Adaptive Convergence Rates for Log-Concave Maximum Likelihood

Gil Kur
ETH Zürich

Adityanand Guntuboyina
UC-Berkeley

Abstract

We study the task of estimating a log-concave density in $\mathbb{R}^d$ using the Maximum Likelihood Estimator, known as the log-concave MLE. We show that for every $d \geq 4$, the log-concave MLE attains an *adaptive rate* when the negative logarithm of the underlying density is the maximum of k affine functions, meaning that the estimation error for such a density is significantly lower than the minimax rate for the class of log-concave densities. Specifically, we prove that for such densities, the risk of the log-concave MLE is of order $c(k) \cdot n^{-\frac{4}{d}}$ in terms of the Hellinger squared distance. This result complements the work of Kim et al. (2018) and Feng et al. (2021), who addressed the cases $d = 1$ and $d \in \{2, 3\}$, respectively. Our proof provides a unified and relatively simple approach for all $d \geq 1$, and is based on techniques from stochastic convex geometry and empirical process theory, which may be of independent interest.[1]

1 Introduction

Density estimation is a fundamental problem in statistics, where the goal is to estimate an unknown density function f_0 on $\mathbb{R}^d$ from data $\mathcal{D} := \{X_1, \ldots X_n\}$ consisting of n i.i.d. observations $X_i \overset{\text{i.i.d.}}{\sim} f_0$. In this paper, we study the problem of *log-concave* density estimation where it is assumed that the unknown density f_0 is log-concave; i.e., f_0 satisfies the inequality

$$f_0(\lambda x + (1-\lambda)y) \geq f_0(x)^\lambda f_0(y)^{1-\lambda}$$

for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0, 1]$.

The class of log-concave densities in $\mathbb{R}^d$, which we denote by $\mathcal{F}_d$, has been studied in many disciplines including statistics (Bagnoli and Bergstrom, 2005), mathematics (Artstein-Avidan et al., 2015; Stanley, 1989), computer science (Balcan and Long, 2013; Lovász and Vempala, 2007) and economics (An, 1997). Furthermore, the problem of log-concave density estimation has received much attention in statistics and computer science over the past two decades; for further details, see the survey of Samworth (2018).

The maximum likelihood (MLE) over the class of the log-concave densities, known as the log-concave MLE, is perhaps the most well-studied method for log-concave density estimation. Specifically, given the data $\mathcal{D}$ drawn from an unknown log-concave density $f_0 \in \mathcal{F}_d$, the log-concave MLE, denoted by $\widehat{f}_n$, is given by

$$\widehat{f}_n := \operatorname*{argmax}_{f \in \mathcal{F}_d} \int \log f d\mathbb{P}_n$$

where $\mathbb{P}_n := n^{-1} \sum_{i=1}^n \delta_{X_i}$ is the random empirical measure over $\{X_i\}_{i=1}^n$.

The study of $\widehat{f}_n$ has been the focus of many papers; here we state a few results that are related to our paper. Cule et al. (2010) showed that $\widehat{f}_n$ exists uniquely and provided an algorithm based on convex optimization to compute it. Efficient computation was addressed in Axelrod et al. (2019). Schuhmacher and Dümbgen (2010) proved that for any $d \geq 1$, the $\widehat{f}_n$ is a consistent estimator of the true density f_0. Specifically, they showed that the *risk* of $\widehat{f}_n$, with respect to the divergence

$$d_X^2(\widehat{f}_n, f_0) := \int \log \frac{\widehat{f}_n}{f_0} d\mathbb{P}_n$$

goes to zero as n grows, i,e.

$$\mathcal{R}_{d_X^2}(\widehat{f}_n, \mathcal{F}_d) := \sup_{f_0 \in \mathcal{F}_d} \mathbb{E}_{\mathcal{D}}\, d_X^2(\widehat{f}_n, f_0) \underset{n \to \infty}{\to} 0.$$

Although the divergence d_X^2 is less widely used in the literature; one can show (see, e.g. Kim et al. (2018, Pg. 2281)) that

$$h^2(\widehat{f}_n, f_0) \leq KL(\widehat{f}_n || f_0) \leq d_X^2(\widehat{f}_n, f_0),$$

[1]Please consult the arXiv version of the paper for an updated (and extended) manuscript.

Proceedings of the 28[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

where $h^2(f,g) := 2^{-1}\int(\sqrt{f}-\sqrt{g})^2$, and $KL(f||g) := \int \log(f/g)f$ denote the Hellinger squared and Kullback-Leibler divergences, respectively. Therefore, consistency in d_X^2 implies consistency in both Hellinger and Kullback-Leibler divergences.

Rates of convergence of $\widehat{f}_n$, along with the associated minimax rates over $\mathcal{F}_d$ were studied in Kim and Samworth (2016), Doss and Wellner (2016), Carpenter et al. (2018) and Kur et al. (2019). Specifically, Kim and Samworth (2016) and Kur et al. (2019) showed that, for all $d \geq 1$, and $n \geq d+1$,

$$\mathcal{R}_{d_X^2}(\widehat{f}_n, \mathcal{F}_d) \underset{d}{\lesssim} \log(n)^{\mathbf{1}_{d\geq 3}} \cdot n^{-\min\{\frac{2}{d+1},\frac{4}{5}\}},$$

where $\underset{d}{\lesssim}$ (or $\underset{d}{\asymp}$) denotes inequality (or equality) up to a multiplicative positive constant that only depends on the dimension d. Also, they proved that the minimax rate with respect to the Hellinger distance that

$$\mathcal{M}_{h^2}(\mathcal{F}_d) := \inf_{\bar{f}_n} \mathcal{R}_{h^2}(\bar{f}_n, \mathcal{F}_d) \gtrsim n^{-\min\{\frac{2}{d+1},\frac{4}{5}\}},$$

where the infimum is taken of all possible estimators that their input is $\mathcal{D}$, and $\lesssim$ (or $\asymp$) denotes inequality (and equality) up to an absolute positive multiplicative constant. Therefore, $\widehat{f}_n$ is minimax optimal in both d_X^2 and h^2 for all $d \geq 1$. Also these results show that the minimax rate in h^2 as well as d_X^2 equals $n^{-4/5}$ for $d=1$ and $n^{-2/(d+1)}$ (up to a possible log factor) for $d \geq 2$. It is very interesting that, for log-concave density estimation, the minimax rate in squared total variation distance (we denote by $d_{TV}(f,g) := \int |f-g|$) is different from the minimax rate for h^2. Indeed, Kur et al. (2019) showed, using metric entropy arguments of Gao and Wellner (2017); Bronshtein (1976), that

$$\mathcal{M}_{d_{TV}^2}(\mathcal{F}_d) \underset{d}{\asymp} n^{-\frac{4}{d+4}}$$

In words, the minimax rate over $\mathcal{F}_d$ for d_{TV}^2 is $n^{-4/(d+4)}$ for all $d \geq 1$, and this rate is strictly smaller than the aforementioned rate for h^2 for $d \geq 3$ (it may be helpful to note here that $d_{TV}(f,g) \leq h(f,g)$ for every pair of densities f and g).

This work begins with the question of whether the log-concave MLE attains *adaptive* rates in all dimensions $d \geq 1$. More precisely, we ask whether there exists an "interesting and natural" subclass of log-concave densities $\mathcal{G}_d \subset \mathcal{F}_d$ such that

$$\mathcal{R}_{d_X^2}(\widehat{f}_n, \mathcal{G}_d) \underset{d}{\asymp} \underbrace{a_n \cdot \mathcal{M}_{d_X^2}(\mathcal{F}_d)}_{:=m_n}, \tag{1}$$

for some sequence of positive numbers $a_n \to 0$. If the answer is positive, we say that the log-concave MLE adapts to the densities in $\mathcal{G}_d$ in the rate of order m_n.

There are several motivations for studying the adaptivity of MLE in statistics (and in particular, of the log-concave density estimation). First, as the minimax risk is considered to be quite a restrictive measure of statistical performance (cf. Tsybakov (2003)), one would like to understand when the estimator achieves faster rates on a certain subclass of interest rather than the entire class. Secondly, in many cases, the restricted MLE over this sub-class is not computationally tractable, and the MLE over the *entire* class can be computed efficiently — and one would hope that the MLE is optimal over this subclass (though it considers more densities).

In the case of log-concavity, a natural candidate to consider is the sub-class of log-concave densities such that their negative logarithm is a k-max affine and their support is a convex polytope (or polyhedron) $\mathcal{P} \subset \mathbb{R}^d$. Formally,

$$\begin{aligned} f_0 \in \mathcal{G}(k,\mathcal{P}) \iff &\exists \{u_i\}_{i=1}^k \in \mathbb{R}^d, \{b_i\}_{i=1}^k \in \mathbb{R} \text{ s.t.} \\ &f_0 = \exp(-\max_{1\leq i\leq k} u_i^\top x + b_i) \cdot \mathbf{1}_{\mathcal{P}}, \end{aligned} \tag{2}$$

where $\mathbf{1}_{\mathcal{P}} : \mathbb{R}^d \to \{0,1\}$ denotes the indicator function on $\mathcal{P}$. The adaptation of $\widehat{f}_n$ on $\mathcal{G}(k,\mathcal{P})$ is studied in Kim et al. (2018); Feng et al. (2021), and its twin task of estimating a regression function that is k-max affine functions (with an incomplete list Kur et al. (2020a); Han and Wellner (2016); Ghosh et al. (2021); Balázs (2022)). We discuss this matter in further detail in §3. Also, note that the restricted MLE over $\mathcal{G}_d(k,\mathcal{P})$, i.e.

$$\bar{f}_{n,k} := \underset{f\in\mathcal{G}_d(k,\mathcal{P})}{\operatorname{argmax}} \int \log f d\mathbb{P}_n$$

is not computationally tractable Ghosh et al. (2019) – as the class $\mathcal{G}_d(k)$ is far than being convex. Furthermore, even estimating k is a hard problem. Therefore, the log-concave MLE can be viewed as a convex relaxation that can be computed efficiently for estimating a k-log affine density.

For simplicity of the introduction and the presentation of our results, we also consider the sub-class

$$f_0 \in \mathcal{G}_d(k) \iff f_0 \in \mathcal{G}(k,\mathcal{C}_d) \text{ and } \inf_{x\in\mathcal{C}_d} f_0(x) \geq c_d, \tag{3}$$

where $\mathcal{C}_d := [-1/2, 1/2]^d$ is the unit-cube, and $c_d \geq 0$ is a constant that only depends on d. Throughout this text, we refer to a $f_0 \in \mathcal{G}_d(k)$ as a k-log-affine density.

Prior to this work, the adaptive rates for log-concave MLE over $\mathcal{G}(k,\mathcal{P})$ were studied only when $d \leq 3$. First, in the univariate case (i.e. when $d=1$), Kim et al. (2018) proved that

$$\mathcal{R}_{d_X^2}(\widehat{f}_n, \mathcal{G}_d(k)) \lesssim (k/n) \cdot \log(n)^{5/4}, \tag{4}$$

note that $\mathcal{P}$ is not playing a key role here as when $d=1$ (the univariate case) all convex polytope $\mathcal{P}$ are intervals. Next, Feng et al. (2021) proved that when $d \in \{2,3\}$

$$\mathcal{R}_{d_X^2}(\widehat{f}_n, \mathcal{G}_d(k)) \lesssim k^{d-1}/n \cdot \log(n)^{\gamma_d}, \tag{5}$$

where $\gamma_2 = 9/2, \gamma_3 = 8$. Also, they showed adaptation over the sub-class of β-Hölder log-concave densities with $\beta \in (1,3)$, and when $d=3$ for log-concave densities that are bounded away from zero on a polytopal support.

Note that the last two equations imply that for $d \le 3$ and $k = \Theta(1)$, the rate of convergence of $\widehat{f}_n$ is of order $1/n$ (e.g. parametric in n), which is faster than the non-parametric minimax rate of $n^{-\frac{4}{d+4}}$, which implies that $\widehat{f}_n$ adapts to such densities. However, it is important to note that the class $\mathcal{G}_d(k)$ is parametric with order $k \cdot d$ "disjoint" parameters, and hence (cf. (Yang and Barron, 1999))

$$\mathcal{M}_{d_X^2}(\mathcal{G}_d(k)) \underset{d}{\asymp} k/n \tag{6}$$

Therefore, when $d \le 2$, the $\widehat{f}_n$ is minimax optimal both in n and k (up log factors) for estimation $\mathcal{G}_d(k)$, and when $d=3$, the bound of (5) is only optimal in n, but not in k.

Before we present our contributions, one may ask why proving these bounds for $d \le 3$ requires two different papers whose proofs do not extend to dimension $d \ge 4$. The main bottleneck in showing the adaptive rates for $\widehat{f}_n$ as it corresponds to bound a suprema of an unbounded empirical process, which is known as a significantly harder problem (cf. Mendelson (2014); Koltchinskii (2011)). The arguments used in these papers use abstract density estimation results from van de Geer (2000)) and bounds the bracketing entropy numbers (see Def. 1 below) over $\mathcal{G}(k,\mathcal{P})$, and unfortunately do not generalize to $d \ge 4$.

Our contributions: Our main results provide sharp bounds on the adaptation of $\widehat{f}_n$ when $d \ge 4$. We follow a different approach from those employed in Kim et al. (2018); Feng et al. (2021), giving optimal rates and a unified proof for all $d \ge 1$. At a high level, we work with expected suprema over log-concave densities and bound these terms using various techniques from empirical processes and convex stochastic geometry (e.g., random polytopes (Schneider and Weil, 2008)). Specifically, in Theorems 1 and 2 below, we prove that up to the following holds when $d \ge 4$:

$$\mathcal{R}_{d_X^2}(\widehat{f}_n, \mathcal{G}_d(k)) \underset{d}{\asymp} C(k,d,n) n^{-\frac{4}{d}}, \tag{7}$$

where $\log(n)^{-4\frac{d-1}{d}} k^{\frac{4}{d}} \underset{d}{\lesssim} C(k,d,n) \underset{d}{\lesssim} k^{4\frac{d-1}{d}} \log(n)^{12}$.

