ON NONPARAMETRIC ESTIMATION OF DENSITY LEVEL SETS

BY A. B. TSYBAKOV

Université Paris VI and Institute for Problems of Information Transmission, Moscow

Let $X_1, \ldots, X_n$ be independent identically distributed observations from an unknown probability density $f(\cdot)$. Consider the problem of estimating the level set $G = G(\lambda) = \{x \in \mathbb{R}^2: f(x) \geq \lambda\}$ from the sample $X_1, \ldots, X_n$, under the assumption that the boundary of $G$ has a certain smoothness. We propose piecewise-polynomial estimators of $G$ based on the maximization of local empirical excess masses. We show that the estimators have optimal rates of convergence in the asymptotically minimax sense within the studied classes of densities. We find also the optimal convergence rates for estimation of convex level sets. A generalization to the $N$-dimensional case, where $N > 2$, is given.

1. Problem statement. Let $X_1, \ldots, X_n$ be independent identically distributed observations from an unknown probability density $f(x), x \in \mathbb{R}^N$, and let $\lambda > 0$ be a fixed number. Consider the $\lambda$-level set of the density $f$:

$$G = G_f(\lambda) = \{x \in \mathbb{R}^N : f(x) \geq \lambda\}.$$ 

The problem is to estimate $G = G_f(\lambda)$ given $X_1, \ldots, X_n$. By an estimator of $G$ we mean an arbitrary closed set $\hat{G}_n$ in $\mathbb{R}^N$, measurable with respect to $X_1, \ldots, X_n$.

Estimation of level sets is useful in various situations. For example, the problem of cluster analysis may be reduced to that of estimation of density level sets [Hartigan (1975)]. The approach based on level sets can be applied to define tests of multimodality, as proposed by Müller and Sawitzki (1991) [see also Müller (1993) and Polonik (1995)]. Another interesting field, where the level sets estimation may be implemented, is nonlinear extension of principal component analysis. By comparing the shapes of a suitable number of level sets, related to different levels, one can look for a hidden common structure in them, which could be interpreted in terms of nonlinear (in general) "principal" curves. Estimation of support of a density $f$ (the "0-level" set of $f$) has been suggested in certain statistical problems as well. Devroye and Wise (1980), Grenander (1981), Cuevas (1990), Cuevas and Fraiman (1993) used density support estimation for pattern recognition and for detection of the abnormal behavior of a system. Also in econometrics one finds some interesting connections to support estimation [see Korostelev and Tsy-
bakov (1993b), Chapter 7 for further references and discussion). In all these
eamples the estimators of general level sets are of interest as well. In fact,
one looks for sets with high concentration of sample points, and it is some-
times better to define them as level sets (and not as supports), to avoid the
effect of possible outliers.

Assume that an estimator \( \hat{f}_n(x) \) of the density \( f(x) \) is available. Then a
straightforward estimator of level set \( G_f(\lambda) \) is \( \hat{G}_n = \{ x \in \mathbb{R}^N : \hat{f}_n(x) \geq \lambda \} \), the
\textit{plug-in estimator}. If the density \( f \) has no “flat parts” at the level \( \lambda \) (for
example, \( \| \nabla f \| \) is bounded away from 0 around the level \( \lambda \), where \( \nabla f \) is the
gradient of \( f \)), then \( \hat{G}_n \) is a consistent estimator of \( G_f \), and it inherits in a
certain sense the convergence rate of \( \hat{f}_n \) [Molchanov (1990, 1993), Cuevas
and Fraiman (1993)]. Thus, the plug-in approach relates the properties of
level set estimators to the smoothness of density \( f \), and does not care about a
specific shape of a level set. However, it is sometimes convenient to introduce
certain restrictions on level sets directly. For example, one could be interested
in estimating ellipsoidal level sets [Nolan (1991)], or convex level sets
[Hartigan (1987)], or the level sets satisfying certain smoothness conditions
on the boundary [Polonik (1995)]. In these examples the behavior of the
density \( f \) is not very important; it may be even discontinuous, and the plug-in
approach does not work. Hartigan (1987) and Müller and Sawitzki (1991)
proposed another approach to level set estimation based on \textit{excess mass}. By
definition, the excess mass of a measurable set \( G \) in \( \mathbb{R}^N \) is the value
\( M(G) = \int_G f(x) \, dx - \lambda \cdot \text{mes}(G) \), where \( \text{mes}(G) \) denotes the Lebesgue
measure of \( G \). It is clear that \( M(G_f(\lambda)) \geq M(G) \) for any \( G \). The empirical excess mass
is defined as \( M_n(G) = (1/n) \sum_{i=1}^n I(X_i \in G) - \lambda \cdot \text{mes}(G) \). A natural estimator
of \( G_f(\lambda) \) is the maximizer of \( M_n(G) \) over a given class of level sets \( G \). Such
estimators, for different classes of level sets, were studied by Hartigan (1987),
who proved consistency and found certain rates of convergence of the estima-
tors. An interesting question is whether the excess mass approach gives
estimators with optimal convergence rates. In this paper we show that in
some cases the answer to this question is positive. However, to get results on
optimal rates in more generality, we need to introduce a new class of
estimators, based on maximization of \textit{local empirical excess masses}, instead of
the usual global maximizers. These estimators have a piecewise-polyno-
mial structure, and they are studied by a technique related to that of
Korostelev and Tsybakov (1993b).

Let us introduce some notation. Let \( w(t) \) be a \textit{loss function}, that is, the
function defined for nonnegative \( t \) and having the following properties: \( w(t) \)
is nonnegative, nondecreasing, \( w(t) \neq 0, w(0) = 0 \), and
\[
w(t) = 1 + t^q, \quad t \geq 0,
\]
for some \( q > 0 \).

