Submitted to the Annals of Statistics
arXiv: math.ST/0911.4151v1

SUPPLEMENT TO "NONPARAMETRIC ESTIMATION OF MULTIVARIATE CONVEX-TRANSFORMED DENSITIES."

By Arseni Seregin* and Jon A. Wellner†

University of Washington

1. Properties of the increasing transformation.

Lemma 1.1. *Let h be a increasing transformation and g be a closed proper convex function with $\operatorname{dom} g = \overline{\mathbb{R}}_+^d$ such that*

$$\int_{\overline{\mathbb{R}}_+^d} h \circ g dx = C < \infty.$$

Then the following are true:

1. *For a sublevel set $\operatorname{lev}_y g$ with $y > y_0$ we have:*
$$\mu[(\operatorname{lev}_y g)^c] \leq C/h(y).$$
2. *For any point $x_0 \in \mathbb{R}_+^d$ and any subgradient $a \in \partial g(x_0)$ all coordinates of a are nonpositive. If in addition $g(x_0) > y_0$ then all coordinates of a are negative.*
3. *For any point $x_0 \in \mathbb{R}_+^d$ such that $g(x_0) > y_0$ we have:*
$$h \circ g(x_0) \leq \frac{Cd!}{d^d V(x_0)},$$
where $V(x) \equiv \prod_{k=1}^d x_k$ for $x \in \mathbb{R}_+^d$.
4. *The function h reverses partial order on $\overline{\mathbb{R}}_+^d$: if $x_1 < x_2$ then $g(x_1) \geq g(x_2)$ and the last inequality is strict if $g(x_1) > y_0$.*
5. *The supremum of g on $\overline{\mathbb{R}}_+^d$ is attained at 0.*

*Research supported in part by NSF grant DMS-0804587

†Research supported in part by NSF grant DMS-0804587 and NI-AID grant 2R01 AI291968-04

AMS 2000 subject classifications: Primary 62G07, 62H12; secondary 62G05, 62G20

Keywords and phrases: consistency, log–concave density estimation, lower bounds, maximum likelihood, mode estimation, nonparametric estimation, qualitative assumptions, shape constraints, strongly unimodal, unimodal

PROOF. 1. Since h is nondecreasing we have $h(y) > 0$ and:

$$C = \int_{\overline{\mathbb{R}}_+^d} h \circ g dx \geq \int_{(\mathrm{lev}_y\, g)^c} h \circ g dx \geq h(y)\mu[(\mathrm{lev}_y\, g)^c].$$

2. Consider the linear function $l(x) = a^T(x - x_0) + g(x_0)$. We have $g \geq l$. If the vector a has a nonnegative coordinate a_i then consider a closed ball $B = \bar{B}(x_0) \subset \mathbb{R}_+^d$. If m is a minimum of the function l on B then the minimum of the function $h \circ l$ on $B + \lambda e_i$ is equal to $h(m + \lambda a_i)$, where e_i is the element of the basis which corresponds to the ith coordinate. For $\lambda > 0$ we have $B + \lambda e_i \subset \mathbb{R}_+^d$.
If $a_i > 0$ then:

$$\int_{\overline{\mathbb{R}}_+^p} h \circ g dx \geq \int_{\overline{\mathbb{R}}_+^p} h \circ l dx \geq \int_B h \circ l dx \ \geq \mu[B] h(m + \lambda a_i) \to +\infty$$

as $\lambda \to \infty$, which contradicts the assumption.
If $a_i = 0$ and $g(x_0) = l(x_0) > y_0$, then we can choose the radius of the ball small enough so that $m > y_0$. Then:

$$\int_{\overline{\mathbb{R}}_+^p} h \circ g dx \geq \int_{\overline{\mathbb{R}}_+^p} h \circ l dx \geq \int_K h \circ l dx \geq \mu[K] h(m) = +\infty$$

where $K \equiv \cup_{\lambda > 0}(B + \lambda e_i)$, and this again contradicts the assumption.

3. Consider the subgradient $a \in \partial g(x_0)$. For the linear function $l(x) = a^T(x - x_0) + g(x_0)$ we have $g \geq l$ and $l(x_0) = g(x_0)$ therefore $(\mathrm{lev}_{g(x_0)}\, l)^c \subseteq (\mathrm{lev}_{g(x_0)}\, g)^c$. From the previous statement we have that $(\mathrm{lev}_{g(x_0)})^c$ is a simplex and using inequality of arithmetic and geometric means we have:

$$\mu[(\mathrm{lev}_{g(x_0)}\, l)^c] = \frac{(a^T x_0)^d}{d! V(a)} \geq \frac{d^d V(x_0)}{d!}$$

which together with 1. proves the statement.

4. Since $x_1 \in \overline{\mathbb{R}}_+^d$ and $x_1 < x_2$ we have $x_2 \in \mathbb{R}_+^d = \mathrm{ri}(\mathrm{dom}\, g)$. For any subgradient $a \in \partial g(x_2)$ we have

$$g(x_1) - g(x_2) \geq a^T(x_1 - x_2) \geq 0$$

from the previous statement. Now, if $g(x_1) > y_0$ then we can assume that $g(x_2) > y_0$ since otherwise the statement is trivial. In this case all coordinates of a are negative and:

$$g(x_1) - g(x_2) \geq a^T(x_1 - x_2) > 0.$$

5. From the previous statement we have that $h \circ g \leq h \circ g(0)$ on $\mathbb{R}^d_+$ which together with continuity of $h \circ g$ implies the statement.

□

LEMMA 1.2. *Let h be an increasing transformation, g be a closed proper convex function on $\overline{\mathbb{R}}^d_+$ and Q be a σ-finite Borel measure on $\overline{\mathbb{R}}^d_+$. Then:*

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

PROOF. Using the Fubini-Tonelli theorem we have:

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

□

LEMMA 1.3. *Let h be an increasing transformation and let g be a polyhedral convex function with* $\operatorname{dom} g = \overline{\mathbb{R}}^d_+$ *such that:*

$$\int_{\overline{\mathbb{R}}^d_+} h \circ g dx < \infty.$$

Then $g(0) < y_\infty$.

PROOF. For $y_\infty = +\infty$ the statement is trivial so we assume that y_∞ is finite. If $g(0) > y_\infty$ then since g is continuous there exists a ball $B \subset \overline{\mathbb{R}}^d_+$ small enough such that $g > y_\infty$ on B and therefore

$$\int_{\overline{\mathbb{R}}^d_+} h \circ g dx = \infty.$$

Let us assume that $g(0) = y_\infty$. By Lemma A.13 there exists $a \in \partial g(0)$ and therefore $g(x) \geq l(x) \equiv a^T x + y_\infty$. Let a_m be the minimum among the coordinates of the vector a and -1. Then on $\overline{\mathbb{R}}^d_+$ we have $l(x) \geq l_1(x) \equiv a_m \mathbf{1}^T x + y_\infty$ where $a_m < 0$ and thus $l_1(x) \leq y_\infty$. By Lemma 1.2 we have:

$$\int_{\overline{\mathbb{R}}^d_+} h \circ g dx \geq \int_{\overline{\mathbb{R}}^d_+} h \circ l_1 dx = \int_{-\infty}^{y_\infty} h'(y)\mu[(\mathrm{lev}_y\, l_1)^c \cap \overline{\mathbb{R}}^d_+]dy.$$

The set $A_y = (\mathrm{lev}_y\, g)^c \cap \overline{\mathbb{R}}^d_+$ is a simplex and:

$$\mu[A_y] = \frac{(y_\infty - y)^d}{d!(-a_m)^d}$$

for $y \leq y_\infty$. By assumption M.I.2 we have $h'(y) \asymp (y_\infty - y)^{-\beta-1}$ as $y \uparrow y_\infty$ where $\beta > d$ and therefore:

$$\int_{\overline{\mathbb{R}}^d_+} h \circ g_1 dx = \int_{\overline{\mathbb{R}}^d_+} h \circ g dx = +\infty.$$

This contradiction proves that $g(0) < y_\infty$. □

LEMMA 1.4. *Let h be an increasing transformation and let $l(x) = a^T x + b$ be a linear function such that all coordinates of a are negative and $b < y_\infty$. Then:*

$$\int_{\overline{\mathbb{R}}^d_+} h \circ l dx < \infty.$$

PROOF. We have $l \leq b$ on $\overline{\mathbb{R}}^d_+$ and by Lemma 1.2:

$$\int_{\overline{\mathbb{R}}^d_+} h \circ l dx = \int_{-\infty}^{b} h'(y)\mu[(\mathrm{lev}_y\, l)^c \cap \overline{\mathbb{R}}^d_+]dy.$$

The set $A_y = (\mathrm{lev}_y\, l)^c \cap \overline{\mathbb{R}}^d_+$ is a simplex and:

$$\mu[A_y] = \frac{(b - y)^d}{d! V(-a)}$$

for $y \leq b$. By assumption M.I.1 we have $h'(y) = o(y^{-\alpha-1})$ as $y \to -\infty$ for $\alpha > d$ and therefore the integral is finite. □

LEMMA 1.5. *Let h be an increasing transformation and suppose that $K \subset \overline{\mathbb{R}}^d_+$ is a compact set. Then there exists a closed proper convex function $g \in \mathcal{G}(h)$ such that $g > y_0$ on K.*

Proof. If $y_0 = -\infty$ then consider the function $T(c)$ defined as:

$$T(c) = \int_{\overline{\mathbb{R}}_+^d} h \circ (-\mathbf{1}^T x + c) dx.$$

By Lemma 1.4, $T(c)$ is finite for $c < y_\infty$, and by Lemma 1.3, we conclude that $T(y_\infty) = +\infty$. By monotone convergence T is left-continuous for $c \in (-\infty, y_\infty]$ and by dominated convergence is right-continuous for $c \in (-\infty, y_\infty)$. Since $T(-\infty) = 0$ there exists $c_1 < y_\infty$ such that $T(c_1) = 1$ and thus the linear function $l(x) = -\mathbf{1}^T x + c_1$ belongs to $\mathcal{G}(h)$.

