The Annals of Statistics
2010, Vol. 38, No. 6, 3751–3781
DOI: 10.1214/10-AOS840

NONPARAMETRIC ESTIMATION OF MULTIVARIATE CONVEX-TRANSFORMED DENSITIES

BY ARSENI SEREGIN[1] AND JON A. WELLNER[1,2]

University of Washington

We study estimation of multivariate densities p of the form $p(x) = h(g(x))$ for $x \in \mathbb{R}^d$ and for a fixed monotone function h and an unknown convex function g. The canonical example is $h(y) = e^{-y}$ for $y \in \mathbb{R}$; in this case, the resulting class of densities

$$\mathcal{P}(e^{-y}) = \{p = \exp(-g) : g \text{ is convex}\}$$

is well known as the class of *log-concave* densities. Other functions h allow for classes of densities with heavier tails than the log-concave class.

We first investigate when the maximum likelihood estimator $\hat{p}$ exists for the class $\mathcal{P}(h)$ for various choices of monotone transformations h, including decreasing and increasing functions h. The resulting models for increasing transformations h extend the classes of *log-convex* densities studied previously in the econometrics literature, corresponding to $h(y) = \exp(y)$.

We then establish consistency of the maximum likelihood estimator for fairly general functions h, including the log-concave class $\mathcal{P}(e^{-y})$ and many others. In a final section, we provide asymptotic minimax lower bounds for the estimation of p and its vector of derivatives at a fixed point x_0 under natural smoothness hypotheses on h and g. The proofs rely heavily on results from convex analysis.

1. Introduction and background.

1.1. *Log-concave and r-concave densities.* A probability density p on $\mathbb{R}^d$ is called *log-concave* if it can be written as

$$p(x) = \exp(-g(x))$$

for some convex function $g : \mathbb{R}^d \to (-\infty, \infty]$. We let $\mathcal{P}(e^{-y})$ denote the class of all log-concave densities on $\mathbb{R}^d$. As shown by Ibragimov (1956), a density function p on $\mathbb{R}$ is log-concave if and only if its convolution with any unimodal density is again unimodal.

Received January 2010; revised May 2010.
[1]Supported in part by NSF Grant DMS-08-04587.
[2]Supported in part by NI-AID Grant 2R01 AI291968-04.
AMS 2000 subject classifications. Primary 62G07, 62H12; secondary 62G05, 62G20.
Key words and phrases. Consistency, log-concave density estimation, lower bounds, maximum likelihood, mode estimation, nonparametric estimation, qualitative assumptions, shape constraints, strongly unimodal, unimodal.

Log-concave densities have proven to be useful in a wide range of statistical problems; see Walther (2010) for a survey of recent developments and statistical applications of log-concave densities on $\mathbb{R}$ and $\mathbb{R}^d$, and see Cule, Samworth and Stewart (2010) for several interesting applications of estimators of such densities in $\mathbb{R}^d$.

Because the class of multivariate log-concave densities contains the class of multivariate normal densities and is preserved under a number of important operations (such as convolution and marginalization), it serves as a valuable nonparametric surrogate or replacement for the class of normal densities. Further study of the class of log-concave densities from this perspective has been undertaken by Schuhmacher, Hüsler and Duembgen (2009).

Log-concave densities have the slight drawback that the tails must be decreasing exponentially, so a number of authors, including Koenker and Mizera (2010), have proposed using generalizations of the log-concave family involving r-concave densities, defined as follows. For $a, b \in \mathbb{R}$, $r \in \mathbb{R}$ and $\lambda \in (0, 1)$, define the generalized mean of order r, $M_r(a, b; \lambda)$, for $a, b \geq 0$, by

$$M_r(a,b;\lambda) = \begin{cases} \big((1-\lambda)a^r + \lambda b^r\big)^{1/r}, & r \neq 0, a, b > 0, \\ 0, & r < 0, ab = 0, \\ a^{1-\lambda}b^{\lambda}, & r = 0. \end{cases}$$

A density function p is then r-concave on $C \subset \mathbb{R}^d$ if and only

$$p\big((1-\lambda)x + \lambda y\big) \geq M_r(p(x), p(y); \lambda) \qquad \text{for all } x, y \in C, \lambda \in (0, 1).$$

We denote the class of all r-concave densities on $C \subset \mathbb{R}^d$ by $\widehat{\mathcal{P}}(y_+^{1/r}; C)$ and write $\widehat{\mathcal{P}}(y_+^{1/r})$ when $C = \mathbb{R}^d$. As noted by Dharmadhikari and Joag-Dev [(1988), page 86], for $r \leq 0$, it suffices to consider $\widehat{\mathcal{P}}(y_+^{1/r})$, and it is almost immediate from the definitions that $p \in \widehat{\mathcal{P}}(y_+^{1/r})$ if and only if $p(x) = (g(x))^{1/r}$ for some convex function g from $\mathbb{R}^d$ to $[0, \infty)$. For $r > 0$, $p \in \widehat{\mathcal{P}}(y_+^{1/r}; C)$ if and only if $p(x) = (g(x))^{1/r}$, where g mapping C into $(0, \infty)$ is concave.

These results motivate definitions of the classes $\mathcal{P}(y_+^{-s}) = \{p(x) = g(x)^{-s} : g$ is convex$\}$ for $s \geq 0$ and, more generally, for a fixed monotone function h from $\mathbb{R}$ to $\mathbb{R}$,

$$\mathcal{P}(h) \equiv \{h \circ g : g \text{ convex}\}.$$

Such generalizations of log-concave densities and log-concave measures based on means of order r have been introduced by a series of authors, sometimes with differing terminology, apparently starting with Avriel (1972), and continuing with Borell (1975), Brascamp and Lieb (1976), Prékopa (1973), Rinott (1976) and Uhrin (1984). A nice summary of these connections is given by Dharmadhikari and Joag-Dev (1988). These authors also present results concerning the preservation of r-concavity under a variety of operations, including products, convolutions and marginalization.

Despite the longstanding and current rapid development of the properties of such classes of densities on the probability side, very little has been done from the standpoint of nonparametric estimation, especially when $d \geq 2$.

Nonparametric estimation of a log-concave density on $\mathbb{R}^d$ was initiated by Cule, Samworth and Stewart (2010). These authors developed an algorithm for computing their estimators and explored several interesting applications. Koenker and Mizera (2010) developed a family of penalized criterion functions related to the Rényi divergence measures and explored duality in the optimization problems. They did not succeed in establishing consistency of their estimators, but did investigate Fisher consistency. Recently, Cule and Samworth (2010) have established consistency of the (nonparametric) maximum likelihood estimator of a log-concave density on $\mathbb{R}^d$, even in a setting of model misspecification: when the true density is not log-concave, the estimator converges to the closest log-concave density to the true density, in the sense of Kullback–Leibler divergence.

In this paper, our goal is to investigate maximum likelihood estimation in the classes $\mathcal{P}(h)$ corresponding to a fixed monotone (decreasing or increasing) function h. In particular, for decreasing functions h, we handle all of the r-concave classes $\mathcal{P}(y_+^{1/r})$ with $r = -1/s$ and $r \leq -1/d$ (or $s \geq d$). On the increasing side, we treat, in particular, the cases $h(y) = y1_{[0,\infty)}(y)$ and $h(y) = e^y$ with $C = \mathbb{R}_+^d$. The first of these corresponds to an interesting class of models which can be thought of as multivariate generalizations of the class of decreasing and convex densities on $\mathbb{R}_+$ treated by Groeneboom, Jongbloed and Wellner (2001), while the second, $h(y) = e^y$, corresponds to multivariate versions of the log-convex families studied by An (1998). Note that our increasing classes $\mathcal{P}(y_+^{1/r}, \mathbb{R}_+^d)$ with $r > 0$ are quite different from the r-concave classes defined above and appear to be completely new, corresponding instead to r-convex densities on $\mathbb{R}_+^d$.

Here is an outline of the rest of the paper. All of our main results are presented in Section 2. Section 2.1 gives definitions and basic properties of the transformations involved. Section 2.2 establishes existence of the maximum likelihood estimators for both increasing and decreasing transformations h under suitable conditions on the function h. In Section 2.3, we give statements concerning consistency of the estimators, both in the Hellinger metric and in uniform metrics under natural conditions. In Section 2.4, we present asymptotic minimax lower bounds for estimation in these classes under natural curvature hypotheses. We conclude the section with a brief discussion of conjectures concerning attainability of the minimax rates by the maximum likelihood estimators. All of the proofs are given in Section 3.

Supplementary material and some proofs omitted here are available in Seregin and Wellner (2010). There, we also summarize a number of definitions and key results from convex analysis in an Appendix, Section A. We use standard notation from convex analysis; see "Notation" for a (partial) list.

1.2. *Convex-transformed density estimation.* Now, let $X_1, \dots, X_n$ be n independent random variables distributed according to a probability density $p_0 =$

$h(g_0(x))$ on $\mathbb{R}^d$, where h is a fixed monotone (increasing or decreasing) function and g_0 is an (unknown) convex function. The probability measure on the Borel sets $\mathcal{B}_d$ corresponding to p_0 is denoted by P_0.

The maximum likelihood estimator (MLE) of a log-concave density on $\mathbb{R}$ was introduced in Rufibach (2006) and Dümbgen and Rufibach (2009). Algorithmic aspects were treated in Rufibach (2007) and, in a more general framework, in Dümbgen, Hüsler and Rufibach (2007), while consistency with respect to the Hellinger metric was established by Pal, Woodroofe and Meyer (2007) and rates of convergence of $\hat{f}_n$ and $\widehat{F}_n$ were established by Dümbgen and Rufibach (2009). Asymptotic distribution theory for the MLE of a log-concave density on $\mathbb{R}$ was established by Balabdaoui, Rufibach and Wellner (2009).

If $\mathcal{C}$ denotes the class of all closed proper convex functions $g : \mathbb{R}^d \to (-\infty, \infty]$, the estimator $\hat{g}_n$ of g_0 is the maximizer of the functional

$$\mathbb{L}_n g \equiv \int (\log h) \circ g \, d\mathbb{P}_n$$

over the class $\mathcal{G}(h) \subset \mathcal{C}$ of all convex functions g such that $h \circ g$ is a density and where $\mathbb{P}_n$ is the empirical measure of the observations. The maximum likelihood estimator of the convex-transformed density p_0 is then $\hat{p}_n := h(\hat{g}_n)$ when it exists and is unique. We investigate conditions for existence and uniqueness in Section 2.

2. Main results.

2.1. *Definitions and basic properties.* To construct the classes of convex-transformed densities of interest here, we first need to define two classes of monotone transformations. An *increasing transformation* h is a nondecreasing function $\overline{\mathbb{R}} \to \overline{\mathbb{R}}_+$ such that $h(-\infty) = 0$ and $h(+\infty) = +\infty$. We define the limit points $y_0 < y_\infty$ of the increasing transformation h as follows: $y_0 = \inf\{y : h(y) > 0\}$, $y_\infty = \sup\{y : h(y) < +\infty\}$. We make the following assumptions about the asymptotic behavior of the increasing transformation:

(I.1) the function $h(y)$ is $o(|y|^{-\alpha})$ for some $\alpha > d$ as $y \to -\infty$;
(I.2) if $y_\infty < +\infty$, then $h(y) \asymp (y_\infty - y)^{-\beta}$ for some $\beta > d$ as $y \uparrow y_\infty$;
(I.3) the function h is continuously differentiable on the interval (y_0, y_∞).

Note that the assumption (I.1) is satisfied if $y_0 > -\infty$.

DEFINITION 2.1. For an increasing transformation h, an *increasing class of convex-transformed densities* or simply an increasing model $\mathcal{P}(h)$ on $\overline{\mathbb{R}}^d_+$ is the family of all bounded densities which have the form $h \circ g \equiv h(g(\cdot))$, where g is a closed proper convex function with $\operatorname{dom} g = \overline{\mathbb{R}}^d_+$.

REMARK 2.2. Consider a density $h \circ g$ from an increasing model $\mathcal{P}(h)$. Since $h \circ g$ is bounded, we have $g < y_\infty$. The function $\tilde{g} = \max(g, y_0)$ is convex and $h \circ \tilde{g} = h \circ g$. Thus, we can assume that $g \geq y_0$.