Define the risk of any estimator \( \hat{G}_n \) as
\[
r_n(G_f(\lambda), \hat{G}_n, \psi_n) = E_f \left[ w \left( \psi_n^{-1} d(G_f(\lambda), \hat{G}_n) \right) \right],
\]
where $d$ is a distance between $G$ and $\hat{G}_n$, $E_f(\cdot)$ denotes the expectation with respect to the probability measure $P_f$ of $X_1, \ldots, X_n$, when the underlying density is $f$ and $\psi_n$ is a normalizing sequence of positive numbers.

We use the following two distances between closed sets $G_1, G_2$: Lebesgue measure of symmetric difference $d = d_1$, with

$$d_1(G_1, G_2) = \text{mes}(G_1 \Delta G_2),$$

and the Hausdorff metric $d = d_x$, with

$$d_x(G_1, G_2) = \max\{ \max_{x \in G_1} \rho(x, G_2); \max_{x \in G_2} \rho(x, G_1) \},$$

where $\rho(x, G) = \min_{y \in G} |x - y|$ is the Euclidean distance between a point $x$ and a closed set $G$.

We study the problem of estimating the level set $G_f(\lambda)$ in the nonparametric minimax setting, in the spirit of Ibragimov and Khasminskii (1981) and Stone (1980). Namely, we introduce a class $\mathcal{F}$ of densities $f$ such that their level sets belong to some class of sets $\mathcal{S}$. We consider the maximal risk of $\hat{G}_n$ over a class $\mathcal{F}$:

$$\mathcal{R}_n(\mathcal{F}, \hat{G}_n, \psi_n) = \sup_{\lambda \in \mathcal{F}} \mathcal{R}_n(G_f(\lambda), \hat{G}_n, \psi_n).$$

Our aim is to find an estimator $G^*_n$ that converges to $G$ as $n \to \infty$ with the best possible rate, that is, the estimator $G^*_n$ satisfying

$$\limsup_{n \to \infty} \mathcal{R}_n(\mathcal{F}, G^*_n, \psi_n) < \infty,$$

where the sequence $\psi_n = \psi_n(\mathcal{F}, d)$ is such that

$$\liminf_{n \to \infty} \inf_{\hat{G}_n} \mathcal{R}_n(\mathcal{F}, \hat{G}_n, \psi_n) > 0$$

($\inf_{\hat{G}_n}$ denotes the infimum over all estimators).

If (1) and (2) hold, we call $\psi_n$ the optimal rate of convergence, and we say that $G^*_n$ has optimal rate of convergence on the class of densities $\mathcal{F}$ (in $d$-metric).

In this paper we find the optimal rates of convergence in $d_1$ and $d_x$-metrics and the optimal level set estimators $G^*_n$ for certain classes of densities $\mathcal{F}$. We mainly consider the case $N = 2$, since the proofs for $N > 2$ are quite similar, but they require more notation. The results for general $N$ are stated in Section 5; elsewhere we assume that $N = 2$.

2. Classes of sets and classes of densities. Let $N = 2$, and let $\gamma$ and $L$ be positive constants. Denote by $\Sigma(\gamma, L)$ the class of functions $g(t), t \in \mathbb{R}^1$, such that the derivatives of $g$ up to the order $k$ exist, are $2\pi$-periodic, and the $k$th derivative satisfies the Hölder condition

$$|g^{(k)}(t) - g^{(k)}(t')| \leq L|t - t'|^{\gamma - k} \quad \forall \ t, t' \in \mathbb{R}^1,$$

where $k = \lfloor \gamma \rfloor$ (i.e., $k$ is the largest integer which is strictly less than $\gamma$).
We call the set $G \subseteq \mathbb{R}^2$ *star-shaped* about the origin if in polar coordinates $(r, \varphi)$, with $0 \leq r < \infty$, $0 \leq \varphi < 2\pi$, it has the form

\begin{equation}
G = \{ x = (r, \varphi) : 0 \leq r \leq g(\varphi), 0 \leq \varphi < 2\pi \},
\end{equation}

where $g(\cdot)$ is a $2\pi$-periodic continuous function on $\mathbb{R}^1$ called the *shape function*.

Denote by $\mathcal{F}(\gamma, L, h)$ the class of all star-shaped sets of the form (3) such that $g \in \Sigma(\gamma, L)$ and $g(\varphi) \geq h$, $\forall \varphi \in [0, 2\pi)$, where $h > 0$ is a given number. We assume that the level set $G_f(\lambda)$ belongs to the class $\mathcal{F}(\gamma, L, h)$, and denote by $g_{\lambda}(\varphi)$ the shape function that corresponds to $G_f(\lambda)$.

Let $f$ be a probability density, with a star-shaped level set $G_f(\lambda)$, and let $0 \leq \alpha < \infty$. For $\alpha > 0$, we say that the density $f$ is $\alpha$-regular around the level $\lambda$, if there exist constants $b_2 > b_1 > 0$, $\delta_0 > 0$ such that

\begin{equation}
b_1 \leq \frac{|f(x) - \lambda|}{|r - g_{\lambda}(\varphi)|^{\alpha}} \leq b_2,
\end{equation}

for all $x = (r, \varphi)$ such that $|f(x) - \lambda| \leq \delta_0$. If the set $\{ x : |f(x) - \lambda| \leq \delta_0 \}$ is empty, we put $\alpha = 0$, and we say that $f$ is $0$-regular around the level $\lambda$. This corresponds to the case where the density $f$ has a jump around the level $\lambda$. In general, condition (4) excludes the possibility for a density $f$ to have flat parts at level $\lambda$. The behavior of the usual densities, such as the Gaussian one, corresponds to $\alpha = 1$. If condition (4) holds only with $\alpha > 1$, it may indicate that there is a local extremum or inflexion point at the level $\lambda$.

