ADAPTATION IN MULTIVARIATE LOG-CONCAVE DENSITY ESTIMATION

By Oliver Y. Feng[§], Adityanand Guntuboyina[*,¶]
Arlene K. H. Kim[†,‖] and Richard J. Samworth[‡,§]

University of Cambridge[§]
University of California, Berkeley[¶] *and*
Korea University[‖]

We study the adaptation properties of the multivariate log-concave maximum likelihood estimator over three subclasses of log-concave densities. The first consists of densities with polyhedral support whose logarithms are piecewise affine. The complexity of such densities f can be measured in terms of the sum $\Gamma(f)$ of the numbers of facets of the subdomains in the polyhedral subdivision of the support induced by f. Given n independent observations from a d-dimensional log-concave density with $d \in \{2,3\}$, we prove a sharp oracle inequality, which in particular implies that the Kullback–Leibler risk of the log-concave maximum likelihood estimator for such densities is bounded above by $\Gamma(f)/n$, up to a polylogarithmic factor. Thus, the rate can be essentially parametric, even in this multivariate setting. For the second type of adaptation, we consider densities that are bounded away from zero on a polytopal support; we show that up to polylogarithmic factors, the log-concave maximum likelihood estimator attains the rate $n^{-4/7}$ when $d = 3$, which is faster than the worst-case rate of $n^{-1/2}$. Finally, our third type of subclass consists of densities whose contours are well-separated; these new classes are constructed to be affine invariant and turn out to contain a wide variety of densities, including those that satisfy Hölder regularity conditions. Here, we prove another sharp oracle inequality, which reveals in particular that the log-concave maximum likelihood estimator attains a risk bound of order $n^{-\min\left(\frac{\beta+3}{\beta+7}, \frac{4}{7}\right)}$ when $d = 3$ over the class of β-Hölder log-concave densities with $\beta \in (1,3]$, again up to a polylogarithmic factor.

1. Introduction. The field of nonparametric inference under shape constraints has witnessed remarkable progress on several fronts over the last decade or so. For instance, the area has been enriched by methodological innovations in new research problems, including convex set estimation (Gardner et al., 2006; Guntuboyina, 2012; Brunel, 2013), shape-constrained dimension reduction (Chen and Samworth, 2016; Xu et al., 2016; Groeneboom and Hendrickx, 2018) and ranking and pairwise comparisons (Shah et al., 2017). Algorithmic advances together with increased computing power now mean that certain estimators have become computationally feasible for much larger sample sizes (Koenker and Mizera, 2014; Mazumder et al., 2018). On the theoretical side, new tools developed in recent years have allowed us to make progress in understanding how shape-constrained procedures behave (Dümbgen et al., 2011; Guntuboyina and Sen, 2013; Cai and Low, 2015). Moreover, minimax rates of convergence are now known[*] for a variety of core problems in the area, including decreasing density estimation on the non-negative half-line (Birgé, 1987), isotonic regression (Zhang, 2002; Chatterjee et al., 2018; Deng and Zhang, 2018; Han et al., 2019) and convex regression (Han and Wellner, 2016). Groeneboom and Jongbloed (2014) provide a book-length introduction to the field; many recent developments are also surveyed in a 2018 special issue of *Statistical Science* devoted to the topic.

One of the most intriguing aspects of many shape-constrained estimators is their ability to *adapt* to unknown features of the underlying data generating mechanism. To illustrate what we mean by this, consider a general setting in which the goal is to estimate a function or parameter that belongs to a class $\mathcal{D}$. Given a subclass $\mathcal{D}' \subseteq \mathcal{D}$, we say that our estimator adapts to $\mathcal{D}'$ with respect to a given loss function if its worst-case rate of convergence over $\mathcal{D}'$ is an improvement on its corresponding worst-case rate over $\mathcal{D}$; in the best case, it may even attain the minimax rates of convergence over both $\mathcal{D}'$ and $\mathcal{D}$, at least up to polylogarithmic factors in the sample size. As a concrete example of this phenomenon, consider independent observations $Y_1, \ldots, Y_n$ with $Y_i \sim N(\theta_{0i}, 1)$, where $\theta_0 := (\theta_{01}, \ldots, \theta_{0n})$ belongs to the monotone cone $\mathcal{D} := \{\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n : \theta_1 \leq \ldots \leq \theta_n\}$. Zhang (2002) established that the least squares estimator $\hat{\theta}_n$ over $\mathcal{D}$ satisfies the worst-case ℓ_2-risk bound

$$\mathbb{E}\{\|\hat{\theta}_n - \theta_0\|^2\} \leq C \left\{ \left(\frac{\theta_{0n} - \theta_{01}}{n}\right)^{2/3} + \frac{\log n}{n} \right\}$$

---

[*]In the interests of transparency, we note that in some of our examples, there remain gaps between the known minimax lower and upper bounds that are polylogarithmic in the sample size.

for some universal constant $C > 0$; thus, in particular, it attains the minimax rate of $O(n^{-2/3})$ for signals $\theta_0 \in \mathcal{D}$ of bounded uniform norm. On the other hand, the fact that the least squares estimator is piecewise constant motivates the thought that $\hat{\theta}_n$ might adapt to piecewise constant signals. More precisely, letting $\mathcal{D}' \equiv \mathcal{D}'_k$ denote the subset of $\mathcal{D}$ consisting of signals with at most k constant pieces, a consequence of Bellec (2018, Theorem 3.2) is that

$$\sup_{\theta_0 \in \mathcal{D}'_k} \mathbb{E}\big\{\|\hat{\theta}_n - \theta_0\|^2\big\} \leq \frac{k}{n} \log\Big(\frac{en}{k}\Big).$$

Note that, up to the logarithmic factor, this rate of convergence (which is parametric when k is a constant) is the same as could be attained by an 'oracle' estimator that had access to the locations of the jumps in the signal. The proof of this beautiful result relies on the characterisation of the least squares estimator as an ℓ_2-projection onto the closed, convex cone $\mathcal{D}$, as well as the notion of such a cone's statistical dimension, which can be computed exactly in the case of the monotone cone (Amelunxen et al., 2014; Soloff et al., 2019).

As a result of intensive work over the past decade, the adaptive behaviour of shape-constrained estimators is now fairly well understood in a variety of univariate problems (Balabdaoui, Rufibach and Wellner, 2009; Dümbgen and Rufibach, 2009; Jankowski, 2014; Chatterjee et al., 2015; Kim et al., 2018; Chatterjee and Lafferty, 2019). Moreover, in the special cases of isotonic and convex regression, very recent work has shown that shape-constrained least squares estimators exhibit an even richer range of adaptation properties in multivariate settings (Han and Wellner, 2016; Chatterjee et al., 2018; Deng and Zhang, 2018; Han et al., 2019; Han, 2019). For instance, Chatterjee et al. (2018) showed that the least squares estimator in bivariate isotonic regression continues to enjoy parametric adaptation up to polylogarithmic factors when the signal is constant on a small number of rectangular pieces. On the other hand, Han et al. (2019) proved that, in general dimensions $d \geq 3$, the least squares estimator in fixed, lattice design isotonic regression[†] adapts at rate $\tilde{O}(n^{-2/d})$ for constant signals, and that it is not possible to obtain a faster rate for this estimator. This is still an improvement on the minimax rate of $\tilde{O}(n^{-1/d})$ over all isotonic signals (in the lexicographic ordering) with bounded uniform norm, but is strictly slower than the parametric rate. We remark that, in addition to the ideas employed by Bellec (2018), these higher-dimensional results rely on an alternative characterisation of the least squares estimator due to Chatterjee (2014), as well

---

[†]Here and below, the $\tilde{O}(\cdot)$ notation is used to denote rates that hold up to polylogarithmic factors in n.

as an argument that controls the statistical dimension of the d-dimensional monotone cone by induction on d; see Han (2019, Theorem 3.9) for an alternative approach to the latter. Given the surprising nature of these results, it is of great interest to understand the extent to which adaptation is possible in other shape-constrained estimation problems.

This paper concerns multivariate adaptation behaviour in log-concave density estimation. The class of log-concave densities lies at the heart of modern shape-constrained nonparametric inference, due to both the modelling flexibility it affords and its attractive stability properties under operations such as marginalisation, conditioning, convolution and linear transformations (Walther, 2009; Saumard and Wellner, 2014; Samworth, 2018). However, the class of log-concave densities is not convex, so the maximum likelihood estimator cannot be regarded as a projection onto a convex set, and the results of Amelunxen et al. (2014), Chatterjee (2014) and Bellec (2018) cannot be applied.

To set the scene, let $\mathcal{F}_d$ denote the class of upper semi-continuous, log-concave densities on $\mathbb{R}^d$, and suppose that $X_1, \ldots, X_n$ are independent and identically distributed random vectors with density $f_0 \in \mathcal{F}_d$. Also, write $d_{\mathrm{H}}(f, g) := \left\{ \int_{\mathbb{R}^d} (f^{1/2} - g^{1/2})^2 \right\}^{1/2}$ for the Hellinger distance between two densities f and g. Kim and Samworth (2016) proved the following minimax lower bound[‡]: for each $d \in \mathbb{N}$, there exists $c_d > 0$ such that

$$
(1) \qquad \inf_{\tilde{f}_n} \sup_{f_0 \in \mathcal{F}_d} \mathbb{E}\{d_{\mathrm{H}}^2(\tilde{f}_n, f_0)\} \geq \left\{ \begin{array}{ll} c_1\, n^{-4/5} & \text{if } d = 1 \\ c_d\, n^{-2/(d+1)} & \text{if } d \geq 2, \end{array} \right.
$$

where the infimum is taken over all estimators $\tilde{f}_n$ of f_0 based on $X_1, \ldots, X_n$. Thus, when $d \geq 3$, there is a more severe curse of dimensionality than for the problem of estimating a density with two bounded derivatives and exponentially decaying tails, for which the corresponding minimax rate is $n^{-4/(d+4)}$ in all dimensions (Goldenshluger and Lepski, 2014). See Section S3.3.1 in the supplementary material (Feng et al., 2020) for further details and discussion. The reason why this comparison is interesting is because any concave function is twice differentiable Lebesgue almost everywhere on its effective domain, while a twice differentiable function is concave if and only if its Hessian matrix is non-positive definite at every point. This observation had led to the prediction that the rates in these problems ought to coincide (e.g. Seregin and Wellner, 2010, page 3778).

The result (1) is relatively discouraging as far as high-dimensional log-concave density estimation is concerned, and has motivated the definition

---

[‡]In fact, more recently, Kur et al. (2019) proved that c_d may be chosen independently of the dimension d.

of alternative procedures that seek improved rates when d is large under additional structure, such as independent component analysis (Samworth and Yuan, 2012) or symmetry (Xu and Samworth, 2019). Nevertheless, in lower-dimensional settings, the performance of the log-concave maximum likelihood estimator $\hat{f}_n := \operatorname{argmax}_{f \in \mathcal{F}_d} \sum_{i=1}^n \log f(X_i)$ has been studied with respect to the divergence $d_X^2(\hat{f}_n, f_0) := n^{-1} \sum_{i=1}^n \log \frac{\hat{f}_n(X_i)}{f_0(X_i)}$ (cf. Kim et al., 2018, page 2281). This loss function is closely related to the Kullback–Leibler divergence $\mathrm{KL}(f, g) := \int_{\mathbb{R}^d} f \log(f/g)$ and Hellinger distance. Indeed, we have $d_{\mathrm{H}}^2(\hat{f}_n, f_0) \leq \mathrm{KL}(\hat{f}_n, f_0) \leq d_X^2(\hat{f}_n, f_0)$, where the the first bound is standard and the second inequality follows by applying Dümbgen et al. (2011, Remark 2.3) to the function $x \mapsto \log\big(f_0(x)/\hat{f}_n(x)\big)$. A small modification of the proof of Kim and Samworth (2016, Theorem 5) yields the following result, which is stated as Theorem S2 in the supplementary material (Feng et al., 2020) for convenience:

$$(2) \qquad \sup_{f_0 \in \mathcal{F}_d} \mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} = \begin{cases} O(n^{-4/5}) & \text{if } d = 1 \\ O(n^{-2/3} \log n) & \text{if } d = 2 \\ O(n^{-1/2} \log n) & \text{if } d = 3; \end{cases}$$

see also Doss and Wellner (2016) for a related result in the univariate case. Moreover, very recently, Kur et al. (2019) proved that[§]

$$(3) \qquad \sup_{f_0 \in \mathcal{F}_d} \mathbb{E}\{d_{\mathrm{H}}^2(\hat{f}_n, f_0)\} = O_d(n^{-2/(d+1)} \log n)$$

for $d \geq 4$, so that, at least in squared Hellinger loss, it follows from (1), (2) and (3) that $\hat{f}_n$ attains the minimax optimal rate in all dimensions, up to a logarithmic factor.

Our goal is to explore the potential of the log-concave maximum likelihood estimator to adapt to three different types of subclass of $\mathcal{F}_d$. The definition of the first of these is motivated by the observation that $\log \hat{f}_n$ is piecewise affine on the convex hull of $X_1, \ldots, X_n$, a polyhedral subset of $\mathbb{R}^d$. It is therefore natural to consider, for $k \in \mathbb{N}$ and $m \in \mathbb{N} \cup \{0\}$, the subclass $\mathcal{F}^k(\mathcal{P}^m) \equiv \mathcal{F}_d^k(\mathcal{P}^m) \subseteq \mathcal{F}_d$ consisting of densities that are both log-k-affine on their support (see Section 1.1), and have the property that this support is a polyhedral set with at most m facets. Note that this class contains densities with unbounded support. By Proposition 1 in Section 2 below, the complexity of such densities f can be measured in terms of the sum $\Gamma(f)$ of the numbers of facets of the subdomains in the polyhedral subdivision of the

---

[§]Here and below, the $O_d(\cdot)$ notation is used as shorthand for an upper bound that holds up to a dimension-dependent quantity.

support induced by f. A consequence of our first main result, Theorem 2, is that for all $f_0 \in \mathcal{F}^k(\mathcal{P}^m)$, we have

$$(4) \qquad \mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} = \tilde{O}\left(\frac{\Gamma(f_0)}{n}\right)$$

when $d \in \{2, 3\}$; moreover, we also show that $\Gamma(f_0)$ is at most of order $k+m$ when $d = 2$, and at most of order $k(k + m)$ when $d = 3$. Thus, when k and m may be regarded as constants, (4) reveals that the log-concave maximum likelihood estimator adapts at a parametric rate to $\mathcal{F}^k(\mathcal{P}^m)$ when $d \in \{2, 3\}$, up to the polylogarithmic term. Moreover, Theorem 2 offers a complete picture for this type of adaptation by providing a sharp oracle inequality that covers the case where f_0 is well approximated (in a Kullback–Leibler sense) by a density in $\mathcal{F}^k(\mathcal{P}^m)$ for some k, m. Unsurprisingly, the proof of this inequality is much more delicate and demanding than the corresponding univariate result given in Kim et al. (2018), owing to the greatly increased geometric complexity of both the boundaries of convex subsets of $\mathbb{R}^d$ for $d \geq 2$ and the structure of the polyhedral subdivisions induced by the densities in $\mathcal{F}^k(\mathcal{P}^m)$. In particular, the parameter m plays no role in the univariate problem, since the boundary of a convex subset of the real line has at most two points, but it turns out to be crucial in this multivariate setting. Indeed, no form of adaptation would be achievable in the absence of restrictions on the shape of the support of $f_0 \in \mathcal{F}_d$; for instance, when f_0 is the uniform density on a closed Euclidean ball in $\mathbb{R}^d$ with $d \geq 2$, consideration of the volume of the convex hull of $X_1, \ldots, X_n$ yields that $\mathbb{E}\{d_H^2(\hat{f}_n, f_0)\} \geq \tilde{c}_d \, n^{-2/(d+1)}$ for some $\tilde{c}_d > 0$ depending only on d (Wieacker, 1987).