First, for $k = \Theta(1)$, the above rate $n^{-4/d}$ is even faster than the minimax rate (in total variation squared) of $n^{-4/(d+4)}$ for all $d \ge 4$ (and in particular of Hellinger squared). Thus, $\widehat{f}_n$ adapts to log-affine densities in *all dimensions* $d \ge 1$. However, note that when $d \ge 3$, these adaptivity bounds have a sub-optimal dependency in k.

Secondly, it is interesting that the rate in (7) of order $n^{-4/d}$ is different when $d \ge 5$ from the parametric rate of n^{-1} of $d \le 4$ appearing in the adaptive risk bounds; which can be seen as a shortcoming of the log-concave MLE for $d \ge 5$. Also, it seems to support the lack of *empirical* success of log-concave MLE in high dimensions (cf. Samworth (2018)). We discuss this in further details in §3 and §4 below.

Organization of this paper: In §2, we state our main results. In §3, we discuss the relations between log-concave density estimation to the task of convex regression. In §4, we provide some future work and open problems. In §5, we prove Theorem 1. Finally, in §6, we provide an overview of the proof of Theorem 2. In the appendix below, we prove Theorem 2 and prove all the loose ends remained in §5.

2 Main Results

Theorem 1. *Let $d \ge 4$ and $n \ge d+1$, and assume that $f_0 \in \mathcal{G}(k,\mathcal{P})$ for some integer $k \ge 1$ and polytope $\mathcal{P} \subset \mathbb{R}^d$. Then, the following holds:*

$$\mathbb{E} d_X^2(\widehat{f}_n, f_0) \lesssim C(f_0) \cdot n^{-\frac{4}{d}} \log(n)^{\alpha_d},$$

where $C(f_0)$ is a constant that only depends on f_0, and $\alpha_d \le 4$.

Our proof holds for all $d \ge 1$ and implies the results of Feng et al. (2021); Kim et al. (2018), and in the case of $f_0 \in \mathcal{G}_d(k)$, our proof implies that $C(f_0) \underset{d}{\lesssim} k^{4\frac{d-1}{d}}$. An explicit bound on $C(f_0)$ appears in §B below. Our next result shows that the bound of Theorem 1 is sharp in the number of samples n when $d \ge 4$ and $0 \le k \underset{d}{\lesssim} \sqrt{n}$, for the sub-class $\mathcal{G}_d(k)$. Specifically, we prove the following:

Theorem 2. *Let $d \ge 4$, $n \ge d+1$, and $\beta_d \le 4$. Then, for every $1 \le k \underset{d}{\lesssim} \sqrt{n}$, there exists an underlying density $f_0 \in \mathcal{G}_d(k)$ such that*

$$\mathbb{E} h^2(\widehat{f}_n, f_0) \underset{d}{\gtrsim} k^{\frac{4}{d}} \cdot n^{-\frac{4}{d}} \log(n)^{-\beta_d}.$$

We remark that in the future work of the authors that analyzes the risk of the log-concave MLE in terms of the total variation, we provide proof that removes the redundant log factors in Theorems 1 and 2 when $d \ge 5$. However, it requires novel and significantly more complicated chaining arguments.

3 Discussion

In this part, we provide an overview of convex regression and its relation to the task of log-concave density estimation. To start this discussion, we remind the definition of the ε-entropy numbers

Definition 1. *The ϵ-entropy of $\mathcal{G}$ with respect to the measure $L_p(\mu)$ is defined as the logarithm of the minimal number M such there exists a set $\{g_i\}_{i=1}^M$ that has the following property: For all $g \in \mathcal{G}$ there exists an $1 \leq i \leq M$ such that $\|g - g_i\|_{L_p(\mu)} \leq \epsilon$.*

The tasks of estimating a (regression) convex or a k-max affine function based on observing n noisy measurements are deeply studied for all dimensions $d \geq 1$ in non-parametric statistics (cf. Hannah and Dunson (2013); Balázs et al. (2015); Brunel (2013, 2016) and references within). In the discussion, we consider two classes. First, given a convex set $K \subset \mathbb{R}^d$, we consider the class of 1-uniformly bounded convex functions,

$$f \in \mathcal{F}_1(K) \iff f : K \to \mathbb{R} \text{ convex and } \|f\|_\infty \leq 1.$$

and the class of (1-uniformly bounded) k-max affine functions restricted to a polytope $\mathcal{P} \subset \mathbb{R}^d$

$$f \in \mathcal{G}_1(k, \mathcal{P}) \iff f \in \mathcal{F}_1(\mathcal{P}) \text{ and } k\text{-max affine.}$$

In the task of regression, we observe that $\mathcal{D}^R = \{(Z_i, Y_i)\}_{i=1}^n$, where

$$Z_i \underset{i.i.d}{\sim} \text{Unif}(K) \text{ and } Y_i = f^*(Z_i) + \xi_i,$$

and $\xi_1, \ldots, \xi_n \sim N(0,1)$. When $f^* \in \mathcal{F}_1(K)$ this is the task of (bounded) convex regression; and when $f^* \in \mathcal{G}_1(k, \mathcal{P})$ it is the task of estimating a (bounded) k-max affine function. In this part, the following rates are in terms of $L_2(\text{Unif}(K))$.

First, Gao and Wellner (2017) showed that ε-entropy numbers in terms of $L_2(\text{Unif}(K))$, of bounded convex functions deeply depends on their support $K \subset \mathbb{R}^d$. Specifically, when $d \geq 3$ and the domain is a smooth convex set (for example when K is the Euclidean ball), it is of order $\varepsilon^{-(d-1)}$; and when it is a $K = \mathcal{P}$ is a polytope, it is of order $C(\mathcal{P})\varepsilon^{-d/2}$. However, the ε entropy in terms of $L_1(\text{Unif}(K))$, does not depend on K and equals to $\epsilon^{-d/2}$.

Next, Han and Wellner (2016) used these to study the minimax rate of (bounded) convex regression, and they showed that when K is smooth, the minimax rate is of order $n^{-\frac{2}{d+1}}$, and of order $n^{-\frac{4}{d+4}}$ when K polytope. Note that these rates equal the minimax rates in total variation and Hellinger squared (respectively) in log-concave density estimation. This is not a coincidence; we address this phenomenon in the extended journal version of the manuscript. At a high level, it follows from the fact that comparing to $L_2(\text{Unif}(K))$ error, the $L_1(\text{Unif}(K))$ is less sensitive to the shape domain as the error follows from the boundary (cf. Kur et al. (2020a)).

It is important to *emphasize* that this sensitivity to the shape of the domain forces us to constrain the k-max affine functions (and the k-log-affine densities) to a polytope and not to an arbitrary convex set. Finally, several works Han and Wellner (2016); Kur et al. (2020a) studied the (bouned) convex LS, defined via

$$\bar{f}_n = \underset{f \in \mathcal{F}_1(K)}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - f(Z_i))^2.$$

They showed that $\bar{f}_n$ attains *adaptive* rates of order $C(\mathcal{P}, k)n^{-\min\{4/d,1\}}$ over the sub-class $\mathcal{G}_1(k, \mathcal{P})$ for all $d \geq 1$. Note that these are the same rates of Theorems 1 and 2 above, and the results of Kim et al. (2018); Feng et al. (2021).

The key challenge in our proofs is the following: the sharp ε-entropy bounds and the chaining arguments that led to the adaptive rates of $\bar{f}_n$ cannot be used to study the log-concave MLE. The main reason is that, unlike $\bar{f}_n$, $\log \widehat{f}_n$ is not uniformly bounded over the entire domain $\mathcal{P}$. To handle this obstacle, greater effort and a different approach are required to use these entropy bounds, as we will see in §5 and §6 below.

4 Future Work and Open Problems

First, we denote the class of the isotropic log-concave densities by

$$\mathcal{F}_d^I := \{f \in \mathcal{F}_d : E_f = \mathbf{0}, \Sigma_f = \mathrm{I}_{d\times d}\}, \tag{8}$$

where E_f and Σ_f are the the mean and the covariance matrix of $X \sim f$ respectively. Now, we introduce the definition of the isotropic constant (cf. (Brazitikos et al., 2014, Chp.2)) of

$$I_d := \max_{f \in \mathcal{F}_d^I} \log f(\mathbf{0}) \asymp \max_{f \in \mathcal{F}_d^I, x \in \mathbb{R}^d} \log f(x),$$

where the last inequality was shown by Fradelizi (cf. (Brazitikos et al., 2014, Thm 2.2.2)). In the recent seminal works of Bizeul (2025); Klartag and Lehec (2024); Guan (2024), it was shown that

$$I_d \asymp d. \tag{9}$$

Using that for every $f \in \mathcal{F}_{d,I}$, it holds that $\max_{x\in\mathbb{R}^d} \log f(x) \gtrsim d$ (Brazitikos et al., 2014, Prop 2.3.12), and the last two equations, we obtain that

$$\forall f \in \mathcal{F}_d^I : \max_{x\in\mathbb{R}^d} \log f(x) \asymp \log f(\mathbf{0}) \asymp d. \tag{10}$$

The last equation and the proof of Theorem 2 and the works of Böröczky Jr (2000); Ludwig et al. (2006), lead us to believe that when $d \geq 5$, the error for the $\widehat{f}_n$ is at least $n^{-4/d}$ for any underlying density. Specifically, we conjecture the following:

Conjecture 1. *For any $d \geq 5$, the following holds:*

$$\inf_{f_0 \in \mathcal{F}_d} \mathbb{E} d_{TV}^2(\widehat{f}_n, f_0) \asymp n^{-4/d}.$$

Our conjecture implies that under the slicing conjecture that if $\widehat{f}_n$ has has an error that is lower than some threshold that is independent of d, then it requires more than an exponential number in d of samples. Also that it cannot attain parametric rates when $d \geq 5$ — supporting empirical evidence of the lack of success of $\widehat{f}_n$ in high dimensions.

Next, note that (6) implies that when $d \geq 5$, $\widehat{f}_n$ is a minimax sub-optimal estimator over the class $\mathcal{G}_d(k)$ both in n, k. In the works of Kur and Putterman (2022); Ghosh et al. (2019, 2021), it is shown that for the related task of estimating a k-max affine regression function based on its noisy measurements, there exists a computationally tractable minimax optimal procedure – namely a minimax optimal estimator that has polynomial runtime in n in the number of samples. It is a natural question to ask if such a procedure exists in estimating a k-log affine density.

Conjecture 2. *For every $d \geq 4$ and $n \geq d+1$, there exists an* efficient *estimator $\bar{f}_n$ that is also minimax optimal, i.e.*

$$\mathcal{R}_{d_X^2}(\bar{f}_n, \mathcal{G}_d(k)) \underset{d}{\lesssim} C(k) \cdot \log(n)^{\gamma_d}/n,$$

where $\gamma_d \geq 0$ only depends on d, and $C(k)$ only depends on the number of pieces k.

5 Proof of Theorem 1

Notations: Throughout this text, the constants $c, c_1, c_2 \in (0,1)$ and $C, C_1, C_2 > 0$ are absolute, and the constants $c(d), c_1(d), c_2(d) \in (0,1)$ and $C(d), C_1(d), C_2(d) \ldots \in \mathbb{R}^+$ only depends on d; all these constants that may change from line to line.For any measurable set $A \subset \mathcal{P}$, we denote by $A^c = \mathcal{P} \setminus A$, also $\text{Vol}(A)$ denotes the Lebesgue measure of A, and $\mathbb{P}_0(A) = \int_A f_0 dx$. Recall that $\mathbb{P}_n$ denotes the empirical measure over $X_1, \ldots, X_n$, and we denote by $\widetilde{\mathbb{P}_n} = \mathbb{P}_n - \mathbb{P}_0$.

5.1 Step I: Reducing to almost isotropic log-concave densities

Recall the definition of the class of isotropic log-concave densities in (8). As d_X is affine invariant, we shall assume without loss of generality that $f_0 \in \mathcal{F}_d^I$, i.e., isotropic log-concave. Yet, there are two delicate points for taking an affine transformation when f_0 is k-log max affine.

Note that an affine transformation does not affect the number of pieces $k \in \mathbb{N}$, i.e., it is still a log k-affine density, and it does not change the combinatorial structure of the polytope domain $\mathcal{P}$, i.e., its number of facets and vertices. We shall also assume that $\text{Vol}(\mathcal{P}) \underset{d}{\asymp} 1$, and that $\inf_{x \in \mathcal{P}} f_0(x) \geq c_d$, for some $c_d \in (0,1)$. Later, these assumptions will be relaxed in §B below. Using (10) above (that follows from the definition of the isotropic constant I_d), and our volume assumption on $\mathcal{P}$, and that the density is a bounded from below by c_d, we know that for each $A \subset \mathcal{P}$, $\text{Vol}(A) \underset{d}{\asymp} \mathbb{P}_0(A)$.

Using the results of (Kim and Samworth, 2016, Lemma 16), with a probability of at least $1 - c_1(d)/n$, $\widehat{f}_n$ is "almost" isotropic (i.e., its almost a zero mean and with identity covariance log-concave density). Therefore, for $n \geq C_3(d)$ (for some sufficiently large $C_3(d) \geq 0$), we may use (10) above, and assume that

$$\sup_{x \in \mathcal{P}} \log \widehat{f}_n(x) \asymp d, \tag{11}$$

under a high probability event (and recall (8) above). Furthermore, note that for every realization of data $\mathcal{D}$, $\widehat{f}_n$ is supported on the convex hull of $X_1, \ldots, X_n$, that lies in the interior of $\mathcal{P}$.