If $y_0 < -\infty$ then choose M such that $\mathbf{1}^T x < M$ on K. Consider the function $T(c)$ defined as:

$$T(c) = \int_{\overline{\mathbb{R}}_+^d} h \circ (c(-\mathbf{1}^T x + M) + y_0) dx.$$

By Lemma 1.4, $T(c)$ is finite for $c < (y_\infty - y_0)/M$ and by Lemma 1.3, $T((y_\infty - y_0)/M) = +\infty$. By monotone and dominated convergence T is continuous for $c \in [0, (y_\infty - y_0)/M]$. Since $T(0) = 0$ there exists $c_1 \in (0, (y_\infty - y_0)/M)$ such that linear function $l(x) = c_1(-\mathbf{1}^T x + M) + y_0$ belongs to $\mathcal{G}(h)$. By construction $l > y_0$ on K. $\square$

Lemma 1.6. *If $X_1, \ldots, X_n$ are i.i.d. $p_0 = h \circ g_0 \in \mathcal{P}(h)$ for a monotone transformation h, then the observations X are in general position with probability* 1.

Proof. Points are not in general position if at least one subset Y of X of size $d+1$ belongs to a proper linear subspace of $\mathbb{R}^d$. This is true if and only if X as a vector in $\mathbb{R}^{nd}$ belongs to a certain non-degenerate algebraic variety. Since with probability 1 we have $X \subset \operatorname{dom} g_0$ and by definition $\dim(\operatorname{dom} g_0) = d$, the statement follows from Okamoto [1973]. $\square$

Below we assume that our observations are in general position for any n. For an increasing model we also assume that all X_i belong to $\mathbb{R}_+^d$. This assumption holds with probability 1 since $\mu\left[\overline{\mathbb{R}}_+^d \setminus \mathbb{R}_+^d\right] = 0$.

Lemma 1.7 (M.3.6). *Consider an increasing transformation h. For any convex function g with* $\operatorname{dom} g = \overline{\mathbb{R}}_+^d$ *such that:*

$$\int_{\overline{\mathbb{R}}_+^d} h \circ g dx \leq 1$$

and $\mathbb{L}_n g > -\infty$, there exists $\tilde{g} \in \mathcal{G}(h)$ such that $\tilde{g} \geq g$ and $\mathbb{L}_n \tilde{g} \geq \mathbb{L}_n g$. The function $\tilde{g}$ can be chosen as a minimal element in $\operatorname{ev}_X^{-1} \tilde{p}$ where $\tilde{p} = \operatorname{ev}_X \tilde{g}$.

PROOF. Let $p = \mathrm{ev}_X g$. Since $\mathbb{L}_n g > -\infty$ we have $g(X_i) > y_0$ for all $1 \le i \le n$ and therefore $g(x) > y_0$ for $x \in \mathrm{conv}(X)$. Consider any minimal element g_1 among convex functions in $\mathrm{ev}_X^{-1} p$ (which exists by Lemma A.15). Then:

$$\int_{\overline{\mathbb{R}}_+^d} h \circ g_1 dx \le \int_{\overline{\mathbb{R}}_+^d} h \circ g dx \le 1.$$

Since g_1 is polyhedral we have $g_1 = \max l_i$ for some linear functions $l_i(x) = a_i^T x + b_i$ and for each function l_i there exists some facet of g_1 such that $g_1 = l_i$ on it.

By Lemma A.15 the interior of the facet of g_1 which corresponds to l_i contains some $X_{j_i} \in X$. We have $\partial g_1(X_{j_i}) = \{a_i\}$ and $g_1(X_{j_i}) = g(X_{j_i}) > y_0$. Thus by Lemma 1.1, all coordinates of a_i are negative and the supremum M of g_1 is attained at 0. Therefore $b_i = l_i(0) \le M$. By Lemma 1.3 we have $M < y_\infty$. Thus by Lemma 1.4 the functions $h \circ (l_i + c)$ are integrable for all $c < y_\infty - M$. Since g_1 has only a finite number of facets we have that $h \circ (g_1 + c)$ is also integrable for all $c < y_\infty - M$. Finally, for $c = y_\infty - M$ the function $h \circ (g_1 + c)$ is not integrable by Lemma 1.3.

The function $T(c)$ defined as:

$$T(c) \equiv \int_{\overline{\mathbb{R}}_+^d} h \circ (g_1 + c) dx$$

is increasing, finite for $c \in [0, y_\infty - M)$ and continuous for $c \in [0, y_\infty - M]$ by monotone and dominated convergence. Since $T(0) \le 1$ and $T(y_\infty - M) = +\infty$, there exists $c_1 \in (0, y_\infty - M)$ such that $T(c_1) = 1$. Since g_1 is the minimal element in $\mathrm{ev}_X^{-1}(p)$, the function $\tilde{g} \equiv g_1 + c_1$ is minimal in $\mathrm{ev}_X^{-1}(p + c_1)$. Then $\tilde{g}$ satisfies the conditions of our lemma. □

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

PROOF. Let $\widehat{g}_0$ be any MLE. Then by Lemma 1.5 applied to $K = \mathrm{conv}(X)$ it follows that $\mathbb{L}_n \widehat{g}_0 > -\infty$. By Lemma 1.7 there exists a function $\widehat{g}_1 \in \mathcal{P}(h)$ such that $\widehat{g}_1$ is a minimal element in $\mathrm{ev}_X^{-1} q_1$ where $q_1 = \mathrm{ev}_X \widehat{g}_1$ and $\widehat{g}_1 \ge \widehat{g}_0$. Since $\widehat{g}_0$ is a MLE we have $\mathrm{ev}_X \widehat{g}_0 = \mathrm{ev}_X \widehat{g}_1$ which together with Lemma A.15 proves the first part of the statement.

By Lemma 1.3 we have $\widehat{g}_0 < y_\infty$ and $\widehat{g}_1 < y_\infty$. Since $h \circ \widehat{g}_0$ and $h \circ \widehat{g}_1$ are continuous functions, for the strictly increasing h the equality:

$$\int_{\overline{\mathbb{R}}_+} (h \circ \widehat{g}_1 - h \circ \widehat{g}_0) dx = 0$$

implies that $\widehat{g}_1(x) = \widehat{g}_0(x)$ for x such that $\widehat{g}_0(x) > y_0$. □

Lemma 1.9 (M.3.8). *Consider a decreasing transformation h. For any convex function g such that:*

$$\int_{\mathbb{R}^d} h \circ g dx \leq 1$$

and $\mathbb{L}_n g > -\infty$ there exists $\tilde{g} \in \mathcal{G}(h)$ such that $\tilde{g} \leq g$ and $\mathbb{L}_n \tilde{g} \geq \mathbb{L}_n g$. The function $\tilde{g}$ can be chosen as the maximal element in $\mathrm{ev}_X^{-1} \tilde{q}$ where $\tilde{q} = \mathrm{ev}_X \tilde{g}$.