In contrast to the isotonic regression problem described above, Theorem 2 indicates that even when $d = 3$, the log-concave maximum likelihood estimator also enjoys essentially parametric adaptation when f_0 is close to a density in $\mathcal{F}^k(\mathcal{P}^m)$ for small k and m. Unfortunately, our arguments do not allow us to extend our results to dimensions $d \geq 4$, where the relevant bracketing entropy integral diverges at a polynomial rate. Recent work by Carpenter et al. (2018) derived worst-case rates in squared Hellinger loss for the log-concave maximum likelihood estimator when $d \geq 4$; the crux of their argument involved using Vapnik–Chervonenkis theory to bound

$$\mathbb{E}\left(\sup_{K \in \mathcal{K}_d^*} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \in K\}} - \mathbb{P}(X_1 \in K) \right|\right),$$

where $\mathcal{K}_d^*$ denotes the set of all closed, convex subsets of $\mathbb{R}^d$. Kur et al. (2019) obtained an improved bound on this quantity of $O_d(n^{-2/(d+1)})$ using

a general chaining argument, and this allowed them to deduce the worst-case guarantees on the performance of the log-concave maximum likelihood estimator stated in (3). Unfortunately, it is unclear whether this approach can provide any adaptation guarantees.

Sections 3 and 4 consider different subclasses of $\mathcal{F}_d$, and are motivated by the hope that if we rule out 'bad' log-concave densities such as the uniform densities with smooth boundaries mentioned above, then we may be able to achieve faster rates of convergence, up to the $n^{-4/(d+4)}$ rate conjectured by Seregin and Wellner (2010). Since this rate already coincides with the worst-case rate for the log-concave maximum likelihood estimator given in (2) when $d = 1, 2$ (up to a logarithmic factor), and since the same entropy integral divergence issues mentioned above apply when $d \geq 4$, we focus on the case $d = 3$ in these sections. In Section 3, we restrict attention to densities with polytopal support (that need not satisfy the log-k-affine condition of Section 2). Theorem 5 therein provides a sharp oracle inequality, which reveals that in such cases, the log-concave maximum likelihood estimator attains the rate $\tilde{O}(n^{-4/7})$ with respect to d_X^2 divergence, at least when the density is bounded away from zero on its support.

In Section 4, we introduce an alternative way to exclude the bad uniform densities mentioned above, namely by considering subclasses of $\mathcal{F}_d$ consisting of densities f whose contours are well-separated in regions where f is small. A major advantage of working with contour separation, as opposed to imposing a conventional smoothness condition such as Hölder regularity, is that we are able to exhibit adaptation over much wider classes of densities, as we illustrate through several examples in Section 4. A consequence of our main theorem in this section (Theorem 9) is that the log-concave maximum likelihood estimator attains the rate $\tilde{O}(n^{-4/7})$ with respect to d_X^2 divergence over the class of Gaussian densities; again, one can think of this result as partially restoring the original conjecture of Seregin and Wellner (2010), in that their rate is achieved with additional restrictions on the class of log-concave densities. A key feature of our definition of contour separation is that it is affine invariant; since the log-concave maximum likelihood estimator is affine equivariant and our loss functions d_H^2, KL and d_X^2 are affine invariant, this allows us to obtain rates that are uniform over classes without any scale restrictions.

We mention that alternative estimators have also been studied for the class of log-concave densities. One such is the smoothed log-concave maximum likelihood estimator (Dümbgen and Rufibach, 2009; Chen and Samworth, 2013), which matches the first two moments of the empirical distribution of the data, but for which results on rates of convergence are less

developed. Another interesting proposal is the ρ-estimation framework of Baraud and Birgé (2016), for which similar adaptation properties as for the log-concave maximum likelihood estimator are known in the univariate case.

Proofs of most of the main results in Section 2 are given in the Appendix (Section 5). The remaining proofs, as well as numerous auxiliary results, are presented in the supplementary material (Feng et al., 2020); these results appear with an 'S' before the relevant label number. In particular, the proofs of all the stated results in Sections 3 and 4 are deferred to Sections S1.4 and S3.1 respectively.

1.1. *Notation and background.* First, we set up some notation and definitions that will be used throughout the main text as well as in the proofs later on. For a fixed $d \in \mathbb{N}$, we write $\{e_1, \ldots, e_d\}$ for the standard basis of $\mathbb{R}^d$ and denote the ℓ_2 norm of $x = (x_1, \ldots, x_d) = \sum_{j=1}^d x_j e_j \in \mathbb{R}^d$ by $\|x\| \equiv \|x\|_2 = \left(\sum_{j=1}^d x_j^2\right)^{1/2}$. For $x, y \in \mathbb{R}^d$, let $[x, y] := \{tx + (1-t)y : t \in [0,1]\}$ denote the closed line segment between them, and define (x, y), $[x, y)$, $(x, y]$ analogously. For $x \in \mathbb{R}^d$ and $r > 0$, let $\bar{B}(x, r) := \{w \in \mathbb{R}^d : \|w - x\| \le r\}$. For $A \subseteq \mathbb{R}^d$, we write $\dim(A)$ for the *affine dimension* of A, i.e. the dimension of the affine hull of A, and for Lebesgue-measurable $A \subseteq \mathbb{R}^d$, we write $\mu_d(A)$ for the d-dimensional Lebesgue measure of A. If $0 < \dim(A) = k < d$, we can view A as a subset of its affine hull and define $\mu_k(A)$ analogously, whilst also setting $\mu_l(A) = 0$ for each integer $l > k$. In addition, we denote the set of positive definite $d \times d$ matrices by $\mathbb{S}^{d \times d}$ and the $d \times d$ identity matrix by $I \equiv I_d$.

Next, let $\Phi \equiv \Phi_d$ be the set of all upper semi-continuous, concave functions $\phi : \mathbb{R}^d \to [-\infty, \infty)$ and let $\mathcal{G} \equiv \mathcal{G}_d := \{e^\phi : \phi \in \Phi\}$. For $\phi \in \Phi$, we write $\operatorname{dom} \phi := \{x \in \mathbb{R}^d : \phi(x) > -\infty\}$ for the *effective domain* of ϕ, and for a general $f : \mathbb{R}^d \to \mathbb{R}$, we write $\operatorname{supp} f := \{x \in \mathbb{R}^d : f(x) \neq 0\}$ for the *support* of f. For $k \in \mathbb{N}$, we say that $f \in \mathcal{G}_d$ is *log-k-affine* if there exist closed sets $E_1, \ldots, E_k$ such that $\operatorname{supp} f = \bigcup_{j=1}^k E_j$ and $\log f$ is affine on each E_j. Moreover, let $\mathcal{F} \equiv \mathcal{F}_d$ be the family of all *densities* $f \in \mathcal{G}_d$, and let $\mu_f := \int_{\mathbb{R}^d} x f(x) \, dx$ and $\Sigma_f := \int_{\mathbb{R}^d} (x - \mu_f)(x - \mu_f)^\top dx$ for each $f \in \mathcal{F}_d$. In addition, we write $\mathcal{F}^{0,I} \equiv \mathcal{F}_d^{0,I} := \{f \in \mathcal{F}_d : \mu_f = 0, \Sigma_f = I\}$ for the class of *isotropic* log-concave densities.

Henceforth, for real-valued functions a and b, we write $a \lesssim b$ if there exists a universal constant $C > 0$ such that $a \le Cb$, and we write $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. More generally, for a finite number of parameters $\alpha_1, \ldots, \alpha_r$, we write $a \lesssim_{\alpha_1, \ldots, \alpha_r} b$ if there exists $C \equiv C_{\alpha_1, \ldots, \alpha_r} > 0$, depending only on $\alpha_1, \ldots, \alpha_r$, such that $a \le Cb$. Also, for $x \in \mathbb{R}$, we write $x^+ := x \vee 0$ and $x^- := (-x)^+$, and for $x > 0$, we define $\log_+ x := 1 \vee \log x$.

To facilitate the exposition in Section 4, we now introduce some additional terminology. We say that the densities f and g on $\mathbb{R}^d$ are *affinely equivalent* if there exist an $\mathbb{R}^d$-valued random variable X and an invertible affine transformation $T \colon \mathbb{R}^d \to \mathbb{R}^d$ such that X has density f and $T(X)$ has density g; in other words, there exist $b \in \mathbb{R}^d$ and an invertible $A \in \mathbb{R}^{d \times d}$ such that $g(x) = |\det A|^{-1} f(A^{-1}(x - b))$ for all $x \in \mathbb{R}^d$. Thus, each $f \in \mathcal{F}_d$ is affinely equivalent to a unique $f_0 \in \mathcal{F}_d^{0,I}$. A class $\mathcal{D}$ of densities is said to be *affine invariant* if it is closed under affine equivalence; in other words, if f belongs to $\mathcal{D}$, then so does every density g that is affinely equivalent to f.

The rest of this subsection is devoted to a review of some convex analysis background used in Section 2. A *closed half-space* is a set of the form $\{x \in \mathbb{R}^d : \alpha^\top x \leq u\}$, where $\alpha \in \mathbb{R}^d \setminus \{0\}$ and $u \in \mathbb{R}$, and the interiors and boundaries of closed half-spaces are known as *open half-spaces* and *affine hyperplanes* respectively. For a non-empty and convex $E \subseteq \mathbb{R}^d$, we say that an affine hyperplane H *supports* E if $H \cap E \neq \emptyset$ and H is the boundary of a closed half-space that contains E. A *face* $F \subseteq E$ is a convex set with the property that if $u, v \in E$ and $tu + (1 - t)v \in F$ for some $t \in (0,1)$, then $u, v \in F$. We say that $x \in E$ is an *extreme point* if $\{x\}$ is a face of E. Also, we say that $F \subseteq E$ is an *exposed face* of E if $F = E \cap H$ for some affine hyperplane H that supports E. Exposed faces of affine dimensions 0, 1 and $\dim(E) - 1$ are also known as *exposed points* (or *vertices*), *edges* and *facets* respectively. We write $\mathcal{F}(E)$ for the set of all facets of E.

A *polyhedral set* is a subset of $\mathbb{R}^d$ that can be expressed as the intersection of finitely many closed half-spaces, and a *polytope* is a bounded polyhedral set, or equivalently the convex hull of a finite subset of $\mathbb{R}^d$; see Theorems 2.4.3 and 2.4.6 in Schneider (2014). As a special case, we also view $\mathbb{R}^d$ as a polyhedral set with 0 facets. Let $\mathcal{P} \equiv \mathcal{P}_d$ denote the collection of all polyhedral sets in $\mathbb{R}^d$ with non-empty interior, and for $m \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, let $\mathcal{P}^m \equiv \mathcal{P}_d^m$ denote the collection of all $P \in \mathcal{P}$ with at most m facets. For $1 \leq k \leq d$, a *k-parallelotope* is the image of $[0,1]^k$ under an injective affine transformation from $\mathbb{R}^k$ to $\mathbb{R}^d$, i.e. a polytope of the form $\{v_0 + \sum_{\ell=1}^k \lambda_\ell v_\ell : 0 \leq \lambda_\ell \leq 1 \text{ for all } \ell\}$, where $v_0, v_1, \ldots, v_k \in \mathbb{R}^d$ and $v_1, \ldots, v_k$ are linearly independent. Recall also that a *k-simplex* is the convex hull of $k + 1$ affinely independent points in $\mathbb{R}^d$. Finally, for $P \in \mathcal{P}_d$, a *(polyhedral) subdivision* of P is a finite collection of sets $E_1, \ldots, E_\ell \in \mathcal{P}_d$ such that $P = \bigcup_{j=1}^\ell E_j$ and $E_i \cap E_j$ is a common face of E_i and E_j for all $i, j \in \{1, \ldots, \ell\}$. A *triangulation* of a polytope $P \in \mathcal{P}_d$ is a subdivision of P consisting solely of d-simplices.

2. Adaptation to log-k-affine densities with polyhedral support.

In order to present the main result of this section, we first need to understand the structure of log-k-affine functions $f \in \mathcal{G}_d$ with polyhedral support. Due to the global nature of the constraints on f, namely that $\log f$ is concave on $\operatorname{supp} f \in \mathcal{P}$ and affine on each of k closed subdomains, the function f necessarily has a simple and rigid structure. More precisely, Proposition 1 below shows that there is a minimal representation of f in which the subdomains are polyhedral sets that form a subdivision of $\operatorname{supp} f$, and the restrictions of $\log f$ to these sets are distinct affine functions. The proof of this result is deferred to Section S2.1.

PROPOSITION 1.   *Suppose that $f \in \mathcal{G}_d$ is log-k-affine for some $k \in \mathbb{N}$ and that $\operatorname{supp} f \in \mathcal{P}$. Then there exist $\kappa(f) \leq k$, $\alpha_1, \ldots, \alpha_{\kappa(f)} \in \mathbb{R}^d$, $\beta_1, \ldots, \beta_{\kappa(f)} \in \mathbb{R}$ and a polyhedral subdivision $E_1, \ldots, E_{\kappa(f)}$ of $\operatorname{supp} f$ such that $f(x) = \exp(\alpha_j^\top x + \beta_j)$ for all $x \in E_j$, and $\alpha_i \neq \alpha_j$ whenever $i \neq j$. Moreover, the triples $(\alpha_j, \beta_j, E_j)_{j=1}^{\kappa(f)}$ are unique up to reordering. In addition, if $\operatorname{supp} f \in \mathcal{P}^m$, then $E_j \in \mathcal{P}^{k+m-1}$ for all j.*

In particular, for each such f, the sum of the numbers of facets of the polyhedral subdomains $E_1, \ldots, E_{\kappa(f)}$, which we denote by

$$(5) \qquad \Gamma(f) := \sum_{j=1}^{\kappa(f)} |\mathcal{F}(E_j)|,$$

is well-defined and can be viewed as a parameter that measures the complexity of f. Now for $k \in \mathbb{N}$ and $P \in \mathcal{P}$, let $\mathcal{F}^k(P)$ denote the collection of all $f \in \mathcal{F}_d$ for which $\kappa(f) \leq k$ and $\operatorname{supp} f = P$, so that $\mathcal{F}^k(\mathcal{P}^m) = \bigcup_{P \in \mathcal{P}^m} \mathcal{F}^k(P)$ for $m \in \mathbb{N}_0$. It is shown in Proposition S21 that $\mathcal{F}^k(\mathcal{P}^m)$ is non-empty if and only if $k + m \geq d + 1$. We remark here that it is more appropriate to quantify the complexity of a polyhedral support in terms of m, which refers to the number of facets of the support, rather than in terms of the number of vertices. Indeed, the former quantity may be much greater than the latter when the support is unbounded; for example, a polyhedral convex cone has just a single vertex but may have arbitrarily many facets. That said, if the support is a polytope with v vertices and m facets, it can be shown that $v = m$ when $d = 2$, and that $v \leq 2m - 4$ and $m \leq 2v - 4$ when $d = 3$; see the proof of Lemma S23 and the subsequent remark.

