The Annals of Statistics
2016, Vol. 44, No. 6, 2756–2779
DOI: 10.1214/16-AOS1480

GLOBAL RATES OF CONVERGENCE IN LOG-CONCAVE DENSITY ESTIMATION

BY ARLENE K. H. KIM AND RICHARD J. SAMWORTH[1]

University of Cambridge

The estimation of a log-concave density on $\mathbb{R}^d$ represents a central problem in the area of nonparametric inference under shape constraints. In this paper, we study the performance of log-concave density estimators with respect to global loss functions, and adopt a minimax approach. We first show that no statistical procedure based on a sample of size n can estimate a log-concave density with respect to the squared Hellinger loss function with supremum risk smaller than order $n^{-4/5}$, when $d = 1$, and order $n^{-2/(d+1)}$ when $d \geq 2$. In particular, this reveals a sense in which, when $d \geq 3$, log-concave density estimation is fundamentally more challenging than the estimation of a density with two bounded derivatives (a problem to which it has been compared). Second, we show that for $d \leq 3$, the Hellinger ε-bracketing entropy of a class of log-concave densities with small mean and covariance matrix close to the identity grows like $\max\{\varepsilon^{-d/2}, \varepsilon^{-(d-1)}\}$ (up to a logarithmic factor when $d = 2$). This enables us to prove that when $d \leq 3$ the log-concave maximum likelihood estimator achieves the minimax optimal rate (up to logarithmic factors when $d = 2, 3$) with respect to squared Hellinger loss.

1. Introduction. Log-concave densities on $\mathbb{R}^d$, namely those expressible as the exponential of a concave function that takes values in $[-\infty, \infty)$, form a particularly attractive infinite-dimensional class. Gaussian densities are of course log-concave, as are many other well-known families, such as uniform densities on convex sets, Laplace densities and many others. Moreover, the class retains several of the properties of normal densities that make them so widely-used for statistical inference, such as closure under marginalisation, conditioning and convolution operations. On the other hand, the set is small enough to allow fully automatic estimation procedures, for example, using maximum likelihood, where more traditional nonparametric methods would require troublesome choices of smoothing parameters. Log-concavity therefore offers statisticians the potential of freedom from restrictive parametric (typically Gaussian) assumptions without paying a hefty price. Indeed, in recent years, researchers have sought to exploit these alluring features to propose new methodology for a wide range of statistical problems, including the

Received April 2014; revised March 2016.

[1]Supported by an EPSRC Early Career Fellowship and a grant from the Leverhulme Trust.

MSC2010 subject classifications. 62G07, 62G20.

Key words and phrases. Bracketing entropy, density estimation, global loss function, log-concavity, maximum likelihood estimation.

detection of the presence of mixing [Walther (2002)], tail index estimation [Müller and Rufibach (2009)], clustering [Cule, Samworth and Stewart (2010)], regression [Dümbgen, Samworth and Schuhmacher (2011)], Independent Component Analysis [Samworth and Yuan (2012)] and classification [Chen and Samworth (2013)].

However, statistical procedures based on log-concavity, in common with other methods based on shape constraints, present substantial theoretical challenges and these have therefore also been the focus of much recent research. For instance, the maximum likelihood estimator of a log-concave density, first studied by Walther (2002) in the case $d = 1$, and by Cule, Samworth and Stewart (2010) for general d, plays a central role in all of the procedures mentioned in the previous paragraph. Through a series of papers [Cule and Samworth (2010), Dümbgen and Rufibach (2009), Dümbgen, Samworth and Schuhmacher (2011), Pal, Woodroofe and Meyer (2007), Schuhmacher and Dümbgen (2010), Seregin and Wellner (2010)], we now have a fairly complete understanding of the global consistency properties of the log-concave maximum likelihood estimator (even under model misspecification).

Results on the global rate of convergence in log-concave density estimation are, however, less fully developed, and in particular have been confined to the case $d = 1$. For a fixed true log-concave density f_0 belonging to a Hölder ball of smoothness $\beta \in [1, 2]$, Dümbgen and Rufibach (2009) studied the supremum distance over compact intervals in the interior of the support of f_0. They proved that the log-concave maximum likelihood estimator $\hat{f}_n$ based on a sample of size n converges in these metrics to f_0 at rate $O_p(\rho_n^{-\beta/(2\beta+1)})$, where $\rho_n := n/\log n$; thus $\hat{f}_n$ attains the same rates in the stated regimes as other adaptive nonparametric estimators that do not satisfy the shape constraint. Very recently, Doss and Wellner (2016) introduced a new bracketing argument to obtain a rate of convergence of $O_p(n^{-4/5})$ in squared Hellinger distance [defined in (3) below] in the case $d = 1$, again for a fixed true log-concave density f_0.

In this paper, we present several new results on global rates of convergence in log-concave density estimation, with a focus on a minimax approach. We begin by proving, in Theorem 1 in Section 2, a minimax lower bound which shows that for the squared Hellinger loss function, no statistical procedure based on a sample of size n can estimate a log-concave density with supremum risk smaller than order $n^{-4/5}$ when $d = 1$, and order $n^{-2/(d+1)}$ when $d \geq 2$. The surprising feature of this result is that it is often thought that estimation of log-concave densities should be similar to the estimation of densities with two bounded derivatives, for which the minimax rate is known to be $n^{-4/(d+4)}$ for all $d \in \mathbb{N}$ [Ibragimov and Khas'minskii (1983)]. The reasoning for this intuition appears to be Aleksandrov's theorem [Aleksandrov (1939)], which states that a convex function on $\mathbb{R}^d$ is twice differentiable (Lebesgue) almost everywhere in its domain, and the fact that for twice continuously differentiable functions, convexity is equivalent to a second derivative condition, namely that the Hessian matrix is nonnegative definite. Thus, the minimax lower bound in Theorem 1 reveals that while this intuition is valid when $d \leq 2$ [note that $4/(d+4) = 2/(d+1) = 2/3$ when $d = 2$], log-concave density estimation in three or more dimensions is fundamentally more challenging in this minimax sense than estimating a density with two bounded derivatives.

The second main purpose of this paper is to provide bounds on the supremum risk with respect to the squared Hellinger loss function of a particular estimator, namely the log-concave maximum likelihood estimator $\hat{f}_n$. The empirical process theory for studying maximum likelihood estimators is well known [e.g., van de Geer (2000), van der Vaart and Wellner (1996)], but relies on obtaining a bracketing entropy bound, which therefore becomes our main challenge. A first step is to show that after standardising the data and using the affine equivariance of the estimator, we can reduce the problem to maximising over a class $\mathcal{G}$ of log-concave densities having a small mean and covariance matrix close to the identity; see Lemma 6 in Section A.2. In Corollary 3 in Section 3, we present an integrable envelope function for such classes.

The first part of Section 4 is devoted to developing the key bracketing entropy results for the class $\mathcal{G}$. In particular, we show that for $d \leq 3$, the ε-bracketing entropy of $\mathcal{G}$ in Hellinger distance h, denoted $\log N_{[\cdot]}(\varepsilon, \mathcal{G}, h)$ and defined at the beginning of Section 4, satisfies

$$\log N_{[\cdot]}(\varepsilon, \mathcal{G}, h) \asymp \max\{\varepsilon^{-d/2}, \varepsilon^{-(d-1)}\} \tag{1}$$

as $\varepsilon \searrow 0$, up to a multiplicative logarithmic factor when $d = 2$. Incidentally, the lower bound in (1) holds for all dimensions d. The second term on the right-hand side of (1), which dominates the first when $d \geq 3$, is somewhat unexpected in view of standard entropy bounds for classes of convex functions on a compact domain taking values in [0, 1] [e.g., Guntuboyina and Sen (2013), van der Vaart and Wellner (1996)], where only the first term on the right-hand side of (1) appears. Roughly speaking, it arises from the potential complexity of the domains of the log-densities and the fact that these log-densities are not bounded below. These upper bounds rely on intricate calculations of the bracketing entropy of classes of bounded, concave functions on an arbitrary closed, convex domain. Further details on these bounds can be found in Section 4.

In the second part of Section 4, we apply the bracketing entropy bounds described above to deduce that

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

where $\mathcal{F}_d$ denotes the set of upper semi-continuous, log-concave densities on $\mathbb{R}^d$. Thus, for $d \leq 3$, the log-concave maximum likelihood estimator attains the minimax optimal rate of convergence with respect to the squared Hellinger loss function, up to logarithmic factors when $d = 2, 3$. The stated rate when $d = 3$ is slower in terms of the exponent of n than had been conjectured in the literature [e.g., Seregin and Wellner (2010), page 3778], and arises as a consequence of the bracketing entropy being of order $\varepsilon^{-(d-1)} = \varepsilon^{-2}$ for this dimension.

It is interesting to note that the logarithmic penalties that appear in (2) when $d = 2, 3$ occur for different reasons. When $d = 2$, the penalty arises from the logarithmic term in the upper bound for the relevant bracketing entropy; cf. Theorem 4. When $d = 3$, the bracketing bound is sharp up to multiplicative constants, and the logarithmic penalty is due to the divergence of the bracketing entropy integral that plays the crucial role in the empirical process theory. The bracketing entropy lower bound in (1) suggests (but does not prove) that the log-concave maximum likelihood estimator will be rate suboptimal for $d \geq 4$; indeed, Birgé and Massart (1993) give an example of a situation where a maximum likelihood estimator has a suboptimal rate of convergence agreeing with that predicted by the same empirical process theory from which we derive our rates.

The proofs of our main results are given in the Appendix, with the exception of the proof of Theorem 1, which is given in the online supplementary material [Kim and Samworth (2016)], hereafter referred to as the online supplement, along with several auxiliary results. We conclude this section with some generic notation used throughout the paper. If $C \subseteq \mathbb{R}^d$ is convex, let C^c, $\mathrm{bd}(C)$ and $\dim(C)$ denote its complement, boundary and dimension, respectively. Let $\bar{B}_d(x_0, \delta)$ denote the closed Euclidean ball in $\mathbb{R}^d$ of radius $\delta > 0$ centred at x_0.

