \log \sum_{y^T} \sup_{\mathbf{w}} P(y^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} = \log S_{\varepsilon}(\mathbf{x}^T) \end{aligned} \tag{9}$$

where (a) follows since $\min_Q$ is attained when $Q = P_{\varepsilon}^*$, and $S_{\varepsilon}(\mathbf{x}^T)$ is defined in (7) and

$$P_{\varepsilon}^*(y^T|\mathbf{x}^T) := \frac{\sup_{\mathbf{w}} P(y^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2}}{\sum_{v^T} \sup_{\mathbf{w}} P(v^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2}}$$

is the *generalized* maximum-likelihood distribution, We note that the regularized minimax regret can also be achieved *precisely* by the following "regularized" Normalized Maximum Likelihood (NML) predictor:

$$\hat{p}_t := P_{\epsilon}^*(y_t \mid y^{t-1}, \mathbf{x}^T) = \frac{\sum_{v^{T-t}} P_{\epsilon}^*(v^{T-t} y^t \mid \mathbf{x}^T)}{\sum_{v^{T-t+1}} P_{\epsilon}^*(v^{T-t+1} y^{t-1} \mid \mathbf{x}^T)}. \tag{10}$$

Our objective is then to find precise asymptotics for the generalized Shtarkov sum $S_{\varepsilon}(\mathbf{x}^T)$ as defined in (7). Note that for a sequence of labels y^T and a sequence of features $\mathbf{x}^T$ we have for any $\varepsilon > 0$

$$P(y^T|\mathbf{x}^T, \mathbf{w}) = \prod_{t=1}^{T} p(y_t\langle \mathbf{x}_t, \mathbf{w}\rangle) \qquad \text{and} \qquad P_{\varepsilon}(y^T|\mathbf{x}^T, \mathbf{w}) = \prod_{t=1}^{T} p(y_t\langle \mathbf{x}_t, \mathbf{w}\rangle) e^{-\varepsilon\|\mathbf{w}\|^2}. \tag{11}$$

We also define $L(y^T|\mathbf{x}^T, \mathbf{w}) = \log P(y^T|\mathbf{x}^T, \mathbf{w})$ and $L_{\varepsilon}(y^T|\mathbf{x}^T, \mathbf{w}) = L(y^T|\mathbf{x}^T, \mathbf{w}) + \varepsilon\|\mathbf{w}\|^2$.

Before we proceed, let us discuss a relation between the regularized Shtarkov sum $S_{\varepsilon}(\mathbf{x}^T)$ as well as regularized regret $r_T^{\varepsilon}(\mathbf{x}^T)$ and the standard Shtarkov $S_R(\mathbf{x}^T)$ as well as the standard regret that we write as $r_T^R(\mathbf{x}^T)$ when $||\mathbf{w}|| \leq R$. Formally, we have the following result.

Lemma 1. *For all $\mathbf{x}^T$, $\varepsilon > 0$ and $R_1, R_2 \geq 0$, we have*

$$e^{-\varepsilon R_1^2} S_{R_1}(\mathbf{x}^T) \leq S_{\varepsilon}(\mathbf{x}^T) \leq S_{R_2}(\mathbf{x}^T) + S_{\infty}(\mathbf{x}^T) e^{-\varepsilon R_2^2}. \tag{12}$$

In particular, if $\log S_{\infty}(\mathbf{x}^T) = O(d \log T)$ and $\varepsilon = 1/R^2$ we find

$$r_T^R(\mathbf{x}^T) \leq r_T^{\varepsilon}(\mathbf{x}^T) \leq r_T^{R\sqrt{d\log T}}(\mathbf{x}^T) + O(1). \tag{13}$$

[2]Following Handmaid's percept: “The shortest paths between two truths on the real line passes through the complex plane."

Proof. The lower bound of (12) follows from

$$\begin{aligned} S_\varepsilon(\mathbf{x}^T) &= \sum_{y^T} \sup_{\mathbf{w}} P(y^T|\mathbf{x}^T,\mathbf{w}) e^{-\varepsilon\|\mathbf{w}\|^2} \\ &\geq \sum_{y^T} \sup_{\|\mathbf{w}\|\leq R_1} P(y^T|\mathbf{x}^T,\mathbf{w}) e^{-\varepsilon\|\mathbf{w}\|^2} \\ &\geq \sum_{y^T} \sup_{\|\mathbf{w}\|\leq R_1} P(y^T|\mathbf{x}^T,\mathbf{w}) e^{-\varepsilon R_1^2} = e^{-\varepsilon R_1^2} S_{R_1}(\mathbf{x}^T). \end{aligned}$$

For the upper bound of (12), we have

$$\begin{aligned} S_\varepsilon(\mathbf{x}^T) &= \sum_{y^T} \sup_{\mathbf{w}} P(y^T|\mathbf{x}^T,\mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} \\ &\leq \sum_{y^T} \sup_{||\mathbf{w}||\leq R_2} P(y^T|\mathbf{x}^T,\mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} + \sum_{y^T} \sup_{R_2<||\mathbf{w}||<\infty} P(y^T|\mathbf{x}^T,\mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} \\ &\leq \sum_{y^T} \sup_{||\mathbf{w}||\leq R_2} P(y^T|\mathbf{x}^T,\mathbf{w}) + e^{-\varepsilon R_2^2} \sum_{y^T} \sup_{|\mathbf{w}||<\infty} P(y^T|\mathbf{x}^T,\mathbf{w}) \\ &= S_{R_2}(\mathbf{x}^T) + S_\infty(\mathbf{x}^T) e^{-\varepsilon R_2^2}. \end{aligned}$$

Taking $\varepsilon = 1/R^2$, $R_1 = R$ and $R_2 = R\sqrt{d\log T}$, we have $e^{-\varepsilon R_1^2} = e^{-1}$ and $e^{-\varepsilon R_2^2} = e^{d\log T}$. The inequality (13) then follows by taking logarithm on both side of (12) and using the fact that $\log S_\infty(\mathbf{x}^T) = O(d\log T)$. □

It was shown in [22, Theorem 2] that for any expert class with a monotone p (including logistic regression), we have $\log S_\infty(\mathbf{x}^T) \leq 2d\log T$. Therefore, (13) implies that

$$r_T^{\epsilon_1}(\mathbf{x}^T) \leq r_T^{R}(\mathbf{x}^T) \leq r_T^{\epsilon_2}(\mathbf{x}^T),$$

by taking

$$\epsilon_1 = \frac{d\log T}{R^2}, \quad \text{and} \quad \epsilon_2 = \frac{1}{R^2}.$$

Thus the $d\log T$ factor contributes only to the *lower-order term* of $r_T^{\epsilon_1}(\mathbf{x}^T)$ as R grows polynomially w.r.t T. Therefore, the regret $r_T^R(\mathbf{x}^T)$ (with a polynomial growth R) can be converted to the *regularized* regret $r_T^{\epsilon}(\mathbf{x}^T)$ with the *same* leading constant as long as $\epsilon = \frac{1}{R^2}$.

We now focus on the logistic regression $p(w) = (1+\exp(-w))^{-1}$ since all interesting behavioral phenomena occur for this function, and its analysis is the key to a general case. For the logistic function we have

$$\nabla L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w}) = -\sum_{t=1}^{T} p(-y_t\langle\mathbf{x}_t,\mathbf{w}\rangle)\, y_t\, \mathbf{x}_t + 2\varepsilon\mathbf{w}$$

and

$$\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w}) = \sum_{t=1}^{T} p(\langle\mathbf{x}_t,\mathbf{w}\rangle)) p(-\langle\mathbf{x}_t,\mathbf{w}\rangle))\, \mathbf{x}_t \otimes \mathbf{x}_t + 2\varepsilon\mathbf{I}, \tag{14}$$

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

To study the Shtarkov sum, and ultimately the minimax regret, we need a better understanding of of the optimal $\mathbf{w}_\varepsilon^*$ defined as

$$\mathbf{w}_\varepsilon^* = \arg\min_{\mathbf{w}\in\mathbb{R}^d} L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})$$

which is the (unique) solution of the equation $\nabla L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w}) = \mathbf{0}$. Similarly, for every $\mathbf{a}\in\mathbb{R}^d$ the equation

$$\nabla L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w}) = \mathbf{a} \tag{15}$$

has a unique solution $\mathbf{w}_\varepsilon^*(\mathbf{a})$. Furthermore, if we denote $G_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{w}) := \nabla L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})$, then we have $\mathbf{w}_\varepsilon^* = G_{y^T|\mathbf{x}^T,\varepsilon}^{-1}(\mathbf{0})$. Note that $\mathbf{w}^*$ can be infinite, but $\mathbf{w}_\varepsilon^* < \infty$.

We shall analyze the Shtarkov sum via a multidimensional Fourier transform method. The first key issue is its existence, which we address next and Appendix A. We set $h_{y^T|\mathbf{x}^T}(\mathbf{a}) = \exp\left(-L_\varepsilon(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right)$. The goal is to show that (for every y^T and $\mathbf{x}^T$) the Fourier transform

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

of $h_{y^T|\mathbf{x}^T}(\mathbf{a})$ exists and that the inverse Fourier transform has an absolute convergent integral representation

$$h_{y^T|\mathbf{x}^T}(\mathbf{a}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z}) e^{i\langle \mathbf{a},\mathbf{y}\rangle}\, d\mathbf{z}.$$

Actually we will need $h_{y^T|\mathbf{x}^T}(\mathbf{0})$ since

$$\sup_{\mathbf{w}\in\mathbb{R}^d} P_\varepsilon(y^T|\mathbf{x}^T, \mathbf{w}) = P_\varepsilon(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{0})) = h_{y^T|\mathbf{x}^T}(\mathbf{0}).$$

Observe that by (15) (for $\mathbf{w} = G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a})$) we have

$$\mathbf{a} = -\sum_{t=1}^{T} p\left(-y_t\langle \mathbf{x}_t, \mathbf{w}\rangle\right) y_t \mathbf{x}_t + 2\varepsilon\mathbf{w} = O(1) + 2\varepsilon\mathbf{w}.$$

Note that the $O(1)$-term depends on y^T and $\mathbf{x}^T$. Thus $G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}) = \frac{1}{2\varepsilon}\mathbf{a} + O(1)$ which directly implies that

$$h_{y^T|\mathbf{x}^T}(\mathbf{a}) = O\left(e^{-\frac{1}{4\varepsilon}\|\mathbf{a}\|^2}\right). \tag{16}$$

Hence the Fourier transform $\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z})$ certainly exists since (16) implies absolute convergence of the corresponding integral. Furthermore, we establish in Appendix A the upper bound

$$\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z}) = O\left(|z_1|^{-k_1} \cdots |z_d|^{-k_d}\right)$$

for all non-negative integers $k_1, \ldots, k_d$ which implies that the inverse Fourier transform is given by

$$h_{y^T|\mathbf{x}^T}(\mathbf{a}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z}) e^{i\langle \mathbf{a},\mathbf{z}\rangle}\, d\mathbf{z}.$$

Consequently we can proceed to evaluate the Shtarkov sum as follows:

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

where we have used the substitution $\mathbf{a} = G_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{w}) = \nabla L_\varepsilon(y^T|\mathbf{x}^T, \mathbf{w})$. To complete our derivation, we observe that

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

where $f(\mathbf{w}, \mathbf{x}, \mathbf{z})$ denotes

$$f(\mathbf{w}, \mathbf{x}, \mathbf{z}) = p(\langle \mathbf{x}, \mathbf{w}\rangle) e^{-ip(-\langle \mathbf{x},\mathbf{w}\rangle)\,\langle \mathbf{x},\mathbf{z}\rangle} + p(-\langle \mathbf{x}, \mathbf{w}\rangle) e^{ip(\langle \mathbf{x},\mathbf{w}\rangle)\,\langle \mathbf{x},\mathbf{z}\rangle}. \tag{17}$$

This leads to our integral representation of the Shtarkov sum

$$S_\varepsilon(\mathbf{x}^T) = (2\pi)^{-d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \prod_{t=1}^{T} f(\mathbf{w}, \mathbf{x}_t, \mathbf{z})\, e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w},\mathbf{z}\rangle} \det\left(\nabla^2 L_\varepsilon(\cdot|\mathbf{x}^T, \mathbf{w})\right)\, d\mathbf{w}\, d\mathbf{z}. \tag{18}$$

We now assume that the vectors $\mathbf{x}_1,\ldots,\mathbf{x}_T$ are iid random vectors $\mathbf{X}_1,\ldots,\mathbf{X}_T$ that follow a probability distribution over bounded support, however, this can be relaxed.[3] Furthermore since P and $\nabla^2 L_\varepsilon$ are bounded in a bounded domain it follows that

$$\mathbb{E}S_\varepsilon(\mathbf{X}^T) = (2\pi)^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}\left[\prod_{t=1}^T f(\mathbf{w},\mathbf{X}_t,\mathbf{z})\det\left(\nabla^2 L_\varepsilon(\cdot|\mathbf{X}^T,\mathbf{w})\right)\right]e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{z}\rangle}\,d\mathbf{w}\,d\mathbf{z}. \tag{19}$$

This expression is the main tool that we will use to study asymptotically the minimax regret. The asymptotic evaluation of (19) is very challenging as we will see in Sections III and in Section IV. We also want to note that in the representation (18) the integral cannot be sees as a double multi-dimensional integral. The representation there is only correct if the integration with respect to $\mathbf{w}$ is done first and the integration with respect to $\mathbf{z}$ in a second step. This is due to the fact that this representation depends on the the inversion of the Fourier transform. For example, by formally exchanging the integration we would have to calculate linear combinations of integrals of the form $\int_{\mathbb{R}^d} e^{ih(\mathbf{w},\mathbf{x}^T)\langle\mathbf{w},\mathbf{z}\rangle}\,d\mathbf{z}$ (for some function $h(\mathbf{w},\mathbf{x}^T)$), however, these integrals do not converge. After taking the expectation the integral representation (19) for $\mathbb{E}S_\varepsilon(\mathbf{X}^T)$ is not only valid as an iterated integration (first with respect to $\mathbf{w}$ and then with respect to $\mathbf{z}$) but converges as a double multi-dimensional integral (this will be shown in Section III for $d=1$ and in Section IV for $d>1$) with different asymptotic behaviors of the integrand for various ranges, in particular if $\mathbf{w}$ and $\mathbf{z}$ are unbounded). Unfortunately this phenomenon does not show up in the computation of the second moment $\mathbb{E}[S_\varepsilon(\mathbf{x}^T)^2]$. Here we have to add a regularization factor of the form $e^{-\eta\|\mathbf{z}\|^2}$ in order to make the appearing multi-integral convergent and by finally setting $\eta>0$ sufficiently small, see Appendix E.

Our first result of this paper can be summarized as follows, which we prove in Sections III and IV.

Theorem 1. *Let $\mathbf{w}\in\mathbb{R}^d$ and $\varepsilon>0$. Assume features $\mathbf{x}_t$ are generated by a uniform distribution over the d-dimensional ball $\mathcal{B}_d$ and $p(w)=(1+\exp\{-w\})^{-1}$ is the logistic function.*

(i) *For $\varepsilon\gg 1/\sqrt{T}$ there exists $\beta(d)>0$ such that*

$$\mathbb{E}[S_\varepsilon(\mathbf{x}^T)] = \left(\frac{T}{2\pi}\right)^{d/2}\int_{\mathbb{R}^d}\sqrt{\det(\bar{\mathbf{B}}(\mathbf{w}))}\,e^{-\varepsilon\|\mathbf{w}\|^2}(1+O(T^{-\beta(d)})) \tag{20}$$

with

$$\bar{\mathbf{B}}(\mathbf{w}) = \frac{1}{\mathrm{Vol}(\mathcal{B}_d)}\int_{\mathcal{B}_d} p(\langle\mathbf{w}|\mathbf{x}\rangle)(1-p(\langle\mathbf{w}|\mathbf{x}\rangle))\mathbf{x}\otimes\mathbf{x}d\mathbf{x}, \tag{21}$$

where $\mathbf{x}\otimes\mathbf{x}=\mathbf{x}\mathbf{x}^\tau$ being the tensor product of $\mathbf{x}_t$ with τ denoting the transpose.

(ii) *Furthermore,*

$$\mathbb{E}[S_\varepsilon(\mathbf{x}^T)^2] = \mathbb{E}[S_\varepsilon(\mathbf{x}^T)]^2(1+O(T^{-\beta(d)})) \tag{22}$$

which implies that

$$\bar{r}_T^\varepsilon = \mathbb{E}\left[\log S_\varepsilon(\mathbf{X}^T)\right] = \log\mathbb{E}[S_\varepsilon(\mathbf{X}^T)](1+O(T^{-\beta(d)})) = \frac{d}{2}\log T + \log C_d(\varepsilon) + O((\log T)T^{-\beta(d)}) \tag{23}$$

where

$$C_d(\varepsilon) = \int_{\mathbb{R}^d}\sqrt{\det(\bar{\mathbf{B}}(\mathbf{w}))}\,e^{-\varepsilon\|\mathbf{w}\|^2}\,d\mathbf{w}$$

for $\varepsilon\gg 1/\sqrt{T}$.

Before we proceed, we first show how to establish $\mathbb{E}\left[\log S_\varepsilon(\mathbf{X}^T)\right] = \log\mathbb{E}[S_\varepsilon(\mathbf{X}^T)](1+O(T^{-\beta(d)}))$ of (23) using the fact that (22) holds which we prove in (A-23) of Appendix E (for simplicity of presentation we provide details only for $d=1$). Thus, let us consider the expected regret $\mathbb{E}[\log S_\varepsilon(x^T)]$. In order to simplify the notation we set $S=S_\varepsilon(x^T)$. Observe that (22) implies $\mathrm{Var}[S]=O((\mathbb{E}[S])^2T^{-\beta})$. By Jensen's inequality we directly have $\mathbb{E}[\log S]\le\log\mathbb{E}[S]$. Thus, it remains to obtain a lower bound. By Chebyshev's inequality

$$\mathbb{P}\left[\left|\frac{S}{\mathbb{E}[S]}-1\right|\ge\frac{1}{2}\right]\le 4\frac{\mathrm{Var}[S]}{(\mathbb{E}[S])^2} = O(T^{-\beta}).$$

[3]In what follows we assume that $\mathbf{X}$ is uniformly distributed in the unit ball. This simplifies several of our computations. However, they can be directly extended to rotation invariant distributions with a sufficiently fast decreasing tail (for example as $O(e^{-c\|\mathbf{x}\|})$ for some constant $c>0$).

Furthermore, for $|x-1| \le \frac{1}{2}$ we have

$$\log x \ge x - 1 - c(x-1)^2$$

for some constant $c > 0$. Hence,

$$\begin{aligned}\mathbb{E}\left[\log \frac{S}{\mathbb{E}[S]}\right] &= \underbrace{\mathbb{E}\left[\mathbf{1}_{[|S/E[S]-1| \le 1/2]} \log \frac{S}{\mathbb{E}[S]}\right]}_{A} + \underbrace{\mathbb{E}\left[\mathbf{1}_{[|S/E[S]-1| > 1/2]} \log \frac{S}{\mathbb{E}[S]}\right]}_{B} \\ &\ge \underbrace{\mathbb{E}\left[\frac{S}{\mathbb{E}[S]} - 1\right] - \mathbb{E}\left[\mathbf{1}_{[|S/E[S]-1| > 1/2]} \left(\frac{S}{\mathbb{E}[S]} - 1\right)\right] - c\,\mathbb{E}\left[\left(\frac{S}{\mathbb{E}[S]} - 1\right)^2\right]}_{A} \\ &\quad + \underbrace{\mathbb{E}\left[\mathbf{1}_{[|S/E[S]-1| > 1/2]} \log \frac{S}{\mathbb{E}[S]}\right]}_{B}.\end{aligned}$$

Trivially

$$\mathbb{E}\left[\frac{S}{\mathbb{E}[S]} - 1\right] = 0$$

and

$$\mathbb{E}\left[\left(\frac{S}{\mathbb{E}[S]} - 1\right)^2\right] = \frac{\mathrm{Var}[S]}{(\mathbb{E}[S])^2} = O(T^{-\beta}).$$

Using Cauchy-Schwarz's inequality, we find

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{[|S/E[S]-1| > 1/2]} \left(\frac{S}{\mathbb{E}[S]} - 1\right)\right] &\le \sqrt{\mathbb{P}\left[\left|\frac{S}{\mathbb{E}[S]} - 1\right| > \frac{1}{2}\right] \cdot \frac{\mathrm{Var}[S]}{(\mathbb{E}[S])^2}} \\ &= O(T^{-\beta}).\end{aligned}$$

Therefore, the term $A \ge -c' T^{-\beta}$ for some $c' > 0$.