A *decreasing transformation* h is a nonincreasing function $\overline{\mathbb{R}} \to \overline{\mathbb{R}}_+$ such that $h(-\infty) = +\infty$ and $h(+\infty) = 0$. We define the limit points $y_0 > y_\infty$ of the decreasing transformation h as follows: $y_0 = \sup\{y : h(y) > 0\}$, $y_\infty = \inf\{y : h(y) < +\infty\}$. We make the following assumptions about the asymptotic behavior of the decreasing transformation:

(D.1) the function $h(y)$ is $o(y^{-\alpha})$ for some $\alpha > d$ as $y \to +\infty$;
(D.2) if $y_\infty > -\infty$, then $h(y) \asymp (y - y_\infty)^{-\beta}$ for some $\beta > d$ as $y \downarrow y_\infty$;
(D.3) if $y_\infty = -\infty$, then $h(y)^\gamma h(-Cy) = o(1)$ for some $\gamma, C > 0$ as $y \to -\infty$;
(D.4) the function h is continuously differentiable on the interval (y_∞, y_0).

Note that the assumption (D.1) is satisfied if $y_0 < +\infty$. We now define the decreasing class of densities $\mathcal{P}(h)$.

DEFINITION 2.3. For a decreasing transformation h, a *decreasing class of convex-transformed densities* or simply a decreasing model $\mathcal{P}(h)$ on $\mathbb{R}^d$ is the family of all bounded densities which have the form $h \circ g$, where g is a closed proper convex function with $\dim(\operatorname{dom} g) = d$.

REMARK 2.4. Consider a density $h \circ g$ from a decreasing model $\mathcal{P}(h)$. Since $h \circ g$ is bounded, we have $g > y_\infty$. For the sublevel set $C = \operatorname{lev}_{y_0} g$, the function $\tilde{g} = g + \delta(\cdot|C)$ is convex and $h \circ \tilde{g} = h \circ g$. Thus, we can assume that $\operatorname{lev}_{y_0} g = \operatorname{dom} g$.

For a monotone transformation h, we denote by $\mathcal{G}(h)$ the class of all closed proper convex functions g such that $h \circ g$ belongs to a monotone class $\mathcal{P}(h)$. The following lemma allows us to compare models defined by increasing or decreasing transformations h.

LEMMA 2.5. *Consider two decreasing* (*or increasing*) *models* $\mathcal{P}(h_1)$ *and* $\mathcal{P}(h_2)$. *If* $h_1 = h_2 \circ f$ *for some convex function* f, *then* $\mathcal{P}(h_1) \subseteq \mathcal{P}(h_2)$.

PROOF. The argument below is for a decreasing model. For an increasing model, the proof is similar. If $f(x) > f(y)$ for some $x < y$, then f is decreasing on $(-\infty, x)$, $f(-\infty) = +\infty$ and therefore h_2 is constant on $(f(x), +\infty)$, and we can redefine $f(y) = f(x)$ for all $y < x$. Thus, we can always assume that f is nondecreasing.

For any convex function g, the function $f \circ g$ is also convex. Therefore, if $p = h_1 \circ g \in \mathcal{P}(h_1)$, then $p = h_2 \circ f \circ g \in \mathcal{P}(h_2)$. $\square$

In this section, we discuss several examples of monotone models. The first two families are based on increasing transformations h.

EXAMPLE 2.6 (Log-convex densities). This increasing model is defined by $h(y) = e^y$. Limit points are $y_0 = -\infty$ and $y_\infty = \infty$. Assumption (I.1) holds for any $\alpha > d$. These classes of densities were considered by An (1998), who established several useful preservation properties. In particular, log-convexity is preserved under mixtures [An (1998), Proposition 3] and under marginalization [An (1998), Remark 8, page 361].

EXAMPLE 2.7 (r-convex densities). This family of increasing models is defined by the transforms $h(y) = \max(y, 0)^s = y_+^s$ with $s > 0$. Limit points are $y_0 = 0$ and $y_\infty = +\infty$. Assumption (I.1) holds for any $\alpha > d$. As noted in Section 1, the model $\mathcal{P}(y_+^{1/r}, \mathbb{R}_+^d)$ corresponds to the class of r-convex densities, with $r = \infty$ corresponding to the log-convex densities of the previous example. For $r < \infty$, these classes appear not to have been previously discussed or considered, except in special cases: the case $r = 1$ and $d = 1$ corresponds to the class of decreasing convex densities on $\mathbb{R}_+$ considered by Groeneboom, Jongbloed and Wellner (2001). It follows from Lemma 2.5 that

$$\mathcal{P}(e^y, \mathbb{R}_+^d) \subset \mathcal{P}(y_+^{s_2}, \mathbb{R}_+^d) \subset \mathcal{P}(y_+^{s_1}, \mathbb{R}_+^d) \qquad \text{for } 0 < s_1 < s_2 < \infty. \tag{2.1}$$

We now consider some models based on decreasing transformations h.

EXAMPLE 2.8 (Log-concave densities). This decreasing model is defined by the transform $h(y) = e^{-y}$. Limit points are $y_0 = +\infty$ and $y_\infty = -\infty$. Assumption (D.1) holds for any $\alpha > d$. Assumption (D.3) holds for any $\gamma > C > 0$.

Many parametric models are subsets of this model: in particular, uniform, Gaussian, gamma, beta, Gumbel, Fréchet and logistic densities are all log-concave.

EXAMPLE 2.9 (r-concave densities and power-convex densities). This family of decreasing models is defined by the transforms $h(y) = y_+^{-s}$ for $s > d$. Limit points are $y_0 = +\infty$ and $y_\infty = 0$. Assumption (D.1) holds for any $\alpha \in (d, s)$. Assumption (D.2) holds for $\beta = s$. As noted in Section 1, the model $\mathcal{P}(y_+^{1/r}) = \mathcal{P}(y_+^{-s})$ (with $r = -1/s < 0$) corresponds to the class of r-concave densities. From Lemma 2.5, we have the following inclusion:

$$\mathcal{P}(e^{-y}) \subset \mathcal{P}(y_+^{-s_2}) \subset \mathcal{P}(y_+^{-s_1}) \qquad \text{for } s_1 < s_2. \tag{2.2}$$

The models defined by power transformations include some parametric models with heavier-than-exponential tails. Several examples, including the multivariate generalizations of Pareto, Student-t, and F-distributions are discussed in Borell (1975)—none of these families are log-concave; see Johnson and Kotz (1972) and Seregin and Wellner (2010) for explicit computations.

Borell (1975) developed a framework which unifies log-concave and power-convex densities and gives an interesting characterization for these classes. Here, we briefly state the main result.

DEFINITION 2.10. Let $C \subseteq \mathbb{R}^d$ be an open convex set and let $s \in \overline{\mathbb{R}}$. We then define $\mathcal{M}_s(C)$ as the family of all positive Radon measures μ on C such that

$$\mu_*(\theta A + (1-\theta)B) \geq [\theta \mu_*(A)^s + (1-\theta)\mu_*(B)^s]^{1/s} \tag{2.3}$$

holds for all $\varnothing \neq A, B \subseteq C$ and all $\theta \in (0,1)$. We define $\mathcal{M}_s^\circ(C)$ as a subfamily of $\mathcal{M}_s(C)$ which consists of probability measures such that the affine hull of its support has dimension d. Here, μ_* is the inner measure corresponding to μ and the cases $s = 0, \infty$ are defined by continuity.

One of the main results of Borell (1975), Prékopa (1973) and Rinott (1976) is as follows.

THEOREM 2.11 (Borell, Prekopa, Rinott). *For $s < 0$, the family $\mathcal{M}_s^\circ(\mathbb{R}^d)$ coincides with the power-convex family $\mathcal{P}(y_+^{-d+1/s})$. For $s = 0$, the family $\mathcal{M}_0^\circ(\mathbb{R}^d)$ coincides with the log-concave family $\mathcal{P}(e^{-y})$. This continues to hold if* (2.3) *holds with $\theta = 1/2$ for all compact* (*or open, or semi-open*) *blocks $A, B \subseteq \Omega$* (*i.e., rectangles with sides parallel to the coordinate axes*).

Theorem 2.11 provides a special case of what has come to be known as the *Borell–Brascamp–Lieb* inequality; see, for example, Dharmadhikari and Joag-Dev (1988) and Brascamp and Lieb (1976). The current terminology is apparently due to Cordero-Erausquin, McCann and Schmuckenschläger (2001).

2.2. *Existence of the maximum likelihood estimators.* Now, suppose that $X_1, \dots, X_n$ are i.i.d. with density $p_0(x) = h(g_0(x))$ for a fixed monotone transformation h and a convex function g_0. As before, $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure of the X_i's and P_0 is the probability measure corresponding to p_0. Then, $\mathbb{L}_n g = \mathbb{P}_n \log h \circ g$ is the log-likelihood function (divided by n) and

$$\hat{p}_n \equiv \arg\max\{\mathbb{L}_n g : h \circ g \in \mathcal{P}(h)\}$$

is the *maximum likelihood estimator of p over the class* $\mathcal{P}(h)$, assuming it exists and is unique. We also write $\hat{g}_n$ for the MLE of g. We first state our main results concerning existence and uniqueness of the MLEs for the classes $\mathcal{P}(h)$.

THEOREM 2.12. *Suppose that h is an increasing transformation satisfying assumptions* (I.1)–(I.3). *The MLE $\hat{p}_n$ then exists almost surely for the model $\mathcal{P}(h)$.*

THEOREM 2.13. *Suppose that h is a decreasing transformation satisfying assumptions* (D.1)–(D.4). *The MLE $\hat{p}_n$ then exists almost surely for the model $\mathcal{P}(h)$ if*

$$n \geq n_d \equiv d + d\gamma 1_{\{-\infty\}}(y_\infty) + \frac{\beta d^2}{\alpha(\beta - d)} 1\{y_\infty > -\infty\}.$$

Uniqueness of the MLE is known for the log-concave model $\mathcal{P}(e^{-y})$; see, for example, Dümbgen and Rufibach (2009) for $d = 1$ and Cule, Samworth and Stewart (2010) for $d \geq 1$. For a brief further comment, see Section 2.5.

2.3. *Consistency of the maximum likelihood estimators*. Once existence of the MLEs is ensured, our attention shifts to other properties of the estimators: our main concern in this subsection is consistency. While, for a decreasing model, it is possible to prove consistency without any restrictions, for an increasing model, we need the following assumptions about the true density $h \circ g_0$:

(I.4) the function g_0 is bounded by some constant $C < y_\infty$;
(I.5) if $d > 1$, then we have, with $V(x) \equiv \prod_{j=1}^d x_j$ for $x \in \mathbb{R}_+^d$,

$$C_g \equiv \int_{\mathbb{R}_+^d} \log\left(\frac{1}{V(x) \wedge 1}\right) dP_0(x) < \infty.$$

REMARK 2.14. Note that for $d = 1$, the assumption (I.5) follows from assumption (I.4) and integrability of $\log(1/x)$ at zero. This assumption is also true if P has finite marginal densities.

(I.6) We have $\int_{\mathbb{R}_+^d} (h|\log h|) \circ g_0(x)\, dx < \infty$.

Let $H(p, q)$ denote the Hellinger distance between two probability measures with densities p and q with respect to Lebesgue measure on $\mathbb{R}^d$:

$$H^2(p, q) \equiv \frac{1}{2} \int_{\mathbb{R}^d} \left(\sqrt{p}(x) - \sqrt{q}(x)\right)^2 dx = 1 - \int_{\mathbb{R}^d} \sqrt{p(x)q(x)}\, dx. \tag{2.4}$$

Our main results about increasing models are as follows.

THEOREM 2.15 (S.2.2). *For an increasing model $\mathcal{P}(h)$, where h satisfies assumptions (I.1)–(I.3) and for the true density $h \circ g_0$ which satisfies assumptions (I.4)–(I.6), the sequence of MLEs $\{\hat{p}_n = h \circ \hat{g}_n\}$ is Hellinger consistent*: $H(\hat{p}_n, p_0) = H(h \circ \hat{g}_n, h \circ g_0) \to_{a.s.} 0$.

THEOREM 2.16. *For an increasing model $\mathcal{P}(h)$, where h satisfies assumptions (I.1)–(I.3), and for the true density $h \circ g_0$ which satisfies assumptions (I.4)–(I.6), the sequence of MLEs $\hat{g}_n$ is pointwise consistent. That is, $\hat{g}_n(x) \to_{a.s.} g_0(x)$ for $x \in \mathrm{ri}(\mathbb{R}_+^d)$ and convergence is uniform on compacta.*

The results about decreasing models can be formulated in a similar way.

