Proceedings of Machine Learning Research vol 196:1–40, 2024 37th Annual Conference on Learning Theory

Precise Regularized Minimax Regret with Unbounded Weights

Michael Drmota MICHAEL.DRMOTA@TUWIEN.AC.AT
TU Wien
Vienna, Austria

Philippe Jacquet PHILIPPE.JACQUT@INRIA.FR
INRIA
Paris, France

Changlong Wu WUCHANGL@HAWAII.EDU
Purdue University
West Lafayette, IN 47907, USA

Wojciech Szpankowski SZPAN@PURDUE.EDU
Purdue University
West Lafayette, IN, USA

Editors: Shipra Agrawal and Aaron Roth

Abstract

In online learning a learner receives data in rounds and at each round predicts a label which is then compared to the true binary label incurring a loss. The total loss over T rounds, when compared to a loss over the best expert from a class of experts/forecasters, is called the regret. In this paper we focus on logarithmic loss for the logistic function with *unbounded* d-dimensional weights, a scenario that was largely unexplored. We introduce a *regularized* version of the average (fixed design) minimax regret by imposing a *soft-constraint* on the weight norm via *precise* analysis of the so-called Shtarkov sum. Our main results provide the first known *precise* characterization of the 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 only for bounded weights, our result implies that the regularized regret grows no faster than $(1/2 + \alpha/4)d \log T$ when the regularization parameter is of order $\Theta(T^{-\alpha})$ for $\alpha \leq 1/2$. We accomplish it using tools from analytic combinatorics, e.g., multidimensional Fourier, saddle point method, and Mellin transform.

1. Introduction

The problem of online learning under logarithmic loss and its regret analysis has been intensively studied over the last decade Grunwald (2007); Rakhlin and Sridharan (2015); Foster et al. (2018); Wu et al. (2022). However, even for logistic regression, there is a lack of precise second-order asymptotics (especially for unbounded weights), with a possible exception of Jacquet et al. (2021) 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 $y_t \in \{-1, 1\}$. The learner makes the prediction $\hat{y}_t = g_t(y^{t-1}, \mathbf{x}^t)$, where

g_t represents the strategy/algorithm of the learner to obtain its prediction based on prior observations y^{t-1} and $\mathbf{x}^t$. Once a prediction is made, nature reveals the true label y_t, and the learner incurs some *loss* evaluated based on a predefined function $\ell : \hat{\mathcal{Y}} \times \mathcal{Y} \to \mathbb{R}^{\geq 0}$, 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 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 $\hat{y}_t = g_t(y^{t-1}, \mathbf{x}^t)$. Throughout, we write $y^t = (y_1, \dots, y_t)$ and $\mathbf{x}^t = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ for $t \in [T]$.

In this paper, we focus specifically on the *logarithmic loss* defined as: $\ell(\hat{y}_t, y_t) = -\frac{1+y_t}{2}\log(\hat{y}_t) - \frac{1-y_t}{2}\log(1-\hat{y}_t)$. Furthermore, we restrict our study to the class of experts:

$$\mathcal{H}_{p,\mathbf{w}} = \{h_{\mathbf{w}}(\mathbf{x}) \stackrel{\text{def}}{=} p(\langle \mathbf{w}, \mathbf{x}\rangle) : \mathbf{w}, \mathbf{x} \in \mathbb{R}^d\}, \tag{1}$$

where $\mathbf{w}$ is a d-dimensional weight vector, $\langle \mathbf{w}, \mathbf{x}\rangle$ is the scalar product of $\mathbf{x}$ and $\mathbf{w}$, and $p(w)$ with $w = \langle \mathbf{w}, \mathbf{x}\rangle$ is a probability function. To ease the presentation, we almost exclusively discuss logistic regression with $p(w) = (1 + \exp(-w))^{-1}$ (Hazan et al., 2014; Shamir, 2020a) (see Appendix F for possible extensions). 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 $\|\mathbf{w}\| \leq R \leq \infty$.

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

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

This notion was also known in the literature as *transductive online learning* Cesa-Bianchi and Shamir (2011). 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.

As discussed in Jacquet et al. (2021); Shamir and Szpankowski (2021) 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

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

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 $r_T(\mathbf{x}^T) = \log S(\mathbf{x}^T)$. 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 $\mathbf{w}^* = \arg\sup_{\|\mathbf{w}\|\leq R} P(y^T|\mathbf{x}^T, \mathbf{w})$ lie on the boundary $\|\mathbf{w}\| = R$. To address these issues, one often resorts to regularization (see Jezequel et al. (2020); Hazan (2012)).

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 $\ell(\hat{y}, y) = -\log P(\hat{y}|\mathbf{x}, \mathbf{w})$ and $\ell(h(\mathbf{x}), y) = -\log P(h(\mathbf{x})|\mathbf{x}, \mathbf{w})$, leading to the *regularized* pointwise regret

$$\mathcal{R}^{\varepsilon}_T(\hat{y}^T, y^T|\mathbf{x}^T) = -\sum_{t=1}^{T} \log P(\hat{y}_t|\mathbf{x}_t, \mathbf{w}) + \sup_{\mathbf{w}} \sum_{t=1}^{T} \log P(y_t|\mathbf{x}_t\mathbf{w})e^{-\varepsilon||\mathbf{w}||^2} \tag{4}$$

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

$$r_T^{\varepsilon}(\mathbf{x}^T) = \inf_{\hat{y}^T} \max_{y^T} \mathcal{R}_T^{\varepsilon}(\hat{y}^T, y^T|\mathbf{x}^T), \qquad \bar{r}_T^{\varepsilon}(\mathcal{H}) := \mathbb{E}_{\mathbf{x}^T}[r_T^{\varepsilon}(\mathcal{H}|\mathbf{x}^T)] \tag{5}$$

and the *generalized* Shtarkov sum is

$$S_{\varepsilon}(\mathbf{x}^T) = \sum_{y^T} \sup_{\mathbf{w}} P(y^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2}. \tag{6}$$

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. We show in Section 2 that $r_T^{\varepsilon}(\mathbf{x}^T) = \log S_{\varepsilon}(\mathbf{x}^T)$ holds as well.

In this paper, we present for the first time precise second-order asymptotics for the average Shtarkov sum and consequently the regularized minimax regret as in (5) with $\varepsilon > 0$. While our derivations do not directly work for $\varepsilon = 0$, we extend our findings to $\varepsilon \to 0$ as long as $\varepsilon \gg T^{-1/2}$ showing a phase transition of the leading term of the regret. This result also shed light on the regular minimax regret (i.e., $\varepsilon = 0$) in (3) when the weight norm R grows as $O(T^{1/4})$.

Related Work. Online learning under logarithmic loss can be viewed as universal compression (source coding) with side information, as discussed in Barron et al. (1998); Takeuchi and Barron (2006) and Drmota and Szpankowski (2024); Jacquet and Szpankowski (2004); Szpankowski and Weinberger (2012); Xie and Barron (1997, 2000). The logistic-type class of experts, as in (1), was studied extensively in Foster et al. (2018); Hazan et al. (2014); Rakhlin and Sridharan (2015); Shamir (2020a); Wu et al. (2022) 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 $(d/2)\log(TR^2/d)$ for the *sequential* regret, i.e., where both $\mathbf{x}^T$ and y^T are selected sequentially (Foster et al., 2018; Shamir, 2020b; Wu et al., 2022). Moreover, Foster et al. (2018) demonstrated that the $\log R$ dependency is tight for sequential regrets. However, for the *fixed design* regret we study here, the precise dependency on the weight norm R is largely unexplored. Several prior results, such as Shamir (2020b); Wu et al. (2022); Mayo et al. (2022), have demonstrated that the regret lower bound grows as $(d/2)\log(T/d^2)$ (with no dependency on R), which can deviate arbitrarily from the generic $(d/2)\log(TR^2/d)$ upper bound for $R \to \infty$. Recently, Drmota et al. (2024) showed that for *fixed design* regret, the upper bound can be improved to $2d\log T$ for a general *monotone class* even with $R = \infty$. They also demonstrated that for $R \geq \Omega(\sqrt{T})$, a *precise* $(1 + o(1))d\log(T/d)$ regret (i.e., including both upper and lower bounds) holds for logistic regression. This leaves open the question of *precisely* characterizing the fixed design regret in the transition region $0 \leq R \leq O(\sqrt{T})$. To the best of our knowledge, Drmota et al. (2024) is the only work that studies the precise characterization of fixed design regret with unbounded weights. We should emphasize that studying the transition region below $O(\sqrt{T})$ 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 Shamir and Szpankowski (2021); Jacquet et al. (2021). In Jacquet et al. (2021), a precise maximal minimax regret is analyzed, but only for a *finite* number of feature values. It should also be mentioned that the general form of our minimax regret (i.e., its second-order term) is related to Fisher information and was already known in information theory Takeuchi and Barron (2006), but only for *fixed* d and *bounded* weights.

Our Contributions. We first represent the regularized minimax regret as the logarithm of the generalized Shtarkov sum. Then in Theorem 2 we present for the first time precise second-order asymptotic expansion of the average Shtarkov sum and hence the regularized minimax regret as (5) for logistic regression (see Appendix F for some extensions) with *unbounded* weights. We prove that for $\varepsilon \gg T^{-1/2}$ the regularized minimax regret grows no faster than $(d/2)\log(2T/\pi) + \log C_d(\varepsilon)$ where $C_d(\varepsilon)$ has a complicated multidimensional integral representation which we explicitly evaluate for $\varepsilon \to 0$. In particular, for $\varepsilon = \Theta(T^{-\alpha})$ with $\alpha \leq 1/2$, we show in Corollary 4 that the leading term grows as $((1/2 + \alpha/4)d - \alpha/2)\log T$. 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 Flajolet and Sedgewick (2008); Szpankowski (2001)), hopefully initiating an analytic learning theory (see Drmota and Szpankowski (2024)) in which problems of machine learning are solved by tools of complex analysis.[2] As mention above, Drmota et al. (2024) proved that for $||\mathbf{w}|| = \Omega(\sqrt{T})$ the regular regret grows no faster than $d \log T + O(d)$, demonstrating a transition from $(d/2)\log T$ to $d \log T$ for unbounded weights. Observe that, our regularized regret in (5) can be interpreted as a *soft-constraint* on $R = ||\mathbf{w}||$ with $R \preceq 1/\sqrt{\varepsilon}$.

2. Main Results

In this section we present our main results with most proofs delegated to the Appendix. Before we start our discussion, we derive the connection between the regularized regret (5) and generalized Shtarkov sum (6). 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

$$\begin{aligned} r_T^{\varepsilon}(\mathcal{H}|\mathbf{x}^T) &= \min_Q \max_{y^T} [-\log Q(y^T|\mathbf{x}^T) + \sup_{\mathbf{w}} \log P(y^T|\mathbf{x}^T) e^{-\varepsilon||\mathbf{w}||^2}] \\ &= \min_Q \max_{y^T} [-\log Q(y^T|\mathbf{x}^T) + \log P_\varepsilon^*(y^T|\mathbf{x}^T)] + \log \sum_{v^T} \sup_{\mathbf{w}} P(v^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} \\ &\overset{(a)}{=} \log \sum_{y^T} \sup_{\mathbf{w}} P(y^T|\mathbf{x}^T, \mathbf{w}) e^{-\varepsilon||\mathbf{w}||^2} = \log S_\varepsilon(\mathbf{x}^T) \end{aligned} \tag{7}$$

where $S_\varepsilon(\mathbf{x}^T)$ is defined in (6) 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, while (a) follows since $\min_Q$ is attained when $Q = P_\varepsilon^*$.