Throughout this proof, we assume that we lie in the event of $\widehat{f}_n$ being almost isotropic, which we denote by $\mathcal{E}_1$.

5.2 Step II: Localizing around the Hellinger ball

For each $t \geq 0$, we consider the "localized ball" and "localized shell" around f_0 that are defined via

$$\underbrace{\{f \in \mathcal{F}_d : \log f : \mathcal{P} \to [-\infty, 2I_d] \text{ and } h(f, f_0) \leq t\}}_{:=\mathcal{F}(t)} \text{ and}$$
$$\underbrace{\{f \in \mathcal{F}(t) : h(f, f_0) \geq t/2\}}_{\mathcal{S}(t)},$$

respectively.

Note that for each realization of $\mathcal{D}$ in $\mathcal{E}_1$, we know that $\widehat{f}_n$ is almost isotropic log-concave, and therefore $\widehat{f}_n \in \mathcal{F}(t)$ for some $t \in (0,1)$ that depends on $\mathcal{D}$. We aim to find a high probability event $\mathcal{E}_2 \subset \mathcal{E}_1$, such that $t \leq t_0$, where

$$t_0 = t_0(f_0, n) \underset{d}{\lesssim} C(f_0) \log(n)^6 n^{-\min\{\frac{2}{d}, \frac{1}{2}\}}, \tag{12}$$

where $C(f_0)$ is a constant that only depends on f_0; an explicit bound appears in (26) below. To prove the last equation, it suffices to show that under $\mathcal{E}_2$, it holds for all $t \geq t_0$:

$$\Psi_n(t) = \Psi(n, f_0, t) := \sup_{f \in \mathcal{S}(t)} \int \log \frac{f}{f_0} d\mathbb{P}_n < 0, \tag{13}$$

and therefore by taking $t = t_0, 2t_0, \dots$, we obtain that

$$\sup_{f \in \mathcal{F}(1) \setminus \mathcal{F}(t_0/2)} \int \log \frac{f}{f_0} d\mathbb{P}_n < 0.$$

Using (12)-(13) and the fact of $\int \log \frac{\widehat{f}_n}{f_0} d\mathbb{P}_n \geq 0$, imply that $\widehat{f}_n \in \mathcal{F}(t_0/2)$, and in particular that $h^2(\widehat{f}_n, f_0) \underset{d}{\lesssim} t_0^2$. In order to prove (13), we shall show that $t \geq 0$, it holds that

$$\Psi_n(t) \leq C_d \cdot \max\{t_0 \cdot t, 1/n\} - ct^2. \tag{14}$$

Furthermore, under $\mathcal{E}_2 \subset \mathcal{E}_1$, the last equations imply that

$$\begin{aligned} d_X^2(\widehat{f}_n, f_0) &= \max_{f \in \mathcal{F}_d} \int \log \frac{f}{f_0} d\mathbb{P}_n \\ &= \max_{t \geq 0} \max_{f \in \mathcal{S}(t)} \int \log \frac{f}{f_0} d\mathbb{P}_n \\ &= \max_{t \geq 0} \Psi_n(t) \lesssim \max_{t \geq 0} C_d \cdot t_0 \cdot t - ct^2 \underset{d}{\asymp} t_0^2. \end{aligned}$$

The remaining parts of the proof are dedicated to show (14), which in particular implies Theorem 1. We recall that $\log f_0$ is uniformly bounded by $C_3(d)$ on $\mathcal{P}$; however, $\log \widehat{f}_n$ is *only* bounded from above on $\mathcal{P}$ by $2I_d$, and unbounded from below (for example, outside the convex hull of $X_1, \dots, X_n$ its values are $-\infty$). To overcome this problem, first for each $f \in \mathcal{S}(t)$ and $t \geq 0$, we define the *convex* set

$$K(f) := \{x \in \mathcal{P} : \log f(x) \geq \min_{x \in \mathcal{P}} \log f_0(x) - \bar{C}_d\}, \tag{15}$$

for $\bar{C}_d > 0$ that is large enough constant that only depends on d, that is chosen later. Note that $K(f)$ is convex as it is a log-level set of a log-concave density. As $f \in \mathcal{S}(t)$ is almost isotropic, we may assume that

$$\|\log(f) \mathbf{1}_{K(f)}\|_\infty \lesssim \bar{C}_d,$$

where we used that $\bar{C}_d$ is large enough, and in particular that $\max\{|\log c_d|, I_d\} \leq \bar{C}_d$.

Next, observe that $\mathbb{P}_0(K(f)^c) \underset{d}{\lesssim} t^2$. To see this, recall the definition of $\mathcal{S}(t)$, and we apply a "Chebyshev's inequality" as follows:

$$\begin{aligned} \mathbb{P}_0(K(f)^c) &= \int_{K(f)^c} f_0 dx \underset{d}{\lesssim} \int_{K(f)^c} (\sqrt{f} - \sqrt{f_0})^2 dx \\ &\underset{d}{\lesssim} h^2(f, f_0) \underset{d}{\lesssim} t^2, \end{aligned} \tag{16}$$

where we used that $\max\{\log \widehat{f}_n, \log f_0\} \lesssim I_d \lesssim d$.

To conclude this part, we showed that $K(f)$ is convex and t^2-close (in terms of $\mathbb{P}_0$-measure) to $\mathcal{P}$, and $\log f \mathbf{1}_{K(f)^c}$ is uniformly bounded by $\bar{C}_d$. The next step is the heart of the proof; we show that there exists a *common* polytope $\mathcal{P}_{t^2}$ that is very close to $\mathcal{P}$ and which is contained in $K(f)$ for every $f \in K(f)$. Therefore, all $\log f \mathbf{1}_{\mathcal{P}_{t^2}}$ (where $f \in \mathcal{S}(t)$) are uniformly bounded by $\bar{C}_d$ on $\mathcal{P}_{t^2}$.

5.3 Step III: the common polytope lemma

For every $1 \leq i \leq d-1$, let $\mathcal{F}_i(\mathcal{P})$ denote the number of the i-dimensional faces of $\mathcal{P}$. We introduce two standard quantities from convex geometry: flag number (see e.g., (Reitzner et al., 2019)) and simplicial polytope.

Definition 2 (Flag Number)**.** *The flag number of the polytope $\mathcal{P}$ is denoted by $\mathcal{F}(\mathcal{P})$ and is defined as the total number of flags of $\mathcal{P}$, where a flag of $\mathcal{P}$ is a sequence $F_1 \subset F_2 \subset \ldots \subset F_{d-2} \subset F_{d-1}$ such that F_i is an i-dimensional face of $\mathcal{P}$ for each i.*

Definition 3 (Simplicial Polytope)**.** *A simplicial polytope $\mathcal{P} \subset \mathbb{R}^d$ is a polytope all of whose facets (i.e., $d-1$-dimensional faces) are simplicies, i.e. each facet is a convex hull of d points.*

We aim to find a polytope set $\mathcal{P}_{t^2} \subset \mathcal{P}$ that is "close as possible" to $\mathcal{P}$ as possible (in terms of measure), and $\mathcal{P}_{t^2}$ has the combinatorial properties of $\mathcal{P}_{t^2}$; and furthermore $\log f \in \mathcal{S}(t)$ *restricted* to $\mathcal{P}_{t^2}$ are uniformly bounded from above and *below*. We shall introduce the definition of the ϵ-floating body (cf. (Schütt and Werner, 1990, pg. 276) or the seminal works of Vu (2005); Dwyer (1988))[2]

Definition 4. *Let $\epsilon \in (0,1)$, and $K \subset \mathbb{R}^d$ be a convex body with volume one. The ϵ-floating body of K, is a convex set that is defined as*

$$K_\epsilon^F := \bigcap_{L \in \mathcal{K}_d(K) : \mathrm{Vol}(L) \geq 1-\epsilon} L,$$

where $\mathcal{K}_d(K)$ denotes the convex sets in $\mathbb{R}^d$ that are contained in K.

The classical results of (Bárány and Buchta, 1993, Thm. 2) and (Bárány and Larman, 1988, Thm. 1) imply that when $K = \mathcal{P}$ is a polytope, then the following holds:

$$\mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_\epsilon^F) \underset{d}{\asymp} (1 + O(\log(\epsilon)^{-1})) \cdot C(\mathcal{P}) \log(1/\epsilon)^{d-1} \epsilon$$

where $C(\mathcal{P}) = \mathcal{F}(\mathcal{P}) \cdot ((d+1)^{d-1}(d-1)!)^{-1}$.

Therefore, we may conclude the following: Any convex set $L \subset \mathcal{P}$, that is ϵ-close $\mathcal{P}$ contains a common *convex* body $\mathcal{P}_\epsilon^F$ that is $C(\mathcal{P}) \log(1/\epsilon)^{d-1} \epsilon$ close to $\mathcal{P}$. However, $\mathcal{P}_\epsilon^F$ may be far from being a polytope. Yet, by

[2] Here we present a different formulation of the ϵ-floating body that is equivalent to the classical one, see Lemma 11 in §B below for completeness.

using the probabilistic method and the rich literature of the random polytopes, we prove the following:

Lemma 1. *Let $\mathcal{P} \subset \mathbb{R}^d$ be a polytope with volume one and $\epsilon \geq 0$. Then, there exists a simplicial polytope $\mathcal{P}_\epsilon \subset \mathcal{P}_\epsilon^F$ such that it both holds that*

$$\mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_\epsilon) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^2 \cdot \log(1/\epsilon)^{2(d-1)}\epsilon$$

and for $1 \leq i \leq d-1$, it holds that

$$\mathcal{F}_i(\mathcal{P}_\epsilon) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P}) \cdot \log(1/\epsilon)^{d-1}.$$

Namely, we show that there exists $\mathcal{P}_\epsilon \subset \mathcal{P}_\epsilon^F$, that is $\mathcal{F}(\mathcal{P})^2 \log(1/\epsilon)^{2(d-1)}\epsilon$-close to $\mathcal{P}$ and has $\mathcal{F}(\mathcal{P}) \cdot \log(1/\epsilon)^{(d-1)}$ i-facets. Namely, this polytope has almost the same combinatorial structure and volume as $\mathcal{P}$.

Proof. Let $k := \epsilon^{-1}$, and $Z_1, \ldots, Z_m \underset{i.i.d}{\sim} \mathrm{Unif}(\mathcal{P})$, and consider the random polytope $P_m = \mathrm{conv}\{Z_i\}_{i=1}^m$. The following is proven in (Bárány and Buchta, 1993, Thm. 2):

$$\mathbb{E}\left[\mathrm{Vol}(\mathcal{P} \setminus P_m)\right] \underset{d}{\asymp} \mathcal{F}(\mathcal{P}) \cdot \log(m)^{d-1}/m, \tag{17}$$

and for all $1 \leq i \leq d-1$

$$\mathbb{E}\mathcal{F}_i(P_m) \underset{d}{\asymp} m \cdot \mathbb{E}\left[\mathrm{Vol}(\mathcal{P} \setminus P_m)\right] \underset{d}{\asymp} \mathcal{F}(\mathcal{P}) \cdot \log(m)^{d-1}, \tag{18}$$

where the expectation is taken over $Z_1, \ldots, Z_m$. Now, we set

$$m \underset{d}{\asymp} \frac{k}{\mathcal{F}(\mathcal{P}) \cdot \log(k)^{d-1}} \underset{d}{\asymp} \frac{1}{\mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_\epsilon^F)}.$$

By using that by independence of $Z_1, \ldots, Z_m$, and that the convex hull of if and only if all vertices lie in

$$\begin{aligned}\Pr(P_m \subset \mathcal{P}_{k^{-1}}^F) &= (1 - \Pr(Z \notin \mathcal{P} \setminus \mathcal{P}_{k^{-1}}^F))^m \\ &\underset{d}{\gtrsim} (1 - \mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_{k^{-1}}^F))^{C_d \mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_{k^{-1}}^F)^{-1}} \\ &\geq \exp(-C_1(d)),\end{aligned} \tag{19}$$

where we used that $(1-x)^{a/x} = (1+O(x)) \cdot e^{-a}$ as $x \to 0$ (or equivalantly as $k \to \infty$).

By using the last equation and Markov's inequality, we know that with constant probability, the random polytope $P_m \subset \mathcal{P}_{k^{-1}}^F$ satisfies the right-hand sides of (17)-(18) (up to a multiplicative constant that only depends on d). Therefore, the claim follows by choosing a P_m in this event and noting that P_m is a simplicial polytope with probability one, for instance see (Ludwig et al., 2006, Eq. (9)). $\square$

Though the proof of this lemma is relatively simple, and surprisingly, it does not appear in the literature, we believe it may be independent of interest.

5.4 Step IV: Bounding the "averaged log-likelihood" of $\widehat{f}_n$

Recall the definition of $K(f)$ in (15), and that

$$\mathbb{P}_0(K(f)^c) \underset{d}{\asymp} \mathrm{Vol}(K(f)^c) \underset{d}{\asymp} t^2.$$

Now, we may apply Lemma 1 with $\epsilon \underset{d}{\asymp} t^2$, it gives us a simplicial polytope $\mathcal{P}_{t^2}$ such that $\mathcal{P}_{t^2} \subset K(f)$ for all $f \in \mathcal{S}(t)$, and furthermore, we have that

$$\mathbb{P}_0(\mathcal{P} \setminus \mathcal{P}_{t^2}) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^2 t^2 \log(1/t)^{2(d-1)},$$

where the first equality follows from $\mathrm{Vol} \underset{d}{\asymp} \mathbb{P}_0$, and the later for the definition of $\mathcal{P}_{t^2}$.