THEOREM 2.17. *For a decreasing model $\mathcal{P}(h)$, where h satisfies assumptions (D.1)–(D.4), the sequence of MLEs $\{\hat{p}_n = h \circ \hat{g}_n\}$ is Hellinger consistent*:

$$H(\hat{p}_n, p_0) = H(h \circ \hat{g}_n, h \circ g_0) \to_{a.s.} 0.$$

THEOREM 2.18. *For a decreasing model $\mathcal{P}(h)$ with h satisfying assumptions (D.1)–(D.4), the sequence of MLEs $\hat{g}_n$ is pointwise consistent in the following sense. Define $g_0^* = g_0 + \delta(\cdot|\mathrm{ri}(\mathrm{dom}\, g_0))$. Then, $g_0^* = g_0$ a.e., $\hat{g}_n \to_{a.s.} g_0^*$ and the convergence is uniform on compacta. Moreover, if* $\mathrm{dom}\, g_0 = \mathbb{R}^d$, *then* $\|h \circ \hat{g}_n - h \circ g_0\|_\infty \to_{a.s.} 0$.

2.4. *Local asymptotic minimax lower bounds.* In this section, we establish local asymptotic minimax lower bounds for any estimator of several functionals of interest on the family $\mathcal{P}(h)$ of convex-transformed densities. We start with several general results following Jongbloed (2000) and then apply them to estimation at a fixed point and to mode estimation.

First, we define minimax risk as in Donoho and Liu (1991).

DEFINITION 2.19. Let $\mathcal{P}$ be a class of densities on $\mathbb{R}^d$ with respect to Lebesgue measure and let T be a functional $T : \mathcal{P} \to \mathbb{R}$. For an increasing convex loss function l on $\mathbb{R}_+$, we define the *minimax risk* as

$$R_l(n; T, \mathcal{P}) = \inf_{t_n} \sup_{p \in \mathcal{P}} \mathrm{E}_{p^{\times n}} l\big(|t_n(X_1, \ldots, X_n) - Tp|\big), \tag{2.5}$$

where t_n ranges over all possible estimators of Tp based on $X_1, \ldots, X_n$.

The main result (Theorem 1) in Jongbloed (2000) can be formulated as follows.

THEOREM 2.20 (Jongbloed). *Let $\{p_n\}$ be a sequence of densities in $\mathcal{P}$ such that* $\limsup_{n\to\infty} \sqrt{n} H(p_n, p) \le \tau$ *for some density p in $\mathcal{P}$. Then,*

$$\liminf_{n\to\infty} \frac{R_l(n; T, \{p, p_n\})}{l((1/4)e^{-2\tau^2}|T(p_n) - T(p)|)} \ge 1. \tag{2.6}$$

It will be convenient to reformulate this result in the following form.

COROLLARY 2.21. *Suppose that for any $\varepsilon > 0$ small enough, there exists $p_\varepsilon \in \mathcal{P}$ such that for some $r > 0$, $\lim_{\varepsilon\to 0} \varepsilon^{-1}|Tp_\varepsilon - Tp| = 1$ and*

$$\limsup_{\varepsilon\to 0} \varepsilon^{-r} H(p_\varepsilon, p) \le c.$$

There then exists a sequence $\{p_n\}$ such that

$$\liminf_{n\to\infty} n^{1/2r} R_1(n; T, \{p, p_n\}) \ge \frac{1}{4(2re)^{1/2r}} c^{-1/r}, \tag{2.7}$$

where R_1 is the risk which corresponds to $l(x) = |x|$.

Corollary 2.21 shows that for a fixed change in the value of the functional T, a family p_ε which is closer to the true density p with respect to Hellinger distance provides a sharper lower bound. This suggests that for the functional T which depends only on the local structure of the density, we would like our family $\{p_\varepsilon\}$ to deviate from p also locally. Below, we formally define such local deviations.

DEFINITION 2.22. We call a family of measurable functions $\{p_\varepsilon\}$ a *deformation* of a measurable function p if p_ε is defined for any $\varepsilon > 0$ small enough, $\lim_{\varepsilon\to 0} \operatorname{ess\,sup} |p - p_\varepsilon| = 0$ and there exists a bounded family of real numbers r_ε and a point x_0 such that

$$\mu[\operatorname{supp}|p_\varepsilon(x) - p(x)|] > 0, \qquad \operatorname{supp}|p_\varepsilon(x) - p(x)| \subseteq B(x_0, r_\varepsilon).$$

If, in addition, we have $\lim_{\varepsilon\to 0} r_\varepsilon = 0$, then we say that $\{p_\varepsilon\}$ is a *local deformation* at x_0.

Since, for a deformation p_ε, we have $\mu[\operatorname{supp}|p_\varepsilon(x) - p(x)|] > 0$ for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\mu\{x : |p_\varepsilon(x) - p(x)| > \delta\} > 0$ and thus the L_r-distance from p_ε to p is positive for all $\varepsilon > 0$. Note that this is always true if p and p_ε are continuous at x_0 and $p_\varepsilon(x_0) \neq p(x_0)$.

We can now state our lower bound for estimation of the convex-transformed density value at a fixed point x_0. This result relies on the properties of strongly convex functions, as described in Appendix S.A.4, and can be applied to both increasing and decreasing classes of convex-transformed densities.

THEOREM 2.23. *Let h be a monotone transformation, let $p = h \circ g \in \mathcal{P}(h)$ be a convex-transformed density and suppose that x_0 is a point in* $\operatorname{ri}(\operatorname{dom} g)$ *such that h is continuously differentiable at $g(x_0)$, $h \circ g(x_0) > 0$, $h' \circ g(x_0) \neq 0$ and* $\operatorname{curv}_{x_0} g > 0$. *Then, for the functional $T(h \circ g) \equiv g(x_0)$, there exists a sequence $\{p_n\} \subset \mathcal{P}(h)$ such that*

$$\liminf_{n\to\infty} n^{2/(d+4)} R_1(n; T, \{h \circ g, p_n\}) \geq C(d) \left[\frac{h \circ g(x_0)^2 \operatorname{curv}_{x_0} g}{h' \circ g(x_0)^4} \right]^{1/(d+4)}, \tag{2.8}$$

where the constant $C(d)$ depends only on the dimension d.

REMARK 2.24. If, in addition, g is twice continuously differentiable at x_0 and $\nabla^2 g(x_0)$ is positive definite, then, by Lemma S.A.22, we have $\operatorname{curv}_{x_0} g = \det(\nabla^2 g(x_0))$.

In Jongbloed (2000), lower bounds were constructed for functionals with values in $\mathbb{R}$. However, it is easy to see that the proof does not change for functionals with values in an arbitrary metric space (V, s) if, instead of $|Tp - Tp_n|$, we consider $s(Tp, Tp_n)$. We define

$$R_s(n; T, \mathcal{P}) = \inf_{t_n} \sup_{p\in\mathcal{P}} \mathrm{E}_{p^{\times n}} s(t_n(X_1, \ldots, X_n), Tp) \tag{2.9}$$

and the analog of Corollary 2.21 then has the following form.

COROLLARY 2.25. *Suppose that for any $\varepsilon > 0$ small enough, there exists $p_\varepsilon \in \mathcal{P}$ such that for some $r > 0$,*

$$\lim_{\varepsilon\to 0} \varepsilon^{-1} s(Tp_\varepsilon, Tp) = 1, \qquad \limsup_{\varepsilon\to 0} \varepsilon^{-r} H(p_\varepsilon, p) \le c.$$

There then exists a sequence $\{p_n\}$ such that

$$\liminf_{n\to\infty} n^{1/2r} R_s(n; T, \{p, p_n\}) \ge \frac{1}{4(2re)^{1/2r}} c^{-1/r}. \tag{2.10}$$

We now consider estimation of the functional $T(h \circ g) = \arg\min(g) \in \mathbb{R}^d$ for the density $p = h \circ g \in \mathcal{P}(h)$, assuming that the minimum is unique. This is equivalent to estimation of the mode of $p = h \circ g$.

The construction of a lower bound for the functional T is similar to the procedure we presented for estimation of $p = h \circ g$ at a fixed point x_0. Again, we use two opposite deformations: one is local and changes the functional value, the other is a convex combination with a fixed deformation and negligible-in-Hellinger-distance computation. However, in this case, the minimax rate also depends on the growth rate of g.

THEOREM 2.26 (S.3.4). *Let h be a decreasing transformation, $h \circ g \in \mathcal{P}(h)$ be a convex-transformed density and a point $x_0 \in \operatorname{ri}(\operatorname{dom} g)$ be a unique global minimum of g such that h is continuously differentiable at $g(x_0)$, $h' \circ g(x_0) \neq 0$ and $\operatorname{curv}_{x_0} g > 0$. In addition, let us assume that g is locally Hölder continuous at x_0, that is, $|g(x) - g(x_0)| \le L\|x - x_0\|^\gamma$ with respect to some norm $\|\cdot\|$. Then, for the functional $T(h \circ g) \equiv \arg\min g$, there exists a sequence $\{p_n\} \in \mathcal{P}(h)$ such that*

$$\begin{aligned} &\liminf_{n\to\infty} n^{2/(\gamma(d+4))} R_s(n; T, \{p, p_n\}) \\ &\qquad \ge C(d) L^{-1/\gamma} \left[\frac{h \circ g(x_0)^2 \operatorname{curv}_{x_0} g}{h' \circ g(x_0)^4}\right]^{1/(\gamma(d+4))}, \end{aligned} \tag{2.11}$$

where the constant $C(d)$ depends only on the dimension d, and the metric $s(x, y)$ is defined as $\|x - y\|$.

REMARK 2.27. If, in addition, g is twice continuously differentiable at x_0 and $\nabla^2 g(x_0)$ is positive definite, then, by Lemma S.A.22, we have $\operatorname{curv}_{x_0} g = \det(\nabla^2 g(x_0))$ and g is locally Hölder continuous at x_0 with exponent $\gamma = 2$ and any constant $L > \|\nabla^2 g(x_0)\|$.

REMARK 2.28. Since $\operatorname{curv}_{x_0} g > 0$, there exists a constant C such that $C\|x - x_0\|^2 \le |g(x) - g(x_0)|$ and thus we have $\gamma \in (0, 2]$.

2.5. *Conjectures concerning uniqueness of MLEs.* There exist counterexamples to uniqueness for nonconvex transformations h which satisfy assumptions (D.1)–(D.4). They suggest that uniqueness of the MLE does not depend on the tail behavior of the transformation h, but rather on the local properties of h in neighborhoods of the optimal values $\hat{g}_n(X_i)$. We conjecture that uniqueness holds for all monotone models if h is convex and $h/|h'|$ is nondecreasing convex. Further work on these uniqueness issues is needed.

2.6. *Conjectures about rates of convergence for the MLEs.* We conjecture that the (optimal) rate of convergence $n^{2/(d+4)}$ appearing in Theorem 2.23 for estimation of $f(x_0)$ will be achieved by the MLE only for $d = 2, 3$. For $d = 4$, we conjecture that the MLE will come within a factor $(\log n)^{-\gamma}$ (for some $\gamma > 0$) of achieving the rate $n^{1/4}$, but for $d > 4$, we conjecture that the rate of convergence will be the suboptimal rate $n^{1/d}$. This conjectured rate-suboptimality raises several interesting further issues:

- Can we find alternative estimators (perhaps via penalization or sieve methods) which achieve the optimal rates of convergence?
- For interesting subclasses, do maximum likelihood estimators remain rate-optimal?

3. Proofs.

3.1. *Preliminaries: Properties of decreasing transformations.*

LEMMA 3.1. *Let h be a decreasing transformation and g be a closed proper convex function such that $\int_{\mathbb{R}^d} h \circ g \, dx = C < \infty$. The following are then true:*

1. *for $y < +\infty$, the sublevel sets* $\operatorname{lev}_y g$ *are bounded and we have*

$$\mu[\operatorname{lev}_y g] \le C/h(y);$$

2. *the infimum of g is attained at some point $x \in \mathbb{R}^d$.*

PROOF. 1. We have

$$C = \int_{\mathbb{R}^d} h \circ g \, dx \ge \int_{\operatorname{lev}_y g} h \circ g \, dx \ge h(y)\mu[\operatorname{lev}_y g],$$

$$\mu[\operatorname{lev}_y g] \le C/h(y).$$

The sublevel set $\operatorname{lev}_y g$ has the same dimension as $\operatorname{dom} g$ [Theorem 7.6 in Rockafellar (1970)], which is d. By Lemma S.A.1, this set is bounded when $y < y_0$. Therefore, it is enough to prove that $\operatorname{lev}_{y_0} g$ is bounded for $y_0 < +\infty$.

Since $h \circ g$ is a density, we have $\inf g < y_0$. If g is constant on $\operatorname{dom} g$, then, for all $y \in [\inf g, +\infty)$, we have $\operatorname{lev}_y g = \operatorname{lev}_{\inf g} h$ and it is therefore bounded.