2. Minimax lower bounds. Let μ_d denote Lebesgue measure on $\mathbb{R}^d$, and recall that $\mathcal{F}_d$ denotes the set of upper semi-continuous, log-concave densities with respect to μ_d, equipped with the σ-algebra it inherits as a subset of $L_1(\mathbb{R}^d)$. Thus, each $f \in \mathcal{F}_d$ can be written as $f = e^{\phi}$, for some upper semi-continuous, concave $\phi : \mathbb{R}^d \to [-\infty, \infty)$; in particular, we do not insist that f is positive everywhere. Let $X_1, \ldots, X_n$ be independent and identically distributed random vectors having some density $f \in \mathcal{F}_d$, and let $\mathbb{P}_f$ and $\mathbb{E}_f$ denote the corresponding probability and expectation operators, respectively. An *estimator* $\tilde{f}_n$ of f is a measurable function from $(\mathbb{R}^d)^{\times n}$ to the class of probability densities with respect to μ_d, and we write $\tilde{\mathcal{F}}_n$ for the class of all such estimators. For $f, g \in L_1(\mathbb{R}^d)$, we define their squared Hellinger distance by

$$h^2(f, g) := \int_{\mathbb{R}^d} (f^{1/2} - g^{1/2})^2 \, d\mu_d. \tag{3}$$

This metric is both affine invariant and particularly convenient for studying maximum likelihood estimators. Adopting a minimax approach, we define the *supremum risk*

$$R(\tilde{f}_n, \mathcal{F}_d) := \sup_{f_0 \in \mathcal{F}_d} \mathbb{E}_{f_0}\{h^2(\tilde{f}_n, f_0)\};$$

our aim in this section is to provide a lower bound for the infimum of $R(\tilde{f}_n, \mathcal{F}_d)$ over $\tilde{f}_n \in \tilde{\mathcal{F}}_n$.

THEOREM 1. *For each $d \in \mathbb{N}$, there exists $c_d > 0$ such that for sufficiently large $n \in \mathbb{N}$,*

$$\inf_{\tilde{f}_n \in \tilde{\mathcal{F}}_n} R(\tilde{f}_n, \mathcal{F}_d) \geq \begin{cases} c_1 n^{-4/5}, & \text{if } d = 1, \\ c_d n^{-2/(d+1)}, & \text{if } d \geq 2. \end{cases}$$

Theorem 1 reveals that when $d \geq 3$, the minimax lower bound rate for squared Hellinger loss is different from that for interior point estimation established under the local strong log-concavity condition in Seregin and Wellner (2010).

In our proof for the case $d = 1$, given in the online supplement, we apply Theorem 1 of Yang and Barron (1999), which provides a minimax lower bound for general parameter spaces and wide classes of squared loss functions L^2. It relies on an upper bound for the ε-covering number of the space with respect to Kullback–Leibler divergence, as well as a lower bound on the ε-packing number of the space with respect to L (which is the Hellinger distance in our case). We can readily obtain such upper and lower bounds, of the same order in ε, for a subset of $\mathcal{F}_1$ consisting of densities that are compactly supported and bounded away from zero on their support. For $d \geq 2$, we can reduce the problem to that of estimating a uniform density on a closed, convex set (since such densities belong to $\mathcal{F}_d$). The lower bound constructions in the convex set estimation proofs of Korostelëv and Tsybakov (1993), Mammen and Tsybakov (1995), Brunel (2013, 2016) can therefore be applied to yield the rate $n^{-2/(d+1)}$.

As can be seen from the above descriptions, the same lower bounds hold for the (smaller) class of upper semi-continuous densities on $\mathbb{R}^d$ that are concave on their support. Moreover, a minimax lower bound can also be obtained for the L_2^2 loss function. Note that in this case, the loss function is not affine invariant, so it makes sense to restrict attention to log-concave densities f with a lower bound on the determinant of the corresponding covariance matrix Σ_f. The result obtained is that there exist $c'_d > 0$ such that for every $\kappa > 0$,

$$\inf_{\tilde{f}_n \in \tilde{\mathcal{F}}_n} \sup_{f_0 \in \mathcal{F}_d : \det(\Sigma_{f_0}) \geq \kappa^2} \mathbb{E}_{f_0} L_2^2(\tilde{f}_n, f_0) \geq \begin{cases} c'_1 n^{-4/5}/\kappa, & \text{if } d = 1, \\ c'_d n^{-2/(d+1)}/\kappa, & \text{if } d \geq 2. \end{cases}$$

3. Integrable envelopes for classes of log-concave densities. In this section, we recall recent results on envelopes for certain classes of log-concave densities developed in the probability literature. The following result, part (a) of which is due to Fresen (2013), Lemma 13 and part (b) of which is due to Lovász and Vempala [(2007), Theorem 5.14(a)], is used in the proof of Lemma 6 in Section A.2. In particular, part (a) gives us uniform control of tail probabilities and moments of log-concave densities with zero mean and identity covariance matrix; part (b) facilitates a lower bound for the smallest eigenvalue of the covariance matrix corresponding to the log-concave projection of a distribution whose own covariance matrix is close to the identity. For $f \in \mathcal{F}_d$, let $\mu_f := \int_{\mathbb{R}^d} x f(x)\, dx$ and

$\Sigma_f := \int_{\mathbb{R}^d}(x-\mu_f)(x-\mu_f)^T f(x)\,dx$. For $\mu \in \mathbb{R}^d$ and a symmetric, positive-definite, $d \times d$ matrix Σ, let

$$\mathcal{F}_d^{\mu,\Sigma} := \{f \in \mathcal{F}_d : \mu_f = \mu, \Sigma_f = \Sigma\}.$$

THEOREM 2. (a) *For each $d \in \mathbb{N}$, there exist $A_{0,d}, B_{0,d} > 0$ such that for all $x \in \mathbb{R}^d$, we have*

$$\sup_{f \in \mathcal{F}_d^{0,I}} f(x) \le e^{-A_{0,d}\|x\|+B_{0,d}}.$$

(b) *We have*

$$\inf_{f \in \mathcal{F}_d^{0,I}} \inf_{x:\|x\|\le 1/9} f(x) > 0.$$

In fact, it will be convenient to have the corresponding envelopes for slightly larger classes in order to establish our bracketing entropy bounds in Section 4. We write $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$ for the smallest and largest eigenvalues respectively of a positive-definite, symmetric $d \times d$ matrix Σ. For $\xi \ge 0$ and $\eta \in (0,1)$, let

$$\tilde{\mathcal{F}}_d^{\xi,\eta} := \{\tilde{f} \in \mathcal{F}_d : \|\mu_{\tilde{f}}\| \le \xi \text{ and } 1-\eta \le \lambda_{\min}(\Sigma_{\tilde{f}}) \le \lambda_{\max}(\Sigma_{\tilde{f}}) \le 1+\eta\}.$$

COROLLARY 3. (a) *For each $d \in \mathbb{N}$, there exist $A_{0,d}, B_{0,d} > 0$ such that for every $\xi \ge 0$, every $\eta \in (0,1)$ and every $x \in \mathbb{R}^d$, we have*

$$\sup_{\tilde{f} \in \tilde{\mathcal{F}}_d^{\xi,\eta}} \tilde{f}(x) \le (1-\eta)^{-d/2} \exp\biggl\{-\frac{A_{0,d}\|x\|}{(1+\eta)^{1/2}} + \frac{A_{0,d}\xi}{(1+\eta)^{1/2}} + B_{0,d}\biggr\}.$$

(b) *For every $\xi \ge 0$ and $\eta \in (0,1)$ satisfying $\xi \le (1-\eta)^{1/2}/9$, we have*

$$\inf_{\tilde{f} \in \tilde{\mathcal{F}}_d^{\xi,\eta}} \inf_{x:\|x\|\le \frac{1}{9}(1-\eta)^{1/2}-\xi} \tilde{f}(x) > 0.$$

4. Bracketing entropy bounds and global rates of convergence of the log-concave maximum likelihood estimator. Let $\mathcal{G}$ be a class of functions on $\mathbb{R}^d$, and let ρ be a semi-metric on $\mathcal{G}$. For $\varepsilon > 0$, let $N_{[\cdot]}(\varepsilon, \mathcal{G}, \rho)$ denote the ε-bracketing number of $\mathcal{G}$ with respect to ρ. Thus, $N_{[\cdot]}(\varepsilon, \mathcal{G}, \rho)$ is the minimal $N \in \mathbb{N}$ such that there exist pairs $\{(g_j^L, g_j^U)\}_{j=1}^N$ with the properties that $\rho(g_j^L, g_j^U) \le \varepsilon$ for all $j = 1, \dots, N$ and, for each $g \in \mathcal{G}$, there exists $j^* \in \{1, \dots, N\}$ satisfying $g_{j^*}^L \le g \le g_{j^*}^U$. We call $\log N_{[\cdot]}(\varepsilon, \mathcal{G}, \rho)$ the ε-bracketing entropy of $\mathcal{G}$. The following entropy bound is key to establishing the rate of convergence of the log-concave maximum likelihood estimator in Hellinger distance.

THEOREM 4. *Let $\eta_d > 0$ be taken from Lemma 6 in Section A.2.*

(i) *There exist $\overline{K}_1, \overline{K}_2, \overline{K}_3 \in (0, \infty)$ such that*

$$\log N_{[\cdot]}\big(\varepsilon, \tilde{\mathcal{F}}_d^{1,\eta_d}, h\big) \leq \begin{cases} \overline{K}_1 \varepsilon^{-1/2}, & \text{when } d = 1, \\ \overline{K}_2 \varepsilon^{-1} \log_{++}^{3/2}(1/\varepsilon), & \text{when } d = 2, \\ \overline{K}_3 \varepsilon^{-2}, & \text{when } d = 3, \end{cases}$$

for all $\varepsilon > 0$, where $\log_{++}(x) := \max(1, \log x)$.

(ii) *For every $d \in \mathbb{N}$, there exist $\varepsilon_d \in (0, 1]$ and $\underline{K}_d \in (0, \infty)$ such that*

$$\log N_{[\cdot]}\big(\varepsilon, \tilde{\mathcal{F}}_d^{1,\eta_d}, h\big) \geq \underline{K}_d \max\big\{\varepsilon^{-d/2}, \varepsilon^{-(d-1)}\big\}$$

for all $\varepsilon \in (0, \varepsilon_d]$.

Note that in this theorem, η_d depends only on d. The proof of the upper bound in Theorem 4 is long, so we give a broad outline here. We first consider the problem of finding a set of Hellinger brackets for the class of restrictions of densities $\tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}$ to $[0, 1]^d$. The main challenge here is that the effective domain of $\tilde{f}$ is unknown, and indeed the shape of this domain affects the bracketing entropy significantly [Gao and Wellner (2015), Guntuboyina and Sen (2013)]. In Proposition 4 in the online supplement, we derive new bracketing entropy bounds for bounded concave functions defined on a general convex domain when $d = 2, 3$. This is achieved by constructing inner layers of convex polyhedral approximations where the number of simplices required to triangulate the region between successive layers can be controlled using results from discrete convex geometry. It is the absence of corresponding convex geometry results for $d \geq 4$ that means we are currently unable to provide bracketing entropy bounds in these higher dimensions.

