Precise Regularized Minimax Regret with Unbounded Weights

Michael Drmota
TU Wien
Vienna, Austria

Philippe Jacquet
INRIA
Paris, France

Changlong Wu
Purdue University
West Lafayette, IN 47907, USA

Wojciech Szpankowski
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 $\hat{y}_t = g_t(y_t^{\perp 1}, \mathbf{x}_t)$, where
$g_t$ represents the strategy/algorithm of the learner to obtain its prediction based on prior observations $y^{t-1}$ and $x^t$. Once a prediction is made, nature reveals the true label $y_t$, and the learner incurs some loss evaluated based on a predefined function $\ell: \hat{Y} \times Y \rightarrow \mathbb{R}_{\geq 0}$, where $\hat{Y} = [0, 1]$ and $Y = \{-1, 1\}$ 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 $h: \mathbb{R}^d \rightarrow \hat{Y}$. After $T$ rounds, the pointwise regret is defined as

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

where $\hat{y}_t = g_t(y^{t-1}, x^t)$. Throughout, we write $y^t = (y_1, \ldots, y_t)$ and $x^t = (x_1, \ldots, x_t)$ for $t \in [T]$.

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

$$H_{p,w} = \{h_w(x) \equiv p(\langle w, x \rangle) : w, x \in \mathbb{R}^d\},$$

where $w$ is a $d$-dimensional weight vector, $\langle w, x \rangle$ is the scalar product of $x$ and $w$, and $p(w)$ with $w = \langle w, x \rangle$ is a probability function. To ease the presentation, we almost exclusively discuss logistic regression with $p(w) = (1 + \exp(-w))^{-1}$ (Hazaran et al., 2014; Shamir, 2020a) (see Appendix F for possible extensions). While we assume that $x_t$ lies on a compact manifold $M_x$ (e.g., $M_x = [-1, 1]^d$, the unit ball $B_d$, or the sphere $S_d$), we do not bound the weights $w \in \mathbb{R}^d$, and this seems to have never been analyzed in depth, to the best of our knowledge. Specifically, we assume that $\|w\| \leq R < \infty$.

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

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

This notion was also known in the literature as transductive online learning Cesa-Bianchi and Shamir (2011). To decouple it from the feature vector $x^T$, one either maximizes over all possible $x^T$ or takes the average over the features. We study here the averaged fixed design minimax regret as $\bar{r}_T(H) := \mathbb{E}_{x^T} [r_T(H|x^T)]$, where the feature vector $x^T$ is generated by an i.i.d. process.

As discussed in Jacquet et al. (2021); Shamir and Szpankowski (2021) the minimax regret $r_T(x^T)$ can be studied through the so called Shtarkov sum which for bounded $\|w\| \leq R$ becomes

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

where $P(y^T|x^T, w) = \prod_{t=1}^{T} p(y_t|x_t, w)$, and the regret is then $r_T(x^T) = \log S(x^T)$. While the Shtarkov sum approach provides an exact solution, there are two main issues: computational and analytical. The optimization problem $\sup_{\|w\| \leq R} P(y^T|x^T, w)$ is non-convex and, more problematically, most of the optimal solutions $w^* = \arg \sup_{\|w\| \leq R} P(y^T|x^T, w)$ lie on the boundary $\|w\| = R$. 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 $\ell(\hat{y}, y) = -\log P(\hat{y}|x, w)$ and $\ell(h(x), y) = -\log P(h(x)|x, w)$, leading to the regularized pointwise regret

$$R^*_T(\hat{y}^T, y^T|x^T) = -\sum_{t=1}^{T} \log P(\hat{y}_t|x_t, w) + \sup_w \sum_{t=1}^{T} \log P(y_t|x_t, w)e^{-\varepsilon\|w\|^2}$$

(4)
where \( \varepsilon \geq 0 \) and \( \sup_{w} \) is \textit{unconstrained}. Then, the regularized minimax regrets is defined as

\[
\hat{r}_T^\varepsilon(x^T) = \inf_{y^T} \max_{y^T} R_T^\varepsilon(y^T, y^T|x^T), \quad \hat{r}_T^\varepsilon(H) := E_{x^T}[\hat{r}_T^\varepsilon(H|x^T)]
\]  

(5)

and the \textit{generalized} Shtarkov sum is

\[
S_\varepsilon(x^T) = \sum_{y^T} \sup_{w} P(y^T|x^T, w)e^{-\varepsilon||w||^2}.
\]  

(6)

Notice that the optimization \( w^* := \arg\sup_{w \in \mathbb{R}^d} P(y^T|x^T, w)e^{-\varepsilon||w||^2} \) is much easier to compute since it is \textit{log-concave} and \( ||w^*|| < \infty \) holds always. We show in Section 2 that \( r_T^\varepsilon(x^T) = \log S_\varepsilon(x^T) \) 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 already known in information theory Takeuchi and Barron (2006), but only for 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 \textit{finite} and \textit{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 \( (d/2) \log(2T/\pi) + \log C_\varepsilon \) where \( C_\varepsilon \) 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 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. As mention above, Drmota et al. (2024) proved that for \( ||w|| = \Omega(\sqrt{T}) \) 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 \( R = ||w|| \) with \( R \leq 1/\sqrt{\varepsilon} \).

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 \( x^T \), the predictor \( \hat{y}_t \) can be compactly represented as a distribution \( Q \) over \( \{1, -1\}^T \) such that \( \hat{y}_t = Q(-1|y_t-1, x^t) \) and \( \ell(y_t, \hat{y}_t) = -\log Q(y_t|y_t-1, x^t) \). Then the regularized regret can be written as

\[
\begin{align*}
\mathcal{R}_\varepsilon^e(\mathcal{H}|x^T) &= \min_Q \max_{y^T} \left[ -\log Q(y^T|x^T) + \sup_w \log P(y^T|x^T)e^{-\varepsilon||w||^2} \right] \\
&= \min_Q \max_{y^T} \left[ -\log Q(y^T|x^T) + \log P^*(y^T|x^T) + \log \sup_w \log P^*(y^T|x^T, w)e^{-\varepsilon||w||^2} \right] \\
&= \log \sum_{y^T} \sup_w P(y^T|x^T, w)e^{-\varepsilon||w||^2} = \log S_\varepsilon(x^T)
\end{align*}
\]

where \( S_\varepsilon(x^T) \) is defined in (6) and \( P^*_\varepsilon(y^T|x^T) := \frac{\sup_w P(y^T|x^T, w)e^{-\varepsilon||w||^2}}{\sum_w \sup_w P(y^T|x^T, w)e^{-\varepsilon||w||^2}} \) 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 \( S_\varepsilon(x^T) \). Note that for a sequence of labels \( y^T \) and a sequence of features \( x^T \) we have for any \( \varepsilon > 0 \)

\[
P(y^T|x^T, w) = \prod_{t=1}^T p(y_t|x_t, w) \quad \text{and} \quad P^*_\varepsilon(y^T|x^T, w) = \prod_{t=1}^T p(y_t|x_t, w)e^{-\varepsilon||w||^2}.
\]

We also define \( L(y^T|x^T, w) = \log P(y^T|x^T, w) \) and \( L^*_\varepsilon(y^T|x^T, w) = L(y^T|x^T, w) + \varepsilon||w||^2 \).

---

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

2. Following Handmaid’s perpect: “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 \( \nabla L_\varepsilon(y^T|x^T, w) = -\sum_{t=1}^{T} p(-y_t(x_t, w)) y_t x_t + 2\varepsilon w \) and

\[
\nabla^2 L_\varepsilon(y^T|x^T, w) = \sum_{t=1}^{T} p((x_t, w))p(-\langle x_t, w \rangle) x_t \otimes x_t + 2\varepsilon I,  
\]

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

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

\[
\nabla L_\varepsilon(y^T|x^T, w) = a
\]

has a unique solution \( w_\varepsilon^*(a) \). Furthermore, if we denote \( G_{y^T|x^T, \varepsilon}(w) := \nabla L_\varepsilon(y^T|x^T, w) \), then we have \( w_\varepsilon^* = G_{y^T|x^T, \varepsilon}^{-1}(0) \).

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 \( h_{y^T|x^T}(a) = \exp\left(-\langle L_\varepsilon(y^T|x^T, G_{y^T|x^T, \varepsilon}^{-1}(a)) \rangle \right) \).

The goal is to show that (for every \( y^T \) and \( x^T \)) the Fourier transform

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

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

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

Observe that by (10) (for \( w = G_{y^T|x^T, \varepsilon}^{-1}(a) \)) we have \( a = -\sum_{t=1}^{T} p(-y_t(x_t, w)) y_t x_t + 2\varepsilon w = O(1) + 2\varepsilon |w| \) Note that the \( O(1) \)-term depends on \( y^T, x^T \).

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

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

Hence the Fourier transform \( \tilde{h}_{y^T|x^T}(z) \) 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 \( y^T, x^T, \varepsilon > 0 \) and for all non-negative integers \( k_1, \ldots, k_d \) we have

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

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

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

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

\[
h_{y^T|x^T}(a) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{h}_{y^T|x^T}(z)e^{i\langle a, z \rangle} \, dz.
\]
Granted the existence of the Fourier \( h_{y^T|X^T}(z) \) we can proceed to evaluate the Shtarkov sum. We first observe that

\[
\sup_{y^T \in \mathbb{R}^d} P_{\varepsilon}(y^T|X^T, w) = P_{\varepsilon}\left(y^T|X^T, G_{y^T | X^T, \varepsilon}(0)\right) = \exp\left(-L(y^T|X^T, G_{y^T | X^T, \varepsilon}(0))\right)
\]

\[
= (2\pi)^{-d} \int_{\mathbb{R}^d} h_{y^T|X^T}(z) \, dz = (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h_{y^T|X^T}(a)e^{-i\langle a, z \rangle} \, da \, dz
\]

\[
= (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \exp\left(-L_{\varepsilon}(y^T|X^T, G_{y^T | X^T, \varepsilon}(a))\right)e^{-i\langle a, z \rangle} \, da \, dz
\]

\[
= (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} P(y^T|X^T, w)e^{-i\langle \nabla L(y^T|X^T, w), z \rangle} e^{-\varepsilon||w||^2-2i\varepsilon\langle w, z \rangle} \det\left(\nabla^2 L_{\varepsilon}(y^T|X^T, w)\right) \, dw \, dz
\]

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

\[
\sum_{y^T \in \{-1,1\}^T} P(y^T|X^T, w) \exp\left(-i\langle \nabla L(y^T|X^T, w), z \rangle\right) = \prod_{t=1}^{T} f(w, x_t, z),
\]

where \( f(w, x_t, z) \) denotes

\[
f(w, x_t, z) = p((x, w))e^{-i\langle w(x, w) \rangle \langle x, z \rangle} + p(-((x, w)))e^{i\langle w(x, w) \rangle \langle x, z \rangle}.
\]