Otherwise, we can choose $\inf h \le y_1 < y_2 < y_0$. Then, $\mu[\mathrm{lev}_{y_2} g] < \infty$ and, by Lemma S.A.3, we have $\mu[\mathrm{lev}_{y_0} g] < \infty$. The argument above shows that $\mathrm{lev}_{y_0} g$ is also bounded.

2. This follows from the fact that g is continuous and $\mathrm{lev}_y g$ is bounded and nonempty for $y > \inf g$. $\square$

LEMMA 3.2. *Let h be a decreasing transformation, let g be a closed proper convex function on $\mathbb{R}^d$ and let Q be a σ-finite Borel measure on $\mathbb{R}^d$. Then*

$$\int_{(\mathrm{lev}_a g)^c} h \circ g \, dQ = -\int_a^{+\infty} h'(y) Q[\mathrm{lev}_y g \cap (\mathrm{lev}_a g)^c] \, dy.$$

PROOF. Using the Fubini–Tonelli theorem, we have, with $L_a^c \equiv (\mathrm{lev}_a g)^c$,

$$\begin{aligned}
\int_{L_a^c} h \circ g \, dQ &= \int_{L_a^c} \int_0^{h(a)} 1\{z \le h \circ g(x)\} \, dz \, dQ(x) \\
&= \int_{L_a^c} \int_0^{h(a)} 1\{h^{-1}(z) \ge g(x)\} \, dz \, dQ(x) \\
&= -\int_{L_a^c} \int_a^{\infty} h'(y) 1\{y \ge g(x)\} \, dy \, dQ(x) \\
&= -\int_a^{\infty} h'(y) \int_{L_a^c} 1\{y \ge g(x)\} \, dQ(x) \, dy \\
&= -\int_a^{\infty} h'(y) Q[\mathrm{lev}_y g \cap L_a^c] \, dy.
\end{aligned}$$

$\square$

LEMMA 3.3. *Let h be a decreasing transformation and let g be a closed proper convex function such that $\int_{\mathbb{R}^d} h \circ g \, dx < \infty$. Then, $\inf g > y_\infty$.*

PROOF. Since g is proper, the statement is trivial for $y_\infty = -\infty$, so we assume that $y_\infty > -\infty$. If, for x_0, we have $g(x_0) = y_\infty$, then there exists a ball $B \equiv B(x; r)$ such that $g < y_\infty + \varepsilon$ on B. Consider the convex function f defined as $f(x) = y_\infty + (\varepsilon/r)\|x - x_0\| + \delta(x \mid B)$. Then, by convexity, $f \ge g$ and $\int_{\mathbb{R}^d} h \circ g \, dx \ge \int_{\mathbb{R}^d} h \circ f \, dx$. We have $\mu[\mathrm{lev}_y f] = S(y - y_\infty)^d$ for $y \in [y_\infty, y_\infty + \varepsilon]$, where S is the Lebesgue measure of a unit ball $B(0; 1)$, and by Lemma 3.2, we can compute

$$\int_{\mathbb{R}^d} h \circ f \, dx = -S \int_{y_\infty}^{y_\infty + \varepsilon} h'(y)(y - y_\infty)^d \, dy.$$

The assumption (D.2) implies that $\int_{\mathbb{R}^d} h \circ g \, dx \ge \int_{\mathbb{R}^d} h \circ f \, dx = \infty$, which proves the statement. $\square$

LEMMA 3.4. *Let h be a decreasing transformation. Then, for any convex function g such that $h \circ g$ belongs to the decreasing model $\mathcal{P}(h)$, we have $\int_{\mathbb{R}^d} [h|\log h|] \circ g\, dx < \infty$.*

PROOF. By assumption (D.1), the function $-[h \log h](y)$ is decreasing to zero as $y \to +\infty$ and we have $0 < -[h \log h](y) < C y^{-d-\alpha'}$ for C large enough and $\alpha' \in (0, \alpha)$ as $y \to +\infty$.

By Lemma 3.1, the level sets $\operatorname{lev}_y g$ are bounded and since $h \circ g \in \mathcal{P}(h)$, we have $\inf g > y_\infty$. Therefore, the integral exists if and only if the integral

$$\int_{(\operatorname{lev}_a g)^c} [h \log h] \circ g\, dx > -\infty$$

for some $a > y_\infty$. Choosing a large enough and using Lemma 3.2 for the decreasing transformation $h_1(y) = y^{-d-\alpha'}$, we obtain

$$\begin{aligned}
0 &\geq \int_{(\operatorname{lev}_a g)^c} [h \log h] \circ g\, dx \\
&\geq -C \int_{(\operatorname{lev}_a g)^c} h_1 \circ g\, dx \geq C \int_a^{+\infty} h_1'(y) \mu[\operatorname{lev}_y g]\, dy \\
&= -C(d + \alpha') \int_a^{+\infty} y^{-d-\alpha'-1} \mu[\operatorname{lev}_y g]\, dy.
\end{aligned}$$

By Lemma S.A.3, we have $\mu(\operatorname{lev}_y g) = O(y^d)$ and therefore the last integral is finite. $\square$

LEMMA 3.5. *Let h be a decreasing transformation and suppose that $K \subset \mathbb{R}^d$ is a compact set. There then exists a closed proper convex function $g \in \mathcal{G}(h)$ such that $g < y_0$ on K.*

PROOF. Let B be a ball such that $K \subset B$. Let c be such that $h(c) = 1/\mu[B]$. The function $g \equiv c + \delta(\cdot \mid B)$ then belongs to $\mathcal{G}(h)$. $\square$

3.2. *Proofs for existence results.* Before giving proofs of Theorems 2.12 and 2.13, we establish two auxiliary lemmas. A set of points $x = \{x_i\}_{i=1}^n$ in $\mathbb{R}^d$ is in *general position* if, for any subset $x' \subseteq x$ of size $d + 1$, the Lebesgue measure of $\operatorname{conv}(x')$ is not zero. It follows from Okamoto (1973) that the observations X are in general position with probability 1 if $X_1, \dots, X_n$ are i.i.d. $p_0 \in \mathcal{P}(h)$. Thus, we may assume in the following that our observations are in general position for every n. For an increasing model, we also assume that all X_i belong to $\mathbb{R}_+^d$.

If an MLE for the model $\mathcal{P}(h)$ exists, then it maximizes the functional

$$\mathbb{L}_n g \equiv \int (\log h) \circ g\, d\mathbb{P}_n$$

over $g \in \mathcal{G}(h)$, where the last integral is over $\overline{\mathbb{R}}_+^d$ for increasing h and over $\mathbb{R}^d$ for decreasing models. The theorem below determines the form of the MLE for an increasing model. We write $\mathrm{ev}_x f = (f(x_1), \dots, f(x_n))$, $x = (x_1, \dots, x_n)$ with $x_i \in \mathbb{R}^d$.

LEMMA 3.6 (S.1.7). *Consider an increasing transformation h. For any convex function g with $\operatorname{dom} g = \overline{\mathbb{R}}_+^d$ such that $\int_{\overline{\mathbb{R}}_+^d} h \circ g \, dx \le 1$ and $\mathbb{L}_n g > -\infty$, there exists $\tilde{g} \in \mathcal{G}(h)$ such that $\tilde{g} \ge g$ and $\mathbb{L}_n \tilde{g} \ge \mathbb{L}_n g$. The function $\tilde{g}$ can be chosen as a minimal element in $\mathrm{ev}_X^{-1} \tilde{p}$, where $\tilde{p} = \mathrm{ev}_X \tilde{g}$.*

THEOREM 3.7 (S.1.8). *If an MLE $\hat{g}_0$ exists for the increasing model $\mathcal{P}(h)$, then there exists an MLE $\hat{g}_1$ which is a minimal element in $\mathrm{ev}_X^{-1} q$, where $q = \mathrm{ev}_X \hat{g}_0$. In other words, $\hat{g}_1$ is a polyhedral convex function such that $\operatorname{dom} g_1 = \overline{\mathbb{R}}_+^d$ and the interior of each facet contains at least one element of X. If h is strictly increasing on $[y_0, y_\infty]$, then $\hat{g}_0(x) = \hat{g}_1(x)$ for all x such that $\hat{g}_0(x) > y_0$ and thus defines the same density from $\mathcal{P}(h)$.*

Here are the corresponding results for decreasing transformations h.

LEMMA 3.8 (S.1.9). *Consider a decreasing transformation h. For any convex function g such that $\int_{\mathbb{R}^d} h \circ g \, dx \le 1$ and $\mathbb{L}_n g > -\infty$, there exists $\tilde{g} \in \mathcal{G}(h)$ such that $\tilde{g} \le g$ and $\mathbb{L}_n \tilde{g} \ge \mathbb{L}_n g$. The function $\tilde{g}$ can be chosen as the maximal element in $\mathrm{ev}_X^{-1} \tilde{q}$, where $\tilde{q} = \mathrm{ev}_X \tilde{g}$.*

THEOREM 3.9 (S.1.10). *If the MLE $\hat{g}_0$ exists for the decreasing model $\mathcal{P}(h)$, then there exists another MLE $\hat{g}_1$ which is the maximal element in $\mathrm{ev}_X^{-1} q$, where $q = \mathrm{ev}_X \hat{g}_0$. In other words, $\hat{g}_1$ is a polyhedral convex function with the set of knots $K_n \subseteq X$ and domain $\operatorname{dom} \hat{g}_1 = \operatorname{conv}(X)$. If h is strictly decreasing on $[y_\infty, y_0]$, then $\hat{g}_0(x) = \hat{g}_1(x)$.*

The bounds provided by the following key lemma are the remaining preparatory work for proving existence of the MLE in the case of increasing transformations.

For an increasing model $\mathcal{P}(h)$, let us denote by $\mathcal{N}(h, X, \varepsilon)$, for $\varepsilon > -\infty$, the family of all convex functions $g \in \mathcal{G}(h)$ such that g is a minimal element in $\mathrm{ev}_X^{-1} q$, where $q = \mathrm{ev}_X g$ and $\mathbb{L}_n g \ge \varepsilon$. By Lemma S.1.5, the family $\mathcal{N}(h, X, \varepsilon)$ is not empty for $\varepsilon > -\infty$ small enough. By construction, for $g \in \mathcal{N}(h, X, \varepsilon)$, we have $g(X_i) > y_0$ for $X_i \in X$.

LEMMA 3.10. *There exist constants $c(x, X, \varepsilon)$ and $C(x, X, \varepsilon) < y_\infty$ which depend only on $x \in \overline{\mathbb{R}}_+^d$, the observations X and ε, such that for any $g \in \mathcal{N}(h, X, \varepsilon)$, we have*

$$c(x, X, \varepsilon) \le g(x) \le C(x, X, \varepsilon).$$

PROOF. By Lemma S.1.1, we have $h \circ g(X_i) \le \frac{d!}{d^d V(X_i)}$, which gives the upper bounds $C(X_i, X, \varepsilon)$. By assumption, we have

$$\big(\max h \circ g(X_i)\big)^{n-1} \min h \circ g(X_i) \ge \prod h \circ g(X_i) \ge e^{n\varepsilon}$$

and therefore

$$\min h \circ g(X_i) \ge \frac{e^{n\varepsilon}}{h(\max C(X_i, X, \varepsilon))^{n-1}},$$

which gives the uniform lower bound $c(X_i, X, \varepsilon)$ for all $X_i \in X$. Since, by Lemma S.1.1, $g(0) \ge g(X_i)$, we also obtain $c(0, X, \varepsilon)$.