Since the logarithms of densities in $\tilde{\mathcal{F}}_d^{1,\eta_d}$ can take the value $-\infty$, we combine an inductive argument with Proposition 4 in the online supplement to derive bracketing bounds for the restrictions of $\tilde{\mathcal{F}}_d^{1,\eta_d}$ to $[0, 1]^d$. Translations of these brackets can be used to cover the restrictions of densities $\tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}$ to other unit boxes. We use our integrable envelope function for the class $\tilde{\mathcal{F}}_d^{1,\eta_d}$ from Corollary 3 to allow us to use fewer brackets as the boxes move further from the origin, yet still cover with higher accuracy, enabling us to obtain the desired conclusion.

We are now in a position to state our main result on the supremum risk of the log-concave maximum likelihood estimator for the squared Hellinger loss function.

THEOREM 5. *Let $X_1, \dots, X_n$ be independent and identically distributed random vectors with density $f_0 \in \mathcal{F}_d$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Then*

$$R(\hat{f}_n, \mathcal{F}_d) = \begin{cases} O\big(n^{-4/5}\big), & \text{if } d = 1, \\ O\big(n^{-2/3} \log n\big), & \text{if } d = 2, \\ O\big(n^{-1/2} \log n\big), & \text{if } d = 3. \end{cases}$$

The proof of this theorem first involves standardising the data and using affine equivariance to reduce the problem to that of bounding the supremum risk over the class of log-concave densities with mean vector 0 and identity covariance matrix. Writing $\hat{g}_n$ for the log-concave maximum likelihood estimator for the standardised data, we show in Lemma 6 in Section A.2 that

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl(\hat{g}_n \notin \tilde{\mathcal{F}}_d^{1,\eta_d}\bigr) = O\bigl(n^{-1}\bigr).$$

As well as using various known results on the relationship between the mean vector and covariance matrix of the log-concave maximum likelihood estimator in relation to its sample counterparts, the main step here is to show that, provided none of the sample covariance matrix eigenvalues are too large, the only way an eigenvalue of the covariance matrix corresponding to the maximum likelihood estimator can be small is if an eigenvalue of the sample covariance matrix is small.

The other part of the proof of Theorem 5 is to control

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{E}\bigl\{h^2(\hat{g}_n, g_0)\mathbb{1}_{\{\hat{g}_n \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}}\bigr\}.$$

This can be done by appealing to empirical process theory for maximum likelihood estimators, and using the Hellinger bracketing entropy bounds developed in Theorem 4.

APPENDIX

A.1. Proofs from Section 3.

PROOF OF COROLLARY 3. (a) Let $\tilde{f} \in \tilde{\mathcal{F}}_d^{\xi,\eta}$. Then we can let $f(x) := |\det \Sigma_{\tilde{f}}|^{1/2} \tilde{f}(\Sigma_{\tilde{f}}^{1/2} x + \mu_{\tilde{f}})$, so that $f \in \mathcal{F}_d^{0,I}$. Thus, by Theorem 2(a), there exist $A_{0,d}, B_{0,d} > 0$ such that

$$f(x) \leq e^{-A_{0,d}\|x\| + B_{0,d}}$$

for all $x \in \mathbb{R}^d$. We deduce that, for all $x \in \mathbb{R}^d$,

$$\begin{aligned}
\tilde{f}(x) &= |\det \Sigma_{\tilde{f}}|^{-1/2} f\bigl(\Sigma_{\tilde{f}}^{-1/2}(x - \mu_{\tilde{f}})\bigr) \\
&\leq (1-\eta)^{-d/2} \exp\biggl\{-\frac{A_{0,d}|\|x\| - \|\mu_{\tilde{f}}\||}{(1+\eta)^{1/2}} + B_{0,d}\biggr\} \\
&\leq (1-\eta)^{-d/2} \exp\biggl\{-\frac{A_{0,d}\|x\|}{(1+\eta)^{1/2}} + \frac{A_{0,d}\xi}{(1+\eta)^{1/2}} + B_{0,d}\biggr\}.
\end{aligned}$$

(b) If $\tilde{f} \in \tilde{\mathcal{F}}_d^{\xi,\eta}$, then as above, we can let $f(x) := |\det \Sigma_{\tilde{f}}|^{1/2} \tilde{f}(\Sigma_{\tilde{f}}^{1/2} x + \mu_{\tilde{f}})$, so that $f \in \mathcal{F}_d^{0,I}$. Moreover, if $\xi \leq (1-\eta)^{1/2}/9$ and $\|x_0\| \leq (1-\eta)^{1/2}/9 - \xi$, then

$$\bigl\|\Sigma_{\tilde{f}}^{-1/2}(x_0 - \mu_{\tilde{f}})\bigr\|^2 \leq \frac{(\|x_0\| + \xi)^2}{1-\eta} \leq \frac{1}{81}.$$

It follows that

$$\tilde{f}(x_0) = |\det \Sigma_{\tilde{f}}|^{-1/2} f\bigl(\Sigma_{\tilde{f}}^{-1/2}(x_0 - \mu_{\tilde{f}})\bigr) \geq (1+\eta)^{-d/2} \inf_{f \in \mathcal{F}_d^{0,I}} \inf_{x:\|x\| \leq 1/9} f(x),$$

so the result follows by Theorem 2(b). □

A.2. Proofs from Section 4.

PROOF OF THEOREM 4. (i) *Step* 1: *Preliminaries*. Let $\varepsilon_{00} \in (0, e^{-1}]$. Fix $\varepsilon \in (0, \varepsilon_{00}]$ and set $y_k := 2^{k/2}$ for $k = 0, 1, \dots, k_0$, where $k_0 := \min\{k \in \mathbb{N} : y_k \geq \log(\varepsilon_{00}/\varepsilon)\}$. Let Φ denote the class of upper semi-continuous, concave functions $\phi : [0,1]^d \to [-\infty, -y_0]$, and let $\mathcal{D}$ denote the class of closed, convex subsets D of $[0,1]^d$. For $D \in \mathcal{D}$, let $\Phi_0(D) = \varnothing$ and for $k = 1, \dots, k_0$, define

$$\Phi_k(D) := \bigl\{\phi \in \Phi : \mathrm{dom}(\phi) = D \text{ and } \phi(x) \geq -y_k \text{ for all } x \in D\bigr\}.$$

Now let $\mathcal{F}_k(\mathcal{D}) := \{e^{\phi} : \phi \in \bigcup_{D \in \mathcal{D}} \Phi_k(D)\}$, where we adopt the convention that $e^{-\infty} = 0$. Write

$$K_{1,k}^* := \Biggl(1 + 5\sum_{j=1}^{k} e^{-y_{j-1}}\Biggr)^{1/2}$$

and

$$K_{2,k,1}^* := \sum_{j=1}^{k} \bigl\{e^{-y_{j-1}/2} K_1 + 8e^{-y_{j-1}/4} + K_1^{\circ} y_j^{1/2} e^{-y_{j-1}/4}\bigr\},$$

$$K_{2,k,2}^* := \sum_{j=1}^{k} \bigl\{K_2 e^{-y_{j-1}/2} + K_2^{\circ} y_j e^{-y_{j-1}/2}\bigr\},$$

$$K_{2,k,3}^* := \sum_{j=1}^{k} \bigl\{K_3 e^{-y_{j-1}} + K_3^{\circ} y_j^2 e^{-y_{j-1}}\bigr\},$$

where K_d and K_d° are the constants defined in the proofs of Propositions 2 and 4 in the online supplement, respectively. Let

$$h_d(\varepsilon) := \begin{cases} \varepsilon^{-1/2}, & \text{when } d = 1, \\ \varepsilon^{-1} \log_{++}^{3/2}(1/\varepsilon), & \text{when } d = 2, \\ \varepsilon^{-2}, & \text{when } d = 3. \end{cases}$$

Step 2. Recall that $h(f,g) = L_2(f^{1/2}, g^{1/2})$ for any $f, g \in L_1(\mathbb{R}^d)$. It will therefore suffice to derive an L_2-bracketing entropy bound for the set $\{\tilde{f}^{1/2} : \tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}$. As a first step towards this goal, we claim that for $k = 1, \dots, k_0$ and $d = 1, 2, 3$, we have

$$\log N_{[\cdot]}\bigl(K_{1,k}^* \varepsilon, \mathcal{F}_k(\mathcal{D}), L_2\bigr) \leq K_{2,k,d}^* h_d(\varepsilon), \tag{4}$$

and prove this by induction. First, consider the case $k = 1$. Let $N_{S,1,1} := \lfloor e^{K_1 - y_0}\varepsilon^{-2}\rfloor$ and $N_{S,1,d} := \lfloor \exp(K_d e^{-(d-1)y_0/2}\varepsilon^{-(d-1)})\rfloor$ for $d = 2, 3$. By Proposition 2 in the online supplement, we can find pairs of measurable subsets $\{(A^L_{j,1}, A^U_{j,1}) : j = 1, \dots, N_{S,1,d}\}$ of $[0, 1]^d$ with the properties that $L_1(\mathbb{1}_{A^U_{j,1}}, \mathbb{1}_{A^L_{j,1}}) \le \varepsilon^2 e^{y_0}$ for $j = 1, \dots, N_{S,1,d}$ and, if A is a closed, convex subset of $[0, 1]^d$, then there exists $j^* \in \{1, \dots, N_{S,1,d}\}$ such that $A^L_{j^*,1} \subseteq A \subseteq A^U_{j^*,1}$. Note that by replacing $A^L_{j,1}$ with the closure of its convex hull if necessary, there is no loss of generality in assuming that each $A^L_{j,1}$ is closed and convex. Moreover, by Proposition 4 in the online supplement, for each $j = 1, \dots, N_{S,1,d}$ for which $A^L_{j,1}$ is d-dimensional, there exists a bracketing set $\{[\psi^L_{j,\ell,1}, \psi^U_{j,\ell,1}] : \ell = 1, \dots, N_{B,1,d}\}$ for $\Phi_1(A^L_{j,1})$, where $N_{B,1,d} := \lfloor \exp\{K^\circ_d h_d(\varepsilon e^{y_0/2}/y_1)\}\rfloor$, such that $-y_1 \le \psi^L_{j,\ell,1} \le \psi^U_{j,\ell,1} \le -y_0$, that $L_2(\psi^U_{j,\ell,1}, \psi^L_{j,\ell,1}) \le 2\varepsilon e^{y_0/2}$ and such that for every $\phi \in \Phi_1(A^L_{j,1})$, we can find $\ell^* \in \{1, \dots, N_{B,1,d}\}$ with $\psi^L_{j,\ell^*,1} \le \phi \le \psi^U_{j,\ell^*,1}$. If $\dim(A^L_{j,1}) < d$, we define a trivial bracketing set $\{[\psi^L_{j,\ell,1}, \psi^U_{j,\ell,1}] : \ell = 1, \dots, N_{B,1,d}\}$ for $\Phi_1(A^L_{j,1})$ by $\psi^L_{j,\ell,1}(x) := -y_1$ and $\psi^U_{j,\ell,1}(x) := -y_0$ for $x \in A^L_{j,1}$. Note that whenever $\dim(A^L_{j,1}) < d$, we have $L_2(\psi^U_{j,\ell,1}, \psi^L_{j,\ell,1}) = 0$. This enables us to define a bracketing set $\{[f^L_{j,\ell,1}, f^U_{j,\ell,1}] : j = 1, \dots, N_{S,1,d}, \ell = 1, \dots, N_{B,1,d}\}$ for $\mathcal{F}_1(\mathcal{D})$ by