This leads to our integral representation of the Shtarkov sum

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

We now assume that the vectors \( x_1, \ldots, x_T \) are iid random vectors \( X_1, \ldots, X_T \) 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

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

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 multidimensional integrals for various ranges of \( w \) and \( z \) 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 \( w \in \mathbb{R}^d \) and \( \varepsilon > 0 \). Assume features \( x_t \) are generated by a uniform distribution over the \( d \)-dimensional ball \( B_d \) 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

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

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.
Then we have the following expression

\[ \bar{\log} \]

where we conjecture the leading term at \( \epsilon \) we observe that our proof Theorem 2 works only for

\( (ii) \) Furthermore,

\[ \bar{r}_T = E \left[ \log S^\epsilon(X^T) \right] \leq \log E[S^\epsilon(X^T)] = \frac{d}{2} \log T + \log C_d(\epsilon) + O(T^{-\beta(d)}) \]

where

\[ C_d(\epsilon) = \int_{\mathbb{R}^d} \sqrt{\det(B(w))} e^{-\epsilon \|w\|^2} dw. \]

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

\[ E \left[ \log S^\epsilon(X^T) \right] \sim \log E[S^\epsilon(X^T)]. \]

Another question is whether from Theorem 2 we can recover the original regret with \( \epsilon = 0 \). First, we observe that our proof Theorem 2 works only for \( \epsilon \gg 1/\sqrt{T} \) which basically translates to the radius \( R \) of \( w \) 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 \( \epsilon \to 0 \) we need some asymptotic information about \( \bar{B}(w) \). We recall that \( X \) is uniformly distributed in the \( d \)-dimensional ball \( B_d \) and that \( \bar{B}(w) = \Theta(1, |w|^{-3}) \) for \( d = 1 \). The proof of the following lemma is given in Appendix C.

**Lemma 3** (i) Suppose that \( d \geq 2 \) and let \( u = w/\|w\| \) and set \( q(x) = p(x)p(-x) = p(x)(1-p(x)) \). Then we have the following expression \( \bar{B}(w) = \phi(w)(I_d - u \otimes u) + \lambda(w)u \otimes u \) where \( I_d \) is the identity operator on \( \mathbb{R}^d \) (thus \( I_d - u \otimes u \) is the identity operator on the hyperplane orthogonal to \( u \)) with

\[ \lambda(w) = \frac{ds_d-1}{s_d} \int_0^1 \int_0^\pi t^{d-1} \cos(\theta)^2 \sin(\theta)^{d-2} q(t \cos(\theta) \|w\|) d\theta dt, \]  

and

\[ \phi(w) = \frac{ds_d-1}{s_d} \int_0^1 \int_0^\pi t^{d-1} \sin(\theta)^{d-1} q(t \cos(\theta) \|w\|) d\theta dt \]

are the eigenvalues of \( \bar{B}(w) \) with multiplicity 1 and \( d-1 \), respectively, and \( s_d \) denotes the area of the unit sphere \( S_d \).

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

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

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

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

$$
\int_{\mathbb{R}^d} \sqrt{\det B(w)} e^{-\varepsilon \|w\|^2} dw \approx \int_{\|w\| \geq 1} \|w\|^{-(d+2)/2} e^{-\varepsilon \|w\|^2} dw \approx \int_1^\infty r^{d-1} r^{-(d+2)/2} e^{-\varepsilon r^2} dr
$$

$$
\approx \varepsilon^{-d/4+1/2} \int_0^\infty u^{d-1} u^{-(d+2)/2} e^{-u^2} du \approx \varepsilon^{-d/4+1/2}
$$

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 $\bar{f}(w, z) = \mathbb{E} f(w, X, z)$. For the case $d = 1$ we have

$$
\bar{f}(w, z) = \mathbb{E} f(w, X, z) = \frac{1}{2} \int_{-1}^1 \left( e^{iwx/(1+e^{iz})} + e^{iwx/(1+e^{-iz})} \right) dx
$$

(19)

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

$$
\mathbb{E} S_\varepsilon (\mathbf{x}^T) = \frac{1}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{E} \left[ \prod_{t=1}^T f(w, X_t, z) \nabla^2 L_\varepsilon (\cdot | X^T, w) \right] e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz
$$

$$
\overset{(a)}{=} \sum_{s=1}^T \frac{1}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{E} \left[ \prod_{t=1}^T f(w, X_t, z) p(x_s w)p(-x_s w) x_s^2 \right] e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz
$$

$$
+ \frac{2\varepsilon}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{E} \left[ \prod_{t=1}^T f(w, X_t, z) \right] e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz
$$

$$
= \frac{T}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty \bar{f}(w, z)^{T-1} B(z, w) e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz + \frac{2\varepsilon}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty \bar{f}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon w z} dw dz,
$$

(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 $\bar{f}(w, z)$ with proofs presented in Appendix B.

---

\(^4\) If requested, we will be happy to provide our notes with detailed derivations.
The case A: The main challenge in the computation of the integral(s) Thus, the integral in Proposition 8 can be replaced by the integral

\[ \int_{-1}^{1} \frac{x^2}{(1 + e^{-xw})(1 + e^{xw})} \, dx. \] (21)

Then we uniformly have \( \overline{B}(w) = \Theta \left( \min(1, |w|^{-3}) \right) \) and for \( c_1 > 0 \)
\[ J(w, z) = 1 - \Theta \left( z^2 \min(1, |w|^{-3}) \right) \quad \text{(for } |z| \leq \max(1, c_1|w|)\text{)} \]
\[ = 1 - \frac{z^2}{2} \overline{B}(w) + O \left( z^3 \min(1, |w|^{-4}) \right) \]
\[ = e^{-\frac{1}{2}z^2\overline{B}(w)} \left( 1 + O \left( z^3 \min(1, |w|^{-4}) \right) + O \left( z^4 \min(1, |w|^{-6}) \right) \right). \]

Lemma 6 If \( |w| \leq 1 \) we uniformly have \( |J(z, w)| \leq \min \left( 1, \frac{C_1}{|w|} \right) \) and for \( |w| \geq 1 \) we have
\[ |J(z, w)| \leq \min \left( 1, \frac{C_1}{\sqrt{|wz|}} + \frac{C_2e^{1/|w|}}{|wz|} \right). \]

Lemma 7 Suppose that \( c_1 > 0 \) is a given constant. Then there exist \( c_2 > 0 \) such that \( |J(z, w)| \leq 1 - \frac{C_2}{|w|} \) uniformly for \(|z| \geq c_1|w|\).

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
\[ \mathbb{E}S_{\varepsilon}(X^T) = \frac{T^{1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \overline{B}(w)^{1/2} e^{-\varepsilon w^2 - 2\varepsilon w^2/(T\overline{B}(w))} \, dw + O \left( \log T \right) \]
provided that \( \varepsilon \gg T^{-1/2} \), where \( \overline{B}(w) \) is given in (21).

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

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

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 \( |w| \leq C_3 \). If \( |w| \leq C_3 \) then we have the uniform bound \( |J(z, w)| \leq C/|z| \). We first look at the case, where \(|z| \leq 2C\). Here we certainly have the uniform representation (see Lemma 5)
\( \overline{f}(z, w) = e^{-\frac{z^2}{T B(w)}} (1 + O(z^3)) \) and by continuity \( |\overline{f}(z, w)| \leq e^{-c_1 z^2} \) for some constant \( c_1 > 0 \). If \( |z| \geq 2C \) then we have the trivial estimate \( |\overline{f}(z, w)| \leq C/|z| \leq 1/2 \) (see Lemma 6). Consequently,

\[
I_1 = \frac{1}{2\pi} \int_{\infty}^{\infty} \int_{-\infty}^{\infty} \overline{f}(w, z) T e^{-\varepsilon w^2 - 2i\varepsilon wz} \, dw \, dz
\]

\[
= \frac{1}{2\pi} \int_{-C_3}^{C_3} \int_{|z| \leq T^{-1/3}} e^{-\frac{T B(w)}{\pi T} \left(1 + O(T z^3)\right)} e^{-\varepsilon w^2 - 2i\varepsilon wz} \, dw \, dz
\]

\[
+ \frac{1}{2\pi} \int_{-C_3}^{C_3} \int_{T^{-1/3} \leq |z| \leq 2C} O(e^{-c_1 T z^3}) e^{-\varepsilon w^2} \, dw \, dz + \frac{1}{2\pi} \int_{-C_3}^{C_3} \int_{|z| \geq 2C} (C/|z|) T e^{-\varepsilon w^2} \, dw \, dz
\]

\[
= \frac{1}{\sqrt{2\pi T}} \int_{-C_3}^{C_3} B(w)^{-1/2} e^{-\varepsilon w^2} \, dw + O(T^{-1}).
\]

**B: The case \( C_3 \leq |w| \leq \eta T \).** Next we consider the range \( C_3 \leq |w| \leq \eta T \), 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 \( |z| \leq z_1 = |w|^{3/2} T^{-1/2} T^{-1/6} \), where we use Lemma 6.