Proof. Let $p = \mathrm{ev}_X g$. Since $\mathbb{L}_n g > -\infty$ we have $g(X_i) < y_0$ for all $1 \leq i \leq n$ and therefore $g(x) < y_0$ for $x \in \mathrm{conv}(X)$. Consider the maximal element g_1 among convex functions in $\mathrm{ev}_X^{-1} p$ (which exists and is unique by Lemma A.14). Then:

$$\int_{\mathbb{R}^d} h \circ g_1 dx \leq \int_{\mathbb{R}^d} h \circ g dx = 1.$$

By Lemma M.3.1 there exists x_0 and $m > -\infty$ such that $g_1 \geq g_1(x_0) = m$. By Lemma M.3.3 we have $m > y_\infty$. By Lemma A.14 we have $\mathrm{dom}\, g_1 = \mathrm{conv}(X)$ and therefore:

$$\int_{\mathbb{R}^d} h \circ (g_1 + c) dx \leq h(m + c)\mu[\mathrm{conv}(X)] < \infty,$$

for $c \in (y_\infty - m, 0]$. By Lemma M.3.3 we have:

$$\int_{\mathbb{R}^d} h \circ (g_1 + y_\infty - m) dx = \infty.$$

Thus the function $T(c)$ defined as:

$$T(c) \equiv \int_{\mathbb{R}^d} h \circ (g_1 + c) dx$$

is decreasing, finite for $c \in (y_\infty - m, 0]$ and continuous for $c \in [y_\infty - m, 0]$ by monotone and dominated convergence. Since $T(0) \leq 1$ and $T(y_\infty - m) = +\infty$, there exists $c_1 \in (y_\infty - m, 0)$ such that $T(c_1) = 1$. Since g_1 is the maximal element in $\mathrm{ev}_X^{-1}(p)$, the function $\tilde{g} \equiv g_1 + c_1$ is maximal in $\mathrm{ev}_X^{-1}(p + c_1)$. Then $\tilde{g}$ satisfies the conditions of our lemma. □

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

PROOF. Let $\widehat{g}_0$ be any MLE. Then by Lemma M.3.5 applied to $K = \mathrm{conv}(X)$ we have that $\mathbb{L}_n \widehat{g}_0 > -\infty$. By Lemma 1.9 there exists a function $\widehat{g}_1 \in \mathcal{G}_h$ such that $\widehat{g}_1$ is the maximal element in $\mathrm{ev}_X^{-1} q_1$ where $q_1 = \mathrm{ev}_X \widehat{g}_1$ and $\widehat{g}_1 \le \widehat{g}_0$. Since $\widehat{g}_0$ is a MLE we have $\mathrm{ev}_X \widehat{g}_0 = \mathrm{ev}_X \widehat{g}_1$, which together with Lemma A.14 proves the first part of the statement.

By Lemma M.3.3 we have $\widehat{g}_0 \ge \widehat{g}_1 > y_\infty$. Since $h \circ \widehat{g}_0$ and $h \circ \widehat{g}_1$ are continuous functions, for the strictly decreasing h, the equality:

$$\int_{\mathbb{R}^d} (h \circ \widehat{g}_1 - h \circ \widehat{g}_0) dx = 0$$

implies that $\widehat{g}_1(x) = \widehat{g}_0(x)$ for $x \in \mathrm{conv}(X)$. Therefore $\widehat{g}_0(x) \ge y_\infty$ for $x \notin \mathrm{conv}(X)$. Since $\widehat{g}_0$ is convex we have $\widehat{g}_0 = \widehat{g}_1$. $\square$

LEMMA 1.11 (M.3.11). *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 $n_k \ge n_d$ such that for some $\varepsilon > -\infty$ and $\rho > 0$ the following is true:*

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

Then there exists $m > y_\infty$ such that $g_k \ge m$ for all k.

PROOF. Suppose, on the contrary, that $m_k \to y_\infty$ where $m_k = \min g_k$. The first condition implies that $X_d \equiv \{X_1, \dots, X_{n_d}\} \in \mathrm{dom}\, h_k$, and therefore by Corollary A.4 the function $\mu[\mathrm{lev}_y g_k]$ as a function of y admits all values in the interval $[\mu[\mathrm{lev}_{m_k} g_k], \mu[\mathrm{conv}(X_d)]]$. If the second condition is true for some ρ then it is also true for all $\rho' \in (0, \rho)$, and therefore we can assume that $\rho < \mu[\mathrm{conv}(X_d)]$.

By Lemma M.3.1 we have $\mu[\mathrm{lev}_{m_k} g_k] \to 0$, and thus there exists such a_k that $\mu[\mathrm{lev}_{a_k} g_k] = \rho$ for all k large enough. We define $A_k = \mathrm{lev}_{a_k} g_k$. By Lemma M.3.1 we have: $h(a_k) \le 1/\rho$ and therefore the sequence $\{a_k\}$ is bounded below by some $a > y_\infty$.

Consider $t_k > m_k$ such that $t_k \to y_\infty$. We will specify the exact form of t_k later in the proof. Since a_k are bounded away from y_∞, it follows that for

k large enough we will have $t_k < a_k$. Using Lemma A.3 we obtain:

$$\rho = \mu[A_k] \le \mu[\mathrm{lev}_{t_k} g_k] \left[\frac{a_k - m_k}{t_k - m_k}\right]^d \le \frac{1}{h(t_k)} \left[\frac{a_k - m_k}{t_k - m_k}\right]^d$$

which implies:

$$a_k \ge m_k + (t_k - m_k)[\rho h(t_k)]^{1/d}.$$

We have:

$$g_k \ge m_k 1\{A_k\} + a_k(1 - 1\{A_k\}),$$

and hence:

$$\begin{aligned}\mathbb{L}_{n_k} g_k &\le \mathbb{P}_{n_k}(A_k) \log h(m_k) + (1 - \mathbb{P}_{n_k}(A_k)) \log h(a_k)\\ &\le \mathbb{P}_{n_k}(A_k) \log h(m_k) + (1 - \mathbb{P}_{n_k}(A_k)) \log h(m_k + (t_k - m_k)[\rho h(t_k)]^{1/d}).\end{aligned}$$

Case $y_\infty = -\infty$. Choose $t_k = (1-\delta)m_k$ where $\delta \in (0,1)$. Then starting from some k we have $m_k < t_k$, $h(m_k) > 1$, $h(-Cm_k) < 1$ and $\delta[\rho h(t_k)]^{1/d} > C+1$. This implies:

$$m_k + (t_k - m_k)[\rho h(t_k)]^{1/d} = m_k(1 - \delta[\rho h(t_k)]^{1/d}) \ge -Cm_k,$$

and hence:

$$\begin{aligned}\mathbb{L}_{n_k} g_k &\le \mathbb{P}_{n_k}(A_k) \log h(m_k) + (1 - \mathbb{P}_{n_k}(A_k)) \log h(-Cm_k)\\ &\le \frac{d}{n_d} \log h(m_k) + \frac{n_d - d}{n_d} \log h(-Cm_k) = \frac{d}{n_d} \log\left[h(m_k) h(-Cm_k)^\gamma\right] \to -\infty.\end{aligned}$$

Case $y_\infty > -\infty$. Without loss of generality we can assume that $y_\infty = 0$. Choose $t_k = (1+\delta)m_k$ where $\delta > 0$. Then:

$$m_k + (t_k - m_k)[\rho h(t_k)]^{1/d} \ge m_k \delta[\rho h((1+\delta)m_k)]^{1/d} \asymp m_k^{-\frac{\beta - d}{d}} \to +\infty$$

which implies

$$h(m_k + (t_k - m_k)[\rho h(t_k)]^{1/d}) = o\left(m_k^{\frac{\alpha(\beta-d)}{d}}\right).$$

This in turn yields

$$\exp(\mathbb{L}_{n_k} g_k) = o\left(m_k^{-\frac{\beta d}{n_d} + \frac{\alpha(\beta-d)(n_d-d)}{d n_d}}\right) = o(1).$$

Therefore in both cases we obtained $\mathbb{L}_{n_k} g_k \to -\infty$. This contradiction concludes the proof. $\square$

LEMMA 1.12 (M.3.13). *Consider a monotone model $\mathcal{P}(h)$. Suppose 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

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

for $\varepsilon > 0$ small enough. Then the sequence of the MLEs is Hellinger consistent:

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

PROOF. For $\varepsilon \in (0,1)$ we have:

$$\begin{aligned}
0 &\geq \int_{\{h \circ g_0(x) \leq 1-\varepsilon\}} \log(\varepsilon + h \circ g_0) dP_0 \geq \log(\varepsilon) P_0\{h \circ g_0(x) \leq 1 - \varepsilon\} > -\infty \\
0 &\leq \int_{\{h \circ g_0(x) \geq 1\}} \log(\varepsilon + h \circ g_0) dP_0 \leq \int_{\{h \circ g_0(x) \geq 1\}} \log(2h \circ g_0) dP_0 \\
&\leq \int (h \log h) \circ g_0(x) dx + \log 2 < \infty.
\end{aligned}$$

Thus the function $\log(\varepsilon + h \circ g_0)$ is integrable with respect to probability measure P_0.