We are now in a position to state our sharp oracle inequality for the risk of the log-concave maximum likelihood estimator when the true $f_0 \in \mathcal{F}_d$ is close to some element of $\mathcal{F}^k(\mathcal{P}^m)$.

THEOREM 2.   *Fix $d \in \{2,3\}$. Let $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} f_0 \in \mathcal{F}_d$ with $n \geq d+1$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Then there exists a universal constant $C > 0$ such that*

$$(6) \quad \mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \inf_{\substack{k \in \mathbb{N},\, m \in \mathbb{N}_0:\\ k+m \geq d+1}} \inf_{f \in \mathcal{F}^k(\mathcal{P}^m)} \left\{ \frac{C\,\Gamma(f)}{n} \log^{\gamma_d} n + \mathrm{KL}(f_0, f) \right\},$$

where $\gamma_2 := 9/2$ and $\gamma_3 := 8$. Moreover, for $d \in \{2,3\}$, we have $\Gamma(f) \lesssim k^{d-2}(k+m)$ for all $f \in \mathcal{F}^k(\mathcal{P}^m)$.

The 'sharpness' in this oracle inequality refers to the fact that the approximation term $\mathrm{KL}(f_0, f)$ has leading constant 1. A consequence of Theorem 2 is that if $d = 2$ and $f_0 \in \mathcal{F}^k(\mathcal{P}^m)$ with $k + m$ small by comparison with $n^{1/3}/\log^{7/2} n$, then the log-concave maximum likelihood estimator attains an adaptive rate that is faster than the rate of decay of the worst-case risk bounds (2) of Kim and Samworth (2016). When $d = 3$, the same conclusion holds when $k(k + m)$ is small by comparison with $n^{1/2}/\log^7 n$.

Theorem 2 is proved in Section 5 by first considering the case $k = 1$, where it turns out that we can prove a slightly stronger version of our result. We therefore state it separately for convenience:

THEOREM 3.   *Fix $d \in \{2,3\}$. Let $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} f_0 \in \mathcal{F}_d$ with $n \geq d+1$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Then there exists a universal constant $\bar{C} > 0$ such that*

$$(7) \qquad \mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \inf_{m \geq d} \left\{ \frac{\bar{C}m}{n} \log^{\gamma_d} n + \inf_{\substack{f \in \mathcal{F}^1(\mathcal{P}^m) \\ \mathrm{supp}\, f_0 \subseteq \mathrm{supp}\, f}} d_{\mathrm{H}}^2(f_0, f) \right\}.$$

We suspect that the restriction on the support of the approximating density f in (7) is an artefact of our proof. Indeed, in the case $d = 1$, Baraud and Birgé (2016) obtain an oracle inequality for their ρ-estimator where the approximating density f need not have this property (although their result is stated for d_{H}^2 rather than d_X^2); moreover, we have been able to strengthen the corresponding univariate result for the log-concave maximum likelihood estimator (Kim et al., 2018, Theorem 5) by removing this restriction.

The proof of Theorem 3 in fact constitutes the main technical challenge in deriving Theorem 2. This entails deriving upper bounds on the (local) Hellinger bracketing entropies of classes of log-concave functions that lie in small Hellinger neighbourhoods of densities in $f \in \mathcal{F}^1(\mathcal{P}^m)$ for each $m \in \mathbb{N}$ with $m \geq d$. Our argument proceeds via a series of steps, the first of which

deals with the case where f is a uniform density on a simplex (Proposition S8); it turns out that any density in a small Hellinger ball around such an f satisfies a uniform upper bound (Lemma S25(ii)), and a pointwise lower bound whose contours are characterised geometrically in Lemma S30 (and illustrated in Figure S5). We proceed by considering a finite nested sequence of polytopal subsets of the simplex, each of which has a controlled number of vertices and approximates the region enclosed by one of the aforementioned contours; see the accompanying Figure S1. After constructing suitable triangulations of the regions between successive polytopes (Corollary S33), we exploit existing bracketing entropy results for classes of bounded log-concave functions (Proposition S7).

In the next step, we consider the uniform density on a polytope in $\mathcal{P}^m$; here, using the fact that there is a triangulation of the support into $O(m)$ simplices (Lemma S23), we apply our earlier bracketing entropy bounds in conjunction with an additional argument which handles carefully the fact that these simplices may have very different volumes (Proposition S9).

Finally, in the proof of Proposition 10 in Section 5, we generalise to settings where f is an arbitrary (not necessarily uniform) log-affine density whose polyhedral support may be unbounded. There, we subdivide the domain by intersecting it with a sequence of parallel half-spaces whose normal vectors are in the direction of the negative log-gradient of the density. Our characterisation of such log-affine densities in Section S2.1 ultimately allows us to apply our earlier results to transformations of the original density and thereby obtain the desired local bracketing entropy bounds (Proposition 10). The conclusion of Theorem 3 then follows from standard empirical process theory arguments (e.g. van de Geer, 2000, Corollary 7.5); see Section 5.

We do not claim any optimality of the polylogarithmic factors in Theorems 2 and 3. In fact, we can improve these exponents in the special case where f_0 is well-approximated by a uniform density $f_P := \mu_d(P)^{-1}\mathbb{1}_P$ on a polytope $P \in \mathcal{P}_d$. Note that every polytope in $\mathcal{P}_d$ has at least as many facets as a d-simplex, namely $d+1$; see for example Lemma S22.

PROPOSITION 4.   *Fix $d \in \{2,3\}$, and for $m \geq d+1$, denote by $\mathcal{F}^{[1]}(\mathcal{P}^m)$ the subclass of all uniform densities on polytopes in $\mathcal{P}^m$. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0 \in \mathcal{F}_d$ with $n \geq d+1$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Then there exists a universal constant $C' > 0$ such that*

$$(8) \qquad \mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \inf_{m \geq d+1} \left\{ \frac{C'm}{n} \log^{\gamma'_d} n + \inf_{\substack{f \in \mathcal{F}^{[1]}(\mathcal{P}^m) \\ \operatorname{supp} f_0 \subseteq \operatorname{supp} f}} d_H^2(f_0, f) \right\},$$

where $\gamma_2' := 3$ and $\gamma_3' := 6$.

3. Adaptation to densities bounded away from zero on a polytopal support. Recall from the discussion in the introduction that in order to observe adaptive behaviour for the log-concave maximum likelihood estimator, we need to exclude uniform densities supported on convex sets with smooth boundaries. In fact, we will see from Proposition 6 below that we also need to rule out subclasses containing sequences of elements of $\mathcal{F}_d$ that approximate such uniform densities. In this section, we continue to work with densities in $\mathcal{F}_d$ that are close to a log-concave density with polyhedral support, but, in contrast to Section 2, now drop the requirement that this approximating density be log-k-affine. In fact, we do not impose any extra structural constraints or smoothness conditions that would regulate further the behaviour of the densities on the interiors of their supports. It will turn out, however, that we will only be able to improve on the worst-case risk bounds of Theorem S2 when the approximating density is also bounded away from zero on its support, which must therefore necessarily be a polytope. The generality of the resulting new classes means that we can no longer expect near-parametric adaptation, and moreover, for the reasons explained in the introduction, our main result of this section (Theorem 5 below) is restricted to the case $d = 3$. As an example of a density that will be covered by this result, we can consider the density of a trivariate Gaussian random vector conditioned to lie in $[-1, 1]^3$.

Following on from Proposition 4, we now extend the definition of $\mathcal{F}^{[1]}(\mathcal{P}^m)$ given above and introduce our new family of subclasses of $\mathcal{F}_d$. For $\theta \in (0, \infty)$ and a polytope $P \in \mathcal{P}_d$, let $\mathcal{F}^{[\theta]}(P) \equiv \mathcal{F}_d^{[\theta]}(P)$ denote the collection of all $f \in \mathcal{F}_d$ for which supp $f = P$ and $f \geq \theta^{-1} f_P$ on P. Then $\mathcal{F}^{[1]}(P) = \{f_P\}$ and $\mathcal{F}^{[\theta]}(P)$ is non-empty if and only if $\theta \geq 1$. For $\theta \in [1, \infty)$ and $m \in \mathbb{N}$ with $m \geq d + 1$, denote by $\mathcal{F}^{[\theta]}(\mathcal{P}^m) \equiv \mathcal{F}_d^{[\theta]}(\mathcal{P}_d^m)$ the union of those $\mathcal{F}^{[\theta]}(P)$ for which P is a polytope in $\mathcal{P}^m \equiv \mathcal{P}_d^m$, and note that this is a non-empty affine invariant subclass of $\mathcal{F}_d$. Indeed, fix $b \in \mathbb{R}^d$ and an invertible $A \in \mathbb{R}^{d \times d}$, and let $T \colon \mathbb{R}^d \to \mathbb{R}^d$ be the invertible affine transformation defined by $T(x) := Ax + b$. If $X \sim f \in \mathcal{F}^{[\theta]}(P)$ for some polytope $P \in \mathcal{P}^m$, then $\mu_d(T(P)) = |\det A| \mu_d(P)$, and so the density g of $T(X)$ satisfies $g(x) = |\det A|^{-1} f(T^{-1}(x)) \geq \{\theta |\det A| \mu_d(P)\}^{-1} = \{\theta \mu_d(T(P))\}^{-1}$ for all $x \in T(P)$. Since supp $g = T(P)$ is also a polytope in $\mathcal{P}^m$, this shows that $g \in \mathcal{F}^{[\theta]}(\mathcal{P}^m)$, as required.

The sharp oracle inequality (9) below may be viewed as complementary to Theorem 3 and Proposition 4.

THEOREM 5.   *Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0 \in \mathcal{F}_3$ with $n \geq 4$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Then there exists a universal constant $C > 0$ such that*

$$
\mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \inf_{\substack{m \geq 4 \\ \theta \in (1,\infty)}} \left\{ C \left(\log^{6/7} \theta \left(\frac{m}{n} \right)^{4/7} \log_+^{17/7} \left(\frac{n}{\log^{3/2} \theta} \right) \right. \right.
$$

$$
\left. + \left(\frac{m}{n} \right)^{20/29} \log^{85/29} n + \theta \log^3(e\theta) \frac{m \log^6 n}{n} \right)
$$

$$
\text{(9)} \qquad\qquad + \inf_{\substack{f \in \mathcal{F}_3^{[\theta]}(\mathcal{P}^m) \\ \text{supp } f_0 \subseteq \text{supp } f}} d_H^2(f_0, f) \Bigg\}.
$$

For a fixed $\theta \in (1, \infty)$, note that if n/m is sufficiently large, then the dominant contribution to the right-hand side of (9) comes from the first term. It follows that for fixed θ, m, the log-concave maximum likelihood estimator $\hat{f}_n$ of $f_0 \in \mathcal{F}_3^{[\theta]}(\mathcal{P}^m)$ converges at rate $\tilde{O}(n^{-4/7})$ as $n \to \infty$, which was the rate originally conjectured by Seregin and Wellner (2010).

Despite the attractions of the adaptation mentioned in the previous paragraph, it is worth considering the bound (9) in the limits as $\theta \searrow 1$ and $\theta \to \infty$. In the first case, owing to the presence of the second term on the right-hand side of (9), we do not recover the bound (8) from Proposition 4 when we take the limit of the right-hand side of (9); see Section S1.3 for further discussion. We also mention here that for a fixed n, the bound in (8) may be stronger than that in (9) if for example $f_0 \in \mathcal{F}_3^{[\theta]}(\mathcal{P}^m)$ for some $\theta \equiv \theta_n \in (1, \infty)$ sufficiently close to 1. To substantiate this remark, we note that if $\theta \in [1, \infty)$ and $P \in \mathcal{P}_3$ is a polytope, then it follows from the proof of Lemma S25(iii) that every $f \in \mathcal{F}_3^{[\theta]}(P)$ satisfies $\theta^{-1} f_P \leq f \lesssim \log^3(e\theta) f_P$ on P. Thus, if $f_0 \in \mathcal{F}_3^{[\theta]}(P)$, then

$$
d_H^2(f_0, f_P) = \int_P \left(\sqrt{f_0} - \sqrt{f_P} \right)^2 \lesssim (1 - \theta^{-1}) \vee \left(\log^3(e\theta) - 1 \right) \lesssim \theta - 1
$$

when $\theta \leq 2$. Consequently, if θ is such that $\theta \leq 1 + n^{-20/29}$ and $m \leq n^{9/29} \log^{-6} n$, then for any $f_0 \in \mathcal{F}_3^{[\theta]}(P)$ with $P \in \mathcal{P}^m$, the bound in (8) is at most a universal constant multiple of $(m/n) \log^6 n + (\theta - 1) \lesssim n^{-20/29}$, while the bound in (9) is at least a universal constant multiple of $n^{-20/29} \log^{85/29} n$.

It is also notable that the bound in (9) diverges to infinity as $\theta \to \infty$. In fact, we will deduce from Proposition 6 below that this is not just an artefact of our analysis; more precisely, the log-concave maximum likelihood

estimator does not adapt uniformly over $\bigcup_{\theta \geq 1} \mathcal{F}_d^{[\theta]}(P)$, or indeed over any subclass of $\mathcal{F}_d$ containing an approximating sequence for a uniform density on a closed Euclidean ball.

PROPOSITION 6.   *Fix $d \in \mathbb{N}$ and $n \geq d + 1$. Let $(f^{(\ell)})$ be a sequence of densities in $\mathcal{F}_d$ for which the corresponding sequence of probability measures $(P^{(\ell)})$ converges weakly to a distribution $P^{(0)}$ with density $f^{(0)} \colon \mathbb{R}^d \to [0, \infty)$. For each $\ell \in \mathbb{N}_0$, let $X_1^{(\ell)}, \ldots, X_n^{(\ell)} \overset{\text{iid}}{\sim} f^{(\ell)}$, and let $\hat{f}_n^{(\ell)}$ denote the corresponding log-concave maximum likelihood estimator. Then*

$$\liminf_{\ell \to \infty} \mathbb{E}\big\{ d_X^2 \big(\hat{f}_n^{(\ell)}, f^{(\ell)} \big) \big\} \geq \mathbb{E}\big\{ d_X^2 \big(\hat{f}_n^{(0)}, f^{(0)} \big) \big\}.$$

To understand the consequences of this lower semi-continuity result, fix any polytope $P \in \mathcal{P}_d$ and a closed Euclidean ball $B \subseteq \mathrm{Int}\, P$. We can find a sequence $(f^{(\ell)})$ in $\bigcup_{\theta \geq 1} \mathcal{F}_d^{[\theta]}(P)$ such that the corresponding probability measures converge weakly to the uniform distribution on B. Such a sequence must necessarily satisfy $\inf_{x \in P} f^{(\ell)}(x) \to 0$, and Proposition 6, together with the result of Wieacker (1987) mentioned in the introduction, then ensures that $\liminf_{\ell \to \infty} \mathbb{E}\big\{ d_X^2 \big(\hat{f}_n^{(\ell)}, f^{(\ell)} \big) \big\} \gtrsim_d n^{-2/(d+1)}$ for $d \geq 2$. Thus, indeed, no adaptation is possible.