Fix positive numbers $\gamma, L, h, \lambda, b_1, \delta_0$ and $b_2 > b_1$, $f_{\text{max}} > \lambda$, $\alpha \geq 0$. Denote by $\mathcal{F}_{\gamma, \alpha}$ the class of all probability densities $f$ on $\mathbb{R}^2$ satisfying simultaneously the following conditions:

C1. $f$ is uniformly bounded by $f_{\text{max}}$ on $\mathbb{R}^2$;
C2. $f$ is $\alpha$-regular around the level $\lambda$;
C3. the level set $G_f(\lambda) \in \mathcal{F}(\gamma, L, h)$.

For brevity, the dependence of $\mathcal{F}_{\gamma, \alpha}$ on $\lambda, L, h, f_{\text{max}}, b_1, b_2, \delta_0$ is not shown explicitly in the notation (these parameters do not influence the optimal rate of convergence $\psi_{\alpha}$). It will be always assumed that $\delta_0$ is sufficiently small (at least, $\delta_0 < \lambda$, $\delta_0 < f_{\text{max}} - \lambda$), without specifying explicitly how small it is. One can easily derive the necessary restrictions on $\delta_0$ in terms of $\lambda, h, \alpha, b_1, b_2$ and $f_{\text{max}}$ but they are cumbersome and of no interest for the results.

**Remark 1.** There exists $R > 0$ that depends only on $\gamma, L$ and $\lambda$, such that $G_f(\lambda) \subseteq B(0, R)$, $\forall f \in \mathcal{F}_{\gamma, \alpha}$, where $B(0, R)$ is the circle of radius $R$ centered at 0. In fact, since $f$ is a density, we have

$$1 \geq \lambda \text{mes}(G_f(\lambda)) = \lambda \int_0^{2\pi} \int_0^g r \, dr \, d\varphi = (\lambda/2) \int_0^{2\pi} g_{\lambda}^2(\varphi) \, d\varphi.$$  

Thus, $g_{\lambda}$ is bounded in $L_2[0, 2\pi]$, uniformly over $\mathcal{F}_{\gamma, \alpha}$. Since also $g_{\lambda} \in \Sigma(\gamma, L)$, it follows that there exists a constant $R = R(\gamma, L, \lambda)$ such that $\sup \psi g_{\lambda}(\varphi) \leq R$. 

EXAMPLE 1 (A family of densities that belong to $\mathcal{S}_{\gamma, 1}$). This example describes densities that have homothetic star-shaped level sets for all levels. As a particular case, one gets elliptically contoured densities, such as the multivariate normal.

Take a monotone decreasing differentiable function $f_*: [0, \infty) \to [0, \infty)$ and a shape function $g_* \in \mathcal{S}(\gamma, L, h)$. Define the bivariate density (in polar coordinates):
\[
f(r, \varphi) = \kappa f_*(r/g_*(\varphi)),
\]
where the constant $\kappa > 0$ is chosen so that $f(r, \varphi)$ integrates to 1 over $\mathbb{R}^2$. Let $\lambda < \kappa f_*(0)$ be given. Denote by $t_\lambda$ the positive number such that $\kappa f_*(t_\lambda) = \lambda$. Clearly, the shape function of the $\lambda$-level set of $f(r, \varphi)$ is $g_*(\varphi) = t_\lambda g_*(\varphi)$. As $\kappa f_*(t_\lambda) \neq 0$, there exist positive constants $b_1', b_2'$ such that
\[
b_1' \leq \frac{|\kappa f_*(t) - \lambda|}{|t - t_\lambda|} \leq b_2',
\]
for all $t$ such that $|\kappa f_*(t) - \lambda|$ is small enough. Substituting here $t = r/g_*(\varphi)$, $g_*(\varphi) = t_\lambda g_*(\varphi)$, one obtains (4) with $\alpha = 1$. Hence, $f \in \mathcal{S}_{\gamma, 1}$.

Along with $\mathcal{S}_{\gamma, \alpha}$ consider a class of densities with convex level sets. Denote
\[
\mathcal{S}_{\gamma, \alpha}^{\text{conv}} = \{G \subseteq B(0, R) : G \text{ is a closed convex set, } G \supseteq B(0, (\mu_0/\pi)^{1/2})\},
\]
where $0 < \mu_0 < \lambda^{-1}$ and $R > (\mu_0/\pi)^{1/2}$ are fixed constants. Clearly, any set $G \in \mathcal{S}_{\gamma, \alpha}^{\text{conv}}$ is star-shaped about the origin.

Define $\mathcal{S}_{\gamma, \alpha}^{\text{conv}}$ as the class of all probability densities $f$ on $\mathbb{R}^2$ that satisfy C1, C2 and the following condition.

C4. The level set $G_f(\lambda) \in \mathcal{S}_{\gamma, \alpha}^{\text{conv}}$.

Again, we keep in the notation only the dependence of $\mathcal{S}$ on $\alpha$, since the values of $\lambda, \mu_0, f_{\text{max}}, b_1, b_2, \delta_0$ do not influence the optimal rate of convergence.

3. Level set estimators for $\mathcal{S}_{\gamma, \alpha}$. In this section we define the estimators of level sets $G_f(\lambda)$ that have optimal convergence rates on the classes of densities $\mathcal{S}_{\gamma, \alpha}$ in $d$-metrics, with $d = d_1$ or $d = d_\infty$. The definition of the estimator depends on $\gamma, \alpha$ and on $d$.