We now prove that there exist $C(0, X, \varepsilon)$. Let l be a linear function which defines any facet of g for which 0 is an element. By Lemma S.A.15, there exists $X_a \in X$ which belongs to this facet. Then, $g(0) = l(0)$ and $g(X_a) = l(X_a)$. Let us denote by S the simplex $\{l = l(X_a)\} \cap \overline{\mathbb{R}}^d_+$, by S^* the simplex $\{l \ge l(X_a)\} \cap \overline{\mathbb{R}}^d_+$ and by l' the linear function which is equal to $c \equiv \min c(X_i, X, \varepsilon)$ on S and to $g(0)$ at 0. By the inequality of arithmetic and geometric means (as in the proof of Lemma S.1.1), we have $\mu[S^*] \ge \frac{d^d V(X_a)}{d!}$. We also have, for $l \ge l'$, $1 = \int_{\overline{\mathbb{R}}^d_+} h \circ g \, dx \ge \int_{S^*} h \circ l' \, dx$. By Lemma S.1.2,

$$\begin{aligned}\int_{S^*} h \circ l' \, dx &= \mu[S^*] \int_c^{g(0)} h'(y) \left(\frac{g(0) - y}{g(0) - c}\right)^d dy \\ &\ge \frac{d^d V(X_a)}{d!} \int_c^{y_\infty} h'(y) 1\{y \le g(0)\} \left(\frac{g(0) - y}{g(0) - c}\right)^d dy.\end{aligned}$$

Consider the function $T(s)$ defined as

$$T(s) = \frac{d^d V(X_a)}{d!} \int_c^{y_\infty} h'(y) 1\{y \le s\} \left(\frac{s - y}{s - c}\right)^d dy.$$

If $y_\infty = +\infty$, then, for a fixed $y \in (c, +\infty)$, we have

$$h'(y) 1\{y \le s\} \left(\frac{s - y}{s - c}\right)^d \uparrow h'(y) \qquad \text{as } s \to y_\infty$$

and, by monotone convergence, we have

$$T(s) \uparrow \int_c^{y_\infty} h'(y) \, dy = +\infty \qquad \text{as } s \to y_\infty.$$

If $y_\infty < +\infty$, then for a fixed $y \in (c, y_\infty]$, we have

$$h'(y) 1\{y \le s\} \left(\frac{s - y}{s - c}\right)^d \uparrow h'(y) \left(\frac{y_\infty - y}{y_\infty - c}\right)^d \qquad \text{as } s \to y_\infty$$

and, by monotone convergence, we have

$$T(s) \uparrow \int_c^{y_\infty} h'(y) \left(\frac{y_\infty - y}{y_\infty - c}\right)^d dy = +\infty \qquad \text{as } s \to y_\infty,$$

by assumption (I.2).

Thus, there exists $s_0 \in (c, y_\infty)$ such that $T(s_0) > 1$. This implies that $g(0) < s_0$. Since s_0 depends only on X_a and $\min c(X_i, X, \varepsilon)$, this gives an upper bound $C(0, X, \varepsilon)$.

By Lemma S.1.1, for any $x_0 \in \overline{\mathbb{R}}^d_+$, we can set $C(x_0, X, \varepsilon) = C(0, X, \varepsilon)$. Let $l(x) = a^T x + l(0)$ be a linear function which defines the facet of g to which x belongs. By Lemma S.A.15, there exists $X_a \in X$ which belongs to this facet and thus $l(X_a) = g(X_a)$. By Lemma S.1.1, we have $a_k < 0$ for all k and, by definition, $l(0) \le g(0)$. We have

$$c(X_a, X, \varepsilon) \le g(X_a) = l(X_a) = a^T X_a + l(0) \le a^T X_a + g(0),$$

therefore

$$a_k \ge \frac{c(X_a, X, \varepsilon) - C(0, X, \varepsilon)}{(X_a)_k} \quad \text{and} \quad l(0) \ge c(X_a, X, \varepsilon).$$

Now,

$$g(x_0) = l(x_0) \ge \frac{c(X_a, X, \varepsilon) - C(0, X, \varepsilon)}{(X_a)_k}(x_0)_k + c(X_a, X, \varepsilon).$$

Since we have only a finite number of possible choices for X_a, we have obtained $c(x_0, X, \varepsilon)$, which completes the proof. □

We are now ready for the proof of Theorem 2.12.

PROOF OF THEOREM 2.12. By Lemma S.1.5, there exists ε small enough such that the family $\mathcal{N}(h, X, \varepsilon)$ is not empty. Clearly, we can restrict MLE candidates $\hat{g}$ to functions in the family $\mathcal{N}(h, X, \varepsilon)$. The set $N = \mathrm{ev}_X \mathcal{N}(h, X, \varepsilon)$ is bounded, by Lemma 3.10. Let us denote by q^* a point in the closure $\bar{N}$ of N which maximizes the continuous function

$$L_n(q) = \frac{1}{n} \sum_{i=1}^n \log h(q_i).$$

Since $q^* \in \bar{N}$, there exists a sequence of functions $g_k \in \mathcal{N}(h, X, \varepsilon)$ such that $\mathrm{ev}_X g_k$ converges to q^*. By Theorem 10.9 in Rockafellar (1970) and Lemma 3.10, there exists a finite convex function g^* on $\overline{\mathbb{R}}^d_+$ such that some subsequence g_l converges pointwise to g^*. Therefore, we have $\mathrm{ev}_X g^* = q^*$. Since $X \subset \mathbb{R}^d_+$, we can assume that g^* is closed. By Fatou's lemma, we have $\int_{\overline{\mathbb{R}}^d_+} h \circ g^* \, dx \le 1$. By Lemma 3.6, there exists $g \in \mathcal{G}(h)$ such that $g \ge g^*$ and $\mathbb{L}_n g \ge \mathbb{L}_n g^* = L_n(q^*)$. By assumption, this implies that $\mathbb{L}_n g = \mathbb{L}_n g^*$. Hence, g is the MLE. Finally, we have to add the "almost surely" clause since we have assumed that the points X_i belong to $\mathbb{R}^d_+$. □

Before proving existence of the MLE for a decreasing transformation family, we need two lemmas.

LEMMA 3.11 (S.1.11). *Consider a decreasing model $\mathcal{P}(h)$. Let $\{g_k\}$ be a sequence of convex functions from $\mathcal{G}(h)$ and let $\{n_k\}$ be a nondecreasing sequence of positive integers $n_k \geq n_d$ such that for some $\varepsilon > -\infty$ and $\rho > 0$, the following is true*:

1. $\mathbb{L}_{n_k} g_k \geq \varepsilon$;
2. *if* $\mu[\mathrm{lev}_{a_k} g_k] = \rho$ *for some* a_k, *then* $\mathbb{P}_{n_k}[\mathrm{lev}_{a_k} g_k] < d/n_d$.

There then exists $m > y_\infty$ such that $g_k \geq m$ for all k.

For a decreasing model $\mathcal{P}(h)$, let us denote by $\mathcal{N}(h, X, \varepsilon)$ for $\varepsilon > -\infty$ the family of all convex functions $g \in \mathcal{G}(h)$ such that g is a maximal element in $\mathrm{ev}_X^{-1} q$, where $q = \mathrm{ev}_X g$ and $\mathbb{L}_n g \geq \varepsilon$. By Lemma 3.5, the family $\mathcal{N}(h, X, \varepsilon)$ is not empty for $\varepsilon > -\infty$ small enough. By construction, for $g \in \mathcal{N}(h, X, \varepsilon)$, we have $g(X_i) < y_0$ for $X_i \in X$.

LEMMA 3.12. *For given observations $X = (X_1, \dots, X_n)$ such that $n \geq n_d$, there exist constants $m > y_\infty$ and M which depend only on observations X and ε such that for any $g \in \mathcal{N}(h, X, \varepsilon)$, we have $m \leq g(x) \leq M$ on* $\mathrm{conv}(X)$.

PROOF. Since, by assumption, the points X are in general position, there exists $\rho > 0$ such that for any d-dimensional simplex S with vertices from X, we have $\mu[S] \geq \rho$. Then, any convex set $C \subseteq \mathrm{conv}(X)$ such that $\mu[C] = \rho$ cannot contain more than d points from X. Therefore, we have $\mathbb{P}_n[C] \leq d/n \leq d/n_d$.

An arbitrary sequence of functions $\{g_k\}$ from $\mathcal{N}(h, X, \varepsilon)$ satisfies the conditions of Lemma 3.11 with $n_k \equiv n$ and the same ε and ρ constructed above. Therefore, the sequence $\{g_k\}$ is bounded below by some constant greater than y_∞. Thus, the family of functions $\mathcal{N}(h, X, \varepsilon)$ is uniformly bounded below by some $m > y_\infty$.

Consider any $g \in \mathcal{N}(h, X, \varepsilon)$. Let M_g be the supremum of g on $\mathrm{dom}\, h$. By Theorem 32.2 in Rockafellar (1970), the supremum is obtained at some $X_M \in X$ and therefore $M_g < y_0$. Let m_g be the minimum of g on X. We have $h(m_g)^{n-1} h(M_g) \geq e^{n\varepsilon}$ and

$$h(M_g) \geq \frac{e^{n\varepsilon}}{h(m_g)^{n-1}} \geq \frac{e^{n\varepsilon}}{h(m)^{n-1}}.$$

Thus, we have obtained an upper bound M which depends only on m and X. □

We are now ready for the proof of Theorem 2.13.

PROOF OF THEOREM 2.13. By Lemma 3.5, there exists ε small enough such that the family $\mathcal{N}(h, X, \varepsilon)$ is not empty. Clearly, we can restrict MLE candidates to the functions in the family $\mathcal{N}(h, X, \varepsilon)$. The set $N = \mathrm{ev}_X \mathcal{N}(h, X, \varepsilon)$ is bounded, by

Lemma 3.12. Let us denote by q^* a point in the closure $\bar{N}$ of N which maximizes the continuous function

$$L_n(q) = \frac{1}{n}\sum_{i=1}^{n} \log h(q_i).$$

Since $q^* \in \bar{N}$, there exists a sequence of functions $g_k \in \mathcal{N}(h, X, \varepsilon)$ such that $\mathrm{ev}_X g_k$ converges to q^*. By Lemma 3.12, the functions $f_k = \sup_{l \geq k} g_l$ are finite convex functions on $\mathrm{conv}(X)$ and the sequence $\{f_k(x)\}$ is monotone decreasing for each $x \in \mathrm{conv}(X)$ and bounded below. Therefore, $f_k \downarrow g^*$ for some convex function g^* and, by construction, $\mathrm{ev}_X g^* = q^*$. We have

$$\int_{\mathbb{R}^d} h \circ f_k \, dx \leq \int_{\mathbb{R}^d} h \circ g_k \, dx = 1$$

and thus, by Fatou's lemma, $\int_{\mathbb{R}^d} h \circ g^* \, dx \leq 1$. By Lemma 3.8, there exists $g \in \mathcal{G}(h)$ such that $g \leq g^*$ and $\mathbb{L}_n g \geq \mathbb{L}_n g^* = L_n(q^*)$. By assumption, this implies that $\mathbb{L}_n g = \mathbb{L}_n g^*$. Thus, the function g is the MLE.

Finally, we have to add the "almost surely" clause since we assumed that the points X_i are in general position. □

3.3. *Proofs for consistency results*. We begin with proofs for some technical results which we will use in the consistency arguments for both increasing and decreasing models. The main argument for proving Hellinger consistency proceeds along the lines of the proof given in the case of $d = 1$ by Pal, Woodroofe and Meyer (2007) and in the log-concave case for $d > 1$ by Schuhmacher and Duembgen (2010).

LEMMA 3.13 (S.1.12). *Consider a monotone model $\mathcal{P}(h)$. Suppose that the true density $h \circ g_0$ and the sequence of MLEs $\{\hat{g}_n\}$ have the following properties*:

$$\int (h|\log h|) \circ g_0(x) \, dx < \infty$$

and

$$\int \log[\varepsilon + h \circ \hat{g}_n(x)] \, d\big(\mathbb{P}_n(x) - P_0(x)\big) \to_{a.s.} 0$$

for $\varepsilon > 0$ small enough. The sequence of the MLEs is then Hellinger consistent: $H(h \circ \hat{g}_n, h \circ g_0) \to_{a.s.} 0$.

The next lemma allows us to obtain pointwise consistency once Hellinger consistency is proved.

LEMMA 3.14. *Suppose that, for a monotone model $\mathcal{P}(h)$, a sequence of MLEs $\hat{g}_n$ is Hellinger consistent. The sequence $\hat{g}_n$ is then pointwise consistent. In other words, $\hat{g}_n(x) \to_{a.s.} g_0(x)$ for $x \in \mathrm{ri}(\mathrm{dom}\, g_0)$ and convergence is uniform on compacta.*

PROOF. Let us denote by L_a^0 and L_a^k the sublevel sets $L_a^0 = \mathrm{lev}_a\, g_0$ and $L_a^n = \mathrm{lev}_a\, \hat{g}_n$, respectively. Consider Ω_0 such that $\Pr[\Omega_0] = 1$ and $H^2(h \circ \hat{g}_n^\omega, h \circ g_0) \to 0$, where $\hat{g}_n^\omega$ is the MLE for $\omega \in \Omega_0$. For all $\omega \in \Omega_0$, we have

$$\begin{aligned}\int [\sqrt{h} \circ g_0 - \sqrt{h} \circ \hat{g}_n]^2\, dx &\geq \int_{L_a^0 \setminus L_{a+\varepsilon}^n} [\sqrt{h} \circ g_0 - \sqrt{h} \circ \hat{g}_n]^2\, dx \\ &\geq (\sqrt{h}(a) - \sqrt{h}(a+\varepsilon))^2 \mu(L_a^0 \setminus L_{a+\varepsilon}^n) \\ &\to 0\end{aligned}$$

and, by Lemma S.A.2, we have $\liminf \mathrm{ri}(L_a^0 \cap L_{a+\varepsilon}^n) = \mathrm{ri}(L_a^0)$. Therefore, $\limsup \hat{g}_n(x) < a + \varepsilon$ for $x \in \mathrm{ri}(L_a^0)$. Since a and ε are arbitrary, we have $\limsup \hat{g}_n \leq g_0$ on $\mathrm{ri}(\mathrm{dom}\, g_0)$.