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

$$P(y^T|\mathbf{x}^T, \mathbf{w}) = \prod_{t=1}^{T} p(y_t\langle \mathbf{x}_t, \mathbf{w}\rangle) \quad \text{and} \quad P_\varepsilon(y^T|\mathbf{x}^T, \mathbf{w}) = \prod_{t=1}^{T} p(y_t\langle \mathbf{x}_t, \mathbf{w}\rangle) e^{-\varepsilon\|\mathbf{w}\|^2}. \tag{8}$$

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

1. A. Odlyzko argued: "Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision."
2. Following Handmaid's percept: "The shortest paths between two truths on the real line passes through the complex plane."

To ease the presentation, we focus on the logistic regression $p(w) = (1 + \exp(-w))^{-1}$ since all interesting behavioral phenomena occurs for this function. 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

$$\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T, \mathbf{w}) = \sum_{t=1}^T p(\langle \mathbf{x}_t, \mathbf{w}\rangle)) p(-\langle \mathbf{x}_t, \mathbf{w}\rangle))\, \mathbf{x}_t \otimes \mathbf{x}_t + 2\varepsilon \mathbf{I}, \tag{9}$$

where $\mathbf{x} \otimes \mathbf{x}$ denotes the matrix $(x_i x_j)_{1,\le i,j \le d}$ 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

$$\nabla L_\varepsilon(y^T|\mathbf{x}^T, \mathbf{w}) = \mathbf{a} \tag{10}$$

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 $\mathbf{w}^*_\varepsilon = G^{-1}_{y^T|\mathbf{x}^T,\varepsilon}(\mathbf{0})$.

We shall analyze the Shtarkov sum via a multidimensional Fourier transform method. The first key issue is its existence, which we address next. 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

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

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

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

Observe that by (10) (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 $y^T, \mathbf{x}^T$.

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

$$h_{y^T|\mathbf{x}^T}(\mathbf{a}) = O\left(e^{-\frac{1}{4\varepsilon}\|\mathbf{a}\|^2}\right). \tag{11}$$

Hence the Fourier transform $\tilde{h}_{y^T|\mathbf{x}^T}(\mathbf{z})$ certainly exists. In the Appendix A we formally establish the following lemma proving the existence of the Fourier transform and its inverse exist for $\varepsilon > 0$.

Lemma 1 *For every fixed y^T, $\mathbf{x}^T$, $\varepsilon > 0$ and for all non-negative integers $k_1, \ldots, k_d$ we have*

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

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

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

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

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

Granted the existence of the Fourier $h_{y^T|\mathbf{x}^T}(\mathbf{z})$ we can proceed to evaluate the Shtarkov sum. We first observe that

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

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

$$f(\mathbf{w},\mathbf{x}_t,\mathbf{z}) = p(\langle\mathbf{x},\mathbf{w}\rangle)e^{-ip(-\langle\mathbf{x},\mathbf{w}\rangle)\,\langle\mathbf{x},\mathbf{z}\rangle} + p(-\langle\mathbf{x},\mathbf{w}\rangle)e^{ip(\langle\mathbf{x},\mathbf{w}\rangle)\,\langle\mathbf{x},\mathbf{z}\rangle}. \tag{12}$$

This leads to our integral representation of the Shtarkov sum

$$S_\varepsilon(\mathbf{x}^T) = (2\pi)^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\prod_{t=1}^T f(\mathbf{w},\mathbf{x}_t,\mathbf{z})\,e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{z}\rangle}\det\left(\nabla^2 L_\varepsilon(\cdot|\mathbf{x}^T,\mathbf{w})\right)\,d\mathbf{w}\,d\mathbf{z}. \tag{13}$$

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. Furthermore since P and $\nabla^2 L_\varepsilon$ are bounded in a bounded domain it follows that

$$\mathbb{E}S_\varepsilon(\mathbf{X}^T) = (2\pi)^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}\left[\prod_{t=1}^T f(\mathbf{w},\mathbf{X}_t,\mathbf{z})\det\left(\nabla^2 L_\varepsilon(\cdot|\mathbf{X}^T,\mathbf{w})\right)\right]e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{z}\rangle}\,d\mathbf{w}\,d\mathbf{z}. \tag{14}$$

This expression is the main tool that we will use to study asymptotically the minimax regret. The asymptotic evaluation of (14) is very challenging due to different behavior of the double multi-dimensional integrals for various ranges of $\mathbf{w}$ and $\mathbf{z}$ which both can be unbounded.

Our main result of his paper can be summarized as follows which we prove in Section 3 and in the Appendix E.

Theorem 2 *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 $\varepsilon \gg \max(1/\sqrt{T}, 1/T^{2/(d+1)})$ there exists $\beta(d) > 0$ such that*[3]

$$\mathbb{E}[S^\varepsilon(\mathbf{x}^T)] = \left(\frac{T}{2\pi}\right)^{d/2}\int_{\mathbb{R}^d}\sqrt{\det(\bar{\mathbf{B}}(\mathbf{w}))}\,e^{-\varepsilon\|\mathbf{w}\|^2}(1+O(T^{-\beta(d)})). \tag{15}$$

3. We expect that the theorem is valid at least for $\varepsilon \gg 1/\sqrt{T}$, however the proof would need some extra work, see Section 3.

with

$$\bar{\mathbf{B}}(\mathbf{w}) = \frac{1}{\mathrm{Vol}(\mathcal{B}_d)} \int_{\mathcal{B}_d} p(\langle \mathbf{w}|\mathbf{x}\rangle)(1 - p(\langle \mathbf{w}|\mathbf{x}\rangle))\mathbf{x} \otimes \mathbf{x} d\mathbf{x}, \tag{16}$$

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,*

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

where

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

Observe that for our second main finding presented in Corollary 4 we only need the upper bound for $\mathbb{E}\left[\log S^\varepsilon(\mathbf{X}^T)\right]$ but there are strong indications that actually (see next section):

$$\mathbb{E}\left[\log S^\varepsilon(\mathbf{X}^T)\right] \sim \log \mathbb{E}[S^\varepsilon(\mathbf{X}^T)].$$

Another question is whether from Theorem 2 we can recover the original regret with $\varepsilon = 0$. First, we observe that our proof Theorem 2 works only for $\varepsilon \gg 1/\sqrt{T}$ which basically translates to the radius R of $\mathbf{w}$ of order $O(T^{1/4})$. Second, from Drmota et al. (2024) we know that for $R \gg \sqrt{T}$ the leading term of the minimax regret is $d \log T$, not $(d/2)\log T$, thus there is a gap for $R \in (0, T^{1/2}]$ where we conjecture the leading term at $\log T$ grows from $d/2$ to d.

In order to study the behaviour of the regularized regret for $\varepsilon \to 0$ we need some asymptotic information about $\overline{B}(\mathbf{w})$. We recall that $\mathbf{X}$ is uniformly distributed in the d-dimensional ball $\mathcal{B}_d$ and that $\overline{B}(w) = \Theta(1, |w|^{-3})$ for $d = 1$. The proof of the following lemma is given in Appendix C.

Lemma 3 (i) *Suppose that* $d \ge 2$ *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* $\mathbf{R}^d$ *(thus* $\mathbf{I}_d - \mathbf{u}\otimes\mathbf{u}$ *is the identity operator on the hyperplane orthogonal to* $\mathbf{u}$*) with*

$$\lambda(\mathbf{w}) = \frac{ds_{d-1}}{s_d}\int_0^1\int_0^\pi t^{d-1}\cos(\theta)^2\sin(\theta)^{d-2} q(t\cos(\theta)\|\mathbf{w}\|)d\theta\, dt, \tag{17}$$

and

$$\phi(\mathbf{w}) = \frac{ds_{d-1}}{s_d}\int_0^1\int_0^\pi t^{d-1}\frac{\sin(\theta)^d}{d-1} q(t\cos(\theta)\|\mathbf{w}\|)d\theta\, dt \tag{18}$$

are the eigenvalues of $\bar{\mathbf{B}}(w)$ *with multiplicity* 1 *and* $d-1$*, respectively, and* s_d *denotes the area of the unit sphere* $\mathcal{S}_d$.

(ii) *The eigenvalue* $\lambda(\mathbf{w})$ *is asympotically of order* $\Theta(\|\mathbf{w}\|^{-2})$ *for* $d = 2$*, of order* $\Theta(\|\mathbf{w}\|^{-3}\log\|\mathbf{w}\|)$ *for* $d = 3$ *and of order* $\Theta(\|\mathbf{w}\|^{-3})$ *for* $d > 3$ *whereas the eigenvalue* $\phi(\mathbf{w})$ *is of order* $\Theta(\|\mathbf{w}\|^{-1})$ *for every* $d \ge 2$*. Consequently the determinant* $\det\bar{\mathbf{B}}(w) = \lambda(\mathbf{w})\cdot\phi^{d-1}(\mathbf{w})$ *is of order* $\Theta(\|\mathbf{w}\|^{-3})$ *for* $d = 2$*, of order* $\Theta(\|\mathbf{w}\|^{-5}\log\|\mathbf{w}\|)$ *for* $d = 3$*, and of order* $O(\|\mathbf{w}\|^{-d-2})$ *for* $d > 3$.

This leads to the following surprising conclusion which is our second main result.

Corollary 4 *Suppose that Theorem 2 holds for $\varepsilon \gg T^{-1/2}$. The we have in this range*

$$\mathbb{E}[S^{\varepsilon}(\mathbf{x}^T)] = \begin{cases} \Theta(T^{\frac{d}{2}}\varepsilon^{-\frac{d}{4}+\frac{1}{2}}) & d > 3, \\ \Theta(T^{1/2}) & d = 1, \\ \Theta(T\varepsilon^{-1/4}) & d = 2, \\ \Theta(T^{3/2}\varepsilon^{-1/4}\sqrt{\log(1/\varepsilon)}) & d = 3. \end{cases}$$

In particular, for $\varepsilon = \Theta(T^{-\alpha})$ with $\alpha \le \frac{1}{2}$, we have $\bar{r}_T^{\varepsilon} \le ((\frac{1}{2} + \frac{\alpha}{4})d - \frac{\alpha}{2})\log T + O(1)$.