We can rearrange:

$$\begin{aligned}
0 \leq \mathbb{L}_n \hat{g}_n - \mathbb{L}_n g_0 &= \int \log[h \circ \hat{g}_n] d\mathbb{P}_n - \int \log[h \circ g_0] d\mathbb{P}_n \\
&\leq \int \log[\varepsilon + h \circ \hat{g}_n] d\mathbb{P}_n - \int \log[h \circ g_0] d\mathbb{P}_n \\
&\leq \int \log[\varepsilon + h \circ \hat{g}_n] d(\mathbb{P}_n - P_0) \quad (1.1) \\
&\quad + \int \log\left[\frac{\varepsilon + h \circ \hat{g}_n}{\varepsilon + h \circ g_0}\right] dP_0 \quad (1.2) \\
&\quad + \int \log[\varepsilon + h \circ g_0] dP_0 - \int \log[h \circ g_0] d\mathbb{P}_n. \quad (1.3)
\end{aligned}$$

The term (1.1) converges almost surely to zero by assumption.
For the term (1.2) we can apply the analogue of Lemma 1 from Pal et al. [2007]:

$$II \equiv \int \log\left[\frac{\varepsilon + h \circ \hat{g}_n}{b + h \circ g_0}\right] dP_0 \leq 2 \int \sqrt{\frac{\varepsilon}{\varepsilon + h \circ g_0}} dP_0 - 2H^2(h \circ \hat{g}_n, h \circ g_0).$$

For the term (1.3), the SLLN implies that:

$$\begin{aligned}
III &= \int \log[\varepsilon + h \circ g_0] dP_0 - \int \log[h \circ g_0] d\mathbb{P}_n \\
&\to_{a.s.} \int \log[\varepsilon + h \circ g_0] dP_0 - \int \log[h \circ g_0] dP_0 = \int \log\left[\frac{\varepsilon + h \circ g_0}{h \circ g_0}\right] dP_0.
\end{aligned}$$

Thus we have:

$$\begin{aligned}
0 &\le \liminf(I + II + III) \\
&\le_{a.s.} - \limsup 2H^2(h \circ \hat{g}_n, h \circ g_0) \\
&\qquad + 2\int \sqrt{\frac{\varepsilon}{\varepsilon + h \circ g_0}} dP_0 + \int \log\left[\frac{\varepsilon + h \circ g_0}{h \circ g_0}\right] dP_0.
\end{aligned}$$

This yields

$$\begin{aligned}
&\limsup H^2(h \circ \hat{g}_n, h \circ g_0) \\
&\quad \le_{a.s.} \int \sqrt{\frac{1}{1 + h \circ g_0/\varepsilon}} dP_0 + \frac{1}{2}\int \log\left[\frac{\varepsilon + h \circ g_0}{h \circ g_0}\right] dP_0 \to 0
\end{aligned}$$

as $\varepsilon \downarrow 0$ by monotone convergence. □

LEMMA 1.13 (M.3.15). *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:

$$\|\mathbb{P}_n - P\|_{\mathcal{A}} \to_{a.s.} 0.$$

PROOF. Let C be an upper bound for the density of P and a be large so that for the set $D \equiv [-a,a]^d$ we have $P([-a,a]^d) > 1 - \varepsilon/2C$. By assumption the class $\mathcal{A} \cap D$ has a finite set of $\varepsilon/2$-brackets $\{[L_i, U_i]\}$. Then for any set $A \in \mathcal{A}$ there exists index i such that:

$$L_i \subseteq A \cap D \subseteq U_i$$

Therefore:

$$L_i \subseteq A \subseteq U_i \cup D^c$$

and:

$$\begin{aligned}\|1\{U_i \cup D^c\} - 1\{L_i\}\|_{L_1(P)} &\leq \|1\{U_i\} - 1\{L_i\}\|_{L_1(P)} + \|1\{D^c\}\|_{L_1(P)} \\ &\leq C(\|1\{U_i\} - 1\{L_i\}\|_{L_1(\lambda)} + \|1\{D^c\}\|_{L_1(\lambda)}) \leq \varepsilon.\end{aligned}$$

Thus the set $\{[L_i, U_i \cup D^c]\}$ is the set of ε-brackets for our class $\mathcal{A}$ in $L_1(P)$. This implies that $\mathcal{A}$ is a Glivenko-Cantelli class and the statement follows from Theorem 2.4.1 van der Vaart and Wellner [1996]. $\square$

2. Consistency of the MLE for an increasing model. To prove consistency for increasing models we begin with a general property of lower layer sets (see Dudley [1999], Chapter 8.3). Recall that a lower lay set $B \subset \mathbb{R}^d$ is a set satisfying $y \leq x$ coordinate-wise with $x \in B$ implies $y \in B$.

LEMMA 2.1. *Let $\mathcal{LL}$ be the class of closed lower layer sets in $\mathbb{R}^d_+$ and P be a Lebesgue absolutely continuous probability measure with bounded density. Then:*

$$\|\mathbb{P}_n - P\|_{\mathcal{LL}} \to_{a.s.} 0.$$

PROOF. By Theorem 8.3.2 Dudley [1999] we have

$$\log N_{[]}(\varepsilon, \mathcal{LL} \cap [0,1]^d, L_1(\mu)) < +\infty.$$

Since the class $\mathcal{LL}$ is invariant under rescaling, the result follows from Lemma 1.13 $\square$

Note that Lemma 1.1 implies that if $h \circ g$ belongs to an increasing model $\mathcal{P}(h)$ then $(\operatorname{lev}_y g)^c$ is a lower layer set and has Lebesgue measure less or equal than $1/h(y)$. Let us denote by A_δ the set $\{V(x) \leq \delta, x \in \mathbb{R}^d_+\}$. Then by Lemma 1.1 part 3 we have:

$$(\operatorname{lev}_y g)^c \subset A_{c/h(y)}, \tag{2.4}$$

for $c \equiv d!/d^d$.

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

PROOF. By Assumption M.I.6 and Lemma 1.12 it is enough to show that:

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

Indeed, applying Lemma 1.2 for the increasing transformation $\log[\varepsilon+h(y)]-\log\varepsilon$ we obtain:

$$\begin{aligned}
&\int \log[\varepsilon + h\circ\hat{g}_n(x)]d(\mathbb{P}_n(x)-P_0(x))\\
&=\int_{-\infty}^{+\infty}\left[\frac{h'(z)}{\varepsilon+h(z)}\right](\mathbb{P}_n-P_0)\left((\operatorname{lev}_z\hat{g}_n)^c\right)dz\\
&\le \|\mathbb{P}_n-P_0\|_{\mathcal{LC}}\int_{-\infty}^{M}\left[\frac{h'(z)}{\varepsilon+h(z)}\right]dz+\int_{M}^{+\infty}\left[\frac{h'(z)}{\varepsilon+h(z)}\right]|\mathbb{P}_n-P_0|\left((\operatorname{lev}_z\hat{g}_n)^c\right)dz\\
&\le \|\mathbb{P}_n-P_0\|_{\mathcal{LC}}\log\left[\frac{\varepsilon+h(M)}{\varepsilon}\right]+\int_{M}^{+\infty}\left[\frac{h'(z)}{\varepsilon+h(z)}\right](\mathbb{P}_n+P_0)\left((\operatorname{lev}_z\hat{g}_n)^c\right)dz.
\end{aligned}$$

The first converges to zero almost surely by Lemma 2.1. For the second term we will use the inclusion 2.4:

$$\int_{M}^{+\infty}\left[\frac{h'(z)}{\varepsilon+h(z)}\right](\mathbb{P}_n+P_0)(\operatorname{lev}_z\hat{g}_n)^c dz\le\int_{M}^{+\infty}\left[\frac{h'(z)}{\varepsilon+h(z)}\right](\mathbb{P}_n+P_0)A_{c/h(z)}\,dz.$$

Now, we can apply Lemma 1.2 again for $g_A(x)=h^{-1}(c/V(x))$. We have $(\operatorname{lev}_z g_A)^c=A_{c/h(z)}$ and therefore:

$$\begin{aligned}
\int_{M}^{+\infty}\left[\frac{h'(z)}{\varepsilon+h(z)}\right](\mathbb{P}_n+P_0)A_{c/h(z)}\,dz&=\int_{A_{c/h(M)}}\log(\varepsilon+c/V(x))d(\mathbb{P}_n+P_0)\\
&\le\int_{A_{c/h(M)}}\log(2c/V(x))d(\mathbb{P}_n+P_0),
\end{aligned}$$

for M large enough. Assumption M.I.5 and the SLLN imply that:

$$\int_{A_{c/h(M)}}\log(2c/V(x))d(\mathbb{P}_n+P_0)\to_{a.s.}2\int_{A_{c/h(M)}}\log(2c/V(x))dP_0.$$

Since M is arbitrary and $A_{c/h(M)}\downarrow\{0\}$ as $M\to+\infty$ the result follows. □