On the other hand, we have

$$\begin{aligned}\int [\sqrt{h} \circ g_0 - \sqrt{h} \circ \hat{g}_n]^2\, dx &\geq \int_{L_{a-\varepsilon}^n \setminus L_a^0} [\sqrt{h} \circ g_0 - \sqrt{h} \circ \hat{g}_n]^2\, dx \\ &\geq (\sqrt{h}(a-\varepsilon) - \sqrt{h}(a))^2 \mu(L_{a-\varepsilon}^n \setminus L_a^0) \\ &\to 0\end{aligned}$$

and by Lemma S.A.2, we have $\limsup \mathrm{cl}(L_{a-\varepsilon}^n \cup L_a^0) = \mathrm{cl}(L_a^0)$. Therefore, $\liminf \hat{g}_n(x) > a - \varepsilon$ for x such that $g_0(x) \geq a$. Since a and ε are arbitrary, we have $\liminf \hat{g}_n \geq g_0$ on $\mathrm{dom}\, g_0$.

Thus, $\hat{g}_n \to g_0$ almost surely on $\mathrm{ri}(\mathrm{dom}\, g_0)$. By Theorem 10.8 in Rockafellar (1970), convergence is uniform on compacta $K \subset \mathrm{ri}(\mathbb{R}_+^d)$. $\square$

We need a general property of the bracketing entropy numbers.

LEMMA 3.15 (S.1.13). *Let $\mathcal{A}$ be a class of sets in $\mathbb{R}^d$ such that class $\mathcal{A} \cap [-a, a]^d$ has finite bracketing entropy with respect to Lebesgue measure μ for any a large enough*: $\log N_{[]}(\varepsilon, \mathcal{A} \cap [-a, a]^d, L_1(\mu)) < +\infty$ *for every* $\varepsilon > 0$. *Then, for any Lebesgue absolutely continuous probability measure P with bounded density, we have that $\mathcal{A}$ is a Glivenko–Cantelli class*: $\|\mathbb{P}_n - P\|_{\mathcal{A}} \to_{a.s.} 0$.

By Lemma S.1.1, we have $\mathrm{ri}(\mathbb{R}_+^d) \subseteq \mathrm{dom}\, g_0$. Thus, Theorem 2.15 and Lemma 3.14 imply Theorem 2.16.

Finally, we prove consistency for decreasing models. We need a general property of convex sets.

LEMMA 3.16. *Let $\mathcal{A}$ be the class of closed convex sets A in $\mathbb{R}^d$ and let P be a Lebesgue absolutely continuous probability measure with bounded density. Then*, $\|\mathbb{P}_n - P\|_{\mathcal{A}} \to_{a.s.} 0$.

PROOF. Let D be a convex compact set. By Theorem 8.4.2 in Dudley (1999), the class $\mathcal{A} \cap D$ has a finite set of ε-brackets. Since the class $\mathcal{A}$ is invariant under rescaling, the result follows from Lemma 3.15. □

LEMMA 3.17. *For a decreasing model $\mathcal{P}(h)$, the sequence of MLEs $\hat{g}_n$ is almost surely uniformly bounded below.*

PROOF. We will apply Lemma 3.11 to the sequences $\hat{g}_n$ and $\{n\}$. By the strong law of large numbers and Lemma 3.4, we have

$$\mathbb{L}_n \hat{g}_n \geq \mathbb{L}_n g_0 \rightarrow_{\text{a.s.}} \int [h \log h] \circ g_0 \, dx > -\infty.$$

Therefore, the sequence $\{\mathbb{L}_n \hat{g}_n\}$ is bounded away from $-\infty$ and the first condition of Lemma 3.11 is true.

Choose some $a \in (0, d/n_d)$. Then, for any set S such that $\mu[S] = \rho \equiv a/h(\min g_0)$, where $\min g_0$ is attained by Lemma 3.1, we have

$$P[S] = \int_S h \circ g_0 \, dx \leq \mu[S] h(\min g_0) = a < d/n_d.$$

Now, let $A_n = \operatorname{lev}_{a_n} \hat{g}_n$ be sets such that $\mu[A_n] = \rho$. Then, by Lemma 3.16, we have

$$|\mathbb{P}_n[A_n] - P[A_n]| \leq \|\mathbb{P}_n - P\|_{\mathcal{A}} \rightarrow_{\text{a.s.}} 0,$$

which implies that $\mathbb{P}_n[A_n] < d/n_d$ almost surely for n large enough. Therefore, the second condition of Lemma 3.11 is true and is applicable to the sequence $\hat{g}_n$ almost surely. □

PROOF OF THEOREM 2.17. By Lemmas 3.4 and 3.13, it is enough to show that

$$\int \log[\varepsilon + h \circ \hat{g}_n(x)] \, d\big(\mathbb{P}_n(x) - P_0(x)\big) \rightarrow_{\text{a.s.}} 0.$$

By Lemma 3.17, we have $\inf \hat{g}_n \geq A$ for some $A > y_\infty$. Therefore, by Lemma 3.2 applied to the decreasing transformation $\log[\varepsilon + h(y)] - \log \varepsilon$, it follows that

$$\begin{aligned}
&\int \log[\varepsilon + h \circ \hat{g}_n(x)] \, d\big(\mathbb{P}_n(x) - P_0(x)\big) \\
&\quad = \int_A^{+\infty} \left[\frac{-h'(z)}{\varepsilon + h(z)}\right] (\mathbb{P}_n - P_0)(\operatorname{lev}_z \hat{g}_n) \, dz \\
&\quad \leq \|\mathbb{P}_n - P_0\|_{\mathcal{A}} \int_A^{+\infty} \left[\frac{-h'(z)}{\varepsilon + h(z)}\right] dz \\
&\quad = \|\mathbb{P}_n - P_0\|_{\mathcal{A}} \log\left[\frac{\varepsilon + h(A)}{\varepsilon}\right] \rightarrow_{\text{a.s.}} 0,
\end{aligned}$$

where the last limit follows from Lemma 3.16. □

PROOF OF THEOREM 2.18. By Lemma 3.14, we have $\hat{g}_n \to g_0$ almost surely on $\operatorname{ri}(\operatorname{dom} g_0)$. Functions g_0 and g_0^* differ only on the boundary $\partial \operatorname{dom} g_0$, which has Lebesgue measure zero, by Lemma S.A.1. Since observations $X_i \in \operatorname{ri}(\operatorname{dom} g_0)$ almost surely, we have $\hat{g}_n = +\infty$ on $\partial \operatorname{dom} g_0$ and thus $\hat{g}_n \to g_0^*$.

Now, we assume that $\operatorname{dom} g_0 = \mathbb{R}^d$. By Lemma 3.1, the function g_0 has bounded sublevel sets and therefore there exists x_0 where g_0 attains its minimum m. Since $h \circ g_0$ is density, we have $h(m) > 0$ and by Lemma 3.3, we have $h(m) < \infty$. Fix $\varepsilon > 0$ such that $h(m) > 3\varepsilon$ and consider a such that $h(a) < \varepsilon$. The set $A = \operatorname{lev}_a g_0$ is bounded and, by continuity, $g_0 = a$ on ∂A. Choose $\delta > 0$ such that $h(a - \delta) < 2\varepsilon < h(m + \delta)$ and

$$\sup_{x \in [m, a+\delta]} |h(x) - h(x - \delta)| \le \varepsilon.$$

The closure $\bar{A}$ is compact and thus, for n large enough, we have, with probability one, $\sup_{\bar{A}} |\hat{g}_n - g_0| < \delta$, which implies that $\sup_{\bar{A}} |h \circ \hat{g}_n - h \circ g_0| < \varepsilon$ since the range of values of g_0 on $\bar{A}$ is $[m, a]$. The set ∂A is compact and therefore $\hat{g}_n$ attains its minimum m_n on this set at some point x_n. By construction,

$$m_n = \hat{g}_n(x_n) > g_0(x_n) - \delta = a - \delta > m + \delta = g_0(x_0) + \delta > \hat{g}_n(x_0).$$

We have $x_0 \in A \cap \operatorname{lev}_{a-\delta} \hat{g}_n$ and $\hat{g}_n \ge m_n > a - \delta$ on ∂A. Thus, by convexity, we have $\operatorname{lev}_{a-\delta} \hat{g}_n \subset A$ and for $x \notin \bar{A}$, we have

$$|h \circ \hat{g}_n(x) - h \circ g_0(x)| \le h \circ \hat{g}_n(x) + h \circ g_0(x) < h(a - \delta) + h(a) < 3\varepsilon.$$

This shows that for any $\varepsilon > 0$ small enough, we will have

$$\|h \circ \hat{g}_n - h \circ g_0\|_\infty < 3\varepsilon$$

with probability one as $n \to \infty$. This concludes the proof. □

3.4. *Proofs for lower bound results.* We will use the following lemma for computing the Hellinger distance between a function and its local deformation.

LEMMA 3.18 (S.3.1). *Let $\{g_\varepsilon\}$ be a local deformation of the function $g : \mathbb{R}^d \to \mathbb{R}$ at the point x_0 such that g is continuous at x_0 and let the function $h : \mathbb{R} \to \mathbb{R}$ be continuously differentiable at the point $g(x_0)$. Then, for any $r > 0$,*

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} |g_\varepsilon(x) - g(x)|^r \, dx = 0, \tag{3.1}$$

$$\lim_{\varepsilon \to 0} \frac{\int_{\mathbb{R}^d} |h \circ g_\varepsilon(x) - h \circ g(x)|^r \, dx}{\int_{\mathbb{R}^d} |g_\varepsilon(x) - g(x)|^r \, dx} = |h' \circ g(x_0)|^r. \tag{3.2}$$

In order to apply Corollary 2.21, we need to construct deformations so that they still belong to the class $\mathcal{G}$. The following lemma provides a technique for constructing such deformations.

LEMMA 3.19 (S.3.2). *Let $\{g_\varepsilon\}$ be a local deformation of the function $g:\mathbb{R}^d \to \mathbb{R}$ at the point x_0 such that g is continuous at x_0 and let the function $h:\mathbb{R}\to\mathbb{R}$ be continuously differentiable at the point $g(x_0)$ so that $h' \circ g(x_0) \neq 0$. Then, for any fixed $\delta > 0$ small enough, the deformation $g_{\theta,\delta} = \theta g_\delta + (1-\theta)g$ and any $r > 0$, we have*

$$\limsup_{\theta\to 0} \theta^{-r} \int_{\mathbb{R}^d} |h \circ g_{\theta,\delta}(x) - h \circ g(x)|^r \, dx < \infty, \tag{3.3}$$

$$\liminf_{\theta\to 0} \theta^{-r} \int_{\mathbb{R}^d} |h \circ g_{\theta,\delta}(x) - h \circ g(x)|^r \, dx > 0. \tag{3.4}$$

Note that $g_{\theta,\delta}$ is not a local deformation of g.

PROOF OF THEOREM 2.23. Our statement is nontrivial only if the curvature $\operatorname{curv}_{x_0} g > 0$ or, equivalently, there exists a positive definite $d \times d$ matrix G such that the function g is locally G-strongly convex. Then, by Lemma S.A.17, this means that there exists a convex function q such that, in some neighborhood $O(x_0)$ of x_0, we have

$$g(x) = \tfrac{1}{2}(x - x_0)^T G(x - x_0) + q(x). \tag{3.5}$$

The plan of the proof is as follows: we introduce families of functions $\{D_\varepsilon(g; x_0, v)\}$ and $\{D^*_\varepsilon(g; x_0)\}$ and prove that these families are local deformations. Using these deformations as building blocks, we construct two types of deformations, $\{h \circ g^+_\varepsilon\}$ and $\{h \circ g^-_\varepsilon\}$, of the density $h \circ g$, which belong to $\mathcal{P}(h)$. These deformations represent positive and negative changes in the value of the function g at the point x_0. We then approximate the Hellinger distances using Lemma 3.18. Finally, applying Corollary 2.21, we obtain lower bounds which depend on G. We complete the proof by taking the supremum of the obtained lower bounds over all $G \in \mathcal{SC}(g; x_0)$. Under the mild assumption of strong convexity of the function g, both deformations give the same rate and structure of the constant $C(d)$. However, it is possible to obtain a larger constant $C(d)$ for the negative deformation if we assume that g is twice differentiable. Note that, by the definition of $\mathcal{P}(h)$, the function g is a closed proper convex function.