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

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

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

for some \( c > 0 \). Similarly we obtain

\[
\int_{|z| \leq z_1} e^{-T B(w)^{2 \over T} T |z^3|/w^4} \, dz \ll \frac{|w|^{1/2}}{T}, \int_{|z| \leq z_1} e^{-T B(w)^{2 \over T} T |z^4|/w^6} \, dz \ll \frac{|w|^{3/2}}{T^{3/2}}.
\]

Summing up we find

\[
I_{21} = \frac{1}{2\pi} \int_{|z| \leq z_1} \int_{C_3 \leq |w| \leq \eta T} \overline{f}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} \, dw \, dz
\]

\[
= \frac{1}{\sqrt{2\pi T}} \int_{|w| \geq C_3} B(w)^{-1/2} e^{-2\varepsilon w^2/(T B(w))} e^{-\varepsilon w^2} \, dw + O \left(e^{-5/4 T^{3/2} \eta T^2} + e^{-3/4} T + e^{-5/4} T^2\right).
\]

Next suppose that \( z_1 = |w|^{3/2} T^{-1/2} \eta^{-1/6} \leq |z| \leq c_1 |w| \), where \( c_1 \) sufficiently small which ensures that (see Lemma 5)

\[
B(w)^{2 \over 2} \geq C \left(\frac{|z|^3}{|w|^4} + \frac{|z|^4}{|w|^6}\right).
\]

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

\[ I_{22} = \frac{1}{2\pi} \int_{|z| \leq c_1|w|} \int_{|w| \leq \eta T} \mathcal{F}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} dw \, dz \]

\[ \ll \frac{\eta^{1/6} e^{-c_3^{-1/3}}}{T^{1/2}} \int_{|z| \leq |w| \leq \eta T} w^{3/2} e^{-\varepsilon w^2} dw \ll \eta^{1/6} e^{-c_3^{-1/3}} \frac{\varepsilon^{5/4}}{e^{5/4}T^{1/2}}. \]

Next suppose that \( c_1|w| \leq |z| \leq c_3 e^{\varepsilon |w|} \) for an arbitrary constant \( c_3 > 0 \). Here we have the upper bound \(|\mathcal{F}(z, w)| \leq 1 - c_2/|w|\) (see Lemma 7) and provided that \( \varepsilon \gg T^{-1/2} \) we have

\[ I_{23} = \frac{1}{2\pi} \int_{c_1|w| \leq |z| \leq c_3 e^{\varepsilon |w|}} \int_{|w| \leq |z| \leq \eta T} \mathcal{F}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} dw \, dz \ll \int_{|z| \leq |w| \leq \eta T} e^{-c_3 T/|w| + |w| - \varepsilon w^2} dw \]

\[ \leq \int_{|w| \leq |z| \leq c_2 \sqrt{T}} e^{-(1-c_2)\sqrt{T} - \varepsilon w^2} dw + \int_{c_2 \sqrt{T} \leq |w| \leq \eta T} e^{-\varepsilon w^2} dw \]

\[ \ll \varepsilon^{-1/2} e^{-(1-c_2)\sqrt{T}} + e^{-\varepsilon c_2^2 T} \ll e^{-c_3 \sqrt{T}} \ll \frac{1}{T}. \]

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

\[ \mathcal{F}(z, w) \leq \tilde{C} \max \left( \frac{1}{|w z|^{3/2}}, \frac{e^{\varepsilon |w|}}{|z| e^{\varepsilon |w|}} \right) \leq \frac{1}{2}. \]

If \( c_3 e^{\varepsilon |w|} \leq |z| \leq e^{2|w|}/|w| \) then the first term \( 1/\sqrt{|w z|} \) dominates and we find

\[ I_{24} = \frac{1}{2\pi} \int_{c_3 e^{\varepsilon |w|} \leq |z| \leq e^{2|w|}/|w|} \int_{|w| \leq \eta T} \mathcal{F}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} dw \, dz \ll \frac{1}{T}. \]

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

\[ I_{25} = \frac{1}{2\pi} \int_{|z| \geq e^{2|w|}/|w|} \int_{1 \leq |w| \leq \eta T} \mathcal{F}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} dw \, dz \ll \frac{1}{T}. \]

Summing up we have

\[ I_2 = \frac{1}{2\pi} \int_{\varepsilon^{-3/4}} \int_{c_3 \leq |w| \leq \eta T} \mathcal{F}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} dw \, dz = I_{21} + I_{22} + I_{23} + I_{24} + I_{25} \]

\[ = \frac{1}{\sqrt{2\pi T}} \int_{|w| \geq c_3} B(w)^{-1/2} e^{-\varepsilon w^2 - 2e^{2\varepsilon^2 w^2}/(T B(w))} dw \]

\[ + O \left( \frac{e^{-3/4}}{T} + \frac{e^{-5/4}}{T^2} + \frac{\varepsilon^{1/6} e^{-c_3^{-1/3}}}{e^{5/4}T^{1/2}} + e^{-5/4 T^{3/2} \varepsilon^{-5/2} T^2} \right). \]

C: The case \(|w| \geq \eta T\). If \(|z| \leq e^{2|w|}/|w|\), then \(|\mathcal{F}(z, w)| \leq 1\) and obtain for \( \varepsilon \gg T^{-1/2} \)

\[ I_{31} = \frac{1}{2\pi} \int_{|z| \leq e^{2|w|}/|w|} \int_{|w| \geq \eta T} \mathcal{F}(w, z)^T e^{-\varepsilon w^2 - 2i\varepsilon wz} dw \, dz \]

\[ \ll \int_{|w| \geq \eta T} e^{2|w| - \varepsilon w^2} dw \ll \int_{|w| \geq \eta T} e^{-\frac{\varepsilon}{2} w^2} dw \ll \frac{1}{\varepsilon \eta T} e^{-\frac{\varepsilon}{2} (\eta T)^2} \ll e^{-\eta^2 T}. \]
If $|z| \geq e^{|w|}/|w|$ we have $|\mathcal{F}(z, w)| \leq \tilde{C}e^{|w|}/|w|$ and, thus,

$$I_{32} = \frac{1}{2\pi} \int_{|z| \geq e^{|w|}/|w|} \int_{|w| \geq \eta T} \mathcal{F}(w, z) T e^{-\varepsilon u^2 - 2\varepsilon w z} \, dw \, dz \ll \frac{1}{\varepsilon \eta T^2} e^{-\frac{\varepsilon}{2}(\eta T)^2} \ll e^{-\eta^2 T}.$$

Consequently,

$$I_3 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{|w| \geq \eta T} \mathcal{F}(w, z) T e^{-\varepsilon u^2 - 2\varepsilon w z} \, dw \, dz \ll e^{-\eta^2 T}.$$

In conclusion, we arrive at

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

provided that $\varepsilon \gg T^{-1/2}$ and where $\eta = \eta(T) = c_0(\log T)^{-3}$ for a sufficiently small positive constant $c_0$.

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:

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

provided that $\varepsilon \gg T^{-1/2}$ and where we have set $\eta = \eta(T) = c_0(\log T)^{-3}$ for a sufficiently small positive constant $c_0$. 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 $\mathbb{E} \left[ \log S_\varepsilon(X_T) \right] \leq \log \mathbb{E} \left[ S_\varepsilon(X_T) \right]$ 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

$$S_{\varepsilon,1}(x_T) = \frac{1}{2\pi} \int_{|z| \leq 1} \int_{|w| \leq T^{1/4}} \prod_{t=1}^{T} f(w, x_t, z) \mathbb{V}^2 L_\varepsilon(-|x_T, w)|e^{-\varepsilon u^2 - 2\varepsilon w z} \, dw \, dz,$$

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 $|S_{\varepsilon,2}(X_T)/S_{\varepsilon,1}(X_T)| \leq y$. Then we certainly have

$$\mathbb{E} \left[ \log S_\varepsilon(X_T) \right] = \mathbb{E} \log |S_\varepsilon(X_T)| \geq \mathbb{E} \left[ 1_{A(y)} \cdot \log |S_{\varepsilon,1}(X_T) + S_{\varepsilon,2}(X_T)| \right] \geq \mathbb{E} \left[ 1_{A(y)} \cdot \log \left( |S_{\varepsilon,1}(X_T)| (1 - y) \right) \right] \geq \mathbb{E} \left[ 1_{A(y)} \cdot \log |S_{\varepsilon,1}(X_T)| \right] - 2y = \mathbb{E} \log |S_{\varepsilon,1}(X_T)| - 2y - \mathbb{E} \left[ 1_{A(y)c} \cdot \log |S_{\varepsilon,1}(X_T)| \right] \geq \mathbb{E} \log |S_{\varepsilon,1}(X_T)| - 2y - \mathbb{P}(A(y)) \sqrt{\mathbb{E} \log^2 |S_{\varepsilon,1}(X_T)|}.$$
Thus, in order to make this estimate work we need an upper bound for
\[
\mathbb{P}(A(y)^c) = \mathbb{P}\left( \frac{S_{\varepsilon,2}(X^T)}{S_{\varepsilon,1}(X^T)} > y \right) \leq \frac{1}{y} \mathbb{E}\left[ \frac{S_{\varepsilon,2}(X^T)}{S_{\varepsilon,1}(X^T)} \right].
\]

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 \(|y| \geq 1\) and satisfies the following growth property: \(|F(y)| \leq C_1|y|^{D_1}\) for positive constants \(C_1, D_1\) and that \(y_1 \sim y_2 \rightarrow \infty\) implies \(F(y_1) \sim F(y_2)\). Then we have
\[
\mathbb{E}[F(S_{\varepsilon,1}(X^T))] \sim F(\mathbb{E}S_{\varepsilon,1}(X^T)), \quad \text{and} \quad \mathbb{E}S_{\varepsilon,1}(X^T) \sim \frac{T^{1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{B(w)} e^{-w^2} dw.
\]

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

**Lemma 10** We have uniformly for \(|z| \leq 1, x \in [0,1]\) 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 \(\mathbb{E}[S_{\varepsilon,2}(X^T)] < \log T\), however the above methods seem to fail for getting upper bounds for \(\mathbb{E}[|S_{\varepsilon,2}(X^T)|]\).

**References**


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 \( \nabla^2 L_\varepsilon(y^T|x^T, w) \) satisfies  
\[
\text{det} \left( \nabla^2 L_\varepsilon(y^T|x^T, w) \right) \geq (2\varepsilon)^d.
\]

**Proof**  
By (9) we have for every vector \( v \)
\[
\langle \nabla^2 L_\varepsilon(y^T|x^T, w)v, v \rangle = \sum_{t=1}^{T} p((x_t, w))p(-\langle x_t, w \rangle)\langle x_t, v \rangle^2 + 2\varepsilon\|v\|^2 
\geq 2\varepsilon\|v\|^2.
\]
In particular if \( v \) is an eigenvector of \( \nabla^2 L_\varepsilon(y^T|x^T, w) \) with eigenvalue \( \lambda \) then
\[
\lambda\|v\|^2 \geq 2\varepsilon\|v\|^2.
\]
Since \( \nabla^2 L_\varepsilon(y^T|x^T, w) \) 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 \( h_y^T|x^T(a) \) are absolutely integrable which is the first part of Lemma 1 which we repeat below.