3. Lower bounds.

3.1. *Local deformations.*

LEMMA 3.1 (M.3.18). *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 $r>0$:*

$$\lim_{\varepsilon\to0}\int_{\mathbb{R}^d}|g_\varepsilon(x)-g(x)|^r dx=0, \tag{3.5}$$

$$\lim_{\varepsilon\to0}\frac{\int_{\mathbb{R}^d}|h\circ g_\varepsilon(x)-h\circ g(x)|^r dx}{\int_{\mathbb{R}^d}|g_\varepsilon(x)-g(x)|^r dx}=|h'\circ g(x_0)|^r. \tag{3.6}$$

PROOF. Since $\{g_\varepsilon\}$ is a local deformation, for $\varepsilon > 0$ small enough we have:

$$\int_{\mathbb{R}^d} |h \circ g_\varepsilon(x) - h \circ g(x)|^r dx = \int_{B(x_0;r_\varepsilon)} |h \circ g_\varepsilon(x) - h \circ g(x)|^r dx,$$
$$\int_{\mathbb{R}^d} |g_\varepsilon(x) - g(x)|^r dx = \int_{B(x_0;r_\varepsilon)} |g_\varepsilon(x) - g(x)|^r dx.$$

Then: $\int_{B(x_0;r_\varepsilon)} |g_\varepsilon - g|^r dx \le \operatorname{ess\,sup} |g_\varepsilon - g|^r \mu[B(x_0; r_\varepsilon)]$ implies (3.5).

Let us define a sequence $\{a_\varepsilon\}$:

$$a_\varepsilon \equiv \operatorname{ess\,sup} |g_\varepsilon - g| + \sup_{x \in B(x_0;r_\varepsilon)} |g(x) - g(x_0)|.$$

For $x \in B(x_0; r_\varepsilon)$ and $y \in [g_\varepsilon(x), g(x)]$ we have a.e.:

$$|y - g(x_0)| \le |g_\varepsilon(x) - g(x)| + |g(x) - g(x_0)| \le a_\varepsilon.$$

Using the mean value theorem we obtain:

$$\int_{\mathbb{R}^d} |h \circ g_\varepsilon(x) - h \circ g(x)|^r dx = \int_{\mathbb{R}^d} |h'(y_x)|^r |g_\varepsilon(x) - g(x)|^r dx$$
$$\inf_{y \in B(g(x_0);a_\varepsilon)} |h'(y)|^r \le \frac{\int_{\mathbb{R}^d} |h \circ g_\varepsilon(x) - h \circ g(x)|^r dx}{\int_{\mathbb{R}^d} |g_\varepsilon(x) - g(x)|^r dx} \le \sup_{y \in B(g(x_0);a_\varepsilon)} |h'(y)|^r$$

Since h' is continuous at $g(x_0)$, to prove (3.6) it is enough to show that $a_\varepsilon \to 0$. By assumption we have: $\lim_{\varepsilon \to 0} \operatorname{ess\,sup} |g_\varepsilon - g| = 0$. Since g is continuous at x_0 and $r_\varepsilon \to 0$ we have: $\lim_{\varepsilon \to 0} \sup_{x \in B(x_0;r_\varepsilon)} |g(x) - g(x_0)| = 0$. Thus $a_\varepsilon \to 0$, which proves (3.6). □

LEMMA 3.2 (M.3.19). *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:*

$$\limsup_{\theta \to 0} \theta^{-r} \int_{\mathbb{R}^d} |h \circ g_{\theta,\delta}(x) - h \circ g(x)|^r dx < \infty, \tag{3.7}$$

$$\liminf_{\theta \to 0} \theta^{-r} \int_{\mathbb{R}^d} |h \circ g_{\theta,\delta}(x) - h \circ g(x)|^r dx > 0. \tag{3.8}$$

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

PROOF. The statement follows from the argument for Lemma 3.1. For a fixed θ the family $\{g_{\theta,\delta}\}$ is a local deformation of g. Thus for $a_{\theta,\varepsilon}$ defined by: $a_{\theta,\varepsilon} \equiv \operatorname{ess\,sup} |g_{\theta,\varepsilon} - g| + \sup_{x \in B(x_0; r_\varepsilon)} |g(x) - g(x_0)|$, it follows that

$$\frac{\int_{\mathbb{R}^d} |h \circ g_{\theta,\varepsilon}(x) - h \circ g(x)|^r dx}{\int_{\mathbb{R}^d} |g_{\theta,\varepsilon}(x) - g(x)|^r dx} \leq \sup_{y \in B(g(x_0); a_{\theta,\varepsilon})} |h'(y)|^r,$$
$$\frac{\int_{\mathbb{R}^d} |h \circ g_{\theta,\varepsilon}(x) - h \circ g(x)|^r dx}{\int_{\mathbb{R}^d} |g_{\theta,\varepsilon}(x) - g(x)|^r dx} \geq \inf_{y \in B(g(x_0); a_{\theta,\varepsilon})} |h'(y)|^r.$$

For $|\theta| < 1$ we have: $|g_{\theta,\varepsilon} - g| = |\theta||g_\varepsilon - g|$ and therefore $a_{\theta,\delta} \leq a_\delta$. Since $a_\varepsilon \to 0$ and h is continuously differentiable for all $\delta > 0$ small enough we have:

$$\sup_{y \in B(g(x_0); a_{\theta,\delta})} |h'(y)|^r \leq \sup_{y \in B(g(x_0); a_\delta)} |h'(y)|^r < \infty,$$
$$\inf_{y \in B(g(x_0); a_{\theta,\delta})} |h'(y)|^r \geq \inf_{y \in B(g(x_0); a_\delta)} |h'(y)|^r > 0.$$

Thus for all θ we obtain:

$$\theta^{-r} \int_{\mathbb{R}^d} |h \circ g_{\theta,\delta} - h \circ g|^r d\mu \leq \sup_{y \in B(g(x_0); a_\delta)} |h'(y)|^r \int_{\mathbb{R}^d} |g_\delta - g|^r d\mu < \infty,$$
$$\theta^{-r} \int_{\mathbb{R}^d} |h \circ g_{\theta,\delta} - h \circ g|^r d\mu \geq \inf_{y \in B(g(x_0); a_\delta)} |h'(y)|^r \int_{\mathbb{R}^d} |g_\delta - g|^r d\mu > 0$$

which proves the lemma. □

LEMMA 3.3 (M.3.22). *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:*

$$g_\varepsilon^+ = (1 - \theta_\varepsilon^+) D_\varepsilon(g; x_0, v_0) + \theta_\varepsilon^+ D_\delta^*(g; x_1)$$
$$g_\varepsilon^- = (1 - \theta_\varepsilon^-) D_\varepsilon^*(g; x_0) + \theta_\varepsilon^- D_\delta(g; x_1; v_1)$$

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

PROOF. By dominated convergence, the function $F(\theta)$ defined by:

$$F(\theta) = \int h \circ ((1 - \theta) D_\varepsilon(g; x_0, v_0) dx + \theta D_\delta^*(g; x_1))\, dx$$

is continuous. We have:

$$F(0) = \int h \circ D_\varepsilon(g; x_0, v_0) dx > \int h \circ g dx = 1,$$
$$F(1) = \int h \circ D_\delta^*(g; x_1) dx < \int h \circ g dx = 1.$$

Therefore there exists $\theta_\varepsilon^+ \in (0, 1)$ such that $F(\theta_\varepsilon^+) = 1$. □

3.2. *Mode estimation.*

THEOREM 3.4 (M.2.26). *Let h be a decreasing transformation, $h \circ g \in \mathcal{P}(h)$ be a convex-transformed density and a point $x_0 \in \mathrm{ri}(\mathrm{dom}\, g)$ 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 $\mathrm{curv}_{x_0} g > 0$. In addition let us assume that g is locally Hölder continuous at x_0: $|g(x) - g(x_0)| \leq L\|x - x_0\|^\gamma$ with respect to some norm $\|\cdot\|$. Then, for the functional $T(h \circ g) \equiv \operatorname{argmin} g$ there exists a sequence $\{p_n\} \in \mathcal{P}(h)$ such that:*

(3.9)

$$\liminf_{n\to\infty} n^{\frac{2}{\gamma(d+4)}} R_s(n; T, \{p, p_n\}) \geq C(d) L^{-\frac{1}{\gamma}} \left[\frac{h \circ g(x_0)^2 \,\mathrm{curv}_{x_0} g}{h' \circ g(x_0)^4}\right]^{\frac{1}{\gamma(d+4)}},$$

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

PROOF. The proof is similar to the proof for a point estimation lower bounds. The deformation we will construct will resemble g^-_ε.