To bound the term B, we note that by the definition of the (regularized) Shtarkov sum, we have $S \ge 1$ (to see this, take $\mathbf{w} = \mathbf{0}$ in (7)). Moreover, by (25) of Theorem 2 we have (uniformly for $\varepsilon \gg 1/\sqrt{T}$) that $\log \mathbb{E}[S] \le c'' d \log T$ for some constant $c'' > 0$. Therefore, we have

$$\log \frac{S}{\mathbb{E}[S]} \ge -\log \mathbb{E}[S] \ge -c'' d \log T.$$

This implies

$$\begin{aligned}B = \mathbb{E}\left[\mathbf{1}_{[|S/E[S]-1| > 1/2]} \log \frac{S}{\mathbb{E}[S]}\right] &\ge -c'' d \log T\ \mathbb{E}[\mathbf{1}_{[|S/E[S]-1| > 1/2]}] \\ &= -c'' d \log T\ \mathbb{P}\left[\left|\frac{S}{\mathbb{E}[S]} - 1\right| \ge \frac{1}{2}\right] \ge -c''' \frac{d \log T}{T^{\beta}}\end{aligned}$$

for some constant $c''' > 0$. Consequently,

$$\mathbb{E}[\log S] - \log \mathbb{E}[S] = \mathbb{E}\left[\log \frac{S}{\mathbb{E}[S]}\right] \ge A + B \ge -\frac{c' + c''' d \log T}{T^{\beta}} \to 0 \text{ as } T \to \infty.$$

Together with the upper bound $\mathbb{E}[\log S] \le \log \mathbb{E}[S]$ we thus obtain

$$\mathbb{E}[\log S] = \log \mathbb{E}[S] + O((\log T) T^{-\beta}) = \log \mathbb{E}[S] \cdot (1 + O(T^{-\beta})).$$

which proves our main result (23) regarding the regret (provided (22) is true established in Appendix E).

Another question is whether Theorem 1 allows us to recover the classical regret with $\varepsilon \to 0$. First, we note that our proof of Theorem 1 works only for $\varepsilon \gg 1/\sqrt{T}$, which corresponds to a radius R of order $O(T^{1/4})$ by Lemma 1. Second, from [22], we know that for $R = \infty$, the leading term of the minimax regret is $d \log T$, not $(d/2) \log T$. This indicates a transition region in which the leading term in $\log T$ grows from $d/2$ to d. In Theorem 2 below we partially fill this gap by providing the asymptotic behavior when $R \in (0, T^{1/4}]$. The main technical ingredient is the following asymptotic analysis of $\det(\bar{\mathbf{B}}(\mathbf{w}))$.

Lemma 2. *Suppose that the matrix $\bar{\mathbf{B}}(\mathbf{w})$ is given by (21). Then, as $\|\mathbf{w}\| \to \infty$, we have*

$$\det(\bar{\mathbf{B}}(\mathbf{w})) = \frac{v_{d-1}}{v_d} \frac{\pi^2}{3(d+2)^{d-1}} \|\mathbf{w}\|^{-d-2} \left(1 + O(\|\mathbf{w}\|^{-2})\right), \tag{24}$$

where $v_d = \pi^{d/2}/\Gamma(1+\frac{d}{2})$ denotes the volume of the d-dimensional unit ball.

The proof of a slightly more refined property is deferred to Appendix B (see Lemma 8). This leads to the following noteworthy conclusion regarding the growth of the regularized Shtarkov sum.

Theorem 2. *Under assumptions of Theorem 1 we have, as $\varepsilon \to 0$ but $\varepsilon \gg T^{-1/2}$,*

$$\mathbb{E}[S_\varepsilon(\mathbf{x}^T)] = \begin{cases} T^{\frac{d}{2}} \bar{C}_d \left(\varepsilon^{-\frac{d}{4}+\frac{1}{2}} + O(1)\right)(1 + O(T^{-\beta(d)})) & d \geq 3, \\ T^{1/2} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\int_{-1}^{1} p(wx)(1-p(wx))x^2 dx}\, dw + O(\varepsilon^{\frac{1}{4}})\right)(1 + O(T^{-\beta(d)})) & d = 1, \\ T\sqrt{\frac{\pi}{24}} \left(\log \varepsilon^{-1} + O(1)\right)(1 + O(T^{-\beta(d)})) & d = 2, \end{cases} \tag{25}$$

where

$$\bar{C}_d = (2\pi)^{-\frac{d}{2}} \sqrt{\frac{v_{d-1}}{v_d} \frac{\pi^2}{3(d+2)^{d-1}}} \frac{d v_d}{2} \Gamma\left(\frac{d-2}{4}\right).$$

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

$$\bar{r}_T^\varepsilon = \mathbb{E}\left[\log S_\varepsilon(\mathbf{X}^T)\right] = \left((\frac{1}{2} + \frac{\alpha}{4})d - \frac{\alpha}{2}\right) \log T + O(1). \tag{26}$$

Proof. Theorem 2 follows from Theorem 1 and Lemma 2. We only do the calculations for $d \geq 3$. The cases $d = 1$ and $d = 2$ are then simple variants.

By Lemma 2 we have for $\|\mathbf{w}\| \geq 1$

$$\det(\bar{\mathbf{B}}(\mathbf{w})) = K_d \|\mathbf{w}\|^{-d-2} \left(1 + O(\|\mathbf{w}\|^{-2})\right)$$

with $K_d = \frac{v_{d-1}}{v_d} \frac{\pi^2}{3(d+2)^{d-1}}$. Consequently we obtain (with $s_d = d v_d$ denoting the surface measure of the d-dimensional unit ball)

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

where we have used the substitution $\varepsilon r^2 = v$ and the property that

$$\int_0^\varepsilon v^{\frac{d}{4}-\frac{3}{2}} e^{-v} dv = O(\varepsilon^{\frac{d}{4}-\frac{1}{2}}).$$

Thus, with the help of Theorem 1 the asymptotic result follows.

$\square$

We highlight the importance of (26). Note that previous results show that $\bar{r}_T^R \sim (d/2)\log T$ for bounded $R = O(1)$, while for $R = \infty$, it was observed in [22] that $\bar{r}_T^R \sim d \log T$. It was conjectured that as R grows to infinity, the leading coefficient in front of $\log T$ also increases. Theorem 2 and Lemma 1 demonstrate that such a growth rate is *linearly* controlled by α. More precisely the relation (8) implies

$$\frac{d}{2}\log T + \frac{d-2}{2}\log R - \frac{d-2}{2}\log\log T + O(1) \le \bar{r}_T^R \le \frac{d}{2}\log T + \frac{d-2}{2}\log R + O(1) \tag{27}$$

for $R \ll T^{1/4}$. It is very likely that the restriction $\varepsilon \gg T^{-1/2}$ is natural and that the behavior for $T^{-1} \ll \varepsilon \ll T^{-1/2}$ will be different. This would be another major challenge to settle in future work.

III. Proof of Theorem 1(i) for $d = 1$

In this and the next sections we present the proof of our main result Theorem 1(i). The proof is long, tedious, and very technical. To help the reader, we focus here on $d = 1$. The extension for $d > 1$ is presented in Section IV.

The second part of Theorem 1 is still more technical. So we will only present in Appendix E the proof of the relation $\mathbb{E}[S_\varepsilon(\mathbf{x}^T)^2] = \mathbb{E}[S_\varepsilon(\mathbf{x}^T)]^2(1 + O(T^{-\beta(d)}))$ for $d = 1$.

We start with a brief road map of the proof. We recall that we already used the following principle to represent the maximum $\max_w f(w) = f(w^*)$ (where $f(w) = P_\varepsilon(y^T|x^T, w)$). By convexity we can determine w^* by the equation $g(w) = \nabla f(w) = 0$ and is, thus, given by $w^* = g^{-1}(0)$ which (by Fourier analysis) can be represented by

$$\max_w f(w) = f \circ g^{-1}(0) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f \circ g^{-1} e^{-iyz}\, dy\, dz = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f(w) e^{-ig(w)z}\nabla^2 f(w)\, dw\, dz.$$

In our case we additionally have the property that $\nabla^2 f(w)$ does not depend on y and that the sum $\sum_{y^T} f(w)e^{-ig(w)z}$ factors nicely.

The main part of the proof is to evaluate the resulting double integral (that we have slightly simplified and where we have taken the sum over all y^T and the expectation with respect to x^T) asymptotically:

$$\frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \overline{f}(w,z)^T e^{-\varepsilon w^2 - 2iwz}\, dw\, dz.$$

For small z and w the first term is approximated by

$$\overline{f}(w,z) \approx e^{-\frac{1}{2}z\overline{B}(w)z}$$

so that the integral over z can be evaluated by a Gauss-like integral. The main problem is that this approximation does not hold for large z and w so it needs subtle analytic methods to show that these parts do not contribute to the main term that is given by the Gauss-like integral. (This is even more involved in the case $d > 1$, see Section IV.)

We now start with the full details. In what follows we will use the notation $\overline{f}(\mathbf{w}, \mathbf{z}) = \mathbb{E} f(\mathbf{w}, \mathbf{X}, \mathbf{z})$. For the case $d = 1$ we have

$$\overline{f}(w,z) = \mathbb{E} f(w, X, z) = \frac{1}{2}\int_{-1}^{1}\left(\frac{e^{-ixz/(1+e^{xw})}}{1+e^{-xw}} + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}}\right) dx, \tag{28}$$

hence $\mathbb{E}S_\varepsilon(\mathbf{X}^T)$ is given by

$$\begin{aligned}
\mathbb{E}S_\varepsilon(X^T) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[\prod_{t=1}^{T} f(w, X_t, z)\,\nabla^2 L_\varepsilon(\cdot|X^T, w)\right] e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\
&\overset{(a)}{=} \sum_{s=1}^{T}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[\prod_{t=1}^{T} f(w, X_t, z)\, p(x_s w)p(-x_s w)x_s^2\right] e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\
&\quad + \frac{2\varepsilon}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[\prod_{t=1}^{T} f(w, X_t, z)\right] e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\
&= \frac{T}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overline{f}(w,z)^{T-1} B(z,w) e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz + \frac{2\varepsilon}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overline{f}(w,z)^{T} e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz, \\
&=: T \cdot J_0 + 2\varepsilon \cdot J_1,
\end{aligned} \tag{29}$$

where $B(z,w)$ abbreviates $B(z,w) = \mathbb{E}\left[f(w,X,z)\, p(Xw)p(-Xw)X^2\right]$ and (a) follows by (14).

We start with three technical lemmas regarding $\overline{f}(w,z)$ with proofs presented in Appendix C.

Lemma 3. *Set*

$$\overline{B}(w) = \mathbb{E}\left[p(xw)p(-xw)x^2\right] = \frac{1}{2}\int_{-1}^{1} \frac{x^2}{(1+e^{-xw})(1+e^{xw})}\, dx. \tag{30}$$

Then we uniformly have $\overline{B}(w) = \Theta\left(\min(1,|w|^{-3})\right)$ *and for* $c_1 > 0$

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

Lemma 4. *For* $|w| \le 1$ *we uniformly have* $|\overline{f}(z,w)| \le \min\left(1, \frac{C}{|z|}\right)$ *and for* $|w| \ge 1$ *we have*

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

Lemma 5. *Suppose that* $c_1 > 0$ *is a given constant. Then there exists* $c_2 > 0$ *such that*

$$\left|\overline{f}(z,w)\right| \le 1 - \frac{c_2}{|w|}$$

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

Granted these lemmas, we now prove our main result for $\mathbb{E}S_\varepsilon(X^T)$ in the case $d = 1$ which we formulate next.

Proposition 1. *Suppose that* $d = 1$ *and that* X *is uniformly distributed on* $[-1,1]$. *Then*

$$\mathbb{E}S_\varepsilon(X^T) = \frac{T^{1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2 - 2\varepsilon^2 w^2/(T\overline{B}(w))}\, dw + O\left(\log T\right)$$

provided that $\varepsilon \gg T^{-1/2}$, *where* $\overline{B}(w)$ *is given in (30).*

Note that

$$\begin{aligned}\int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2 - 2\varepsilon^2 w^2/(T\overline{B}(w))}\, dw &= \int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2}\, dw + O\left(T^{-1/10}\right)\\ &= \int_{-\infty}^{\infty} \overline{B}(w)^{1/2}\, dw + O\left(\varepsilon^{1/4} + T^{-1/10}\right)\end{aligned} \tag{31}$$

as long as $\varepsilon \to 0$ such that $\varepsilon \gg T^{-1/2}$. Thus, the integral in Proposition 1 can be replaced by the integral

$$\int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2}\, dw.$$

The rest of this subsection is devoted to the proof of the Proposition 1. We recall from (29) the representation $\mathbb{E}S_\varepsilon(X^T) = T \cdot J_0 + 2\varepsilon \cdot J_1$, where

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

The main challenge in the computation of the integral(s) J_0 and J_1 are the parts that correspond to large w. We only discuss the integral J_1 in detail. For any constant $C_3 > 0$, we consider the following cases:

A: *The case* $|w| \le C_3$. If $|w| \le C_3$ then we have the uniform bound $|\overline{f}(z,w)| \le C/|z|$ by Lemma 4. We first look at the case, where $|z| \le 2C$. Here we certainly have the uniform representation (see Lemma 3) $\overline{f}(z,w) =$

$e^{-\frac{z^2}{2}\overline{B}(w)}\left(1+O(z^3)\right)$ and by continuity $|\overline{f}(z,w)| \le e^{-c_1 z^2}$ for some constant $c_1>0$. If $|z|\ge 2C$ then we have the trivial estimate $|\overline{f}(z,w)| \le C/|z| \le 1/2$ (see Lemma 4). Consequently,

$$\begin{aligned}
I_1 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-C_3}^{C_3} \overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz \\
&= \frac{1}{2\pi}\int_{-C_3}^{C_3}\int_{|z|\le T^{-1/3}} e^{-T\overline{B}(w)\frac{z^2}{2}}\left(1+O(Tz^3)\right)e^{-\varepsilon w^2-2i\varepsilon wz}\,dz\,dw \\
&+ \frac{1}{2\pi}\int_{-C_3}^{C_3}\int_{T^{-1/3}\le|z|\le 2C} O\left(e^{-c_1Tz^2}\right)e^{-\varepsilon w^2}\,dz\,dw + \frac{1}{2\pi}\int_{-C_3}^{C_3}\int_{|z|\ge 2C}(C/|z|)^T e^{-\varepsilon w^2}\,dz\,dw \\
&= \int_{-C_3}^{C_3}\frac{1}{\sqrt{2\pi T\overline{B}(w)}}e^{-2\varepsilon^2w^2/(T\overline{B}(w))}e^{-\varepsilon w^2}\,dw + O\left(T^{-1}\right) + O\left(e^{-c_1T^{1/3}}\right) + O(2^{-T}) \\
&= \frac{1}{\sqrt{2\pi T}}\int_{-C_3}^{C_3}\overline{B}(w)^{-1/2}e^{-\varepsilon w^2}\,dw + O\left(T^{-1}\right).
\end{aligned}$$

B: *The case $C_3 \le |w| \le \eta T$ where $\eta = \eta(T) \to 0$.* Furthermore we divide the integral over z into several parts.

The first part is the interval $|z| \le z_1 = |w|^{3/2}T^{-1/2}\eta^{-1/6}$, where we use Lemma 3

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

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

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

for some $c>0$. Similarly we obtain

$$\int_{|z|\le z_1} e^{-T\overline{B}(w)\frac{z^2}{2}}T|z^3/w^4|\,dz = O\left(\frac{|w|^{1/2}}{T}\right), \int_{|z|\le z_1} e^{-T\overline{B}(w)\frac{z^2}{2}}T|z^4/w^6|\,dz = O\left(\frac{|w|^{3/2}}{T^{3/2}}\right).$$

Summing up we find

$$\begin{aligned}
I_{21} &= \frac{1}{2\pi}\int_{|z|\le z_1}\int_{C_3\le|w|\le\eta T}\overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz \\
&= \frac{1}{\sqrt{2\pi T}}\int_{|w|\ge C_3}\overline{B}(w)^{-1/2}e^{-2\varepsilon^2w^2/(T\overline{B}(w))}e^{-\varepsilon w^2}\,dw + O\left(\varepsilon^{-5/4}T^{3/2}e^{-\varepsilon\eta^2T^2} + \frac{\varepsilon^{-3/4}}{T} + \frac{\varepsilon^{-5/4}}{T^2}\right).
\end{aligned}$$

Next suppose that $z_1 = |w|^{3/2}T^{-1/2}\eta^{-1/6} \le |z| \le c_1|w|$, where c_1 sufficiently small which ensures that (see Lemma 3)

$$\overline{B}(w)\frac{z^2}{2} \ge C\left(\frac{|z|^3}{|w|^4}+\frac{|z|^4}{|w|^6}\right).$$

Recall that $\overline{B}(w) = \Theta(|w|^{-3})$ for $|w|\ge 1$. Hence, there exists a constant $c>0$ such that $|\overline{f}(z,w)| \le e^{-cz^2/|w|^3}$ uniformly for $z_1 \le |z| \le c_1|w|$. This implies that the corresponding integral is upper bounded by

$$\begin{aligned}
I_{22} &= \frac{1}{2\pi}\int_{z_1\le|z|\le c_1|w|}\int_{C_3\le|w|\le\eta T}\overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz \\
&= O\left(\frac{\eta^{1/6}e^{-c\eta^{-1/3}}}{T^{1/2}}\int_{C_3\le|w|\le\eta T} w^{3/2}e^{-\varepsilon w^2}\,dw\right) = O\left(\frac{\eta^{1/6}e^{-c\eta^{-1/3}}}{\varepsilon^{5/4}T^{1/2}}\right).
\end{aligned}$$

Next suppose that $c_1|w| \le |z| \le c_3 e^{|w|}$ for an arbitrary constant $c_3 > 0$. Here we have the upper bound $|\overline{f}(z,w)| \le 1 - c_2/|w| \le e^{-c_2/|w|}$ (see Lemma 5) and **provided** that $\varepsilon \gg T^{-1/2}$ we have

$$I_{23} = \frac{1}{2\pi}\int_{c_1|w|\le|z|\le c_3 e^{|w|}}\int_{C_3\le|w|\le\eta T}\overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz = O\left(\int_{C_3\le|w|\le\eta T} e^{-c_2T/|w|+|w|-\varepsilon w^2}\,dw\right)$$
$$= O\left(\int_{C_3\le|w|\le c_2\sqrt{T}} e^{-(1-c_2)\sqrt{T}-\varepsilon w^2}\,dw\right) + O\left(\int_{c_2\sqrt{T}\le|w|\le\eta T} e^{-\frac{\varepsilon}{2}w^2}\,dw\right)$$
$$= O\left(\varepsilon^{-1/2}e^{-(1-c_2)\sqrt{T}} + e^{-\frac{\varepsilon}{2}c_2^2T}\right) = O\left(e^{-c_5\sqrt{T}}\right) = O\left(\frac{1}{T}\right).$$

Finally if $|z| \ge c_3 e^{|w|}$, where c_3 is chosen sufficiently large, we have (for some constant $\tilde{C} \le C_3$)

$$\overline{f}(z,w) \le \tilde{C}\max\left(\frac{1}{\sqrt{|wz|}}, \frac{e^{|w|}}{|zw|}\right) \le \frac{1}{2}.$$

If $c_3 e^{|w|} \le |z| \le e^{2|w|}/|w|$ then the first term $1/\sqrt{|wz|}$ dominates and we find

$$I_{24} = \frac{1}{2\pi}\int_{c_3 e^{|w|}\le|z|\le e^{2|w|}/|w|}\int_{C_3\le|w|\le\eta T}\overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz = O\left(\frac{1}{T}\right).$$

Similarly we have for $|z| \ge e^{2|w|}/|w|$, where the second term dominates

$$I_{25} = \frac{1}{2\pi}\int_{|z|\ge e^{2|w|}/|w|}\int_{1\le|w|\le\eta T}\overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz = O\left(\frac{1}{T}\right).$$

Summing up we have provided that $\varepsilon \gg T^{-1/2}$

$$I_2 = \frac{1}{\sqrt{2\pi T}}\int_{|w|\ge C_3}\overline{B}(w)^{-1/2}e^{-\varepsilon w^2-2\varepsilon^2w^2/(T\overline{B}(w))}\,dw$$
$$+ O\left(\frac{\varepsilon^{-3/4}}{T} + \frac{\varepsilon^{-5/4}}{T^2} + \frac{\eta^{1/6}e^{-c\eta^{-1/3}}}{\varepsilon^{5/2}T^{1/2}} + \varepsilon^{-5/4}T^{3/2}e^{-\varepsilon\eta^2T^2}\right).$$

C: *The case* $|w| \ge \eta T$. If $|z| \le e^{2|w|}/|w|$, then $|\overline{f}(z,w)| \le 1$ and obtain for $\varepsilon \gg T^{-1/2}$

$$I_3 = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{|w|\ge\eta T}\overline{f}(w,z)^T e^{-\varepsilon w^2-2i\varepsilon wz}\,dw\,dz = O\left(e^{-\eta^2T}\right).$$

In conclusion, we arrive at

$$J_1 = I_1 + I_2 + I_3 = \frac{1}{\sqrt{2\pi T}}\int_{-\infty}^{\infty}\overline{B}(w)^{-1/2}e^{-\varepsilon w^2-2\varepsilon^2w^2/(T\overline{B}(w))}\,dw + O\left(T^{-5/8}\right)$$

provided that $\varepsilon \gg T^{-1/2}$ and where $\eta = \eta(T) = c_6(\log T)^{-3}$ for a sufficiently small positive constant c_6.