$$
\begin{aligned}
f^L_{j,\ell,1}(x) &:= e^{\psi^L_{j,\ell,1}(x)}\mathbb{1}_{\{x\in A^L_{j,1}\}},\\
f^U_{j,\ell,1}(x) &:= e^{\psi^U_{j,\ell,1}(x)}\mathbb{1}_{\{x\in A^L_{j,1}\}} + e^{-y_0}\mathbb{1}_{\{x\in A^U_{j,1}\setminus A^L_{j,1}\}}
\end{aligned}
$$

for $x \in [0, 1]^d$. Note that

$$
\begin{aligned}
L_2^2(f^U_{j,\ell,1}, f^L_{j,\ell,1}) &= \int_{A^L_{j,1}} (e^{\psi^U_{j,\ell,1}} - e^{\psi^L_{j,\ell,1}})^2\, d\mu_d + e^{-2y_0}\mu_d(A^U_{j,1} \setminus A^L_{j,1})\\
&\le e^{-2y_0} L_2^2(\psi^U_{j,\ell,1}, \psi^L_{j,\ell,1}) + e^{-2y_0} L_1(\mathbb{1}_{A^U_{j,1}}, \mathbb{1}_{A^L_{j,1}})\\
&\le (K^*_{1,1})^2\varepsilon^2.
\end{aligned}
$$

Moreover, when $d = 1$ the cardinality of this bracketing set is

$$
\begin{aligned}
N_{S,1,1}N_{B,1,1} &\le e^{K_1 - y_0}\varepsilon^{-2}\exp\left\{K^\circ_1 h_1\left(\frac{\varepsilon e^{y_0/2}}{y_1}\right)\right\}\\
&\le \exp\left\{e^{-y_0/2}K_1\varepsilon^{-1/2} + 8e^{-y_0/4}\varepsilon^{-1/2} + K^\circ_1 h_1\left(\frac{\varepsilon e^{y_0/2}}{y_1}\right)\right\}\\
&\le e^{K^*_{2,1,1}\varepsilon^{-1/2}},
\end{aligned}
$$

where we have used the facts that $e^{y_0/2}\varepsilon^{1/2} \le e^{y_{k_0-1}/2}\varepsilon^{1/2} \le \varepsilon_{00}^{1/2} \le 1$ and $2e^{y_0/4}\varepsilon^{1/2}\log(1/\varepsilon) \le 8e^{y_{k_0-1}/4}\varepsilon^{1/4} \le 8\varepsilon_{00}^{1/4} \le 8$. When $d = 2$,

$$N_{S,1,2}N_{B,1,2} \le \exp\biggl\{K_2 e^{-y_0/2}\varepsilon^{-1} + K_2^\circ h_2\biggl(\frac{\varepsilon e^{y_0/2}}{y_1}\biggr)\biggr\}$$
$$\le e^{K_{2,1,2}^*\varepsilon^{-1}\log_{++}^{3/2}(1/\varepsilon)}.$$

Finally, when $d = 3$, the cardinality of the bracketing set is

$$N_{S,1,3}N_{B,1,3} \le \exp\biggl\{K_3 e^{-y_0}\varepsilon^{-2} + K_3^\circ h_3\biggl(\frac{\varepsilon e^{y_0/2}}{y_1}\biggr)\biggr\} \le e^{K_{2,1,3}^*\varepsilon^{-2}}.$$

This proves the claim (4) when $k = 1$. Now suppose the claim is true for some $k-1 < k_0 - 1$, so there exist brackets $\{[f_{j',k-1}^L, f_{j',k-1}^U] : j' = 1, \dots, N'_{k-1,d}\}$ for $\mathcal{F}_{k-1}(\mathcal{D})$, where $N'_{k-1,d} := \lfloor\exp\{K_{2,k-1,d}^* h_d(\varepsilon)\}\rfloor$, such that $L_2(f_{j',k-1}^U, f_{j',k-1}^L) \le K_{1,k-1}^*\varepsilon$, and for every $f \in \mathcal{F}_{k-1}(\mathcal{D})$, there exists $(j')^* \in \{1, \dots, N'_{k-1,d}\}$ such that $f_{(j')^*,k-1}^L \le f \le f_{(j')^*,k-1}^U$. Let $B_{j',k-1}^U := \{x \in [0,1]^d : f_{j',k-1}^U(x) > 0\}$. We also define $N_{S,k,1} := \lfloor e^{K_1 - y_{k-1}}\varepsilon^{-2}\rfloor$ and $N_{S,k,d} := \lfloor\exp(K_d e^{-y_{k-1}(d-1)/2}\varepsilon^{-(d-1)})\rfloor$ for $d = 2, 3$. Using Proposition 2 in the online supplement again, we can find pairs of measurable subsets $\{(A_{j,k}^L, A_{j,k}^U) : j = 1, \dots, N_{S,k,d}\}$ of $[0,1]^d$, where $A_{j,k}^L$ is closed and convex, with the properties that $L_1(\mathbb{1}_{A_{j,k}^U}, \mathbb{1}_{A_{j,k}^L}) \le \varepsilon^2 e^{y_{k-1}}$ for $j = 1, \dots, N_{S,k,d}$ and, if A is a closed, convex subset of $[0,1]^d$, then there exists $j^* \in \{1, \dots, N_{S,k,d}\}$ such that $A_{j^*,k}^L \subseteq A \subseteq A_{j^*,k}^U$. Using Proposition 4 in the online supplement again, for each $j = 1, \dots, N_{S,k,d}$ for which $\dim(A_{j,k}^L) = d$, there exists a bracketing set $\{[\psi_{j,\ell,k}^L, \psi_{j,\ell,k}^U] : \ell = 1, \dots, N_{B,k,d}\}$ for $\Phi_k(A_{j,k}^L)$, where $N_{B,k,d} := \lfloor\exp\{K_d^\circ h_d(\frac{\varepsilon e^{y_{k-1}/2}}{y_k})\}\rfloor$, such that $-y_k \le \psi_{j,\ell,k}^L \le \psi_{j,\ell,k}^U \le -y_0$, that $L_2(\psi_{j,\ell,k}^U, \psi_{j,\ell,k}^L) \le 2\varepsilon e^{y_{k-1}/2}$ and that for every $\phi \in \Phi_k(A_{j,k}^L)$, we can find $\ell^* \in \{1, \dots, N_{B,k,d}\}$ with $\psi_{j,\ell^*,k}^L \le \phi \le \psi_{j,\ell^*,k}^U$. Similar to the $k = 1$ case, whenever $\dim(A_{j,k}^L) < d$, we define $\psi_{j,\ell,k}^L(x) := -y_k$ and $\psi_{j,\ell,k}^U(x) := -y_0$ for $x \in A_{j,k}^L$. We can now define a bracketing set $\{[f_{j,\ell,j',k}^L, f_{j,\ell,j',k}^U] : j = 1, \dots, N_{S,k,d}, \ell = 1, \dots, N_{B,k,d}, j' = 1, \dots, N'_{k-1,d}\}$ for $\mathcal{F}_k(\mathcal{D})$ by

$$f_{j,\ell,j',k}^L(x) := e^{\min\{-y_{k-1}, \psi_{j,\ell,k}^L(x)\}}\mathbb{1}_{\{x \in A_{j,k}^L \setminus B_{j',k-1}^U\}} + f_{j',k-1}^L(x)\mathbb{1}_{\{x \in B_{j',k-1}^U\}},$$

$$f_{j,\ell,j',k}^U(x) := e^{\min\{-y_{k-1}, \psi_{j,\ell,k}^U(x)\}}\mathbb{1}_{\{x \in A_{j,k}^L \setminus B_{j',k-1}^U\}} + f_{j',k-1}^U(x)\mathbb{1}_{\{x \in B_{j',k-1}^U\}}$$
$$+ e^{-y_{k-1}}\mathbb{1}_{\{x \in A_{j,k}^U \setminus (B_{j',k-1}^U \cup A_{j,k}^L)\}}$$

for $x \in [0,1]^d$. Again, we can compute

$$L_2^2(f^U_{j,\ell,j',k}, f^L_{j,\ell,j',k}) \le e^{-2y_{k-1}} L_2^2(\psi^U_{j,\ell,k}, \psi^L_{j,\ell,k}) + \varepsilon^2\left(1 + 5\sum_{j=1}^{k-1} e^{-y_{j-1}}\right) \\ + e^{-2y_{k-1}} L_1(\mathbb{1}_{A^U_{j,k}}, \mathbb{1}_{A^L_{j,k}}) \le (K^*_{1,k})^2 \varepsilon^2.$$

When $d = 1$, the cardinality of this bracketing set is

$$N'_{k-1,1} N_{S,k,1} N_{B,k,1} \le e^{K^*_{2,k-1,1} h_1(\varepsilon)} e^{K_1 - y_{k-1}} \varepsilon^{-2} e^{K_1^\circ h_1(\frac{\varepsilon e^{y_{k-1}/2}}{y_k})} \le e^{K^*_{2,k,1}\varepsilon^{-1/2}},$$

as required. When $d = 2$, the cardinality is

$$\begin{aligned} &N'_{k-1,2} N_{S,k,2} N_{B,k,2} \\ &\quad \le \exp\left\{K^*_{2,k-1,2} h_2(\varepsilon) + K_2 e^{-y_{k-1}/2} \varepsilon^{-1} + K_2^\circ h_2\left(\frac{\varepsilon e^{y_{k-1}/2}}{y_k}\right)\right\} \\ &\quad \le e^{K^*_{2,k,2}\varepsilon^{-1}\log^{3/2}_{++}(1/\varepsilon)}. \end{aligned}$$

Finally, when $d = 3$, the cardinality of the bracketing set is

$$\begin{aligned} &N'_{k-1,3} N_{S,k,3} N_{B,k,3} \\ &\quad \le \exp\left\{K^*_{2,k-1,3} h_3(\varepsilon) + K_3 e^{-y_{k-1}} \varepsilon^{-2} + K_3^\circ h_3\left(\frac{\varepsilon e^{y_{k-1}/2}}{y_k}\right)\right\} \\ &\quad \le e^{K^*_{2,k,3}\varepsilon^{-2}}. \end{aligned}$$

This establishes the claim (4) by induction.