Most importantly, recall that all $f \in \mathcal{S}(t)$ are *uniformly* bounded from both sides by constant that only depends on d on $\mathcal{P}_{t^2}$ — justifying its name the common polytope. Now, we define the "averaged log-likelihood" of f on some measurable set $A \subset \mathcal{P}$ by

$$s_n(f, A) := \int_A \log(f/f_0) d\mathbb{P}_n.$$

Also for every $\mathcal{H} \subset \{\mathcal{P} \to \mathbb{R}\}$, we define

$$G_n(\mathcal{H}) := \sup_{f \in \mathcal{H}} \int f d\widetilde{\mathbb{P}_n}. \tag{20}$$

Also, let

$$\bar{B}_t := \{\log(f/f_0) \cdot \mathbf{1}_{K(f) \setminus P_{t^2}} : f \in \mathcal{S}(t)\}. \tag{21}$$

Now, we state the following lemma:

Lemma 2. *Let $t \geq 0$, an denote by*

$$\bar{s}_n(f) := s_n(f, K(f)^c) + G_n(\bar{B}_t).$$

Then, the following holds

$$\bar{s}_n(f) \underset{d}{\lesssim} \bar{C}_d \cdot \left(t_1 \cdot t^{\max\{2 \cdot \frac{d-1}{d+1}, 1\}} - \mathbb{P}_0(K(f)^c)\right), \tag{22}$$

with a probability of at least $1 - \exp(-cnt)$ uniformly $f \in \mathcal{S}(t)$, where

$$t_1 := t_1(n, \mathcal{P}) \underset{d}{\lesssim} C_1(\mathcal{P}) \cdot n^{-\min\{\frac{2}{d+1}, \frac{1}{2}\}} \cdot \log(n)^{2d}, \tag{23}$$

and $C_1(\mathcal{P})$ can be a upper bounded by

$$\mathcal{F}(\mathcal{P})^2 \mathbf{1}_{d \geq 4} + \mathcal{F}_{d-1}(\mathcal{P})^{\frac{d}{2}} \mathbf{1}_{d \in \{2,3\}} + \mathbf{1}_{d=1}.$$

In other words, it means that the "averaged log-likelihood" of $\widehat{f}_n$ is affected by the points in $\mathcal{D}$ that lie in the interior of $\mathcal{P}$ rather than its boundary. The proof of this lemma appears in §B below. We remark that it is based on L_1-chaining type arguments (cf. Han (2021)) and known entropy bounds. Next, we denote by

$$B_t := \{\log(f/f_0) \cdot \mathbf{1}_{\mathcal{P}_{t^2}} : f \in \mathcal{S}(t)\}. \tag{24}$$

Now, we state the following lemma:

Lemma 3. *For all $t \geq 0$, then the following holds with probability of $1 - C\exp(-cnt^2)$:*

$$G_n(B_t) \underset{d}{\lesssim} \bar{C}_d \cdot t_0 \cdot t \tag{25}$$

where

$$t_0 := C(f_0) \cdot n^{-\min\{\frac{1}{2}, \frac{2}{d}\}} \log(n)^6,$$

and $C(f_0)$ can upper be bounded by

$$\mathcal{F}(\mathcal{P})^{4/d} k^{2\frac{d-1}{d}} \cdot \mathbf{1}_{d \geq 4} + k^{\frac{d-1}{2}} \cdot \mathcal{F}_{d-1}(\mathcal{P}) \cdot \mathbf{1}_{d \in \{2,3\}} + \sqrt{k} \cdot \mathbf{1}_{d=1}. \tag{26}$$

The proof of this lemma appears in §5.6 below. We remark that the proof is based on known entropy bounds and by apply the the classical Ossiander's bound (cf. Pollard (2002)) which provides an upper bound on suprema of an empirical process over a *unfiormly bounded* class; see Lemma 5 in the Appendix A below for completeness. Finally, note that for $f \in \mathcal{S}(t)$, it holds that

$$KL(f_0||f) = \int_{\mathcal{P}} \log \frac{f_0}{f} d\mathbb{P}_0 \geq h^2(f, f_0) \geq t^2/4.$$

Yet, for our analysis, we need to consider the KL-divergence to $K(f)$, i.e $\int_{K(f)} \log \frac{f}{f_0} d\mathbb{P}_0$. Using standard calculations, we prove in §B that the following holds:

Lemma 4. *Let $t \in (0, c_d)$. Then, the following holds for all $f \in \mathcal{S}(t)$:*

$$\int_{K(f)} \log \frac{f}{f_0} d\mathbb{P}_0 \leq -\frac{t^2}{16} + C_4(d) \cdot \mathbb{P}_0(K(f)^c). \tag{27}$$

5.5 Step V: Concluding the theorem

Now, fix some $t \in (C_d n^{-1/2} \log(n)^{2d}, 1)$, and consider the case of $d \geq 4$. Also, we assume that we are in the event defined by all the previous steps. Therefore, by combining Lemmas 3 and 4, we obtain that the $s_n(f, \mathcal{P})$ is upper bounded by

$$\begin{aligned}
&G_n(B_t) + G_n(\bar{B}_t) + \int_{K(f)} \log \frac{f}{f_0} d\mathbb{P}_0 + s_n(f, K(f)^c) \\
&\underset{d}{\lesssim} \bar{C}_d \cdot (t_0 t + t_1 t^{\frac{2\cdot(d-1)}{d+1}}) - \frac{t^2}{16} - (\bar{C}_d - C_1(d))\mathbb{P}_0(K(f)^c) \\
&\underset{d}{\lesssim} \bar{C}_d \cdot \left(t_0 t + t_1 t^{\frac{2\cdot(d-1)}{d+1}}\right) - \frac{t^2}{16},
\end{aligned}$$

where in the last inequality we set $\bar{C}_d \geq C_1(d)$ to be large enough. Now, recall the definitions of $t_0 = t_0(n, f_0), t_1 = t_1(n, \mathcal{P})$ above, and note that

$$t_0 t \underset{d}{\gtrsim} t_1 t^{\frac{2(d-1)}{d+1}} \iff t \underset{d}{\gtrsim} n^{-1/2} \log(n)^{2d}.$$

Therefore, by using the last two equations, we conclude that for any fixed $t \geq t_4 := C_d n^{-1/2} \log(n)^{-2d}$, it holds

$$\Psi_n(t) = \sup_{f \in \mathcal{S}(t)} s_n(f, \mathcal{P}) \leq C_3(d) \cdot t_0 \cdot t - t^2/16,$$

with probability of $1 - \exp(-cnt)$. Note that $t \leq t_4$, then we know that

$$\sup_{f \in \mathcal{F}(t_4)} s_n(f, \mathcal{P}) \leq G_n(B_{t_4}) + G_n(\bar{B_{t_4}}) \underset{d}{\lesssim} n^{-1}.$$

Hence, (14) follows. Namely, when $d \geq 4$ we obtained with a high probability that

$$d_X^2(\widehat{f}_n, f_0) \lesssim C(f_0)^2 n^{-\frac{4}{d}} \log(n)^{12},$$

where $C(f_0)$ is defined in Lemma 3, and the theorem follows. Repeating this step, when $d \leq 3$, would recover (up to log-factors in n), the results of (Feng et al., 2021; Kim et al., 2018), for completeness; see §B.

5.6 Proof of Lemma 3

First, recall the definition ϵ-entropy with bracketing (see Definition 6 in the appendix), and we denote by $\|\cdot\|_{\mathbb{P}_0}$ to be the $L_2(\mathbb{P}_0)$ norm, and for every $\mathcal{G} \subset L_2(\mathbb{P}_0)$ denote by $\mathrm{D}(\mathcal{G}) := \sup_{g_1, g_2 \in \mathcal{G}} \|g_1 - g_2\|_{\mathbb{P}_0}$. Now, we state the classical Ossiander's bound (cf. Pollard (2002); van de Geer (2000)):

Lemma 5. *Let $\mathbb{P}$ by a probability distribution, and $\mathcal{G} \subset L_2(\mathbb{P})$ be uniformly bounded function class (i.e. $\sup_{g \in \mathcal{G}} \|g\|_\infty \lesssim 1$). Denote by $\epsilon_* = \epsilon_*(n)$, to be stationary point of*

$$\epsilon\sqrt{n} \asymp \int_{\epsilon}^{\mathrm{D}(\mathcal{G})} \sqrt{\log \mathcal{N}_{2,[]}(u, \mathcal{G}, \mathbb{P}_0)} du. \tag{28}$$

Then, $\sup_{g \in \mathcal{G}} |\int g d(\mathbb{P}_n - \mathbb{P})| \lesssim \epsilon_*$ with probability of at least $1 - C\exp(-cn\epsilon_*^2/\mathrm{D}(\mathcal{G}))^2)$.

Next, let $\mathcal{F}(\mathcal{Q}, \Gamma)$ be all the convex functions supported on a polytope $\mathcal{Q} \subset \mathbb{R}^d$ that are uniformly bounded by Γ. For any $r \geq 0$, denote by

$$B_{\mathcal{Q},\Gamma}(f, r) := \{g \in \mathcal{F}(\mathcal{Q}, \Gamma) : \int (f - g)^2 dx \leq r^2\}.$$

The following lemma that is proven in Kur et al. (2020a):

Lemma 6. *For any $0 \leq \epsilon \leq r$ and f be a max affine function that is supported on some polytope $\mathcal{Q} \subset \mathbb{R}^d$ with volume one. Then,*

$$\begin{aligned}
&\log \mathcal{N}_{2,[]}(\epsilon, B_{\mathcal{Q},\Gamma}(f, r), \mathrm{Vol}) \\
&\qquad \underset{d}{\lesssim} \triangle(f, \mathcal{Q}) \cdot (r/\epsilon)^{\frac{d}{2}} \log(\max\{\Gamma, \epsilon^{-1}\})^{\alpha_d}
\end{aligned} \tag{29}$$

where $\alpha_d \leq d + 1$, $\triangle(f, \mathcal{Q})$ is the triangulation number of f w.r.t. $\mathcal{Q}$ (see Def. 7 in §A).

First, we upper bound $\triangle(\log f_0, P_{t^2})$ in a naive way. When $d = 1$, as $\mathcal{P}_{t^2}$ is an interval, we obtain that $\triangle(\log f_0, P_{t^2}) \lesssim k$. When $d = 2$, it is easy to note that any k max affine function is also $2k$-simplicial, a by simple combinatorial fact, and also the $\mathcal{F}(\mathcal{P}) \asymp \mathcal{F}_1(\mathcal{P})$. Therefore, as each piece of the max affine can intersect at most $\mathcal{F}_1(\mathcal{P})$ facets, we conclude that

$$\triangle(\log f_0, P_{t^2}) \lesssim k \cdot \mathcal{F}(\mathcal{P})^2 \cdot \log(1/t).$$

By induction, we obtain that for all $d \geq 3$,

$$\triangle(\log f_0, P_{t^2}) \lesssim k^{d-1} \cdot \mathcal{F}(\mathcal{P})^2 \cdot \log(1/t)^{d-1}.$$

Now, recall that $\log f \in B_t$ (see (24) above) is uniformly bounded over $\mathcal{P}_{t^2}$ by $\bar{C}_d$, and $\mathbb{P}_0$ is bounded from above by I_d. Hence, we obtain that

$$\begin{aligned}\int_{\mathcal{P}_{t^2}} (\log f - \log f_0)^2 dx &\lesssim_d \|\log f - \log f_0\|_{\mathbb{P}_0}^2 \\ &\lesssim_d \bar{C}_d^{\,2} h^2(f, f_0) \lesssim_d \bar{C}_d^2 t^2.\end{aligned} \tag{30}$$

Now, Lemma 6 and the last equation imply $B_t \subset B_{\mathcal{P}_{t^2}, \bar{C}_d}(f_0, \bar{C}_d t)$ and the lemma implies:

$$\begin{aligned}&\log \mathcal{N}_{2,[]}(\epsilon, B_t, \mathbb{P}_0|_{\mathcal{P}_{t^2}}) \lesssim_d \\ &\triangle(\log f_0, \mathcal{P}_{t^2}) \left(\frac{\bar{C}_d t}{\epsilon}\right)^{\frac{d}{2}} \log(\max\{\epsilon^{-1}, \bar{C}_d\})^{d+1},\end{aligned} \tag{31}$$

where

$$\triangle(\log f_0, P_{t^2}) \lesssim_d k^{d-1} \mathcal{F}(\mathcal{P})^2 \cdot \log(n)^{2(d-1)+2} 1_{d \geq 2} + k.$$

Finally, it is straighten forward to see that Lemma 5, the entropy bound of the last equation, and our triangulation bound imply that to show that for $d \geq 1$ and for any fixed $t \in (t_1, 1)$ that with probability of at least $1 - n^{-10}$,

$$\begin{aligned}&\sup_{f \in \mathcal{S}(t)} \left| \int_{P_{t^2}} \log \frac{f}{f_0} d\widetilde{\mathbb{P}_n} \right| \lesssim_d \\ &\bar{C}_d \cdot \triangle(\log f_0, P_{t^2})^{2/d} \cdot n^{-\min\{\frac{2}{d}, \frac{1}{2}\}} \log(n)^{2 \cdot \frac{d+1}{d}} t,\end{aligned} \tag{32}$$

for completeness see in §B.3.2 below.