**Lemma 12**  
For every fixed \( y^T, x^T, \varepsilon > 0 \) and for all non-negative integers \( k_1, \ldots, k_d \) we have
\[
\frac{\partial^{k_1+\cdots+k_d}}{\partial a_1^{k_1} \cdots \partial a_d^{k_d}} h_y^T|x^T(a) = O \left( e^{-\frac{1}{4}\|a\|^2\|a\|^{k_1+\cdots+k_d}} \right).
\]

**Proof**  
By (11) the case \( k_1 = \cdots = k_d = 0 \) is already covered. Next let us consider the first derivatives of \( e^{-L(y^T|x^T,w)} \):
\[
\nabla \exp \left( -L(y^T|x^T,G^{-1}_{y^T|x^T,\varepsilon}(a)) \right) = -\exp \left( -L(y^T|x^T,G^{-1}_{y^T|x^T,\varepsilon}(a)) \right) \nabla L(y^T|x^T,G^{-1}_{y^T|x^T,\varepsilon}(a)) \nabla G^{-1}_{y^T|x^T,\varepsilon}(a).
\]
Clearly we have
\[
\exp \left( -L(y^T|x^T,G^{-1}_{y^T|x^T,\varepsilon}(a)) \right) \leq 1.
\]
Since
\[
\nabla L(y^T|x^T,w) = -\sum_{t=1}^{T} p(-y_t(x_t, w)) y_t x_t = O(1)
\]
uniformly for all \( w \in \mathbb{R}^d \) we also have
\[
\nabla L(y^T|x^T,G^{-1}_{y^T|x^T,\varepsilon}(a)) = O(1).
\]
Finally
\[
\nabla G^{-1}_{y^T|x^T,\varepsilon}(a) = \left( \nabla^2 L(y^T|x^T,G^{-1}_{y^T|x^T,\varepsilon}(a)) \right)^{-1}.
\]
All entries of the matrix $\nabla^2 L(y^T|x^T, w)$ are uniformly bounded as well as the reciprocal of its determinant (see Lemma 11). This proves that

$$\nabla G_{y^T|x^T, \epsilon}(a) = O(1)$$

and hence

$$\nabla \exp \left(-L(y^T|x^T, G_{y^T|x^T, \epsilon}(a))\right)$$

$$= \exp \left(-\varepsilon \|G_{y^T|x^T, \epsilon}(a)\|^2\right) \nabla \exp \left(-L(y^T|x^T, G_{y^T|x^T, \epsilon}(a))\right)$$

$$- \exp \left(-\varepsilon \|G_{y^T|x^T, \epsilon}(a)\|^2 - L(y^T|x^T, G_{y^T|x^T, \epsilon}(a))\right) 2\varepsilon G_{y^T|x^T, \epsilon}(a) \nabla G_{y^T|x^T, \epsilon}(a)$$

$$= O \left(e^{-\frac{1}{4\varepsilon} \|a\|^2 \|a\|^k_1 + \cdots + k_d}\right)$$

as proposed.

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

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

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

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

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 $y^T, x^T$, and $\epsilon > 0$ the Fourier transform

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

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

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

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

$$h_{y^T|x^T}(a) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{h}_{y^T|x^T}(z)e^{i(a,z)} \, dz.$$
and the latter integral exists by Lemma 12 it follows that the integral
\[ \int_{\mathbb{R}^d} \tilde{h}_{y^T|\mathbf{x}_T}(\mathbf{z})e^{i(a^T\mathbf{z})} \, d\mathbf{z} \]
exists. Since all involved functions are continuously differentiable this integral equals (up to the factor \((2\pi)^{-d}\)) the original function \(h_{y^T|\mathbf{x}_T}(\mathbf{a})\).

\[ \square \]

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
\[ \mathcal{F}(w, z) = \int_0^1 \left( \frac{e^{-ixz/(1+e^{xw})}}{1+e^{-xw}} + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \right) \, dx. \]

By differentiation we directly obtain
\[ \mathcal{F}(w, 0) = \int_0^1 \left( \frac{1}{1+e^{-xw}} + \frac{1}{1+e^{xw}} \right) \, dx = \int_0^1 1 \, dx = 1, \]
\[ \frac{\partial \mathcal{F}}{\partial z}(w, z) = \int_0^1 \left( \frac{e^{-ixz/(1+e^{xw})}}{1+e^{-xw}} \frac{-ix}{1+e^{xw}} + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \frac{ix}{1+e^{-xw}} \right) \, dx, \]
\[ \frac{\partial \mathcal{F}}{\partial z}(w, 0) = \int_0^1 0 \, dx = 0, \]
\[ \frac{\partial^2 \mathcal{F}}{\partial z^2}(w, z) = \int_0^1 \left( \frac{e^{-ixz/(1+e^{xw})}}{1+e^{-xw}} \left( \frac{-ix}{1+e^{xw}} \right)^2 + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \left( \frac{ix}{1+e^{-xw}} \right)^2 \right) \, dx, \]
\[ \frac{\partial^2 \mathcal{F}}{\partial z^2}(w, 0) = \int_0^1 \left( \frac{1}{1+e^{-xw}} \left( \frac{-ix}{1+e^{xw}} \right)^2 + \frac{1}{1+e^{xw}} \left( \frac{ix}{1+e^{-xw}} \right)^2 \right) \, dx, \]
\[ = -\int_0^1 \frac{x^2}{(1+e^{-xw})(1+e^{xw})} \left( \frac{1}{1+e^{-xw}} + \frac{1}{1+e^{xw}} \right) \, dx \]
\[ = -\int_0^1 \frac{x^2}{(1+e^{-xw})(1+e^{xw})} \, dx = -\mathcal{B}(w), \]
\[ \frac{\partial^3 \mathcal{F}}{\partial z^3}(w, z) = \int_0^1 \left( \frac{e^{-ixz/(1+e^{xw})}}{1+e^{-xw}} \left( \frac{-ix}{1+e^{xw}} \right)^3 + \frac{e^{ixz/(1+e^{-xw})}}{1+e^{xw}} \left( \frac{ix}{1+e^{-xw}} \right)^3 \right) \, dx. \]

Clearly we have
\[ \mathcal{B}(w) = \Theta \left( \int_0^1 x^2 e^{-|xw|} \, dx \right) \]
\[ = \Theta \left( \frac{1}{|w|^2} \int_0^{|w|} v^3 e^{-v} \, dv \right) \]
\[ = \Theta \left( \min(1, |w|^{-3}) \right) . \]
Similarly it follows that
\[ \frac{\partial^3}{\partial z^3} \mathcal{F}(w, z) \ll \int_0^1 x^2 e^{-|xw|} dx \ll \min(1, |w|^{-4}). \]
Thus, it immediately follows that
\[ \mathcal{F}(z, w) = 1 - \frac{z^2}{2} \mathcal{B}(w) + O \left( z^3 \min(1, |w|^{-4}) \right). \]
and by expanding \( \mathcal{F}(z, w) = e^{\log \mathcal{F}(z, w)} \) we obtain the third representation for \( \mathcal{F}(z, w) \) (where we use \( z^2 \min(1, |w|^{-6}) \) as the order of \( z^2 \mathcal{B}(w)^2 \)).

Finally, if \( |z| \leq \max(1, c_1 |w|) \) (for some sufficiently small constant \( c_1 > 0 \)) it follows that
\[ z^2 \mathcal{B}(w) \geq z^3 \min(1, |w|^{-4}) \]
Thus we also get
\[ \mathcal{F}(w, z) = 1 - \Theta \left( z^2 \min(1, |w|^{-3}) \right). \]

**Lemma 14** Let \( \beta_1, \beta \) 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
\[ \left| \int_{\beta_1}^{\beta_2} g(x) e^{ih(x)} dx \right| \leq \frac{2 \max_{[\beta_1, \beta_2]} |g| + V_{\beta_1}^{\beta_2}(g)}{\min_{[\beta_1, \beta_2]} |h'|}, \tag{23} \]
where \( V_{\beta_1}^{\beta_2}(g) \) denotes the total variation of \( g \) on \( [\beta_1, \beta_2] \).

**Lemma 15** Let \( \beta_1, \beta \) 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
\[ \left| \int_{\beta_1}^{\beta_2} g(x) e^{ih(x)} dx \right| \leq \frac{8 \max_{[\beta_1, \beta_2]} |g| + V_{\beta_1}^{\beta_2}(g)}{\min_{[\beta_1, \beta_2]} \sqrt{|h''|}}. \tag{24} \]

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

For the proof of Lemma 6 we need to further properties (that can be found in Huxley (1996)).
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 \( w \geq 0 \) (so that \( u = xw \geq 0 \)). Here we certainly have
\[
|h'(x)| \geq \frac{|z|}{4} \tag{25}
\]
and that \( h''(x) \) has the same sign as \( z \). Hence, by a direct application of Lemma 14 we obtain
\[
\left| \int_0^1 p(-xw)e^{ih(x)} \, dx \right| \leq \frac{8}{|z|} \tag{26}
\]
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
\[
2(1 + e^{-u}) + u(1 - e^{-u}) \geq 2 - e^{-1} > 0
\]
for \( u \leq 0 \). Thus, if \( 0 \leq w \leq 1 \) and \( 0 \leq x \leq 1 \) we have
\[
|h'(-x)| \geq \frac{|z|}{(1 + e)^2} \tag{27}
\]
and consequently we get
\[
\left| \int_0^1 p(xw)e^{ih(-x)} \, dx \right| \leq \frac{2(1 + e)^2}{|z|}
\]
which implies
\[
\overline{f}(z, w) \ll \frac{1}{|z|}.
\]
Trivially we have \( |\overline{f}(z, w)| \leq 1 \). The case \( -1 \leq w < 0 \) can be handled in completely the same way. Thus, we have completed the case \( |w| \leq 1 \).

If \( |w| \geq 1 \) 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 \( 0 \leq x \leq 1/w \) then we get again (27) and, thus,
\[
\left| \int_0^{1/w} p(xw)e^{ih(-x)} \, dx \right| \leq \frac{2(1 + e)^2}{|z|}
\]
Secondly we consider the interval \( 1/w \leq x \leq (|u_0| + \kappa)/w \) (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 \( (|u_0| + \kappa)/w \leq x \leq 1 \) we again apply Lemma 14. In this range we have

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

which gives

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

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 \( |x| \geq c_1' |w| \), where \( c_1' \) will be chosen sufficiently large. As in the proof of Lemma 5 it follows (if \( w \geq 0 \))

\[
\left| \int_{0}^{1} p(-xw)e^{ih(x)} \, dx \right| \ll \frac{1}{|z|}
\]

and

\[
\left| \int_{0}^{1} p(xw)e^{ih(-x)} \, dx \right| = \left| \left( \int_{0}^{1/w} + \int_{1/w}^{(|u_0|+\eta)/w} + \int_{(|u_0|+\eta)/w}^{1} \right) p(xw)e^{ih(-x)} \, dx \right|
\]

\[
\leq 1 - \frac{|u_0| + \eta}{w} + O \left( \frac{1}{|z|^w} + \frac{1}{\sqrt{|z|^w}} \right).
\]

Thus, if \( |z| \geq c_1' w \) for a sufficiently large constant \( c_1' \) we have

\[
\left| \int_{0}^{1} \left( p(-xw)e^{ih(x)} + p(xw)e^{ih(-x)} \right) \, dx \right| \leq 1 - \frac{c_2}{w}.
\]