Our statement is not trivial only if the curvature $\mathrm{curv}_{x_0} g > 0$ or equivalently there exists such positive definite $d \times d$ matrix G so that the function g is locally G-strongly convex. For $a > 0$ small enough $h' \circ g(x)$ is negative and decomposition (M.3.16) is true on $B(x_0; a)$. Let us fix some $v_0 \in \partial g(x_0)$, some $x_1 \in B(x_0; a)$ such that $x_1 \neq x_0$ and some $y_1 \in \partial g(x_1)$. We fix δ such that equation (M.3.14) of Lemma M.3.19 is true for the transformation $\sqrt{h}$ and $r = 2$ and also $x_0 \notin \overline{B_G(x_1; \sqrt{2\delta})}$.

Let us consider the deformation $D^*_{\xi(\varepsilon)}(g; x_0 + \varepsilon u)$ where $u \in \mathbb{R}^d$ is an arbitrary fixed vector in $\mathbb{R}^d$ with $\|u\| = 1$ and

$$\xi(\varepsilon) = g(x_0) - g(x_0 + \varepsilon u) + \varepsilon^{\gamma+1}.$$

Since the value of $D^*_{\xi(\varepsilon)}(g; x_0 + \varepsilon u)$ at any point x is a convex combination of $g(y)$ for some y, $g(x) \geq g(x_0)$ and

$$D^*_{\xi(\varepsilon)}(g; x_0 + \varepsilon u)(x_0 + \varepsilon u) = g(x_0) + \varepsilon^{\gamma+1}$$

the global minimum of $D^*_{\xi(\varepsilon)}(g; x_0 + \varepsilon u)$ is $x_0 + \varepsilon u$. By Lemma M.3.21 for all $\varepsilon > 0$ small enough we have

$$\mathrm{supp}[D^*_{\xi(\varepsilon)}(g; x_0 + \varepsilon u) - g] \subseteq B_G(x_0 + \varepsilon u, \sqrt{2\xi(\varepsilon)}).$$

Since, by assumption

$$\xi(\varepsilon) \leq L\varepsilon^\gamma + \varepsilon^{\gamma+1}$$

the support of $\operatorname{supp}[D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u) - g]$ converges to a point x_0 and thus does not intersect $\operatorname{supp}[D_\varepsilon(g; x_1, y_1) - g]$ for ε small enough i.e. these two deformations do not interfere.

The same argument as in Lemma M.3.22 shows that there exists $\theta^m_\varepsilon \in (0,1)$ such that the deformation g^m_ε defined as:

$$g^m_\varepsilon = (1-\theta^m_\varepsilon) D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u) + \theta^m_\varepsilon D_\delta(g; x_1, y_1)$$

belongs to $\mathcal{P}(h)$. Also $g^m_\varepsilon \geq D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u)$ and the global minimum of g^m_ε is $x_0+\varepsilon u$. We have:

$$\varepsilon^{-1} s(Tg^m_\varepsilon, Tg) \equiv 1.$$

Next, we will show that θ^m_ε goes to zero fast enough so that g^m_ε is very close to $D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u)$. We have:

$$\begin{aligned}
0 &= \int (h\circ g^m_\varepsilon - h\circ g)dx \\
&= -\int \Big(h\circ g - h\circ((1-\theta^m_\varepsilon)D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u) + \theta^m_\varepsilon g)\Big)\, dx \\
&\quad + \int \left(h\circ(\theta^m_\varepsilon D_\delta(g; x_1, y_1) + (1-\theta^m_\varepsilon)g) - h\circ g\right) dx,
\end{aligned}$$

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

$$\begin{aligned}
&\int \Big|h\circ g - h\circ((1-\theta^m_\varepsilon)D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u) + \theta^m_\varepsilon g)\Big|\, dx, \\
&\leq \int \Big|h\circ g - h\circ D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u)\Big|\, dx \sim \int \Big|D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u) - g\Big|\, dx \\
&\leq \xi(\varepsilon)\mu[B_G(x_0+\varepsilon u; \sqrt{2\xi(\varepsilon)})] = O(\xi(\varepsilon)^{1+d/2})
\end{aligned}$$

The second integral is monotone in θ^m_ε and by Lemma M.3.19 we have:

$$\int \left(h\circ(\theta^m_\varepsilon D_\delta(g; x_1, y_1) + (1-\theta^m_\varepsilon)g) - h\circ g\right) dx \sim \theta^m_\varepsilon,$$

thus we have $\theta^+_\varepsilon = O(\varepsilon^{\gamma(1+d/2)})$.

For Hellinger distance we have:

$$\begin{aligned}
H(h\circ g^m_\varepsilon, h\circ g) &= H(h\circ((1-\theta^m_\varepsilon)D^*_{\xi(\varepsilon)}(g; x_0+\varepsilon u) + \theta^m_\varepsilon g), h\circ g) \\
&\quad + H(h\circ(\theta^m_\varepsilon D_\delta(g; x_1, y_1) + (1-\theta^m_\varepsilon)g)), h\circ g).
\end{aligned}$$

For the first part we can apply Lemma M.3.18:

$$H^2(h\circ((1-\theta_\varepsilon^m)D^*_{\xi(\varepsilon)}(g;x_0+\varepsilon u)+\theta_\varepsilon^m g),h\circ g)\le H^2(h\circ D^*_{\xi(\varepsilon)}(g;x_0+\varepsilon u),h\circ g)$$

$$\lim_{\varepsilon\to 0}\frac{H^2(h\circ D^*_{\xi(\varepsilon)}(g;x_0+\varepsilon u),h\circ g)}{\int(D^*_{\xi(\varepsilon)}(g;x_0+\varepsilon u)-g)^2dx}=\frac{h'\circ g(x_0)^2}{4h\circ g(x_0)}\quad\text{and}$$

$$\begin{aligned}\int(D^*_{\xi(\varepsilon)}(g;x_0+\varepsilon u)-g)^2dx &\le \xi(\varepsilon)^2\mu[B_G(x_0;\sqrt{2\xi(\varepsilon)})]\\ &=\xi(\varepsilon)^{2+d/2}\frac{2^{d/2}\mu[S(0,1)]}{\sqrt{\det G}}\end{aligned}$$

which gives:

$$\begin{aligned}&\limsup_{\varepsilon\to 0}\varepsilon^{-\gamma(1+d/4)}H(h\circ((1-\theta_\varepsilon^m)D^*_{\xi(\varepsilon)}(g;x_0+\varepsilon u)+\theta_\varepsilon^m g),h\circ g)\\ &\quad\le C(d)L^{1+d/4}\left[\frac{h'\circ g(x_0)^4}{h\circ g(x_0)^2\det G}\right]^{1/4}\end{aligned}$$

where $S(0,1)$ is d-dimensional sphere of radius 1.
For the second part by Lemma M.3.19 we obtain:

$$\begin{aligned}&\limsup_{\varepsilon\to 0}(\theta_\varepsilon^+)^{-2}H^2(h\circ((1-\theta_\varepsilon^+)g+\theta_\varepsilon^+D_\delta(g;x_1,y_1)),h\circ g)<\infty\\ &H(h\circ((1-\theta_\varepsilon^+)g+\theta_\varepsilon^+D_\delta(g;x_1,y_1)),h\circ g)=O(\varepsilon^{\gamma(1+d/2)}).\end{aligned}$$

Thus:

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

Finally, we apply Corollary M.2.21:

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

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

APPENDIX A: SOME RESULTS FROM CONVEX ANALYSIS

We will use the following general properties of convex sets and convex functions. We use Rockafellar [1970] as a reference.

LEMMA A.1. *For any convex set A in $\mathbb{R}^d$ we have:*

1. *The boundary of A has Lebesgue measure zero.*
2. *A has Lebesgue measure zero if and only if it belongs to a $d-1$ dimensional affine subspace.*
3. *A has Lebesgue measure $+\infty$ if and only if it is unbounded and has dimension d.*

PROOF.

1. If A is such that $\mathrm{cl}(A)$ has finite Lebesgue measure then:
$$\partial A \subseteq (1+\varepsilon)\,\mathrm{cl}(A) \setminus (1-\varepsilon)\,\mathrm{cl}(A), \quad \varepsilon \in (0,1)$$
$$\mu[\partial A] \leq 2\varepsilon \mu[\mathrm{cl}(A)]$$
and thus $\mu[\partial A] = 0$. Since $\mathbb{R}^d$ is a countable union of closed convex cubes B_i the result for an arbitrary convex set A follows from:
$$\partial A \subseteq \bigcup_i \partial(A \cap B_i).$$
2. If A has dimension $k \leq d$ then its affine hull V has dimension k and A contains a k-dimensional simplex D (Theorem 2.4 Rockafellar [1970]). Then if $k = d$ we have $\mu[D] > 0$ and if $k < d$ we have $\mu[V] = 0$.
3. Part 1 implies that it is enough to consider closed convex sets. Part 2 implies that it is enough to prove that an unbounded closed convex set of dimension d has Lebesgue measure $+\infty$. Let A be such a set; i.e. an unbounded closed convex set. Then A contains d-dimensional simplex D (Theorem 2.4 Rockafellar [1970]) which has non-zero Lebesgue measure. Since A is unbounded then its recession cone is non-empty (Theorem 8.4 Rockafellar [1970]) and therefore we can choose a direction v such that $D + \lambda v \subset A$ for all $\lambda \geq 0$ which implies $\mu[A] = +\infty$.