The proof of Theorem 5 follows a similar approach to that set out after the statement of Theorem 3. The key intermediate results are the local bracketing entropy bounds in Propositions S10 and S11 in Section S1.3, which are analogous to the Propositions S8 and S9 that prepare the ground for the proof of Theorem 3. As we explain in the discussion before the proof of Proposition S8, some modifications to the previous arguments are necessary, but we once again draw heavily on the technical apparatus developed in Section S2.2. The key reason we are able to apply these techniques here is that the densities in $\mathcal{F}^{[\theta]}(\mathcal{P}^m)$ are bounded away from zero, as evidenced by the fact that the bound (9) diverges as $\theta \to \infty$. Once we have obtained Proposition S11, all that remains is to appeal to standard empirical process theory (van de Geer, 2000, Corollary 7.5), from which the desired conclusion (9) follows readily; see Section S1.4. In contrast to the proof of Theorem 3, we do not require an additional argument along the lines of the proof of Proposition 10 given in Section 5, which is specific to the log-1-affine densities (with possibly unbounded polyhedral support) studied in Section 2.

4. Adaptation to densities with well-separated contours.   In this section, we consider adaptation of the log-concave maximum likelihood estimator over yet further subclasses of $\mathcal{F}_d$. As discussed in Examples 4 and 5

below, these are designed to generalise notions of Hölder smoothness, while at the same time satisfying our key property of affine invariance. Given $S \in \mathbb{S}^{d \times d}$ and $x \in \mathbb{R}^d$, we write $\|x\|_S := (x^\top S^{-1} x)^{1/2}$ for its S-*Mahalanobis norm*.

DEFINITION 1.   *For $\beta \geq 1$ and $\Lambda, \tau > 0$, let $\mathcal{F}^{(\beta,\Lambda,\tau)} \equiv \mathcal{F}_d^{(\beta,\Lambda,\tau)}$ denote the collection of all $f \in \mathcal{F}_d$ that are continuous on $\mathbb{R}^d$ and satisfy*

$$(10) \qquad \|x - y\|_{\Sigma_f} \geq \frac{\{f(x) - f(y)\} \det^{1/2} \Sigma_f}{\Lambda \left\{ f(x) \det^{1/2} \Sigma_f \right\}^{1 - 1/\beta}}$$

whenever $x, y \in \mathbb{R}^d$ are such that $f(y) < f(x) < \tau \det^{-1/2} \Sigma_f$. In addition, we define $\mathcal{F}^{(\beta,\Lambda)} := \bigcap_{\tau > 0} \mathcal{F}^{(\beta,\Lambda,\tau)}$.

The defining condition (10) imposes a separation condition on contours below some fixed level. For instance, when f is isotropic, the condition asks that for all small $t > 0$, the contours of f at levels t and $2t$ are at least a distance of order $\Lambda^{-1} t^{1/\beta}$ apart. See the motivating examples below for further discussion. We now collect together some basic properties of the classes $\mathcal{F}^{(\beta,\Lambda,\tau)}$.

PROPOSITION 7.   *For $\beta \geq 1$ and $\Lambda, \tau > 0$, we have the following:*

- *(i) $\mathcal{F}^{(\beta,\Lambda,\tau)}$ is* affine invariant; *i.e. if $X \sim f \in \mathcal{F}^{(\beta,\Lambda,\tau)}$ and $T \colon \mathbb{R}^d \to \mathbb{R}^d$ is an invertible affine transformation, then the density g of $T(X)$ also lies in $\mathcal{F}^{(\beta,\Lambda,\tau)}$.*
- *(ii) $\mathcal{F}^{(\beta,\Lambda,\tau)} \subseteq \mathcal{F}^{(\beta,\Lambda^*)}$ for all $\Lambda^* \geq \Lambda(B_d/\tau)^{1/\beta}$, where we set $B_d := \sup_{h \in \mathcal{F}_d^{0,I}} \sup_{x \in \mathbb{R}^d} h(x) \in (0, \infty)$.*
- *(iii) If $\alpha \in [1, \beta)$, then $\mathcal{F}^{(\beta,\Lambda,\tau)} \subseteq \mathcal{F}^{(\alpha,\Lambda',\tau)}$ for all $\Lambda' \geq B_d^{1/\alpha - 1/\beta} \Lambda$.*
- *(iv) There exists $\Lambda_{0,d} > 0$, depending only on d, such that $\mathcal{F}^{(\beta,\Lambda)}$ is non-empty only if $\Lambda \geq \Lambda_{0,d}$.*

Note in particular that since the log-concave maximum likelihood estimator $\hat{f}_n$ is affine equivariant (Dümbgen et al., 2011, Remark 2.4), and since our loss functions d_H^2, KL and d_X^2 are affine invariant, property (i) above allows us to restrict attention to *isotropic* $f \in \mathcal{F}^{(\beta,\Lambda,\tau)}$, namely those belonging to $\mathcal{F}_d^{0,I}$. Property (iii) indicates that the classes $\mathcal{F}^{(\beta,\Lambda,\tau)}$ are nested with respect to the exponent $\beta \geq 1$.

In addition, by taking $\alpha = 1$ in (iii) and then applying (ii), we deduce that the densities in $\mathcal{F}^{(\beta,\Lambda,\tau)}$ are all Lipschitz on $\mathbb{R}^d$, but as we will see in Examples 2 and 4, they need not be differentiable everywhere. In cases where

$f \in \mathcal{F}_d$ is differentiable on an open set of the form $\{x \in \mathbb{R}^d : f(x) < \tau^*\}$ for some $\tau^* > 0$, the necessary and sufficient condition in the following proposition provides us with a simpler way of checking whether f belongs to $\mathcal{F}^{(\beta,\Lambda,\tau)}$. For $w \in \mathbb{R}^d$ and $S \in \mathbb{S}^{d\times d}$, let $\|w\|'_S := (w^\top S^{-1} w)^{1/2} \det^{-1/2} S$ denote its *scaled S-Mahalanobis norm*.

PROPOSITION 8.   *Suppose that there exists $\tau^* > 0$ such that $f \in \mathcal{F}_d$ is continuous on $\mathbb{R}^d$ and differentiable at every $x \in \mathbb{R}^d$ satisfying $f(x) < \tau^*$. Then for $\beta \geq 1$ and any $\tau \leq \tau^* \det^{1/2} \Sigma_f$, we have $f \in \mathcal{F}^{(\beta,\Lambda,\tau)}$ if and only if*

$$(11) \qquad \|\nabla f(x)\|'_{\Sigma_f^{-1}} \leq \Lambda \left\{ f(x) \det^{1/2} \Sigma_f \right\}^{1-1/\beta}$$

for all $x \in \mathbb{R}^d$ with $f(x) < \tau \det^{-1/2} \Sigma_f$.

Our main result in this section is a sharp oracle inequality for the performance of the log-concave maximum likelihood estimator when the true log-concave density is close to $\mathcal{F}_d^{(\beta,\Lambda)}$ when $d = 3$. In view of Proposition 7(ii), we work here with the classes $\mathcal{F}_3^{(\beta,\Lambda)}$ rather than the more general classes $\mathcal{F}_3^{(\beta,\Lambda,\tau)}$ for ease of presentation. Let $\Lambda_0 \equiv \Lambda_{0,3} > 0$ be the universal constant from Proposition 7(iv) and its proof, and for each $\beta \geq 1$, let $r_\beta := \frac{\beta+3}{\beta+7} \wedge \frac{4}{7}$.

THEOREM 9.   *Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0 \in \mathcal{F}_3$ for some $n \geq 4$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Then there exists a universal constant $C > 0$ such that*
$$(12)$$
$$\mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \inf_{\beta \geq 1, \Lambda \geq \Lambda_0} \left\{ C\Lambda^{\frac{4\beta}{\beta+7} \wedge 1} n^{-r_\beta} \log^{\frac{16\beta+39}{2(\beta+3)} r_\beta} n + \inf_{f \in \mathcal{F}_3^{(\beta,\Lambda)}} d_{\mathrm{H}}^2(f_0, f) \right\}.$$

Ignoring polylogarithmic factors and focusing on the case where $f_0 \in \mathcal{F}_3^{(\beta,\Lambda)}$ for some $\beta \geq 1$ and $\Lambda > 0$, Theorem 9 presents a continuum of rates that interpolate between the worst-case rate of $\tilde{O}(n^{-1/2})$, corresponding to rate when $\beta = 1$, and $\tilde{O}(n^{-4/7})$, again matching the rate conjectured by Seregin and Wellner (2010).

As mentioned in the introduction, the main attraction of working with the general contour separation condition (10) is that we can give several examples of classes of densities contained within $\mathcal{F}^{(\beta,\Lambda,\tau)}$ for suitable β, Λ and τ. Since each of the conditions (10) and (11) are affine invariant, it suffices to check these conditions for the isotropic elements of the relevant classes (or for any other convenient choice of scaling). Moreover, to verify (10) for

densities that are spherically symmetric, it suffices to consider pairs x, y of the form $y = \lambda x$ for some $\lambda > 0$; in other words, if $f(x) = g(\|x\|)$, then it is enough to verify the contour separation condition (10) for g.

EXAMPLE 1 (Gaussian densities).   Writing $f \colon x \mapsto (2\pi)^{-d/2} e^{-\|x\|^2/2}$ for the standard Gaussian density on $\mathbb{R}^d$ and fixing an arbitrary $\beta \geq 1$, we have

$$\|\nabla f(x)\|'_I = \|\nabla f(x)\| = \frac{\|x\|}{(2\pi)^{d/2}} e^{-\|x\|^2/2} = 2^{1/2} f(x) \log^{1/2} \left(\frac{1}{(2\pi)^{d/2} f(x)} \right)$$

$$\leq \frac{\beta^{1/2}}{(2\pi)^{d/(2\beta)}} e^{-1/2} f(x)^{1-1/\beta}$$

for all $x \in \mathbb{R}^d$. Hence, it follows from Proposition 8 that $f \in \mathcal{F}^{(\beta,\Lambda)}$ for all $\beta \geq 1$, with $\Lambda = \beta^{1/2} e^{-1/2} (2\pi)^{-d/(2\beta)}$. Thus, Theorem 9 implies that when $d = 3$, the log-concave maximum likelihood estimator attains the rate $\tilde{O}(n^{-4/7})$ in d_X^2 divergence uniformly over the class of Gaussian densities.

EXAMPLE 2 (Spherically symmetric Laplace density).   Writing $V_d := \mu_d(\bar{B}(0,1)) = \pi^{d/2}/\Gamma(1+d/2)$, we see that $f \colon x \mapsto (d!\, V_d)^{-1} e^{-\|x\|}$ is a density in $\mathcal{F}_d$ with corresponding covariance matrix $\Sigma \equiv \Sigma_f = (d+1)I$. For $\tau \leq (d+1)^{d/2} (d!\, V_d)^{-1}$ and any $\beta \geq 1$, we have

$$\|\nabla f(x)\|'_{\Sigma^{-1}} = (d+1)^{(d+1)/2} f(x) \leq \frac{(d+1)^{(d+1)/2}}{(d!\, V_d)^{1/\beta}} f(x)^{1-1/\beta}$$

for all $x \in \mathbb{R}^d$ with $f(x) < \tau \det^{-1/2} \Sigma = \tau(d+1)^{-d/2}$. Hence, when $d = 3$, the log-concave maximum likelihood estimator attains the rate $\tilde{O}(n^{-4/7})$ in d_X^2 divergence uniformly over the class of densities that are affinely equivalent to f, even though f is not differentiable at 0. A similar conclusion holds for the densities f_1, f_2 satisfying $f_1(x) \propto \exp(-e^{\|x\|})$ and $f_2(x) \propto \exp(-e^{e^{\|x\|}})$.

EXAMPLE 3 (Spherically symmetric bump function density).   Consider the smooth density $f \colon x \mapsto C e^{-1/(1-\|x\|^2)} \mathbb{1}_{\{\|x\|<1\}}$, where $C > 0$ is a normalisation constant. By Xu and Samworth (2019, Proposition 2), f is log-concave. Writing $\Sigma \equiv \Sigma_f = \sigma^2 I$ for the covariance matrix corresponding to f, and again fixing an arbitrary $\beta \geq 1$, we see that each $x \in \mathbb{R}^d$ with $\|x\| < 1$ satisfies

$$\|\nabla f(x)\|'_{\Sigma^{-1}} = \sigma^{d+1} \|\nabla f(x)\| = \sigma^{d+1} \frac{2C\|x\|}{(1-\|x\|^2)^2} e^{-1/(1-\|x\|^2)}$$

$$\leq 2\sigma^{d+1} f(x) \log^2 \left(\frac{C}{f(x)} \right) \leq \Lambda_\beta \left\{ f(x) \det^{1/2} \Sigma \right\}^{1-1/\beta},$$

where $\Lambda_\beta := 8C^{1/\beta}\beta^2 e^{-2}\sigma^{1+d/\beta}$. Thus, again by Proposition 8, we deduce that $f \in \mathcal{F}^{(\beta,\Lambda_\beta)}$ for all $\beta \geq 1$. Consequently, when $d = 3$, the log-concave maximum likelihood estimator attains the rate $\tilde{O}(n^{-4/7})$ in d_X^2 divergence uniformly over the class of densities that are affinely equivalent to f.

EXAMPLE 4 (Hölder condition on the log-density).   For $\gamma \in (1,2]$ and $L > 0$, let $\tilde{\mathcal{H}}^{\gamma,L} \equiv \tilde{\mathcal{H}}_d^{\gamma,L}$ denote the subset of densities $f \in \mathcal{F}_d$ such that $\phi := \log f$ is differentiable and

$$(13) \qquad \|\nabla\phi(y) - \nabla\phi(x)\|_{\Sigma_f^{-1}} \leq L\|y - x\|_{\Sigma_f}^{\gamma-1}$$

for all $x, y \in \mathbb{R}^d$. We extend this definition to $\gamma = 1$ by writing $\tilde{\mathcal{H}}^{1,L}$ for the subset of densities $f \in \mathcal{F}_d$ for which $\phi = \log f$ satisfies

$$(14) \qquad |\phi(y) - \phi(x)| \leq L\|y - x\|_{\Sigma_f}.$$

for all $x, y \in \mathbb{R}^d$. Note that the densities in $\tilde{\mathcal{H}}^{1,L}$ can have points of non-differentiability for arbitrarily small values of the density. For instance, if we define $f \in \mathcal{F}_d$ by

$$f(x) \propto \exp\left(-\sum_{r=0}^\infty \frac{\|x\| - r}{2^r} \mathbb{1}_{\{\|x\| \geq r\}}\right),$$

which is not differentiable at any $x \in \mathbb{R}^d$ with integer Euclidean norm, then $f \in \tilde{\mathcal{H}}^{1,L}$ for suitably large $L > 0$.