Proof This follows by Lemma 3 and the following simple calculations (that we do only for $d > 3$)

$$\int_{\mathbb{R}^d} \sqrt{\det B(\mathbf{w})} e^{-\varepsilon\|\mathbf{w}\|^2} d\mathbf{w} \approx \int_{\|\mathbf{w}\|\ge 1} \|\mathbf{w}\|^{-(d+2)/2} e^{-\varepsilon\|\mathbf{w}\|^2} d\mathbf{w} \approx \int_1^{\infty} r^{d-1} r^{-(d+2)/2} e^{-\varepsilon r^2} dr$$
$$\approx \varepsilon^{-d/4+1/2} \int_0^{\infty} u^{d-1} u^{-(d+2)/2} e^{-u^2} du \approx \varepsilon^{-d/4+1/2}$$

by using substitution $u = r\sqrt{\varepsilon}$ for $\varepsilon \gg T^{-1/2}$. The last part then follows from Theorem 2 (ii). ■

3. Sketch of the Proof of Theorem 2

In this subsection we present ingredients of the proof of our main result. The proof is long, tedious, and very technical[4]. To help the reader, we focus here on $d = 1$. Extension for $d > 1$ is presented in Appendix E

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

$$\overline{f}(w, z) = \mathbb{E}f(w, X, z) = \frac{1}{2}\int_{-1}^{1} \left(\frac{e^{-ixz/(1+e^{xw})}}{1 + e^{-xw}} + \frac{e^{ixz/(1+e^{-xw})}}{1 + e^{xw}} \right) dx \tag{19}$$

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

$$\begin{aligned} \mathbb{E}S_{\varepsilon}(X^T) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[\prod_{t=1}^{T} f(w, X_t, z)\, \nabla^2 L_{\varepsilon}(\cdot|X^T, w)\right] e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\ &\overset{(a)}{=} \sum_{s=1}^{T} \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[\prod_{t=1}^{T} f(w, X_t, z)\, p(x_s w) p(-x_s w) x_s^2\right] e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\ &\quad + \frac{2\varepsilon}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[\prod_{t=1}^{T} f(w, X_t, z)\right] e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\ &= \frac{T}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overline{f}(w, z)^{T-1} B(z, w) e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz + \frac{2\varepsilon}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overline{f}(w, z)^{T} e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz, \\ &=: T \cdot J_0 + 2\varepsilon \cdot J_1, \end{aligned} \tag{20}$$

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 (9).

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

4. If requested, we will be happy to provide our notes with detailed derivations.

Lemma 5 *Set*

$$\overline{B}(w) = \mathbb{E}\left[p(xw)p(-xw)x^2\right] = \frac{1}{2}\int_{-1}^{1} \frac{x^2}{(1+e^{-xw})(1+e^{xw})}\, dx. \tag{21}$$

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 6 *If* $|w| \le 1$ *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 7 *Suppose that* $c_1 > 0$ *is a given constant. Then there exist* $c_2 > 0$ *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 first prove our main result for $\mathbb{E}S_\varepsilon(X^T)$ in the case $d = 1$ which we formulate next.

Proposition 8 *Suppose that* $d = 1$ *and that* X *is uniformly distributed on* $[-1,1]$. *Then*

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

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

Note that

$$\int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2 - 2\varepsilon^2 w^2/(T\overline{B}(w))}\, dw = \int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2}\, dw + O\left(T^{-1/10}\right).$$

Thus, the integral in Proposition 8 can be replaced by the integral $\int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2}\, dw$ as stated in Theorem 2.

The rest of this subsection is devoted to the proof of the Proposition 8. We recall from (20) 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 first 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 $|\overline{f}(z,w)| \le C/|z|$. We first look at the case, where $|z| \le 2C$. Here we certainly have the uniform representation (see Lemma 5)

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

$$
\begin{aligned}
I_1 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-C_3}^{C_3} \overline{f}(w,z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dw\, dz \\
&= \frac{1}{2\pi}\int_{-C_3}^{C_3}\int_{|z|\le T^{-1/3}} e^{-T\overline{B}(w)\frac{z^2}{2}}\left(1+O(Tz^3)\right) e^{-\varepsilon w^2 - 2i\varepsilon wz}\, dz\, dw \\
&+ \frac{1}{2\pi}\int_{-C_3}^{C_3}\int_{T^{-1/3}\le|z|\le 2C} O\left(e^{-c_1 T z^2}\right) e^{-\varepsilon w^2}\, dz\, dw + \frac{1}{2\pi}\int_{-C_3}^{C_3}\int_{|z|\ge 2C} (C/|z|)^T e^{-\varepsilon w^2}\, dz\, dw \\
&= \frac{1}{\sqrt{2\pi T}}\int_{-C_3}^{C_3} \overline{B}(w)^{-1/2} e^{-\varepsilon w^2}\, dw + O\left(T^{-1}\right).
\end{aligned}
$$

B: *The case* $C_3 \le |w| \le \eta T$. Next we consider the range $C_3 \le |w| \le \eta T$, where $\eta = \eta(T) \to 0$ will be fixed in the sequel. 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 5

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

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

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

for some $c > 0$. Similarly we obtain

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

Summing up we find

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

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 5)

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

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

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

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

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

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

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

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

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

Summing up we have

$$I_2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{C_3 \le |w| \le \eta T} \overline{f}(w,z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} \, dw \, dz = I_{21} + I_{22} + I_{23} + I_{24} + I_{25}$$
$$= \frac{1}{\sqrt{2\pi T}} \int_{|w| \ge C_3} \overline{B}(w)^{-1/2} e^{-\varepsilon w^2 - 2\varepsilon^2 w^2/(T\overline{B}(w))} \, dw$$
$$+ O\left(\frac{\varepsilon^{-3/4}}{T} + \frac{\varepsilon^{-5/4}}{T^2} + \frac{\eta^{1/6} e^{-c\eta^{-1/3}}}{\varepsilon^{5/2} T^{1/2}} + \varepsilon^{-5/4} T^{3/2} e^{-\varepsilon\eta^2 T^2} \right).$$

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

$$I_{31} = \frac{1}{2\pi} \int_{|z| \le e^{2|w|}/|w|} \int_{|w| \ge \eta T} \overline{f}(w,z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} \, dw \, dz$$
$$\ll \int_{|w| \ge \eta T} e^{2|w| - \varepsilon w^2} \, dw \ll \int_{|w| \ge \eta T} e^{-\frac{\varepsilon}{2} w^2} \, dw \ll \frac{1}{\varepsilon \eta T} e^{-\frac{\varepsilon}{2} (\eta T)^2} \ll e^{-\eta^2 T}.$$

If $|z| \ge e^{2|w|}/|w|$ we have $|\overline{f}(z,w)| \le \tilde{C}e^{|w|}/|zw|$ and, thus,

$$I_{32} = \frac{1}{2\pi}\int_{|z|\ge e^{2|w|}/|w|}\int_{|w|\ge \eta T} \overline{f}(w,z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz}\,dw\,dz \ll \frac{1}{\varepsilon\eta T^2} e^{-\frac{\varepsilon}{2}(\eta T)^2} \ll e^{-\eta^2 T}.$$

Consequently,

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

In conclusion, we arrive at

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

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.

In a similar manner, with another few pages of calculations (see also Appendix E), we obtain the asymptotics of J_0 which we summarize as follows:

$$J_0 = \frac{1}{\sqrt{2\pi T}}\int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2 - 2\varepsilon^2 w^2/(T\overline{B}(w))}\,dw + O\left(\frac{\log T}{T}\right)$$

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. In summary, we prove Proposition 8 (and hence the first part of Theorem 2 for $d = 1$).

Proof of Theorem 2(ii): Since the logarithm is concave it directly follows that $\mathbb{E}\left[\log\, S_\varepsilon(X^T)\right] \le \log \mathbb{E}\left[S_\varepsilon(X^T)\right]$ which provides the desired upper bound.

We now discuss a (potential) lower bound, but again to ease the presentation we just consider the case $d = 1$ (with some details deferred to Appendix D). We partition $S_\varepsilon(x^T)$ into two parts $S_\varepsilon(x^T) = S_{\varepsilon,1}(x^T) + S_{\varepsilon,2}(x^T)$, where $S_{\varepsilon,1}(x^T)$ should play the dominant rôle. More precisely we set

$$S_{\varepsilon,1}(x^T) = \frac{\mathbf{1}_{v\ge cT^{1/4}}}{2\pi}\int_{|z|\le 1}\int_{|w|\le T^{1/4}} \prod_{t=1}^{T} f(w,x_t,z)\,\nabla^2 L_\varepsilon(\cdot|x^T,w) e^{-\varepsilon w^2 - 2i\varepsilon wz}\,dw\,dz,$$

where $c > 0$ is a proper constant and v abbreviates $v = \sum_{t=1}^{T} x_t^2 p(x_t T^{1/4})p(-x_t T^{1/4})$. Suppose that y is a small real number and denote by $A(y)$ the event that $\left|S_{\varepsilon,2}(X^T)/S_{\varepsilon,1}(X^T)\right| \le y$. Then we certainly have

$$\begin{aligned}
\mathbb{E}\left[\log\, S_\varepsilon(X^T)\right] &= \mathbb{E}\log\left|S_\varepsilon(X^T)\right| \ge \mathbb{E}\left[\mathbf{1}_{A(y)}\cdot\log\left|S_{\varepsilon,1}(X^T)+S_{\varepsilon,2}(X^T)\right|\right] \\
&\ge \mathbb{E}\left[\mathbf{1}_{A(y)}\cdot\log\left(\left|S_{\varepsilon,1}(X^T)\right|(1-y)\right)\right] \ge \mathbb{E}\left[\mathbf{1}_{A(y)}\cdot\log\left|S_{\varepsilon,1}(X^T)\right|\right] - 2y \\
&= \mathbb{E}\left[\log\left|S_{\varepsilon,1}(X^T)\right|\right] - 2y - \mathbb{E}\left[\mathbf{1}_{A(y)^c}\cdot\log\left|S_{\varepsilon,1}(X^T)\right|\right] \\
&\ge \mathbb{E}\left[\log\left|S_{\varepsilon,1}(X^T)\right|\right] - 2y - \mathbb{P}(A(y)^c)\sqrt{\mathbb{E}\left[\log^2\left|S_{\varepsilon,1}(X^T)\right|\right]}.
\end{aligned}$$

Thus, in order to make this estimate work we need an upper bound for

$$\mathbb{P}(A(y)^c) = \mathbb{P}\left(\left|\frac{S_{\varepsilon,2}(X^T)}{S_{\varepsilon,1}(X^T)}\right| > y\right) \le \frac{1}{y}\mathbb{E}\left|\frac{S_{\varepsilon,2}(X^T)}{S_{\varepsilon,1}(X^T)}\right|.$$

The main term $S_{\varepsilon,1}(X^T)$ behaves quite nicely. In particular we can prove the following property (see Appendix D).

Lemma 9 *Suppose that $F(y)$ is defined for $|y| \ge 1$ and satisfies the following growth property: $|F(y)| \le C_1|y|^{D_1}$ for positive constants C_1, D_1 and that $y_1 \sim y_2 \to \infty$ implies $F(y_1) \sim F(y_2)$. Then we have*

$$\mathbb{E}\left[F(S_{\varepsilon,1}(X^T))\right] \sim F\left(\mathbb{E}\, S_{\varepsilon,1}(X^T)\right), \;\; \textit{and} \;\; \mathbb{E}\, S_{\varepsilon,1}(X^T) \sim \frac{T^{1/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sqrt{\overline{B}(w)}e^{-\varepsilon w^2}\, dw.$$

The basis of the proof of Lemma 9 is the following formula (that follows directly from the definition).

Lemma 10 *We have uniformly for $|z| \le 1$, $x \in [0,1]$ and all $w \in \mathbb{R}$*

$$\left(\prod_{t=1}^{T} f(w, x_t, z)\right)\nabla^2 L_\varepsilon(\cdot|x^T, w) = e^{-\frac{1}{2}Uz^2(1+O(z))}(U + 2\varepsilon),$$

where U abbreviates

$$U = \sum_{t=1}^{T} x_t^2 p(x_t w)p(-x_t w) =: \nabla^2 L(\cdot|x_t, w) \tag{22}$$

By applying the methods from above it is easy to obtain an upper bound for $\mathbb{E}\left[S_{\varepsilon,2}(X^T)\right] \ll \log T$, however the above methods seem to fail for getting upper bounds for $\mathbb{E}\left[|S_{\varepsilon,2}(X^T)|\right]$.

References

A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inf. Theory*, 44(6):2743–2760, Oct. 1998.