6 Overview on the proof of Theorem 2

Consider the case of $d \geq 5$, and recall the definitions of B_t and $G_n(\cdot)$ in (20),(24), and consider the sub-class $\mathcal{G}_d(k)$. In Theorem 1, we showed that

$$G_n(B_t) \lesssim_d k^{2 \cdot \frac{d-1}{d}} n^{-\frac{2}{d}} \log(n)^6 t.$$

In Theorem 2, we construct a $f_0 \in \mathcal{G}_d(k)$ such that the last equation is "tight", i.e.

$$G_n(B_t) \gtrsim_d k^{\frac{2}{d}} \cdot n^{-\frac{2}{d}} \log(n)^{\frac{2(d-1)}{d}} t := t_2 t.$$

This bound and following somehow classical ideas that appear in Birgé and Massart (1993); Birgé (2006); Chatterjee (2014), imply Theorem 2.

To explain the key challenge of proving the last equation, we state some notation and definitions. Let $w_n(\mathcal{H})$ denote the Rademacher complexity of $\mathcal{H} \subset \{\mathcal{C}_d \to \mathbb{R}\}$, i.e.

$$w_n(\mathcal{H}) = \mathbb{E} \sup_{f \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(X_i) \right|,$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. Rademacher random variables; and the expectation is taken over $\{(X_i, \epsilon_i)\}_{i=1}^n$. The classical symmetrization and de-symmetrization lemmas states that for any $\mathcal{H} \subset \{\mathcal{C}_d \to \mathbb{R}\}$ (cf. (Koltchinskii, 2011, Thm 2.1)) it holds:

$$2^{-1} w_n(\mathcal{H}^c) \leq \mathbb{E} \sup_{f \in \mathcal{H}} \left| \int f d(\mathbb{P}_n - \mathbb{P}_0) \right| \leq 2 w_n(\mathcal{H}),$$

where $\mathcal{H}^c := \{f - \int f d\mathbb{P}_0 : f \in \mathcal{H}\}$. The key problem is that the absolute value is *essential* for the desymmetrization inequality to be valid. That is, one may construct a pathological function class $\mathcal{H}$, and a distribution $\mathbb{P}_0$ such that

$$\mathbb{E} \sup_{f \in \mathcal{H}} \int f d(\mathbb{P}_n - \mathbb{P}_0) \ll w_n(\mathcal{H}).$$

However, since $\widehat{f}_n$ maximizes $\int_{\mathcal{C}_d} \log f d\mathbb{P}_n$, the absolute value matters, and therefore, we cannot lower the bound $G_n(B_t)$ by the Rademacher complexity of $w_n(B_t)$. The key challenge of Theorem 2 is to show that for all $t \in (0, t_2)$

$$w_n(B_t) \asymp_d \mathbb{E} G_n(B_t).$$

We call such an inequality a *positive*-desymmetrization inequality.

The proof of this claim is based on constructive proof that deeply relies on the geometry of log-concave densities and our underlying f_0. We remark that in convex regression, this problem does not appear as the noise is symmetric, and sharp estimates for the Gaussian or Rademacher complexity can be obtained via Sudakov's inequality and Dudley's integral (cf. (Koltchinskii, 2011, Thms 3.1 and 3.2)) via the ε-entropy of the class.

Acknowledgments

GK would like to acknowledge his PhD advisor, **Prof. Alexander Rakhlin**, for his valuable discussions. Currently, GK is supported by the **ETH Computer Science Department**. The majority of GK's work was supported by the **National Science Foundation (NSF)** and the **Simons Foundation** through the Collaboration on the Theoretical Foundations of Deep Learning, under awards **DMS-2031883** and **#814639**. AG's work was supported by **NSF grant DMS-2210504**.

References

An, M. Y. (1997). Log-concave probability distributions: Theory and statistical testing. *Duke University Dept of Economics Working Paper*.

Artstein-Avidan, S., Giannopoulos, A., and Milman, V. D. (2015). *Asymptotic geometric analysis, Part I*, volume 202. American Mathematical Soc.

Axelrod, B., Diakonikolas, I., Stewart, A., Sidiropoulos, A., and Valiant, G. (2019). A polynomial time algorithm for log-concave maximum likelihood via locally exponential families. *Advances in Neural Information Processing Systems*, 32.

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic theory*, 26(2):445–469.

Balázs, G. (2022). Adaptively partitioning max-affine estimators for convex regression. In *International Conference on Artificial Intelligence and Statistics*, pages 860–874. PMLR.

Balázs, G., György, A., and Szepesvári, C. (2015). Near-optimal max-affine estimators for convex regression. In *Artificial Intelligence and Statistics*, pages 56–64. PMLR.

Balcan, M.-F. and Long, P. (2013). Active and passive learning of linear separators under log-concave distributions. In *Conference on Learning Theory*, pages 288–316.

Bárány, I. and Buchta, C. (1993). Random polytopes in a convex polytope, independence of shape, and concentration of vertices. *Mathematische Annalen*, 297(1):467–497.

Bárány, I. and Larman, D. G. (1988). Convex bodies, economic cap coverings, random polytopes. *Mathematika*, 35(2):274–291.

Birgé, L. (2006). Model selection via testing: an alternative to (penalized) maximum likelihood estimators. In *Annales de l'IHP Probabilités et statistiques*, pages 273–325.

Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields*, 97(1-2):113–150.

Bizeul, P. (2025). The slicing conjecture via small ball estimates. *arXiv preprint arXiv:2501.06854*.

Böröczky Jr, K. (2000). Approximation of general smooth convex bodies. *Advances in Mathematics*, 153(2):325–341.

Brazitikos, S., Giannopoulos, A., Valettas, P., and Vritsiou, B.-H. (2014). *Geometry of isotropic convex bodies*, volume 196. American Mathematical Soc.

Bronshtein, E. (1976). ε-entropy of convex sets and functions. *Siberian Mathematical Journal*, 17(3):393–398.

Brunel, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electronic Journal of Statistics*, 7:1301–1327.

Brunel, V.-E. (2016). Adaptive estimation of convex and polytopal density support. *Probability Theory and Related Fields*, 164(1-2):1–16.

Carpenter, T., Diakonikolas, I., Sidiropoulos, A., and Stewart, A. (2018). Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. In *Conference On Learning Theory*, pages 1234–1262.

Chatterjee, S. (2014). A new perspective on least squares under convex constraint. *The Annals of Statistics*, 42(6):2340–2381.

Cule, M., Samworth, R., and Stewart, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(5):545–607.

Doss, C. R. and Wellner, J. A. (2016). Global rates of convergence of the mles of log-concave and s-concave densities. *Annals of statistics*, 44(3):954.

Dwyer, R. A. (1988). On the convex hull of random points in a polytope. *Journal of Applied Probability*, 25(4):688–699.

Feng, O. Y., Guntuboyina, A., Kim, A. K. H., and Samworth, R. J. (2021). Adaptation in multivariate log-concave density estimation. *The Annals of Statistics*, 49(1):129 – 153.

Gao, F. and Wellner, J. A. (2017). Entropy of convex functions on $\mathbb{R}^d$. *Constructive approximation*, 46(3):565–592.

Ghosh, A., Pananjady, A., Guntuboyina, A., and Ramchandran, K. (2019). Max-affine regression: Provable, tractable, and near-optimal statistical estimation. *arXiv preprint arXiv:1906.09255*.

Ghosh, A., Pananjady, A., Guntuboyina, A., and Ramchandran, K. (2021). Max-affine regression: Parameter estimation for gaussian designs. *IEEE Transactions on Information Theory*, 68(3):1851–1885.

Guan, Q. (2024). A note on bourgain's slicing problem. *arXiv preprint arXiv:2412.09075*.

Han, Q. (2021). Set structured global empirical risk minimizers are rate optimal in general dimensions. *The Annals of Statistics*, 49(5):2642–2671.

Han, Q. and Wellner, J. A. (2016). Multivariate convex regression: global risk bounds and adaptation. *arXiv preprint arXiv:1601.06844*.

Hannah, L. A. and Dunson, D. B. (2013). Multivariate convex regression with adaptive partitioning. *The Journal of Machine Learning Research*, 14(1):3261–3294.

Kabatiansky, G. A. and Levenshtein, V. I. (1978). On bounds for packings on a sphere and in space. *Problemy Peredachi Informatsii*, 14(1):3–25.

Kim, A. K., Guntuboyina, A., and Samworth, R. J. (2018). Adaptation in log-concave density estimation. *The Annals of Statistics*, 46(5):2279–2306.

Kim, A. K. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *The Annals of Statistics*, 44(6):2756–2779.

Klartag, B. and Lehec, J. (2024). Affirmative resolution of bourgain's slicing problem using guan's bound. *arXiv preprint arXiv:2412.15044*.

Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. Springer Science & Business Media.

Kur, G., Dagan, Y., and Rakhlin, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. *arXiv preprint arXiv:1903.05315*.

Kur, G., Gao, F., Guntuboyina, A., and Sen, B. (2020a). Convex regression in multidimensions: Suboptimality of least squares estimators. *arXiv preprint arXiv:2006.02044*.

Kur, G. and Putterman, E. (2022). An efficient minimax optimal estimator for multivariate convex regression. In *Conference on Learning Theory*, pages 1510–1546. PMLR.

Kur, G., Rakhlin, A., and Guntuboyina, A. (2020b). On suboptimality of least squares with application to estimation of convex bodies. In *Conference on Learning Theory*, pages 2406–2424. PMLR.

Lee, Y. T. and Vempala, S. S. (2018). Stochastic localization+ stieltjes barrier= tight bound for log-sobolev. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1122–1129. ACM.

Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358.

Ludwig, M., Schütt, C., and Werner, E. (2006). *Approximation of the Euclidean ball by polytopes*. Citeseer.

Mendelson, S. (2014). Learning without concentration. In *Conference on Learning Theory*, pages 25–39.

Paouris, G. (2006). Concentration of mass on convex bodies. *Geometric & Functional Analysis GAFA*, 16(5):1021–1049.

Pollard, D. (2002). Maximal inequalities via bracketing with adaptive truncation. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, pages 1039–1052. Elsevier.

Reitzner, M., Schuett, C., and Werner, E. (2019). The convex hull of random points on the boundary of a simple polytope. *arXiv preprint arXiv:1911.05917*.

Samworth, R. J. (2018). Recent progress in log-concave density estimation. *Statistical Science*, 33(4):493–509.

Schneider, R. and Weil, W. (2008). *Stochastic and integral geometry*. Springer Science & Business Media.

Schuhmacher, D. and Dümbgen, L. (2010). Consistency of multivariate log-concave density estimators. *Statistics & probability letters*, 80(5-6):376–380.

Schütt, C. and Werner, E. (1990). The convex floating body. *Mathematica Scandinavica*, pages 275–290.

Schütt, C. and Werner, E. (2003). Polytopes with vertices chosen randomly from the boundary of a convex body. In *Geometric aspects of functional analysis*, pages 241–422. Springer.

Stanley, R. P. (1989). Log-concave and unimodal sequences in algebra, combinatorics, and geometry. *Annals of the New York Academy of Sciences*, 576(1):500–535.

Tsybakov, A. B. (2003). *Introduction to Nonparametric Estimation*. Springer.

van de Geer, S. A. (2000). *Empirical Processes in M-estimation*, volume 6. Cambridge university press.

Vu, V. H. (2005). Sharp concentration of random polytopes. *Geometric & Functional Analysis GAFA*, 15(6):1284–1318.

Yang, Y. and Barron, A. (1999). Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, pages 1564–1599.

A Definitions and Preliminaries

A.1 Notations

$\|\cdot\|_\mu$ denotes the $L_2(\mu)$ norm. $\mathcal{K}_d$ denotes the collection of convex sets in $\mathbb{R}^d$.We also use the notation of $f(n) = O(g(n))$ meaning $f(n) \lesssim g(n)$, and $f(n) = O_d(g(n))$, meaning $f(n) \underset{d}{\lesssim} g(n)$ ($\Omega, \Omega_d, \Theta, \Theta_d$ are defined in a similar fashion), finally $f(n) = \tilde{O}_d(g(n))$ meaning $f(n) \underset{d}{\lesssim} \log(n)^{C_d} g(n)$ for some $C_d \geq 0$. Also, recall that $\widetilde{\mathbb{P}_n} = \mathbb{P}_n - \mathbb{P}_0$.

A.2 Definitions

Definition 5. *Let $\triangle(\mathcal{P})$ be the triangulation* number *of a polytope $\mathcal{P} \subset \mathbb{R}^d$, that is the minimal number such that*

$$\bigcup_{i=1}^{M} S_i = \mathcal{P} \textit{ where each } S_i \subset \mathcal{P} \textit{ is a simplex}$$

The entropy with bracketing is defined as follows:

Definition 6. *The ε-entropy with bracketing numbers of $\mathcal{G}$ w.r.t. to the measure $L_p(\mu)$ as the logarithm of the minimal number M such there exists a set $\{(g_{i,-}, g_{i,+})\}_{i=1}^{M}$ that has the following property: For all $g \in \mathcal{G}$ there exists an $1 \leq i \leq M$ such that $g_{i,-} \leq g \leq g_{i,+}$ and $\|g_{i,-} - g_{i,+}\|_{L_p(\mu)}$.*

We denote the ϵ-entropy with bracketing by $\log \mathcal{N}_{[],p}(\epsilon, \mathcal{G}, \mu)$. Next, we define a k-simplicial function:

Definition 7. *Let $\mathcal{P} \subset \mathbb{R}^d$ be a polytope, and a convex function $h : \mathcal{P} \to \mathbb{R}$ is said to be k-simplicial when it can be presented as*

$$\forall i \in 1, \ldots, k \quad h : S_i \to \mathbb{R} \textit{ is an affine function and } \mathcal{P} = \bigcup_{i=1}^{k} S_i,$$

where $S_1, \ldots, S_k$ are simplicies in $\mathbb{R}^d$. We denote this minimal $k \geq 0$; by $\triangle(h, \mathcal{P})$.