IV. Proof of Theorem 1(i) for general d

Now we present detailed proof of Theorem 1(i) for general d. The proof is very long with many complicated estimates of multi dimensional integrals. We keep most details in this section for clarity, however, very technical derivaties we delay till Appendix D.

We fix $d > 1$ and consider first the expected value $\mathbb{E}S_\varepsilon(\mathbf{X}^T)$. By using the representation (19) and by expanding the determinant we obtain (very similarly as in the case $d = 1$)

$$\mathbb{E}S_\varepsilon(\mathbf{X}^T) = \sum_{j=0}^{d} T^{d-j}\tilde{J}_j(\varepsilon),$$

where $\tilde{J}_j(\varepsilon)$, $j = 0, \ldots, d$, are proper linear combinations of integrals of the forms similar to J_0 and J_1 (from the case $d = 1$) together with proper powers of ε. In particular the dominant term $\tilde{J}_0 = J_0(\varepsilon)$ is given by

$$\tilde{J}_0 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \overline{f}(\mathbf{w}, \mathbf{z})^{T-d} \det(B(\mathbf{z}, \mathbf{w})) e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w}, \mathbf{z}\rangle}\, d\mathbf{w}\, d\mathbf{z},$$

where

$$B(\mathbf{z}, \mathbf{w}) = \mathbb{E}\left[f(\mathbf{w}, \mathbf{X}, \mathbf{w}) p(\langle \mathbf{x}, \mathbf{w}\rangle) p(-\langle \mathbf{x}, \mathbf{w}\rangle) \mathbf{X} \otimes \mathbf{X}\right].$$

For the sake of brevity we will only consider the term $\tilde{J}_0$ (the other terms are similar). First, however, in Appendix D we derive several upper bounds on $\overline{f}(\mathbf{w}, \mathbf{z})$. Granted this, we focus now on estimating $\tilde{J}_0$.

We prove the following main technical result of this section.

Proposition 2. *We have*

$$\tilde{J}_0 \sim \frac{1}{(2\pi T)^{d/2}} \int_{\mathbb{R}^d} \sqrt{\det\left(\overline{B}(\mathbf{w})\right)} e^{-\varepsilon\|\mathbf{w}\|^2 - \frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1} \mathbf{w}}\, d\mathbf{w}$$

provided that $\varepsilon \gg T^{-1/2}$, *where* $\overline{B}(\mathbf{w})$ *is given by (21).*

We recall that

$$B(\mathbf{z}, \mathbf{w}) = \int_{\mathcal{B}_d} f(\mathbf{w}, \mathbf{x}, \mathbf{z}) p(\langle \mathbf{x}, \mathbf{w}\rangle) p(-\langle \mathbf{x}, \mathbf{w}\rangle) \mathbf{x} \otimes \mathbf{x}\, d\mathbf{x}$$

and we note that $B(0, \mathbf{w}) = \overline{B}(\mathbf{w})$. Moreover $B(0, \mathbf{w})$ is a positive matrix and all entries of $B(\mathbf{z}, \mathbf{w})$ satisfy

$$|B(\mathbf{z}, \mathbf{w})_{i,j}| \le \overline{B}(\mathbf{w})_{i,j}, \qquad 1 \le i, j \le d.$$

Furthermore, it is an easy exercise to show (by expanding the determinant and estimating all parts absolutely) that (see also Lemma 2)

$$\det\left(B(\mathbf{z}, \mathbf{w})\right) \le \min\left(1, \|\mathbf{w}\|^{-d-2}\right).$$

Furthermore by Taylor expansion (and similar computations) we have

$$\det\left(B(\mathbf{z}, \mathbf{w})\right) = \det\left(\overline{B}(\mathbf{w})\right) \cdot \left(1 + O\left(\|\mathbf{z}\|^2 \min(1, \|\mathbf{w}\|^{-1})\right)\right) \tag{32}$$

$$= \det\left(\overline{B}(\mathbf{w})\right) + O\left(\|\mathbf{z}\|^2 \min(1, \|\mathbf{w}\|^{-d-3})\right). \tag{33}$$

As in the case $d = 1$ we partition the $2d$-dimensional integral into several parts.

A: *The case* $\|\mathbf{w}\| \le 1$. First suppose that $\|\mathbf{z}\| \le T^{-1/3}$. By Lemma 11 we have

$$\begin{aligned}
I_{11} &= \frac{1}{(2\pi)^d} \int_{\|\mathbf{z}\| \le T^{-1/3}} \int_{\|\mathbf{w}\| \le 1} \det(B(\mathbf{z}, \mathbf{w})) \overline{f}(\mathbf{w}, \mathbf{z})^{T-d} e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w}, \mathbf{z}\rangle}\, d\mathbf{w}\, d\mathbf{z} \\
&= \frac{1}{(2\pi)^d} \int_{\|\mathbf{w}\| \le 1} \det(\overline{B}(\mathbf{w})) \\
&\qquad \int_{\|\mathbf{z}\| \le T^{-1/3}} e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z}} \left(1 + O(T\|\mathbf{z}\|^3) + O(\|\mathbf{z}\|^2)\right) e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w}, \mathbf{z}\rangle}\, d\mathbf{w}\, d\mathbf{z} \\
&= \frac{1}{(2\pi)^d} \int_{\|\mathbf{w}\| \le 1} \det(\overline{B}(\mathbf{w})) e^{-\varepsilon\|\mathbf{w}\|^2} \left(\int_{\mathbb{R}^d} e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z} - 2i\varepsilon\langle \mathbf{w}, \mathbf{z}\rangle}\, d\mathbf{z} + O(e^{-cT^{1/3}})\right) d\mathbf{w} \\
&+ O\left(\int_{\|\mathbf{w}\| \le 1} \det(\overline{B}(\mathbf{w})) e^{-\varepsilon\|\mathbf{w}\|^2} \int_{\mathbb{R}^d} e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z}} T\|\mathbf{z}\|^3\, d\mathbf{z}\, d\mathbf{w}\right) \\
&+ O\left(\int_{\|\mathbf{w}\| \le 1} \det(\overline{B}(\mathbf{w})) e^{-\varepsilon\|\mathbf{w}\|^2} \int_{\mathbb{R}^d} e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z}} \|\mathbf{z}\|^2\, d\mathbf{z}\, d\mathbf{w}\right) \\
&= \frac{1}{(2\pi T)^{d/2}} \int_{\|\mathbf{w}\| \le 1} \sqrt{\det(\overline{B}(\mathbf{w}))} e^{-\varepsilon\|\mathbf{w}\|^2 - \frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1} \mathbf{w}}\, d\mathbf{w} + O\left(e^{-cT^{1/3}} + T^{-\frac{d+1}{2}} + T^{-\frac{d+2}{2}}\right) \\
&= \frac{1}{(2\pi T)^{d/2}} \int_{\|\mathbf{w}\| \le 1} \sqrt{\det(\overline{B}(\mathbf{w}))} e^{-\varepsilon\|\mathbf{w}\|^2 - \frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1} \mathbf{w}}\, d\mathbf{w} + O\left(T^{-\frac{d+1}{2}}\right)
\end{aligned}$$

for some constant $c > 0$.

Next we consider the case $T^{-1/3} \le \|\mathbf{z}\| \le C$, where C is chosen in a way that $C/\log C \ge 2C_1$, where C_1 is the constant from the inequality (A-15):

$$\begin{aligned} I_{12} &= \frac{1}{(2\pi)^d} \int_{T^{-1/3} \le \|\mathbf{z}\| \le C} \int_{\|\mathbf{w}\| \le 1} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon \|\mathbf{w}\|^2 - 2i\varepsilon \langle \mathbf{w},\mathbf{z}\rangle} \, d\mathbf{w} \, d\mathbf{z} \\ &= O\left(\int_{T^{-1/3} \le \|\mathbf{z}\| \le C} \int_{\|\mathbf{w}\| \le 1} e^{-c_1 T \|\mathbf{z}\|^2} e^{-\varepsilon \|\mathbf{w}\|^2} \, d\mathbf{w} \, d\mathbf{z} \right) = O\left(e^{-c_1 T^{1/3}} \right) = O\left(T^{-d} \right) \end{aligned}$$

for some constant $c_1 > 0$. Finally by Lemma 13 for $\|\mathbf{z}\| \ge C$ we have $|\overline{f}(z,w)| \le C_1 \log(\|\mathbf{z}\|)/\|\mathbf{z}\| \le \frac{1}{2}$ and consequently

$$\begin{aligned} I_{13} &= \frac{1}{(2\pi)^d} \int_{\|\mathbf{z}\| \ge C} \int_{\|\mathbf{w}\| \le 1} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon \|\mathbf{w}\|^2 - 2i\varepsilon \langle \mathbf{w},\mathbf{z}\rangle} \, d\mathbf{w} \, d\mathbf{z} \\ &= O\left(\int_{\|\mathbf{z}\| \ge C} \left(\frac{C_1 \log(\|\mathbf{z}\|)}{\|\mathbf{z}\|} \right)^{T-d} d\mathbf{z} \right) \\ &= O\left(\int_C^\infty r^{d-1} \left(\frac{C_1 \log r}{r} \right)^{T-d} dr \right) \\ &= O\left(\frac{1}{T 2^T} \right) = O\left(T^{-d} \right). \end{aligned}$$

Summing up we have

$$\begin{aligned} I_1 &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\|\mathbf{w}\| \le 1} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon \|\mathbf{w}\|^2 - 2i\varepsilon \langle \mathbf{w},\mathbf{z}\rangle} \, d\mathbf{w} \, d\mathbf{z} \\ &= \frac{1}{(2\pi T)^{d/2}} \int_{\|\mathbf{w}\| \le 1} \sqrt{\det(\overline{B}(\mathbf{w}))} e^{-\varepsilon \|\mathbf{w}\|^2 - \frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1} \mathbf{w}} \, d\mathbf{w} + O\left(T^{-\frac{d+1}{2}} \right). \end{aligned}$$

B: *The case* $1 \le \|\mathbf{w}\| \le \eta T$.

We again assume that $\eta = \eta(T) = c_6 (\log T)^{-3} \to 0$ for a sufficiently small positive constant c_6. We start with the case $\|\mathbf{z}\| \le z_1 = \|\mathbf{w}\|^{3/2} T^{-1/2} \eta^{-1/6}$. By Lemma 11 and by (33) we have (by distinguishing between the cases $|\varphi| \le 1/\|\mathbf{w}\|$ and $1/\|\mathbf{w}\| |\varphi| \le \pi$, where φ denote the angle between $\mathbf{z}$ and $\mathbf{w}$)

$$\begin{aligned} &\int_{\|\mathbf{z}\| \le z_1} \det(B(\mathbf{w},\mathbf{z})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-2i\varepsilon \langle \mathbf{z},\mathbf{w}\rangle} \, dz \\ &\quad = \det(\overline{B}(\mathbf{w})) \int_{\|\mathbf{z}\| \le z_1} e^{-(T-d)\frac{1}{2} \mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z} - 2i\varepsilon \langle \mathbf{z},\mathbf{w}\rangle} \, dz \\ &\quad + O\left(\|\mathbf{w}\|^{-d-3} \int_{\|\mathbf{z}\| \le z_1} \|\mathbf{z}\|^2 \overline{f}(\mathbf{w},\mathbf{z})^{T-d} \, d\mathbf{z} \right) \\ &\quad = \sqrt{\det(\overline{B}(\mathbf{w}))} \left(\frac{2\pi}{T} \right)^{d/2} e^{-\frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1} \mathbf{w}} \\ &\quad + O\left(T^{-\frac{d}{2}} \|\mathbf{w}\|^{-\frac{d}{2}-1} e^{-c\eta^{-1/3}} \right) + O\left(T^{-\frac{d}{2}-1} \|\mathbf{w}\|^{-\frac{d}{2}+1} \right). \end{aligned}$$

We consider, for example, the following error term:

$$
\begin{aligned}
&\|\mathbf{w}\|^{-d-3}\int_{\|\mathbf{z}\|\le z_1}\|\mathbf{z}\|^2\overline{f}(\mathbf{w},\mathbf{z})^{T-d}\,d\mathbf{z}\\
&\quad=\|\mathbf{w}\|^{-d-3}\int\limits_{\|\mathbf{z}\|\le z_1,\,|\varphi|\le 1/\|\mathbf{w}\|}\|\mathbf{z}\|^2\overline{f}(\mathbf{w},\mathbf{z})^{T-d}\,d\mathbf{z}\\
&\quad+\|\mathbf{w}\|^{-d-3}\int\limits_{\|\mathbf{z}\|\le z_1,\,1/\|\mathbf{w}\|\le|\varphi|\le\pi}\|\mathbf{z}\|^2\overline{f}(\mathbf{w},\mathbf{z})^{T-d}\,d\mathbf{z}\\
&\quad=O\left(\|\mathbf{w}\|^{-d-3}\int_0^{1/\|\mathbf{w}\|}\int_0^\infty\varphi^{d-2}r^{d+1}e^{-c_1Tr^2/\|\mathbf{w}\|^3}\,dr\,d\varphi\right)\\
&\quad+O\left(\|\mathbf{w}\|^{-d-3}\int_{1/\|\mathbf{w}\|}^{\pi}\int_0^\infty\varphi^{d-2}r^{d+1}e^{-c_1Tr^2\varphi^2/\|\mathbf{w}\|}\,dr\,d\varphi\right)\\
&\quad=O\left(T^{-\frac{d}{2}-1}\|\mathbf{w}\|^{-\frac{d}{2}+1}\right)
\end{aligned}
$$

where we use the polar coordinates with $\|\mathbf{z}\|=r$.

This implies

$$
\begin{aligned}
I_{21}&=\frac{1}{(2\pi)^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\int_{\|\mathbf{z}\|\le z_1}\det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{z},\mathbf{w}\rangle}\,d\mathbf{w}\,d\mathbf{z}\\
&=\frac{1}{(2\pi T)^{d/2}}\int_{1\le\|\mathbf{w}\|\le\eta T}\sqrt{\det(\overline{B}(\mathbf{w}))}e^{-\varepsilon\|\mathbf{w}\|^2-\frac{2\varepsilon^2}{T}\mathbf{w}^\tau\overline{B}(\mathbf{w})^{-1}\mathbf{w})}\,d\mathbf{w}\\
&+O\left(\frac{\varepsilon^{-\frac{d}{4}+\frac12}e^{-c\eta^{1/3}}}{T^{d/2}}\right)+O\left(\frac{\varepsilon^{-\frac{d}{4}-\frac12}}{T^{(d+2)/2}}\right)\\
&=\frac{1}{(2\pi T)^{d/2}}\int_{\|\mathbf{w}\|\ge 1}\sqrt{\det(\overline{B}(\mathbf{w}))}e^{-\varepsilon\|\mathbf{w}\|^2-\frac{2\varepsilon^2}{T}\mathbf{w}^\tau\overline{B}(\mathbf{w})^{-1}\mathbf{w})}\,d\mathbf{w}\\
&+O\left(\frac{\varepsilon^{-\frac{d}{4}+\frac12}e^{-c\eta^{1/3}}}{T^{d/2}}\right)+O\left(\frac{\varepsilon^{-\frac{d}{4}-\frac12}}{T^{(d+2)/2}}\right).
\end{aligned}
$$

Note that the integral is of order

$$
\Theta\left(\frac{\varepsilon^{-d/4+1/2}}{T^{d/2}}\right)
$$

that is certainly asymptotically leading if $\varepsilon\gg T^{-1/2}$.

Next suppose that $z_1=\|\mathbf{w}\|^{3/2}T^{-1/2}\eta^{-1/6}\le\|\mathbf{z}\|\le c_1\|\mathbf{w}\|$, where c_1 sufficiently small. Here we know that (for a suitable constant $c>0$)

$$
|\overline{f}(\mathbf{z},\mathbf{w})|\le e^{-c(\|\mathbf{z}\|^2/\|\mathbf{w}\|^3+\|\mathbf{z}\|^2\varphi^2/\|\mathbf{w}\|)}
$$

uniformly for $z_1\le\|\mathbf{z}\|\le c_1\|\mathbf{w}\|$. This implies that the corresponding integral is upper bounded by

$$
\begin{aligned}
I_{22}&=\frac{1}{(2\pi)^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\int_{z_1\le\|\mathbf{z}\|\le c_1\|\mathbf{w}\|}\det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{z},\mathbf{w}\rangle}\,d\mathbf{w}\,d\mathbf{z}\\
&=O\left(\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}e^{-\varepsilon w^2}\int_{z_1\le\|\mathbf{z}\|\le c_1\|\mathbf{w}\|}e^{-c(T-d)(\|\mathbf{z}\|^2/\|\mathbf{w}\|^3+\|\mathbf{z}\|^2\varphi^2/\|\mathbf{w}\|)}\,d\mathbf{z}\,d\mathbf{w}\right)\\
&=O\left(\frac{\eta^{-(d-1)/6}e^{-c\eta^{-1/3}}}{T^{d/2}}\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d/2-1}e^{-\varepsilon\|\mathbf{w}\|^2}\,dw\right)\\
&=O\left(\frac{\eta^{-(d-1)/6}e^{-c\eta^{-1/3}}}{T^{d/2}}\varepsilon^{-d/4+1}\right).
\end{aligned}
$$

Recall that $e^{-c\eta^{-1/3}} = e^{-cc_6^{-1/3}\log T} = T^{-cc_6^{-1/3}}$, where we can choose c_3 arbitrarily small.