R. Bellman. *Introduction to Matrix Analysis*. SIAM, 1997.

Nicolo Cesa-Bianchi and Ohad Shamir. Efficient transductive online learning via randomized rounding. *arXiv preprint arXiv:1106.2429*, 2011.

M. Drmota and W. Szpankowski. *Analytic Information Theory: From Compression to learning*. Cambridge University Press, Cambridge, 2024.

M. Drmota, P. Jacquet, C. Wu, and W. Szpankowski. Minimax regret with unbounded weights, 2024. URL https://www.cs.purdue.edu/homes/spa/papers/isit24-unbounded.pdf.

P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2008.

Dylan J Foster, Satyen Kale, Haipeng Luo andMehryar Mohri, and Karthik Sridharan. Logistic regression: The importance of being improper. In *COLT - Conference on Learning Theory*, 2018.

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

E. Hazan. The convex optimization approach to regret minimization. In *S. Sra, S. Nowozin, and S. Wright, editors, Optimization for Machine Learning*, pages 287–303. MIT press, 2012.

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*, pages 197–209. MIT press, 2014.

M. N. Huxley. *Area, lattice points, and exponential sums*, volume 13 of *Lond. Math. Soc. Monogr., New Ser.* Oxford: Clarenton Press, 1996. ISBN 0-19-853466-3.

P. Jacquet and W. Szpankowski. Markov types and minimax redundancy for markov sources. *IEEE Trans. on Information Theory*, 5o:1393–1402, 2004.

P. Jacquet, G. I. Shamir, and W. Szpankowski. Precise minimax regret for logistic regression with categorical feature values. In *PRML: ALT'21*, volume 132, pages 755–771, 2021.

R. Jezequel, P. Gillard, and A. Rudi. Efficient improper learning for online logistic regression. In *PMLR, COLT'20*, volume 125, 2020.

J. Mayo, H. Hadiji, and T. van Erven. Scale-free unconstrained online learning for curved losses. In *COLT*, volume 178, pages 1–34, 2022.

Alexander Rakhlin and Karthik Sridharan. Sequential probability assignment with binary alphabets and large classes of experts. *arXiv preprint arXiv:1501.07340*, 2015.

G. I. Shamir. Logistic regression regret: What's the catch? In *COLT*, 2020a.

G. I. Shamir and W. Szpankowski. Low complexity approximate bayesian logistic regression for sparse online learning. In *ISIT21 and ArXiv: http://arxiv.org/abs/2101.12113*, 2021.

Gil I Shamir. Logistic regression regret: What's the catch? In *Conference on Learning Theory*, pages 3296–3319. PMLR, 2020b.

W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.

W. Szpankowski and M. Weinberger. Minimax pointwise redundancy for memoryless models over large alphabets. *IEEE Trans. Information Theory*, 58:4094–4104, 2012.

J. Takeuchi and A. Barron. Asymptotically minimax regret by bayes mixtures for non-exponential families. In *IEEE Information Theory Workshop*, 2006.

Changlong Wu, Mohsen Heidari, Ananth Grama, and Wojciech Szpankowski. Precise regret bounds for log-loss via a truncated bayesian algorithm. In *Advances in Neural Information Processing Systems*, 2022.

Q. Xie and A. Barron. Minimax redundancy for the class of memoryless sources. *IEEE Trans. Information Theory*, pages 647–657, 1997.

Q. Xie and A. Barron. Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Trans. Information Theory*, 46:431–445, 2000.

Appendix A. Proof of Lemma 1

We start the proof of Lemma 1 with the following result.

Lemma 11 *The determinant of the matrix matrix* $\nabla^2 L_\varepsilon(y^T|\mathbf{x}^T,\mathbf{w})$ *satisfies*

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

Proof By (9) we have for every vector $\mathbf{v}$

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

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

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

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. ■

We can now show that all derivatives of $h_{y^T|\mathbf{x}^T}(\mathbf{a})$ are absolutely integrable which is the first part of Lemma 1 which we repeat below.

Lemma 12 *For every fixed* y^T, $\mathbf{x}^T$, $\varepsilon > 0$ *and for all non-negative integers* $k_1,\dots,k_d$ *we have*

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

Proof By (11) 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

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

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 11). 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 $k_1, \dots k_d$ 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 the lemma. ■

This allows us to show that the Fourier transform exists proving Lemma 1 which we repeat below.

Lemma 13 *For every fixed y^T, $\mathbf{x}^T$, and $\varepsilon > 0$ the Fourier transform*

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

of $h_{y^T|\mathbf{x}^T}(\mathbf{a})$ exists and satisfies

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

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

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

Proof Since

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

and the latter integral exists by Lemma 12 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. Proof of Lemma 5–7

We start with the proof of Lemma 5.

Proof [Lemma 5] We use the representation (19) 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

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

Similarly it follows that

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

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 6 we need to further properties (that can be found in Huxley (1996)),

Lemma 14 *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*

$$\left|\int_{\beta_1}^{\beta_2} g(x)e^{ih(x)}\,dx\right| \le 2\frac{\max_{[\beta_1,\beta_2]}|g| + V_{\beta_1}^{\beta_2}(g)}{\min_{[\beta_1,\beta_2]}|h'|}, \tag{23}$$

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

Lemma 15 *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*

$$\left|\int_{\beta_1}^{\beta_2} g(x)e^{ih(x)}\,dx\right| \le 8\frac{\max_{[\beta_1,\beta_2]}|g| + V_{\beta_1}^{\beta_2}(g)}{\min_{[\beta_1,\beta_2]}\sqrt{|h''|}}. \tag{24}$$

The proof of Lemma 6 runs as follows.

Proof [Lemma 6] 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

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

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

$$|h'(x)| \ge \frac{|z|}{4} \tag{25}$$

and that $h''(x)$ has the same sign as z. Hence, by a direct application of Lemma 14 we obtain

$$\left| \int_0^1 p(-xw)e^{ih(x)}\,dx \right| \le \frac{8}{|z|}. \tag{26}$$

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

$$|h'(-x)| \ge \frac{|z|}{(1+e)^2} \tag{27}$$

and consequently we get

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

which implies

$$\overline{f}(z,w) \ll \frac{1}{|z|}.$$

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 (25) which implies (26).

However, for the second integral we have to distinguish between three ranges. If $0 \le x \le 1/w$ then we get again (27) 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 14 we apply Lemma 15 and obtain

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

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 14. In this range we have

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

which gives

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

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 7.

Proof [Lemma 7] We consider first the case $|x| \ge c_1'|w|$, where c_1' will be chosen sufficiently large. As in the proof of Lemma 5 it follows (if $w \ge 0$)

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

and

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

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 C. Proof of Lemma 3

We prove here Lemma 3. We recall that

$$\overline{B}(\mathbf{w}) = \frac{d}{s_d} \int_{\mathcal{B}_d} q(\langle \mathbf{x}, \mathbf{w} \rangle) \mathbf{x} \otimes \mathbf{x} \, d\mathbf{x}.$$

We start with part (i). Let θ be the angle between $\mathbf{x}$ and $\mathbf{u} = \mathbf{w}/\|\mathbf{w}\|$. We have the decomposition $\mathbf{x} = t(\cos(\theta)\mathbf{u} + \mathbf{b})$ with $\mathbf{b} \in \sin\theta \mathcal{S}_{d-1}(\mathbf{u})$ where $\mathcal{S}_{d-1}(\mathbf{u})$ is the unit hypersphere orthogonal to $\mathbf{u}$ and $t \in [0,1]$. Since $\mathbf{x}$'s have a spheric symmetry in its distribution, so it is the case for the $\mathbf{b}$'s in $\sin\theta \mathcal{S}_{d-1}(\mathbf{u})$ for any given angle θ. Thus

$$\begin{aligned}
\bar{\mathbf{B}}(\mathbf{w}) &= \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(t\|\mathbf{w}\| \cos\theta) d\theta \int_{\sin\theta \mathcal{S}_{d-1}(\mathbf{u})} (\mathbf{b} + \cos\theta \mathbf{u}) \otimes (\mathbf{b} + \cos\theta \mathbf{u}) d\mathbf{b} \, dt \qquad (28) \\
&= \frac{d}{s_d} \int_0^\pi t^{d-1} q(t\|\mathbf{w}\| \cos\theta) d\theta \int_{\sin\theta \mathcal{S}_{d-1}(\mathbf{u})} (\mathbf{b} \otimes \mathbf{b} + (\cos\theta)^2 \mathbf{u} \otimes \mathbf{u}) d\mathbf{b} \, dt \\
&+ \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(\|\mathbf{w}\| \cos\theta) d\theta \int_{\sin\theta \mathcal{S}_{d-1}(\mathbf{u})} \cos\theta (\mathbf{b} \otimes \mathbf{u} + \mathbf{u} \otimes \mathbf{b}) d\mathbf{b}. \qquad (29)
\end{aligned}$$

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

$$\begin{aligned}
\bar{\mathbf{B}}(\mathbf{w}) &= \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(t\|\mathbf{w}\| \cos\theta) d\theta \int_{\sin\theta \mathcal{S}_{d-1}(\mathbf{u})} (\mathbf{b} \otimes \mathbf{b} + (\cos\theta)^2 \mathbf{u} \otimes \mathbf{u} d\mathbf{b} \, dt \\
&= \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(t\|\mathbf{w}\| \cos\theta) (\sin\theta)^{d-1} d\theta \, dt \\
&\qquad \int_{\mathcal{S}_{d-1}(\mathbf{u})} ((\sin\theta)^2 \mathbf{b} \otimes \mathbf{b} + (\cos\theta)^2 \mathbf{u} \otimes \mathbf{u}) d\mathbf{b}. \qquad (30)
\end{aligned}$$

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

The quantity $\int_{\mathcal{S}_{d-1}(\mathbf{u})} \mathbf{b} \otimes \mathbf{b}$ is the $(d-1) \times (d-1)$ matrix whose (i,j) coefficient is $\int_{\mathcal{S}_{d-1}} b_i b_j d\mathbf{b}$. Clearly, by spheric symmetry of the $\mathbf{b}$ vectors $\int_{\mathcal{S}_{d-1}} b_i b_j d\mathbf{b} = 0$ when $i \neq j$. We also have for all $i \neq j$:

$$\int_{\mathcal{S}_{d-1}} (b_i)^2 d\mathbf{b} = \int_{\mathcal{S}_{d-1}} (b_j)^2 d\mathbf{b} = \frac{1}{d-1} \int_{\mathcal{S}_{d-1}} \|\mathbf{b}\|^2 d\mathbf{b} = \frac{s_{d-1}}{d-1}. \qquad (31)$$

Thus

$$\int_{\mathcal{S}_{d-1}(\mathbf{u})} \mathbf{b} \otimes \mathbf{b} d\mathbf{b} = \frac{s_{d-1}}{d-1} \mathbf{I}_{d-1}(\mathbf{u}) \qquad (32)$$

and similarly

$$\int_{\mathcal{S}_{d-1}(\mathbf{u})} \mathbf{u} \otimes \mathbf{u} d\mathbf{b} = s_{d-1} \mathbf{u} \otimes \mathbf{u} \qquad (33)$$

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

Now we move to part (ii) of Lemma 3. 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 Mellin transform which is an effective tool of analytic combinatorics for complex asymptotics. The reader is refereed to Flajolet and Sedgewick (2008) and Szpankowski (2001) for detailed discussions.