A.3 Preliminaries

We state two classical lemmas, first known as the ε-entropy bound for convex sets due to Bronshtein (1976). The second one is a standard combining entropy lemma (cf. Gao and Wellner (2017))

Lemma 7. *Let $K \subset \mathbb{R}^d$ be a convex body. Then, the following holds for every $0 < \epsilon \leq \mathrm{Vol}(K)$:*

$$\log \mathcal{N}_{1,[]}(\epsilon, \mathcal{K}_d(K), \mathrm{Vol}) \lesssim d^{9/2} (\mathrm{Vol}(K)/\epsilon)^{(d-1)/2}.$$

Lemma 8. *Let $\epsilon > 0$, and $\epsilon_1, \ldots, \epsilon_k > 0$ such that $\sum_{i=1}^{k} \epsilon_i = \epsilon$. Assume that $K = \bigcup_{i=1}^{k} K_i \subset \mathbb{R}^d$, then, the following holds:*

$$\log \mathcal{N}_{1,[]}(\epsilon, \mathcal{K}_d(K), \mathrm{Vol}) \leq \sum_{i=1}^{k} \log \mathcal{N}_{1,[]}(\epsilon_i, \mathcal{K}_d(K_i), \mathrm{Vol}).$$

The following is proven in Kur et al. (2019); Han (2021).

Lemma 9. *Let $\mathcal{A}$ be a collection of sets, and let $\sigma^2 = \mathbb{P}_0(\bigcup_{A \in \mathcal{A}} A)$. Denote by $\epsilon_* = \epsilon_*(n)$ to be minimizer of*

$$\epsilon + \frac{c}{\sqrt{n}} \int_{\epsilon}^{\sigma^2} u^{-1/2} \sqrt{\log \mathcal{N}_{1,[]}(u, \mathcal{A}, \mathbb{P}_0)} du.$$

Then, with probability of at least $1 - C \exp(-cn\epsilon_*)$, it holds that

$$\sup_{A \in \mathcal{A}} |\widetilde{\mathbb{P}_n}(A)| \lesssim \epsilon_*.$$

Finally, we denote by $B_{\mathrm{Vol}}(\mathcal{P}, r) := \{K \in \mathcal{K}_d : \mathrm{Vol}(\mathcal{P} \triangle K) \leq r\}$. The following lemma is proven in (Kur et al., 2020b) (the proof is mainly based on Lemma 6).

Lemma 10. *Let $\mathcal{P} \subset \mathbb{R}^d$ be a polytope with volume one. Then, for any $0 < \epsilon \leq r \leq 1$, the following holds:*

$$\log \mathcal{N}_{1,[]}(\epsilon, B_{\mathrm{Vol}}(\mathcal{P}, r), \mathrm{Vol}) \underset{d}{\lesssim} \triangle(\mathcal{P}) \cdot \left(\frac{r}{\epsilon}\right)^{\frac{d-1}{2}} \log(\epsilon^{-1})^d,$$

where $\triangle(\mathcal{P})$ is the triangulation number of $\mathcal{P}$.

B Missing parts of Theorem 1

B.1 Proof of Lemma 2

Recall the definition of $\mathcal{S}(t)$, and the convex set, defined for every $f \in \mathcal{S}(t)$ by

$$K(f) := \{x \in \mathcal{P} : \log f(x) \geq \underbrace{\min_{x \in \mathcal{P}} \log f_0(x)}_{:=-u_d} - \bar{C}_d\},$$

where $\bar{C}_d$ is chosen in Step V in the proof.

Now, we prove that the following holds For any $t \in (C_1(d) n^{-1/2} \log(n)^d, 1)$, the following holds with a probability of at least $1 - n^{-1}$

$$\int_{K(f)^c} \log \frac{f}{f_0} d\mathbb{P}_n \underset{d}{\lesssim} \epsilon_*^2 - c_1(d) \bar{C}_d \cdot \mathbb{P}_0(K(f)^c), \tag{33}$$

and

$$\int_{K(f) \setminus \mathcal{P}_{t^2}} \log \frac{f}{f_0} d\widetilde{\mathbb{P}_n} \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^{\frac{4d}{d+1}} \cdot \epsilon'_*, \tag{34}$$

where ϵ_* and ϵ'_* are defined in (35) and (39) respectively.

By combining the last two equations, the lemma follows. It remains to prove (33),(34).

Proof of (33). Note that as $f \in \mathcal{S}(t)$, we know that $K(f) \in B_{\mathrm{Vol}}(\mathcal{P}, O_d(t^2))$, where B_{Vol} is defined in §A above. To see this, recall that in Step II, we showed that $K(f)$ is convex and its volume $O_d(t^2)$ close to $\mathcal{P}$. Therefore, by Lemma 10 and that $\mathbb{P}_0 \underset{d}{\asymp} \mathrm{Vol}$, we obtain that

$$\log \mathcal{N}_{1,[]}(\epsilon, \{K(f) : f \in \mathcal{S}(t)\}, \mathbb{P}_0) \underset{d}{\lesssim} \triangle(\mathcal{P}) \cdot \left(\frac{t^2}{\epsilon}\right)^{\frac{d-1}{2}} \log(\epsilon^{-1})^d, \tag{35}$$

where $\triangle(\mathcal{P})$ is the triangulation number of $\mathcal{P}$.

By applying an L_1-chaining argument of Lemma 9, with the entropy bound of the last equation, it follows that

$$\epsilon_* := \epsilon_*(t, n, \mathcal{P}) \underset{d}{\asymp} \left(\frac{\triangle(\mathcal{P})}{n}\right)^{\frac{2}{d+1}} t^{2 \cdot \frac{d-1}{d+1}} \log(n)^{\frac{2d}{d+1}} 1_{d \geq 4} + \sqrt{\frac{\triangle(\mathcal{P})}{n}} \log(n)^{d/2} t 1_{d \leq 3}, \tag{36}$$

for completeness, he proof of this equation appears in §B.3.1 below. Using the last equation, and definition of $K(f)$, we obtain that uniformly for all $f \in \mathcal{S}(t)$ with probability $1 - \exp(-cn\epsilon_*)$ it holds

$$\begin{aligned} \int_{K(f)^c} \log \frac{f}{f_0} d\mathbb{P}_n &\leq -\frac{\bar{C}_d}{2} \cdot \mathbb{P}_0(K(f)^c) + \mathbb{E} \sup_{f \in \mathcal{S}(t)} |\widetilde{\mathbb{P}_n}(K(f)^c)| \\ &\geq -\frac{\bar{C}_d}{2} \cdot \mathbb{P}_0(K(f)^c) + \epsilon_*. \end{aligned} \tag{37}$$

where we used in the first inequality that $\bar{C}_d$ is large enough, and the claim follows. □

Proof of (34). For each $t \geq n^{-1/2}$, we define the collection of sets

$$\mathcal{L}(t) := \{\{\log f \geq u\} \setminus \mathcal{P}_{t^2} : f \in \mathcal{S}(t), -(\bar{C}_d + u_d) \leq u \leq I_d\}.$$

Below, we prove the following entropy bound:

$$\log \mathcal{N}_{1,[]}(\epsilon, \mathcal{L}(t), \mathbb{P}_0) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^{2d} \log(1/\epsilon)^{2(d-1)^2+2(d-1)} (t^2/\epsilon)^{(d-1)/2}. \tag{38}$$

Then, using Lemma 9 with the above entropy bound, we obtain that (in similar fashion as in the proof of (36)) that

$$\mathbb{E} \sup_{S\in\mathcal{L}(t)} |\widetilde{\mathbb{P}_n}(S)| \underset{d}{\lesssim} \epsilon'_* \underset{d}{\asymp} \mathcal{F}(\mathcal{P})^{\frac{4d}{d+1}} \log(n)^{\frac{4d}{d+1}} t^{2\cdot\frac{d-1}{d+1}} n^{-\frac{2}{d+1}} \mathbf{1}_{d\geq 4} + \mathcal{F}(\mathcal{P})^d n^{-\frac{1}{2}} \log(n)^3 t \mathbf{1}_{d\leq 3}. \tag{39}$$

Now, we are ready to conclude the lemma. By using the tail's formula and that each $\log f \in \log \mathcal{S}(t)$ is uniformly from above by I_d and below by $-2\bar{C}_d$ on $K(f)$, we obtain

$$\begin{aligned}
\mathbb{E} \sup_{f\in\mathcal{S}(t)} \int \log f \mathbf{1}_{K(f)^c\setminus\mathcal{P}_{t^2}} d\widetilde{\mathbb{P}_n} &\leq \int_{-2\bar{C}(d)}^{I_d} \mathbb{E} \sup_{f\in\mathcal{S}(t)} |\widetilde{\mathbb{P}_n}(\{\log f \geq u\} \setminus \mathcal{P}_{t^2}\})| du \\
&\leq \int_{-2\bar{C}(d)}^{I_d} \mathbb{E} \sup_{S\in\mathcal{L}(t)} |\widetilde{\mathbb{P}_n}(S)| du \\
&\underset{d}{\lesssim} \bar{C}_d \cdot \epsilon'_*,
\end{aligned} \tag{40}$$

where we used that $\bar{C}_d \gtrsim I_d$.

Next, as $-\log f_0$ is uniformly bounded and convex, we may apply the tail again as we did in the last equation, and obtain that

$$\mathbb{E} \sup_{f\in\mathcal{S}(t)} \left| \int \log f_0 \mathbf{1}_{K(f)^c\setminus\mathcal{P}_{t^2}} d\widetilde{\mathbb{P}_n} \right| \underset{d}{\lesssim} \max\{I_d, u_d\} \cdot \mathbb{E} \sup_{S\in\mathcal{S}(t)} |\widetilde{\mathbb{P}_n}|(S) \underset{d}{\lesssim} \bar{C}_d \epsilon'_*, \tag{41}$$

where we used our assumption that $\bar{C}_d \underset{d}{\gtrsim} \max\{I_d, u_d\}$. Finally, the high probability bound follows from Talagrands's bound (cf. Koltchinskii (2011)), i.e.

$$\sup_{f\in\mathcal{S}(t)} \left| \int \log f_0 \mathbf{1}_{K(f)^c\setminus\mathcal{P}_{t^2}} d\widetilde{\mathbb{P}_n} \right| \underset{d}{\lesssim} \epsilon'_*$$

with probability of at least $1 - C\exp(-cnt^{-2}\epsilon'^2_*)$, and the claim follows. It remains to prove (38). □

Proof of (38). Recall that $\log f$ is a concave function, but restricted to $\{\log f \geq u\} \setminus \mathcal{P}_{t^2}$ which is not convex. Still, it is a difference with a convex set with $\mathcal{P}_{t^2}$ that satisfies

$$\mathcal{F}_{d-1}(\mathcal{P}_{t^2}) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^2 \log(n)^{2(d-1)}.$$

Therefore, $\{\log f \geq u\} \setminus \mathcal{P}_{t^2}$ can be presented as a union of at most $\mathcal{F}_{d-1}(\mathcal{P}_{t^2})$-convex sets (which we denote by $K_1, \ldots, K_{\mathcal{F}_{d-1}(\mathcal{P}_{t^2})}$) that their combined volume is at most

$$\mathcal{F}_{d-1}(\mathcal{P}_{t^2}) \cdot \mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_{t^2}) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^4 \log(n)^{4(d-1)} t^2,$$

where we used that $\mathcal{F}_{d-1}(\mathcal{P}_{t^2}) \underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^2 \log(n)^{2(d-1)}$.

Using the last observation, we apply the classical Bronstien's entropy bound for convex sets (Bronshtein, 1976), and a standard "combining" entropy bound (Gao and Wellner, 2017) (Lemmas 7 and 8 respectively above), we obtain

$$\begin{aligned}
\log \mathcal{N}_{1,[]}(\epsilon, \mathcal{L}(t), \mathbb{P}_0) &\underset{d}{\lesssim} \sum_{j=1}^{\mathcal{F}_{d-1}(\mathcal{P}_{t^2})} \log \mathcal{N}_{1,[]}\Big(\frac{\epsilon \,\mathrm{Vol}(K_j)}{\mathrm{Vol}(\mathcal{P} \setminus \mathcal{P}_{t^2})}, \mathcal{K}_d(K_j), \mathrm{Vol}\Big) \\
&\underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^2 \log(1/\epsilon)^{2(d-1)} (t^2 \mathcal{F}(\mathcal{P})^4 \log(1/\epsilon)^{4(d-1)}/\epsilon)^{\frac{(d-1)}{2}} \\
&\underset{d}{\lesssim} \mathcal{F}(\mathcal{P})^{2d} \log(1/\epsilon)^{2(d-1)^2+2(d-1)} (t^2/\epsilon)^{(d-1)/2},
\end{aligned} \tag{42}$$

where we used that $\mathbb{P}_0 \underset{d}{\asymp} \text{Vol}$, and (38) follows. Therefore, the claim follows. We remark that with much more effort, this bound can be improved to

$$\mathcal{F}(\mathcal{P})^{d+1} \log(1/\epsilon)^{(d-1)^2+(d-1)}(t^2/\epsilon)^{(d-1)/2}$$

as the logarithmic factors do not play a key role in our analysis, we do not pursue this claim. $\square$

B.2 Proof of Lemma 4

Recall the definition of the isotropic constant, $I_d \geq 0$ above. Let $N_f = \int_{K(f)} f dx$, and consider the normalized densities $\bar{f}_0 := \frac{f_0 \mathbf{1}_{K(f)}}{N_{f_0}}$, and $\bar{f} := \frac{f \mathbf{1}_{K(f)}}{N_f}$. Using that both f, f_0 are almost isotropic we know that $N_f - 1 \lesssim I_d \cdot \mathbb{P}_0(K(f)^c)$, and by applying the classical identity of $h^2(p,q) \leq KL(p||q)$, for $p = \bar{f}_0$ and $q = \bar{f}$, we obtain

$$\begin{aligned}
h^2(\bar{f}_0, \bar{f}) &\leq \int_{K(f)} \log \frac{\bar{f}_0}{\bar{f}} d\frac{\mathbb{P}_0}{N_{f_0}} \\
&= \frac{1}{N_{f_0}} \cdot \left(\int_{K(f)} \log \frac{f_0}{f} d\mathbb{P}_0 + \log\left(\frac{N_f}{N_{f_0}}\right) \right) \\
&\leq (1 + CI_d\mathbb{P}_0(K(f)^c)) \left(\int_{K(f)} \log \frac{f_0}{f} d\mathbb{P}_0 + C_1 I_d \mathbb{P}_0(K(f)^c) \right) \\
&\leq (1 + C_2 I_d \mathbb{P}_0(K(f)^c)) \cdot \left(\int_{K(f)} \log \frac{f_0}{f} d\mathbb{P}_0 \right)
\end{aligned}$$

where we used $\frac{1}{1-x} = 1 + x + O(x^2)$ and $\log(1+x) = x + O(x^2)$, when $x \in (0,1)$; and lastly that the densities are upper bounded by $2I_d$.