Let us define a function $D_\varepsilon(g; x_0, v_0)$ for a given $\varepsilon > 0$, $x_0 \in \operatorname{dom} g$ and $v_0 \in \partial g(x_0)$ as follows: $D_\varepsilon(g; x_0, v_0)(x) = \max(g(x), l_0(x) + \varepsilon)$, where $l_0(x) = \langle v_0, x - x_0\rangle + g(x_0)$ is a support plane to g at x_0 (see Figure 1). Since $l_0 + \varepsilon$ is a support plane to $g + \varepsilon$, we have $g \le D_\varepsilon(g; x_0, v_0) \le g + \varepsilon$ and thus $\operatorname{dom} D_\varepsilon(g; x_0, v_0) = \operatorname{dom} g$. As a maximum of two closed convex functions, $D_\varepsilon(g; x_0, v_0)$ is a closed convex function. For a given x_1, we have $D_\varepsilon(g; x_0, v_0)(x_1) = g(x_1)$ if and only if

$$g(x_1) - \varepsilon \ge \langle v_0, x_1 - x_0\rangle + g(x_0). \tag{3.6}$$

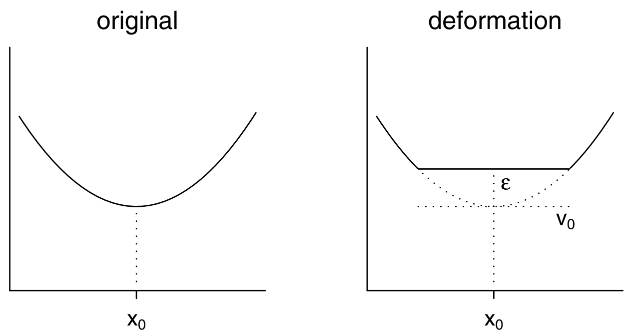

FIG. 1. *Example of the deformation $D_\varepsilon(g; x_0, v_0)$.*

We also define a function $D_\varepsilon^*(g; x_0)$ for a given $\varepsilon > 0$ and $x_0 \in \operatorname{dom} g$ as a maximal convex minorant (Appendix S.A.1) of the function $\tilde{g}_\varepsilon$, defined as

$$\tilde{g}_\varepsilon(x) = g(x)1_{\{x_0\}^c}(x) + \big(g(x_0) - \varepsilon\big)1_{\{x_0\}}(x),$$

see Figure 2. Both functions $D_\varepsilon(g; x_0, v_0)$ and $D_\varepsilon^*(g; x_0)$ are convex by construction and, as the next lemma shows, have similar properties. However, the argument for $D_\varepsilon^*(g; x_0)$ is more complicated.

LEMMA 3.20. *Let g be a closed proper convex function, g^* its convex conjugate and $x_0 \in \operatorname{ri}(\operatorname{dom} g)$. Then*:

1. *$D_\varepsilon^*(g; x_0)$ is a closed proper convex function such that $g - \varepsilon \le D_\varepsilon^*(g; x_0) \le g$ and $\operatorname{dom} D_\varepsilon^*(g; x_0) = \operatorname{dom} g$;*
2. *for a given $x_1 \in \operatorname{ri}(\operatorname{dom} g)$, we have $D_\varepsilon^*(g; x_0)(x_1) = g(x_1)$ if and only if there exists $v \in \partial g(x_1)$ such that*

$$g(x_1) + \varepsilon \le \langle v, x_1 - x_0 \rangle + g(x_0); \tag{3.7}$$

3. *if $v_0 \in \partial g(x_0)$, then $x_0 \in \partial g^*(v_0)$ and $D_\varepsilon(g; x_0, v_0) = (D_\varepsilon^*(g^*; v_0))^*$.*

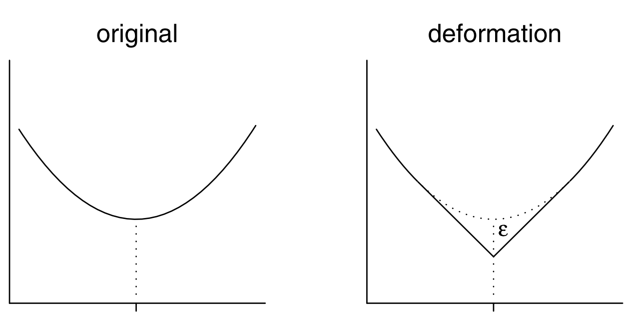

FIG. 2. *Example of the deformation $D_\varepsilon^*(g; x_0)$.*

PROOF. Obviously, $\tilde{g}_\varepsilon \geq g - \varepsilon$. Since $g - \varepsilon$ is a closed proper convex function, it is equal to the supremum of all linear functions l such that $l \leq h - \varepsilon$. Thus, $g - \varepsilon \leq D_\varepsilon^*(g; x_0)$, which implies that $D_\varepsilon^*(g; x_0)$ is a proper convex function and $\operatorname{dom} D_\varepsilon^*(g; x_0) \subseteq \operatorname{dom}(g - \varepsilon) = \operatorname{dom} g$. By Lemma S.A.10, we have $D_\varepsilon^*(g; x_0) \leq g$ and therefore $\operatorname{dom} g \subseteq \operatorname{dom} D_\varepsilon(g; x_0)$, which proves item 1.

If $v \in \partial g(x_1)$, then $l_v(x) = \langle v, x - x_1\rangle + g(x_1)$ is a support plane to $g(x)$ and $l_v \leq g$. If inequality (3.7) holds true, then $l_v(x)$ is majorized by $\tilde{g}_\varepsilon$ and we have $D_\varepsilon(g; x_0)(x_1) \leq g(x_1) = l_v(x_1) \leq D_\varepsilon(g; x_0)(x_1)$. On the other hand, by item 1, we have $x_1 \in \operatorname{ri}(\operatorname{dom} D_\varepsilon(g; x_0))$, hence there exists $v \in \partial D_\varepsilon(g; x_0)(x_1)$ and

$$\begin{aligned} g(x) \geq \tilde{g}_\varepsilon(x) \geq D_\varepsilon(g; x_0)(x) &\geq \langle v, x_0 - x_1\rangle + D_\varepsilon(g; x_0)(x_1) \\ &= \langle v, x_0 - x_1\rangle + g(x_1). \end{aligned}$$

Therefore, $v \in \partial g(x_1)$. In particular,

$$\begin{aligned} \tilde{g}_\varepsilon(x_0) = g(x_0) - \varepsilon \geq D_\varepsilon(g; x_0)(x_0) &\geq \langle v, x_0 - x_1\rangle + D_\varepsilon(g; x_0)(x_1) \\ &= \langle v, x_0 - x_1\rangle + g(x_1), \end{aligned}$$

which proves item 2.

We can represent $D_\varepsilon^*(g^*; x_0)$ as the maximal convex minorant of g defined by $g = \min(g, g(x_0) - \varepsilon + \delta(\cdot|x_0))$. For $x \in \operatorname{dom} g$, by Lemma S.A.10, $g^*(v_0) + g(x_0) = \langle v_0, x_0\rangle$. Thus,

$$\big(g(x_0) - \varepsilon + \delta(\cdot|x_0)\big)^*(y) = \langle x_0, y\rangle - g(x_0) + \varepsilon = \langle x_0, y - v\rangle + \varepsilon$$

for some $v \in \partial g(x_0)$. By Lemma S.A.7, we have

$$D_\varepsilon^*(g^*; x_0)^* = \max(g^*, l_0^*), \qquad l_0^*(y) \equiv \langle x_0, y - v\rangle + \varepsilon,$$

which concludes the proof the lemma. $\square$

Since the domain of the quadratic part of equation (3.5) is $\mathbb{R}^d$, by Lemma S.A.11, we have that for $x_0 \in \operatorname{dom} g$ and $v_0 \in \partial g(x)$, there exists $w_0 \in \partial q(x)$ such that

$$v_0 = G(x - x_0) + w_0. \tag{3.8}$$

Therefore, for the point x_1 in the neighborhood $O(x_0)$ where the decomposition (3.5) is true, condition (3.6) is equivalent to

$$\tfrac{1}{2}(x_1 - x_0)^T G(x_1 - x_0) + q(x_1) - \varepsilon \geq \langle w_0, x_1 - x_0\rangle + q(x_0).$$

Since $\langle w_0, x_1 - x_0\rangle + q(x_0)$ is a support plane to $q(x)$, the inequality (3.6) is satisfied if $2^{-1}(x_1 - x_0)^T G(x_1 - x_0) \geq \varepsilon$, which is the complement of an open ellipsoid $B_G(x_0, \sqrt{2\varepsilon})$ defined by G with center at x_0. For ε small enough, this ellipsoid will belong to the neighborhood $O(x_0)$. Since $|D_\varepsilon(g; x_0, v_0) - g| \leq \varepsilon$, this proves that the family $D_\varepsilon(g; x_0, v_0)$ is a local deformation.

In the same way, the condition (3.7) is equivalent to

$$\tfrac{1}{2}(x_1 - x_0)^T G(x_1 - x_0) + q(x_1) + \varepsilon \le \langle G(x_1 - x_0) + w_1, x_1 - x_0\rangle + q(x_0)$$

or $2^{-1}(x_1 - x_0)^T G(x_1 - x_0) + q(x_0) - \varepsilon \ge \langle w_1, x_0 - x_1\rangle + q(x_1)$, which is satisfied if we have $2^{-1}(x_1 - x_0)^T G(x_1 - x_0) \ge \varepsilon$. Since $|D^*_\varepsilon(g; x_0) - g| \le \varepsilon$, this proves that the family $D^*_\varepsilon(g; x_0)$ is also a local deformation. Thus, we have proven the following.

LEMMA 3.21. *Let g be a closed proper convex function, locally G-strongly convex at some $x_0 \in \operatorname{ri}\operatorname{dom} g$ and $v_0 \in \partial g(x_0)$. The families $D_\varepsilon(g; x_0, v_0)$ and $D^*_\varepsilon(g; x_0)$ are then local deformations for all $\varepsilon > 0$ small enough. Moreover, the condition $2^{-1}(x - x_0)^T G(x - x_0) \ge \varepsilon$ implies that $D_\varepsilon(g; x_0, v_0)(x) = D^*_\varepsilon(g; x_0)(x) = g(x)$; equivalently,* $\operatorname{supp}[D_\varepsilon(g; x_0, v_0) - g]$ *and* $\operatorname{supp}[D^*_\varepsilon(g; x_0) - g]$ *are subsets of $B_G(x_0, \sqrt{2\varepsilon})$.*

For $r > 0$ small enough, $h' \circ g(x)$ is nonzero and the decomposition (3.5) is true on $B(x_0; r)$. Let us fix some $v_0 \in \partial g(x_0)$, some $x_1 \in B(x_0; r)$ such that $x_1 \ne x_0$ and some $y_1 \in \partial g(x_1)$. We fix δ such that equation (3.3) of Lemma 3.19 is true for the transformation $\sqrt{h}$ and $r = 2$, and also $x_0 \notin \overline{B_G(x_1; \sqrt{2\delta})}$. Then, by Lemma 3.21, for all $\varepsilon > 0$ small enough, the support sets $\operatorname{supp}[D_\varepsilon(g; x_0, v_0) - g]$ and $\operatorname{supp}[D^*_\delta(g; x_1) - g]$ do not intersect; that is, these two deformations do not interfere.

We can now prove Theorem 2.23. The argument below is identical for g^+_ε and g^-_ε, so we will give the proof only for g^+_ε. We define deformations g^+_ε and g^-_ε by means of the following lemma.

LEMMA 3.22 (S.3.3). *For all $\varepsilon > 0$ small enough, there exist $\theta^+_\varepsilon, \theta^-_\varepsilon \in (0, 1)$ such that the functions g^+_ε and g^-_ε defined by*

$$\begin{aligned} g^+_\varepsilon &= (1 - \theta^+_\varepsilon) D_\varepsilon(g; x_0, v_0) + \theta^+_\varepsilon D^*_\delta(g; x_1), \\ g^-_\varepsilon &= (1 - \theta^-_\varepsilon) D^*_\varepsilon(g; x_0) + \theta^-_\varepsilon D_\delta(g; x_1; v_1) \end{aligned}$$

belong to $\mathcal{P}(h)$.