Step 3. For $\tilde{b} > 0$, write $\mathcal{G}_{d,[0,1]^d,\tilde{b}}$ for the set of functions on $[0,1]^d$ of the form $f^{1/2}$, where f is an upper semi-continuous, log-concave function whose domain is a closed, convex subset of $[0,1]^d$, and for which $f^{1/2} \le \tilde{b}$. Our next goal is to derive an L_2-bracketing entropy bound for $\mathcal{G}_{d,[0,1]^d,e^{-1}}$. Writing $\bar{\mathcal{F}}_{k_0}(\mathcal{D}) := \{e^\phi : \phi \in \Phi \setminus \bigcup_{D\in\mathcal{D}} \Phi_{k_0}(D)\}$, we note that since square roots of log-concave functions are log-concave,

$$\mathcal{G}_{d,[0,1]^d,e^{-1}} \subseteq \{e^\phi : \phi \in \Phi\} = \mathcal{F}_{k_0}(\mathcal{D}) \cup \bar{\mathcal{F}}_{k_0}(\mathcal{D}).$$

We derived brackets $[f^L_{j,\ell,j'}, f^U_{j,\ell,j'}]$ for $\mathcal{F}_{k_0}(\mathcal{D})$ in Step 2 above, and moreover, a bracketing set for $\bar{\mathcal{F}}_{k_0}(\mathcal{D})$ is given by $\{[\bar{f}^L_{j,\ell,j'}, \bar{f}^U_{j,\ell,j'}] : j = 1, \ldots, N_{S,k_0,d}, \ell = 1, \ldots, N_{B,k_0,d}, j' = 1, \ldots, N'_{k_0-1,d}\}$, where

$$\begin{aligned} &\bar{f}^L_{j,\ell,j'}(x) := f^L_{j,\ell,j',k_0}(x), \\ &\bar{f}^U_{j,\ell,j'}(x) := f^U_{j,\ell,j',k_0}(x)\mathbb{1}_{\{\log f^U_{j,\ell,j',k_0}(x) \ge -y_{k_0}\}} + e^{-y_{k_0}}\mathbb{1}_{\{\log f^U_{j,\ell,j',k_0}(x) < -y_{k_0}\}} \end{aligned}$$

for $x \in [0,1]^d$. Observe that

$$L_2^2(\bar{f}^U_{j,\ell,j'}, \bar{f}^L_{j,\ell,j'}) \le (K^*_{1,k_0})^2\varepsilon^2 + e^{-2y_{k_0}} \le \left(K^*_{1,k_0} + \frac{1}{\varepsilon_{00}}\right)^2 \varepsilon^2.$$

Since k_0 depends on ε, it is important to observe that for all $k = 1, \dots, k_0$,

$$\begin{aligned} K^*_{1,k} &\le 4, \\ K^*_{2,k,1} &\le 2K_1 + 32 + 8K_1^\circ =: \bar{K}^*_{2,1} - \log 2, \\ K^*_{2,k,2} &\le 2K_2 + K_2^\circ(8e^{1/2}+1) =: \bar{K}^*_{2,2} - \log 2, \\ K^*_{2,k,3} &\le K_3 + K_3^\circ(8e+1) =: \bar{K}^*_{2,3} - \log 2. \end{aligned}$$

In particular, these bounds do not depend on ε, and since $\varepsilon \in (0, \varepsilon_{00}]$ was arbitrary, we conclude that

$$\begin{aligned} \log N_{[\cdot]}\big((4+\varepsilon_{00}^{-1})\varepsilon, \mathcal{G}_{d,[0,1]^d,e^{-1}}, L_2\big) &\le \log N_{[\cdot]}\big((4+\varepsilon_{00}^{-1})\varepsilon, \{e^\phi : \phi \in \Phi\}, L_2\big) \\ &\le \bar{K}^*_{2,d} h_d(\varepsilon) \end{aligned}$$

for all $\varepsilon \in (0, \varepsilon_{00}]$ and $d = 1, 2, 3$. By a simple scaling argument, we deduce that for any $b > 0$,

$$\log N_{[\cdot]}\big((4+\varepsilon_{00}^{-1})\varepsilon b^{1/2}, \mathcal{G}_{d,[0,1]^d,be^{-1}}, L_2\big) \le \bar{K}^*_{2,d} h_d\big(\varepsilon/b^{1/2}\big)$$

for all $\varepsilon \in (0, b^{1/2}\varepsilon_{00}]$.

Step 4. We now show how to translate and scale brackets appropriately for other cubes, and combine the results to obtain the final bracketing entropy bound for $\tilde{\mathcal{F}}_d^{1,\eta_d}$. Let $A_{0,d}, B_{0,d} > 0$ be as in Corollary 3(a). Define

$$T_d := \frac{A_{0,d}(d^{1/2}+1)}{(1+\eta_d)^{1/2}} + B_{0,d} + \frac{d}{2}\log\left(\frac{1}{1-\eta_d}\right) + d + 1,$$

set $\varepsilon_{01,d} := \min\{e^{-T_d}, \frac{1}{d^d}\varepsilon_{00}^4\}$ and fix $\varepsilon \in (0, \varepsilon_{01,d}]$. For $\mathbf{j} = (j_1, \dots, j_d) \in \mathbb{Z}^d$, let

$$C_{\mathbf{j}}^2 := \exp\left(-\frac{A_{0,d}\|\mathbf{j}\|}{(1+\eta_d)^{1/2}} + T_d\right),$$

where $\|\mathbf{j}\|^2 := \sum_{k=1}^d j_k^2$. Note from Corollary 3(a) that

$$\sup_{\tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}} \sup_{x \in [j_1, j_1+1] \times \cdots \times [j_d, j_d+1]} \tilde{f}(x)^{1/2} \le C_{\mathbf{j}} e^{-1}.$$

Let $j_0 := \max\{\|\mathbf{j}\| : \mathbf{j} \in \mathbb{Z}^d, C_{\mathbf{j}} \ge \varepsilon\{\log(1/\varepsilon)\}^{-(d-1)/2}\}$, so we may assume $j_0 \ge 1$. For $\mathbf{j} = (j_1, \dots, j_d) \in \mathbb{Z}^d$ such that $\|\mathbf{j}\| \le j_0$, let $N_{\mathbf{j}} := N_{[\cdot]}((4 +$

$\varepsilon_{00}^{-1})\varepsilon C_{\mathbf{j}}^{1/2}, \mathcal{G}_{d,[0,1]^d,C_{\mathbf{j}}e^{-1}}, L_2)$, and let $\{[f_{\mathbf{j},\ell}^L, f_{\mathbf{j},\ell}^U], \ell = 1, \ldots, N_{\mathbf{j}}\}$, denote a bracketing set for $\mathcal{G}_{d,[0,1]^d,C_{\mathbf{j}}e^{-1}}$ with $L_2(f_{\mathbf{j},\ell}^U, f_{\mathbf{j},\ell}^L) \leq (4+\varepsilon_{00}^{-1})\varepsilon C_{\mathbf{j}}^{1/2}$. Such a bracketing set can be found because when $\|\mathbf{j}\| \leq j_0$, we have

$$\varepsilon \leq C_{\mathbf{j}}^{1/2}\varepsilon^{1/2}\{\log(1/\varepsilon)\}^{d/4} \leq C_{\mathbf{j}}^{1/2}\varepsilon^{1/2}(d\varepsilon^{-(1/d)})^{d/4} \leq C_{\mathbf{j}}^{1/2}\varepsilon_{00}.$$

Finally, for $\{\boldsymbol{\ell} = (\ell_{\mathbf{j}}) \in \times_{\mathbf{j}:\|\mathbf{j}\|\leq j_0}\{1, \ldots, N_{\mathbf{j}}\}\}$, we define a bracketing set for $\{\tilde{f}^{1/2} : \tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}$ by

$$\begin{aligned} f_{\boldsymbol{\ell}}^L(x) &:= \sum_{\mathbf{j}:\|\mathbf{j}\|\leq j_0} f_{\mathbf{j},\ell_{\mathbf{j}}}^L(x-\mathbf{j})\mathbb{1}_{\{x\in[j_1,j_1+1)\times\cdots\times[j_d,j_d+1)\}}, \\ f_{\boldsymbol{\ell}}^U(x) &:= \sum_{\mathbf{j}:\|\mathbf{j}\|\leq j_0} f_{\mathbf{j},\ell_{\mathbf{j}}}^U(x-\mathbf{j})\mathbb{1}_{\{x\in[j_1,j_1+1)\times\cdots\times[j_d,j_d+1)\}} \\ &\quad + e^{-1}\sum_{\mathbf{j}:\|\mathbf{j}\|>j_0} C_{\mathbf{j}}\mathbb{1}_{\{x\in[j_1,j_1+1)\times\cdots\times[j_d,j_d+1)\}} \end{aligned}$$

for $x \in \mathbb{R}^d$. Note that

$$\begin{aligned} L_2(f_{\boldsymbol{\ell}}^U, f_{\boldsymbol{\ell}}^L) &\leq (4+\varepsilon_{00}^{-1})\varepsilon\biggl(\sum_{\mathbf{j}\in\mathbb{Z}^d} C_{\mathbf{j}}\biggr)^{1/2} + \biggl(\sum_{\mathbf{j}:\|\mathbf{j}\|>j_0} C_{\mathbf{j}}^2\biggr)^{1/2} e^{-1} \\ &\leq (4+\varepsilon_{00}^{-1})\varepsilon\frac{e^{\frac{A_{0,d}d^{1/2}}{4(1+\eta_d)^{1/2}}+\frac{T_d}{4}}d^{1/2}\pi^{d/4}}{\Gamma(1+d/2)^{1/2}}\biggl\{\int_0^\infty r^{d-1}e^{-\frac{rA_{0,d}}{2(1+\eta_d)^{1/2}}}\,dr\biggr\}^{1/2} \\ &\quad + \frac{e^{\frac{A_{0,d}d^{1/2}}{2(1+\eta_d)^{1/2}}+\frac{T_d}{2}-1}d^{1/2}\pi^{d/4}}{\Gamma(1+d/2)^{1/2}}\biggl\{\int_{j_0}^\infty r^{d-1}e^{-\frac{rA_{0,d}}{(1+\eta_d)^{1/2}}}\,dr\biggr\}^{1/2} \\ &\leq \varepsilon(B_1+B_2), \end{aligned}$$

where

$$\begin{aligned} B_1 &:= (4+\varepsilon_{00}^{-1})\frac{e^{\frac{A_{0,d}d^{1/2}}{4(1+\eta_d)^{1/2}}+\frac{T_d}{4}}d^{1/2}\pi^{d/4}}{\Gamma(1+d/2)^{1/2}} \\ &\quad\times\frac{\{(d-1)!\}^{1/2}2^{d/2}(1+\eta_d)^{d/4}}{A_{0,d}^{d/2}}, \end{aligned}$$

$$B_2 := \frac{e^{\frac{A_{0,d}d^{1/2}}{2(1+\eta_d)^{1/2}}+\frac{T_d}{2}-1}d^{1/2}\pi^{d/4}}{\Gamma(1+d/2)^{1/2}}\frac{(1+\eta_d)^{d/4}}{A_{0,d}^{d/2}}e^{-\frac{T_d}{2}+\frac{A_{0,d}}{2(1+\eta_d)^{1/2}}}(d+2)^{d/2}.$$