Next, note that

$$h^2(f, f_0) \leq h^2(\bar{f}, \bar{f}_0) + C_1 I_d \mathbb{P}_0(K(f)^c).$$

To see this,

$$h^2(f_0, f) \leq \int_{K(f)} (\sqrt{f_0} - \sqrt{f})^2 dx + CI_d \leq (1 + N_f) h^2(\bar{f}, \bar{f}_0) + CI_d \leq (1 + C_1 I_d) h^2(\bar{f}, \bar{f}_0)$$

where we used that the Hellinger squared is homogenous in the normalization constant, and that $N_f - 1 \lesssim I_d \cdot \mathbb{P}_0(K(f))$. Combining the last two equations, we obtain

$$\begin{aligned}
\int_{K(f)} \log \frac{f_0}{f} d\mathbb{P}_0 &\geq h^2(\bar{f}_0, \bar{f}) - 4I_d \mathbb{P}_0(K^c(f)) \\
&\geq h^2(f_0, f) - 8I_d \cdot \mathbb{P}_0(K(f)^c) \\
&\geq t^2 - 8I_d \cdot \mathbb{P}_0(K(f)^c),
\end{aligned}$$

where we used the fact that $\mathbb{P}_0 \underset{d}{\asymp} \text{Vol}$, and the definition of $\mathcal{S}(t)$.

B.3 Loose Ends

B.3.1 Proofs of (36) and (39)

Recall that above we showed

$$\log \mathcal{N}_{1,[]}(\epsilon, \{K(f) : f \in \mathcal{S}(t)\}, \mathbb{P}_0) \underset{d}{\lesssim} \triangle(\mathcal{P}) \cdot \left(\frac{t^2}{\epsilon}\right)^{\frac{d-1}{2}} \log(\epsilon^{-1})^d.$$

Now, by using Lemma 9, we obtain that with probability of at least $1 - \exp(-cn\epsilon_*)$ that

$$\sup_{f \in \mathcal{S}(t)} \widetilde{\mathbb{P}_n}(K(f)) \underset{d}{\lesssim} \epsilon_*$$

where ϵ_* minimizes

$$\epsilon + \frac{c}{\sqrt{n}} \int_\epsilon^{t^2} u^{-1/2} \sqrt{\triangle(\mathcal{P}) \cdot \left(\frac{t^2}{u}\right)^{\frac{d-1}{2}} \log(u^{-1})^d du}.$$

Note that $u \cdot u^{-1/2} \cdot u^{-(d-1)/4}$ diverges as $\epsilon \to 0$ when $d \geq 4$, and therefore

$$\frac{1}{\sqrt{n}} \int_\epsilon^{t^2} u^{-1/2} \sqrt{\triangle(\mathcal{P}) \cdot \left(\frac{t^2}{u}\right)^{\frac{d-1}{2}} \log(u^{-1})^d du} \asymp \epsilon^{1/2} \cdot \sqrt{\triangle(\mathcal{P}) \cdot \left(\frac{t^2}{\epsilon}\right)^{\frac{d-1}{2}} \log(\epsilon)^d}$$

i.e., we need blalace

$$\epsilon \underset{d}{\asymp} \frac{\triangle(\mathcal{P}) \cdot \left(\frac{t^2}{\epsilon}\right)^{\frac{d-1}{2}} \log(1/\epsilon)^d}{n}$$

Therefore, we obtain

$$\epsilon_* \underset{d}{\asymp} t^{2 \cdot \frac{d-1}{d+1}} \left(\frac{\triangle(\mathcal{P})}{n}\right)^{\frac{2}{d+1}} \log(n)^{\frac{2d}{d+1}}.$$

For $d \leq 3$, it is easy to see that

$$\frac{1}{\sqrt{n}} \int_0^{t^2} u^{-1/2} \sqrt{\triangle(\mathcal{P}) \cdot \left(\frac{t^2}{u}\right)^{\frac{d-1}{2}} \log(u^{-1})^d du} \asymp \sqrt{\frac{\triangle(\mathcal{P}) \log(1/t)^d}{n}} \cdot t$$

and (36) follows. Similarly, (39) follows by repeating the same arguments with the entropy bound

$$\mathcal{F}(\mathcal{P})^{2d} \log(1/\epsilon)^{2(d-1)^2+2(d-1)} (t^2/\epsilon)^{(d-1)/2}.$$

B.3.2 Proof of (32)

By using Lemma 5 with the entropy bound of (31), we need to minimize

$$\epsilon + \underbrace{\frac{c}{\sqrt{n}} \int_\epsilon^{\bar{C}_d t} \sqrt{\triangle(\log f_0, \mathcal{P}_{t^2}) \left(\frac{\bar{C}_d t}{u}\right)^{\frac{d}{2}} \log(u^{-1})^{d+1} du}}_{(*)}.$$

When $d \geq 4$, the integral diverges, hence we need to balance $(*) \asymp \epsilon$, which corresponds to the equation

$$\sqrt{\frac{\triangle(\log f_0, \mathcal{P}_{t^2}) \left(\frac{\bar{C}_d t}{\epsilon}\right)^{\frac{d}{2}} \log(1/\epsilon)^{d+1}}{n}} \cdot \epsilon \log(1/\epsilon)^{1_{d=4}} \asymp \epsilon \iff$$
$$\epsilon \underset{d}{\asymp} \bar{C}_d \cdot (\triangle(\log f_0, \mathcal{P}_{t^2})/n)^{2/d} \log(n)^{\frac{d+1}{2d}} \log(n)^{1_{d=4}} t$$

and when $d \leq 3$ the integral converges, we set $\epsilon = 0$, and obtain that

$$(*) \asymp \sqrt{\frac{\triangle(\log f_0, \mathcal{P}_{t^2}) \left(\frac{\bar{C}_d t}{\epsilon}\right)^{\frac{d}{2}} \log(1/\epsilon)^{d+1}}{n}} \cdot \epsilon \asymp \epsilon \iff \epsilon \underset{d}{\asymp} \bar{C}_d \cdot (\triangle(\log f_0, \mathcal{P}_{t^2})/n)^{1/2} \log(n)^{\frac{d+1}{2d}} t$$

and the claim follows.

B.3.3 Floating body

We use the notation of $H^-_{x,t} = \{z \in \mathbb{R}^d, \langle x, z \rangle \leq t\}$, and $H^+_{x,t} = \{z \in \mathbb{R}^d, \langle x, z \rangle \geq t\}$. The following is a classical definition of the floating body (cf. Vu (2005)):

Definition 8. *Let $\epsilon \in (0,1)$, and $K \subset \mathbb{R}^d$ be a convex body with volume one. The ϵ-floating body of K, is a convex set that is defined as*

$$K^F_\epsilon := \bigcap_{\mathrm{Vol}(H^+_{x,t} \cap K) = 1-\epsilon} (H^-_{x,t} \cap K).$$

In the following lemma, we connect the definition from the main body to the classical one.

Lemma 11.

$$\bigcap_{L\in\mathcal{K}_d(K):\mathrm{Vol}(L)\geq 1-\epsilon} L = \bigcap_{\mathrm{Vol}(H^+_{x,t}\cap K)=1-\epsilon} (H^-_{x,t}\cap K), \tag{43}$$

Proof. First, note that $K\cap H^+_{x,t}$ is a convex set, we know that

$$\bigcap_{L\in\mathcal{K}_d(K):\mathrm{Vol}(L)\geq 1-\epsilon} L \subseteq \bigcap_{\mathrm{Vol}(H^+_{x,t}\cap K)=1-\epsilon} (H^-_{x,t}\cap K),$$

and note that any $L\subset K$ such that $\mathrm{Vol}(L)\geq 1-\epsilon$, can be presented as

$$L = \bigcap_{\mathrm{Vol}(H^+_{x,t}\cap K)\geq 1-\epsilon} (H^-_{x,t}\cap K)$$

and therefore

$$\bigcap_{\mathrm{Vol}(H^+_{x,t}\cap K)\geq 1-\epsilon} (H^-_{x,t}\cap K) = \bigcap_{\mathrm{Vol}(H^+_{x,t}\cap K)= 1-\epsilon} (H^-_{x,t}\cap K) \subseteq \bigcap_{L\in\mathcal{K}_d(K):\mathrm{Vol}(L)\geq 1-\epsilon} L,$$

where we used the continuity of the volume of the caps in t.

□

B.3.4 Relaxing the uniform lower bound and the volume assumption

In this part, we only assume that

$$-\log f_0(x) = \left(\max_{1\leq i\leq k} a_i^\top x + b_i\right)\cdot \mathbf{1}_{\mathcal{P}}(x),$$

where f_0 is not bounded from below, and $\mathcal{P}$ can be an arbitrary polytope or polyhedron in $\mathbb{R}^d$.

We follow an approach that originally appeared in Kur et al. (2019) to estimate the minimax rate of log-concave density estimation in total variation. We provide a loose bound up to a $\log(n)^{\gamma_d}$ factor, where $\gamma_d = O(d)$ with a simple proof. To remove this redundant term, one should carefully peel the log-concave densities into shells. However, it is more technical and does not provide any major insight.

The main idea is to show that we may reduce the problem to show that we can always assume

$$-\log(n) \underset{d}{\lesssim} \log f_0(x)\mathbf{1}_{\mathcal{P}'} \leq 2I_d$$

where $\mathcal{P}'\subset\mathbb{R}^d$ is a polytope such that $\mathrm{Vol}(\mathcal{P}')\underset{d}{\lesssim}\log(n)^d$, and that

$$\mathcal{F}_{d-1}(\mathcal{P}') \leq k + \mathcal{F}_{d-1}(\mathcal{P}) + 2d.$$

Then, one could repeat the proof of Theorem 1, with these changes and observe that the bound by $\log(n)^{\gamma_d}$-factor, and the dependency in $\mathcal{P}$.

Now, we provide this reduction. First , as we may assume that f_0 is isotropic as in Step I. We know that all $f\in\mathcal{S}(t)$ are uniformly bounded from above.

Now, we remind a classical result in high dimensional geometry. First, any isotropic log-concave density, satisfies the following for $t\gtrsim 1$:

$$\mathbb{P}_0(\|x\|_2 \geq C\sqrt{d}t) \leq \exp(-c\sqrt{d}t) \tag{44}$$

this is shown in the seminal work of (Paouris, 2006, Thm 1.1) and later in (Lee and Vempala, 2018, Thm. 9). It implies that with probability of $1-n^{-10}$ that it holds:

$$\mathrm{conv}\{X_i\}_{i=1}^n \subset \mathcal{P}' := \{\log f_0 \geq n^{-2}\}\cap\mathcal{P}\cap C(d)\log(n)\mathcal{C}_d. \tag{45}$$

Proof of (44)**.** , That identity of $\|\cdot\|_\infty \le \|\cdot\|_2 \le \sqrt{d}\|\cdot\|_\infty$, and (44) imply that with probability of at least $1-n^{-10}$ that

$$\mathrm{conv}\{X_i\}_{i=1}^n \subset \mathcal{P} \cap C(d)\log(n)\cdot\mathcal{C}_d.$$

To see this, note that sufficiently large $C \ge 0$, it holds

$$\mathbb{P}_0(\|x\|_\infty \ge Cd\log(n)) \le n^{-100}.$$

Now, as f_0 is supported on $\mathcal{P}$, we have that

$$\mathrm{Vol}(C_d\log(n)\mathcal{C}_d \cap \mathcal{P}) \underset{d}{\lesssim} \log(n)^d$$

implies that

$$\mathbb{P}_0(\{f_0(x) \le n^{-100}\} \cap C_d\log(n)\mathcal{C}_d \cap \mathcal{P}) \underset{d}{\lesssim} \mathrm{Vol}(\log(n)\cdot\mathcal{C}_d)\cdot n^{-100} \underset{d}{\lesssim} n^{-99},$$

where we used that the density $\mathbb{P}_0$ is bounded from above by I_d; therefore (45) follows. □

Concluding the reduction: Finally, using (45), and

$$\mathcal{L}(t) = \{\log f_0 \ge t\} = \mathcal{P} \cap \bigcap_{i=1}^k \{a_i^\top x + b_i \le \log t\},$$

we may assume that with high probability

$$\mathrm{conv}\{X_i\}_{i=1}^n \subset \bigcap_{i=1}^k \{a_i^\top x + b_i \le 100\log(n)\} \cap C_d\log(n)\mathcal{C}_d \cap \mathcal{P} := \mathcal{P}'$$

and clearly $\mathrm{Vol}(\mathcal{P}') \underset{d}{\lesssim} \log(n)^d$, and the claim follows.