The careful and non-standard choice of norms in (13) and (14) ensures that the classes $\tilde{\mathcal{H}}^{\gamma,L}$ are affine invariant. Moreover, Proposition S41(iv) in Section S3.1 shows that for each $\beta \geq 1$, there exists $\Lambda' \equiv \Lambda'(\beta, L)$ such that $\bigcup_{\gamma\in[1,2]} \tilde{\mathcal{H}}^{\gamma,L} \subseteq \mathcal{F}^{(\beta,\Lambda')}$. Thus, when $d = 3$, the log-concave maximum likelihood estimator attains the rate $\tilde{O}(n^{-4/7})$ in d_X^2 divergence uniformly over $\bigcup_{\gamma\in[1,2]} \tilde{\mathcal{H}}^{\gamma,L}$.

A related result in the literature is Dümbgen and Rufibach (2009, Theorem 4.1), which applies when $d = 1$, $\gamma \in (1,2]$ and the logarithm of the true fixed $f_0 \in \mathcal{F}_1$ is γ-Hölder on some compact subinterval T of the interior of supp f_0. In this case, the corresponding $\hat{f}_n$ is shown to achieve an adaptive rate of order $\left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\gamma+1}}$ with respect to the supremum norm over certain compact subintervals of the interior of T. We remark that this is not entirely comparable with the rate we obtain in the paragraph above, especially since our loss function d_X^2 is rather different.

Observe that the densities in the classes $\tilde{\mathcal{H}}^{\gamma,L}$ must be supported on the whole of $\mathbb{R}^d$, and that conditions (13) and (14) imply that the rate of tail

decay of f is 'super-Gaussian'. This is quite a stringent restriction; note for example that the density f satisfying $f(x) \propto \exp(-e^{\|x\|})$ does not feature in any of the classes $\tilde{\mathcal{H}}^{\gamma,L}$. Another drawback of this definition of smoothness is that the classes are not nested with respect to the Hölder exponent $\gamma \in (1,2]$; this can be seen by considering a density f satisfying $f(x) \propto \exp(-\|x\|^{\gamma})$, which belongs to $\tilde{\mathcal{H}}^{\tilde{\gamma},L}$ for some $L > 0$ if and only if $\tilde{\gamma} = \gamma$.

EXAMPLE 5 (Hölder condition on the density). To remedy the issues mentioned in the previous example, fix $\beta \in (1,2]$ and $L > 0$ and let $\mathcal{H}^{\beta,L} \equiv \mathcal{H}_d^{\beta,L}$ denote the set of $f \in \mathcal{F}_d$ such that f is differentiable on $\mathbb{R}^d$ and

$$(15) \qquad \|\nabla f(y) - \nabla f(x)\|'_{\Sigma_f^{-1}} \leq L\|y - x\|_{\Sigma_f}^{\beta-1}$$

for all $x, y \in \mathbb{R}^d$. Again, it can be shown that the classes $\mathcal{H}^{\beta,L}$ are affine invariant, and if $f \in \mathcal{F}_d$ is β-Hölder in the usual Euclidean sense, i.e. $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|^{\beta-1}$ for all $x, y \in \mathbb{R}^d$, then $f \in \mathcal{H}^{\beta,\tilde{L}}$ with $\tilde{L} := L\,\lambda_{\max}^{\beta/2}(\Sigma_f)\det^{1/2}\Sigma_f$, where $\lambda_{\max}(\Sigma_f)$ denotes the maximum eigenvalue of Σ_f. This follows from the facts that $\|w\|_{\Sigma_f} \geq \|w\|\,\lambda_{\max}^{-1/2}(\Sigma_f)$ and $\|w\|'_{\Sigma_f^{-1}} \leq \|w\|\,\lambda_{\max}^{1/2}(\Sigma_f)\det^{1/2}\Sigma_f$ for all $w \in \mathbb{R}^d$. Moreover, Proposition S40 shows that the classes $\mathcal{H}^{\beta,L}$ are nested with respect to the Hölder exponent β; more precisely, if β, L are as above, then there exists $\tilde{L} \equiv \tilde{L}(d,\beta,L) > 0$ such that $\mathcal{H}^{\beta,L} \subseteq \mathcal{H}^{\alpha,\tilde{L}}$ for all $\alpha \in (1,\beta]$.

The condition (15) can in fact be extended to an affine invariant notion of β-Hölder regularity for all $\beta > 1$; see Section S3.3 for full technical details. Here, we present the analogue of (15) for $\beta \in (2,3]$ and $L > 0$, for which we require the following additional notation. First, if $g\colon \mathbb{R}^d \to \mathbb{R}$ is twice differentiable at $x \in \mathbb{R}^d$, then denote by $Hg(x) \in \mathbb{R}^{d \times d}$ the Hessian of g at x. In addition, for each $S \in \mathbb{S}^{d \times d}$, define a norm $\|\cdot\|'_S$ on $\mathbb{R}^{d \times d}$ by $\|M\|'_S := \|S^{-1/2}MS^{-1/2}\|_{\mathrm{F}}\det^{-1/2}S$, where $\|A\|_{\mathrm{F}} := \mathrm{tr}(A^\top A)^{1/2}$ denotes the Frobenius norm of $A \in \mathbb{R}^{d \times d}$. We now define $\mathcal{H}^{\beta,L}$ to be the collection of $f \in \mathcal{F}_d$ for which f is twice differentiable on $\mathbb{R}^d$ and

$$(16) \qquad \|Hf(y) - Hf(x)\|'_{\Sigma_f^{-1}} \leq L\|y - x\|_{\Sigma_f}^{\beta-2}$$

for all $x, y \in \mathbb{R}^d$. In Section S3.3, we present a unified argument that establishes the affine invariance of the classes $\mathcal{H}^{\beta,L}$ defined by (15) and (16); see the proof of Lemma S39.

In addition, for each $\beta \in (1,3]$ and $L > 0$, parts (i) and (iii) of Proposition S41 imply that $\mathcal{H}^{\beta,L} \subseteq \mathcal{F}^{(\beta,\Lambda)}$ for some $\Lambda \equiv \Lambda(\beta,L)$; when $\beta \in (1,2]$,

we can take $\Lambda(\beta, L) := L^{1/\beta}(1 - 1/\beta)^{-1+1/\beta}$. It was this fact that motivated our choice of parametrisation in β in (10). Theorem 9 therefore yields the rate $\tilde{O}\big(n^{-\min\left\{\frac{\beta+3}{\beta+7}, \frac{4}{7}\right\}}\big)$ for the log-concave maximum likelihood estimator, uniformly over $\mathcal{H}^{\beta,L}$. An interesting feature of this rate is that, when $\beta \in (1, 9/5)$, it is faster than the rate $O(n^{-\frac{2\beta}{2\beta+3}})$ that can be obtained in squared Hellinger distance for β-Hölder densities that satisfy a 'tail dominance' condition (Goldenshluger and Lepski, 2014, Section 4). For further details of this comparison, see Section S3.3.1. Thus, in this range of β, the log-concavity shape constraint results in a strict improvement in the rates attainable.

5. Appendix: Proofs of main results in Section 2.   The following notation is used in this section and in the supplementary material.

To define bracketing entropy, let $S \subseteq \mathbb{R}^d$ and let $\mathcal{G}$ be a class of non-negative functions whose domains contain S. For $\varepsilon > 0$ and a semi-metric ρ on $\mathcal{G}$, let $N_{[]}(\varepsilon, \mathcal{G}, \rho, S)$ denote the smallest $M \in \mathbb{N}$ for which there exist pairs of functions $\{[g_j^L, g_j^U] : j = 1, \ldots, M\}$ such that $\rho(g_j^U, g_j^L) \leq \varepsilon$ for every $j = 1, \ldots, M$, and such that for every $g \in \mathcal{G}$, there exists $j^* \in \{1, \ldots, M\}$ with $g_{j^*}^L(x) \leq g(x) \leq g_{j^*}^U(x)$ for every $x \in S$. We then define the ε-*bracketing entropy of $\mathcal{G}$ over S with respect to* ρ by $H_{[]}(\varepsilon, \mathcal{G}, \rho, S) := \log N_{[]}(\varepsilon, \mathcal{G}, \rho, S)$ and write $H_{[]}(\varepsilon, \mathcal{G}, \rho) := H_{[]}(\varepsilon, \mathcal{G}, \rho, \mathbb{R}^d)$ when $S = \mathbb{R}^d$.

For each $f_0 \in \mathcal{F}_d$ and $\delta > 0$, let $\mathcal{G}(f_0, \delta) \equiv \mathcal{G}_d(f_0, \delta) := \{f \mathbb{1}_{\operatorname{supp} f_0} : f \in \mathcal{G}_d, d_{\mathrm{H}}(f, f_0) \leq \delta\}$. In addition, let $\mathcal{F}(f_0, \delta) \equiv \mathcal{F}_d(f_0, \delta) = \mathcal{F}_d \cap \mathcal{G}_d(f_0, \delta)$ and let $\tilde{\mathcal{F}}(f_0, \delta) \equiv \tilde{\mathcal{F}}_d(f_0, \delta) := \{f \in \mathcal{F}_d : d_{\mathrm{H}}(f, f_0) \leq \delta\}$. Writing $\|M\| \equiv \|M\|_{\mathrm{op}} := \sup_{\|u\| \leq 1} \|Mu\|$ for the operator norm of a matrix $M \in \mathbb{R}^{d \times d}$, we denote by $\tilde{\mathcal{F}}^{1,\eta} \equiv \tilde{\mathcal{F}}_d^{1,\eta_d} := \{f \in \mathcal{F}_d : \|\mu_f\| \leq 1, \|\Sigma_f - I\| \leq \eta_d\}$ the class of *'near-isotropic'* log-concave densities, where the constant $\eta \equiv \eta_d \in (0, 1)$ is taken from Kim and Samworth (2016, Lemma 6) and depends only on d. Finally, we define $h_2, h_3 \colon (0, \infty) \to (0, \infty)$ by $h_2(x) := x^{-1} \log_+^{3/2}(x^{-1})$ and $h_3(x) := x^{-2}$ respectively.

The proof of Proposition 1 is lengthy and is deferred to Section S2.1. The main goal of this subsection, therefore, is to prove Theorem 2, which proceeds via several intermediate results, including Theorem 3. We begin by stating our main local bracketing entropy result, whose proof is summarised at the end of Section 2. Note that by Proposition S21, the subclass $\mathcal{F}^1(\mathcal{P}^m)$ is non-empty if and only if $m \geq d$.

PROPOSITION 10.   *Let $d \in \{2, 3\}$ and fix $m \in \mathbb{N}$ with $m \geq d$. Then there exist universal constants $\varrho_2, \varrho_3 > 0$ such that whenever $0 < \varepsilon < \delta < \varrho_d$ and*

$f_0 \in \mathcal{F}^1(\mathcal{P}^m)$, *we have*

$$(17) \qquad H_{[]}(2^{1/2}\varepsilon, \mathcal{G}(f_0, \delta), d_{\mathrm{H}}) \lesssim m\left(\frac{\delta}{\varepsilon}\right) \log^3\left(\frac{1}{\delta}\right) \log^{3/2}\left(\frac{\log(1/\delta)}{\varepsilon}\right)$$

when $d = 2$ and

$$(18) \quad H_{[]}(2^{1/2}\varepsilon, \mathcal{G}(f_0, \delta), d_{\mathrm{H}}) \lesssim m\left\{\left(\frac{\delta}{\varepsilon}\right)^2 \log^6\left(\frac{1}{\delta}\right) + \left(\frac{\delta}{\varepsilon}\right)^{3/2} \log^7\left(\frac{1}{\delta}\right)\right\}$$

when $d = 3$.

See Propositions S8 and S9 for details of the initial stages of the proof, which deal with the case where f_0 is the uniform density $f_K := \mu_d(K)^{-1}\mathbb{1}_K$ on some polytope $K \in \mathcal{P}^m$. Here, we turn our attention to the general non-uniform case, where the support of f_0 may be unbounded. Writing $\mathcal{F}^1$ for the subclass of all log-1-affine densities in $\mathcal{F}_d$, we note that any $f \in \mathcal{F}^1$ must take the form $x \mapsto f_{K,\alpha}(x) := c_{K,\alpha}^{-1}\exp(-\alpha^\top x)\mathbb{1}_{\{x \in K\}}$, where $K \subseteq \mathbb{R}^d$ and $\alpha \in \mathbb{R}^d$ are the support and negative log-gradient of f respectively, and $c_{K,\alpha} := \int_K \exp(-\alpha^\top x)\,dx \in (0, \infty)$; see (S75). It follows from the characterisation of $\mathcal{F}^1$ given in Proposition S15 that K and α satisfy the conditions of Proposition S13(ii), which in turn implies that $m_{K,\alpha} := \inf_{x \in K} \alpha^\top x$ is finite. In addition, let $M_{K,\alpha} := \sup_{x \in K} \alpha^\top x \in (-\infty, \infty]$, and for $t \in \mathbb{R}$, define the convex sets

$$
\begin{aligned}
K_{\alpha,t} &:= K \cap \{x \in \mathbb{R}^d : \alpha^\top x = t\}, \\
K_{\alpha,t}^+ &:= K \cap \{x \in \mathbb{R}^d : \alpha^\top x \le t\}, \\
\check{K}_{\alpha,t} &:= K \cap \{x \in \mathbb{R}^d : t - 1 \le \alpha^\top x \le t\},
\end{aligned}
$$

which are all compact by Proposition S13; see Figure S2 for an illustration. Finally, we denote by $\mathcal{F}_\star^1$ the collection of all $f = f_{K,\alpha} \in \mathcal{F}^1$ for which $m_{K,\alpha} = 0$.

PROOF OF PROPOSITION 10. For a fixed $d \in \{2, 3\}$, let $C \equiv C_d := 8d+7$, $\upsilon \equiv \upsilon_d := 2^{-3/2} \wedge \{d^{-1/2}(d+1)^{-(d-1)/2}\}$ and $\varrho \equiv \varrho_d := \{\upsilon_d e^{-C/2}\gamma^{1/2}\} \wedge \nu_d$, where $\gamma \equiv \gamma(d, C)$ and $\nu \equiv \nu_d$ are taken from Lemmas S17 and S26 respectively. For $0 < \varepsilon < \delta < \upsilon$, the important quantity $H_d(\delta, \varepsilon)$ is defined in Proposition S8.