In the next step we consider the case $c_1\|\mathbf{w}\| \le \|\mathbf{z}\| \le c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\|$. Here we have the upper bound $|\overline{f}(\mathbf{z},\mathbf{w})| \le 1 - c_4/(\|\mathbf{w}\|\log^2\|\mathbf{w}\|)$ and, thus, we find

$$\begin{aligned}
I_{23} &= \frac{1}{(2\pi)^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\int_{c_1\|\mathbf{w}\|\le\|\mathbf{z}\|\le c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\|} \det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{z},\mathbf{w}\rangle}\,d\mathbf{w}\,d\mathbf{z}\\
&= O\left(\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}e^{-\varepsilon\|\mathbf{w}\|^2}\int_{c_1\|\mathbf{w}\|\le\|\mathbf{z}\|\le c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\|} e^{-Tc_4/(\|\mathbf{w}\|\log^2\|\mathbf{w}\|)}\,d\mathbf{z}\,d\mathbf{w}\right)\\
&= O\left(\int_{1\le\|\mathbf{w}\|\le\eta T}(\log\|\mathbf{w}\|)^{2d}\|\mathbf{w}\|^{-2}e^{-Tc_4/(\|\mathbf{w}\|\log^2\|\mathbf{w}\|)-\varepsilon\|\mathbf{w}\|^2}\,d\mathbf{w}\right)\\
&= O\left(\int_1^\infty(\log r)^{2d}r^{d-3}e^{-c_4T/(r\log r)-\varepsilon r^2}\,dr\right).
\end{aligned}$$

We split the integral at $r_0 = (T/\varepsilon)^{1/3}$ so that

$$\frac{T}{r_0\log^2 r_0} = 9\frac{(T^2\varepsilon)^{1/3}}{\log^2(T/\varepsilon)} \quad\text{and}\quad \varepsilon r_0^2 = (T^2\varepsilon)^{1/3}.$$

Since $\varepsilon \gg T^{1/2}$ we obtain the upper bound

$$I_{23} = O\left(e^{-c''(T^2\varepsilon)^{1/3}}\right) = O\left(e^{-c'''T^{1/2}}\right)$$

for some constants $c'', c''' > 0$.

Next suppose that $c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\| \le \|\mathbf{z}\| \le c_3e^{4\|\mathbf{w}\|}$ for an arbitrary constant $c_3 > 0$. Here we have the upper bound $|\overline{f}(\mathbf{z},\mathbf{w})| \le 1 - c_2/\|\mathbf{w}\| \le e^{-c_2/\|\mathbf{w}\|}$ and consequently

$$\begin{aligned}
I_{24} &= \frac{1}{(2\pi)^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\int_{c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\|\le\|\mathbf{z}\|\le c_3e^{4\|\mathbf{w}\|}} \det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{z},\mathbf{w}\rangle}\,d\mathbf{w}\,d\mathbf{z}\\
&= O\left(\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}e^{-c_2T/\|\mathbf{w}\|+4d\|\mathbf{w}\|-\varepsilon\|\mathbf{w}\|^2}\,dw\right)\\
&= O\left(\int_{1\le\|\mathbf{w}\|\le c_2\sqrt{T}}\|\mathbf{w}\|^{-d-2}e^{-(1-c_2)\sqrt{T}-\varepsilon\|\mathbf{w}\|^2}\,d\mathbf{w} + \int_{c_2\sqrt{T}\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}e^{-\varepsilon/2\|\mathbf{w}\|^2}\,dw\right)\\
&= O\left(e^{-(1-c_2)\sqrt{T}} + e^{-(c_2^2/2)\varepsilon T}\right)
\end{aligned}$$

provided that $\varepsilon \gg T^{-1/2}$.

Finally if $\|\mathbf{z}\| \ge c_3e^{4\|\mathbf{w}\|}$, where c_3 is chosen sufficiently large, we have

$$\overline{f}(\mathbf{z},\mathbf{w}) \le \tilde{C}\frac{\max\left(\log(\|\mathbf{w}\|),\log(\|\mathbf{z}\|),e^{\|\mathbf{w}\|\|}\right)}{\sqrt{\|\mathbf{w}\|\,\|\mathbf{z}\|}} \le \frac{1}{2}. \tag{34}$$

In particular we assume that $\tilde{C}/\sqrt{c_3} \leq \frac{1}{2}$. Thus we find

$$\begin{aligned}
I_{25} &= \frac{1}{(2\pi)^d} \int_{1 \leq \|\mathbf{w}\| \leq \eta T} \int_{\|\mathbf{z}\| \geq c_3 e^{4\|\mathbf{w}\|}} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w},\mathbf{z}\rangle} d\mathbf{z}\, d\mathbf{w} \\
&= O\left(\tilde{C}^{T-d} \int_{1 \leq \|\mathbf{w}\| \leq \eta T} \|\mathbf{w}\|^{-d-2} \frac{(\log(\|\mathbf{w}\|))^{T-d}}{\|\mathbf{w}\|^{(T-d)/2}} \int_{\|\mathbf{z}\| \geq c_3 e^{4\|\mathbf{w}\|}} \|\mathbf{z}\|^{-(T-d)/2} d\mathbf{z}\, d\mathbf{w} \right) \\
&+ O\left(\tilde{C}^{T-d} \int_{1 \leq \|\mathbf{w}\| \leq \eta T} \|\mathbf{w}\|^{-d-2-(T-d)/2} \int_{\|\mathbf{z}\| \geq c_3 e^{4\|\mathbf{w}\|}} \frac{(\log(\|\mathbf{z}\|))^{T-d}}{\|\mathbf{z}\|^{(T-d)/2}} d\mathbf{z}\, d\mathbf{w} \right) \\
&+ O\left(\tilde{C}^{T-d} \int_{1 \leq \|\mathbf{w}\| \leq \eta T} \|\mathbf{w}\|^{-d-2-(T-d)/2} e^{(T-d)\|\mathbf{w}\|} \int_{\|\mathbf{z}\| \geq c_3 e^{4\|\mathbf{w}\|}} \|\mathbf{z}\|^{-(T-d)/2} d\mathbf{z}\, d\mathbf{w} \right) \\
&= O\left(\frac{\tilde{C}^T}{T c_3^{T/2}} \int_1^\infty r^{\frac{d}{2}-3-(T-d)} (\log r)^{T-d} e^{-2(T-d)r+4r}\, dr \right) \\
&+ O\left(\frac{\tilde{C}^T}{T c_3^{T/2}} \int_1^\infty r^{\frac{d}{2}-3-(T-d)} (\log(c_3 e^{4r}))^{T-d} e^{-2(T-d)r+4r}\, dr \right) \\
&+ O\left(\frac{\tilde{C}^T}{T c_3^{T/2}} \int_1^\infty r^{\frac{d}{2}-3-(T-d)} e^{-(T-d)r+4r}\, dr \right) \\
&= O\left(\frac{1}{T 2^T} \right).
\end{aligned}$$

Summing up we have

$$\begin{aligned}
J_2 &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{1 \leq \|\mathbf{w}\| \leq \eta T} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w},\mathbf{z}\rangle}\, d\mathbf{w}\, d\mathbf{z} \\
&= J_{21} + J_{22} + J_{23} + J_{24} + J_{25} \\
&= \frac{1}{\sqrt{2\pi T}} \int_{\|\mathbf{w}\| \geq 1} \sqrt{\overline{\det}(B(\mathbf{w}))} e^{-\frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1} \mathbf{w}} e^{-\varepsilon\|\mathbf{w}\|^2}\, d\mathbf{w} + O\left(\frac{\varepsilon^{-\frac{d}{4}-\frac{1}{2}}}{T^{d/2+1}} \right).
\end{aligned}$$

provided that $\varepsilon \gg T^{-1/2}$ and the constant c_6 is chosen sufficiently small.

C: *The case* $\|\mathbf{w}\| \geq \eta T$. If $\|\mathbf{z}\| \leq e^{4\|\mathbf{w}\|}$ then we use the trivial bound $|\overline{f}(z,w)| \leq 1$ and obtain for $\varepsilon \gg T^{-1/2}$

$$\begin{aligned}
I_{31} &= \frac{1}{(2\pi)^d} \int_{\|\mathbf{w}\| \geq \eta T} \int_{\|\mathbf{z}\| \leq e^{4\|\mathbf{w}\|}} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w},\mathbf{z}\rangle}\, d\mathbf{w}\, d\mathbf{z} \\
&= O\left(\int_{\|\mathbf{w}\| \geq \eta T} \|\mathbf{w}\|^{-d-2} e^{4d\|\mathbf{w}\| - \varepsilon\|\mathbf{w}\|^2}\, d\mathbf{w} \right) \\
&= O\left(\int_{r \geq \eta T} e^{-\frac{\varepsilon}{2} r^2}\, dr \right) = O\left(\frac{1}{\varepsilon \eta T} e^{-\frac{\varepsilon}{2}(\eta T)^2} \right) = O\left(e^{-\eta^2 T} \right).
\end{aligned}$$

If $\|\mathbf{z}\| \geq e^{4\|\mathbf{w}\|}$ we again use the upper bound (34) and obtain

$$\begin{aligned}
I_{32} &= \frac{1}{(2\pi)^d} \int_{\|\mathbf{w}\| \geq \eta T} \int_{\|\mathbf{z}\| \geq e^{4\|\mathbf{w}\|}} \det(B(\mathbf{z},\mathbf{w})) \overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon\|\mathbf{w}\|^2 - 2i\varepsilon\langle \mathbf{w},\mathbf{z}\rangle}\, d\mathbf{w}\, d\mathbf{z} \\
&= O\left(\tilde{C}^T \int_{|w| \geq \eta T} \|\mathbf{w}\|^{-d-2-(T-d)/2} e^{(T-d)\|\mathbf{w}\| - \varepsilon\|\mathbf{w}\|^2} \int_{\|\mathbf{z}\| \geq e^{4\|\mathbf{w}\|}} \|\mathbf{z}\|^{-(T-d)/2}\, d\mathbf{z}\, d\mathbf{w} \right) \\
&= O\left(\frac{\tilde{C}^T}{T} \int_{|w| \geq \eta T} \|\mathbf{w}\|^{-d-2-(T-d)/2} e^{(T-d)\|\mathbf{w}\|} e^{(d-(T-d)/2)4\|\mathbf{w}\|}\, d\mathbf{w} \right) \\
&= O\left(\frac{\tilde{C}^T}{T} \int_{r \geq \eta T} e^{(3d-T)r} \right) = O\left(e^{-\frac{\eta}{2} T^2} \right).
\end{aligned}$$

Consequently,

$$I_3 = \frac{1}{2\pi}\int_{\mathbb{R}^d}\int_{\|\mathbf{w}\|\geq\eta T} \det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{z}\rangle}\,d\mathbf{w}\,d\mathbf{z} = O\left(e^{-\eta^2 T}\right).$$

D: *The whole range*.

Summing up we arrive at

$$\tilde{J}_0 = \frac{1}{(2\pi T)^{d/2}}\int_{\mathbb{R}^d}\sqrt{\det(\overline{B}(\mathbf{w})}e^{-\varepsilon\|\mathbf{w}\|^2-\frac{2\varepsilon^2}{T}\mathbf{w}^\tau\overline{B}(\mathbf{w})^{-1}\mathbf{w})}\,d\mathbf{w} + O\left(\frac{\varepsilon^{-\frac{d}{4}-\frac{1}{2}}}{T^{d/2+1}}\right)$$

provided that $\varepsilon \gg T^{-1/2}$ and where we have set

$$\eta = \eta(T) = c_6(\log T)^{-3}$$

for a sufficiently small positive constant c_6. This completes the proof.

V. Other Losses

Finally, let us briefly discuss some other losses for which our methodology may still apply. For example, in [28] it was studied the case where $p(\langle\mathbf{w},\mathbf{x}\rangle) = \langle\mathbf{w},\mathbf{x}\rangle^2$ and the weight $\mathbf{w}$ as well as the feature $\mathbf{x}$ are unitary vectors on the unit circle. Here, $p(\langle\mathbf{w},\mathbf{x}\rangle) = \cos(\theta-\theta_Q-x_t)^2$ and $q(\langle\mathbf{w},\mathbf{x}\rangle) = 1-p(\langle\mathbf{w},\mathbf{x}\rangle) = \sin(\theta-\theta_Q-x_t)^2$ where θ is the argument of $\mathbf{w}$, x an element of $\mathbf{x}$, and θ_Q the unknown polarization. We still have the expression via Fourier transform:

$$S(\mathbf{x}^T) = \sum_{y^T}\frac{1}{2\pi}\int_0^{2\pi}\ell''_{y^T}(\theta)d\theta\int_{\mathbb{R}}e^{-i\ell'_{y^T}(\theta)z}dz$$

and the Hessian is strongly dependent of y^T. In fact,

$$\ell''_{y^T}(\theta) = \sum_{y_t<0}\frac{2}{\sin(\theta-\theta_Q-x_t)^2} + \sum_{y_t>1}\frac{2}{\cos(\theta-\theta_Q-x_t)^2}$$

and $\ell'_{y^T}(\theta) = 2\sum_{y_t>0}\tan(\theta-\theta_Q-x_t) + 2\sum_{y_t<0}\cot(\theta-\theta_Q-x_t)$. The terms may be regrouped leading to

$$S(\mathbf{x}^T) = \frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{R}}\tilde{\ell}''_T(\theta,z)\exp(L_T(\theta,z))d\theta dz$$

for

$$\tilde{\ell_T}''(\theta,z) = \sum_t \frac{2}{p_t(\mathbf{w},z)+q_t(\mathbf{w},z)}\left(\frac{p_t(\mathbf{w},z)}{\cos(\theta-\theta_Q-x_t)^2} + \frac{q_t(\mathbf{w},z)}{\sin(\theta-\theta_Q-x_t)^2}\right)$$

and

$$L_T(\theta,z) = \sum_t \log(p_t(\mathbf{w},z)+q_t(\mathbf{w},z))$$

with

$$\begin{array}{rcl} p_t(\mathbf{w},z) &=& p(\langle\mathbf{w}|\mathbf{x}_t\rangle)\exp(-i\tan(\theta-\theta_Q-x_t)z) \\ q_t(\mathbf{w},z) &=& q(\langle\mathbf{w}|\mathbf{x}_t\rangle)\exp(-i\cot(\theta-\theta_Q-x_t)z). \end{array}.$$

Observe that $\ell''_T(\theta,z)$ and $L_T(\theta,z)$ are both $O(T)$, so that we can apply the saddle point method. Notice that in this case there is no need for a normalized regret because $\mathbf{x}$ and $\mathbf{w}$ are unitary vectors.

VI. Conclusion

In this paper we studied the regularized regret under logarithmic loss for the logistic function with *unbounded* d-dimensional weights, via a new complexity named *regularized* Shtarkov sum. Our main results provide the first *precise* characterization of the regularized Shtarkov sum and consequently the regularized regret with unbounded weights up to second order asymptotics. Furthermore, we also show how the leading asymptotics of the standard regret can be inferred from the regularized regret. Notably, unlike the $d/2\log T$ regret growth known only for bounded weights, our result implied that the regularized regret grows as $(1/2+\alpha/4)d\log T$ when the regularization parameter is of order $\Theta(T^{-\alpha})$ for $\alpha \leq 1/2$. This provides the first known *fine-grained* characterization of the minimax regret with an unbounded weight norm. We accomplish this using tools from analytic combinatorics, such as multidimensional Fourier, saddle point method, and Mellin transform, which we believe is of independent interests.

There are several directions for future work. First, we need an extension to non-logistic functions. Next, we may relax our assumption on ε when it goes to zero. A bigger challenge is to extend our Fourier approach to obtain second order asymptotics of the regular regret with a hard truncation on the weight norm (unlike our ε-regularized constraints).

Acknowledgment

We would like to thank the referees for their valuable suggestions, which have improved the presentation of this paper.

Appendix A
Existence of the Inverse Fourier Transform

In this Appendix we prove the existence of the Fourier and its inverse.

Lemma 6. *For every fixed y^T, $\mathbf{x}^T$, $\varepsilon > 0$ and for all non-negative integers $k_1, \dots, k_d$ we have*

$$\frac{\partial^{k_1+\cdots k_d}}{\partial a_1^{k_1}\cdots\partial a_d^{k_d}} h_{y^T|\mathbf{x}^T}(\mathbf{a}) = O\left(e^{-\frac{1}{4\varepsilon}\|\mathbf{a}\|^2}\|\mathbf{a}\|^{k_1+\cdots+k_d}\right). \tag{A-1}$$

Furthermore, $\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z}) = \int_{\mathbb{R}^d} h_{y^T|\mathbf{x}^T}(\mathbf{a})e^{-i\langle\mathbf{a},\mathbf{z}\rangle}\,d\mathbf{a}$ of $h_{y^T|\mathbf{x}^T}(\mathbf{a})$ exists and satisfies

$$\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z}) = O\left(|z_1|^{-k_1}\cdots|z_d|^{-k_d}\right) \tag{A-2}$$

for all non-negative integers $k_1, \dots, k_d$. Consequently the inverse Fourier transform is given by

$$h_{y^T|\mathbf{x}^T}(\mathbf{a}) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z})e^{i\langle\mathbf{a},\mathbf{z}\rangle}\,d\mathbf{z}.$$

We start the proof of Lemma 6 with the following result.

Lemma 7. *The determinant of the matrix matrix $\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})$ satisfies*

$$\det\left(\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})\right) \geq (2\varepsilon)^d.$$

Proof. By (14) we have for every vector $\mathbf{v}$

$$\begin{aligned}\left\langle\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})\mathbf{v},\mathbf{v}\right\rangle &= \sum_{t=1}^{T} p(\langle\mathbf{x}_t,\mathbf{w}\rangle))p(-\langle\mathbf{x}_t,\mathbf{w}\rangle))\langle\mathbf{x}_t,\mathbf{v}\rangle^2 + 2\varepsilon\|\mathbf{v}\|^2 \\ &\geq 2\varepsilon\|\mathbf{v}\|^2.\end{aligned}$$

In particular if $\mathbf{v}$ is an eigenvector of $\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})$ with eigenvalue λ then

$$\lambda\|\mathbf{v}\|^2 \geq 2\varepsilon\|\mathbf{v}\|^2.$$

Since $\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})$ is real symmetric, its determinant is just the product of all its eigenvalues. This completes the proof of the lemma. □

Proof of Lemma 6. We start with the proof of (A-1). By (16) the case $k_1 = \cdots = k_d = 0$ is already covered. Next let us consider the first derivatives of $e^{-L(y^T|\mathbf{x}^T,\mathbf{w})}$:

$$\begin{aligned}&\nabla \exp\left(-L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right)\\ &\qquad = -\exp\left(-L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right)\nabla L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\nabla G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}).\end{aligned}$$

Clearly we have

$$\exp\left(-L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right) \leq 1.$$

Since

$$\nabla L(y^T|\mathbf{x}^T, \mathbf{w}) = -\sum_{t=1}^{T} p(-y_t\langle \mathbf{x}_t, \mathbf{w}\rangle)\, y_t\, \mathbf{x}_t = O(1)$$

uniformly for all $\mathbf{w} \in \mathbb{R}^d$ we also have

$$\nabla L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a})) = O(1).$$

Finally

$$\nabla G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}) = \left(\nabla^2 L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right)^{-1}.$$

All entries of the matrix $\nabla^2 L(y^T|\mathbf{x}^T, \mathbf{w})$ are uniformly bounded as well as the reciprocal of its determinant (see Lemma 7). This proves that

$$\nabla G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}) = O(1)$$

and hence

$$\begin{aligned}&\nabla \exp\left(-L_\varepsilon(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right)\\ &\quad = \exp\left(-\varepsilon\|G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\|^2\right)\nabla \exp\left(-L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right)\\ &\quad - \exp\left(-\varepsilon\|G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\|^2 - L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right) 2\varepsilon G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\nabla G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a})\\ &\quad = O\left(e^{-\frac{1}{4\varepsilon}\|\mathbf{a}\|^2}\|\mathbf{a}\|\right)\end{aligned}$$

as proposed.

In a similar way we can compute higher derivatives. For all non-negative integers $k_1, \ldots k_d$ we find

$$\frac{\partial^{k_1+\cdots k_d}}{\partial a_1^{k_1}\cdots\partial a_d^{k_d}}\exp\left(-L(y^T|\mathbf{x}^T, G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\right) = O(1).$$

The computations follow the same lines as above. It remains to note that

$$\frac{\partial^{k_1+\cdots k_d}}{\partial a_1^{k_1}\cdots\partial a_d^{k_d}}\exp\left(-\varepsilon\|G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{a}))\|^2\right) = O\left(e^{-\frac{1}{4\varepsilon}\|\mathbf{a}\|^2}\|\mathbf{a}\|^{k_1+\cdots+k_d}\right)$$

and to apply the product rule. This completes the proof of (A-1).