C Proof of Theorem 2

C.1 Step I: Construction of the underlying densities

First, define $f_G(x)$ as the Gaussian measure γ_d, which is restricted to the unit cube, i.e .

$$f_G(x) := \gamma_d(\mathcal{C}_d)^{-1}\cdot\gamma_d(x)\cdot\mathbf{1}_{\mathcal{C}_d}(x).$$

Then, we consider its logarthim:

$$\log f_G(x) := -\left(\sum_{i=1}^d x_i^2 + \log N_d\right)\cdot\mathbf{1}_{\mathcal{C}_d}(x) = -\left(\sum_{i=1}^d x_i^2\mathbf{1}_{|x_i|\le 1/2} + \log N_d\right),$$

where N_d is the normalization constant. It is easy to see that this density is defined by d independent random variables with zero mean, and that $\Sigma_{f_G} = \lambda_d I_{d\times d}$, where $\lambda_d \in (0,1)$. Therefore, f_G is almost isotropic. The independency of the coordinates implies that $N_d = (N_1)^d$, therefore a simple volume considerations shows that

$$\|\log f_G(x)\mathbf{1}_{\mathcal{C}_d}\|_\infty \lesssim d, \tag{46}$$

For each $k \ge (Cd)^{d/2}$, we approximate $\log f_G$ by a concave function $\log f_0$ that is supported on the unit cube $\mathcal{C}_d$, that has the following properties:

1. $\|\log f_0 - \log f_G\|_1 \underset{d}{\asymp} \|\log f_0 - \log f_G\|_2 \underset{d}{\asymp} k^{-2/d}$.
2. $\|\log f_0 - \log f_G\|_\infty \underset{d}{\lesssim} 1$.
3. $\log f_0$ is k-simplicial on a domain $\mathcal{D}_k \subset \mathcal{C}_d$ such that $\mathbb{P}_0(\mathcal{C}_d \setminus \mathcal{D}_k) \underset{d}{\lesssim} \log(k)^{d-1}\cdot k^{-\frac{d+2}{d}}$.
4. The diameter of all simplicies is at most $O_d((\log k/k)^{1/d})$.

The existence of such a paper follows from many classical results (Bronshtein, 1976; Bárány and Larman, 1988; Schütt and Werner, 2003); see also (Kur and Putterman, 2022, Thm 10).

For simplicity of our analysis, we consider the Gaussian restricted to $\mathcal{D}_k$. For $k \underset{d}{\gtrsim} 1$ large enough, we know that the normalized f_G over $\mathcal{D}_k$ is "almost" identical to f_G. With some ambiguity in the notation, we shall assume f_G is normalized Gaussian over $\mathcal{D}_k$.

C.2 Step II: Positive score around f_G

Our proof is based on the following on the following lemma:

Lemma 12. *With probability of at least $1 - C\exp(-cn)$, there exists a (random) concave function $\log h_n$ such that*

- $\int_{\mathcal{D}_k} \log \frac{f_G}{h_n} d\mathbb{P}_0 \underset{d}{\asymp} n^{-2/d}$.
- $\forall X_1, \ldots, X_n : \quad \log h_n(X_i) = \log f_G(X_i)$.
- $\| \log \frac{f_G}{h_n} \mathbf{1}_{\mathcal{D}_k} \|_\infty \underset{d}{\asymp} n^{-\frac{2}{d}}$

We emphasize that h_n log-concave but not a *density* of a probability distribution. up to a normalization of $1 + O_d(n^{-\frac{2}{d}})$. Next, we denote the event of the lemma by A, using the the fact that $c \le \frac{d}{dx}\mathbb{P}_0(x) \le Cd$ for all $x \in \mathcal{D}_k$, we obtain that for $n \underset{d}{\gtrsim} 1$ that is large enough

$$\begin{aligned}
\mathbb{E} \int \log h_n d\widetilde{\mathbb{P}_n} &= \mathbb{E}\left[\int \log f_G d\mathbb{P}_n - \int \log h_n d\mathbb{P}_0 | A\right] - \mathbb{E}\left[\int \log f_G d\mathbb{P}_n - \int \log h_n d\mathbb{P}_0 | A^c\right] \\
&= \int \log(f_G/h_n) d\mathbb{P}_0 + e^{-c_d n} \underset{d}{\gtrsim} \exp(-c_3 d) n^{-2/d},
\end{aligned}$$

where we first used the fact $h_n = f_G$ on the n points $X_1, \ldots, X_n$, then we used the first item and the high probability bound of the lemma. It remains to prove Lemma 12.

Proof of Lemma 12. Assume that $n \ge (Cd)^{d/2}$ and let $B_2(c, r)$ to be the Euclidean ball with center c and radius r in $\mathbb{R}^d$. Using the classical bound on packing numbers (cf. (Kabatiansky and Levenshtein, 1978) or (Artstein-Avidan et al., 2015, Chp. 4)) there is a collection of balls

$$\mathcal{A}_d := \{B_2(x_1, s(d)n^{-1/d}), \ldots, B_2(x_n, s(d)n^{-1/d})\},$$

that satisfies:

$$\forall 1 \le i < j \le n \;\; B_2(x_i, s(d)n^{-1/d}) \cap B_2(x_j, s(d)n^{-1/d}) = \emptyset$$

where $s(d)$ is defined in way that $\text{Vol}(B_2(x_i, s(d)n^{-1/d})) = (C \cdot I_d)^{-d} n^{-1}$, and note that these balls lie in $\mathcal{D}_k$. Now, recall that $\log f_0(x) \lesssim I_d \lesssim d$ for all $x \in \mathcal{D}_k$. Hence, by the independence of $X_1, \ldots, X_n$, we obtain that

$$\forall 1 \le i \le n \quad \Pr(\nexists j \text{ such that } X_j \in B(x_i, s(d)n^{-1/d})) \le (1 - C^{-d}n^{-1})^n \le e^{-C^d} < 1/10$$

Therefore, by linearity of expectation, we know that

$$\mathbb{E}[\sum_{i=1}^n \mathbf{1}_{\exists X_j \in B_2(x_i, s(d))}] \le (1 - C^{-d}) \cdot n \le n/3.$$

Using Mcdiarmid's inequality, we know that with a probability of at least $1 - \exp(-c_1(d)n)$, there exists a *random* subset of balls $\mathcal{S} \subset \mathcal{A}_d$ of the cardinality of at least $n/2$, that do not contain any $X_1, \ldots, X_n$. Without loss of generality, we consider $\mathcal{S}$ to be the balls with the centers $X_1, \ldots, X_{n/2}$. Now, we follow the strategy of the proof of (Gao and Wellner, 2017, Thm 4.). We denote the epigraph of $\log f_G : \mathcal{D}_k \to \mathbb{R}$ by

$$E_{f_G} := \text{epigraph}(\log f_G) \subset \mathcal{D}_k \times \mathbb{R}.$$

Now, for every $x_i \in \mathcal{D}_k$, we choose the tangent plane to $(x_i, \log f_G(x_i))$ in epigraph, and choose the maximal $\lambda_i \ge 0$ such that

$$\begin{aligned}
C_i &:= \text{Prj}_{\mathbb{R}^d} \left(\{z \in E_{f_G} : (1 - \lambda_i) \cdot (-\|x_i\|_2^2 + \log N_d) \le (-x_i, 1)^\top \cdot z \le (-\|x_i\|_2^2 + \log N_d)\}\right) \\
&\subseteq B_2(x_i, s(d)n^{-1/d}),
\end{aligned}$$

where we used that $\nabla \log f_G = \mathbf{x}$.

In other words, we truncate the maximal cap C_i from the epigraph of $\log f_G$ (that is the paraboloid restricted to the cube) such that its projection to $\mathcal{D}_k$ is contained in $B_2(x_i, s(d)n^{-1/d})$. Most importantly, note that $C_i \cap C_j = \emptyset$ for all $1 \leq i < j \leq n/2$ as their projection is disjoint.

As $\log f_G$ is a paraboloid restricted to $\mathcal{D}_k$ is a , we remove $n/2$ disjoint caps from the Euclidean ball in $\mathbb{R}^{d+1}$ with volume of $c_2^{-d} n^{-1+2/d}$ — which follows via standard volume estimation as the Hessian of the paraboloid is the identity matrix (cf. Artstein-Avidan et al. (2015)), and therefore the height of each cap is of order $c_3 n^{-2/d}$. We define $\log h_n$ by the epigraph of the concave function

$$\bigcap_{i=1}^{n/2} \{z \in E_{f_G} : (-x_i, 1)^\top \cdot z \leq (1-\lambda_i) \cdot (-\|x\|_2^2 + \log N_d)\}$$

Therefore, we removed from the epigraph of $\log f_G$ a volume of $\Theta_d(n^{-\frac{2}{d}})$, and clearly $\log h_n(X_i) = \log f_G(X_i)$ for all $1 \leq i \leq n$, as we removed caps that their projection to $\mathcal{D}_k$ do not contain $X_1, \ldots, X_n$. Clearly, we have that $\log f_G \geq \log h_n$, and we can conclude that

$$\int_{\mathcal{D}_k} \log(f_G/h_n) d\mathbb{P}_0 \underset{d}{\asymp} n^{-2/d},$$

where we used that $\mathbb{P}_0 \underset{d}{\asymp} \mathrm{Vol}$, and the claim follows. □

C.3 Step III: Lower bound for the score function

First, recall that we defined $\log f_k$ in a way that

$$\| \log \frac{f_0}{f_G} \|_1 \underset{d}{\asymp} \| \log \frac{f_0}{f_G} \|_2 \underset{d}{\asymp} k^{-\frac{2}{d}}.$$

Now, for each $t \in (0, c_2(d)k^{-2/d})$, define

$$\log h_n(t) := \log f_0 + k^{2/d} t (\log h_n - \log f_0),$$

and denote by $\widetilde{h_n(t)}$ the log-concave density that is induced by normalizing the log-concave function $h_n(t) : \mathcal{D}_k \to \mathbb{R}^+$. Note that by Taylor expansion of $\log(\cdot)$ and by the uniform boundedness of the densities, we obtain that

$$\begin{aligned} KL(f_0 || \widetilde{h_n(t)}) &\underset{d}{\lesssim} \int (f_0 - \widetilde{h_n(t)})^2 d\mathbb{P}_0 \underset{d}{\lesssim} \int (\log f_0 / \widetilde{h_n(t)})^2 d\mathbb{P}_0 \\ &\underset{d}{\lesssim} k^{4/d} t^2 \int (\log(h_n/f_0))^2 d\mathbb{P}_0 + (\log \int h_n(t))^2 \\ &\underset{d}{\lesssim} k^{4/d} t^2 \cdot k^{-4/d} + t^2 \underset{d}{\lesssim} t^2. \end{aligned} \tag{47}$$

where we used both that $\| \log(f_0/f_g) \|_{L_2(\mathbb{P}_0)}^2 \lesssim k^{-4/d}$, and that $\| \log(f_0/h_n) \|_{L_2(\mathbb{P}_0)}^2 \lesssim n^{-4/d}$.

Finally, we conclude that the following holds for $t \in (0, c_1(d)k^{-2/d})$

$$\begin{aligned}
\mathbb{E}\left[\int \log\left(\frac{\widetilde{h_n(t)}}{f_0}\right) d\mathbb{P}_n\right] &= \mathbb{E}\left[\int \log \widetilde{h_n(t)} d\mathbb{P}_n\right] - \int \log f_0 d\mathbb{P}_0 \\
&= \mathbb{E}\left[\int \log \widetilde{h_n(t)} d\widetilde{\mathbb{P}_n} - KL(f_0 || \widetilde{h_n(t)})\right] \\
&\geq \mathbb{E}\left[\int \log \widetilde{h_n(t)} d\widetilde{\mathbb{P}_n}\right] - c_d t^2 \\
&= \mathbb{E}\left[\int \log h_n(t) d\widetilde{\mathbb{P}_n}\right] - c_d t^2 \\
&\geq t k^{\frac{2}{d}} \mathbb{E}\left[\int \log h_n d\widetilde{\mathbb{P}_n}\right] + (1-t)k^{2/d} \underbrace{\mathbb{E}\left[\int \log f_0 d\widetilde{\mathbb{P}_n}\right]}_{=0} - Ct^2 \\
&\geq c_3(d)(k/n)^{2/d} t - C_3(d)t^2,
\end{aligned} \tag{48}$$

where in the last inequality, we used Lemma 12. Note that the random variable $|\log(\widetilde{h_n(t)}/f_0)|$ is bounded by $C_3 n^{-1/2d}$. Therefore, by Hoeffding's inequality, we obtain that with high probability and uniformly for $t \in (0, k^{-2/d})$ that

$$\Psi_{f_0,n}(t) \geq \int \log(\widetilde{h_n(t)}/f_0) d\mathbb{P}_n \geq c_3(d)(k/n)^{2/d} t - C_3(d)t^2,$$

where $\Psi_{f_0,n}(t)$ is defined in (13) above.

Now, we may apply Theorem 1 on f_0 (and optimize the dependencies on the logarithmic factors in n and in k), which implies that

$$\Psi_{f_0,n}(t) \leq C_3(d)(k/n)^{2/d} \log(n)^{2\frac{d-1}{d}} t - t^2/16,$$

As $\widehat{f}_n$ lies in $\mathcal{S}(t)$ that maximizes $\Psi_{f_0,n}(t)$, it lies in some $t \geq 0$ such that

$$\Psi_{f_0,n}(t) \underset{d}{\gtrsim} (k/n)^{\frac{4}{d}}.$$

By using the last three equations, we conclude a t that maximizes $\Psi_{f_0,n}(\cdot)$ must be larger than

$$t_2(n,k,d) \underset{d}{\gtrsim} (k/n)^{2/d} \log(n)^{-2\frac{d-1}{d}}$$

and the claim follows.