Let
\[
\delta_n = \begin{cases} 
(1/n)^{1/(2(\alpha + 1)\gamma + 1)}, & \text{if } d = d_1, \\
(\log n/n)^{1/(2(\alpha + 1)\gamma + 1)}, & \text{if } d = d_\infty.
\end{cases}
\]
Denote $M = M_n = 1/\delta_n$ and assume without loss of generality that $M$ is an integer. Let $\varphi_l = 2\pi l\delta_n, l = 0, 1, \ldots, M$, and define the sectors
\[
S_l = \{(r, \varphi) : 0 \leq r \leq R, \varphi_{l-1} \leq \varphi < \varphi_l\}, \quad l = 1, \ldots, M.
\]
In each sector $S_l$, define the parametric family of subsets

$$B_l(\theta) = \{(r, \varphi) \in S_l : 0 \leq r \leq \theta_0 + \theta_1(\varphi - \varphi_{l-1}) + \cdots + \theta_k(\varphi - \varphi_{l-1})^k\},$$

where $k = \lfloor \gamma \rfloor$, and $\theta = (\theta_0, \ldots, \theta_k)$ is a vector of real parameters. Denote

$$\Theta^* = \{\theta : h \leq \theta_0 + \theta_1(\varphi - \varphi_{l-1}) + \cdots + \theta_k(\varphi - \varphi_{l-1})^k \leq R, \forall \varphi : \varphi_{l-1} \leq \varphi < \varphi_l\},$$

where $R$ is the radius of the circle containing the level sets (see Remark 1). The condition $\theta \in \Theta^*$ guarantees that $B_l(\theta) \subseteq S_l$. Note that $\Theta^*$ does not depend on $l$. Consider the following discrete subset of $\Theta^*$:

$$\Theta_n = \{\theta = (m_0 \delta_n^\gamma, m_1 \delta_n^{\gamma-1}, \ldots, m_k \delta_n^{\gamma-k}) \} \cap \Theta^*,$$

where $m = (m_0, m_1, \ldots, m_k)$ is a $(k + 1)$-tuple of integers.

Define the piecewise-polynomial estimator of the shape function $g(\varphi)$:

$$\hat{g}(\varphi) = \hat{\theta}_{0l} + \hat{\theta}_{1l}(\varphi - \varphi_{l-1}) + \cdots + \hat{\theta}_{kl}(\varphi - \varphi_{l-1})^k,$$

$$\varphi_{l-1} \leq \varphi < \varphi_l, \ l = 1, \ldots, M,$$

where the vector $(\hat{\theta}_{0l}, \ldots, \hat{\theta}_{kl}) = \hat{\theta}_l \in \Theta_n$ is a solution of the following maximization problem:

$$\hat{\theta}_l = \arg \max_{\theta \in \Theta_n} J_l(\theta),$$

where $J_l(\theta)$ is the empirical excess mass corresponding to the parametric set $B_l(\theta)$; that is,

$$J_l(\theta) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \in B_l(\theta)\} - \lambda \text{mes}(B_l(\theta)).$$

The set $\{(r, \varphi) : 0 \leq r \leq \hat{g}(\varphi), 0 \leq \varphi < 2\pi\}$ is not necessarily closed. Let us take as an estimator of $G = G_\ell(\lambda)$ the closure of this set:

$$(5) \quad G_n^* = \text{Cl}\{(r, \varphi) : 0 \leq r \leq \hat{g}(\varphi), 0 \leq \varphi < 2\pi\}.$$

**Theorem 1.** For any loss function $\omega$ the $\lambda$-level set estimator $G_n^*$ has optimal rate of convergence on the class of densities $\mathcal{F}_{\gamma, \alpha}$ in $d_1$- and $d_\infty$-metrics respectively, depending on the choice of $\delta_n$. The optimal rate has the following form:

$$\psi_n(\mathcal{F}_{\gamma, \alpha}, d_1) = n^{-\gamma/(2\alpha + 1)\gamma + 1},$$

$$\psi_n(\mathcal{F}_{\gamma, \alpha}, d_\infty) = (n/\log n)^{-\gamma/(2\alpha + 1)\gamma + 1}.$$  

The proof of Theorem 1 is given in Section 6.

**Remark 2.** For the typical case where $\alpha = 1$, one gets the optimal rate of convergence $n^{-\gamma/(3\gamma + 1)}$ in $d_1$, and by a log-factor worse rate in $d_\infty$. For the
extremal case $\alpha = 0$, Theorem 1 gives the optimal rate $n^{-\gamma/(\gamma+1)}$, the same as for estimation of support of a uniform density [Korostelev and Tsybakov (1993a, b)]. This coincidence is natural, since in both situations one deals with the jump behavior of a density at the contour representing the boundary of either a level set or of a support.

**Remark 3.** To prove that (6) gives optimal rates of convergence, one needs to show the relation (1) (called upper bound) and (2) (called lower bound). Here we prove that the upper bound is attained by the estimator (5). In the special case $d = d_1$, $\gamma > 1$ the correct upper bound of order $n^{-\gamma/((2\alpha+1)\gamma+1)}$ is attained also on global excess mass estimates [Polonik (1995)].

4. Estimators of convex level sets. In this section we show that certain estimators of convex level sets have optimal rates of convergence. We consider the classes of densities $\mathcal{F}_{\text{conv}, \alpha}$. The case of convex level sets appears to be quite different from that studied in Section 3. For the estimation of convex sets, the dependence of optimal rates $\psi_n$ on the metric $d$ is more important: one loses a polynomial factor rather than a logarithmic one when passing from $d_1$- to $d_2$-metric.

Hartigan (1987) introduced the following estimator of convex level sets based on the maximization of the empirical excess mass:

$$G_n^H = \arg \max_{G \text{ convex}} \left[ \frac{1}{n} \sum_{i=1}^{n} I\{X_i \in G\} - \lambda \mes(G) \right].$$

He proposed a computational algorithm for $G_n^H$ and conjectured that $G_n^H$ converges to $G_f(\lambda)$ with the rate $(n/\log n)^{-2/7}$ in the Hausdorff metric $d_a$. We will show that this conjecture is not true for general convex level sets $G_f(\lambda)$. But first we consider the behavior of $G_n^H$ with respect to the distance $d_1$.