Next we consider the interval \( c_1 w \leq |z| \leq c_1' w \). With \( c = z/w \) we have

\[
\overline{f}(z, w) = \int_{0}^{1} \left( p(-xw)e^{-ixzp(xw)} + p(xw)e^{ixzp(-xw)} \right) \, dx
\]

\[
= \frac{1}{w} \int_{0}^{w} \left( p(-v)e^{-icvp(v)} + p(v)e^{icvp(-v)} \right) \, dv.
\]

By continuity it follows that uniformly for \( c_1 \leq c \leq c_1' \)

\[
\left| \int_{0}^{1} p(-v)e^{-icvp(v)} \, dv \right| \leq \int_{0}^{1} p(-v) \, dv - c_2
\]

for some constant \( c_2 > 0 \). Hence

\[
|\overline{f}(z, w)| \leq \frac{1}{w} \int_{0}^{w} \left( p(-v) + p(v) \right) \, dv - \frac{c_2}{w} = 1 - \frac{c_2}{w},
\]

as proposed. (The case \( w < 0 \) is completely similar.)
Appendix C. Proof of Lemma 3

We prove here Lemma 3. We recall that

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

We start with part (i). Let \( \theta \) 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 \( \mathbf{b} \in \sin \theta S_{d-1}(\mathbf{u}) \) where \( S_{d-1}(\mathbf{u}) \) 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 \( \sin \theta S_{d-1}(\mathbf{u}) \) for any given angle \( \theta \). Thus

\[
\mathcal{B}(\mathbf{w}) = \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(t\|\mathbf{w}\| \cos \theta) \, d\theta \int_{\sin \theta S_{d-1}(\mathbf{u})} (\mathbf{b} + \cos \theta \mathbf{u}) \otimes (\mathbf{b} + \cos \theta \mathbf{u}) \, d\mathbf{b} \, dt
\]

\[ = \frac{d}{s_d} \int_0^\pi \int_0^\pi t^{d-1} q(\|\mathbf{w}\| \cos \theta) \, d\theta \int_{\sin \theta S_{d-1}(\mathbf{u})} (\mathbf{b} \otimes \mathbf{b} + (\cos \theta)^2 \mathbf{u} \otimes \mathbf{u}) \, d\mathbf{b} \, dt
\]

\[ + \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(\|\mathbf{w}\| \cos \theta) \, d\theta \int_{\sin \theta S_{d-1}(\mathbf{u})} \cos \theta (\mathbf{b} \otimes \mathbf{u} + \mathbf{u} \otimes \mathbf{b}) \, d\mathbf{b}. \]

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

\[
\mathcal{B}(\mathbf{w}) = \frac{d}{s_d} \int_0^1 \int_0^\pi t^{d-1} q(\|\mathbf{w}\| \cos \theta) \, d\theta \int_{\sin \theta S_{d-1}(\mathbf{u})} (\mathbf{b} \otimes \mathbf{b} + (\cos \theta)^2 \mathbf{u} \otimes \mathbf{u}) \, d\mathbf{b} \, dt
\]

\[ = \frac{d}{s_d} \int_0^\pi t^{d-1} q(\|\mathbf{w}\| \cos \theta) \, d\theta \int_{\sin \theta S_{d-1}(\mathbf{u})} (\sin \theta)^{d-1} \, d\theta \, dt
\]

\[ \int_{S_{d-1}(\mathbf{u})} ((\sin \theta)^2 \mathbf{b} \otimes \mathbf{b} + (\cos \theta)^2 \mathbf{u} \otimes \mathbf{u}) \, d\mathbf{b}. \]

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

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

\[
\int_{S_{d-1}} (b_i)^2 d\mathbf{b} = \int_{S_{d-1}} (b_j)^2 d\mathbf{b} = \frac{1}{d-1} \int_{S_{d-1}} \|\mathbf{b}\|^2 d\mathbf{b} = \frac{s_{d-1}}{d-1}
\]

Thus

\[
\int_{S_{d-1}(\mathbf{u})} \mathbf{b} \otimes \mathbf{b} d\mathbf{b} = \frac{s_{d-1}}{d-1} I_{d-1}(\mathbf{u})
\]

and similarly

\[
\int_{S_{d-1}(\mathbf{u})} \mathbf{u} \otimes \mathbf{u} d\mathbf{b} = s_{d-1} \mathbf{u} \otimes \mathbf{u}
\]

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 \( \mathbf{w} = \|\mathbf{w}\| \). 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

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

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

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

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

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

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 $(1-y^2)^{(d-3)/2}$ defined over $[0, 1]$. The Mellin transform $\beta_1^*(s)$ is defined for 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 Re$(s) > 0$. The Mellin transform $q^*(s)$ is defined for 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 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

$$
\phi^*(s) = 2\frac{ds_{d-1}}{(d-1)s_d} \beta_2^*(1-s) \frac{\beta_1^*(1-s)}{d-s} q^*(s),
$$

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 $\phi(w)$ as $w \to \infty$.

The Mellin transform $\phi^*(s)$ is also defined on 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 $|z| \leq 1$, $x \in [0,1]$ and all $w \in \mathbb{R}$

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

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

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

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

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

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

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

**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 $c > 0$)

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

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

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

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

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

for some constant $c'' > 0$.

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


Now, if we condition on $B$ we certainly have

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

Thus we obtain uniformly

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

and consequently

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

In general we certainly have the trivial upper bound

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

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 $v \geq c T^{1/4}$. Recall that $v$ is $U$ for $w = T^{1/4}$ so that we have $U \geq v \geq c T^{1/4}$ for all $|w| \leq T^{1/4}$. We compute the Gaussian type integral more precisely

$$\int_{|z| \leq 1} e^{-\frac{1}{2} U z^2 (1 + O(z)) - 2i \varepsilon z w} \, dz = \int_{-\infty}^{\infty} e^{-\frac{1}{2} U z^2 (1 + O(z)) - 2i \varepsilon z w} \, dz + O \left( \frac{1}{U} \right)$$

$$= \sqrt{\frac{2\pi}{U}} e^{-2\varepsilon^2 w^2 / U} + O \left( \frac{1}{U} \right).$$

Now we observe that $v \geq c T^{1/4}$ 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 $v \geq c T^{1/4}$ 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 $E[F(S_{\epsilon,1}(X^T))]$ we consider the following partition:

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

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

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

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

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

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

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

The lower bound ensures that $|S_{\epsilon,1}(X^T)| \geq 1$ 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 $\mathbb{E}S_{\epsilon}(X^T)$. By using the representation (14) and by expanding the determinant we obtain (very similarly as in the case $d = 1$)

$$\mathbb{E}S_{\epsilon}(X^T) = \sum_{j=0}^{d} T^{d-j} J_j(\epsilon),$$

where $J_j(\epsilon)$, $j = 0, \ldots, d$, 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 $\epsilon$. In particular the dominant term $J_0 = J_0(\epsilon)$ is given by

$$J_0 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \overline{f(w, z)} (B(z, w)) e^{-\epsilon \|w\|^2 - 2i\epsilon \langle w, z \rangle} \, dw \, dz,$$

where

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

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 $\overline{f}(w, z)$

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

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

$$z^T \overline{B}(w) z \gg \begin{cases} \|z\|^2 & \text{for } \|w\| \leq 1, \\ \|z\|^2 \left(\|w\| \cos \varphi + \|w\| \sin \varphi\right)^2 & \text{for } \|w\| > 1, \end{cases}$$

where $\varphi$ denotes the angle between $w$ and $z$, that is $\cos \varphi = \langle w, z \rangle / (\|w\| \|z\|)$. Furthermore

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

**Proof** The case $\|w\| \leq 1$ is easy to handle. We just use Taylor expansion and the property that

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

In the case $\|w\| > 1$ we have to be more careful. Let $w_0 = w / \|w\|$ and $w_1 = \tilde{w} / \|\tilde{w}\|$, where $	ilde{w} = z - \langle z, w \rangle / \|w\|^2 w$ is orthogonal to $w$. We now represent $x$ as $x = x_1 w_0 + x_2 w_1 + x_3$, where $x_2$ is in the orthogonal to $w$ and $z$. With the help of this notation we have

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

We also note that

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

and that the integral

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

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

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

as proposed.
The second part of the lemma follows by applying first Taylor expansion for $f(w, z)$ and then by taking the logarithm. The computations are very similar to the preceding ones. For example we have for the third derivative

$$\frac{\partial^3 f(w, z)}{\partial z_j \partial z_k \partial z_\ell} =
$$

$$i \mathbb{E} \left[ X_j X_k X_\ell p((X, w)) p(-(X, w)) \left( p(-(X, w))^2 e^{-i \langle X, z \rangle p(-(X, w))} - p((X, w))^2 e^{i \langle X, z \rangle p((X, w))} \right) \right]$$

which gives

$$\sum_{i,j,k} \frac{\partial^3 f(w, z)}{\partial z_i \partial z_j \partial z_k} z_i z_j z_k \ll \mathbb{E} \left[ p((X, w)) p(-(X, w)) \left| \langle X, z \rangle \right|^3 \right]$$

$$\ll \frac{1}{\| w \|^4} \left( | \cos \varphi | + \| w \| | \sin \varphi | \right)^3.$$  

This directly leads to the proposed asymptotic relation.

We recall that $X$ is uniformly distributed on the unit ball $B_d$. The idea is to parametrise the unit ball with the help of spherical coordinates

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

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

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

Note that for $t = 1$ we also get a parametrisation of the sphere $S_d$.

We start with a simple lemma.

**Lemma 18** Suppose that $x \in S_d$, that is $\| x \| = 1$. If $| \langle x, w \rangle | \leq 1$ then we have

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

whereas if $| \langle x, w \rangle | \geq 1$ we have

$$\int_0^1 t^{d-1} f(w, tx, z) dt \ll \frac{1}{\sqrt{| \langle x, z \rangle \langle x, \bar{w} \rangle |}} + \frac{e^{\langle x, w \rangle}}{| \langle x, z \rangle \langle x, w \rangle |}$$
**Proof** The proof follows that same lines as the proof of Lemma 6. We just use the (auxiliary) function

\[ h(t) = p(\langle tx, w \rangle) \langle tx, z \rangle \]

that satisfies

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

and

\[ h''(t) = \langle x, z \rangle \langle x, w \rangle e^{-\langle tx, w \rangle} \frac{2(1 + e^{-\langle tx, w \rangle}) + \langle tx, w \rangle (e^{-\langle tx, w \rangle} - 1)}{(1 + e^{-\langle tx, w \rangle})^3}. \]

As a corollary we obtain the following upper bounds for \( f(z, w) \).