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

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

Observe now that

$$\lambda(w) = 2\frac{ds_{d-1}}{s_d} \int_0^1 \int_0^{\pi/2} t^{d-1} q(t\cos(\theta)w) \cos^2(\theta) \sin^{d-2}(\theta) d\theta\, dt \quad (34)$$

$$= \frac{2ds_{d-1}}{s_d} \int_0^1 \int_0^1 t^{d-1} y^2 (1-y^2)^{(d-3)/2} q(tyw) dy \quad (35)$$

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

$$\lambda^*(s) = \frac{2ds_{d-1}}{s_d} \int_0^1 \int_0^1 t^{d-1} (1-y^2)^{(d-3)/2} y^2 \int_0^\infty q(tyw) x^{s-1} dy\, dy\, dt \quad (36)$$

$$= \frac{2ds_{d-1}}{s_d} q^*(s) \int_0^1 t^{d-1-s}\, dt \int_0^1 (1-y^2)^{(d-3)/2} y^{2-s} dy \quad (37)$$

$$= \frac{2ds_{d-1}}{s_d} \frac{q^*(s)}{d-s} \beta_1^*(3-s) \quad (38)$$

where $q^*(s)$ is the Mellin transform of $q(x) = p(x)(1-p(x)$ and $\beta_1(s)$ is the Mellin transform of the function $(1-y^2)^{(d-3)/2} y$ defined over $[0,1]$.

The Mellin transform $\beta_1^*(s)$ is defined for $\mathrm{Re}(s) > 0$ and being locally analytical it has poles on the negative even integers, corresponding to the Taylor expansion of $(1-y^2)^{(d-3)/2}$. Actually it can be written in terms of the Beta function and, thus, in terms of the Gamma function, and these representations directly provide the corresponding meromorphic extension. The Mellin transform $q^*(s)$ of function $q(x)$ is

$$q^*(s) = (s-1)^2 (s-2) h^*(s-2)$$

where $h^*(s)$ is the Mellin transform of function $h(x) = \log(1+e^{-x})$ defined for $\mathrm{Re}(s) > 0$. The Mellin transform $q^*(s)$ is defined for $\mathrm{Re}(s > 2$ but the simple poles at $s = 1$ and $s = 2$ are canceled by the factor $(s-1)(s-2)$ thus is finally defined for $\mathrm{Re}(s) > 0$. More precisely, we have

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

where $\Gamma(s)$ is the Euler gamma function and $\zeta(s)$ is the Riemann zeta function.

If $d = 2$ the simple pole coming from the factor $1/(2-s)$ dominates and shows that $\lambda(w)$ behaves as w^{-2} as $w \to \infty$. If $d = 2$ the two simple poles at $s = 3$ (coming from $1/(3-s)$ and $\beta_1^*(3-s)$) correspond to the leading asymptotic behavior of the form $w^{-3} \log w$. Finally, for $d > 3$ we just have a simple pole at $s = 3$ coming from $\beta_1^*(3-s)$ and so $\lambda(w)$ behaves as w^{-3} as $w \to \infty$.

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

$$\phi^*(s) = 2\frac{ds_{d-1}}{(d-1)s_d} \frac{\beta_2^*(1-s)}{d-s} q^*(s),$$

where $\beta_2^*(s)$ is the Mellin transform of the function $(1-y^2)^{(d-1)/2}$. For $d \geq 2$ there is always a dominant simple pole at $s = 1$ coming from $\beta_2^*(s)$ which is reflected by the asymptotic order w^{-1} for $\phi(w)$ we $w \to \infty$.

The Mellin transform $\phi^*(s)$ is also defined on $\mathrm{Re}(s) \in]0,3[$ and has a simple pole at $s = 3$ with residue $-\zeta(2)\Gamma(3)\frac{d-1}{2} = -\pi^2/6(d-1)$. For both $\lambda^*(s)$ and $\phi^*(s)$ the next pole is at $s = 5$. This completes the proof of Lemma 3.

Appendix D. Proof of Lemma 9

Before we prove Lemma 9 we recall the following fact (see Lemma 10): we have uniformly for $|z| \le 1$, $x \in [0,1]$ and all $w \in \mathbb{R}$

$$\left(\prod_{t=1}^{T} f(w, x_t, z) \right) \nabla^2 L_\varepsilon(\cdot|x^T, w) = e^{-\frac{1}{2}Uz^2(1+O(z))}(U + 2\varepsilon),$$

where U abbreviates $U = \sum_{t=1}^T x_t^2 p(x_t w) p(-x_t w) =: \nabla^2 L(\cdot|x_t, w)$.

We also recall that

$$S_{\varepsilon,1}(x^T) = \frac{\mathbf{1}_{v \ge cT^{1/4}}}{2\pi} \int_{|z|\le 1} \int_{|w| \le T^{1/4}} \prod_{t=1}^{T} f(w, x_t, z)\, \nabla^2 L_\varepsilon(\cdot|x^T, w) e^{-\varepsilon w^2 - 2i\varepsilon w z}\, dw\, dz.$$

Now we are ready to prove Lemma 9 which we repeat below.

Lemma 16 *Suppose that $F(y)$ is defined for $|y| \ge 1$ and satisfies the following growth property:*

$$|F(y)| \le C_1 |y|^{D_1}$$

for positive constants C_1, D_1 and that $y_1 \sim y_2 \to \infty$ implies $F(y_1) \sim F(y_2)$. Then we have for positive constants C_1, D_1 and that $y_1 \sim y_2 \to \infty$ implies $F(y_1) \sim F(y_2)$. Then we have

$$\mathbb{E}\left[F(S_{\varepsilon,1}(X^T))\right] \sim F\left(\mathbb{E}\, S_{\varepsilon,1}(X^T)\right)$$

and

$$\mathbb{E}\, S_{\varepsilon,1}(X^T) \sim \frac{T^{1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\overline{B}(w)} e^{-\varepsilon w^2}\, dw.$$

Proof We consider now U (defined in (22)) as a random variable (actually a sum of T iid random variables). By applying a proper Chernov bound it follows that (for some constant $c > 0$)

$$\mathbb{P}\left(|U - T\overline{B}(w)| \ge y\right) \ll e^{-cy^2 \max(1,|w|^5)/T}$$

If we assume that $|w| \le T^{1/4}$ and $y = \overline{B}(w)T^{3/4}$ it follows that

$$\mathbb{P}\left(|U - T\overline{B}(w)| \ge \overline{B}(w)T^{3/4}\right) \ll e^{-c'T^{1/12}}$$

for some constant $c' > 0$. Now we use the monotonicity of U in w and the values $w_j = j/T^2$, $0 \le j \le T^{9/4}$ to conclude that

$$\mathbb{P}\left(\exists\, |w| \le T^{1/4} : |U - T\overline{B}(w)| \ge \overline{B}(w)T^{3/4}/2\right) \ll e^{-c''T^{1/12}}$$

for some constant $c'' > 0$.

Denote by B the event that $|U - T\overline{B}(w)| < \overline{B}(w)T^{3/4}/2$ for all $|w| \le T^{1/4}$. Then we know that $\mathbb{P}(B) \ge 1 - e^{-c''T^{1/12}}$. Note that v is exactly U if $w = T^{1/4}$. Thus, if B holds then we also have $v \ge cT^{2/5}$ for a proper constant c. So we do not have to take of the indicator function $\mathbf{1}_{v \ge cT^{2/5}}$.

Now, if we condition on B we certainly have

$$U = T\overline{B}(w)(1 + O(T^{-1/4}).$$

Thus we obtain uniformly

$$\int_{|z|\le 1} e^{-\frac{1}{2}Uz^2(1+O(z))-2i\varepsilon zw}(U + 2\varepsilon)\, dz \sim \sqrt{2\pi U} \sim \sqrt{2\pi T\overline{B}(w)}$$

and consequently

$$\begin{aligned} S_{\varepsilon,1}(X^T) &\sim \frac{T^{1/2}}{\sqrt{2\pi}} \int_{|w|\le T^{1/4}} \sqrt{\overline{B}(w)} e^{-\varepsilon w^2}\, dw \\ &\sim \frac{T^{1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\overline{B}(w)} e^{-\varepsilon w^2}\, dw. \end{aligned}$$

In general we certainly have the trivial upper bound

$$\int_{|z|\le 1} e^{-\frac{1}{2}Uz^2(1+O(z))}(U + 2\varepsilon)\, dz \le \eta(U + 2\varepsilon) \ll T$$

and so we find

$$S_{\varepsilon,1}(X^T) \ll T^{5/4}.$$

We now prove a lower lower bound for which we need the introduced condition $v \ge cT^{1/4}$. Recall that v is U for $w = T^{1/4}$ so that we have $U \ge v \ge cT^{1/4}$ for all $|w| \le T^{1/4}$. We compute the Gaussian type integral more precisely

$$\begin{aligned} \int_{|z|\le 1} e^{-\frac{1}{2}Uz^2(1+O(z))-2i\varepsilon zw}\, dz &= \int_{-\infty}^{\infty} e^{-\frac{1}{2}Uz^2(1+O(z))-2i\varepsilon zw}\, dz + O\left(\frac{1}{U}\right) \\ &= \sqrt{\frac{2\pi}{U}} e^{-2\varepsilon^2 w^2/U} + O\left(\frac{1}{U}\right). \end{aligned}$$

Now we observe that $v \ge cT^{1/4}$ implies that $2\varepsilon^2 w^2/U = o(\varepsilon w^2)$. Thus, the factor $e^{-2\varepsilon^2 w^2/U}$ will not play a rôle when we multiply with $(U + 2\varepsilon)e^{-\varepsilon w^2}$ and integrate with respect to w. Consequently we get a general lower bound for $S_{\varepsilon,1}(X^T)$ of the form

$$S_{\varepsilon,1}(X^T) \gg \sqrt{v} \int_0^{T^{1/4}} e^{-\varepsilon w^2/2}\, dw \gg \sqrt{\frac{v}{\varepsilon}} \gg T^{1/8}/\sqrt{\varepsilon}$$

provided that $v \ge cT^{1/4}$ and $\varepsilon \gg T^{-1/2}$. We note again that the error term

$$O\left(\int_0^{T^{1/4}} e^{-\varepsilon w^2}\, dw\right) = O(1/\sqrt{\varepsilon})$$

is negligible.

Finally in order to compute the expected value $\mathbb{E}F(S_{\varepsilon,1}(X^T))$ we consider the following partition:

$$\mathbb{E}\left[F(S_{\varepsilon,1}(X^T))\right] = \mathbb{E}\left[\mathbf{1}_{v\geq cT^{1/4}}\cdot \mathbf{1}_B \cdot F(S_{\varepsilon,1}(X^T))\right], \\ + \mathbb{E}\left[\mathbf{1}_{v\geq cT^{1/4}}\cdot \mathbf{1}_{B^c} \cdot F(S_{\varepsilon,1}(X^T))\right].$$

We now note that $\mathbb{P}(v \leq cT^{1/4}) \geq 1 - e^{-c''T^{3/4}}$ if c is properly chosen. Thus by assumption and the properties of the events B and $\{v \geq cT^{1/4}\}$ it follows that

$$\mathbb{E}\left[\mathbf{1}_{v\geq cT^{1/4}}\cdot \mathbf{1}_B \cdot F(S_{\varepsilon,1}(X^T))\right] \sim F\left(\frac{T^{1/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sqrt{\overline{B}(w)}e^{-\varepsilon w^2}\,dw\right).$$

For the second term we use Cauchy-Schwarz's inequality so that we have

$$\mathbb{E}\left[\mathbf{1}_{v\geq cT^{1/4}}\cdot \mathbf{1}_{B^c} \cdot F(S_{\varepsilon,1}(X^T))\right] \leq \sqrt{\mathbb{P}(B^c)}\sqrt{\mathbb{E}\left[\mathbf{1}_{v\geq cT^{1/4}}\cdot |F(S_{\varepsilon,1}(X^T))|^2\right]}.$$

Now we use the general upper bounds for $S_{\varepsilon,1}(X^T))$ with gives

$$\mathbb{E}\left[\mathbf{1}_{v\geq cT^{1/4}}\cdot |F(S_{\varepsilon,1}(X^T))|^2\right] \ll T^{5D_1/2}$$

The lower bound ensures that $|S_{\varepsilon,1}(X^T)| \geq 1$ so that everything is well defined. This completes the proof of the lemma. ■

Appendix E. Proof of Theorem 2 for general d

We fix $d > 1$ and consider first the expected value $\mathbb{E}S_\varepsilon(\mathbf{X}^T)$. By using the representation (14) 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 $d = 0$) 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 J_0 (the other terms are similar but do not have such a nice explicit form).