□

The following lemma shows that convergence of convex sets in measure implies pointwise convergence.

LEMMA A.2. *Let A be a convex set in $\mathbb{R}^d$ such that $\dim(A) = d$ and $\mathrm{ri}(A) \neq \emptyset$. Then:*

1. *Suppose a sequence of convex sets B_n is such that $A \subseteq B_n$ and $\lim \mu[B_n \setminus A] = 0$ then $\limsup \mathrm{cl}(B_n) = \mathrm{cl}(A)$;*
2. *Suppose a sequence of convex sets B_n is such that $C_n \subseteq A$ and $\lim \mu[A \setminus C_n] = 0$ then $\liminf \mathrm{ri}(C_n) = \mathrm{ri}(A)$.*

PROOF. By Lemma A.1 we can assume that A, B_n and C_n are closed convex sets.

1. If on the contrary, there exists a subsequence $\{k\}$ such that for some $x \in A^c$ we have $x \in \cap_{k\geq 1} B_k$ then for $xA = \text{conv}(\{x\} \cup A)$ we have:
$$\begin{aligned}&xA \subseteq B_k\\&\mu[B_k \setminus A] \geq \mu[xA \setminus A].\end{aligned}$$
Since A is closed there exists a ball $B(x)$ such that $B(x) \cap A = \emptyset$. Since $\text{ri}(A) \neq \emptyset$ there exists a ball $B(x_0)$ such that $B(x_0) \subseteq A$ for some $x_0 \in \text{ri}(A)$. Then for $xB = \text{conv}(\{x\} \cup B(x_0))$ we have:
$$\begin{aligned}&xB \subseteq xA\\&\mu[xA \setminus A] \geq \mu[xB \cap B(x)] > 0.\end{aligned}$$
This contradiction implies $\limsup B_i = A$.
2. If on the contrary, there exists a point $x \in \text{ri}(A)$ and subsequence $\{k\}$ such that $x \notin C_k$ for all k then for each C_k there exists a half-space L_k such that $x \in L_k$ and $C_k \subseteq L_k^c$. Let $B(x)$ be a ball such that $B(x) \subseteq A$. We have:
$$\mu[A \setminus C_i] \geq \mu[A \cap L_k] \geq \mu[B(x) \cap L_k] = \mu[B(x)]/2 > 0.$$
This contradiction implies $\text{ri}(A) \subseteq \liminf C_i$.

□

Our next lemma shows that the Lebesgue measure of sublevel sets of a convex function grows at most polynomially.

LEMMA A.3. *Let g be a convex function and values $y_1 < y_2 < y_3$ are such that $\text{lev}_{y_1} g \neq \emptyset$. Then we have:*

$$\mu[\text{lev}_{y_3} g] \leq \left[\frac{y_3 - y_1}{y_2 - y_1}\right]^d \mu[\text{lev}_{y_2} g]. \tag{A.10}$$

PROOF. By assumption we have:
$$\mu[\text{lev}_{y_3} g] \geq \mu[\text{lev}_{y_2} g] \geq \mu[\text{lev}_{y_1} g] > 0.$$
Let us consider the set L defined as:
$$L = \{x_1 + k(x - x_1) \,|\, x \in \text{lev}_{y_2} g\},$$

where x_1 is any fixed point such that $g(x_1) = y_1$ and

$$k = \frac{y_3 - y_1}{y_2 - y_1} > 1.$$

Then:

$$\mu[L] = k^d \mu[\operatorname{lev}_{y_2} g].$$

and therefore it is enough to prove that $\operatorname{lev}_{y_3} g \subseteq L$.

If $x_3 \in \operatorname{lev}_{y_3} g$ then for $x_2 = x_1 + (x_3 - x_1)/k$ we have:

$$\begin{aligned} &x_3 = x_1 + k(x_2 - x_1), \\ &g(x_2) \leq (1 - 1/k)g(x_1) + (1/k)g(x_3) = y_2 \end{aligned}$$

and thus $x_2 \in \operatorname{lev}_{y_2} g$. □

COROLLARY A.4. *If g is a convex function then function $\mu[\operatorname{lev}_y g]$ is continuous on $(\inf g, \sup g)$.*

A.1. Maximal convex minorant. In this section we describe the convex function f_c which is in some sense the closest to a given function f.

DEFINITION A.5. *The maximal convex minorant f_c of a proper function f is a supremum of all linear functions l such that $l \leq f$.*

It is possible that f_c does not majorate any linear function and then $f_c = -\infty$. However if it is not the case the following properties of the maximal convex minorant hold. Recall that for any function f, the convex conjugate f^* of f is defined by $f^*(y) \equiv \sup_{x \in \mathbb{R}^d} (\langle y, x \rangle - f(x))$.

LEMMA A.6. *Let f be a function and $f_c \neq -\infty$ its maximal convex minorant. Then:*

1. *f_c is a closed proper convex function;*
2. *if f is proper convex function then f_c is its closure;*
3. *$f_c \leq f$;*
4. *$(f_c)^*(y) = f^*(y)$.*

PROOF. This follows from Corollary 12.1.1 Rockafellar [1970]. □

The maximal convex minorant allows us to see an important duality between operations of pointwise minimum and pointwise maximum.

Lemma A.7. *Let f_i be a proper convex functions and let $g = \inf_i f_i$ be the pointwise infinum of f_i. Then $(g_c)^* = \sup_i f_i^*$.*

Proof. This follows from Theorem 16.5 Rockafellar [1970]. □

A.2. Subdifferential.

Definition A.8. *The subdifferential $\partial h(x)$ of a convex function h at the point x is the set of all vectors v which satisfy the inequality*

$$h(z) \geq \langle v, z - x\rangle + h(x) \qquad \textit{for all} \;\; x.$$

Obviously $\partial h(x)$ is a closed convex set. It might be empty, but if it is not, the function h is called *subdifferentiable* at x.

Lemma A.9. *Let h be a proper convex function then for $x \in \operatorname{ri}\operatorname{dom} h$ subdifferential $\partial h(x)$ is not empty.*

Proof. This follows from Theorem 23.4 Rockafellar [1970]. □

Lemma A.10. *Let h be a closed proper convex function. Then the following conditions on x and x^* are equivalent:*

1. *$x^* \in \partial h(x)$;*
2. *$l(z) = \langle x^*, z\rangle - h^*(x^*)$ is a support plane for $\operatorname{epi}(h)$ at x;*
3. *$h(x) + h^*(x^*) = \langle x^*, x\rangle$;*
4. *$x \in \partial h^*(x^*)$;*
5. *$l(z) = \langle x, z\rangle - h(x)$ is a support plane for $\operatorname{epi}(h^*)$ at x^*;*

Proof. This follows from Theorem 23.5 Rockafellar [1970]. □

Lemma A.11. *Let h_1 and h_2 be proper convex functions such that $\operatorname{ri}\operatorname{dom} h_1 \cap \operatorname{ri}\operatorname{dom} h_2 \neq \emptyset$. Then $\partial(h_1 + h_2) = \partial h_1 + \partial h_2$ for all x.*

Proof. This follows from Theorem 23.8 Rockafellar [1970]. □

A.3. Polyhedral functions.