**Lemma 19** If \( \|w\| \leq 1 \) then we have

\[ |f(z, w)| \leq \min \left( 1, C_1 \frac{\log(\|z\|)}{\|z\|} \right) \]

for some constant \( C_1 > 0 \), whereas if \( \|w\| \geq 1 \) we have

\[ |f(z, w)| \leq \min \left( 1, C_2 \frac{\log(\|z\| \|w\|) + e \|w\|}{\sqrt{\|z\| \|w\|}} \right). \]

for some constant \( C_2 > 0 \).

**Proof** We start with the case \( \|w\| \leq 1 \). Note that \( \|w\| \leq 1 \) implies \( \langle x, w \rangle \leq 1 \) for all \( x \in S_d \). By Lemma 18 we have

\[ \bar{f}(z, w) \ll \int_{S_d} \min \left( 1, \frac{1}{\|x, z\|} \right) dx, \]

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

\[ \bar{f}(z, w) \ll \int_{-1}^{1} (1 - x_1^2)^{d-3 \frac{1}{2}} \min \left( 1, \frac{1}{\|x_1\| \|z\|} \right) dx_1 \]

\[ \ll \frac{1}{\|z\|} + \frac{1}{\|z\|} \int_{1/\|z\|}^{1} (1 - x_1^2)^{d-3 \frac{1}{2}} dx_1 \]

\[ \ll \frac{\log(\|z\|)}{\|z\|} \]

as proposed.

Now suppose that \( \|w\| \geq 1 \). Then either (39) or (40) holds. But since \( \|w\| \geq 1 \) then (39) implies (40). Thus we have (40) in all cases. So we have to consider the two \((d - 1)\)-dimensional integrals

\[ K_1 = \int_{S_d} \min \left( 1, \frac{1}{\sqrt{\|x, z\| \langle x, w \rangle}} \right) dx \]

29
\[ K_2 = \int_{S_d} \min \left( 1, \frac{e^{\|w\|}}{\langle x, z \rangle \langle x, w \rangle} \right) \, dx. \]

Note that \( |\langle x, w \rangle| \leq \|w\| \) if \( x \in S_d \).

We start with \( K_1 \) and suppose first that \( z \) and \( w \) are parallel. Then we are in the same situation as in the previous case and, thus, we obtain
\[ K_1 \ll \frac{\log(\|z\| \|w\|)}{\sqrt{\|z\| \|w\|}}. \]

In general we distinguish between the cases
\[ \frac{\langle x, z \rangle}{\|z\|} \leq \frac{\langle x, w \rangle}{\|w\|} \quad \text{and} \quad \frac{\langle x, z \rangle}{\|z\|} > \frac{\langle x, w \rangle}{\|w\|} \]

and obtain
\[ \frac{1}{\sqrt{\langle x, z \rangle \langle x, w \rangle}} \leq \frac{1}{\|x, z\| \|w\|} + \frac{1}{\|x, w\| \|z\|}. \]

Thus, we get
\[
K_1 \ll \int_{S^{d-1}} \min \left( 1, \frac{1}{\langle x, z \rangle \|w\|} \right) \, dx + \int_{S_d} \min \left( 1, \frac{1}{\langle x, w \rangle \|z\|} \right) \, dx
\]
\[
\ll \int_{-1}^{1} (1 - x^2)^{d-3} \min \left( 1, \frac{1}{x_1 \sqrt{\|w\| \|z\|}} \right) \, dx_1
\]
\[
\ll \frac{\log(\|z\| \|w\|)}{\sqrt{\|z\| \|w\|}}
\]
as proposed.

Finally we consider the integral \( K_2 \), where we use the inequality
\[ \frac{1}{\langle x, z \rangle \langle x, w \rangle} \leq \frac{1}{\langle x, z \rangle^2 \|w\|} + \frac{1}{\langle x, w \rangle^2 \|z\|} \]

and use the property
\[
\int_{-1}^{1} (1 - x^2)^{d-3} \min \left( 1, \frac{1}{x_1^2 \|w\| \|z\|} \right) \, dx_1 \ll \frac{1}{\sqrt{\|z\| \|w\|}}
\]

This completes the proof of the lemma.

**Lemma 20** There exist constants \( c_1 > 0 \) and \( c_2 > 0 \) such that
\[ |\mathcal{F}(z, w)| \leq 1 - \frac{c_2}{\|w\|} \quad (43) \]

uniformly for \( \|z\| \geq c_1 \|w\| \log^2 \|w\| \).
Proof We consider first the integral $\int_0^1 t^{d-1} f(w, tx, z) dt$ and assume that $|\langle x, w \rangle| > u_0 + \eta$. We split up the integral into three intervals of the form (compare also with the proofs of Lemma 6 and 7):

$[0, 1/|\langle x, w \rangle|], ~ [1/|\langle x, w \rangle|, (|u_0| + \eta)/|\langle x, w \rangle|], ~ [(|u_0| + \eta)/|\langle x, w \rangle|, 1]$

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

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

Note that we used a the trivial bound $|f(w, tx, z)| \leq 1$ in the third interval.

We already observed that

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

Furthermore we have

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

This directly leads to

$$|\mathcal{F}(z, w)| \leq 1 - \frac{d_1}{\|w\|} + \frac{d_2 \log(\|z\|/\|w\|)}{\sqrt{\|z\|/\|w\|}}$$

for proper constants $d_1, d_2 > 0$. Clearly if $\|z\| \geq c_1 \|w\| \log^2 \|w\|$ 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

$$|\mathcal{F}(z, w)| \leq 1 - \frac{c_4}{\|z\|^d} \quad (44)$$

uniformly for $\|z\| \geq c_3 \|w\|$. In particular it follows that

$$|\mathcal{F}(z, w)| \leq 1 - \frac{c_5}{(\|w\| \log^2 \|w\|)^d}$$

uniformly for $c_3 \|w\| \leq \|z\| \leq c_1 \|w\| \log^2 \|w\|$. 

Proof The idea is to show that for some constant $c_3 > 0$ we have

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

uniformly for all $z$ with $\|z\| \geq c_3 \|w\|$. Clearly (45) implies (44).
We apply the substitution $u = \|z\| x$. Furthermore we set $c = \|w\| \|z\|$, $z_0 = z / \|z\|$ and $w_0 = w / \|w\|$, and we also represent $z_0$ as $z_0 = S w_0$ for some rotation $S$. With these notations in inequality (45) is equivalent to the inequality

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

that should now hold uniformly in $c \in [0, 1/c_2]$, $w_0$, and $S$. However, this is trivial by a compactness argument. Clearly for every choice of $c \in [0, 1/c_2]$, $w_0$, and $S$ the left hand side is smaller than 1, and the left hand side is continuous in $c \in [0, 1/c_2]$, $w_0$, 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

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

provided that $\varepsilon \gg \max(T^{-1/2}, T^{-2/(d+1)})$, where $\mathcal{B}(\mathbf{w})$ is given by (16).

We recall that

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

and we note that

$$\mathcal{B}(0, \mathbf{w}) = \mathcal{B}(\mathbf{w}).$$

Moreover $\mathcal{B}(0, \mathbf{w})$ is a positive matrix and all entries of $\mathcal{B}(\mathbf{z}, \mathbf{w})$ satisfy

$$|\mathcal{B}(\mathbf{z}, \mathbf{w})_{i,j}| \leq \mathcal{B}(\mathbf{w})_{i,j}, \quad 1 \leq i, j \leq d.$$ 

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

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

Furthermore by Taylor expansion (and similar computations) we have

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

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

As in the case $d = 1$ we partition the $2d$-dimensional integral into several parts.
A: The case $\|w\| \leq 1$. First suppose that $\|z\| \leq T^{-1/3}$:

$$I_{11} = \frac{1}{(2\pi)^d} \int_{\|z\| \leq T^{-1/3}} \int_{\|w\| \leq 1} \det(B(z, w)) \mathcal{J}(w, z)^{T-d} e^{-\varepsilon \|w\|^2 - 2i\varepsilon \langle w, z \rangle} \, dw \, dz$$

$$= \frac{1}{(2\pi)^d} \int_{\|w\| \leq 1} \det(B(w)) e^{-(T-d)\frac{1}{2}z^T B(w) z} (1 + O(T\|z\|^3) + O(\|z\|^2)) \, dw$$

$$+ O \left( \int_{\|w\| \leq 1} \det(B(w)) e^{-\varepsilon \|w\|^2} \int_{\mathbb{R}^d} e^{-(T-d)\frac{1}{2}z^T B(w) z} \|z\|^3 \, dz \, dw \right)$$

$$+ O \left( \int_{\|w\| \leq 1} \det(B(w)) e^{-\varepsilon \|w\|^2} \int_{\mathbb{R}^d} e^{-(T-d)\frac{1}{2}z^T B(w) z} \|z\|^2 \, dz \, dw \right)$$

$$= \frac{1}{(2\pi T)^{d/2}} \int_{\|w\| \leq 1} \sqrt{\det(B(w))} e^{-\varepsilon \|w\|^2 - \frac{2}{T} w^T B(w)^{-1} w} \, dw$$

$$+ O \left( e^{-cT^{1/3}} + T^{-\frac{d+1}{2}} + T^{-\frac{d+2}{2}} \right)$$

for some constant $c > 0$.

Next we consider the case $T^{-1/3} \leq \|z\| \leq C$, where $C$ is chosen in a way that $C/\log C \geq 2C_1$, where $C_1$ is the constant from the inequality (41):

$$I_{12} = \frac{1}{(2\pi)^d} \int_{T^{-1/3} \leq \|z\| \leq C} \int_{\|w\| \leq 1} \det(B(z, w)) \mathcal{J}(w, z)^{T-d} e^{-\varepsilon \|w\|^2 - 2i\varepsilon \langle w, z \rangle} \, dw \, dz$$

$$\ll \int_{T^{-1/3} \leq \|z\| \leq C} \int_{\|w\| \leq 1} e^{-c_1 T \|z\|^2} e^{-\varepsilon \|w\|^2} \, dw \, dz \ll e^{-c_1 T^{1/3}} \ll T^{-d}$$

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

$$I_{13} = \frac{1}{(2\pi)^d} \int_{\|z\| \geq C} \int_{\|w\| \leq 1} \det(B(z, w)) \mathcal{J}(w, z)^{T-d} e^{-\varepsilon \|w\|^2 - 2i\varepsilon \langle w, z \rangle} \, dw \, dz$$

$$\ll \int_{\|z\| \geq C} \left( \frac{C_1 \log(\|z\|)}{\|z\|} \right)^{T-d} \, dz$$

$$\ll \int_{C}^\infty r^{d-1} \left( \frac{C_1 \log r}{r} \right)^{T-d} \, dr$$

$$\ll \frac{1}{T^{2T}} \ll T^{-d}.$$
Summing up we have

\[ I_1 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\|w\| \leq 1} \det(B(z, w)) f(w, z)^{T-d} e^{-\|w\|^2 - 2\varepsilon \langle w, w \rangle} \, dw \, dz \]

\[ = \frac{1}{(2\pi T)^{d/2}} \int_{\|w\| \leq 1} \sqrt{\det(B(w))} e^{-\|w\|^2 - \frac{2\varepsilon^2}{T} w^T B(w)^{-1} w} \, dw + O \left( T^{-\frac{d+1}{2}} \right). \]

**B: The case** \( 1 \leq \|w\| \leq \eta T \).

We again assume that \( \eta = \eta(T) c_6 (\log T)^{-3} \rightarrow 0 \) for a sufficiently small positive constant \( c_6 \).