E.1. Upper bounds for $\overline{f}(\mathbf{w}, \mathbf{z})$

We need analogous for Lemmas 5–7 for $d > 1$. Actually the situation is slightly more difficult.

Lemma 17 *We have uniformly for* $\mathbf{z} \in \mathbb{R}^d$

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

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*

$$\log \overline{f}(\mathbf{w}, \mathbf{z}) = -\frac{1}{2}\mathbf{z}^\tau \overline{B}(\mathbf{w})\mathbf{z} + \begin{cases} O(\|\mathbf{z}\|^3 + \|\mathbf{z}\|^4) & \textit{for } \|\mathbf{w}\| \le 1, \\ O\left(\frac{\|\mathbf{z}\|^3}{\|\mathbf{w}\|^4} \left(|\cos\varphi| + \|\mathbf{w}\| |\sin\varphi|\right)^3 + \frac{\|\mathbf{z}\|^4}{\|\mathbf{w}\|^6} \left(|\cos\varphi| + \|\mathbf{w}\| |\sin\varphi|\right)^4 \right) & \textit{for } \|\mathbf{w}\| > 1. \end{cases}$$

Proof The case $\|\mathbf{w}\| \le 1$ is easy to handle. We just use Taylor expansion and the property that

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

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 $\mathbf{x} = x_1 \mathbf{w}_0 + x_2 \mathbf{w}_1 + \mathbf{x}_2$, where $\mathbf{x}_2$ is in the 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 the integral

$$\begin{aligned} \mathbf{z}^\tau \overline{B}(\mathbf{w})\mathbf{z} &= \mathbb{E}\left[p(\langle \mathbf{X}, \mathbf{w}\rangle) p(-\langle \mathbf{X}, \mathbf{w}\rangle) \langle \mathbf{X}, \mathbf{z}\rangle^2 \right] \\ & \mathbb{E}\left[p(x_1\|\mathbf{w}\|) p(-x_1\|\mathbf{w}\|) (A x_1 + B x_2)^2 \right] \end{aligned}$$

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

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

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. The computations are very similar to the preceding ones. For example we have for the third derivative

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

which gives

$$\left|\sum_{i,j,k} \frac{\partial^3 \overline{f}(\mathbf{w},\mathbf{z}}{\partial z_i \partial z_j \partial z_k} z_i z_j z_k\right| \ll \mathbb{E}\left[p(\langle \mathbf{X},\mathbf{w}\rangle)p(-\langle \mathbf{X},\mathbf{w}\rangle)\,|\langle \mathbf{X},\mathbf{z}\rangle|^3\right]$$
$$\ll \frac{1}{\|\mathbf{w}\|^4}\left(|\cos\varphi| + \|\mathbf{w}\||\sin\varphi|\right)^3.$$

This directly leads to the proposed asymptotic relation. ■

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

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

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

$$\int_0^1 t^{d-1} f(\mathbf{w}, t\mathbf{x}, \mathbf{z})\, dt \ll \frac{1}{|\langle \mathbf{x},\mathbf{z}\rangle|} \tag{39}$$

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

$$\int_0^1 t^{d-1} f(\mathbf{w}, t\mathbf{x}, \mathbf{z})\, dt \ll \frac{1}{\sqrt{|\langle \mathbf{x},\mathbf{z}\rangle\langle \mathbf{x},\mathbf{w}\rangle|}} + \frac{e^{|\langle \mathbf{x},\mathbf{w}\rangle|}}{|\langle \mathbf{x},\mathbf{z}\rangle\langle \mathbf{x},\mathbf{w}\rangle|} \tag{40}$$

Proof The proof follows that same lines as the proof of Lemma 6. 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 a corollary we obtain the following upper bounds for $\overline{f}(\mathbf{z}, \mathbf{w})$.

Lemma 19 *If* $\|\mathbf{w}\| \le 1$ *then we have*

$$|\overline{f}(\mathbf{z}, \mathbf{w})| \le \min\left(1, C_1 \frac{\log(\|\mathbf{z}\|)}{\|\mathbf{z}\|}\right) \tag{41}$$

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

$$|\overline{f}(\mathbf{z}, \mathbf{w})| \le \min\left(1, C_2 \frac{\log(\|\mathbf{z}\|\,\|\mathbf{w}\|) + e^{\|\mathbf{w}\|}}{\sqrt{\|\mathbf{z}\|\,\|\mathbf{w}\|}}\right). \tag{42}$$

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 18 we have

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

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

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

as proposed.

Now suppose that $\|\mathbf{w}\| \ge 1$. Then either (39) or (40) holds. But since $\|\mathbf{w}\| \ge 1$ then (39) implies (40). Thus we have (40) in all cases. So we 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

$$K_1 \ll \frac{\log(\|\mathbf{z}\|\,\|\mathbf{w}\|)}{\sqrt{\|\mathbf{z}\|\,\|\mathbf{w}\|}}.$$

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

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

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

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

This completes the proof of the lemma. ∎

Lemma 20 *There exist constants* $c_1 > 0$ *and* $c_2 > 0$ *such that*

$$|\overline{f}(\mathbf{z}, \mathbf{w})| \le 1 - \frac{c_2}{\|\mathbf{w}\|} \tag{43}$$

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 6 and 7):

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

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

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

Furthermore we have

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

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 (43) for some constant $c_2 > 0$. ■

Lemma 21 *Suppose that $c_3 > 0$ is a given constant. Then there exists $c_4 > 0$ such that*

$$|\overline{f}(\mathbf{z}, \mathbf{w})| \le 1 - \frac{c_4}{\|\mathbf{z}\|^d} \tag{44}$$

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

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

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 $c_3 > 0$ we have

$$\left| \int_{[0,1/\|\mathbf{z}\|]^d} f(\mathbf{w}, \mathbf{x}, \mathbf{z}) \, d\mathbf{x} \right| \le \frac{1 - c_4}{\|\mathbf{z}\|^d} \tag{45}$$

uniformly for all $\mathbf{z}$ with $\|\mathbf{z}\| \ge c_3 \|\mathbf{w}$. Clearly (45) implies (44).

We apply the substitution $\mathbf{u} = \|\mathbf{z}\|\mathbf{x}$. Furthermore we set $c = \|\mathbf{w}\|/\|\mathbf{z}\|$, $\mathbf{z}_0 = \mathbf{z}/\|\mathbf{z}\|$ and $\mathbf{w}_0 = \mathbf{w}/\|\mathbf{w}\|$, and we also represent $\mathbf{z}_0$ as $\mathbf{z}_0 = S\mathbf{w}_0$ for some rotation S. With these notations in inequality (45) is equivalent to the inequality

$$\left| \int_{[0,1]^d} \left(p(c\langle \mathbf{u}, \mathbf{w}_0\rangle) e^{-ip(-c\langle \mathbf{u},\mathbf{w}_0\rangle)\langle \mathbf{u}, S\mathbf{w}_0\rangle} + p(-c\langle \mathbf{u}, \mathbf{w}_0\rangle) e^{ip(c\langle \mathbf{u},\mathbf{w}_0\rangle)\langle \mathbf{u}, S\mathbf{w}_0\rangle} \right) d\mathbf{u} \right| \le 1 - c_4$$

that should now hold uniformly in $c \in [0, 1/c_2]$, $\mathbf{w}_0$, and S. However, this is trivial by a compactness argument. Clearly for every choice of $c \in [0, 1/c_2]$, $\mathbf{w}_0$, and S the left hand side is smaller than 1, and the left hand side is continous in $c \in [0, 1/c_2]$, $\mathbf{w}_0$, and S. ■

E.2. Asymptotics of $\tilde{J}_0$

Based on the above properties we derive an asymptotic representation for the integral J_0.

Proposition 22 *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 $\varepsilon \gg \max(T^{-1/2}, T^{-2/(d+1)})$*, where* $\overline{B}(w)$ *is given by (16).*

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 by estimating all parts absolutely) that

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

Furthermore by Taylor expansion (and similar computations) we have

$$\det\left(B(\mathbf{z}, \mathbf{w})\right) = \det\left(\overline{B}(\mathbf{w})\right) \cdot \left(1 + O\left(\|\mathbf{z}\|^2 \min(1, \|\mathbf{w}\|^{-1})\right)\right) \tag{46}$$

$$= \det\left(\overline{B}(\mathbf{w})\right) + O\left(\|\mathbf{z}\|^2 \min(1, \|\mathbf{w}\|^{-d-3})\right) \tag{47}$$

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

A: *The case* $\|\mathbf{w}\| \leq 1$. First suppose that $\|\mathbf{z}\| \leq T^{-1/3}$:

$$\begin{aligned}
I_{11} &= \frac{1}{(2\pi)^d} \int_{\|\mathbf{z}\| \leq T^{-1/3}} \int_{\|\mathbf{w}\| \leq 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}\| \leq 1} \det(\overline{B}(\mathbf{w})) \\
&\qquad \int_{\|\mathbf{z}\| \leq 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}\| \leq 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}\| \leq 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}\| \leq 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}\| \leq 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}\| \leq 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 $T^{-1/3} \leq \|\mathbf{z}\| \leq C$, where C is chosen in a way that $C/\log C \geq 2C_1$, where C_1 is the constant from the inequality (41):

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

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

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

Summing up we have

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

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 17 and by(47) we have

$$\int_{\|\mathbf{z}\|\le z_1}\det(B(\mathbf{w},\mathbf{z}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-2i\varepsilon\langle\mathbf{w},\mathbf{w}\rangle}\,dz$$
$$= \det(\overline{B}(\mathbf{w}))\int_{\|\mathbf{z}\|\le z_1} e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau\overline{B}(\mathbf{w})\mathbf{z}-2i\varepsilon wz}\,dz$$
$$+ O\left(\|\mathbf{w}\|^{-d-3}\int_{\|\mathbf{z}\|\le z_1}\|\mathbf{z}\|^2e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau\overline{B}(\mathbf{w})\mathbf{z}}\,d\mathbf{z}\right)$$
$$+ O\left(\|\mathbf{w}\|^{-d-6}\int_{\|\mathbf{z}\|\le z_1}\|\mathbf{z}\|^3e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau\overline{B}(\mathbf{w})\mathbf{z}}(1+\|\mathbf{w}\|)^3\,d\mathbf{z}\right)$$
$$+ O\left(\|\mathbf{w}\|^{-d-8}\int_{\|\mathbf{z}\|\le z_1}\|\mathbf{z}\|^4e^{-(T-d)\frac{1}{2}\mathbf{z}^\tau\overline{B}(\mathbf{w})\mathbf{z}}(1+\|\mathbf{w}\|)^4\,d\mathbf{z}\right)$$
$$= \sqrt{\det(\overline{B}(\mathbf{w}))}\left(\frac{2\pi}{T}\right)^{d/2}e^{-\frac{2\varepsilon^2}{T}\mathbf{w}^\tau\overline{B}(\mathbf{w})^{-1}\mathbf{w}}$$
$$+ O\left(\frac{\|\mathbf{w}\|^{(d-4)/2}}{T^{d/2}}e^{-c\eta^{-1/3}}\right) + O\left(\frac{\|\mathbf{w}\|^{d/2}}{T^{(d+2)/2}}\right)$$
$$+ O\left(\frac{\|\mathbf{w}\|^{(d+3)/2}}{T^{(d+3)/2}}\right) + O\left(\frac{\|\mathbf{w}\|^{(d+4)/2}}{T^{(d+4)/2}}\right).$$

This implies

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

Note that the integral is of order

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

that is asymptotically leading if $\varepsilon \gg T^{-2/(d+1)}$.