In a second step we prove (A-2). Since

$$\begin{aligned}\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z}) &= \int_{\mathbb{R}^d} h_{y^T|\mathbf{x}^T}(\mathbf{a})e^{-i\langle\mathbf{a},\mathbf{z}\rangle}\, d\mathbf{a}\\ &\qquad = (iz_1)^{-k_1}\cdots(iz_d)^{-k_d}\int_{\mathbb{R}^d}\frac{\partial^{k_1+\cdots k_d}}{\partial a_1^{k_1}\cdots\partial a_d^{k_d}}h_{y^T|\mathbf{x}^T}(\mathbf{a})e^{-i\langle\mathbf{a},\mathbf{z}\rangle}\, d\mathbf{a}.\end{aligned}$$

It follows that the integral

$$\int_{\mathbb{R}^d}\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z})e^{i\langle\mathbf{a},\mathbf{z}\rangle}\, d\mathbf{z}$$

exists. Since all involved functions are continuously differentiable this integral equals (up to the factor $(2\pi)^{-d}$) the original function $h_{y^T|\mathbf{x}^T}(\mathbf{a})$. □

Appendix B
Eigenvalues of $\bar{\mathbf{B}}(\mathbf{w})$

We recall the notation $\mathcal{B}_d$ for a d-dimensional unit ball. The d-dimensional volume v_d of $\mathcal{B}_d$ is given by

$$v_d = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}.$$

We also note that $\mathcal{S}_d$ denotes the corresponding sphere that has ($(d-1)$-dimensional) surface measure $s_d = d\,v_d = 2\pi^{d/2}/\Gamma\left(\frac{d}{2}\right)$.

Lemma 8. (i) *Suppose that $d \geq 2$ and let $\mathbf{u} = \mathbf{w}/||\mathbf{w}||$ and set $q(x) = p(x)p(-x) = p(x)(1-p(x))$. Then we have the following expression*

$$\bar{\mathbf{B}}(\mathbf{w}) = \phi(\mathbf{w})(\mathbf{I}_d - \mathbf{u}\otimes\mathbf{u}) + \lambda(\mathbf{w})\mathbf{u}\otimes\mathbf{u}$$

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

$$\lambda(\mathbf{w}) = \frac{v_{d-1}}{v_d}\int_0^\pi \cos(\theta)^2\sin(\theta)^d q(\cos(\theta)\|\mathbf{w}\|)d\theta \tag{A-3}$$

and

$$\phi(\mathbf{w}) = \frac{1}{d+2}\int_0^\pi \sin(\theta)^{d+2} q(\cos(\theta)\|\mathbf{w}\|)d\theta \tag{A-4}$$

are the eigenvalues of $\bar{\mathbf{B}}(w)$ with multiplicity 1 and $d-1$.

(ii) *The eigenvalue $\lambda(\mathbf{w})$ is asympotically*

$$\lambda(\mathbf{w}) = \frac{v_{d-1}\pi^2}{3v_d}\|\mathbf{w}\|^{-3}\left(1+O(\|\mathbf{w}\|^{-2})\right)$$

whereas the eigenvalue $\phi(\mathbf{w})$ is asympotically

$$\phi(\mathbf{w}) = \frac{1}{(d+2)\|\mathbf{w}\|}\left(1+O(\|\mathbf{w}\|^{-2})\right).$$

Consequently

$$\det\bar{\mathbf{B}}(\mathbf{w}) = \lambda(\mathbf{w})\phi^{d-1}(\mathbf{w}) = \frac{v_{d-1}}{v_d}\frac{\pi^2}{3(d+2)^{d-1}}\|\mathbf{w}\|^{-d-2}\left(1+O(\|\mathbf{w}\|^{-2})\right).$$

Note that Lemma 8 is consistent with the case $d=1$. Here we have, as $w\to\infty$.

$$\overline{B}(w) = \int_0^1 q(wx)x^2\,dx \sim \frac{\pi^2}{6}.$$

Proof. We start with part (i). Let θ be the angle between $\mathbf{x}$ and $\mathbf{u} = \mathbf{w}/\|\mathbf{w}\|$. We have the decomposition $\mathbf{x} = \cos(\theta)\mathbf{u} + \mathbf{b}$ with $\mathbf{b}\in\sin\theta\mathcal{B}_{d-1}(\mathbf{u})$ where $\mathcal{B}_{d-1}(\mathbf{u})$ is the unit hypersphere orthogonal to $\mathbf{u}$. Since $\mathbf{x}$'s have a spherical symmetry in its distribution, so it is the case for the $\mathbf{b}$'s in $\sin\theta\mathcal{B}_{d-1}(\mathbf{u})$ for any given angle θ. Thus

$$\begin{aligned}\bar{\mathbf{B}}(\mathbf{w}) &= \frac{1}{v_d}\int_0^\pi q(\|\mathbf{w}\|\cos\theta)\sin\theta\,d\theta\int_{\sin\theta\mathcal{B}_{d-1}(\mathbf{u})}(\mathbf{b}+\cos\theta\mathbf{u})\otimes(\mathbf{b}+\cos\theta\mathbf{u})d\mathbf{b} \\ &= \frac{1}{v_d}\int_0^\pi q(\|\mathbf{w}\|\cos\theta)\sin\theta\,d\theta\int_{\sin\theta\mathcal{B}_{d-1}(\mathbf{u})}(\mathbf{b}\otimes\mathbf{b}+(\cos\theta)^2\mathbf{u}\otimes\mathbf{u})d\mathbf{b} \\ &+ \frac{1}{v_d}\int_0^\pi q(\|\mathbf{w}\|\cos\theta)\sin\theta\,d\theta\int_{\sin\theta\mathcal{B}_{d-1}(\mathbf{u})}\cos\theta(\mathbf{b}\otimes\mathbf{u}+\mathbf{u}\otimes\mathbf{b})d\mathbf{b}.\end{aligned} \tag{A-5}$$

Again due to the spherical symmetry of $\mathbf{b}$ we also have $\int_{\sin\theta\mathcal{B}_{d-1}(\mathbf{u})}\mathbf{b} = 0$ leading to

$$\begin{aligned}\bar{\mathbf{B}}(\mathbf{w}) &= \frac{1}{v_d}\int_0^\pi q(\|\mathbf{w}\|\cos\theta)\sin\theta\,d\theta\int_{\sin\theta\mathcal{B}_{d-1}(\mathbf{u})}(\mathbf{b}\otimes\mathbf{b}+(\cos\theta)^2\mathbf{u}\otimes\mathbf{u}d\mathbf{b} \\ &= \frac{1}{v_d}\int_0^\pi q(\|\mathbf{w}\|\cos\theta)(\sin\theta)^d d\theta\int_{\mathcal{B}_{d-1}(\mathbf{u})}((\sin\theta)^2\mathbf{b}\otimes\mathbf{b}+(\cos\theta)^2\mathbf{u}\otimes\mathbf{u})d\mathbf{b}.\end{aligned}$$

The $(\sin\theta)^{d-1}$ factor and the factor $(\sin\theta)^2$ arise from the change of integration domain from $\sin\theta\mathcal{B}_{d-1}(\mathbf{u})$ to $\mathcal{B}_{d-1}(\mathbf{u})$.

It is an easy exercise to show that

$$\int_{\mathcal{B}_{d-1}(\mathbf{u})} \mathbf{b}\otimes\mathbf{b}d\mathbf{b} = \frac{v_d}{d+2}\mathbf{I}_{d-1}(\mathbf{u}) = \frac{v_d}{d+2}(\mathbf{I}_d - \mathbf{u}\otimes\mathbf{u}). \tag{A-6}$$

Furthermore we obviously have

$$\int_{\mathcal{B}_{d-1}(\mathbf{u})} \mathbf{u}\otimes\mathbf{u}d\mathbf{b} = v_{d-1}\mathbf{u}\otimes\mathbf{u} \tag{A-7}$$

which completes the proof of part (i) of the lemma.

Now we move to part (ii) of Lemma 8. Both $\lambda(\mathbf{w})$ and $\phi(\mathbf{w})$ are functions of $w = \|\mathbf{w}\|$. We write $\lambda(w) = \lambda(\|\mathbf{w}\|)$ and $\phi(w) = \phi(\|\mathbf{w}\|)$. To capture the asymptotics of these functions we apply the Mellin transform which is an effective tool of analytic combinatorics for complex asymptotics. The reader is refereed to [24] and [25] for detailed discussions.

The Mellin transforms $\lambda^*(s)$ and $\phi^*(s)$ of $\lambda(w)$ and $\phi(w)$ are defined, respectively, as

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

Observe now that

$$\begin{aligned}\lambda(w) &= \frac{2v_{d-1}}{v_d}\int_0^{\pi/2} q(\cos(\theta)w)\cos^2(\theta)\sin^d(\theta)d\theta \\ &= \frac{2v_{d-1}}{v_d}\int_0^1 y^2(1-y^2)^{(d-1)/2}q(yw)dy\end{aligned}$$

via the change of variable $y = \cos(\theta)$. Thus we find

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

where

$$q^*(s) = \Gamma(s)\zeta(s-1)(1-2^{-(s-2)})$$

is the Mellin transform of $q(x) = p(x)(1-p(x))$ and

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

is the Mellin transform of the function $(1-y^2)^{(d-1)/2}$ defined over $[0,1]$. Note that $\Gamma(s)$ is the Euler gamma function and $\zeta(s)$ is the Riemann zeta function. Hence,

$$\lambda^*(s) = \frac{v_{d-1}}{v_d}\Gamma(s)\zeta(s-1)(1-2^{-(s-2)})\frac{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{3-s}{2}\right)}{\Gamma\left(\frac{d+4-s}{2}\right)}.$$

Note that $q^*(s)$ and $\beta_1^*(s)$ are analytic for $\mathrm{Re}(s) > 0$. Both functions can be meromorphically continued. The function $q^*(s)$ has poles on the non-positive integers and the function $\beta_1^*(s)$ has poles on the non-positive even integers. Thus, the dominating singularity of $\lambda^*(s)$ is a simple pole at $s = 3$ (coming from $\Gamma\left(\frac{3-s}{2}\right)$) with residue

$$\mathrm{res}(\lambda^*, s=3) = -\frac{2v_{d-1}}{v_d}\zeta(2).$$

This implies that

$$\lambda(w) = \frac{2v_{d-1}}{v_d}\zeta(2)w^{-3}(1+O(w^{-2})) = \frac{v_{d-1}\pi^2}{3v_d}w^{-3}(1+O(w^{-2}))$$

as $w \to \infty$, since $s = 5$ is the next pole of $\lambda^*(s)$.

We can make a similar analysis for $\phi^*(s)$ and we 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,

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

Here the dominating singularity is a simple pole at $s = 1$ with residue

$$\operatorname{res}(\phi^*, s = 1) = \frac{-1}{d+2}.$$

(Note that $\zeta(0) = -\frac{1}{2}$.) The next pole is $s = 3$. Thus, we have

$$\phi(w) = \frac{1}{(d+2)w}(1 + O(w^{-3})).$$

This completes the proof of the lemma. □

Appendix C
Proof of Lemma 3–5

We start with the proof of Lemma 3.

Proof of Lemma 3. We use the representation (28) that we can rewrite to

$$\overline{f}(w,z) = \int_0^1 \left(\frac{e^{-ixz/(1+e^{xw})}}{1+e^{-xw}} + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \right) dx.$$

By differentiation we directly obtain

$$\begin{aligned}
\overline{f}(w,0) &= \int_0^1 \left(\frac{1}{1+e^{-xw}} + \frac{1}{1+e^{xw}} \right) dx = \int_0^1 1\, dx = 1,\\
\frac{\partial \overline{f}}{\partial z}(w,z) &= \int_0^1 \left(\frac{e^{-ixz/(1+e^{x}w)}}{1+e^{-xw}} \frac{-ix}{1+e^{xw}} + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \frac{ix}{1+e^{-xw}} \right) dx,\\
\frac{\partial \overline{f}}{\partial z}(w,0) &= \int_0^1 0\, dx = 0,\\
\frac{\partial^2 \overline{f}}{\partial z^2}(w,z) &= \int_0^1 \left(\frac{e^{-ixz/(1+e^{x}w)}}{1+e^{-xw}} \left(\frac{-ix}{1+e^{xw}} \right)^2 + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \left(\frac{ix}{1+e^{-xw}} \right)^2 \right) dx,\\
\frac{\partial^2 \overline{f}}{\partial z^2}(w,0) &= \int_0^1 \left(\frac{1}{1+e^{-xw}} \left(\frac{-ix}{1+e^{xw}} \right)^2 + \frac{1}{1+e^{xw}} \left(\frac{ix}{1+e^{-xw}} \right)^2 \right) dx,\\
&= -\int_0^1 \frac{x^2}{(1+e^{-xw})(1+e^{xw})} \left(\frac{1}{1+e^{-xw}} + \frac{1}{1+e^{xw}} \right) dx\\
&= -\int_0^1 \frac{x^2}{(1+e^{-xw})(1+e^{xw})}\, dx = -\overline{B}(w),\\
\frac{\partial^3 \overline{f}}{\partial z^3}(w,z) &= \int_0^1 \left(\frac{e^{-ixz/(1+e^{x}w)}}{1+e^{-xw}} \left(\frac{-ix}{1+e^{xw}} \right)^3 + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \left(\frac{ix}{1+e^{-xw}} \right)^3 \right) dx.
\end{aligned}$$

Clearly we have

$$\overline{B}(w) = \Theta\left(\int_0^1 x^2 e^{-|xw|}\,dx\right)$$
$$= \Theta\left(\frac{1}{|w|^3}\int_0^{|w|} v^3 e^{-v}\,dv\right)$$
$$= \Theta\left(\min(1,|w|^{-3})\right).$$

Similarly it follows that

$$\frac{\partial^3 \overline{f}}{\partial z^3}(w,z) = O\left(\int_0^1 x^2 e^{-|xw|}\,dx\right) = O\left(\min(1,|w|^{-4})\right).$$

Thus, it immediately follows that

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

and by expanding $\overline{f}(z,w) = e^{\log \overline{f}(z,w)}$ we obtain the third representation for $\overline{f}(z,w)$ (where we use $z^2\min(1,|w|^{-6})$ as the order of $z^2\overline{B}(w)^2$).

Finally, if $|z| \le \max(1, c_1|w|)$ (for some sufficiently small constant $c_1 > 0$) it follows that

$$z^2\overline{B}(w) \ge z^3\min(1,|w|^{-4}).$$

Thus we also get

$$\overline{f}(w,z) = 1 - \Theta\left(z^2\min(1,|w|^{-3})\right).$$

□

For the proof of Lemma 4 we need to further properties (that can be found in [29]),

Lemma 9. *Let β_1, β be real numbers with $\beta_1 < \beta_2$. Assume that h is continuously differentiable on $[\beta_1, \beta_2]$ and has a monotone nonvanishing derivative. Then for each continuous function g we have*

$$\left|\int_{\beta_1}^{\beta_2} g(x)e^{ih(x)}\,dx\right| \le 2\frac{\max_{[\beta_1,\beta_2]}|g| + V_{\beta_1}^{\beta_2}(g)}{\min_{[\beta_1,\beta_2]}|h'|}, \tag{A-8}$$

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

Lemma 10. *Let β_1, β be real numbers with $\beta_1 < \beta_2$. 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 g we have*

$$\left|\int_{\beta_1}^{\beta_2} g(x)e^{ih(x)}\,dx\right| \le 8\frac{\max_{[\beta_1,\beta_2]}|g| + V_{\beta_1}^{\beta_2}(g)}{\min_{[\beta_1,\beta_2]}\sqrt{|h''|}}. \tag{A-9}$$

The proof of Lemma 4 runs as follows.

Proof of Lemma 4. We consider the function

$$h(x) = \frac{xz}{1+e^{-xw}} = xzp(xw)$$

that satisfies

$$h'(x) = z\frac{1+e^{-xw}+xwe^{-xw}}{(1+e^{-xw})^2} = z\frac{1+e^{-u}+ue^{-u}}{(1+e^{-u})^2}$$

and

$$h''(x) = z\left(2\frac{we^{-xw}}{(1+e^{-xw})^2} + w^2xe^{-xw}\frac{1-e^{-xw}}{(1+e^{-xw})^3}\right)$$
$$= \frac{zue^{-u}}{x}\,\frac{2(1+e^{-u})+u(1-e^{-u})}{(1+e^{-u})^3},$$

where u abbreviates $u = xw$. Note that

$$\overline{f}(x,w) = \int_0^1 p(-xw)e^{ih(x)}\,dx + \int_0^1 p(xw)e^{ih(-x)}\,dx.$$

First we consider the case $w \ge 0$ (so that $u = xw \ge 0$). Here we certainly have

$$|h'(x)| \ge \frac{|z|}{4} \tag{A-10}$$

and that $h''(x)$ has the same sign as z. Hence, by a direct application of Lemma 9 we obtain

$$\left|\int_0^1 p(-xw)e^{ih(x)}\,dx\right| \le \frac{8}{|z|}. \tag{A-11}$$

Note that the function $p(-xw)$ is monotone and bounded by 1.

Next observe that there is $u_0 < -1$ such that $1 + e^{-u} + ue^{-u}$. Furthermore we also have that

$$2(1+e^{-u}) + u(1-e^{-u}) \ge 2 - e^{-1} > 0$$

for $u \le 0$. Thus, if $0 \le w \le 1$ and $0 \le x \le 1$ we have

$$|h'(-x)| \ge \frac{|z|}{(1+e)^2} \tag{A-12}$$

and consequently we get

$$\left|\int_0^1 p(xw)e^{ih(-x)}\,dx\right| \le \frac{2(1+e)^2}{|z|}$$

which implies

$$\overline{f}(z,w) = O\left(\frac{1}{|z|}\right).$$

Trivially we have $|\overline{f}(z,w)| \le 1$. The case $-1 \le w < 0$ can be handled in completely the same way. Thus, we have completed the case $|w| \le 1$.

If $|w| \ge 1$ we have to be more careful. First we again have (A-10) which implies (A-11).

However, for the second integral we have to distinguish between three ranges. If $0 \le x \le 1/w$ then we get again (A-12) and, thus,

$$\left|\int_0^{1/w} p(xw)e^{ih(-x)}\,dx\right| \le \frac{2(1+e)^2}{|z|}$$

Secondly we consider the interval $1/w \le x \le (|u_0| + \kappa)/w$ (for some $\kappa > 0$) then $h'(-x)$ is very close to 0 (and actually equal to 0 for $x = |u_0|/w$). So instead of Lemma 9 we apply Lemma 10 and obtain

$$\left|\int_{1/w}^{(|u_0|+\kappa)/w} p(xw)e^{ih(-x)}\,dx\right| = O\left(\frac{1}{\sqrt{|zw|}}\right)$$

since

$$|h''(x)| = \Theta\left(\frac{|z|}{x}\right) = \Theta(|zw|)$$

in this range.

Finally if $(|u_0| + \kappa)/w \le x \le 1$ we again apply Lemma 9. In this range we have

$$|h'(x)| \ge c|z|\frac{w}{e^w}$$

(for a proper constant $c > 0$) which gives

$$\left|\int_{(|u_0|+\kappa)/w}^{1} p(xw)e^{ih(-x)}\,dx\right| = O\left(\frac{e^w}{|zw|}\right).$$

This completes the proof of the lemma since the case $w < -1$ can be handled in completely the same way. □

Finally we give a proof of Lemma 5.