**Theorem 2.** For any loss function $w$ the $\lambda$-level set estimator $G_n^H$ has optimal rate of convergence on the class of densities $\mathcal{F}_{\text{conv},\alpha}$ in $d_1$-metric. The optimal rate is

$$\psi_n(\mathcal{F}_{\text{conv},\alpha}, d_1) = n^{-2/(4\alpha+3)}.$$  

The proof of Theorem 2 is given in Section 6. In fact, we will prove only the lower bound for the maximal risks. Proof of the upper bound, that is, of the fact that $G_n^H$ attains the rate (7), is an easy consequence of a result of Polonik (1995), and it is therefore omitted.

Note that for the typical case $\alpha = 1$, Hartigan's estimator $G_n^H$ converges with the rate $n^{-2/7}$ in $d_1$-metric, and this is the optimal rate. On the other hand, in the extreme case $\alpha = 0$ (the density $f$ has a jump at the level $\lambda$), Theorem 2 gives $\psi_n = n^{-2/3}$, which is shown by Mammen and Tsybakov (1995) to be the optimal rate in estimation of a convex support of a uniform density in $\mathbb{R}^2$ (cf. Remark 2 in Section 3).
The behavior of convex level sets estimators in the Hausdorff metric is substantially worse, which is shown by the following lower bound.

**Theorem 3.** Let $\psi_n = n^{-1/(2a+2)}$ and let $d = d_\infty$. Then for any loss function $w$ the following lower bound holds:

$$\liminf_{n \to \infty} \inf_{\hat{G}_n} \mathcal{R}_n(\mathcal{F}_{\text{conv}, \alpha}, \hat{G}_n, \psi_n) > 0.$$  

The proof of Theorem 3 is given in Section 6.

The rate $\psi_n = n^{-1/(2a+2)}$ is optimal for $\mathcal{F}_{\text{conv}, \alpha}$ in $d_\infty$-metric, up to a log-factor. This follows from Theorems 1 and 3. In fact, if $f \in \mathcal{F}_{\text{conv}, \alpha}$, then $f \in \mathcal{F}_{1, \alpha}$, since the boundary of a convex set is Lipschitz (it is not hard to find the expression for the Lipschitz constant $L$ in terms of $\mu_0$). Hence, one can use the estimator $G^*_n$ with $\delta_n = (n/\log n)^{-1/(2a+2)}$ (corresponding to $\mathcal{F}_{1, \alpha}$ and $d = d_\infty$), and apply the upper bound of Theorem 1. We find that $G^*_n$ has the convergence rate $(n/\log n)^{-1/(2a+2)}$. (Note, however, that $G^*_n$, which is thus used to estimate a convex level set, is not a convex set itself.) This argument and Theorem 3 lead to the following corollary.

**Corollary.** For any loss function $w$ the optimal convergence rate on the class of densities $\mathcal{F}_{\text{conv}, \alpha}$ in $d_\infty$-metric satisfies

$$n^{-1/(2a+2)} \leq \psi_n(\mathcal{F}_{\text{conv}, \alpha}, d_\infty) \leq (n/\log n)^{-1/(2a+2)}.$$  

Comparing this with Theorems 1 and 2, we find that, qualitatively, the behavior of the best convex level set estimators in $d_1$-metric is similar to those of smooth level sets, with smoothness $\gamma = 2$, while the behavior in $d_\infty$-metric corresponds roughly to smoothness $\gamma = 1$. This effect is analogous to that discovered by Korostelev and Tsybakov (1995) in the problem of estimating the convex support of a uniform density. Seemingly, these results are particular cases of a general breakdown effect in optimal convergence rates [cf. Nemirovskii (1985), Donoho, Johnstone, Kerkyacharian and Picard (1995), who discuss similar phenomena in the classical nonparametric regression and density estimation problems].

**5. Extensions.** Let us mention some extensions of the results of Sections 3 and 4.

**E1.** Condition (4) can be given in a slightly different form (which is easier to interpret, but harder to handle analytically):

$$b_1 \leq \frac{|f(x) - \lambda|}{\rho(x, \partial G_f(\lambda)^\alpha} \leq b_2,$$

where $\partial G_f(\lambda)$ denotes the boundary of $G_f(\lambda)$. All the results remain valid if (4') is replaced by (4). This condition can be weakened even further if one is
interested in the rates of convergence in $d_1$-metric only. In fact, for getting the correct upper bounds on the risks in $d_1$-metric (up to a log-factor) one can apply the method of Polonik (1995) that works under the following condition:

$$\text{mes}\{x: \left| f(x) - \lambda \right| \leq \delta \} \leq b\delta^{1/\alpha}, \quad 0 < \delta < \delta_0,$$

where $b > 0$, $\delta_0 > 0$ and $\alpha > 0$ are constants. This condition is a weaker analog of the left inequalities in (4) and (4').

E2. The results can be made uniform in $\lambda$ (i.e., one can add $\sup_{\lambda \in \Lambda}$ in the definition of the maximal risk $\mathcal{R}_n$), if it is assumed that (4) or (4') holds uniformly in $\lambda \in \Lambda$, where $\Lambda$ is a compact interval.

E3. Theorem 1 can be extended to the dimension $N > 2$ in the following way. Let $\Sigma_N(\gamma, L)$ be the Hölder class of functions on the unit sphere in $\mathbb{R}^N$, and let $(r, \Phi)$ be the polar coordinates in $\mathbb{R}^N$, where $\Phi$ is an $(N-1)$-vector of polar angles. Denote by $\mathcal{F}_N(\gamma, L, h)$ the class of all star-shaped sets of the form $G = \{x = (r, \Phi): 0 < r < g(\Phi), \Phi \in [0, 2\pi)^{N-1}\}$, such that $g \in \Sigma_N(\gamma, L)$ and $g(\Phi) \geq h, \, h > 0$.

Define the class $\mathcal{F}_{\gamma, \alpha, N}$ of all probability densities $f$ on $\mathbb{R}^N$ bounded by $f_{\text{max}}$, satisfying (4') and such that $G_f(\lambda) \in \mathcal{F}_N(\gamma, L, h)$. As in the two-dimensional case, there exists $R > 0$ such that $G_f(\lambda) \subseteq B(0, R)$, where $B(0, R)$ is a Euclidean ball of radius $R$ centered at 0 (cf. Remark 1).