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 $c > 0$)

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

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

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

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 $|\overline{f}(\mathbf{z},\mathbf{w})| \le 1 - c_4/(\|\mathbf{w}\|\log^2\|\mathbf{w}\|)^d$ and, thus, we get

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

We split the integral at $r_0 = (T/\varepsilon)^{1/(d+2)}$ so that

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

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

$$I_{23} \ll e^{-c''T^\kappa}$$

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

Next suppose that $c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\| \le \|\mathbf{z}\| \le c_3 e^{4\|\mathbf{w}\|}$ 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

$$
\begin{aligned}
I_{24} &= \frac{1}{(2\pi)^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\int_{c_1'\|\mathbf{w}\|\log^2\|\mathbf{w}\|\le\|\mathbf{z}\|\le c_3 e^{4\|\mathbf{w}\|}} \det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d} e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{w}\rangle}\,d\mathbf{w}\,d\mathbf{z}\\
&\ll \int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}e^{-c_2T/\|\mathbf{w}\|+4d\|\mathbf{w}\|-\varepsilon\|\mathbf{w}\|^2}\,dw\\
&\le \int_{1\le\|\mathbf{w}\|\le c_2\sqrt{T}}\|\mathbf{w}\|^{-d-2}e^{-(1-c_2)\sqrt{T}-\varepsilon\|\mathbf{w}\|^2}\,d\mathbf{w} + \int_{c_2\sqrt{T}\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}e^{-\varepsilon/2\|\mathbf{w}\|^2}\,dw\\
&\ll e^{-(1-c_2)\sqrt{T}} + e^{-(c_2^2/2)\varepsilon T}
\end{aligned}
$$

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

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

$$
\overline{f}(\mathbf{z},\mathbf{w}) \le \tilde{C}\frac{\max\left(\log(\|\mathbf{w}\|),\log(\|\mathbf{z}\|),e^{\|\mathbf{w}\|\|}\right)}{\sqrt{\|\mathbf{w}\|\,\|\mathbf{z}\|}} \le \frac{1}{2}. \tag{48}
$$

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

$$
\begin{aligned}
I_{25} &= \frac{1}{(2\pi)^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\int_{\|\mathbf{z}\|\ge c_3e^{4\|\mathbf{w}\|}}\det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{z}\rangle}d\mathbf{z}\,d\mathbf{w}\\
&\ll \tilde{C}^{T-d}\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2}\frac{(\log(\|\mathbf{w}\|))^{T-d}}{\|\mathbf{w}\|^{(T-d)/2}}\int_{\|\mathbf{z}\|\ge c_3e^{4\|\mathbf{w}\|}}\|\mathbf{z}\|^{-(T-d)/2}d\mathbf{z}\,d\mathbf{w}\\
&+ \tilde{C}^{T-d}\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2-(T-d)/2}\int_{\|\mathbf{z}\|\ge c_3e^{4\|\mathbf{w}\|}}\frac{(\log(\|\mathbf{z}\|))^{T-d}}{\|\mathbf{z}\|^{(T-d)/2}}d\mathbf{z}\,d\mathbf{w}\\
&+ \tilde{C}^{T-d}\int_{1\le\|\mathbf{w}\|\le\eta T}\|\mathbf{w}\|^{-d-2-(T-d)/2}e^{(T-d)\|\mathbf{w}\|}\int_{\|\mathbf{z}\|\ge c_3e^{4\|\mathbf{w}\|}}\|\mathbf{z}\|^{-(T-d)/2}d\mathbf{z}\,d\mathbf{w}\\
&\ll \frac{\tilde{C}^T}{Tc_3^{T/2}}\int_1^\infty r^{\frac{d}{2}-3-(T-d)}(\log r)^{T-d}e^{-2(T-d)r+4r}\,dr\\
&+ \frac{\tilde{C}^T}{Tc_3^{T/2}}\int_1^\infty r^{\frac{d}{2}-3-(T-d)}(\log\left(c_3e^{4r}\right))^{T-d}e^{-2(T-d)r+4r}\,dr\\
&+ \frac{\tilde{C}^T}{Tc_3^{T/2}}\int_1^\infty r^{\frac{d}{2}-3-(T-d)}e^{-(T-d)r+4r}\,dr\\
&\ll \frac{1}{T2^T}.
\end{aligned}
$$

Summing up we have

$$
\begin{aligned}
J_2 &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{1\le\|\mathbf{w}\|\le\eta T}\det(B(\mathbf{z},\mathbf{w}))\overline{f}(\mathbf{w},\mathbf{z})^{T-d}e^{-\varepsilon\|\mathbf{w}\|^2-2i\varepsilon\langle\mathbf{w},\mathbf{z}\rangle}\,d\mathbf{w}\,d\mathbf{z}\\
&= J_{21}+J_{22}+J_{23}+J_{24}+J_{25}\\
&= \frac{1}{\sqrt{2\pi T}}\int_{\|\mathbf{w}\|\ge 1}\sqrt{\overline{\det}(B(\mathbf{w}))}e^{-\frac{2\varepsilon^2}{T}\mathbf{w}^\tau\overline{B}(\mathbf{w})^{-1}\mathbf{w}}e^{-\varepsilon\|\mathbf{w}\|^2}\,d\mathbf{w} + O\left(\frac{\varepsilon^{-3d/4}}{T^{(d+2)/2}}\right).
\end{aligned}
$$

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

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

If $\|\mathbf{z}\| \geq e^{4\|\mathbf{w}\|}$ we again use the upper bound (48) and obtain

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

Consequently,

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

D: *The whole range*.

Summing up we arrive at

$$\tilde{J}_0 = \frac{1}{(2\pi T)^{d/2}} \int_{\mathbb{R}^d} \sqrt{\det(\overline{B}(\mathbf{w})} e^{-\varepsilon\|\mathbf{w}\|^2 - \frac{2\varepsilon^2}{T} \mathbf{w}^\tau \overline{B}(\mathbf{w})^{-1}\mathbf{w})} \, d\mathbf{w} + +O\left(\frac{\varepsilon^{-3d/4}}{T^{(d+2)/2}}\right)$$

provided that

$$\varepsilon \gg \max\left(T^{-1/2}, T^{-2/(d+1)}\right)$$

and where we have set

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

for a sufficiently small positive constant c_6.

Appendix F. Extension: General Class $\mathcal{H}_{p,\mathbf{w}}$

We now discuss some possible extensions of Theorem 2 to a larger class of hypothesis class $\mathcal{H}_{p,\mathbf{w}}$ under the assumption that the functions $-\log p(\langle \mathbf{w}|\mathbf{x}\rangle)$ and $-\log(1 - p(\langle \mathbf{w}|\mathbf{x}\rangle))$ are convex with bounded gradient and Hessian.

In the logistic case for all y^T we have $\nabla^2 L(y^T|\mathbf{w}) = \mathbf{B}(\mathbf{w})$. In general, this is not the case. In fact, $\nabla^2 L(y^T|\mathbf{w})$ varies with y^T, and when it is viewed as a random variable we have $\mathbb{E}[Y_t] = 1 - 2p(w)$ and $\mathbb{E}[\nabla^2 L(y^T|\mathbf{w})] = \mathbf{B}(\mathbf{w})$. Indeed,

$$\mathbb{E}[\nabla^2 L(y^T|\mathbf{w})] = \sum_{t=1}^{T} \frac{(p'(\langle \mathbf{w}|\mathbf{x}_t\rangle)^2}{(1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle))p(\langle \mathbf{w}|\mathbf{x}_t\rangle)} = \mathbf{B}(\mathbf{w}, \mathbf{x}^T).$$

We now split $\nabla^2 L(y^T|\mathbf{w})$ as $\nabla^2 L(y^T|\mathbf{w}) = \mathbf{A}(\mathbf{w}) + \mathbf{F}(y^T)$ with $\mathbf{F}(y^T) = \sum_{t=1}^{T} y_t \mathbf{F}_t$ where

$$\mathbf{A}(\mathbf{w}) = \frac{1}{2}\sum_t \left(\frac{(p'(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2}{(p(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2} + \frac{(p'(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2}{(1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2} - p''(\langle \mathbf{w}|\mathbf{x}_t\rangle)\left(\frac{1}{p(\langle \mathbf{w}|\mathbf{x}_t\rangle)} - \frac{1}{1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle)}\right)\right) \mathbf{x}_t \otimes \mathbf{x}_t$$

$$\mathbf{F}_t = \frac{1}{2}\left(\frac{-(p'(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2}{(p(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2} + \frac{(p'(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2}{(1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle))^2} ip''(\langle \mathbf{w}|\mathbf{x}_t\rangle)\left(\frac{1}{p(\langle \mathbf{w}|\mathbf{x}_t\rangle)} + \frac{1}{1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle)}\right)\right) \mathbf{x}_t \otimes \mathbf{x}_t.$$

We also have the following identity $\mathbb{E}[\mathbf{F}(y^T)] = \sum_{t=1}^{T}(1 - 2p(\langle \mathbf{w}|\mathbf{x}_t\rangle))\mathbf{F}_t$.

We now need a large deviation result for $\nabla^2 L(y^T|\mathbf{w})$, that is, $\mathbf{F}(y^T)$ proved in the next lemma.

Lemma 23 *If $\mathbf{F}^T$ is a uniformly bounded sequence for all T, then for any $A > 0$ and $\alpha > 1/2$ there exists B such that*

$$P\left(\|\mathbf{F}(y^T) - \mathbb{E}[\mathbf{F}(y^T)]\| > AdT^\alpha\right) \le 2d^2 e^{-BT^{2\alpha-1}} \tag{49}$$

where $\|.\|$ is an arbitrary metric.