We start with the case \( \|z\| \leq z_1 = \|w\|^{3/2} T^{-1/2} \eta^{-1/6} \). By Lemma 17 and by (47) we have

\[ \int_{\|z\| \leq z_1} \det(B(w, z)) f(w, z)^{T-d} e^{-2\varepsilon \langle w, w \rangle} \, dz \]

\[ = \det(B(w)) \int_{\|z\| \leq z_1} e^{-(T-d) \frac{1}{2} z^T B(w) z - 2\varepsilon w^T z} \, dz \]

\[ + O \left( \|w\|^{-d-3} \int_{\|z\| \leq z_1} \|z\|^2 e^{-(T-d) \frac{1}{2} z^T B(w) z} \, dz \right) \]

\[ + O \left( \|w\|^{-d-6} \int_{\|z\| \leq z_1} \|z\|^3 e^{-(T-d) \frac{1}{2} z^T B(w) z} (1 + \|w\|)^3 \, dz \right) \]

\[ + O \left( \|w\|^{-d-8} \int_{\|z\| \leq z_1} \|z\|^4 e^{-(T-d) \frac{1}{2} z^T B(w) z} (1 + \|w\|)^4 \, dz \right) \]

\[ = \sqrt{\det(B(w))} \left( \frac{2\pi}{T} \right)^{d/2} e^{-\frac{2\varepsilon^2}{T} w^T B(w)^{-1} w} \]

\[ + O \left( \frac{\|w\|^{(d-4)/2}}{T^{d/2}} e^{-\varepsilon \|w\|^{-1/3}} \right) + O \left( \frac{\|w\|^{d/2}}{T^{(d+2)/2}} \right) \]

\[ + O \left( \frac{\|w\|^{(d+3)/2}}{T^{(d+3)/2}} \right). \]

This implies

\[ I_{21} = \frac{1}{(2\pi)^d} \int_{1 \leq \|w\| \leq \eta T} \int_{\|z\| \leq z_1} \det(B(z, w)) f(w, z)^{T-d} e^{-\|w\|^2 - 2\varepsilon \langle w, w \rangle} \, dw \, dz \]

\[ = \frac{1}{(2\pi T)^{d/2}} \int_{1 \leq \|w\| \leq \eta T} \sqrt{\det(B(w))} e^{-\|w\|^2 - \frac{2\varepsilon^2}{T} w^T B(w)^{-1} w} \, dw \]

\[ + O \left( \frac{e^{-3d/4 + 1} e^{-\varepsilon \|w\|^{-1/3}}}{T^{d/2}} \right) + O \left( \frac{e^{-3d/4 - 1}}{T^{(d+2)/2}} \right) + O \left( \frac{e^{-3d/4 + 1}}{T^{(d+3)/2}} \right) \]

\[ = \frac{1}{(2\pi T)^{d/2}} \int_{\|w\| \geq 1} \sqrt{\det(B(w))} e^{-\|w\|^2 - \frac{2\varepsilon^2}{T} w^T B(w)^{-1} w} \, dw \]

\[ + O \left( \frac{e^{-3d/4 + 1} e^{-\varepsilon \|w\|^{-1/3}}}{T^{d/2}} \right) + O \left( \frac{e^{-3d/4 - 1}}{T^{(d+2)/2}} \right) + O \left( \frac{e^{-3d/4 + 1}}{T^{(d+3)/2}} \right) + O \left( \frac{e^{-3d/4 + 1}}{T^{(d+4)/2}} \right). \]
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 \( z_1 = \|w\|^3/2T^{-1/2} \eta^{-1/6} \leq \|z\| \leq c_1 \|w\| \), where \( c_1 \) sufficiently small. Here we know that (for a suitable constant \( c > 0 \))
\[ |\mathcal{f}(z, w)| \leq e^{-c\|z\|^2/\|w\|^3} \]
uniformly for \( z_1 \leq \|z\| \leq c_1 \|w\| \). This implies that the corresponding integral is upper bounded by
\[
I_{22} = \frac{1}{(2\pi)^d} \int_{1 \leq \|w\| \leq \eta T} \int_{c_1 \|w\| \leq \|z\| \leq c_1 \|w\|} \det(B(z, w))\mathcal{f}(z, w) T^{-d} e^{-\varepsilon \|z\|^2 - 2i \varepsilon \langle w, z \rangle} \, dw \, dz
\]
\[ \ll \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d-2} e^{-\varepsilon \|w\|^2} \int_{c_1 \|w\| \leq \|z\| \leq c_1 \|w\|} e^{-c(T-d)\|z\|^2/\|w\|^3} \, dz \, dw
\]
\[ \ll \frac{\eta^{-(d-2)/6} e^{-c\eta^{-1/3}}}{T^{d/2}} \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d/2-2} e^{-\varepsilon \|w\|^2} \, dw
\]
\[ \ll \frac{\eta^{-(d-2)/6} e^{-c\eta^{-1/3}}}{T^{d/2}} \varepsilon^{-3d/4+1}. \]

In the next step we consider the case \( c_1 \|w\| \leq \|z\| \leq c'_1 \|w\| \log^2 \|w\| \). Here we have the upper bound \( |\mathcal{f}(z, w)| \leq 1 - c_4/(\|w\| \log^2 \|w\|)^d \) and, thus, we get
\[
I_{23} = \frac{1}{(2\pi)^d} \int_{1 \leq \|w\| \leq \eta T} \int_{c_1 \|w\| \leq \|z\| \leq c'_1 \|w\| \log^2 \|w\|} \det(B(z, w))\mathcal{f}(z, w) T^{-d} e^{-\varepsilon \|z\|^2 - 2i \varepsilon \langle w, z \rangle} \, dw \, dz
\]
\[ \ll \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d-2} e^{-\varepsilon \|w\|^2} \int_{c_1 \|w\| \leq \|z\| \leq c'_1 \|w\| \log^2 \|w\|} e^{-Tc_4/(\|w\| \log^2 \|w\|)^d} \, dz \, dw
\]
\[ \ll \int_{1 \leq \|w\| \leq \eta T} (\log \|w\|)^2 \|w\|^{-d-2} e^{-Tc_4/(\|w\| \log^2 \|w\|)^d - \varepsilon \|w\|^2} \, dw
\]
\[ \ll \int_{1}^{\infty} (\log r)^2 r^{-d-3} e^{-c_4 T/(r \log r)^d - \varepsilon r^2} \, dr. \]

We split the integral at \( r_0 = (T/\varepsilon)^{1/(d+2)} \) so that
\[ \frac{T}{(r_0 \log^2 r_0)^d} = (d + 2)^2 \frac{T^{2d}}{\log^2(T/\varepsilon)} \quad \text{and} \quad \varepsilon r_0^2 = (T^2 \varepsilon)^{1/(d+2)}. \]

Since \( \varepsilon \gg T^{2/(d+1)} \) we obtain the upper bound
\[ I_{23} \ll e^{-c''T^\kappa} \]
for some constants \( c'', \kappa > 0 \).

35
Next suppose that $c_1' \|w\| \log^2 \|w\| \leq \|z\| \leq c_3 e^4 \|w\|$ for an arbitrary constant $c_3 > 0$. Here we have the upper bound $|\mathcal{F}(z, w)| \leq 1 - c_2/\|w\| \leq e^{-c_2/\|w\|}$ and consequently

$$I_{24} = \frac{1}{(2\pi)^d} \int_{1 \leq \|w\| \leq \eta T} \int \det(B(z, w)) \mathcal{F}(w, z)^{T-d} e^{-\varepsilon \|w\|^2 - 2\varepsilon \langle w, z \rangle} \, dw \, dz$$

$$\ll \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d-2} e^{-c_2 T/\|w\| + 4d \|w\| - \varepsilon \|w\|^2} \, dw$$

$$\leq \int_{1 \leq \|w\| \leq c_2 \sqrt{T}} \|w\|^{-d-2} e^{-(1-c_2)\sqrt{T} - \varepsilon \|w\|^2} \, dw + \int_{c_2 \sqrt{T} \leq \|w\| \leq \eta T} \|w\|^{-d-2} e^{-\varepsilon/2} \|w\|^2 \, dw$$

$$\ll e^{-(1-c_2)\sqrt{T} + e^{-(c_2^2/2)\varepsilon T}}$$

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

Finally if $\|z\| \geq c_3 e^4 \|w\|$, where $c_3$ is chosen sufficiently large, we have

$$\mathcal{F}(z, w) \leq C \max \left( \log(\|w\|), \log(\|z\|), e^{\|w\|}\right) \leq \frac{1}{\sqrt{\|w\| \|z\|}} \leq 2.$$ (48)