Proof of Lemma 5. We consider first the case $|x| \ge c_1'|w|$, where c_1' will be chosen sufficiently large. As in the proof of Lemma 3 it follows (if $w \ge 0$)

$$\left| \int_0^1 p(-xw) e^{ih(x)}\, dx \right| = O\left(\frac{1}{|z|}\right)$$

and

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

Thus, if $|z| \ge c_1' w$ for a sufficiently large constant c_1' we have

$$\left| \int_0^1 \left(p(-xw) e^{ih(x)} + p(xw) e^{ih(-x)} \right) dx \right| \le 1 - \frac{c_2}{w}.$$

Next we consider the interval $c_1 w \le |z| \le c_1' w$. With $c = z/w$ we have

$$\begin{aligned} \overline{f}(z,w) &= \int_0^1 \left(p(-xw) e^{-ixzp(xw)} + p(xw) e^{ixzp(-xw)} \right) dx \\ &= \frac{1}{w} \int_0^w \left(p(-v) e^{-icvp(v)} + p(v) e^{icvp(-v)} \right) dv. \end{aligned}$$

By continuity it follows that uniformly for $c_1 \le c \le c_1'$

$$\left| \int_0^1 p(-v) e^{-icvp(v)}\, dv \right| \le \int_0^1 p(-v)\, dv - c_2$$

for some constant $c_2 > 0$. Hence

$$|\overline{f}(z,w)| \le \frac{1}{w} \int_0^w \left(p(-v) + p(v) \right) dv - \frac{c_2}{w} = 1 - \frac{c_2}{w},$$

as proposed. (The case $w < 0$ is completely similar.) □

Appendix D
Upper bounds for $\overline{f}(\mathbf{w}, \mathbf{z})$

We need analogous for Lemmas 3–5 for $d > 1$. Actually the situation is slightly more difficult.

Lemma 11. *We have uniformly for $\mathbf{z} \in \mathbb{R}^d$ for some $\tau > 0$*

$$\mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z} \ge \begin{cases} c_1 \|\mathbf{z}\|^2 & \text{for } \|\mathbf{w}\| \le 1, \\ c_2 \frac{\|\mathbf{z}\|^2}{\|\mathbf{w}\|^3} \left(|\cos\varphi| + \|\mathbf{w}\| |\sin\varphi| \right)^2 & \text{for } \|\mathbf{w}\| > 1 \end{cases}$$

for proper constants $c_1, c_2 > 0$, where φ denotes the angle between $\mathbf{w}$ and $\mathbf{z}$, that is $\cos\varphi = \langle \mathbf{w}, \mathbf{z} \rangle / (\|\mathbf{w}\|\, \|\mathbf{z}\|)$. Furthermore

$$\begin{aligned} \log \overline{f}(\mathbf{w}, \mathbf{z}) &= -\frac{1}{2} \mathbf{z}^\tau \overline{B}(\mathbf{w}) \mathbf{z} \\ &\quad + \begin{cases} O(\|\mathbf{z}\|^3 + \|\mathbf{z}\|^4) & \text{for } \|\mathbf{w}\| \le 1, \\ O\left(\frac{\|\mathbf{z}\|^3}{\|\mathbf{w}\|^4} + \frac{\|\mathbf{z}\|^2 |\sin\varphi|^2}{\|\mathbf{w}\|} \right) & \text{for } \|\mathbf{w}\| > 1. \end{cases} \end{aligned}$$

In particular if $\|\mathbf{z}\| \le c\|\mathbf{w}\|$ for a sufficiently small constant $c > 0$ and if $\|\mathbf{w}\| > 1$ we uniformly have

$$\mathrm{Re}\left(\log \overline{f}(\mathbf{w}, \mathbf{z}) \right) \le -C \left(\frac{\|\mathbf{z}\|^2}{\|\mathbf{w}\|^3} + \frac{\|\mathbf{z}\|^2 |\sin\varphi|^2}{\|\mathbf{w}\|} \right)$$

for a proper constant $C > 0$.

Proof. The case $\|\mathbf{w}\| \le 1$ is easy to handle. We just use Taylor expansion and the property that (for proper constants $c_3, c_4 > 0$)

$$\begin{aligned}\mathbf{z}^\tau \overline{B}(\mathbf{w})\mathbf{z} &= \mathbb{E}\left[p(\langle \mathbf{X}, \mathbf{w}\rangle)p(-\langle \mathbf{X}, \mathbf{w}\rangle)\langle \mathbf{X}, \mathbf{z}\rangle^2\right] \\ &\ge c_3 \mathbb{E}\left[\langle \mathbf{X}, \mathbf{z}\rangle^2\right] \\ &\ge c_4 \|\mathbf{z}\|^2.\end{aligned}$$

In the case $\|\mathbf{w}\| > 1$ we have to be more careful. Let $\mathbf{w}_0 = \mathbf{w}/\|\mathbf{w}\|$ and $\mathbf{w}_1 = \tilde{\mathbf{w}}/\|\tilde{\mathbf{w}}\|$, where $\tilde{\mathbf{w}} = \mathbf{z} - \langle \mathbf{z}, \mathbf{w}\rangle/\|\mathbf{w}\|^2\, \mathbf{w}$ is orthogonal to $\mathbf{w}$. We now represent $\mathbf{x}$ as $\mathbf{x} = x_1\mathbf{w}_0 + x_2\mathbf{w}_1 + \mathbf{x}_3$, where $\mathbf{x}_3$ is orthogonal to $\mathbf{w}$ and $\mathbf{z}$. With the help of this notation we have

$$\langle \mathbf{x}, \mathbf{z}\rangle = x_1 \frac{\langle \mathbf{z}, \mathbf{w}\rangle}{\|\mathbf{w}\|^2} + x_2 \|\tilde{\mathbf{w}}\|.$$

We also note that

$$A = \frac{\langle \mathbf{z}, \mathbf{w}\rangle}{\|\mathbf{w}\|} = \|z\|\, \cos\varphi \quad \text{and} \quad B = \|\tilde{\mathbf{w}}\| = \|z\|\, |\sin\varphi|.$$

and that

$$\begin{aligned}\mathbf{z}^\tau \overline{B}(\mathbf{w})\mathbf{z} &= \mathbb{E}\left[p(\langle \mathbf{X}, \mathbf{w}\rangle)p(-\langle \mathbf{X}, \mathbf{w}\rangle)\langle \mathbf{X}, \mathbf{z}\rangle^2\right] \\ & \mathbb{E}\left[p(x_1\|\mathbf{w}\|)p(-x_1\|\mathbf{w}\|)(Ax_1 + Bx_2)^2\right]\end{aligned}$$

is a positive definite quadratic form in A, B. Thus, we get the lower bound

$$\begin{aligned}\mathbf{z}^\tau \overline{B}(\mathbf{w})\mathbf{z} &\ge c_5 A^2 \mathbb{E}\left[p(x_1\|\mathbf{w}\|)p(-x_1\|\mathbf{w}\|)x_1^2\right] + B^2 \mathbb{E}\left[p(x_1\|\mathbf{w}\|)p(-x_1\|\mathbf{w}\|)x_2^2\right] \\ &\ge c_6 \frac{A^2}{\|\mathbf{w}\|^3} + \frac{B^2}{\|\mathbf{w}\|} \ge c_7 \frac{\|\mathbf{z}\|^2}{\|\mathbf{w}\|^3}\left(|\cos\varphi| + \|\mathbf{w}\||\sin\varphi|\right)^2\end{aligned}$$

(for proper constants $c_5, c_5, c_7 > 0$) as proposed.

The second part of the lemma follows by applying first Taylor expansion for $\overline{f}(\mathbf{w}, \mathbf{z})$ and then by taking the logarithm. However, the computations are more involved. For the third derivative we obtain

$$\frac{\partial^3 \overline{f}(\mathbf{w}, \mathbf{z})}{\partial z_j \partial z_k \partial z_\ell} = i\mathbb{E}\left[X_j X_k X_\ell p(\langle \mathbf{X}, \mathbf{w}\rangle)p(-\langle \mathbf{X}, \mathbf{w}\rangle)\left(p(-\langle \mathbf{X}, \mathbf{w}\rangle)^2 e^{-i\langle X, \mathbf{z}\rangle p(-\langle \mathbf{X}, \mathbf{w}\rangle)} - p(\langle \mathbf{X}, \mathbf{w}\rangle)^2 e^{i\langle X, \mathbf{z}\rangle p(\langle \mathbf{X}, \mathbf{w}\rangle)}\right)\right]$$

which gives

$$\frac{\partial^3 \overline{f}(\mathbf{w}, \theta\mathbf{z})}{\partial \theta^3} = i\mathbb{E}\left[p(\langle \mathbf{X}, \mathbf{w}\rangle)p(-\langle \mathbf{X}, \mathbf{w}\rangle)\left(p(-\langle \mathbf{X}, \mathbf{w}\rangle)^2 e^{-i\langle X, \theta\mathbf{z}\rangle p(-\langle \mathbf{X}, \mathbf{w}\rangle)} - p(\langle \mathbf{X}, \mathbf{w}\rangle)^2 e^{i\langle \mathbf{X}, \theta\mathbf{z}\rangle p(\langle \mathbf{X}, \mathbf{w}\rangle)}\right)\langle \mathbf{X}, \mathbf{z}\rangle^3\right].$$

If $\|\mathbf{w}\||\sin\varphi| \le 1$ it is easy (by using the methods from above) to obtain an upper bound for this integral of the form $O(\|\mathbf{z}\|^3/\|\mathbf{w}\|^4)$.

However, if $\|\mathbf{w}\||\sin\varphi| \ge 1$ we have to be more careful. By using the parametrization from above and the integral representation of the remainder term in Taylor's theorem it turns out that the error term is essentially bounded above by the integral

$$\tilde{I} = \|\mathbf{z}\|^3 \int_0^1 \iint\limits_{x_1^2 + x_2^2 \le 1} p(x_1\|\mathbf{w}\|)^3 p(-x_1\|\mathbf{w}\|) e^{i\theta\|\mathbf{z}\| p(x_1\|\mathbf{w}\|)(x_1\cos\varphi + x_2\sin\varphi)} (\sin\varphi)^3 x_2^3 (1-\theta)^2 \, dx_1\, dx_2\, d\theta.$$

The integral with respect to x_2 can be upper bounded by

$$O\left(\frac{1}{1 + \theta\|\mathbf{z}\| p(x_1\|\mathbf{w}\|)|\sin\varphi|}\right).$$

Furthermore we have (uniformly for $A > 0$)

$$\int_0^1 \frac{1}{1 + A\theta}(1-\theta)^2\, d\theta = O\left(\frac{1}{A^2} + \frac{\log(1+A)}{A}\right).$$

By simple computations this leads to an upper bound for $\tilde{I}$ of the form

$$\tilde{I} = O\left(\frac{\|\mathbf{z}\|^2|\sin\varphi|^2}{\|\mathbf{w}\|}\right)$$

as proposed.

Note that the lower bound for $\frac{1}{2}\mathbf{z}^\tau\overline{B}(\mathbf{w})\mathbf{z}$ and the upper bound for the error term are of the same order, namely $\|\mathbf{z}\|^2|\sin\varphi|^2/\|\mathbf{w}\|$. Finally a slightly more careful analysis shows that the constant in the lower bound for $\frac{1}{2}\mathbf{z}^\tau\overline{B}(\mathbf{w})\mathbf{z}$ is bigger than the constant of the upper bound of the error term. Thus, we finally find

$$\operatorname{Re}\left(\log\overline{f}(\mathbf{w},\mathbf{z})\right) \le -C\left(\frac{\|\mathbf{z}\|^2}{\|\mathbf{w}\|^3} + \frac{\|\mathbf{z}\|^2|\sin\varphi|^2}{\|\mathbf{w}\|}\right)$$

for a proper constant $C > 0$. □

We recall that $\mathbf{X}$ is uniformly distributed on the unit ball $\mathcal{B}_d$. The idea is to parametrise the unit ball with the help of spherical coordinates

$$\begin{aligned}
x_1 &= t\cos(\phi_1)\\
x_2 &= t\sin(\phi_1)\cos(\phi_2)\\
x_3 &= t\sin(\phi_1)\sin(\phi_2)\cos(\phi_3)\\
&\vdots\\
x_{d-1} &= t\sin(\phi_1)\cdots\sin(\phi_{d-2})\cos(\phi_{d-1})\\
x_d &= r\sin(\phi_1)\cdots\sin(\phi_{d-2})\sin(\phi_{d-1}),
\end{aligned}$$

where $0 \le t \le 1$, $0 \le \phi_j \le \pi$ $(1 \le j \le d-2)$, $\le \phi_{d-1} \le 2\pi$, and the determinant of the Jacobian is given by

$$t^{d-1}\cdot\prod_{k=2}^{d-1}\left(\sin(\phi_{d-k})\right)^{k-1}.$$

Note that for $t = 1$ we also get a parametrisation of the sphere $\mathcal{S}_d$.

We start with a simple lemma.

Lemma 12. *Suppose that $\mathbf{x} \in \mathcal{S}_d$, that is $\|\mathbf{x}\| = 1$. If $|\langle\mathbf{x},\mathbf{w}\rangle| \le 1$ then*

$$\int_0^1 t^{d-1} f(\mathbf{w}, t\mathbf{x}, \mathbf{z})\,dt = O\left(\frac{1}{|\langle\mathbf{x},\mathbf{z}\rangle|}\right) \tag{A-13}$$

whereas if $|\langle\mathbf{x},\mathbf{w}\rangle| \ge 1$ we find

$$\int_0^1 t^{d-1} f(\mathbf{w}, t\mathbf{x}, \mathbf{z})\,dt = O\left(\frac{1}{\sqrt{|\langle\mathbf{x},\mathbf{z}\rangle\langle\mathbf{x},\mathbf{w}\rangle|}} + \frac{e^{|\langle\mathbf{x},\mathbf{w}\rangle|}}{|\langle\mathbf{x},\mathbf{z}\rangle\langle\mathbf{x},\mathbf{w}\rangle|}\right). \tag{A-14}$$

Proof. The proof follows that same lines as the proof of Lemma 4. We just use the (auxiliary) function

$$h(t) = p(\langle t\mathbf{x},\mathbf{w}\rangle)\langle t\mathbf{x},\mathbf{z}\rangle$$

that satisfies

$$h'(t) = \langle\mathbf{x},\mathbf{z}\rangle\frac{1 + e^{-\langle t\mathbf{x},\mathbf{w}\rangle} + \langle t\mathbf{x},\mathbf{w}\rangle e^{-\langle t\mathbf{x},\mathbf{w}\rangle}}{(1+e^{-\langle t\mathbf{x},\mathbf{w}\rangle})^2}$$

and

$$h''(t) = \langle\mathbf{x},\mathbf{z}\rangle\langle\mathbf{x},\mathbf{w}\rangle e^{-\langle t\mathbf{x},\mathbf{w}\rangle}\frac{2(1+e^{-\langle t\mathbf{x},\mathbf{w}\rangle}) + \langle t\mathbf{x},\mathbf{w}\rangle(e^{-\langle t\mathbf{x},\mathbf{w}\rangle}-1)}{(1+e^{-\langle t\mathbf{x},\mathbf{w}\rangle})^3}$$

as needed. □

As a corollary we obtain the following upper bounds for $\overline{f}(\mathbf{z},\mathbf{w})$.

Lemma 13. *If $\|\mathbf{w}\| \le 1$ then we have*

$$|\overline{f}(\mathbf{z},\mathbf{w})| \le \min\left(1, C_1 \frac{\log(\|\mathbf{z}\|)}{\|\mathbf{z}\|}\right) \tag{A-15}$$

for some constant $C_1 > 0$, whereas if $\|\mathbf{w}\| \ge 1$ we obtain

$$|\overline{f}(\mathbf{z},\mathbf{w})| \le \min\left(1, C_2 \frac{\log(\|\mathbf{z}\|\,\|\mathbf{w}\|) + e^{\|\mathbf{w}\|}}{\sqrt{\|\mathbf{z}\|\,\|\mathbf{w}\|}}\right). \tag{A-16}$$

for some constant $C_2 > 0$.

Proof. We start with the case $\|\mathbf{w}\| \le 1$. Note that $\|\mathbf{w}\| \le 1$ implies $|\langle \mathbf{x}, \mathbf{w}\rangle| \le 1$ for all $\mathbf{x} \in \mathcal{S}_d$. By Lemma 12

$$\overline{f}(\mathbf{z},\mathbf{w}) = O\left(\int_{\mathcal{S}_d} \min\left(1, \frac{1}{|\langle \mathbf{x}, \mathbf{z}\rangle|}\right) d\mathbf{x}\right),$$

where the integral is considered as an $(d-1)$-dimensional integral. Due to rotation symmetry we can assume that $\mathbf{z}$ is parallel to the first axis. Thus, we have $\langle \mathbf{x}, \mathbf{z}\rangle = x_1\|\mathbf{z}\|$ and consequently

$$\begin{aligned}
\overline{f}(\mathbf{z},\mathbf{w}) &= O\left(\int_{-1}^{1} (1-x_1^2)^{\frac{d-3}{2}} \min\left(1, \frac{1}{|x_1|\,\|\mathbf{z}\|}\right) dx_1\right) \\
&= O\left(\frac{1}{\|\mathbf{z}\|} + \frac{1}{\|\mathbf{z}\|}\int_{1/\|\mathbf{z}\|}^{1} (1-x_1^2)^{\frac{d-3}{2}}\, dx_1\right) \\
&= O\left(\frac{\log(\|\mathbf{z}\|)}{\|\mathbf{z}\|}\right)
\end{aligned}$$

as proposed.

Now suppose that $\|\mathbf{w}\| \ge 1$. Then either (A-13) or (A-14) holds. But since $\|\mathbf{w}\| \ge 1$ then (A-13) implies (A-14). Thus we have (A-14) in all cases. To complete have to consider the two ($(d-1)$-dimensional) integrals

$$K_1 = \int_{\mathcal{S}_d} \min\left(1, \frac{1}{\sqrt{|\langle \mathbf{x}, \mathbf{z}\rangle\langle \mathbf{x}, \mathbf{w}\rangle|}}\right) d\mathbf{x}$$

and

$$K_2 = \int_{\mathcal{S}_d} \min\left(1, \frac{e^{\|\mathbf{w}\|}}{|\langle \mathbf{x}, \mathbf{z}\rangle\langle \mathbf{x}, \mathbf{w}\rangle|}\right) d\mathbf{x}.$$

Note that $|\langle \mathbf{x}, \mathbf{w}\rangle| \le \|\mathbf{w}\|$ if $\mathbf{x} \in \mathcal{S}_d$.

We start with K_1 and suppose first that $\mathbf{z}$ and $\mathbf{w}$ are parallel. Then we are in the same situation as in the previous case and, thus, we obtain

$$K_1 = O\left(\frac{\log(\|\mathbf{z}\|\,\|\mathbf{w}\|)}{\sqrt{\|\mathbf{z}\|\,\|\mathbf{w}\|}}\right).$$

In general we distinguish between the cases

$$\frac{|\langle \mathbf{x}, \mathbf{z}\rangle|}{\|\mathbf{z}\|} \le \frac{|\langle \mathbf{x}, \mathbf{w}\rangle|}{\|\mathbf{w}\|} \quad \text{and} \quad \frac{|\langle \mathbf{x}, \mathbf{z}\rangle|}{\|\mathbf{z}\|} > \frac{|\langle \mathbf{x}, \mathbf{w}\rangle|}{\|\mathbf{w}\|}$$

and obtain

$$\frac{1}{\sqrt{|\langle \mathbf{x}, \mathbf{z}\rangle\langle \mathbf{x}, \mathbf{w}\rangle|}} \le \frac{1}{|\langle \mathbf{x}, \mathbf{z}\rangle|}\sqrt{\frac{\|\mathbf{z}\|}{\|\mathbf{w}\|}} + \frac{1}{|\langle \mathbf{x}, \mathbf{w}\rangle|}\sqrt{\frac{\|\mathbf{w}\|}{\|\mathbf{z}\|}}.$$

Thus, we get (with proper constants $C', C'' > 0$)

$$\begin{aligned} K_1 &\le C' \int_{S^{d-1}} \min\left(1, \frac{1}{|\langle \mathbf{x}, \mathbf{z}\rangle|}\sqrt{\frac{\|\mathbf{z}\|}{\|\mathbf{w}\|}}\right) \, d\mathbf{x} + C' \int_{\mathcal{S}_d} \min\left(1, \frac{1}{|\langle \mathbf{x}, \mathbf{w}\rangle|}\sqrt{\frac{\|\mathbf{w}\|}{\|\mathbf{z}\|}}\right) \, d\mathbf{x} \\ &\le C'' \int_{-1}^{1} (1-x_1^2)^{\frac{d-3}{2}} \min\left(1, \frac{1}{|x_1|\sqrt{\|\mathbf{w}\| \, \|\mathbf{z}\|}}\right) \, dx_1 \\ &= O\left(\frac{\log(\|\mathbf{z}\| \, \|\mathbf{w}\|)}{\sqrt{\|\mathbf{z}\| \, \|\mathbf{w}\|}}\right) \end{aligned}$$

as proposed.