Fix $0 < \varepsilon < \delta < \varrho$ and $m \in \mathbb{N}$ with $m \ge d$. It follows from Corollary S16 and the affine invariance of the Hellinger distance that we need only consider densities $f_0 = f_{K,\alpha} \in \mathcal{F}_\star^1 \cap \mathcal{F}^1(\mathcal{P}^m)$, which have the property that $K \in \mathcal{P}^m$ and $m_{K,\alpha} = 0$. Since Proposition S9 handles the case $\alpha = 0$, we fix an

arbitrary $f_{K,\alpha} \in \mathcal{F}_\star^1 \cap \mathcal{F}^1(\mathcal{P}^m)$ with $\alpha \neq 0$, and set $L := \lceil M_{K,\alpha} \rceil \in \mathbb{N} \cup \{\infty\}$. Now define

$$K_j' := \begin{cases} K_{\alpha,C}^+ & \text{for } j = C \\ \breve{K}_{\alpha,j} & \text{for each } j \in \mathbb{N} \text{ with } C + 1 \leq j \leq L, \end{cases}$$

which is compact for all integers $C \leq j \leq L$. Note also that since $K \in \mathcal{P}^m$, it follows from Bruns and Gubeladze (2009, Theorem 1.6) that $K_C' \in \mathcal{P}^{m+1}$ and $K_j' \in \mathcal{P}^{m+2}$ for all integers $C + 1 \leq j \leq L$.

In addition, let a_+ be the smallest integer $C + 1 \leq j \leq L$ such that $\delta^2 e^{j+1} \mu_d(\breve{K}_{\alpha,j})^{-1} c_{K,\alpha} \geq \upsilon^2$ if such a j exists, and let $a_+ = L + 1$ otherwise. Since $(1/\tilde{\delta})^{d-1} \geq \log^{d-1}(1/\tilde{\delta}) \geq d(d+1)^{d-1} \upsilon^2 \log^{d-1}(1/\tilde{\delta})$ for all $\tilde{\delta} \in (0,1)$, we deduce from (S78) in Lemma S17 that

$$\frac{\delta^2 e^{t+1} c_{K,\alpha}}{\mu_d(\breve{K}_{\alpha,t})} \geq \frac{\delta^2 e^{t+1} c_{K,\alpha}}{dt^{d-1} \mu_d(K_{\alpha,1}^+)} \geq \frac{\delta^2 e^t}{dt^{d-1}} \geq \upsilon^2$$

for all $t \geq (d+1) \log(1/\delta)$, and hence that $a_+ \lesssim \log(1/\delta)$. Next, set $u_j^2 := c \exp\{-(j - a_+)/2\}$ for each integer $a_+ \leq j \leq L$, where $c := 1 - e^{-1/2}$ is chosen to ensure that $\sum_{j=a_+}^L u_j^2 \leq 1$, and also define

$$\varepsilon_j^2 := \begin{cases} 2\varepsilon^2/3 & \text{for } j = C \\ 2\varepsilon^2(a_+ - C)^{-1}/3 & \text{for } j = C + 1, \ldots, a_+ - 1 \\ 2u_j^2 \varepsilon^2/3 & \text{for } j = a_+, \ldots, L. \end{cases}$$

Since $K = \bigcup_{j=C}^L K_j'$ and $\sum_{j=C}^L \varepsilon_j^2 \leq 2\varepsilon^2$, we can write

(19) $H_{[]}(2^{1/2}\varepsilon, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}) \leq H_{[]}(\varepsilon_C, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_C')$

(20) $+ \sum_{j=C+1}^{a_+ - 1} H_{[]}(\varepsilon_j, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_j')$

(21) $+ \sum_{j=a_+}^L H_{[]}(\varepsilon_j, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_j')$,

and we now address each of the terms (19), (20) and (21) in turn. Note that while there are infinitely many summands in (21) when $M_{K,\alpha} = L = \infty$, it will follow from the bounds we obtain that the series remains summable.

For (19), let $A_C := c_{K,\alpha}/\mu_d(K_C')$, which by Lemma S17 satisfies $e^{-C} \leq A_C \leq \gamma^{-1}$. Now for $f \in \mathcal{G}(f_{K,\alpha}, \delta)$, define $\tilde{f}_C \colon \mathbb{R}^d \to [0, \infty)$ by $\tilde{f}_C(x) := A_C \exp(\alpha^\top x) f(x) \mathbb{1}_{\{x \in K_C'\}}$ and observe that

$$\delta^2 \geq \int_{K_C'} (f^{1/2} - f_{K,\alpha}^{1/2})^2 = \int_{K_C'} \frac{e^{-\alpha^\top x}}{A_C} \{\tilde{f}_C^{1/2}(x) - f_{K_C'}^{1/2}(x)\}^2 \, dx$$

$$\geq \frac{e^{-C}}{A_C} \int_{K_C'} (\tilde{f}_C^{1/2} - f_{K_C'}^{1/2})^2,$$

which shows that $\tilde{f}_C \in \mathcal{G}(f_{K'_C}, A_C^{1/2} e^{C/2}\delta)$. Since $\delta < \varrho < \upsilon e^{-C/2}\gamma^{1/2}$, it follows from the above bounds on A_C that

$$(22) \qquad \delta \leq A_C^{1/2} e^{C/2}\delta < \upsilon < 2^{-3/2} \quad \text{and} \quad A_C^{-1/2}\varepsilon_C^{-1} \lesssim \varepsilon^{-1}.$$

Recalling that $K'_C \in \mathcal{P}^{m+1}$, we can now apply Proposition S9 to deduce that there exists an $(A_C^{1/2}\varepsilon_C)$-Hellinger bracketing set $\{[\tilde{g}_\ell^L, \tilde{g}_\ell^U] : 1 \leq \ell \leq N_C\}$ for $\mathcal{G}(f_{K'_C}, A_C^{1/2} e^{C/2}\delta)$ such that

$$(23) \qquad \log N_C \lesssim (m+1) H_d(A_C^{1/2} e^{C/2}\delta, A_C^{1/2}\varepsilon_C) \lesssim m H_d(\delta, \varepsilon).$$

We see that $\{f\mathbb{1}_{K'_C} : f \in \mathcal{G}(f_{K,\alpha}, \delta)\}$ is covered by the brackets $\{[g_\ell^L, g_\ell^U] : 1 \leq \ell \leq N_C\}$ defined by

$$g_\ell^L(x) := A_C^{-1} \exp(-\alpha^\top x)\tilde{g}_\ell^L(x); \quad g_\ell^U(x) := A_C^{-1} \exp(-\alpha^\top x)\tilde{g}_\ell^U(x).$$

Moreover, $\exp(-\alpha^\top x) \leq 1$ for all $x \in K'_C$, so

$$\int_{K'_C}\left(\sqrt{g_\ell^U} - \sqrt{g_\ell^L}\right)^2 = A_C^{-1}\int_{K'_C}\left(\sqrt{\tilde{g}_\ell^U(x)} - \sqrt{\tilde{g}_\ell^L(x)}\right)^2 \exp(-\alpha^\top x)\,dx \leq \varepsilon_C^2$$

for all $1 \leq \ell \leq N_C$. Together with (23), this implies that

$$(24) \qquad H_{[]}(\varepsilon_C, \mathcal{G}(f_{K,\alpha}, \delta), d_H, K'_C) \leq \log N_C \lesssim m H_d(\delta, \varepsilon).$$

For (20), fix an integer $C + 1 \leq j \leq a_+ - 1$ (if such a j exists) and let $A_j := c_{K,\alpha}/\mu_d(K'_j)$. For $f \in \mathcal{G}(f_{K,\alpha}, \delta)$, define $\tilde{f}_j \colon \mathbb{R}^d \to [0, \infty)$ by $\tilde{f}_j(x) := A_j \exp(\alpha^\top x)f(x)\mathbb{1}_{\{x \in K'_j\}}$. Now

$$\delta^2 \geq \int_{K'_j}\left(f^{1/2} - f_{K,\alpha}^{1/2}\right)^2 = \int_{K'_j}\frac{e^{-\alpha^\top x}}{A_j}\left\{\tilde{f}_j^{1/2}(x) - f_{K'_j}^{1/2}(x)\right\}^2 dx$$
$$\geq \frac{e^{-j}}{A_j}\int_{K'_j}\left(\tilde{f}_j^{1/2} - f_{K'_j}^{1/2}\right)^2,$$

so $\tilde{f}_j \in \mathcal{G}(f_{K'_j}, A_j^{1/2} e^{j/2}\delta)$. Since $j \leq a_+ - 1$, it follows from the definition of a_+ that $A_j < \delta^{-2}\upsilon^2 e^{-(j+1)}$. In addition, since $K'_j \subseteq K_{\alpha,j}^+$, we can apply Lemma S17 to deduce that $A_j \geq c_{K,\alpha}/\mu_d(K_{\alpha,j}^+) \geq e^{-j}$. Therefore,

$$\delta \leq A_j^{1/2} e^{j/2}\delta < \upsilon < 2^{-3/2} \quad \text{and} \quad A_j^{-1/2} e^{-(j-1)/2} \lesssim 1.$$

Since $K_j' \in \mathcal{P}^{m+2}$, we can apply Proposition S9 to deduce that there exists an $(A_j^{1/2} e^{(j-1)/2} \varepsilon_j)$-Hellinger bracketing set $\{[\tilde{g}_\ell^L, \tilde{g}_\ell^U] : 1 \leq \ell \leq N_j\}$ for $\mathcal{G}(f_{K_j'}, A_j^{1/2} e^{j/2} \delta)$ such that

$$(25) \qquad \log N_j \lesssim (m+2) H_d\big(A_j^{1/2} e^{j/2} \delta, A_j^{1/2} e^{(j-1)/2} \varepsilon_j\big) \lesssim m H_d(\delta, \varepsilon_j).$$

We see that $\{f \mathbb{1}_{K_j'} : f \in \mathcal{G}(f_{K,\alpha}, \delta)\}$ is covered by the brackets $\{[g_\ell^L, g_\ell^U] : 1 \leq \ell \leq N_j\}$ defined by

$$g_\ell^L(x) := A_j^{-1} \exp(-\alpha^\top x) \tilde{g}_\ell^L(x); \quad g_\ell^U(x) := A_j^{-1} \exp(-\alpha^\top x) \tilde{g}_\ell^U(x).$$

Moreover, $\exp(-\alpha^\top x) \leq e^{-(j-1)}$ for all $x \in K_j'$, so

$$\int_{K_j'} \left(\sqrt{g_\ell^U} - \sqrt{g_\ell^L}\right)^2 = A_j^{-1} \int_{K_j'} \left(\sqrt{\tilde{g}_\ell^U(x)} - \sqrt{\tilde{g}_\ell^L(x)}\right)^2 \exp(-\alpha^\top x)\, dx \leq \varepsilon_j^2$$

for all $1 \leq \ell \leq N_j$. Together with (25) and the fact that $a_+ \lesssim \log(1/\delta)$, this implies that

$$\sum_{j=C+1}^{a_+-1} H_{[]}(\varepsilon_j, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_j') \lesssim \log(1/\delta)\, m H_d\big(\delta,\, \varepsilon/\log(1/\delta)^{1/2}\big),$$

which is bounded above up to a universal constant by

$$(26) \qquad m \left(\frac{\delta}{\varepsilon}\right) \log^3\left(\frac{1}{\delta}\right) \log^{3/2}\left(\frac{\log(1/\delta)}{\varepsilon}\right)$$

when $d = 2$ and

$$(27) \qquad m \left\{\left(\frac{\delta}{\varepsilon}\right)^2 \log^6\left(\frac{1}{\delta}\right) + \left(\frac{\delta}{\varepsilon}\right)^{3/2} \log^7\left(\frac{1}{\delta}\right)\right\}$$

when $d = 3$.

For (21), if $L \geq C+1$, consider $f = e^\phi \in \mathcal{G}(f_{K,\alpha}, \delta)$ and define $\psi \equiv \tilde{\phi}_{K,\alpha} \colon \mathbb{R}^d \to [-\infty, \infty)$ by $\psi(x) := \phi(x) + \alpha^\top x + \log c_{K,\alpha}$, as in the statement of Lemma S26. First, we claim that

$$(28) \qquad \psi(x) \leq \frac{4d+2}{a_+ - 2} \alpha^\top x$$

for all $x \in K \setminus K_{\alpha, a_+-1}^+$. To see this, first set $\tilde{K} := K_{\alpha, a_+-1}^+$ and $\tilde{A} := c_{K,\alpha}/\mu_d(\tilde{K})$, and define $\tilde{f} \colon \mathbb{R}^d \to [0, \infty)$ by $\tilde{f}(x) := \tilde{A} \exp(\alpha^\top x) f(x) \mathbb{1}_{\{x \in \tilde{K}\}}$. Observe that

$$\log \tilde{f}(x) = \log f(x) + \alpha^\top x + \log c_{K,\alpha} - \log \mu_d(\tilde{K}) = \psi(x) - \log \mu_d(\tilde{K}).$$

Then by similar arguments to those given above, we deduce that $\tilde{f} \in \mathcal{G}\big(f_{\tilde{K}}, \tilde{A}^{1/2} e^{(a_+ - 1)/2} \delta\big)$. Moreover, if $a_+ \geq C + 2$, then it follows from the definitions of a_+ and υ that

$$\tilde{A} e^{a_+ - 1} \delta^2 \leq \mu_d(\check{K}_{\alpha, a_+ - 1})^{-1} c_{K,\alpha} e^{a_+ - 1} \delta^2 < e^{-1} \upsilon^2 < 2^{-3}.$$

Otherwise, if $a_+ = C + 1$, then recall from (22) that

$$\tilde{A} e^{a_+ - 1} \delta^2 = A_C e^C \delta^2 < \upsilon^2 < 2^{-3}.$$

Thus, in all cases, Lemma S25(ii) implies that

$$\log \tilde{f}(x) \leq 2^{7/2} d \, (\tilde{A} e^{a_+ - 1} \delta^2)^{1/2} - \log \mu_d(\tilde{K})$$

for all $x \in \tilde{K}$, and hence that $\psi \leq 4d$ on $K_{\alpha, a_+ - 1}$. On the other hand, we know from Lemma S26 that there exists some $x_- \in K_{\alpha, 1}^+$ such that $\psi(x_-) > -2$. Now if $x \in K$ and $\alpha^\top x > a_+ - 1$, then $s := (a_+ - 1 - \alpha^\top x_-)/(\alpha^\top x - \alpha^\top x_-)$ satisfies $1 \geq s \geq (a_+ - 2)/(\alpha^\top x - 1) > 0$, and $w := sx + (1 - s)x_-$ lies in $K_{\alpha, a_+ - 1}$. It then follows from the concavity of ψ that

$$\psi(x) \leq \frac{1}{s} \psi(w) - \frac{1 - s}{s} \psi(x_-) \leq \frac{4d}{s} + \frac{2\,(1 - s)}{s} = \frac{4d + 2}{s} - 2 < \frac{4d + 2}{a_+ - 2} \alpha^\top x,$$

which yields (28), as required.