Definition A.12. *A polyhedral convex set is a set which can be expressed as an intersection of finitely many half-spaces. A polyhedral convex function is a convex function whose epigraph is polyhedral.*

From Theorem 19.1 Rockafellar [1970] we have that the epigraph of the polyhedral function $h : \mathbb{R}^d \to \mathbb{R}$ has finite number of extremal points and faces. We call projections of extremal points the *knots* of h and projections of the nonvertical d-dimensional faces the *facets* of h. Thus the set of knots and the set of facets of polyhedral function are always finite. Moreover, by Theorem 18.3 Rockafellar [1970] the knots are the extremal points of the facets. Finally, let $\{C_i\}$ be the set of facets of a polyhedral function h then:

$$\begin{aligned} &\operatorname{dom} h = \bigcup_i C_i \\ &\operatorname{ri}(C_i) \cap \operatorname{ri}(C_j) = \emptyset, \end{aligned}$$

and on $\operatorname{dom} h$ we have $h = \max(l_i)$ where l_i are linear functions. For each C_i there exists l_i such that:

$$C_i = \{x \,|\, h(x) = l_i(x)\}.$$

LEMMA A.13. *Let f be a polyhedral convex function and $x \in \operatorname{dom} h$ then $\partial h(x) \neq \emptyset$.*

PROOF. This follows from Theorem 23.10 Rockafellar [1970]. □

LEMMA A.14. *For the set of points $x = \{x_i\}_{i=1}^n$ such that $x_i \in \mathbb{R}^d$ and any point $p \in \mathbb{R}^n$ consider a family of all convex functions h with $\operatorname{ev}_x h = p$. The unique maximal element U_x^p in this family is a polyhedral convex function with domain $\operatorname{dom} U_x^p = \operatorname{conv}(x)$ and the set of knots $K \subseteq x$.*

PROOF. Points (x_i, p_i) and direction $(0, 1)$ belong to the epigraph of any convex function h in our family and so does convex hull U of these points and direction. By construction U is an epigraph of some closed proper convex function U_x^p such that $\operatorname{dom} U_x^p = \operatorname{conv}(x)$, by Theorem 19.1 Rockafellar [1970] this function is polyhedral, by Corollary 18.3.1 Rockafellar [1970] the set of its knots K belongs to x and since $\operatorname{epi}(U_x^p) = U \subseteq \operatorname{epi}(h)$ we have $h \leq U_x^p$. On the other hand, since $(x_i, p_i) \in U$ we have

$$p_i = h(x_i) \leq U_x^p(x_i) \leq p_i$$

and therefore $U_x^p(x_i) = p_i$ which proves the lemma. □

LEMMA A.15. *For the set of points $x = \{x_i\}_{i=1}^n$, convex set C such that $x_i \in \operatorname{ri}(C)$ and any point $p \in \mathbb{R}^n$ consider a family of all convex functions h with $\operatorname{ev}_x h = p$ and $C \subseteq \operatorname{dom} h$. Any minimal element L_x^p in this family is a polyhedral convex function with $\operatorname{dom} L_x^p = \mathbb{R}^d$. For each facet C of L_x^p, $\operatorname{ri}(C)$ contains at least one element of x.*

PROOF. For any function h in our family let us consider the set of linear functions l_i such that $l_i(x_i) = h(x_i) = p_i$ and $l_i \leq h$ and which correspond to arbitrarily chosen nonvertical support planes for $\mathrm{epi}(h)$ at x_i. Then $L = \max(l_i)$ is polyhedral and since $l_j(x_i) \leq h(x_i) = p_i$ we have $L(x_i) = p_i$. We also have $\mathrm{dom}\, L = \mathbb{R}^d$. If the interior of any facet C_i of L does not contain elements of x we can exclude corresponding linear function l_i from maximum. For the new polyhedral function $L' = \max_{j \neq i} l_j$ we still have $\mathrm{ev}_x L' = p$. Now, we repeat this procedure until interior of each facet contains at least one element of x and denote the function we obtained by L_x^p. If a closed proper convex function h is such that $\mathrm{ev}_x h = p$ and $h \leq L_x^p$, then consider for any facet C_i and corresponding linear function l_i we have $h \leq l_i$ on C_i and the supremum of h on the convex set C_i is obtained in interior point $x_j \in x$. By Theorem 32.1 Rockafellar [1970] $h \equiv L_x^p$ on C_i. Thus $h \equiv L_x^p$ and L_x^p is the minimal element of our family. □

LEMMA A.16. *For linear function $l(x) = a^T x + b$ the polyhedral set $A = \{l \geq c\} \cap \mathbb{R}_+^d$ is bounded if and only if all coordinates of a are negative. In this case, if $b \geq c$ the set A is a simplex with vertices $p_i = ((c-b)/a_i)e_i$ and 0, where e_i are basis vectors. Otherwise, A is empty.*

PROOF. If coordinate a_i is nonnegative then the direction $\{\lambda e_i\}$, $\lambda > 0$ belongs to the recession cone of A and thus it is unbounded. If all coordinates a_i are negative and $b \leq c$ the set A is either empty or consists of zero vector 0. Finally, if a_i are negative and $b > c$ then for $x \in A$ we can define $\theta_i = a_i x_i/(c-b) > 0$. Then $1 \geq \sum_i \theta_i$ and $x = \sum_i \theta_i p_i$, which proves that A is simplex. □

A.4. Strong convexity. Following Rockafellar and Wets [1998] page 565 we say that a proper convex function $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ is *strongly convex* if there exists a constant σ such that:

$$h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta)h(y) - \frac{1}{2}\sigma\theta(1-\theta)\|x-y\|^2 \tag{A.11}$$

for all x, y and $\theta \in (0,1)$. There is a simple characterization of strong convexity:

LEMMA A.17. *A proper convex function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is strongly convex if and only if the function $f(x) - \frac{1}{2}\sigma\|x\|^2$ is convex.*

Since we need a more precise control over the curvature of a convex function we define a generalization of strong convexity based on the characterization above:

Definition A.18. *We say that a proper convex function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ *is* G*-strongly convex if there exists a point* x_0*, a positive semidefinite* $d \times d$ *matrix* G *and a convex function* q *such that:*

$$h(x) = \frac{1}{2}(x - x_0)^T G(x - x_0) + q(x) \qquad \textit{for all} \;\; x. \tag{A.12}$$

Obviously, strong convexity is equivalent to σI-strong convexity where I denotes the $d \times d$ identity matrix. Note that the definition does not depend on the choice of x_0.

Definition A.19. *We say that a proper convex function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ *is locally* G*-strongly convex at a point* x_0 *if there exist an open neighborhood of* x_0*, a positive semidefinite* $d \times d$ *matrix* G *and a convex function* q *such that* (A.12) *holds for any* x *in this neighborhood.*

We can relate G-strong convexity to the Hessian of a smooth convex function:

Lemma A.20. *If a proper convex function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ *is continuously twice differentiable at* x_0 *then* h *is locally* $(1-\varepsilon)\nabla^2 h$*-strongly convex for any* $\varepsilon \in (0, 1)$.

The last result suggests the following definition:

Definition A.21. *For a proper convex function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ *we define a curvature* $\operatorname{curv}_{x_0} h$ *at a point* x_0 *as:*

$$\operatorname{curv}_{x_0} h = \sup_{G \in \mathcal{SC}(h; x_0)} \det(G) \tag{A.13}$$

where $\mathcal{SC}(h; x_0)$ *is the set of all positive semidefinite matrices* G *such that* h *is locally* G*-strong convex at* x_0.

Lemma A.20 implies that:

Lemma A.22. *If a proper convex function* $h : \mathbb{R}^d \to \overline{\mathbb{R}}$ *is continuously twice differentiable at* x_0 *and Hessian* $\nabla^2 h(x_0)$ *is positive definite then*

$$\operatorname{curv}_{x_0} h = \det(\nabla^2 h(x_0)). \tag{A.14}$$

NOTATION

$$
\begin{array}{lcl}
\mathbb{R} & = & (-\infty, +\infty) \\
\overline{\mathbb{R}} & = & [-\infty, +\infty] \\
\mathbb{R}_+ & = & [0, +\infty) \\
\overline{\mathbb{R}}_+ & = & [0, +\infty] \\
V(x) & = & \prod_{k=1}^d x_k, \quad x \in \mathbb{R}_+^d \\
\mathcal{C} & = & \{f : \mathbb{R}^d \to \overline{\mathbb{R}} \,|\, f \text{ closed proper convex function}\} \\
\mathcal{D} & = & \{p : \mathbb{R}^d \to \overline{\mathbb{R}} \,|\, p \text{ density}\} \\
\mathcal{G}(h) & = & \{h : | \, g \in \mathcal{C},\, h \circ g \in \mathcal{D}\} \\
\mathbb{L}_n g & = & \mathbb{P}_n h \circ g \\
\operatorname{ev}_x f & = & (f(x_1), \dots f(x_n)), \quad x_i \in \mathbb{R}^d \\
\operatorname{supp}(f) & = & \{x \,|\, f(x) \neq 0\} \\
\delta(\cdot \,|\, C) & = & \infty \cdot 1_{C^c} + 0 \cdot 1_C \\
\operatorname{lev}_y g & = & \{x \,|\, g(x) \leq y\} \\
\mu[S] & = & \text{Lebesgue measure of } S \\
\{f \lesseqqgtr a\} & = & \{x \in X \,|\, f(x) \lesseqqgtr a\} \\
B(x_0; r) & = & \{x : \|x - x_0\| < r\} \\
B_H(x_0; r) & = & \{x : (x - x_0)^T H (x - x_0) < r^2\} \\
\operatorname{curv}_x h & = & \text{curvature of a convex function } h \text{ at a point } x \\
\operatorname{ri}(A) & = & \text{the relative interior of the set } A
\end{array}
$$

REFERENCES

DUDLEY, R. M. (1999). *Uniform central limit theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

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

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*, vol. 54 of *IMS Lecture Notes-Monograph Series*. IMS, 239–249.

ROCKAFELLAR, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J.

ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational analysis*, vol. 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

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

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

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