Consider the partition of $[0, 2\pi)^{N-1}$ into $M = s_n^{N-1}$ cubes $Q_1, \ldots, Q_M$ with equal edges of length $2\pi/s_n$, where

$$s_n = \begin{cases} n^{1/(2\alpha + 1)\gamma + N - 1}, & \text{if } d = d_1, \\ \left[\frac{n}{\log n}\right]^{1/(2\alpha + 1)\gamma + N - 1}, & \text{if } d = d_\alpha. \end{cases}$$

Let $p(\Phi)$ be a polynomial (in dimension $N - 1$) of order $k = \lfloor \gamma \rfloor$, and let $\theta$ be the vector of its coefficients. We write then $p(\Phi) = p(\Phi; \theta)$. Denote by $\mathcal{R}_{n\gamma}$ a minimal $s_n^{-\gamma}$-net in Euclidean metric on the set of coefficients

$$\{\theta: \theta \leq p(\Phi; \theta) \leq R, \forall \Phi \in Q_i\}.$$

Let $J_i(\theta)$ be the empirical excess mass of the set

$$B_i(\theta) = \{x = (r, \Phi): 0 \leq r < p(\Phi; \theta), \Phi \in Q_i\},$$

and let $\hat{\theta}_i = \arg\max_{\theta \in \mathcal{R}_{n\gamma}} J_i(\theta)$. The estimator $G_{n, N}^*$ ($N$-dimensional analog to $G_n^*$) is defined as

$$G_{n, N}^* = \text{Cl}\left\{(r, \Phi): 0 \leq r \leq \sum_{l=1}^{M} p(\Phi; \hat{\theta}_l) I[\Phi \in Q_i]\right\}.$$

The following result is the $N$-dimensional analog to Theorem 1.

**THEOREM 4.** For any loss function $w$ the $\lambda$-level set estimator $G_{n, N}^*$ has optimal rate of convergence on the class of densities $\mathcal{F}_{\gamma, \alpha, N}$ in $d_1$- and $d_\alpha$-metrics, respectively, depending on the choice of $s_n$. The optimal rate has
the following form:

\[
\psi_n(\mathcal{F}_\gamma, a, N, d_1) = n^{-\gamma/(2a+1)\gamma+N-1},
\]

\[
\psi_n(\mathcal{F}_\gamma, a, N, d_\omega) = (n/\log n)^{-\gamma/(2a+1)\gamma+N-1}.
\]

Theorems 2 and 3 are generalized to the \( N \)-dimensional case similarly. In particular, \( \psi_n(\mathcal{F}_{\text{conv}, a, N}, d_1) = n^{-2/(4a+N+1)} \), and (up to a log-factor)

\[
\psi_n(\mathcal{F}_{\text{conv}, a, N}, d_\omega) = n^{-1/(2a+N)}, \quad \text{where} \quad \mathcal{F}_{\text{conv}, a, N} \quad \text{is the \( N \)-dimensional analog to} \quad \mathcal{F}_{\text{conv}, a}.
\]

E4. In this paper we assume that \( \lambda > 0 \). As \( \lambda \to 0 \), we get the problem of density support estimation which is quite different. In fact, let, by definition, \( G_f(0) = \text{Cl} \{ x \in \mathbb{R}^N : f(x) > 0 \} \), and assume that \( G_f(0) \in \mathcal{F}_N(\gamma, L, h) \) and C1 and C2 of Section 3 hold, with the necessary modifications [\( \alpha \)-regularity around the level 0 means that (4') is satisfied, with \( \lambda = 0 \) for \( x \in G_f(0) \) such that \( \vert f(x) \vert \leq \delta_0 \)]. Then there exist estimators of \( G_f(0) \) that have the \( d_1 \) rate of convergence \( n^{-\gamma/(a+1)\gamma+N-1} \) if \( N \geq 2 \); this rate is optimal [Härdle, Park and Tsybakov (1995)]. We see that, except for the case \( \alpha = 0 \) (which corresponds to the jump behavior of a density near the boundary of its support or level set), the estimators of support converge faster than the estimators of level sets. This fact is easy to understand by comparing the likelihood ratios (or Hellinger distances) for the "worst hypotheses" appearing in the lower bounds. If \( \lambda > 0 \), the squared Hellinger distance between a density and its perturbation by a "small" function \( \tilde{f} \geq 0 \) around the level \( \lambda \) is roughly \( \int [(\lambda + \tilde{f}(x))^{1/2} - \lambda^{1/2}]^2 \, dx \sim \int \tilde{f}^2(x) \, dx \), while for \( \lambda = 0 \) this distance is proportional to \( \int \tilde{f}(x) \, dx \). Thus, if \( \lambda = 0 \), the perturbations of smaller order than for \( \lambda > 0 \) can be distinguished with the same accuracy.

6. Proofs.

PROOF OF THEOREM 1. We introduce some notation. By definition, there is a one-to-one correspondence between the vectors \( \theta = (\theta_0, \ldots, \theta_k) \in \Theta_n \) and the vectors of integers \( m = (m_0, \ldots, m_k) \). It is therefore convenient to write \( \theta = \theta(m) \).

DEFINITION 1. Denote by \( \hat{m}_l \) the vector of integers such that \( \hat{\theta}_l = \theta(\hat{m}_l) \).