Now fix an integer $a_+ \leq j \leq L$ (if such a j exists). First, recalling the definition of a_+, we deduce from the bound (S78) in Lemma S17 that

$$(29) \qquad \frac{\mu_d(K_j')}{c_{K,\alpha} e^{a_+}} \leq \left(\frac{j}{a_+ - 1}\right)^{d-1} \frac{\mu_d(\check{K}_{\alpha, a_+})}{c_{K,\alpha} e^{a_+}} \leq e \upsilon^{-2} \delta^2 \left(\frac{j}{a_+ - 1}\right)^{d-1}.$$

Also, it follows from (28) that if $f \in \mathcal{G}(f_{K,\alpha}, \delta)$, then the function $\tilde{f}_j \colon \mathbb{R}^d \to [0, \infty)$ defined by $\tilde{f}_j(x) := c_{K,\alpha} \exp(\alpha^\top x) f(x) \mathbb{1}_{\{x \in K_j'\}}$ belongs to the set $\mathcal{G}_{-\infty, B_j}(K_j') := \big\{ g \mathbb{1}_{K_j'} : g \in \mathcal{G}, \, g \mathbb{1}_{K_j'} \leq e^{B_j} \big\}$, where $B_j := (4d + 2)j/(a_+ - 2)$. Now if $\{[\tilde{g}_\ell^L, \tilde{g}_\ell^U] : 1 \leq \ell \leq N\}$ is a $(c_{K,\alpha}^{1/2} e^{(j-1)/2} \varepsilon_j)$-Hellinger bracketing set for $\mathcal{G}_{-\infty, B_j}(K_j')$, then $\{f \mathbb{1}_{K_j'} : f \in \mathcal{G}(f_{K,\alpha}, \delta)\}$ is covered by the brackets $\{[g_\ell^L, g_\ell^U] : 1 \leq \ell \leq N\}$ defined by

$$g_\ell^L(x) := c_{K,\alpha}^{-1} \exp(-\alpha^\top x) \tilde{g}_\ell^L(x); \quad g_\ell^U(x) := c_{K,\alpha}^{-1} \exp(-\alpha^\top x) \tilde{g}_\ell^U(x).$$

Moreover, $\exp(-\alpha^\top x) \leq e^{-(j-1)}$ for all $x \in K_j'$, so

$$\int_{K_j'} \left(\sqrt{g_\ell^U} - \sqrt{g_\ell^L}\right)^2 = c_{K,\alpha}^{-1} \int_{K_j'} \left(\sqrt{\tilde{g}_\ell^U(x)} - \sqrt{\tilde{g}_\ell^L(x)}\right)^2 \exp(-\alpha^\top x) \, dx \leq \varepsilon_j^2$$

for all $1 \le \ell \le N$. Recalling that $a_+ \ge C = 8d + 7$ and that h_d is a decreasing function for $d = 2, 3$, we now apply (29) and the bound (S35) from Proposition S7 to deduce that

$$H_{[]}\big(\varepsilon_j, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_j'\big) \le H_{[]}\big(c_{K,\alpha}^{1/2} e^{(j-1)/2}\, \varepsilon_j, \mathcal{G}_{-\infty, B_j}(K_j'), d_{\mathrm{H}}\big)$$

$$\lesssim h_d \left(\frac{c_{K,\alpha}^{1/2} e^{(j-1)/2}\, \varepsilon_j}{\mu_d(K_j')^{1/2} \exp\left\{ \frac{(4d+2)\, j}{2(a_+-2)} \right\}} \right)$$

$$\lesssim h_d \left(\left\{ \frac{c_{K,\alpha}}{\mu_d(K_j')} \right\}^{1/2} \frac{e^{(j-a_+)/2} e^{a_+/2} e^{-(j-a_+)/4}\, \varepsilon}{\exp\left\{ \frac{(4d+2)(j-a_+)}{2(a_+-2)} + \frac{(4d+2)\, a_+}{2(a_+-2)} \right\}} \right)$$

$$\lesssim h_d \left(\left\{ \frac{c_{K,\alpha}\, e^{a_+}}{\mu_d(K_j')} \right\}^{1/2} \varepsilon \exp\left\{ -\left(\frac{4d+2}{2(a_+-2)} - \frac{1}{4} \right)(j - a_+) \right\} \right)$$

$$\lesssim h_d \left(\frac{\varepsilon}{\delta} \left(\frac{a_+ - 1}{j} \right)^{\frac{d-1}{2}} \exp\left\{ -\left(\frac{4d+2}{2(a_+-2)} - \frac{1}{4} \right)(j - a_+) \right\} \right)$$

$$\lesssim h_d \left(\frac{\varepsilon}{\delta} \left(\frac{a_+ - 1}{j} \right)^{\frac{d-1}{2}} \exp\left\{ -\left(\frac{4d+2}{2(8d+5)} - \frac{1}{4} \right)(j - a_+) \right\} \right),$$

and we note that $\frac{4d+2}{2(8d+5)} - \frac{1}{4} < 0$. Thus, when $d = 2$, the final expression above is bounded above by a constant multiple of

$$\frac{\delta}{\varepsilon} \log^{3/2}\left(\frac{\delta}{\varepsilon} \right) j^{1/2} \left(\log^{3/2} j \right) \exp\left\{ -\left(\frac{1}{4} - \frac{4d+2}{2(8d+5)} \right)(j - a_+) \right\},$$

where we have used the fact that $\log_+(ax) \le (1 + \log a) \log_+ x$ for all $x > 0$ and $a \ge 1$. It follows that

$$(30) \qquad \sum_{j=a_+}^{L} H_{[]}(\varepsilon_j, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_j') \lesssim \frac{\delta}{\varepsilon} \log^{3/2}\left(\frac{\delta}{\varepsilon} \right)$$

when $d = 2$. Similarly, when $d = 3$, we conclude that

$$(31) \qquad \sum_{j=a_+}^{L} H_{[]}(\varepsilon_j, \mathcal{G}(f_{K,\alpha}, \delta), d_{\mathrm{H}}, K_j') \lesssim \left(\frac{\delta}{\varepsilon} \right)^2.$$

The result follows upon combining the bounds (19), (20), (21), (24), (26), (27), (30) and (31). □

We are now in a position to give the proof of Theorem 3.

PROOF OF THEOREM 3. By the affine equivariance of the log-concave maximum likelihood estimator (Dümbgen et al., 2011, Remark 2.4) and the affine invariance of d_{H}, we may assume without loss of generality that $f_0 \in \mathcal{F}_d^{0,I}$. In addition, by Kim and Samworth (2016, Lemma 6), we have

$$(32) \qquad \sup_{f_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}(\hat{f}_n \notin \tilde{\mathcal{F}}_d^{1,\eta_d}) = O(n^{-1}),$$

where $\tilde{\mathcal{F}}_d^{1,\eta_d}$ is the class of 'near-isotropic' log-concave densities defined at the start of Section 5. For fixed $f_0 \in \mathcal{F}_d$ and $m \geq d$, let

$$\Delta := \inf_{\substack{f \in \mathcal{F}^1(\mathcal{P}^m) \\ \mathrm{supp}\, f_0 \subseteq \mathrm{supp}\, f}} d_{\mathrm{H}}^2(f_0, f).$$

First we consider the case $d = 2$ and assume for the time being that $\Delta \leq \varrho_2/2$, where ϱ_2 is taken from Proposition 10. If $\delta \in (0, \varrho_2 - \Delta)$, then for all $\eta' \in (0, \varrho_2 - \Delta - \delta)$, there exists $f \in \mathcal{F}^1(\mathcal{P}^m)$ with $\mathrm{supp}\, f_0 \subseteq \mathrm{supp}\, f$ such that $d_{\mathrm{H}}(f_0, f) \leq \Delta + \eta'$. It follows from the triangle inequality that $\mathcal{F}(f_0, \delta) \subseteq \mathcal{F}(f, \delta + \Delta + \eta') \subseteq \mathcal{F}(f, \varrho_2)$, and we deduce from the first bound (17) in Proposition 10 that

$$H_{[]}(2^{1/2}\varepsilon, \mathcal{F}(f_0, \delta), d_{\mathrm{H}}) \lesssim m\left(\frac{\delta + \Delta + \eta'}{\varepsilon}\right) \log^3\left(\frac{1}{\delta}\right) \log^{3/2}\left(\frac{\log(1/\delta)}{\varepsilon}\right).$$

But since $\eta' \in (0, \varrho_2 - \Delta - \delta)$ was arbitrary, it follows that

$$(33) \quad H_{[]}(2^{1/2}\varepsilon, \mathcal{F}(f_0, \delta), d_{\mathrm{H}}) \lesssim m\left(\frac{\delta + \Delta}{\varepsilon}\right) \log^3\left(\frac{1}{\delta}\right) \log^{3/2}\left(\frac{\log(1/\delta)}{\varepsilon}\right)$$

and hence that

$$\int_{\delta^2/2^{13}}^{\delta} H_{[]}^{1/2}(\varepsilon, \mathcal{F}(f_0, \delta), d_{\mathrm{H}}) \, d\varepsilon$$

$$(34) \qquad \lesssim m^{1/2}(\delta + \Delta)^{1/2} \log^{3/2}\left(\frac{1}{\delta}\right) \int_0^{\delta} \varepsilon^{-1/2} \log^{3/4}\left(\frac{\log(1/\delta)}{\varepsilon}\right) d\varepsilon.$$

Now for any $a > e\delta$, we can integrate by parts to establish that

$$\int_0^{\delta} \varepsilon^{-1/2} \log^{3/4}\left(\frac{a}{\varepsilon}\right) d\varepsilon = a^{1/2} \int_{\log(a/\delta)}^{\infty} u^{3/4} e^{-u/2} \, du$$

$$= 2\delta^{1/2} \log^{3/4}\left(\frac{a}{\delta}\right) + \frac{3a^{1/2}}{2} \int_{\log(a/\delta)}^{\infty} \frac{e^{-u/2}}{u^{1/4}} \, du$$

$$(35) \qquad \leq 5\delta^{1/2} \log^{3/4}(a/\delta).$$

Thus, setting $a := \log(1/\delta)$ and combining the bounds in (34) and (35), we see that

$$\frac{1}{\delta^2} \int_{\delta^2/2^{13}}^{\delta} H_{[]}^{1/2}(\varepsilon, \mathcal{F}(f_0, \delta) \cap \tilde{\mathcal{F}}^{1, \eta_2}, d_{\mathrm{H}}) \, d\varepsilon \lesssim m^{1/2} \left(\frac{\delta + \Delta}{\delta^3} \right)^{1/2} \log^{9/4} \left(\frac{1}{\delta} \right),$$

where the right-hand side is a decreasing function of $\delta \in (0, \varrho_2 - \Delta)$. On the other hand, if $\delta \geq \varrho_2 - \Delta$, which is at least $\varrho_2/2$, then it follows from Kim and Samworth (2016, Theorem 4) that

$$H_{[]}(\varepsilon, \tilde{\mathcal{F}}^{1, \eta_2}, d_{\mathrm{H}}) \lesssim h_2(\varepsilon) \lesssim \frac{1}{\varepsilon} \log_+^{3/2} \left(\frac{1}{\varepsilon} \right) \lesssim \frac{\delta}{\varepsilon} \log_+^{3/2} \left(\frac{1}{\varepsilon} \right)$$

and hence that

$$\frac{1}{\delta^2} \int_{\delta^2/2^{13}}^{\delta} H_{[]}^{1/2}(\varepsilon, \mathcal{F}(f_0, \delta) \cap \tilde{\mathcal{F}}^{1, \eta_2}, d_{\mathrm{H}}) \, d\varepsilon \lesssim \frac{1}{\delta} \log_+^{3/4} \left(\frac{1}{\delta} \right).$$

Consequently, there exists a universal constant $C_2' > 0$ such that the function $\Psi_2 : (0, \infty) \to (0, \infty)$ defined by

$$\Psi_2(\delta) := C_2' \, m^{1/2} \, \delta^{1/2} (\delta + \Delta)^{1/2} \, \log_+^{9/4}(1/\delta)$$

satisfies $\Psi_2(\delta) \geq \delta \vee \int_{\delta^2/2^{13}}^{\delta} H_{[]}^{1/2}(\varepsilon, \mathcal{F}(f_0, \delta) \cap \tilde{\mathcal{F}}^{1, \eta_2}, d_{\mathrm{H}}) \, d\varepsilon$ for all $\delta > 0$ and has the property that $\delta \mapsto \delta^{-2} \Psi_2(\delta)$ is decreasing. Setting $c_2 := 2^{69/4} C_2' \vee 1$ and $\delta_n := (c_2^2 m n^{-1} \log^{9/2} n + \Delta^2)^{1/2}$, we have $\Delta \leq \delta_n$ and $\delta_n^{-1} \leq c_2^{-1} m^{-1/2} n^{1/2} \log^{-9/4} n \leq n^{1/2}$, so

$$(36) \qquad \delta_n^{-2} \Psi_2(\delta_n) \leq 2^{1/2} C_2' m^{1/2} \delta_n^{-1} \log^{9/4}(n^{1/2}) \leq 2^{-19} n^{1/2}.$$

We are now in a position to apply van de Geer (2000, Corollary 7.5), which is restated as Theorem 10 in the online supplement to Kim et al. (2018). It follows from this, (32) and the bound (S2) from Lemma S1 that there are universal constants $\bar{C}, c, c', c'' > 0$ such that

$$\mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \int_0^{8d \log n} \mathbb{P}\big[\{d_X^2(\hat{f}_n, f_0) \geq t\} \cap \{\hat{f}_n \in \tilde{\mathcal{F}}^{1, \eta_2}\}\big] \, dt$$

$$+ (8d \log n) \, \mathbb{P}(\hat{f}_n \notin \tilde{\mathcal{F}}^{1, \eta_2}) + \int_{8d \log n}^{\infty} \mathbb{P}\{d_X^2(\hat{f}_n, f_0) \geq t\} \, dt$$

$$\leq \delta_n^2 + \int_{\delta_n^2}^{\infty} c \exp(-nt/c^2) \, dt + c' n^{-1} \log n + c'' n^{-3}$$

$$(37) \qquad \leq \delta_n^2 + 2c' n^{-1} \log n \leq \frac{\bar{C} m}{n} \log^{9/2} n + \Delta^2$$

for all $n \geq 3$, provided that $\Delta \leq \varrho_2/2$. On the other hand, when $\Delta > \varrho_2/2$, observe that by Theorem S2, which is a small modification of Kim and Samworth (2016, Theorem 5), we have $\mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \lesssim n^{-2/3} \log n \lesssim (\varrho_2/2)^2 \leq \Delta^2$. We have now established the $d = 2$ case of the desired result.