Finally we consider the integral K_2, where we use the inequality

$$\frac{1}{|\langle \mathbf{x}, \mathbf{z}\rangle\langle \mathbf{x}, \mathbf{w}\rangle|} \le \frac{1}{|\langle \mathbf{x}, \mathbf{z}\rangle|^2}\frac{\|\mathbf{z}\|}{\|\mathbf{w}\|} + \frac{1}{|\langle \mathbf{x}, \mathbf{w}\rangle|^2}\frac{\|\mathbf{w}\|}{\|\mathbf{z}\|}$$

and use the property

$$\int_{-1}^{1} (1-x_1^2)^{\frac{d-3}{2}} \min\left(1, \frac{1}{|x_1|^2\|\mathbf{w}\| \, \|\mathbf{z}\|}\right) \, dx_1 \frac{1}{\sqrt{\|\mathbf{z}\| \, \|\mathbf{w}\|}}.$$

This completes the proof of the lemma. □

Lemma 14. *There exist constants* $c_1 > 0$ *and* $c_2 > 0$ *such that*

$$|\overline{f}(\mathbf{z}, \mathbf{w})| \le 1 - \frac{c_2}{\|\mathbf{w}\|} \tag{A-17}$$

uniformly for $\|\mathbf{z}\| \ge c_1 \|\mathbf{w}\| \log^2 \|\mathbf{w}\|$.

Proof. We consider first the integral $\int_0^1 t^{d-1} f(\mathbf{w}, t\mathbf{x}, \mathbf{z}) dt$ and assume that $|\langle \mathbf{x}, \mathbf{w}\rangle| > |u_0| + \eta$. we split up the integral into three intervals of the form (compare also with the proofs of Lemma 4 and 5):

$$[0, 1/|\langle \mathbf{x}, \mathbf{w}\rangle|], \quad [1/|\langle \mathbf{x}, \mathbf{w}\rangle|, (|u_0| + \eta)/|\langle \mathbf{x}, \mathbf{w}\rangle|], \quad [(|u_0| + \eta)/|\langle \mathbf{x}, \mathbf{w}\rangle|, 1]$$

and obtain (for some constant $C' > 0$)

$$\left| \int_0^1 t^{d-1} f(\mathbf{w}, t\mathbf{x}, \mathbf{z}) dt \right| \le C' \max\left(1, \frac{1}{\sqrt{|\langle \mathbf{x}, \mathbf{w}\rangle \, \langle \mathbf{x}, \mathbf{z}\rangle|}}\right) + \frac{1}{d}\left(1 - \left(\frac{|u_0| + \eta}{|\langle \mathbf{x}, \mathbf{w}\rangle|}\right)^d\right).$$

Note that we used a the trivial bound $|f(\mathbf{w}, t\mathbf{x}, \mathbf{z})| \le 1$ in the third interval.

We already observed that

$$\int_{\mathcal{S}_d} \max\left(1, \frac{1}{\sqrt{|\langle \mathbf{x}, \mathbf{w}\rangle \, \langle \mathbf{x}, \mathbf{z}\rangle|}}\right) \mathbf{x} = O\left(\frac{\log(\|\mathbf{z}\| \, \|\mathbf{w}\|)}{\sqrt{\|\mathbf{z}\| \, \|\mathbf{w}\|}}\right).$$

Furthermore we have (for proper constants $C', C'' > 0$)

$$\begin{aligned} \int_{\mathcal{S}_d, |\langle \mathbf{x}, \mathbf{w}\rangle| > |u_0| + \eta} |\langle \mathbf{x}, \mathbf{w}\rangle|^{-d} \, d\mathbf{x} &\ge C' \int_{(|u_0| + \eta)/\|\mathbf{w}\|}^{1} (x_1 \|\mathbf{w}\|)^{-d} (1-x_1^2)^{\frac{d-3}{2}} \, dx_1 \\ &\ge C'' \frac{1}{\|\mathbf{w}\|}. \end{aligned}$$

This directly leads to

$$|\overline{f}(\mathbf{z}, \mathbf{w})| \le 1 - \frac{d_1}{\|\mathbf{w}\|} + \frac{d_2 \log(\|\mathbf{z}\| \, \|\mathbf{w}\|)}{\sqrt{\|\mathbf{z}\| \, \|\mathbf{w}\|}}$$

for proper constants $d_1, d_2 > 0$. Clearly if $\|\mathbf{z}\| \ge c_1 \|\mathbf{w}\| \log^2 \|\mathbf{w}\|$ for a properly chosen constant $c_1 > 0$ we obtain (A-17) for some constant $c_2 > 0$. □

Lemma 15. *Suppose that* $c_3 > 0$ *is a given constant. Then there exists* $c_4 > 0$ *such that*

$$|\overline{f}(\mathbf{z},\mathbf{w})| \le 1 - \frac{c_4}{\|\mathbf{z}\|} \tag{A-18}$$

uniformly for $\|\mathbf{z}\| \ge c_3\|\mathbf{w}\|$*. In particular it follows that*

$$|\overline{f}(\mathbf{z},\mathbf{w})| \le 1 - \frac{c_5}{(\|\mathbf{w}\| \log^2 \|\mathbf{w}\|)}$$

uniformly for $c_3\|\mathbf{w}\| \le \|\mathbf{z}\| \le c_1\|\mathbf{w}\|\log^2\|\mathbf{w}\|$.

Proof. The idea is to show that for some constant $c > 0$ we have

$$\int_U p(\langle \mathbf{x},\mathbf{w}\rangle) e^{-ip(-\langle \mathbf{x},\mathbf{w}\rangle)\langle \mathbf{x},\mathbf{z}\rangle}\, d\mathbf{x} \le (1-c)\int_U p(\langle \mathbf{x},\mathbf{w}\rangle)\, d\mathbf{x} \tag{A-19}$$

uniformly for all $\mathbf{z}$ with $\|\mathbf{z}\| \ge c_3\|\mathbf{w}\|$, and where

$$U = \{\mathbf{x} = (x_1,\ldots,x_d) \in \mathcal{S}_d : |x_1| \le 1/\|\mathbf{z}\|,\ |x_2| \le 1/2\}$$

is a subset of $\mathcal{S}_d$ of measure $\ge c'/\|\mathbf{z}\|$ (for a proper constant $c' > 0$). Clearly (A-19) implies (A-18).

We set $\mathbf{w}_0 = \frac{1}{\|\mathbf{z}\|}\mathbf{w}$ and represent $\mathbf{z}$ and $\mathbf{x}$ in the form $\mathbf{z} = z_1\mathbf{w}_0 + \mathbf{z}_2$ and $\mathbf{x} = x_1\mathbf{w}_0 + x_2\mathbf{z}_{2,0} + \mathbf{x}_3$, where $\mathbf{z}_2$ is orthogonal to $\mathbf{w}_0$, $\mathbf{w}_{2,0} = \frac{1}{\|\mathbf{w}_2\|}\mathbf{w}_2$, and $\mathbf{x}_3$ is orthogonal to $\mathbf{w}$ and $\mathbf{z}_{2,0}$. With the help of these representations we have

$$\langle \mathbf{x},\mathbf{w}\rangle = x_1\,\|\mathbf{w}\| \quad \text{and} \quad \langle \mathbf{x},\mathbf{z}\rangle = x_1 z_1 + x_2\|\mathbf{z}_2\|.$$

Furthermore we denote by $U(x_1)$ (for $|x_1| \le\le 1/\|\mathbf{z}\|$) and $U(x_1,x_2)$ (for $|x_1| \le\le 1/\|\mathbf{z}\|$ and $x_2 \le 1/2$) the sets

$$U(x_1) = \{(x_2,\mathbf{x}_3) : |x_2| \le 1/2,\ \|(x_1,x_2,\mathbf{x}_3)\| \le 1\} \in \mathbb{R}^{d-1}$$

and

$$U(x_1,x_2) = \{\mathbf{x}_3 : \|(x_1,x_2,\mathbf{x}_3)\| \le 1\} \in \mathbb{R}^{d-2}.$$

It is easy to show that

$$\mathrm{Vol}_{d-1}(U(x_1)) \to C_1 \quad \text{and} \quad \mathrm{Vol}_{d-2}(U(x_1,x_2)) \to C_2$$

as $x_1 \to 0$ for proper constants $C_1 > 0, C_2 > 0$.

We now have

$$\begin{aligned}\int_U p(\langle \mathbf{x},\mathbf{w}\rangle)\, d\mathbf{x} &= \int\limits_{|x_1|\le 1/\|\mathbf{z}\|} p(x_1\|\mathbf{w}\|)\mathrm{Vol}_{d-1}(U(x_1))\, dx_1 \\ &= \int\limits_{|x_1|\le 1/\|\mathbf{z}\|} \int_{-1/2}^{1/2} p(x_1\,\|\mathbf{w}\|)\mathrm{Vol}_{d-2}(U(x_1,x_2))\, dx_2\, dx_1\end{aligned}$$

and

$$\begin{aligned}R &:= \int_U p(\langle \mathbf{x},\mathbf{w}\rangle) e^{-ip(-\langle \mathbf{x},\mathbf{w}\rangle)\langle \mathbf{x},\mathbf{z}\rangle}\, d\mathbf{x} \\ &= \int\limits_{|x_1|\le 1/\|\mathbf{z}\|} \int_{-1/2}^{1/2} p(x_1\,\|\mathbf{w}\|) e^{-ip(-x_1\|\mathbf{w}\|)(x_1z_1 + x_2\|\mathbf{z}_2\|)}\mathrm{Vol}_{d-2}(U(x_1,x_2))\, dx_2\, dx_1.\end{aligned}$$

First suppose that $\|\mathbf{z}_2\| \ge 1$. By assumption we have $|x_1|\|\mathbf{w}\| \le \|\mathbf{w}\|/\|\mathbf{z}\| \le 1/c_3$ which implies that $A = p(-x_1\|\mathbf{w}\|)\|\mathbf{z}_2\|$ satisfies

$$|A| = p(-x_1\|\mathbf{w}\|)\|\mathbf{z}_2\| \ge c''$$

uniformly for some constant $c'' > 0$. Thus, the following integral satisfies

$$\begin{aligned}\left|\int_{-1/2}^{1/2} e^{-ip(-x_1\|\mathbf{w}\|)\|\mathbf{z}_2\|x_2}\,dx_2\right| &= \left|\int_{-1/2}^{1/2} e^{-Ax_2}\,dx_2\right| \\ &= \left|\frac{\sin(A/2)}{A/2}\right| \le 1-c\end{aligned}$$

for some constant $c > 0$ (provided that $|A| \ge c''$). Consequently (A-19) follows in this case.

If $\|\mathbf{z}_2\| \ge 1$ we can use a continuity and a compactness argument. We rewrite the integral R as

$$R = \frac{1}{\|\mathbf{z}\|}\int\limits_{|u|\le 1}\int_{-1/2}^{1/2} p(u\,\|\mathbf{w}\|/\|\mathbf{z}\|)e^{-ip(-u\|\mathbf{w}\|/\|\mathbf{z}\|)(uz_1/\|\mathbf{z}\|+x_2\|\mathbf{z}_2\|)}\mathrm{Vol}_{d-2}(U(x_1,x_2))\,dx_2\,dx_1$$

and recall that $0 \le \|\mathbf{w}\|/\|\mathbf{z}\| \le 1/c_2$, $-1 \le z_1/\|\mathbf{z}\| \le 1$, and $\|\mathbf{z}_2\| \ge 1$ vary in a compact set. Thus, we certainly have uniformly in that range

$$J \le \frac{1-c}{\|\mathbf{z}\|}\int\limits_{|u|\le 1}\int_{-1/2}^{1/2} p(u\,\|\mathbf{w}\|/\|\mathbf{z}\|)\mathrm{Vol}_{d-2}(U(x_1,x_2))\,dx_2\,dx_1$$

for some constant $c > 0$, and we are done. □

Appendix E
Proof of Theorem 1(ii) for $d = 1$

The proof of Theorem 1(ii) runs along similar lines as that of Theorem 1(i) but it is more technical. We first note that the second moment $\mathbb{E}[S_\varepsilon(\mathbf{x}^T)^2]$ cannot be explicitly represented as a convergent multi-dimensional integral as it is the case for the first moment. We use a regularized version of $S_\varepsilon(\mathbf{x}^T)$ of the form

$$S_\varepsilon^\eta(x^T) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}\prod_{t=1}^{T} f(w,x_t,z)\,e^{-\varepsilon w^2-2i\varepsilon wz}\nabla^2 L_\varepsilon(\cdot|x^T,w)e^{-\eta z^2}\,dw\,dz \tag{A-20}$$

that is absolutely convergent as a double integral if $\eta > 0$ and has the property that

$$S_\varepsilon(x^T) = \lim_{\eta\to 0} S_\varepsilon^\eta(x^T).$$

Actually we can be more explicit. By slightly extending the computations of Appendix A, that is, by using the more precise relation $a = 2\varepsilon w + O(T)$ that holds uniformly in y^T, x^T and T, where $w = G^{-1}_{y^T|x^T}(a)$ and by using $(a+O(T))^2 \ge \frac{1}{2}a^2 + O(T^2)$ and $2^T = O(e^{O(T^2/\varepsilon)})$ it follows that

$$\sum_{y^T} h_{y^T|x^T}(a) = O\left(e^{-\frac{1}{8\varepsilon}(a^2+O(T^2))}\right)$$

and

$$\sum_{y^T} h_{y^T|x^T}(a)'' = O\left(\frac{T^2+a^2}{\varepsilon^3}e^{-\frac{1}{8\varepsilon}(a^2+O(T^2))}\right).$$

Consequently we have

$$\begin{aligned}\sum_{y^T}\tilde{h}_{y^T|x^t}(z) &= \frac{1}{2\pi}\int_{\mathbb{R}}\sum_{y^T} h_{y^T|x^T}(a)e^{-iaz} \\ &= \frac{-1}{2\pi z^2}\int_{\mathbb{R}}\sum_{y^T} h_{y^T|x^T}(a)''e^{-iaz} \\ &= O\left(\varepsilon^{-\frac{1}{2}}\min\left(1,\frac{1}{\varepsilon z^2}\right)e^{O(T^2/\varepsilon)}\right).\end{aligned}$$

In other words,

$$\left|\int_{\mathbb{R}}\prod_{t=1}^{T} f(w,x_t,z)\,e^{-\varepsilon w^2-2i\varepsilon wz}\nabla^2 L_\varepsilon(\cdot|x^T,w)\,dw\right| = O\left(\varepsilon^{-\frac{1}{2}}\min\left(1,\frac{1}{\varepsilon z^2}\right)e^{O(T^2/\varepsilon)}\right).$$

This upper bound immediately implies

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

and consequently (since $S_\varepsilon(x^T)=O(2^T)=e^{O(T^2/\varepsilon)}$)

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

In particular, if we choose

$$\eta = e^{-CT^2/\varepsilon} > 0 \tag{A-21}$$

for a sufficiently large constant C we find

$$\left|\mathbb{E}[S_\varepsilon(x^T)^2]-\mathbb{E}[S_\varepsilon^\eta(x^T)^2]\right| = O\left(\varepsilon^{-2}e^{-C'T^2/\varepsilon}\right) = O\left(Te^{-C'T^{3/2}}\right)$$

if $\varepsilon \gg T^{-1/2}$.

Thus, it suffices to compute $\mathbb{E}[S_\varepsilon^\eta(x^T)^2]$. Since $\eta>0$, we know that $S_\varepsilon^\eta(x^T)$ is represented as an absolute convergent double integral (A-20), by dominated convergence we obtain

$$\begin{aligned}
\mathbb{E}[S_\varepsilon^\eta(x^T)^2] &= \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4}\mathbb{E}\left[\prod_{t=1}^{T}\left(f(w_1,x_t,z_1)\overline{f(w_2,x_t,z_2)}\right)\nabla^2 L_\varepsilon(\cdot|x^T,w_1)\nabla^2 L_\varepsilon(\cdot|x^T,w_2)\right]\\
&\qquad\times e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1z_1-w_2z_2)}e^{-\eta(z_1^2+z_2^2)}\,dw_1\,dw_2\,dz_1\,dz_2\\
&= \frac{T^2-T}{(2\pi)^2}\int_{\mathbb{R}^4}\overline{\overline{f}}(w_1,w_2,z_1,z_2)^{T-1}B(z_1,w_1)B(z_2,w_2)e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1z_1-w_2z_2)}e^{-\eta(z_1^2+z_2^2)}\,dw_1\,dw_2\,dz_1\,dz_2\\
&+\frac{T}{(2\pi)^2}\int_{\mathbb{R}^4}\overline{\overline{f}}(w_1,w_2,z_1,z_2)^{T-1}B_2(z_1,z_2,w_1,w_2)e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1z_1-w_2z_2)}e^{-\eta(z_1^2+z_2^2)}\,dw_1\,dw_2\,dz_1\,dz_2\\
&+2\frac{2\varepsilon T}{(2\pi)^2}\int_{\mathbb{R}^4}\overline{\overline{f}}(w_1,w_2,z_1,z_2)^{T-1}\overline{f}(w_1,z_1)B(z_2,w_2)e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1z_1-w_2z_2)}e^{-\eta(z_1^2+z_2^2)}\,dw_1\,dw_2\,dz_1\,dz_2\\
&+\frac{(2\varepsilon)^2}{(2\pi)^2}\int_{\mathbb{R}^4}\overline{\overline{f}}(w_1,w_2,z_1,z_2)^{T}e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1z_1-w_2z_2)}e^{-\eta(z_1^2+z_2^2)}\,dw_1\,dw_2\,dz_1\,dz_2\\
&= (T^2-T)\cdot\overline{J}_0^\eta + T\cdot\overline{J}_{0,2}^\eta + 2\varepsilon T\cdot\overline{J}_1^\eta + (2\varepsilon)^2\overline{J}_2^\eta,
\end{aligned}$$

where

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

and

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

For the sake of simplicity we only consider (as in Section III) the integral

$$\overline{J}_2^\eta := \int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\overline{\overline{f}}(w_1,w_2,z_1,z_2)^T e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1z_1-w_2z_2)}e^{-\eta(z_1^2+z_2^2)}dw_1\,dw_2\,dz_1\,dz_2.$$

We will show that

$$\overline{J}_2^\eta = J_1^2(1+O(T^{-\beta}))+O((1-\kappa)^T\log\eta^{-1}) \tag{A-22}$$

for some $\beta > 0$ and for some $\kappa > 0$. By applying the same kind of calculations to the other parts of the integral we find in particular for the asmyptotically leading terms

$$\overline{J}_0^\eta = J_0^2(1 + O(T^{-\beta})) + O((1-\kappa)^T \log \eta^{-1})$$

(and similar properties for $\overline{J}_{0,2}^\eta$ and $\overline{J}_1^\eta$).

By setting $\eta = \exp\bigl(-CT^2/\varepsilon\bigr)$, see (A-21), we thus obtain

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

and consequently

$$\mathbb{E}[S_\varepsilon(x^T)^2] = \mathbb{E}[S_\varepsilon(x^T)]^2(1 + O(T^{-\beta})). \tag{A-23}$$

which proves (22) of Theorem 1.

For the proof of (A-22) we need some information on $\overline{\overline{f}}(w_1, w_2, z_1, z_2)$. The first property is a direct extension of Lemma 3.

Lemma 16. *We have uniformly for $|z_1| \le \max(1, c_1|w_1|)$ and $|z_2| \le \max(1, c_1|w_2|)$*

$$\overline{\overline{f}}(w_1, w_2, z_1, z_2) = e^{-\frac{1}{2}z_1^2\overline{B}(w_1) - \frac{1}{2}z_2^2\overline{B}(w_2)}\Bigl(1 + O\left(z_1^3 \min(1, |w_1|^{-4})\right) + O\left(z_1^4 \min(1, |w_1|^{-6})\right) + O\left(z_2^3 \min(1, |w_2|^{-4})\right) + O\left(z_2^4 \min(1, |w_2|^{-6})\right)\Bigr).$$

The next two lemmas require some more work.