DEFINITION 2 (Intersection of the slice \( S_l \) with the level set \( G_f(\lambda) \)).

\[
D_l = G \cap S_l, \quad l = 1, \ldots, M.
\]

For brevity we write

\[
g(\varphi) = g_\lambda(\varphi).
\]

DEFINITION 3 (The Taylor approximation of the set \( D_l \)).

\[
B_l^{(0)} = \{(r, \varphi) \in S_l : 0 \leq r \leq g_l^{(0)}(\varphi) \},
\]

where \( g_l^{(0)}(\varphi) = \theta_0^{(0)} + \theta_1^{(0)}(\varphi - \varphi_{l-1}) + \cdots + \theta_k^{(0)}(\varphi - \varphi_{l-1})^k \), \( \theta_j^{(0)} = m^{(0,j)} \delta_n^{-j}, \)

\[
m^{(0,j)} = \left[ g_l^{(j)}(\varphi_{l-1}) / (j! \delta_n^{-j}) \right], \quad j = 0, \ldots, k. \]

Here \( [\cdot] \) is the integer part of a real
number. Denote

\[ m_i^{(0)} = (m_i^{(0,0)}, m_i^{(0,1)}, \ldots, m_i^{(0,k)}). \]

For any vector of integers \( m = (m_0, \ldots, m_k) \) set

\[ ||m|| = \max_{0 \leq j \leq k} |m_j|. \]

Since \( g \in \Sigma(\gamma, L) \), we have

\[ |g(\varphi) - g_i^{(0)}(\varphi)| \leq \delta_n^\gamma L/k! + (k + 1) \delta_n^\gamma = c_* \delta_n^\gamma, \quad \varphi_{l-1} \leq \varphi < \varphi_l, \]

where the summand \( (k + 1) \delta_n^\gamma \) is due to the discretization, and \( c_* = L/k! + (k + 1) \).

Lemma 1. For any Lebesgue measurable set \( B \subseteq S_l \) we have

\[
\int (f(x) - \lambda) I\{x \in B\} \, dx \\
= \int (f(x) - \lambda) I\{x \in D_l\} \, dx - \int |f(x) - \lambda I\{x \in B \triangle D_l\} | \, dx.
\]

Proof.

\[
\int (f(x) - \lambda) I\{x \in B\} \, dx - \int (f(x) - \lambda) I\{x \in D_l\} \, dx \\
= \int (f(x) - \lambda) I\{x \in B \triangle D_l\} \, dx - \int (f(x) - \lambda) I\{x \in D_l \triangle B\} \, dx \\
= -\int |f(x) - \lambda I\{x \in B \triangle D_l\} | \, dx. \quad \Box
\]

Lemma 2. There exist positive constants \( c_1 \) and \( c_2 \), such that for any vector of integers \( m \) with \( \theta(m) \in \Theta_n \) we have

\[ \text{mes}(D_l \Delta B_i^{(0)}) \leq c_1 \delta_n^{\gamma+1}, \]

\[ c_2 \delta_n^{\gamma+1} ||m - m_i^{(0)}|| \leq \text{mes}(B_l(\theta(m)) \Delta B_l^{(0)}) \leq c_1 \delta_n^{\gamma+1} ||m - m_i^{(0)}||, \]

\[ \text{mes}(B_l(\theta(m)) \Delta D_l) \geq \delta_n^{\gamma+1}(c_2 ||m - m_i^{(0)}|| - c_1), \quad l = 1, \ldots, M. \]

The constants \( c_1 \) and \( c_2 \) do not depend on \( l, m \) or on the density \( f \in \mathcal{F}_{\gamma, \alpha} \).

Proof. Using (9) and the inequality \( g(\varphi) \leq R \), we get

\[
\text{mes}(D_l \Delta B_i^{(0)}) = \int_{\varphi_{l-1}}^{\varphi_l} \int_{\min(g, g_l^{(0)})}^{\max(g, g_l^{(0)})} r \, dr \, d\varphi \\
\leq (R + c_* \delta_n^\gamma) \int_{\varphi_{l-1}}^{\varphi_l} |g(\varphi) - g_i^{(0)}(\varphi)| \, d\varphi \leq c_1 \delta_n^{\gamma+1},
\]
which gives (10). Similarly, if \( \theta(m) \in \Theta_n \),

\[
\text{mes}(B_i(\theta(m)) \Delta B_i^{(0)}) \leq (R + c_* \delta_n^\gamma) \left| \int_{\varphi_{l-1}}^{\varphi_l} \delta_n^\gamma \left| \sum_{j=0}^{k} (m_j - m^{(0,j)})(\varphi - \varphi_{l-1})/\delta_n \right| d\varphi \right|
\]

\[
= (R + c_* \delta_n^\gamma) \delta_n^{\gamma+1} \mathcal{J},
\]

where

\[
\mathcal{J} = \int_0^1 \left| \sum_{j=0}^{k} (m_j - m^{(0,j)}) y^j \right| dy
\]

and

\[
\text{mes}(B_i(\theta(m)) \Delta B_i^{(0)}) \geq (h - c_* \delta_n^\gamma) \delta_n^{\gamma+1} \mathcal{J}.
\]

Since all the norms in the space of polynomials of order \( k \) are equivalent, the integral \( \mathcal{J} \) is bounded from above and from below by the expressions of the form \( C\|m - m_l^{(0)}\| \), where \( C > 0 \) is a constant depending on \( k \) only. This proves (11). Now, (12) follows from (10) and (11) by use of the triangle inequality. □

Denote

\[
\zeta_{1l} = I\{X_i \in B_i(\theta(m))\} - I\{X_i \in B_i^{(0)}\} - \lambda \left[ \text{mes}(B_i(\theta(m))) - \text{mes}(B_i^{(0)}) \right].
\]

**Lemma 3.** There exist positive constants \( c_3 \) and \( c_4 \), such that for any vector of integers \( m \) with \( \theta(m) \in \Theta_n \) we have

\[
E_f(\zeta_{1l}) \leq -\delta_0 \delta_n^{(\alpha+1)+1}(c_2\|m - m_l^{(0)}\| - c_3),
\]

\[
\text{Var}_f(\zeta_{1l}) = E_f\left( \left[ E_f(\zeta_{1l}) - \zeta_{1l} \right]^2 \right) \leq c_4 \delta_n^{\gamma+1}\|m - m_l^{(0)}\|, \quad l = 1, \ldots, M,
\]

where \( \delta_0 > 0 \) is the constant used in the definition (4) of \( \alpha \)-regularity, and the constants \( c_3 \) and \( c_4 \) do not depend on \( l, m \) or on the density \( f \in \mathcal{F}_{\gamma, \alpha} \).