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

$$I_{25} = \frac{1}{(2\pi)^d} \int_{1 \leq \|w\| \leq \eta T} \int \det(B(z, w)) \mathcal{F}(w, z)^{T-d} e^{-\varepsilon \|w\|^2 - 2\varepsilon \langle w, z \rangle} \, dz \, dw$$

$$\ll \tilde{C}^{T-d} \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d-2} \left( \log(\|w\|)^{(T-d)/2} \right) \|z\|^{-(T-d)/2} \, dz \, dw$$

$$+ \tilde{C}^{T-d} \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d-2 -(T-d)/2} \int_{\|z\| \geq c_3 e^4 \|w\|} \left( \log(\|z\|) \right)^{(T-d)/2} \|z\|^{-(T-d)/2} \, dz \, dw$$

$$+ \tilde{C}^{T-d} \int_{1 \leq \|w\| \leq \eta T} \|w\|^{-d-2 -(T-d)/2} e^{(T-d) \|w\|} \int_{\|z\| \geq c_3 e^4 \|w\|} \|z\|^{-(T-d)/2} \, dz \, dw$$

$$\ll \frac{\tilde{C}^T}{T c_3^{T/2}} \int_1^\infty \frac{r^{d-3-(T-d)} (\log r)^{T-d} e^{-2(T-d) r + 4r}}{r^{(T-d)/2}} \, dr$$

$$+ \frac{\tilde{C}^T}{T c_3^{T/2}} \int_1^\infty \frac{r^{d-3-(T-d)} (\log(c_3 e^4 r)^{T-d} e^{-2(T-d) r + 4r}}{r^{(T-d)/2}} \, dr$$

$$+ \frac{\tilde{C}^T}{T c_3^{T/2}} \int_1^\infty \frac{r^{d-3-(T-d)} e^{-2(T-d) r + 4r}}{r^{(T-d)/2}} \, dr$$

$$\ll \frac{1}{T^{2T}}.$$

Summing up we have

$$J_2 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{1 \leq \|w\| \leq \eta T} \det(B(z, w)) \mathcal{F}(w, z)^{T-d} e^{-\varepsilon \|w\|^2 - 2\varepsilon \langle w, z \rangle} \, dw \, dz$$

$$= J_{21} + J_{22} + J_{23} + J_{24} + J_{25}$$

$$= \frac{1}{\sqrt{2\pi T}} \int_{\|w\| \geq 1} \sqrt{\det(B(w))} e^{-\frac{2\varepsilon^2}{T} w^T B(w)^{-1} w} e^{-\varepsilon \|w\|^2} \, dw + O \left( \frac{e^{-3d/4}}{T^{(d+2)/2}} \right).$$
We now discuss some possible extensions of Theorem 2 to a larger class of hypothesis class for a sufficiently small positive constant $c$ 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 $\mathcal{H}_{p,w}$

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

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

$$I_{31} = \frac{1}{(2\pi)^d} \int_{\|w\| \geq \eta T} \int_{\|z\| \leq e^4 \|w\|} \det (B(z,w)) f(w,z)^{T-d} e^{-\varepsilon \|w\|^2 - 2\varepsilon \langle w,z \rangle} \, dw \, dz$$

$$\ll \int_{\|w\| \geq \eta T} \|w\|^{-d-2} e^{4d\|w\| - \varepsilon \|w\|^2} \, dw$$

$$\ll \int_{r \geq \eta T} e^{-\frac{7}{16} r^2} \, dr \ll \frac{1}{\varepsilon \eta T} e^{-\frac{7}{8} (\eta T)^2} \ll e^{-\eta^2 T}.$$

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

$$I_{32} = \frac{1}{(2\pi)^d} \int_{\|w\| \geq \eta T} \int_{\|z\| \geq e^4 \|w\|} \det (B(z,w)) f(w,z)^{T-d} e^{-\varepsilon \|w\|^2 - 2\varepsilon \langle w,z \rangle} \, dw \, dz$$

$$\ll \tilde{C}^T \int_{\|w\| \geq \eta T} \|w\|^{-d-2-(T-d)/2} e^{(T-d)\|w\| - \varepsilon \|w\|^2} \int_{\|z\| \geq e^4 \|w\|} \|z\|^{-(T-d)/2} \, dz \, dw$$

$$\ll \frac{\tilde{C}^T}{T} \int_{\|w\| \geq \eta T} \|w\|^{-d-2-(T-d)/2} e^{(T-d)/4 \|w\|} e^{(d-(T-d)/2)4\|w\|} \, dw$$

$$\ll \frac{\tilde{C}^T}{T} \int_{r \geq \eta T} e^{(3d-T)r} \ll e^{-\frac{7}{4} T^2}.$$ Consequently,

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

D: The whole range.

Summing up we arrive at

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

provided that

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

and where we have set

$$\eta = \eta(T) = c_6 (\log T)^{-3}$$
In the logistic case for all $y^T$ we have $\nabla^2 L(y^T|w) = B(w)$. In general, this is not the case. In fact, $\nabla^2 L(y^T|w)$ varies with $y^T$, and when it is viewed as a random variable we have $\mathbb{E}[Y] = 1 - 2p(w)$ and $\mathbb{E}[\nabla^2 L(y^T|w)] = B(w)$. Indeed,

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

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

$$
A(w) = \frac{1}{2} \sum_{t} \left( \frac{(p'((w|x_t)))^2}{(p((w|x_t)))^2} + \frac{(p'((w|x_t)))^2}{(1 - p((w|x_t)))^2} - \frac{p''((w|x_t))}{1 - p((w|x_t))} \right) x_t \otimes x_t,
$$

$$
F_t = \frac{1}{2} \left( -\frac{(p'((w|x_t)))^2}{(p((w|x_t)))^2} + \frac{(p'((w|x_t)))^2}{(1 - p((w|x_t)))^2} + p''((w|x_t)) \left( \frac{1}{p((w|x_t))} - \frac{1}{p((w|x_t))} \right) \right) x_t \otimes x_t.
$$

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

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

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

$$
P \left( \|F(y^T) - \mathbb{E}[F(y^T)]\| > Ad^{T\alpha} \right) \leq 2d^{T}e^{-BT^{2\alpha-1}}$$

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

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

$$
\tilde{F}_T(\Theta) = \prod_{t} \left( p((w|x_t))e^{-\text{Tr}(\Theta F_t)} + (1 - p((w|x_t)))e^{\text{Tr}(\Theta F_t)} \right).
$$

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

$$
p((w|x_t))e^{-\text{Tr}(\Theta F_t)} + (1 - p((w|x_t)))e^{\text{Tr}(\Theta F_t)} = e^{\text{Tr}(\Theta F_t)} \left( 1 + p((w|x_t))(e^{-2\text{Tr}(\Theta F_t)} - 1) \right)
$$

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

$$
|p((w|x_t))(e^{-2\text{Tr}(\Theta F_t)} - 1)| < 1/2 \quad \text{and} \quad p((w|x_t))(e^{-2\text{Tr}(\Theta F_t)} - 1) \neq 0.
$$

Since $\mathcal{U}$ is simply connected, the logarithm of $\tilde{F}_T(\Theta)$ exists, and it turns out that the logarithm is always bounded by some $C > 0$, hence the logarithm of the product satisfies: $|\log \tilde{F}(\Theta)| \leq TC$. As a consequence, $\log \tilde{F}(\Theta)$ is an analytic function on $\mathcal{U}$ and its derivatives are also $O(T)$, in particular the second derivative. Since the first derivative of $\log \tilde{F}_T(\Theta)$ at its derivatives are also
\(O(T)\), in particular the second derivative. Since the first derivative of \(\log \tilde{F}_T(\Theta)\) at \(\Theta = 0\) is exactly \(\mathbb{E}[F(y^T)]\), we have the following Taylor expansion

\[
\log \tilde{F}(\Theta) = \text{Tr} \left( \Theta \mathbb{E}[F(y^T)] \right) + O(T\|\Theta\|^2)
\]

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

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

Since \(\mathbb{E}[e^{\theta \text{Tr}(e_{ij} F(y^T))}] = \tilde{F}(\theta e_{ij})\), and thanks to the estimate (51) with \(\|e_{ij}\|^2 = 1\), the right-hand side is upper bounded by \(\exp(RT\theta^2 - \theta AT^\alpha)\) with the minimum \(\exp(-(\alpha T^2(\log d - 1))/(4R))\). Also with the minimum \(\exp(\alpha T^2(\log d - 1)/(4R))\). Also \(P(\tilde{F}(y^T)_{ij} < \mathbb{E}[\tilde{F}(y^T)]_{ij} - AT^\alpha)\) 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 \(x_t\) is generated by a distribution \(\mu\) over the cube \([-1, 1]^d\). Then one should expect

\[
\mathbb{E}[S^e(X^T)] = \left(\frac{T}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} \sqrt{\text{det}(B(w))} e^{-\frac{1}{2}||w||^2} \frac{1}{2} e^{-\frac{1}{2}||w||^2} d\mu(w) \times (1 + O(dT^\alpha - 1))
\]

with

\[
\tilde{B}(w) = \int_{\mathbb{R}^d} \mu(x) \frac{p'(\langle w | x \rangle)^2}{p(\langle w | x \rangle)(1 - p(\langle w | x \rangle))} x \otimes x dx.
\]

Furthermore it should hold that

\[
\tilde{v}_T = \mathbb{E}[\log S(X^T)] \sim \log \mathbb{E}[S^e_T(X^T)].
\]

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

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

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

\[
|\text{det}(\nabla^2 L(y^T | w)) - \text{det}(B(w, x^T))| = |\text{det}(B(w, x^T)) \cdot \|B(w, x^T)^{-1}\| O(\|\nabla^2 L(y^T | w) - B(w, x^T)\|)) = |\text{det}(B(w, x^T)) \cdot \|B(w, x^T)^{-1}\| O(dAT^\alpha)
\]

\[
= \text{det}(B(w, x^T)) O(dT^\alpha - 1)
\]
since \( \|B(w, x^T)^{-1}\| = O(d/T) \) (as long as the matrix \( B(w, x^T)/T \) is not degenerate). Thus 
\[
\det(\nabla^2 L(y^T|w)) = \det(B(w, x^T))(1 + O(d^2T^{\alpha-1})).
\]
We call \( J_\alpha \) the set of sequences \( y^T \) that satisfies this property. We already know that \( \sum_{y^T \notin J_\alpha} P(y^T|w) < 2d^2e^{-BT^{2\alpha-1}} \) that exponentially decays since \( \alpha > 1/2 \). Therefore with \( O(d^\alpha) \) we can approximate \( \det(\nabla^2 L(y^T|x^T, w)) \) by \( \det(B(w, x^T)) \) which allows us to write
\[
S(x^T|w) = \frac{1}{(2\pi)^d} \sum_{y^T} \det(\nabla^2 L(y^T|x^T, w)) \int_{\mathbb{R}^d} \exp(L(y^T|x^T, w) - i\langle \nabla L(y^T|x^T, w)|z \rangle) dz
\]
\[
= \frac{\det(B(w, x^T))}{(2\pi)^d} \sum_{y^T} \int_{\mathbb{R}^d} \exp(L(y^T|x^T, w) - i\langle \nabla L(y^T|x^T, w)|z \rangle) dz \left(1 + O(dT^{\alpha-1})\right)
\]
\[
= \frac{\det(B(w, x^T))}{(2\pi)^d} \prod_{t} f(w, x_t, z) \left(1 + O(dT^{\alpha-1})\right).
\]
Since 
\[
S^\varepsilon(x^T) = (2\pi)^{-d} \int S(x^T|w) e^{-\varepsilon \|w\|^2} dw
\]
we, thus, obtain a similar expression as in the logistic case. Hence, this is a strong indication that (52) holds.