Note that to obtain the expression for B_2, we have used the fact that

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{j_0}^{\infty} r^{d-1} e^{-\frac{rA_{0,d}}{(1+\eta_d)^{1/2}}}\,dr \\
&\quad = \frac{(1+\eta_d)^{d/4}}{A_{0,d}^{d/2}}\{(d-1)!\}^{1/2} e^{-\frac{j_0 A_{0,d}}{2(1+\eta_d)^{1/2}}}\left\{\sum_{k=0}^{d-1}\frac{j_0^k A_{0,d}^k}{(1+\eta_d)^{k/2}k!}\right\}^{1/2}\varepsilon^{-1} \\
&\quad \le \frac{(1+\eta_d)^{d/4}}{A_{0,d}^{d/2}} e^{-\frac{T_d}{2}+\frac{A_{0,d}}{2(1+\eta_d)^{1/2}}}(d+2)^{d/2},
\end{aligned}$$

using the definition of j_0 and $\varepsilon_{01,d}$. Moreover, the cardinality of the bracketing set is

$$\prod_{\mathbf{j}:\|\mathbf{j}\|\le j_0} N_{\mathbf{j}} = \exp\left\{\bar{K}_{2,d}^* \sum_{\mathbf{j}:\|\mathbf{j}\|\le j_0} h_d\left(\frac{\varepsilon}{C_{\mathbf{j}}^{1/2}}\right)\right\} \le \exp\{\bar{K}_{2,d}^* B_{3,d} h_d(\varepsilon)\},$$

where

$$\begin{aligned}
B_{3,1} &:= \sum_{\mathbf{j}:\|\mathbf{j}\|\le j_0} C_{\mathbf{j}}^{1/4} \le e^{T_1/8} e^{\frac{A_{0,1}}{8(1+\eta_d)^{1/2}}}\frac{16(1+\eta_d)^{1/2}}{A_{0,1}}, \\
B_{3,2} &:= 2^{3/2}\sum_{\mathbf{j}:\|\mathbf{j}\|\le j_0} C_{\mathbf{j}}^{1/2} \le e^{T_2/4} 2^{5/2}\pi e^{\frac{A_{0,2}}{2^{3/2}(1+\eta_d)^{1/2}}}\frac{16(1+\eta_d)}{A_{0,2}^2}, \\
B_{3,3} &:= \sum_{\mathbf{j}:\|\mathbf{j}\|\le j_0} C_{\mathbf{j}} \le e^{T_3/2} 4\pi e^{\frac{3^{1/2}A_{0,3}}{2(1+\eta_d)^{1/2}}}\frac{8(1+\eta_d)^{3/2}}{A_{0,3}^3}.
\end{aligned}$$

Since $\varepsilon \in (0, \varepsilon_{01,d}]$ was arbitrary, we conclude that

$$\log N_{[\cdot]}\big(\varepsilon, \tilde{\mathcal{F}}_d^{1,\eta_d}, h\big) = \log N_{[\cdot]}\big(\varepsilon, \{\tilde{f}^{1/2} : \tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}, L_2\big) \le \overline{\overline{K}}_d h_d(\varepsilon),$$

for all $\varepsilon \in (0, \varepsilon_{02,d}]$, where $\varepsilon_{02,d} := \varepsilon_{01,d}(B_1 + B_2)$ and where

$$\overline{\overline{K}}_d := \bar{K}_{2,d}^* B_{3,d}\max\{(B_1+B_2)^{d/2}, (B_1+B_2)^{d-1}\}\left\{2 + \frac{2\log_{++}(B_1+B_2)}{\log_{++}(e/(B_1+B_2))}\right\},$$

where, as in the proof of Proposition 2 in the online supplement, we have used the fact that $\log_{++}(a/\varepsilon) \le \{2 + \frac{2\log_{++}(a)}{\log_{++}(e/a)}\}\log_{++}(1/\varepsilon)$ for all $a, \varepsilon > 0$. Now let

$$\varepsilon_{03,d} := \max\left\{\varepsilon_{02,d}, \left[\frac{(1+\eta_d)^{d/2}}{(1-\eta_d)^{d/2}} e^{\frac{A_{0,d}}{(1+\eta_d)^{1/2}}+B_{0,d}}\frac{d!\pi^{d/2}}{\Gamma(1+d/2)A_{0,d}^d}\right]^{1/2}\right\},$$

and let $\overline{K}_d := \overline{\overline{K}}_d h_d(\varepsilon_{02,d})/h_d(\varepsilon_{03,d})$. For $\varepsilon \in (\varepsilon_{02,d}, \varepsilon_{03,d}]$, we have

$$\begin{aligned}
\log N_{[\cdot]}\big(\varepsilon, \tilde{\mathcal{F}}_d^{1,\eta_d}, h\big) &\le \log N_{[\cdot]}\big(\varepsilon_{02,d}, \tilde{\mathcal{F}}_d^{1,\eta_d}, h\big) \le \overline{\overline{K}}_d h_d(\varepsilon_{02,d}) = \overline{K}_d h_d(\varepsilon_{03,d}) \\
&\le \overline{K}_d h_d(\varepsilon).
\end{aligned}$$

Finally, if $\varepsilon > \varepsilon_{03,d}$, we can use a single bracketing pair $\{f^L, f^U\}$, with $f^L(x) := 0$ and $f^U(x)$ defined to be the integrable envelope function from Corollary 3(a) with $\xi = 1$ and $\eta = \eta_d$ there. Note that $h(f^U, f^L) \le \varepsilon_{03,d}$. This proves the upper bound.

(ii) For this part of the proof, we use the Gilbert–Varshamov theorem, treating $d = 1$ and $d \ge 2$ separately, to construct a finite subset of $\tilde{\mathcal{F}}_d^{1,\eta_d}$ of the desired cardinality where each pair of functions is well separated in Hellinger distance. In the case $d = 1$, this is achieved by constructing densities that are perturbations of a semicircle (it is convenient to raise the semicircle to be bounded away from zero on its domain). In the case $d \ge 2$, we instead construct uniform densities on perturbations of a closed Euclidean ball B, in an almost identical fashion to Brunel (2013) (we simply need to choose the radius to ensure that the mean and variance restrictions are satisfied). Further details can be found in the `arxiv` version of this paper [Kim and Samworth (2015), Theorem 8(ii)]. □

PROOF OF THEOREM 5. Let $\mu := \mathbb{E}(X_1)$ and $\Sigma := \mathrm{Cov}(X_1)$. Note that since $f_0 \in \mathcal{F}_d$, we have that Σ is a finite, positive definite matrix. We can therefore define $Z_i := \Sigma^{-1/2}(X_i - \mu)$ for $i = 1, \dots, n$, so that $\mathbb{E}(Z_1) = 0$ and $\mathrm{Cov}(Z_1) = I$. We also set $g_0(z) := (\det \Sigma)^{1/2} f_0(\Sigma^{1/2} z + \mu)$, so $g_0 \in \mathcal{F}_d^{0,I}$, and let $\hat{g}_n(z) := (\det \Sigma)^{1/2} \hat{f}_n(\Sigma^{1/2} z + \mu)$, so by affine equivariance [Dümbgen, Samworth and Schuhmacher (2011), Remark 2.4], $\hat{g}_n$ is the log-concave maximum likelihood estimator of g_0 based on $Z_1, \dots, Z_n$.

Let $\hat{\mu}_n := \int_{\mathbb{R}^d} z \hat{g}_n(z)\, dz$ and $\hat{\Sigma}_n := \int_{\mathbb{R}^d} (z - \hat{\mu}_n)(z - \hat{\mu}_n)^T \hat{g}_n(z)\, dz$ respectively denote the mean vector and covariance matrix corresponding to $\hat{g}_n$. Then by Lemma 6 below, there exists $\eta_d \in (0, 1)$ and $n_0 \in \mathbb{N}$, depending only on d, such that for $n \ge n_0$, we have

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\big(\hat{g}_n \notin \tilde{\mathcal{F}}_d^{1,\eta_d}\big) \le \frac{1}{n^{4/5}}.$$

We can now apply Theorem 5 in Section 3 in the online supplement, which provides an exponential tail inequality controlling the performance of a maximum likelihood estimator in Hellinger distance in terms of a bracketing entropy integral. It is an immediate consequence of Theorem 7.4 of van de Geer (2000), although our notation is slightly different (in particular her definition of Hellinger distance is normalised with a factor of $1/\sqrt{2}$) and we have used the fact (apparent from her proofs) that, in her notation, we may take $C = 2^{13/2}$.