Lemma 17. *Suppose that $|w_1| \le C_1$ and $|w_2| \le C_1$ (for a proper constant $C_1 > 0$) and $z_1 \ne z_2$. Then we have (for some constant $C > 0$).*

$$\left|\overline{\overline{f}}(w_1, w_2, z_1, z_2)\right| \le \min\left(1, \frac{C}{\sqrt{|z_1 - z_2| + |z_2||w_1 - w_2|}}\right).$$

If $|w_1| \le C_1$ and $|w_2| \le C_1$ then

$$\left|\overline{\overline{f}}(w_1, w_2, z_1, z_2)\right| \le \min\left(1, \frac{C}{\sqrt{|z_1| + |z_2 w_2| e^{-|w_2|}}}\right).$$

Finally, if $|w_1| \le C_1$ and $|w_2| \le C_1$, then

$$\left|\overline{\overline{f}}(w_1, w_2, z_1, z_2)\right| \le \min\left(1, \frac{C}{\sqrt{|z_1 w_1| e^{-|w_1|} + |z_2 w_2| e^{-|w_2|}}}\right).$$

Lemma 18. *Suppose that $c_1 > 0$ is a given constant. Then there exists $c_2 > 0$ such that*

$$\left|\overline{\overline{f}}(w_1, w_2, z_1, z_2)\right| \le 1 - \frac{c_2}{\max\{|w_1|, |w_2|\}}$$

uniformly for (w_1, w_2, z_1, z_2) with $|w_1| \ge C_1$, $|z_1| \ge c_1|w_1|$ or $|w_2| \ge C_2$, $|z_2| \ge c_1|w_3|$.

Proof. We give only a detailed proof of the first part of Lemma 17. The proof of the second and third part of Lemma 17 are then quite similar. Finally, the proof of Lemma 18 is very close to that of Lemma 5.

By definition we see that $\overline{\overline{f}}(w_1, w_2, z_1, z_2)$ consists of four terms of the form (or of a very similar form)

$$P := \frac{1}{2}\int_{-1}^{1} p(-xw_1)p(-xw_2)e^{ih(x)},$$

where

$$h(x) = \frac{xz_1}{1 + e^{-xw_1}} - \frac{xz_2}{1 + e^{-xw_2}} = xz_1 p(xw_1) - xz_2 p(xw_2).$$

As in the proof of Lemma 3, we make use of Lemmas 9 and 10. In particular we need the first and second derivative of $h(x)$ that we represent in the following way:

$$\begin{aligned} h'(x) &= (p'(xw_1)xw_1 + p(xw_1))(z_1 - z_2) \\ &\quad + (p'(xw_1)xw_1 + p(xw_1) - p'(xw_2)xw_2 - p(xw_2))z_2, \\ h''(x) &= (p''(xw_1)xw_1^2 + 2p'(xw_1)w_1)(z_1 - z_2) \\ &\quad + (p''(xw_1)xw_1^2 + 2p'(xw_1)w_1 - p''(xw_2)xw_2^2 - 2p'(xw_2)w_2)z_2. \end{aligned}$$

It is immediate that $h'(x) = O(|z_1 - z_2| + |w_1 - w_2||z_2|)$ and $h''(x) = O(|z_1 - z_2| + |w_1 - w_2||z_2|)$. Thus, it remains that these upper bounds are – more or less – tight.

We suppose (for the sake of simplicity) that $x \ge 0$, $w_1 > 0$, and $w_2 > 0$. If

$$H_1(x) = (p'(xw_1)xw_1 + p(xw_1))(z_1 - z_2)$$

and

$$H_2(x) = (p'(xw_1)xw_1 + p(xw_1) - p'(xw_2)xw_2 - p(xw_2))z_2$$

then then we trivially get the lower bound $|h'(x)| \ge c|z_1 - z_2|$ (for some constant $c > 0$) which proves (by Lemma 9) that

$$P = O\left(\frac{1}{|z_1 - z_2|}\right) = O\left(\frac{1}{|z_1 - z_2| + |w_1 - w_2||z_2|}\right) = O\left(\frac{1}{\sqrt{|z_1 - z_2| + |w_1 - w_2||z_2|}}\right).$$

as proposed. Here we used the fact that $H_1(x) = \Theta(|z_1 - z_2|)$ and $H_2(x) = \Theta(x|w_1 - w_2||z_2|)$. Actually if $H_1(x)$ and $H_2(x)$ have different signs and if we have $|H_1(x)| \ge 2|H_2(x)|$ for all $x \in [0, 1]$ then it also follows that $|h'(x)| \ge \frac{1}{2}|H_1(x)| = \Theta(|z_1 - z_2|)$. Thus, we get the same upper bound. Finally suppose that $H_1(x)$ and $H_2(x)$ have different signs and that $|H_1(x)| \ge 2|H_2(x)|$ holds only for $0 \le x \le x_1$ for some $x_1 \in (0, 1)$ which means that $|z_1 - z_2| \le c'|w_1 - w_2||z_2|$ for some constant $c' > 0$. We now consider the second derivative $h''(x) = H_1'(x) + H_2'(x)$. Here we have $H_1'(x) = \Theta(|z_1 - z_2||w_1|)$ and $H_2'(x) = \Theta(|w_1 - w_2||z_2|)$. Consequently, if $|w_1|$ is sufficiently small then $|H_2'(x)| \ge 2|H_1(x)|$ for all $x \in [0, 1]$ and we find (by Lemma 10)

$$P = O\left(\frac{1}{\sqrt{|w_1 - w_2||z_2|}}\right) = O\left(\frac{1}{\sqrt{|z_1 - z_2| + |w_1 - w_2||z_2|}}\right).$$

The other cases can be handled in a very similar way (just with slightly more care) and completes the proof of the first part of Lemma 17. □

We split now the integration over $\mathbb{R}^2 \times \mathbb{R}^2$ in the integral $\overline{J}_2^\eta$ into several parts. Actually, with the help of the above properties (which are all very similar to those that have been used for evaluating the expected value, see Section III) we can cover all but one region (for the sake of brevity we omit the easy technical calculations in the first five cases):

- $|z_1| \le \max(1, c_1|w_1|)$ and $|z_2| \le \max(1, c_1|w_2|)$. In this case, we apply Lemma 16 and obtain the leading term by Gaussian approximation. We note that the regularization factor $e^{-\eta z^2}$ does not change the leading term. Since $\eta = e^{-CT^2/\varepsilon}$ is extremely small, we can use the approximation $e^{-\eta z^2} = 1 - O(\eta z^2)$ and obtain (uniformly for $\varepsilon \gg T^{-1/2}$):
$$\frac{1}{2\pi T}\left(\int_{-\infty}^{\infty} \overline{B}(w)^{-1/2} e^{-\varepsilon w^2}\, dw\right)^2 \left(1 + O\left(T^{-\frac{1}{2}}\right) + O\left(\eta\, \varepsilon^{-1}\right)\right).$$
Clearly, this leading term matches the square of the leading term of J_1.
- $|w_1| \ge C_1$, $c_1|w_1| \le |z_1| \le e^{|w_1|}$ or $|w_2| \ge C_1$, $c_1|w_2| \le |z_2| \le e^{|w_2|}$.
Here we use Lemma 18 and argue as in Section III This gives an error term of magnitude $O(e^{-c_5\sqrt{T}})$.
- $|w_1| \le C_1$, $|z_1| \ge e^{|w_1|}$ or $|w_2| \ge C_1$, $|z_2| \ge e^{|w_2|}$.
Here we use again Lemma 18 and obtain an error term of magnitude $O(T^{-\frac{3}{2}})$.
- $|w_1| \le C_1$, $|w_2| \le C_1$, $|z_1| \le 1$, $|z_2| \ge 1$ (or the other way round).
Here we use Lemma 17 and obtain (again) an error term of order $O(T^{-\frac{3}{2}})$

- $|w_1| \le C_1$, $|w_2| \le C_1$, $|z_1| \ge 1$, $|z_2| \ge 1$, and $|z_1 - z_2| + |w_1 - w_2||z_1| \ge C_3$ (for C_3 sufficiently large). Here we use the first part of Lemma 17 and get an error term of order $O(T^{-\frac{3}{2}})$.
- $|w_1| \le C_1$, $|w_2| \le C_1$, $|z_1| \ge 1$, $|z_2| \ge 1$, and $|z_1 - z_2| + |w_1 - w_2||z_1| \le C_3$.
  In this case we have to argue separately, as mentioned above. The methods from Section III do not apply.

The essential point in the last case is that the corresponding part in the integral $\overline{J}_2^{\eta}$, when we set $\eta = 0$, is not convergent. So we will need some precise information on $\overline{\overline{f}}(w_1, w_2, z_1, z_2)$ in this range and then we use the regularizing factor $e^{-\eta(z_1^2+z_2^2)}$ to get convergence and proper upper bounds. This will then complete the proof.

Lemma 19. *Suppose that* $|w_1| \le C_1$, $|w_2| \le C_1$, $|z_1| \ge 1$, $|z_2| \ge 1$, *and* $|z_1 - z_2| + |w_1 - w_2||z_1| \le C_3$. *Then*

$$\overline{\overline{f}}(w_1, w_2, z_1, z_2) = \frac{1}{2} + O(|w_1| + |w_2| + |z_1 - z_2| + |w_1 - w_2||z_2|) + O\left(\frac{1}{|z_1|}\right). \tag{A-24}$$

Proof. We split up $\overline{\overline{f}}(w_1, w_2, z_1, z_2)$ into four parts:

$$\begin{aligned}
\overline{\overline{f}}_1 &:= \frac{1}{2}\int_{-1}^{1} p(xw_1)p(xw_2)e^{-ip(-xw_1)xz_1+ip(-xw_2)xz_2}\,dx,\\
\overline{\overline{f}}_2 &:= \frac{1}{2}\int_{-1}^{1} p(xw_1)p(-xw_2)e^{-ip(-xw_1)xz_1-ip(xw_2)xz_2}\,dx,\\
\overline{\overline{f}}_3 &:= \frac{1}{2}\int_{-1}^{1} p(-xw_1)p(xw_2)e^{ip(xw_1)xz_1+ip(-xw_2)xz_2}\,dx,\\
\overline{\overline{f}}_4 &:= \frac{1}{2}\int_{-1}^{1} p(-xw_1)p(-xw_2)e^{ip(xw_1)xz_1-ip(xw_2)xz_2}\,dx.
\end{aligned}$$

We first study $\overline{\overline{f}}_1$, where we replace $p(u)$ by $p(u) = \frac{1}{2} + r(u)$. Note that $r(u) = O(u)$ for $u = O(1)$. In particular we have

$$\begin{aligned}
H &:= -ip(-xw_1)xz_1 + ip(-xw_2)xz_2\\
&= -\frac{i}{2}x(z_1 - z_2) - ir(-xw_1)xz_1 + ir(-xw_2)xz_2\\
&= -\frac{i}{2}x(z_1 - z_2) - ix(r(-xw_1) - r(-xw_2))z_2 - ixr(-xw_1)(z_1 - z_2).
\end{aligned}$$

By assumption $|z_1 - z_2| + |w_1 - w_2||z_2|$ is bounded. Hence, H is bounded, too. We then obtain

$$\begin{aligned}
\overline{\overline{f}}_1 &= \frac{1}{2}\int_{-1}^{1} p(xw_1)p(xw_2)e^{-ip(-xw_1)xz_1+ip(-xw_2)xz_2}\,dx,\\
&= \frac{1}{2}\int_{-1}^{1}\frac{1}{4}\,dx + \frac{1}{2}\int_{-1}^{1} r(xw_1)p(xw_2)e^{iH}\,dx\\
&+ \frac{1}{2}\int_{-1}^{1} r(xw_2)p(xw_1)e^{iH}\,dx + \frac{1}{2}\int_{-1}^{1} p(xw_1)p(xw_2)(e^{iH} - 1)\,dx\\
&= \frac{1}{4} + O(|w_1| + |w_2| + |z_1 - z_2| + |w_1 - w_2||z_2|).
\end{aligned}$$

Similarly we find

$$\overline{\overline{f}}_4 = \frac{1}{4} + O(|w_1| + |w_2| + |z_1 - z_2| + |w_1 - w_2||z_2|).$$

The behavior of $\overline{\overline{f}}_2$ and $\overline{\overline{f}}_3$ is different. Since

$$-ip(-xw_1)xz_1 - ip(xw_2)xz_2 = -i(1 - p(xw_1))xz_1 - ip(xw_2)xz_2 = -ixz_1 + H$$

we have

$$\overline{\overline{f}}_2 = \frac{1}{2}\int_{-1}^{1} p(xw_1)p(-xw_2)e^{-ixz_1+H}\,dx,$$

and by applying Lemma 9 we obtain

$$\overline{\overline{f}}_2 = O\left(\frac{1}{|z_1|}\right).$$

Similarly

$$\overline{\overline{f}}_3 = O\left(\frac{1}{|z_1|}\right).$$

Thus (A-24) follows. □

It is an immediate consequence of Lemma 19 (together with a simple continuity argument) that

$$\left|\overline{\overline{f}}(w_1, w_2, z_1, z_2)\right| \le 1 - \kappa$$

for some $\kappa > 0$ provided that $|w_1| \le C_1$, $|w_2| \le C_1$, $|z_1| \ge 1$, $|z_2| \ge 1$, and $|z_1 - z_2| + |w_1 - w_2||z_1| \le C_3$. For notational convenience, we denote the subset of (w_1, w_2, z_1, z_2) satisfying these conditions by R. Furthermore we denote by R' the set of (w_1, w_2, z_1) with $|w_1| \le C_1$, $|w_2| \le C_1$, $|z_1| \ge 1$, and $|w_1 - w_2||z_1| \le C_3$, and by R'' the set of (w, z_1) with $|w| \le 2C_1$ and $|z_1| \ge 1$, and $|wz_1| \le C_3$. Then we get

$$\begin{aligned}
&\iiiint_R \overline{\overline{f}}(w_1, w_2, z_1, z_2)^T e^{-\varepsilon(w_1^2+w_2^2)-2i\varepsilon(w_1 z_1 - w_2 z_2)} e^{-\eta(z_1^2+z_2^2)}\, dw_1\, dw_2\, dz_1\, dz_2 \\
&= O\left((1-\kappa)^T \iiiint_R e^{-\eta z_1^2}\, dw_1\, dw_2\, dz_1\, dz_2\right) \\
&= O\left((1-\kappa)^T \iiint_{R'} e^{-\eta z_1^2}\, dw_1\, dw_2\, dz_1\right) \\
&= O\left((1-\kappa)^T \iint_{R''} e^{-\eta z_1^2}\, dw\, dz_1\right).
\end{aligned}$$

Finally we have

$$\begin{aligned}
\iint_{R''} e^{-\eta z_1^2}\, dw\, dz_1 &= O\left(\int_1^\infty e^{-\eta z_1^2} \min\left(\frac{1}{z}, 2C_1\right) dz\right) \\
&= O\left(\int_1^{\eta^{-1/2}} \frac{1}{z}\, dz\right) + O\left(\int_{\eta^{-1/2}}^\infty e^{-\eta z_1^2} \frac{1}{z}\, dz\right) \\
&= O\left(\log\frac{1}{\eta}\right) + O(1).
\end{aligned}$$

This implies an upper bound of the form

$$O\left((1-\kappa)^T \log\frac{1}{\eta}\right).$$

In summary, we prove (A-22) which implies (22) and completes the proof of Theorem 1.

REFERENCES

[1] P. D. Grunwald, *The Minimum Description Length Principle*. MIT Press, 2007.

[2] A. Rakhlin and K. Sridharan, "Sequential probability assignment with binary alphabets and large classes of experts," *arXiv preprint arXiv:1501.07340*, 2015.

[3] D. J. Foster, S. Kale, H. L. andMehryar Mohri, and K. Sridharan, "Logistic regression: The importance of being improper," in *COLT - Conference on Learning Theory*, 2018.

[4] C. Wu, M. Heidari, A. Grama, and W. Szpankowski, "Precise regret bounds for log-loss via a truncated bayesian algorithm," in *Advances in Neural Information Processing Systems*, 2022.

[5] ——, "Regret bounds for log-loss via bayesian algorithms," *Trans. Information Theory*, vol. 69, no. 9, pp. 5971–5989, 2023.

[6] P. Jacquet, G. I. Shamir, and W. Szpankowski, "Precise minimax regret for logistic regression with categorical feature values," in *PRML: ALT'21*, vol. 132, 2021, pp. 755–771.

[7] E. Hazan, T. Koren, and K. Y. Levy, "Logistic regression: Tight bounds for stochastic and online optimization," in *The 27th Conference on Learning Theory, COLT 2014*. MIT press, 2014, pp. 197–209.

[8] G. I. Shamir, "Logistic regression regret: What's the catch?" in *Conference on Learning Theory*. PMLR, 2020, pp. 3296–3319.

[9] N. Cesa-Bianchi and O. Shamir, "Efficient transductive online learning via randomized rounding," *arXiv preprint arXiv:1106.2429*, 2011.

[10] C. Wu, M. Heidari, A. Grama, and W. Szpankowski, "Expected worst case regret via stochastic sequential covering," *Transactions on Machine Learning Research*, 2023.
[11] G. I. Shamir and W. Szpankowski, "Low complexity approximate bayesian logistic regression for sparse online learning," in *ISIT24 and ArXiv: http://arxiv.org/abs/2101.12113*, 2021.
[12] R. Jezequel, P. Gillard, and A. Rudi, "Efficient improper learning for online logistic regression," in *PMLR, COLT'20*, vol. 125, 2020.
[13] E. Hazan, "The convex optimization approach to regret minimization," in *S. Sra, S. Nowozin, and S. Wright, editors, Optimization for Machine Learning*. MIT press, 2012, pp. 287–303.
[14] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, Oct. 1998.
[15] J. Takeuchi and A. Barron, "Asymptotically minimax regret by bayes mixtures for non-exponential families," in *IEEE Information Theory Workshop*, 2006.
[16] M. Drmota and W. Szpankowski, *Analytic Information Theory: From Compression to Learning*. Cambridge: Cambridge University Press, 2023.
[17] P. Jacquet and W. Szpankowski, "Markov types and minimax redundancy for markov sources," *IEEE Trans. on Information Theory*, vol. 5o, pp. 1393–1402, 2004.
[18] W. Szpankowski and M. Weinberger, "Minimax pointwise redundancy for memoryless models over large alphabets," *IEEE Trans. Information Theory*, vol. 58, pp. 4094–4104, 2012.
[19] Q. Xie and A. Barron, "Minimax redundancy for the class of memoryless sources," *IEEE Trans. Information Theory*, pp. 647–657, 1997.
[20] ——, "Asymptotic minimax regret for data compression, gambling, and prediction," *IEEE Trans. Information Theory*, vol. 46, pp. 431–445, 2000.
[21] J. Mayo, H. Hadiji, and T. van Erven, "Scale-free unconstrained online learning for curved losses," in *COLT*, vol. 178, 2022, pp. 1–34.
[22] M. Drmota, P. Jacquet, C. Wu, and W. Szpankowski, "Minimax regret with unbounded weights," in *ISIT*, 2024. [Online]. Available: https://www.cs.purdue.edu/homes/spa/papers/isit24-unbounded.pdf
[23] J. Qian, A. Rakhlin, and N. Zhivotovskiy, "Refined risk bounds for unbounded losses via transductive priors," *arXiv preprint arXiv:2410.21621*, 2024.
[24] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge University Press, 2008.
[25] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.
[26] ——, "On asymptotics of certain recurrences arising in universal coding," *Problems of Information Transmission*, vol. 34, pp. 55–61, 1998.
[27] P. Flajolet and W. Szpankowski, "Analytic variations on redundancy rates of renewal processes," *IEEE Trans. Information Theory*, vol. 48, pp. 2911–2921, 2002.
[28] P. Jacquet, "Is quantum tomography a difficult problem for machine learning?" in *Physical Sciences Forum*, vol. 5, no. 1. MDPI, 2023, p. 47.
[29] M. N. Huxley, *Area, lattice points, and exponential sums*, ser. Lond. Math. Soc. Monogr., New Ser. Oxford: Clarenton Press, 1996, vol. 13.