Next, we will show that θ^+_ε goes to zero fast enough so that g^+_ε is very close to $D_\varepsilon(g; x_0, v_0)$. Since supports do not intersect, we have

$$\begin{aligned} 0 &= \int (h \circ g^+_\varepsilon - h \circ g)\, dx \\ &= \int \bigl(h \circ \bigl((1 - \theta^+_\varepsilon) D_\varepsilon(g; x_0, v_0) + \theta^+_\varepsilon g\bigr) - h \circ g\bigr)\, dx \\ &\quad - \int \bigl(h \circ g - h \circ \bigl((1 - \theta^+_\varepsilon) g + \theta^+_\varepsilon D^*_\delta(g; x_1)\bigr)\bigr)\, dx, \end{aligned}$$

where both integrals have the same sign. For the first integral, by Lemma 3.18, we have

$$\begin{aligned}\int \big|h\circ\big((1-\theta_\varepsilon^+)D_\varepsilon(g;x_0,v_0)+\theta_\varepsilon^+ g\big)-h\circ g\big|\,dx \\ \le \int |h\circ D_\varepsilon(g;x_0,v_0)-h\circ g|\,dx \\ \asymp \int \big(g-D_\varepsilon(g;x_0,v_0)\big)\,dx \le \varepsilon\mu\big[B_G\big(x_0;\sqrt{2\varepsilon}\big)\big].\end{aligned}$$

The second integral is monotone in θ_ε^+ and, by Lemma 3.19, we have

$$\int \big(h\circ g - h\circ\big((1-\theta_\varepsilon^+)g+\theta_\varepsilon^+ D_\delta^*(g;x_1)\big)\big)\,dx \asymp \theta_\varepsilon^+.$$

Thus, we have $\theta_\varepsilon^+ = O(\varepsilon^{1+d/2})$ and

$$\lim_{\varepsilon\to 0}\varepsilon^{-1}\big(g_\varepsilon^+(x_0)-g(x_0)\big)=\lim_{\varepsilon\to 0}(1-\theta_\varepsilon^+)=1.$$

For Hellinger distance, we have

$$\begin{aligned}H(h\circ g_\varepsilon^+, h\circ g) = H\big(h\circ\big((1-\theta_\varepsilon^+)D_\varepsilon(g;x_0,v_0)+\theta_\varepsilon^+ g\big), h\circ g\big) \\ + H\big(h\circ\big((1-\theta_\varepsilon^+)g+\theta_\varepsilon^+ D_\delta^*(g;x_1)\big), h\circ g\big).\end{aligned}$$

We can now apply Lemma 3.18:

$$H^2\big(h\circ\big((1-\theta_\varepsilon^+)D_\varepsilon(g;x_0,v_0)+\theta_\varepsilon^+ g\big), h\circ g\big) \le H^2\big(h\circ D_\varepsilon(g;x_0,v_0), h\circ g\big),$$

$$\lim_{\varepsilon\to 0}\frac{H^2(h\circ D_\varepsilon(g;x_0,v_0), h\circ g)}{\int (D_\varepsilon(g;x_0,v_0)-g)^2\,dx} = \frac{h'\circ g(x_0)^2}{4h\circ g(x_0)}$$

and

$$\int \big(D_\varepsilon(g;x_0,v_0)-g\big)^2\,dx \le \varepsilon^2\mu\big[B_G\big(x_0;\sqrt{2\varepsilon}\big)\big] = \varepsilon^{2+d/2}\frac{2^{d/2}\mu[S(0,1)]}{\sqrt{\det G}}.$$

This yields

$$\begin{aligned}\limsup_{\varepsilon\to 0}\varepsilon^{-(d+4)/4}H\big(h\circ\big((1-\theta_\varepsilon^+)D_\varepsilon(g;x_0,v_0)+\theta_\varepsilon^+ g\big), h\circ g\big) \\ \le C(d)\bigg(\frac{h'\circ g(x_0)^4}{h\circ g(x_0)^2\det G}\bigg)^{1/4},\end{aligned}$$

where $S(0,1)$ is the d-dimensional sphere of radius 1.

For the second part, by Lemma 3.19, we obtain

$$\limsup_{\varepsilon\to 0}(\theta_\varepsilon^+)^{-2}H^2\big(h\circ\big((1-\theta_\varepsilon^+)g+\theta_\varepsilon^+ D_\delta^*(g;x_1)\big), h\circ g\big) < \infty$$

and

$$H\big(h\circ\big((1-\theta_\varepsilon^+)g+\theta_\varepsilon^+ D_\delta^*(g;x_1)\big), h\circ g\big) = O\big(\varepsilon^{(d+2)/2}\big).$$

Thus,

$$\limsup_{\varepsilon\to 0} \varepsilon^{-(d+4)/4} H(h\circ g_\varepsilon^+, h\circ g) \le C(d)\left(\frac{h'\circ g(x_0)^4}{h\circ g(x_0)^2 \det G}\right)^{1/4}.$$

Finally, we apply Corollary 2.21:

$$\liminf_{n\to\infty} n^{2/(d+4)} R_1(n; T, \{g, g_n\}) \ge C(d)\left[\frac{h\circ g(x_0)^2 \det G}{h'\circ g(x_0)^4}\right]^{1/(d+4)}.$$

Taking the supremum over all $G \in \mathcal{SC}(g; x_0)$, we obtain the statement of the theorem. $\square$

3.5. *Indications of proofs for conjectured rates.* From Birgé and Massart (1993) and van der Vaart and Wellner (1996), we expect that the global rate of convergence of the MLE $\hat{p}_n$ of $p_0 = h \circ g_0$ in the class $\mathcal{P}(h)$ will be determined by the entropy of the class of convex and Lipschitz functions g on convex bounded domains C, as given by Bronšteĭn (1976) and Dudley (1999): if $\mathcal{F}_{L,C}$ is the class of all convex functions defined on a compact convex set $C \subset \mathbb{R}^d$ such that $|f(x) - f(y)| \le L\|x - y\|$ for all $x, y \in C$, then the covering numbers for $\mathcal{F}_{L,C}$ satisfy

$$\log N(\epsilon, \mathcal{F}_{L,C}, \|\cdot\|_\infty) \le K(1+L)^{d/2}\epsilon^{-d/2} \tag{3.9}$$

for all (small) $\epsilon > 0$, for a constant K depending only on C and d. Then, after an argument to transfer this covering number bound to a bracketing entropy bound for $\mathcal{P}(h)$ with respect to Hellinger distance H, it follows from oscillation bounds for empirical processes [cf. van der Vaart and Wellner (1996), Theorems 3.4.1 and 3.4.4] that rates of convergence of $\hat{p}_n$ with respect to Hellinger distance are determined by $r_n^2\phi(1/r_n) \asymp \sqrt{n}$ with

$$\phi(\delta) \equiv \int_{c\delta}^{\delta} \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{P}(h), H)}\, d\epsilon. \tag{3.10}$$

Assuming that the bound of (3.9) can be carried over to $\log N_{[]}(\epsilon, \mathcal{P}(h), H)$ sufficiently closely, routine calculations show that the expected rates of convergence of $\hat{p}_n$ to $p_0 = h(g_0)$ with respect to Hellinger distance H are

$$r_n = \begin{cases} n^{2/(d+4)}, & \text{if } d \in \{1,2,3\},\\ (n/(\log n)^2)^{1/4}, & \text{if } d = 4,\\ n^{1/d}, & \text{if } d > 4.\end{cases}$$

Based on these heuristics, we expect that the MLE $\hat{p}_n$ will be rate efficient if $d \le 3$, but rate inefficient (not attaining the optimal rate $n^{2/(d+4)}$) if $d \ge 4$.

Some details.

Case 1: $d \leq 3$. In this case, we find that

$$\begin{aligned}\phi(\delta) &= \int_{c\delta^2}^{\delta} \sqrt{1+\log N_{[]}(\epsilon, \mathcal{F}_{L,C}, \|\cdot\|)}\, d\epsilon \\ &\asymp \int_{c\delta^2}^{\delta} \sqrt{K(1+L)^{d/2}}\epsilon^{d/4} d\epsilon \\ &\asymp M_1 \delta^{1-d/4},\end{aligned}$$

where $M_1 \equiv (K(1+L)^{d/2})^{1/2}/(1-d/4)$. Solving the relation $r_n^2 \phi(1/r_n) \asymp \sqrt{n}$ for r_n yields $r_n = n^{2/(d+4)}$ up to a constant.

Case 2: $d = 4$. In this case, we find that

$$\phi(\delta) \asymp M_2 \log\left(\frac{1}{c\delta}\right),$$

where $M_2 \equiv (K(1+L)^{d/2})^{1/2}$. Solving the relation $r_n^2 \phi(1/r_n) \asymp \sqrt{n}$ for r_n yields $r_n = (n/(\log n)^2)^{1/4}$ up to a constant.

Case 3: $d > 4$. In this case, we calculate

$$\phi(\delta) \asymp M_2 \delta^{2(1-d/4)},$$

where $M_3 \equiv (K(1+L)^{d/2})^{1/2}/(d/4-1)$. Solving the relation $r_n^2 \phi(1/r_n) \asymp \sqrt{n}$ for r_n yields $r_n = n^{1/d}$ up to a constant.

Acknowledgments. This research is part of the Ph.D. dissertation of the first author at the University of Washington. We would like to thank two referees for a number of helpful suggestions.

SUPPLEMENTARY MATERIAL

Supplement: Omitted Proofs and Some Facts from Convex Analysis (DOI: 10.1214/10-AOS840; .pdf). In the supplement, we provide omitted proofs and some basic facts from convex analysis used in this paper.

REFERENCES

AN, M. Y. (1998). Logconcavity versus logconvexity: A complete characterization. *J. Econom. Theory* **80** 350–369. MR1637480

AVRIEL, M. (1972). r-convex functions. *Math. Program.* **2** 309–323. MR0301151

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. Available at http://dx.doi.org/10.1007/BF01199316. MR1240719

BORELL, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.* **6** 111–136. MR0404559

BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480

BRONŠTEĬN, E. M. (1976). ε-entropy of convex sets and functions. *Sibirsk. Mat. Ž.* **17** 508–514, 715. MR0415155

CORDERO-ERAUSQUIN, D., MCCANN, R. J. and SCHMUCKENSCHLÄGER, M. (2001). A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* **146** 219–257. MR1865396

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

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density (with discussion). *J. Roy. Statist. Soc. Ser. B* **72** 1–32.

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity and Applications*. Academic Press, Boston, MA. MR0954608

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. MR1105839

DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712

DÜMBGEN, L., HÜSLER, A. and RUFIBACH, K. (2007). Active set and EM algorithms for log-concave densities based on complete and censored data. Technical report, Univ. Bern. Available at arXiv:0707.4643.

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

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

IBRAGIMOV, I. A. (1956). On the composition of unimodal distributions. *Teor. Veroyatnost. i Primenen.* **1** 283–288. MR0087249

JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York. MR0418337

JONGBLOED, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett.* **50** 279–284. MR1792307

KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027.

OKAMOTO, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Statist.* **1** 763–765. MR0331643

PAL, J. K., WOODROOFE, M. B. and MEYER, M. C. (2007). Estimating a Polya frequency function. In *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459192

PRÉKOPA, A. (1973). On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)* **34** 335–343. MR0404557

RINOTT, Y. (1976). On convexity of measures. *Ann. Probab.* **4** 1020–1026. MR0428540

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton. MR0274683

RUFIBACH, K. (2006). Log-concave density estimation and bump hunting for I.I.D. observations. Ph.D. thesis, Univ. Bern and Göttingen.

RUFIBACH, K. (2007). Computing maximum likelihood estimators of a log-concave density function. *J. Stat. Comput. Simul.* **77** 561–574. MR2407642

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

SCHUHMACHER, D., HÜSLER, A. and DUEMBGEN, L. (2009). Multivariate log-concave distributions as a nearly parametric model. Technical report, Univ. Bern. Available at arXiv:0907.0250v1.

SEREGIN, A. and WELLNER, J. A. (2010). Supplement to "Nonparametric estimation of multivariate convex-transformed densities." DOI: 10.1214/10-AOS840SUPP.

UHRIN, B. (1984). Some remarks about the convolution of unimodal functions. *Ann. Probab.* **12** 640–645. MR0735860

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

WALTHER, G. (2010). Inference and modeling with log-concave distributions. *Statist. Sci.* **24** 319–327.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195-4322
USA
E-MAIL: arseni@stat.washington.edu
jaw@stat.washington.edu