In Theorem 5 in the online supplement, we take $\bar{\mathcal{F}} := \{\frac{\tilde{f}+g_0}{2} : \tilde{f} \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}$. Note that if $[f^L, f^U]$ are elements of a bracketing set for $\tilde{\mathcal{F}}_d^{1,\eta_d}$, and we set $\bar{f}^L := \frac{f^L+g_0}{2}$ and $\bar{f}^U := \frac{f^U+g_0}{2}$, then

$$h^2(\bar{f}^U, \bar{f}^L) = \frac{1}{2}\int_{\mathbb{R}^d} \{(f^U + g_0)^{1/2} - (f^L + g_0)^{1/2}\}^2 \le \frac{1}{2} h^2(f^U, f^L).$$

It follows from this and our bracketing entropy bound (Theorem 4) that

$$\log N_{[\cdot]}(u, \bar{\mathcal{F}}, h) \le \log N_{[\cdot]}\big(2^{1/2}u, \tilde{\mathcal{F}}_d^{1,\eta_d}, h\big) \le \begin{cases} 2^{-1/4}\overline{K}_1 u^{-1/2}, & \text{for } d=1, \\ 2^{-1/2}\overline{K}_2 u^{-1}\log_{++}^{3/2}(1/u), & \text{for } d=2, \\ 2^{-1}\overline{K}_3 u^{-2}, & \text{for } d=3. \end{cases}$$

We now consider three different cases, assuming throughout that $n \ge d+1$ so that, with probability 1, the log-concave maximum likelihood estimator exists and is unique:

1. For $d=1$, we define $\delta_n := 2^{-1/2}M_1^{1/2}n^{-2/5}$, where we let $M_1 := \max\{(\frac{2^{37/2}}{3})^{8/5}\overline{K}_1^{4/5}, 2^{33}\}$. Then

$$\int_{\delta_n^2/2^{13}}^{\delta_n} \sqrt{\log N_{[\cdot]}(u, \bar{\mathcal{F}}, h)}\, du \le \frac{4}{2^{1/2}3}\overline{K}_1^{1/2}M_1^{3/8}n^{-3/10} \le 2^{-16}n^{1/2}\delta_n^2.$$

   Moreover, $\delta_n \le 2^{-17}M_1 n^{-3/10} = 2^{-16}n^{1/2}\delta_n^2$. We conclude by Theorem 5 in the online supplement that for $t \ge M_1$,

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\big[\{n^{4/5}h^2(\hat{g}_n, g_0) \ge t\} \cap \{\hat{g}_n \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}\big] \le 2^{13/2}\sum_{s=0}^{\infty}\exp\biggl(-\frac{2^{2s}tn^{1/5}}{2^{28}}\biggr) \le 2^{15/2}\exp\biggl(-\frac{tn^{1/5}}{2^{28}}\biggr),$$

   where the final bound follows because $tn^{1/5}/2^{28} \ge \log 2$.

2. For $d=2$, we define $\delta_n := 2^{-1/2}M_2^{1/2}n^{-1/3}\log^{1/2} n$, where $M_2 := \max\{2^{23}\overline{K}_2^{2/3}5^{4/3}/3, 2^{33}\}$. Let $n_{0,2}$ be large enough that $\delta_n \le 1/e$ for $n \ge n_{0,2}$. Then, for such n,

$$\begin{aligned}
&\int_{\delta_n^2/2^{13}}^{\delta_n} \sqrt{\log N_{[\cdot]}(u, \bar{\mathcal{F}}, h)}\, du \\
&\le 2^{-1/4}\overline{K}_2^{1/2}\int_0^{\delta_n} u^{-1/2}\log^{3/4}(1/u)\, du \\
&= 2^{-1/4}\overline{K}_2^{1/2}\int_{\log(1/\delta_n)}^{\infty} s^{3/4}e^{-s/2}\, ds \\
&= 2^{-1/4}\overline{K}_2^{1/2}\biggl\{2\delta_n^{1/2}\log^{3/4}\biggl(\frac{1}{\delta_n}\biggr) + \frac{3}{2}\int_{\log(1/\delta_n)}^{\infty} s^{-1/4}e^{-s/2}\, ds\biggr\} \\
&\le 2^{-1/4}\overline{K}_2^{1/2}5\delta_n^{1/2}\log^{3/4}(1/\delta_n) \le 2^{1/2}3^{-3/4}\overline{K}_2^{1/2}5\delta_n^{1/2}\log^{3/4} n \\
&\le 2^{-16}n^{1/2}\delta_n^2,
\end{aligned}$$

where we have used the fact that $2^{1/2}M_2^{-1/2}\log^{-1/2} n \le n^{1/3}$ in the penultimate inequality. We conclude that for $n \ge n_{0,2}$ and $t \ge M_2$, we have

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\biggl[\biggl\{\frac{n^{2/3}}{\log n} h^2(\hat{g}_n, g_0) \ge t\biggr\} \cap \{\hat{g}_n \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}\biggr] \le 2^{15/2} \exp\biggl(-\frac{tn^{1/3}\log n}{2^{28}}\biggr).$$

3. For $d = 3$, the entropy integral diverges as $\delta \searrow 0$, so we cannot bound the bracketing entropy integral by replacing the lower limit with zero. Nevertheless, we can set $\delta_n := 2^{-1/2}M_3^{1/2}n^{-1/4}\log^{1/2} n$, where $M_3 := \{2^{33/2}10\overline{K}_3^{1/2}, 2^{33}\}$. For $t \ge M_3$, we have

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\biggl[\biggl\{\frac{n^{1/2}}{\log n} h^2(\hat{g}_n, g_0) \ge t\biggr\} \cap \{\hat{g}_n \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}\biggr] \le 2^{15/2} \exp\biggl(-\frac{tn^{1/2}\log n}{2^{28}}\biggr).$$

Let $\rho_{n,1}^2 := n^{4/5}$, $\rho_{n,2}^2 := n^{2/3}(\log n)^{-1}$ and $\rho_{n,3}^2 := n^{1/2}(\log n)^{-1}$. We conclude that if $n \ge \max(n_0, d+1)$ (and also $n \ge n_{0,2}$ when $d = 2$), then

$$\begin{aligned}
&\rho_{n,d}^2 \sup_{f_0 \in \mathcal{F}_d} \mathbb{E}_{f_0}\{h^2(\hat{f}_n, f_0)\} \\
&= \rho_{n,d}^2 \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{E}_{g_0}\{h^2(\hat{g}_n, g_0)\} \\
&\le \sup_{g_0 \in \mathcal{F}_d^{0,I}} \int_0^\infty \mathbb{P}_{g_0}[\{\rho_{n,d}^2 h^2(\hat{g}_n, g_0) \ge t\} \cap \{\hat{g}_n \in \tilde{\mathcal{F}}_d^{1,\eta_d}\}]\,dt \\
&\quad + 2\rho_{n,d}^2 \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}(\hat{g}_n \notin \tilde{\mathcal{F}}_d^{1,\eta_d}) \le M_d + 2^{71/2} + 2,
\end{aligned}$$

as required. $\square$

LEMMA 6. *There exists $\eta_d \in (0,1)$ such that*

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}(\hat{g}_n \notin \tilde{\mathcal{F}}_d^{1,\eta_d}) = O(n^{-1})$$

as $n \to \infty$, where $\hat{g}_n$ denotes the log-concave maximum likelihood estimator based on a random sample $Z_1, \ldots, Z_n$ from g_0.

PROOF. For $g \in \mathcal{F}_d$, we write $\mu_g := \int_{\mathbb{R}^d} z g(z)\,dz$ and $\Sigma_g := \int_{\mathbb{R}^d} (z-\mu_g)(z-\mu_g)^T g(z)\,dz$. Note that for $n \geq d+1$, and for any $\eta_d \in (0,1)$,

$$\begin{aligned}
\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl(\hat{g}_n \notin \tilde{\mathcal{F}}_d^{1,\eta_d}\bigr) \leq{}& \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl(\|\mu_{\hat{g}_n}\| > 1\bigr) \\
&+ \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl\{\lambda_{\max}(\Sigma_{\hat{g}_n}) > 1+\eta_d\bigr\} \\
&+ \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl\{\lambda_{\min}(\Sigma_{\hat{g}_n}) < 1-\eta_d\bigr\}.
\end{aligned} \tag{5}$$

We treat the three terms on the right-hand side of (5) in turn. By Remark 2.3 of Dümbgen, Samworth and Schuhmacher (2011), we have that $\mu_{\hat{g}_n} = n^{-1} \times \sum_{i=1}^n Z_i =: \bar{Z}$, where the density of $n^{1/2}\bar{Z} := n^{1/2}(\bar{Z}_1,\ldots,\bar{Z}_d)^T$ belongs to $\mathcal{F}_d^{0,I}$. Taking $A_{0,d}, B_{0,d} > 0$ from Theorem 2(a), it follows that for any $t \geq 0$ and $j = 1,\ldots,d$,

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl(n^{1/2}|\bar{Z}_j| > t\bigr) \leq 2\int_t^\infty e^{-A_{0,d}x + B_{0,d}}\,dx = \frac{2}{A_{0,d}} e^{-A_{0,d}t + B_{0,d}}.$$

Hence,

$$\begin{aligned}
\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl(\|\mu_{\hat{g}_n}\| > 1\bigr) &\leq \sup_{g_0 \in \mathcal{F}_d^{0,I}} \sum_{j=1}^d \mathbb{P}_{g_0}\left(n^{1/2}|\bar{Z}_j| > \frac{n^{1/2}}{d^{1/2}}\right) \\
&\leq \frac{2d}{A_{0,d}} e^{-\frac{A_{0,d}n^{1/2}}{d^{1/2}} + B_{0,d}} = O(n^{-1}).
\end{aligned}$$

For the second term, we use Remark 2.3 of Dümbgen, Samworth and Schuhmacher (2011) again to see that $\lambda_{\max}(\Sigma_{\hat{g}_n}) \leq \lambda_{\max}(\tilde{\Sigma}_n)$, where $\tilde{\Sigma}_n := n^{-1}\sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^T = n^{-1}\sum_{i=1}^n Z_i Z_i^T - \bar{Z}\bar{Z}^T$ denotes the sample covariance matrix. For each $j = 1,\ldots,d$,

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \int_{\mathbb{R}^d} z_j^4 g_0(z)\,dz \leq 2\int_0^\infty z_j^4 e^{-A_{0,1}z_j + B_{0,1}}\,dz_j = \frac{48 e^{B_{0,1}}}{A_{0,1}^5}.$$