Proof Let $\mathbf{\Theta}$ be a complex $d \times d$ matrix, and let $\widetilde{\mathbf{F}}_T(\mathbf{\Theta})$ be the Laplace transform of $\mathbf{F}(y^T)$, that is, $\widetilde{\mathbf{F}}_T(\mathbf{\Theta}) = \mathbb{E}[e^{\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}(y^T))}]$, where the trace $\mathrm{Tr}(\cdot)$ is the classical expression for the dot product of matrices. We have, when y^T is viewed as a random variable with probability $P(y^T|\mathbf{w})$,

$$\widetilde{\mathbf{F}}_T(\mathbf{\Theta}) = \prod_t \left(p(\langle \mathbf{w}|\mathbf{x}_t\rangle)e^{-\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)} + (1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle))e^{\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)}\right). \tag{50}$$

We will show that there exists a simply connected complex neighborhood $\mathcal{U}$ of the null matrix, such that for all T the following $\mathbf{\Theta} \in \mathcal{U}$ implies that $\log \widetilde{\mathbf{F}}_T(\mathbf{\Theta})$ exists and is uniformly $O(T)$. Indeed by rewriting (50)

$$p(\langle \mathbf{w}|\mathbf{x}_t\rangle)e^{-\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)} + (1-p(\langle \mathbf{w}|\mathbf{x}_t\rangle))e^{\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)} = e^{\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)}\left(1 + p(\langle \mathbf{w}|\mathbf{x}_t\rangle)(e^{-2\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)} - 1)\right)$$

and notice that $\|\mathbf{F}_t\| < F$ for some $F > 0$ (here we take $\|\mathbf{F}_t\| = \sqrt{\mathrm{Tr}(\mathbf{F}_t^2)}$ with the classic matrix dot product expression). Then taking $\|\mathbf{\Theta}\| < \frac{1}{20F}$ we have

$$\left|p(\langle \mathbf{w}|\mathbf{x}_t\rangle)(e^{-2\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)} - 1)\right| < 1/2 \quad \text{and} \quad p(\langle \mathbf{w}|\mathbf{x}_t\rangle)(e^{-2\mathrm{Tr}(\mathbf{\Theta}\mathbf{F}_t)} - 1) \neq 0.$$

Since $\mathcal{U}$ is simply connected, the logarithm of $\widetilde{\mathbf{F}}_T(\mathbf{\Theta})$ exists, and it turns out that the logarithm is always bounded by some $C > 0$, hence the logarithm of the product satisfies: $|\log \widetilde{\mathbf{F}}(\mathbf{\Theta})| \le TC$. As a consequence, $\log \widetilde{\mathbf{F}}(\mathbf{\Theta})$ is an analytic function on $\mathcal{U}$ and its derivatives are also $O(T)$, in particular the second derivative. Since the first derivative of $\log \widetilde{\mathbf{F}}_T(\mathbf{\Theta})$ at its derivatives are also

$O(T)$, in particular the second derivative. Since the first derivative of $\log \widetilde{\mathbf{F}}_T(\mathbf{\Theta})$ at $\mathbf{\Theta} = 0$ is exactly $\mathbb{E}[\mathbf{F}(y^T)]$, we have the following Taylor expansion

$$\log \widetilde{\mathbf{F}}(\mathbf{\Theta}) = \mathrm{Tr}\left(\mathbf{\Theta}\mathbb{E}[\mathbf{F}(y^T)]\right) + O(T\|\mathbf{\Theta}\|^2) \tag{51}$$

with $O(T\|\mathbf{\Theta}\|^2) \leq RT\|\mathbf{\Theta}\|^2$ for some $R > 0$. We will use this estimate via the Chebychev inequality. Having $\|\mathbf{F}(y^T) - \mathbb{E}[\mathbf{F}(y^T)]\| > AdT^\alpha$ implies to have one of d^2 component of $\mathbf{F}(y^T) - \mathbb{E}[\mathbf{F}(y^T)]$ greater than AT^α or smaller than $-AT^\alpha$. For $(i,j) \in \{1,\dots,d\}^2$ let $\mathbf{F}(y^T)_{ij}$ denote the ij component of $\mathbf{F}(y^T)$. We look at $P\left(\mathbf{F}(y^T)_{ij} > \mathbb{E}[\mathbf{F}(y^T)]_{ij} + AT^\alpha\right)$. If $\mathbf{e}_{ij}$ is the matrix with all components equal to zero, except the (i,j)-th element which is equal to 1, then $\mathbf{F}(y^T)_{ij} = \mathrm{Tr}(\mathbf{e}_{ij}\mathbf{F}(y^T))$ and for all $\theta > 0$ by Markov inequality

$$P\left(\mathbf{F}(y^T)_{ij} > \mathbb{E}[\mathbf{F}(y^T)]_{ij} + AT^\alpha\right) \leq \frac{\mathbb{E}[e^{\theta \mathrm{Tr}(\mathbf{e}_{ij}\mathbf{F}(y^T))}]}{\exp(\theta \mathrm{Tr}(\mathbf{e}_{ij}\mathbf{F}(y^T)) + \theta AT^\alpha)}.$$

Since $\mathbb{E}[e^{\theta \mathrm{Tr}(\mathbf{e}_{ij}\mathbf{F}(y^T))}] = \widetilde{\mathbf{F}}(\theta \mathbf{e}_{ij})$, and thanks to the estimate (51) with $\|\mathbf{e}_{ij}\|^2 = 1$, the right-hand side is upper bounded by $\exp\left(RT\theta^2 - \theta AT^\alpha\right)$ with the minimum $\exp\left(-A^2T^{2\alpha-1}/(4R)\right)$. Also with the minimum $\exp\left(-A^2T^{2\alpha-1}/(4R)\right)$. Also $P\left(\mathbf{F}(y^T)_{ij} < \mathbb{E}[\mathbf{F}(y^T)]_{ij} - AT^\alpha\right)$ but with $\theta < 0$. This concludes the proof with $B = A/(4R)$. ■

As a consequence of the above lemma and our previous analysis, we envision the following extension of Theorem 2: Assume the sequences $\mathbf{x}_t$ is generated by a distribution μ over the cube $[-1,1]^d$. Then one should expect

$$\mathbb{E}[S^\varepsilon(\mathbf{x}^T)] = \left(\frac{T}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} \sqrt{\det(\bar{\mathbf{B}}(\mathbf{w}))} e^{-\varepsilon\|\mathbf{w}\|^2 - \frac{2\varepsilon^2}{T}\langle \mathbf{w}|\bar{\mathbf{B}}(\mathbf{w})^{-1}\mathbf{w}\rangle} d\mathbf{w} \times (1 + O(d^2T^{\alpha-1})) \tag{52}$$

with

$$\bar{\mathbf{B}}(\mathbf{w}) = \int_{\mathbb{R}^d} \mu(\mathbf{x}) \frac{p'(\langle \mathbf{w}|\mathbf{x}\rangle)^2}{p(\langle \mathbf{w}|\mathbf{x}\rangle)(1 - p(\langle \mathbf{w}|\mathbf{x}\rangle))} \mathbf{x} \otimes \mathbf{x} d\mathbf{x}.$$

Furthermore it should hold that

$$\bar{r}_T^\varepsilon = \mathbb{E}[\log S(\mathbf{X}^T)] \sim \log \mathbb{E}[S_T^\varepsilon(\mathbf{X}^T)]. \tag{53}$$

To see this, we notice that $\mathbb{E}[\mathbf{F}(y^T)] = \mathbf{B}(\mathbf{w}) - \mathbf{A}(\mathbf{w})$ and by Lemma 23

$$P\left(\|\nabla^2 L(y^T|\mathbf{w}) - \mathbf{B}(\mathbf{w}))\| > AdT^\alpha\right) \leq 2d^2 e^{-BT^{2\alpha-1}}.$$

We have $\nabla^2 L(y^T|\mathbf{w})$ and $\mathbf{B}(\mathbf{w}, \mathbf{x}^T)$ both of order T thus $\det(\nabla^2 L(y^T|\mathbf{w}))$ and $\det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T)$ is of order T^d. When we choose $\alpha < 1$, $\|\nabla^2 L(y^T|\mathbf{w}) - \mathbf{B}(\mathbf{w})\|$ is of order dT^α hence of order smaller than $\nabla^2 L(y^T|\mathbf{w})$. By Jacobi formula for any matrix A we have $\nabla(\det(A)) = \det(A)A^{-1}$, as in Bellman (1997). Thus for another matrix B we can write $\det(\mathbf{B}) = \det(\mathbf{A}) + O(\det(\mathbf{A})\|\mathbf{A}^{-1}\| \cdot \|\mathbf{B} - \mathbf{A}\|)$ or more precisely $O(|\det(\mathbf{A})| + |\det(\mathbf{B})|)$ but $\det(\mathbf{A})$ suffices when $det(\mathbf{A}) \neq 0$ and when $\|\mathbf{A} - \mathbf{B}\| o(\|\mathbf{A}\|)$. We now can write

$$\begin{aligned} \left|\det(\nabla^2 L(y^T|\mathbf{w})) - \det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T)\right| &= \left|\det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T)\right| \cdot \|\mathbf{B}(\mathbf{w}, \mathbf{x}^T)^{-1}\| O\left(\|\nabla^2 L(y^T|\mathbf{w}) - \mathbf{B}(\mathbf{w}, \mathbf{x}^T)\|\right) \\ &= \left|\det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T)\right| \cdot \|\mathbf{B}(\mathbf{w}, \mathbf{x}^T)^{-1}\| O(dAT^\alpha) \\ &= \det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T) O(d^2T^{\alpha-1}) \end{aligned}$$

since $\|\mathbf{B}(\mathbf{w}, \mathbf{x}^T)^{-1}\| = O(d/T)$ (as long as the matrix $\mathbf{B}(\mathbf{w}, \mathbf{x}^T)/T$ is not degenerate). Thus $\det(\nabla^2 L(y^T|\mathbf{w})) = \det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T)(1 + O(d^2 T^{\alpha-1}))$. We call $\mathcal{J}_\alpha$ the set of sequences y^T that satisfies this property. We already know that $\sum_{y^T \notin \mathcal{J}_\alpha} P(y^T|\mathbf{w}) < 2d^2 e^{-BT^{2\alpha-1}}$ that exponentially decays since $\alpha > 1/2$. Therefore with $O(d^2 T^\alpha)$ we can approximate $\det(\nabla^2 L(y^T|\mathbf{x}^T, \mathbf{w}))$ by $\det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T))$ which allows us to write

$$
\begin{aligned}
S(\mathbf{x}^T|\mathbf{w}) &= \frac{1}{(2\pi)^d} \sum_{y^T} \det(\nabla^2 L(y^T|\mathbf{x}^T, \mathbf{w})) \int_{\mathbb{R}^d} \exp\bigl(L(y^T|\mathbf{x}^T, \mathbf{w}) - i\langle \nabla L(y^T|\mathbf{x}^T, \mathbf{w})|\mathbf{z}\rangle\bigr) d\mathbf{z} \\
&= \frac{\det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T))}{(2\pi)^d} \sum_{y^T} \int_{\mathbb{R}^d} \exp\bigl(L(y^T|\mathbf{x}^T, \mathbf{w}) - i\langle \nabla L(y^T|\mathbf{x}^T, \mathbf{w})|\mathbf{z}\rangle\bigr) d\mathbf{z} \left(1 + O(d^2 T^{\alpha-1})\right) \\
&= \frac{\det(\mathbf{B}(\mathbf{w}, \mathbf{x}^T))}{(2\pi)^d} \prod_t f(\mathbf{w}, \mathbf{x}_t, \mathbf{z}) \left((1 + O(d^2 T^{\alpha-1})\right).
\end{aligned}
$$

Since

$$
S^\varepsilon(\mathbf{x}^T) = (2\pi)^{-d} \int S(\mathbf{x}^T|\mathbf{w}) e^{-\varepsilon\|\mathbf{w}\|^2} d\mathbf{w}
$$

we, thus, obtain a similar expression as in the logistic case. Hence, this is a strong indication that (52) holds.