The proof for the case $d = 3$ is very similar in most respects, except that the first term in the local bracketing entropy bound (18) from Proposition 10 gives rise to a divergent entropy integral. If $\Delta \leq \varrho_3/2$, then

$$\frac{1}{\delta^2} \int_{\delta^2/2^{13}}^{\delta} H_{[]}^{1/2}(\varepsilon, \mathcal{F}(f_0, \delta) \cap \tilde{\mathcal{F}}^{1, \eta_3}, d_{\mathrm{H}}) \, d\varepsilon \lesssim m \left(\frac{\delta + \Delta}{\delta^2} \right) \log_+^4 \left(\frac{1}{\delta} \right)$$

for all $\delta > 0$, where we once again appeal to the global entropy bound

$$H_{[]}(\varepsilon, \tilde{\mathcal{F}}^{1, \eta_3}, d_{\mathrm{H}}) \lesssim h_3(\varepsilon) \lesssim \frac{1}{\varepsilon^2}$$

from Kim and Samworth (2016, Theorem 4) to handle the case $\delta \geq \varrho_3 - \Delta$. We conclude as above that there exists $C_3' > 0$ such that the function $\Psi_3 \colon (0, \infty) \to (0, \infty)$ defined by

$$\Psi_3(\delta) := C_3' \, m^{1/2} (\delta + \Delta) \log_+^4 (1/\delta)$$

has all the required properties. Also, if we set $c_3 := 2^{16} C_3' \vee 1$, then $\delta_n := \left(c_3^2 \, mn^{-1} \log^8 n + \Delta^2 \right)^{1/2}$ satisfies $\delta_n^{-2} \Psi_3(\delta_n) \leq 2^{-19} n^{1/2}$ for all $n \geq 4$. The rest of the argument above then goes through, and we once again use the worst-case bound $\mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \lesssim n^{-1/2} \log n$ from Theorem S2 to handle the case where $\Delta > \varrho_3/2$. $\qquad \square$

PROOF OF PROPOSITION 4. Observe that in Proposition S9, the poly-logarithmic exponents in the local bracketing entropy bounds for uniform densities on polytopes in $\mathcal{P}^m$ are smaller than those that appear in Proposition 10. We can therefore exploit this and deduce Proposition 4 from Proposition S9 in the same way as Theorem 3 is derived from Proposition 10. We omit the details for brevity. $\qquad \square$

Now that we have established our main novel results of this section, the proof of Theorem 2 is broadly similar to that of the univariate oracle inequality stated as Theorem 3 in Kim et al. (2018), so our exposition will be brief, and we will seek to emphasise the main points of difference.

PROOF OF THEOREM 2. Fix $f_0 \in \mathcal{F}$ and an arbitrary $f \in \bigcup_{m \in \mathbb{N}} \mathcal{F}^k(\mathcal{P}^m)$ such that $\mathrm{KL}(f_0, f) < \infty$. Note that we must have $\mathrm{supp}\, f_0 \subseteq \mathrm{supp}\, f$.

Proposition 1 yields a polyhedral subdivision $E_1, \ldots, E_\ell$ of $\operatorname{supp} f \in \mathcal{P}$ with $\ell := \kappa(f) \le k$ such that $\log f$ is affine on each E_j, and recall that $\Gamma(f) = \sum_{j=1}^{\ell} d_j$, where $d_j := |\mathcal{F}(E_j)|$. Setting $p_j := \int_{E_j} f_0$ and $q_j := \int_{E_j} f$ for each $j \in \{1, \ldots, \ell\}$, we see that $\sum_{j=1}^{\ell} p_j = \sum_{j=1}^{\ell} q_j = 1$. Moreover, let $N_j := \sum_{i=1}^{n} \mathbb{1}_{\{X_i \in E_j\}}$ for each $j \in \{1, \ldots, \ell\}$, and partition the set of indices $\{1, \ldots, \ell\}$ into the subsets $J_1 := \{j : N_j \ge d+1\}$ and $J_2 := \{j : N_j \le d\}$. Then $|J_2| \le d\ell$ and

$$(38) \qquad d_X^2(\hat{f}_n, f_0) \le \frac{1}{n} \sum_{j \in J_1} \sum_{i:X_i \in E_j} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} + \frac{d\ell}{n} \max_{1 \le i \le n} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)}.$$

The bound (S1) from Lemma S1 controls the expectation of the second term on the right-hand side of (38), so it remains to handle the first term. For each $j \in J_1$, let $f_0^{(j)}, f^{(j)} \in \mathcal{F}$ be the functions defined by $f_0^{(j)}(x) := p_j^{-1} f_0(x) \mathbb{1}_{\{x \in E_j\}}$ and $f^{(j)}(x) := q_j^{-1} f(x) \mathbb{1}_{\{x \in E_j\}}$. We also denote by $\hat{f}^{(j)}$ the maximum likelihood estimator based on $\{X_1, \ldots, X_n\} \cap E_j$, which exists and is unique with probability 1 for each $j \in J_1$ (Dümbgen et al., 2011, Theorem 2.2). Writing $M_1 := \sum_{j \in J_1} N_j$ and arguing as in Kim et al. (2018), we find that

$$\sum_{j \in J_1} \sum_{i:X_i \in E_j} \hat{f}_n(X_i) \le \sum_{j \in J_1} \sum_{i:X_i \in E_j} \frac{N_j}{M_1} \hat{f}^{(j)}(X_i).$$

It follows that

$$\frac{1}{n} \mathbb{E} \left\{ \sum_{j \in J_1} \sum_{i:X_i \in E_j} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} \right\} \le \frac{1}{n} \mathbb{E} \left\{ \sum_{j \in J_1} \sum_{i:X_i \in E_j} \log \frac{N_j \hat{f}^{(j)}(X_i)/M_1}{p_j f_0^{(j)}(X_i)} \right\}$$

$$(39) \qquad = \frac{1}{n} \mathbb{E} \left\{ \sum_{j \in J_1} \sum_{i:X_i \in E_j} \log \frac{\hat{f}^{(j)}(X_i)}{f_0^{(j)}(X_i)} \right\} + \mathbb{E} \left(\sum_{j \in J_1} \frac{N_j}{n} \log \frac{N_j}{np_j} \right)$$

$$+ \mathbb{E} \left(\frac{M_1}{n} \log \frac{n}{M_1} \right)$$

$$=: r_1 + r_2 + r_3.$$

To bound r_1, we observe that $f^{(j)} \in \mathcal{F}^1(\mathcal{P}^{d_j})$ and $\operatorname{supp} f_0^{(j)} \subseteq \operatorname{supp} f^{(j)}$ for each $j \in J_1$. Consequently, after conditioning on the set of random variables $\{N_j : j = 1, \ldots, \ell\}$, we can apply the risk bound in Theorem 3 to each $f_0^{(j)}$

and the corresponding $\hat{f}^{(j)}$ to deduce that

$$r_1 \leq \frac{1}{n} \mathbb{E}\left(\sum_{j \in J_1} N_j \left\{ \frac{\bar{C} d_j}{N_j} \log^{\gamma_d} N_j + \inf_{\substack{f_1 \in \mathcal{F}^1(\mathcal{P}^{d_j}) \\ \operatorname{supp} f_0^{(j)} \subseteq \operatorname{supp} f_1}} d_{\mathrm{H}}^2\big(f_0^{(j)}, f_1\big) \right\} \right)$$

$$\leq \frac{\bar{C}\, \Gamma(f)}{n} \log^{\gamma_d} n + \sum_{j=1}^{\ell} p_j\, d_{\mathrm{H}}^2\big(f_0^{(j)}, f^{(j)}\big)$$

$$(40) \qquad \leq \frac{\bar{C}\, \Gamma(f)}{n} \log^{\gamma_d} n + \mathrm{KL}(f_0, f),$$

where the penultimate inequality follows as in the proof of Kim et al. (2018, Theorem 3). Moreover,

(41)

$$r_2 \leq \sum_{j=1}^{\ell} \mathbb{E}\left\{ \frac{N_j}{n}\left(\frac{N_j}{np_j} - 1\right) \right\} - \mathbb{E}\left(\sum_{j \in J_2} \frac{N_j}{n} \log \frac{N_j}{np_j} \right) \leq \frac{\ell}{n} + \frac{d\ell}{n} \log n.$$

Finally, for r_3, we first suppose that $d\ell < n/2$, in which case $M_1/n \geq 1 - (d\ell)/n > 1/2$. Thus, arguing as in Kim et al. (2018), we deduce that $r_3 \leq (2\ell d)/n$. Together with (39), (40), (41) and the fact that $\ell \leq \Gamma(f)$, this implies that the desired bound (6) holds whenever $d\ell < n/2$. On the other hand, if $d\ell \geq n/2$, then $\Gamma(f)/n \gtrsim 1$ and we can apply Lemma S1 again to conclude that

$$\mathbb{E}\{d_X^2(\hat{f}_n, f_0)\} \leq \mathbb{E}\left\{ \max_{1 \leq i \leq n} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} \right\} \lesssim \log n \lesssim \frac{\Gamma(f)}{n} \log^{\gamma_d} n.$$

This completes the proof of (6). The final assertion of Theorem 2 follows from Lemma S23 in the case $d = 2$ and from the final assertion of Proposition 1 in the case $d = 3$. □

References.

Amelunxen, D., Lotz, M., McCoy, M. B. and Tropp, J. A. (2014). Living on the edge: phase transition in convex programs with random data. *Inf. Inference*, **3**, 224–294.

Balabdaoui, F., Rufibach, K. and Wellner, J. A. (2009). Limit distribution theory for pointwise maximum likelihood estimation of a log-concave density. *Ann. Statist.*, **37**, 1299–1331.

Baraud, Y. and Birgé, L. (2016). Rates of convergence of rho-estimators for sets of densities satisfying shape constraints. *Stoch. Proc. Appl.*, **12**, 3888–3912.

Bellec, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.*, **46**, 745–780.

Birgé, L. (1987). Estimating a density under order restrictions: nonasymptotic minimax risk. *Ann. Statist.*, **15**, 995–1012.

Brunel, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Statist.*, **7**, 1301–1327.

Bruns, W. and Gubeladze, J. (2009). *Polytopes, Rings and K-theory*. Springer–Verlag, New York.

Cai, T. T. and Low, M. (2015). A framework for estimation of convex functions. *Statist. Sinica*, **25**, 423–456.

Carpenter, T., Diakonikolas, I., Sidiropoulos, A. and Stewart A. (2018). Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. *Proc. Mach. Learn. Res.*, **75**, 1–29.

Chatterjee, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.*, **42**, 2340–2381.

Chatterjee, S., Guntuboyina, A. and Sen, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.*, **43**, 1774–1800.

Chatterjee, S., Guntuboyina, A. and Sen, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli*, **24**, 1072–1100.

Chatterjee, S. and Lafferty, J. (2019). Adaptive risk bounds in unimodal regression. *Bernoulli*, **25**, 1–25.

Chen, Y. and Samworth, R. J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica*, **23**, 1373–1398.

Chen, Y. and Samworth, R. J. (2016). Generalised additive and index models with shape constraints. *J. Roy. Statist. Soc., Ser. B*, **78**, 729–754.

Deng, H. and Zhang, C.-H. (2018). Isotonic regression in multi-dimensional spaces and graphs. Available at https://arxiv.org/abs/1812.08944.

Doss, C. R. and Wellner, J. A. (2016). Global rates of convergence of the MLEs of log-concave and *s*-concave densities. *Ann. Statist.*, **44**, 954–981.

Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, **15**, 40–68.

Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.*, **39**, 702–730.

Feng, O. Y., Guntuboyina, A., Kim, A. K. H. and Samworth, R. J. (2020). Supplementary material for 'Adaptation in multivariate log-concave density estimation'. *Submitted*.

Gardner, R. J., Kiderlen, M. and Milanfar, P. (2006). Convergence of algorithms for reconstructing convex bodies and directional measures. *Ann. Statist.*, **34**, 1331–1374.

Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on $\mathbb{R}^d$. *Probab. Theory Related Fields*, **159**, 479–543.

Groeneboom, P. and Hendrickx, K. (2018). Current status linear regression. *Ann. Statist.*,

46, 1415–1444.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints.* Cambridge University Press, Cambridge.

Guntuboyina, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.*, **40**, 385–411.

Guntuboyina, A. and Sen, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inf. Th.*, **59**, 1957–1965.

Han, Q. (2019). Global empirical risk minimizers with "shape constraints" are rate optimal in general dimensions. Available at https://arxiv.org/abs/1905.12823v1.

Han, Q., Wang, T., Chatterjee, S. and Samworth, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.*, **47**, 2440–2471.

Han, Q. and Wellner, J. A. (2016). Multivariate convex regression: global risk bounds and adaptation. Available at https://arxiv.org/abs/1601.06844.

Jankowski, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.*, **42**, 635–653.

Kim, A. K. H. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.*, **44**, 2756–2779.

Kim, A. K. H., Guntuboyina, A. and Samworth, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.*, **46**, 2279–2306.

Koenker, R. and Mizera, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Amer. Statist. Assoc.*, **109**, 674–685.

Kur, G., Dagan, Y. and Rakhlin, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. Available at https://arxiv.org/abs/1903.05315.

Lovász, L. and Vempala, S. (2006). The geometry of logconcave functions and sampling algorithms. *Random Struct. Alg.*, **30**, 307–358.

Mazumder, R., Choudhury, A., Iyengar, G. and Sen, B. (2018). A computational framework for multivariate convex regression and its variants. *J. Amer. Statist. Assoc.*, **114**, 318–331.

Samworth, R. J. (2018). Recent progress in log-concave density estimation. *Stat. Sci.*, **33**, 493–509.

Samworth, R. J. and Yuan, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.*, **40**, 2973–3002.

Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: a review. *Statist. Surveys*, **8**, 45–114.

Schneider, R. (2014). *Convex Bodies: The Brunn–Minkowski Theory*, 2nd edition. Cambridge University Press, Cambridge.

Seregin, A. and Wellner, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.*, **38**, 3751–3781.

Shah, N., Balakrishnan, S., Guntuboyina, A. and Wainwright, M. J. (2017). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Trans. Inf. Th.*, **63**, 934–959.

Soloff, J. A., Guntuboyina, A. and Pitman, J. (2019). Distribution-free properties of isotonic regression. *Electron. J. Statist.*, **13**, 3243–3253.

van de Geer, S. A. (2000). *Empirical Processes in M-Estimation.* Cambridge University Press, Cambridge.

Walther, G. (2009). Inference and modeling with log-concave densities. *Stat. Sci.*, **24**, 319–327.

Wieacker, J. A. (1987). *Einige Probleme der polyedrischen Approximation.* Diplomarbeit, Freiburg.

Xu, M., Chen, M. and Lafferty, J. (2016). Faithful variable screening for high-dimensional convex regression. *Ann. Statist.*, **44**, 2624–2660.

Xu, M. and Samworth, R. J. (2019). High-dimensional nonparametric density estimation via symmetry and shape constraints. Available at https://arxiv.org/abs/1903.06092.

Zhang, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.*, **30**, 528–555.

STATISTICAL LABORATORY
WILBERFORCE ROAD
CAMBRIDGE
CB3 0WB
UNITED KINGDOM
E-MAIL: oyf20@cam.ac.uk
URL: https://www.dpmms.cam.ac.uk/people/oyf20

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
423 EVANS HALL
BERKELEY, CA - 94720
UNITED STATES OF AMERICA
E-MAIL: aditya@stat.berkeley.edu
URL: http://www.stat.berkeley.edu/~aditya/

DEPARTMENT OF STATISTICS
KOREA UNIVERSITY
145 ANAM-RO, SEONGBUK-GU
SEOUL 02841
SOUTH KOREA
E-MAIL: arlenent@korea.ac.kr
URL: http://sites.google.com/site/kyoungheearlene/home

STATISTICAL LABORATORY
WILBERFORCE ROAD
CAMBRIDGE
CB3 0WB
UNITED KINGDOM
E-MAIL: r.samworth@statslab.cam.ac.uk
URL: http://www.statslab.cam.ac.uk/~rjs57