Writing $Z_i := (Z_{i1},\ldots,Z_{id})^T$, we deduce from the Gerschgorin circle theorem [Gerschgorin (1931), Gradshteyn and Ryzhik (2007)], Chebychev's inequality and Cauchy–Schwarz that

$$\begin{aligned}
&\sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl\{\lambda_{\max}(\Sigma_{\hat{g}_n}) > 1+\eta_d\bigr\} \\
&\quad\leq \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\bigl\{\lambda_{\max}(\tilde{\Sigma}_n) > 1+\eta_d\bigr\}
\end{aligned}$$

$$\leq \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\left(\bigcup_{j=1}^{d}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_{ij}^2 - 1\right\} > \frac{\eta_d}{3}\right)$$

$$+ \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\left(\bigcup_{1\leq j<k\leq d}\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ij}Z_{ik}\right| > \frac{\eta_d}{3d}\right)$$

$$+ \sup_{g_0 \in \mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\left(\|\bar{Z}\|^2 > \frac{\eta_d}{3}\right)$$

$$\leq \frac{432de^{B_{0,1}}}{A_{0,1}^5\eta_d^2 n} + \frac{216d^3(d-1)e^{B_{0,1}}}{A_{0,1}^5\eta_d^2 n} + \frac{2d}{A_{0,d}} e^{-\frac{A_{0,d}\eta_d^{1/2}n^{1/2}}{3^{1/2}d^{1/2}} + B_{0,d}}$$

$$= O(n^{-1}).$$

The third term on the right-hand side of (5) is the most challenging to handle. Let $\mathcal{P}^{1/10,1/2}$ denote the class of probability distributions P on $\mathbb{R}^d$ such that $\mu_P := \int_{\mathbb{R}^d} x\, dP(x)$ and $\Sigma_P := \int_{\mathbb{R}^d}(x-\mu_P)(x-\mu_P)^T\, dP(x)$ satisfy $\|\mu_P\| \leq 1/10$ and $1/2 \leq \lambda_{\min}(\Sigma_P) \leq \lambda_{\max}(\Sigma_P) \leq 3/2$, and such that

$$\int_{\mathbb{R}^d} \|x\|^4\, dP(x) \leq \frac{2d\pi^{d/2}\Gamma(d+4)}{\Gamma(1+d/2)} \frac{e^{B_{0,d}}}{A_{0,d}^{d+4}} =: \tau_{4,d},$$

say, where $A_{0,d}$ and $B_{0,d}$ are taken from Theorem 2(a). By Theorem 2(a),

$$\sup_{g_0 \in \mathcal{F}_d^{0,I}} \int_{\mathbb{R}^d} \|x\|^4 g_0(x)\, dx \leq \int_{\mathbb{R}^d} \|x\|^4 e^{-A_{0,d}\|x\|+B_{0,d}}\, dx$$

$$= \frac{d\pi^{d/2}e^{B_{0,d}}}{\Gamma(1+d/2)} \int_0^\infty r^{d+3}e^{-A_{0,d}r}\, dr = \frac{\tau_{4,d}}{2}.$$

Recall from Theorem 2.2 of Dümbgen, Samworth and Schuhmacher (2011) that for $P \in \mathcal{P}^{1/10,1/2}$, there exists a unique log-concave projection $\psi^*(P) \in \mathcal{F}_d$ given by

$$\psi^*(P) := \operatorname*{argmax}_{f \in \mathcal{F}_d} \int_{\mathbb{R}^d} \log f\, dP.$$

Our first claim is that there exists $M_{0,d} > 0$, depending only on d, such that

$$\sup_{P \in \mathcal{P}^{1/10,1/2}} \sup_{x \in \mathbb{R}^d} \log \psi^*(P)(x) \leq M_{0,d}.$$

To see this, suppose that there exist $(P_n) \in \mathcal{P}^{1/10,1/2}$ such that

$$\sup_{x \in \mathbb{R}^d} \log \psi^*(P_n)(x) \to \infty.$$

Note that for any $R > 0$,

$$\begin{aligned}\sup_{n\in\mathbb{N}} P_n\big(\bar{B}(0, R)^c\big) &\leq \sup_{n\in\mathbb{N}} \frac{1}{R^2}\int_{\mathbb{R}^d} \|x\|^2\, dP_n(x)\\ &\leq \sup_{n\in\mathbb{N}} \frac{d\lambda_{\max}(\Sigma_{P_n}) + \|\mu_{P_n}\|^2}{R^2}\\ &\leq \frac{3d}{2R^2} + \frac{1}{100R^2} \to 0\end{aligned}$$

as $R \to \infty$, so the sequence (P_n) is tight. We deduce from Prohorov's theorem that there exists a subsequence (P_{n_k}) and a probability measure P on $\mathbb{R}^d$ such that $P_{n_k} \stackrel{d}{\to} P$. If (Y_{n_k}) is a sequence of random vectors on the same probability space with $Y_{n_k} \sim P_{n_k}$, then $\{\|Y_{n_k}\| : k \in \mathbb{N}\}$ is uniformly integrable, because $\mathbb{E}(\|Y_{n_k}\|^2) \leq 3d/2 + 1/100$. We deduce that $\int_{\mathbb{R}^d} \|x\|\,dP_{n_k}(x) \to \int_{\mathbb{R}^d} \|x\|\, dP(x)$. Together with the weak convergence, this means that P_{n_k} converges to P in the Wasserstein distance. Moreover, for any unit vector $u \in \mathbb{R}^d$, the family $\{(u^T Y_{n_k})^2 : k \in \mathbb{N}\}$ is uniformly integrable, because $\mathbb{E}\{(u^T Y_{n_k})^4\} \leq \mathbb{E}(\|Y_{n_k}\|^4) \leq \tau_{4,d}$. Thus, $u^T \Sigma_P u = \lim_{k\to\infty} u^T \Sigma_{P_{n_k}} u \geq 1/2$, so in particular, $P(H) < 1$ for every hyperplane H in $\mathbb{R}^d$. We conclude by Theorem 2.15 and Remark 2.16 of Dümbgen, Samworth and Schuhmacher (2011) that $\psi^*(P_{n_k})$ converges to $\psi^*(P)$ uniformly on closed subsets of $\mathbb{R}^d \setminus \mathrm{disc}(\psi^*(P))$, where $\mathrm{disc}(\psi^*(P))$ denotes the set of discontinuity points of $\psi^*(P)$. In turn, this implies that

$$\sup_{x\in\mathbb{R}^d} \psi^*(P_{n_k})(x) \leq \sup_{x\in\mathbb{R}^d} \psi^*(P)(x) + 1$$

for sufficiently large k, which establishes our desired contradiction.

Moreover, by Theorem 2(b), there exists $a_{0,d} > 0$, depending only on d, such that

$$\inf_{f\in\mathcal{F}_d^{0,I}} f(0) \geq a_{0,d}.$$

It follows that for any $\mu \in \mathbb{R}^d$,

$$\inf_{f\in\mathcal{F}_d^{\mu,\Sigma}} \sup_{x\in\mathbb{R}^d} f(x) \geq a_{0,d}(\det\Sigma)^{-1/2}.$$

Thus, using our claim, if $\det\Sigma < a_{0,d}^2 e^{-2M_{0,d}}$, then $\{\psi^*(P) : P \in \mathcal{P}^{1/10,1/2}\} \cap (\bigcup_{\mu\in\mathbb{R}^d} \mathcal{F}_d^{\mu,\Sigma}) = \varnothing$. Since $\sup_{P\in\mathcal{P}^{1/10,1/2}} \lambda_{\max}(\Sigma_P) \leq 3/2$, we deduce that if $\lambda_{\min}(\Sigma) < 2^{d-1}a_{0,d}^2 e^{-2M_{0,d}}/3^{d-1}$, then

$$\{\psi^*(P) : P \in \mathcal{P}^{1/10,1/2}\} \cap \bigg(\bigcup_{\mu\in\mathbb{R}^d} \mathcal{F}_d^{\mu,\Sigma}\bigg) = \varnothing.$$

Finally, we conclude that if we define $\eta_d := 1 - \frac{2^{d-2}a_{0,d}^2 e^{-2M_{0,d}}}{3^{d-1}}$, then

$$\begin{aligned}
&\sup_{g_0\in\mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\{\lambda_{\min}(\Sigma_{\hat{g}_n}) < 1-\eta_d\} \\
&\quad\le \sup_{g_0\in\mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\{\lambda_{\min}(\tilde{\Sigma}_n) < 1/2\} \\
&\qquad + \sup_{g_0\in\mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\{\lambda_{\max}(\tilde{\Sigma}_n) > 3/2\} + \sup_{g_0\in\mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}(\|\bar{Z}\| > 1/10) \\
&\qquad + \sup_{g_0\in\mathcal{F}_d^{0,I}} \mathbb{P}_{g_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \{\|Z_i\|^4 - \mathbb{E}(\|Z_1\|^4)\}\right| > \frac{\tau_{4,d}}{2}\right) \\
&\quad= O(n^{-1}),
\end{aligned}$$

using very similar arguments to those used above, as well as Chebychev's inequality for the last term. □

Acknowledgements. The authors are very grateful for helpful comments on an earlier draft from Charles Doss, Roy Han and Jon Wellner, as well as anonymous reviewers.

SUPPLEMENTARY MATERIAL

Supplementary material to "Global rates of convergence in log-concave density estimation" (DOI: 10.1214/16-AOS1480SUPP; .pdf). Proof of Theorem 1 and auxiliary results.

REFERENCES

ALEKSANDROV, A. D. (1939). Almost everywhere existence of the second differential of a convex functions and related properties of convex surfaces. *Uchenye Zapisky Leningrad. Gos. Univ. Math. Ser.* **37** 3–35.

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719

BRUNEL, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Stat.* **7** 1301–1327. MR3063609

BRUNEL, V.-E. (2016). Adaptive estimation of convex and polytopal density support. *Probab. Theory Related Fields* **164** 1–16. MR3449384

CHEN, Y. and SAMWORTH, R. J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica* **23** 1373–1398. MR3114718

CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 545–607. MR2758237

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. MR3485950

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. MR2546798

DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336

FRESEN, D. (2013). A multivariate Gnedenko law of large numbers. *Ann. Probab.* **41** 3051–3080. MR3127874

GAO, F. and WELLNER, J. A. (2015). Entropy of convex functions on $\mathbb{R}^d$. Available at http://arxiv.org/abs/1502.01752.

GERSCHGORIN, S. (1931). Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk* **6** 749–754.

GRADSHTEYN, I. S. and RYZHIK, I. M. (2007). *Table of Integrals, Series, and Products*, 7th ed. Elsevier/Academic Press, Amsterdam. MR2360010

GUNTUBOYINA, A. and SEN, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inform. Theory* **59** 1957–1965. MR3043776

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1983). Estimation of distribution density. *J. Sov. Math.* **25** 40–57.

KIM, A. K. H. and SAMWORTH, R. J. (2015). Global rates of convergence in log-concave density estimation. Available at http://arxiv.org/abs/1404.2298v2.

KIM, A. K. H. and SAMWORTH, R. J. (2016). Supplement to "Global rates of convergence in log-concave density estimation." DOI:10.1214/16-AOS1480SUPP.

KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. Springer, New York. MR1226450

LOVÁSZ, L. and VEMPALA, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. MR1332579

MÜLLER, S. and RUFIBACH, K. (2009). Smooth tail-index estimation. *J. Stat. Comput. Simul.* **79** 1155–1167. MR2572422

PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459196

SAMWORTH, R. J. and YUAN, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40** 2973–3002. MR3097966

SCHUHMACHER, D. and DÜMBGEN, L. (2010). Consistency of multivariate log-concave density estimators. *Statist. Probab. Lett.* **80** 376–380. MR2593576

SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. MR2766867

VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500

STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE
CB3 0WB
UNITED KINGDOM
E-MAIL: r.samworth@statslab.cam.ac.uk
a.kim@statslab.cam.ac.uk
URL: http://www.statslab.cam.ac.uk/~rjs57
http://sites.google.com/site/kyoungheearlene/home