**Proof.** Denote for brevity \( B = B_i(\theta(m)) \), \( B^0 = B_i^{(0)} \). Using Lemma 1 we get

\[
E_f(\zeta_{1l}) = \int (f(x) - \lambda) I\{x \in B\} dx - \int (f(x) - \lambda) I\{x \in B^0\} dx
\]

(13)

\[
= \int |f(x) - \lambda| \{I\{x \in B^0 \Delta D_i\} - I\{x \in B \Delta D_i\}\} dx.
\]

By virtue of (4), (9) and (10) we have

\[
\int |f(x) - \lambda| I\{x \in B^0 \Delta D_i\} dx \leq \text{mes}(B^0 \Delta D_i) \sup_{x=\varphi, |r-g(\varphi)| \leq c_* \delta_n^\gamma} |f(x) - \lambda|
\]

(14)

\[
\leq c_1 b_2c_*^\alpha \delta_n^{(\alpha+1)+1}
\]
if \( n \) is large enough. Let \( \delta_0 > 0 \) be the constant used in the definition (4) of \( \alpha \)-regularity. For any Lebesgue measurable set \( A \subseteq S_t \),

\[
\int |f(x) - \lambda|I\{x \in A\} \, dx \\
\geq \int |f(x) - \lambda|I\{x \in A, |f(x) - \lambda| \geq \delta_0 \delta_n^{\alpha\gamma}\} \, dx \\
\geq \delta_0 \delta_n^{\alpha\gamma} \text{mes}\{A \cap \{x \in S_t : |f(x) - \lambda| \geq \delta_0 \delta_n^{\alpha\gamma}\}\} \\
\geq \delta_0 \delta_n^{\alpha\gamma}(\text{mes}(A) - \text{mes}\{x \in S_t : |f(x) - \lambda| \leq \delta_0 \delta_n^{\alpha\gamma}\})�.
\]

Let \( \alpha > 0 \). Then, in view of (4),

\[
\text{mes}\{x \in S_t : |f(x) - \lambda| \leq \delta_0 \delta_n^{\alpha\gamma}\} \leq \text{mes}\{(r, \varphi) \in S_t : b_1|r - g(\varphi)|^{\alpha} \leq \delta_0 \delta_n^{\alpha\gamma}\} \\
\leq 4\pi R(\delta_0/b_1)^{1/\alpha}\delta_n^{\gamma+1}.
\]

Thus, the last expression in (15) is bounded from below by

\[
\delta_0 \delta_n^{\alpha\gamma}(\text{mes}(A) - 4\pi R(\delta_0/b_1)^{1/\alpha}\delta_n^{\gamma+1}).
\]

If \( \alpha = 0 \), then by the definition of \( \alpha \)-regularity (see Section 2) the set \( \{x : |f(x) - \lambda| \leq \delta_0\} \) is empty. Using this, we get

\[
\text{mes}\{x \in S_t : |f(x) - \lambda| \leq \delta_0 \delta_n^{\alpha\gamma}\} = 0,
\]

and thus, if \( \alpha = 0 \), the last expression in (15) is bounded from below by \( \delta_0 \delta_n^{\alpha\gamma} \text{mes}(A) \). Hence,

\[
\int |f(x) - \lambda|I\{x \in A\} \, dx \geq \delta_0 \delta_n^{\alpha\gamma}(\text{mes}(A) - b\delta_n^{\gamma+1}),
\]

where \( b = 4\pi R(\delta_0/b_1)^{1/\alpha}I(\alpha > 0) \). Now, set \( A = B\Delta D_t \), and use the last inequality and (12). We get

\[
\int |f(x) - \lambda|I\{x \in B\Delta D_t\} \, dx \geq \delta_0 \delta_n^{\alpha\gamma}(\alpha+1)(c_2\|m - m_t^\alpha\| - c_1 - b),
\]

which, together with (13) and (14), yields the first inequality of Lemma 3.

To prove the second inequality, note that

\[
\text{Var}_f(\xi_{11}) = \text{Var}_f(I\{X_1 \in B\} - I\{X_1 \in B^0\}) \\
\leq E_f[|I\{X_1 \in B\} - I\{X_1 \in B^0\}|^2] \\
= E_f[I\{X_1 \in B\Delta B^0\}] \leq f_{\text{max}} \text{mes}(B\Delta B^0) \\
\leq f_{\text{max}} c_1 \delta_n^{\gamma+1}\|m - m_t^\alpha\|,
\]

where we used (11). \( \square \)
PROOF OF THE UPPER BOUND OF THEOREM 1. For any $t > 0$, by definition of $\hat{\theta}_i$ in Section 3 and Definition 1 of $\hat{m}_i$ in terms of $\hat{\theta}_i$, we have

$$P_f(\|\hat{m}_i - m_i^{(0)}\| > t) \leq P_f\left( \max_{\|m - m_i^{(0)}\| > t} (J_i(\theta(m)) - J_i(\theta(m_i^{(0)}))) > 0 \right)$$

$$\leq \sum_{r=1}^{\infty} P_f\left( \max_{m \in \mathcal{L}(r)} (J_i(\theta(m)) - J_i(\theta(m_i^{(0)}))) > 0 \right)$$

$$\leq \sum_{r=1}^{\infty} \text{card } \mathcal{L}(r) \max_{m \in \mathcal{L}(r)} P_f\left( \frac{1}{n} \sum_{i=1}^{n} \xi_{il} > 0 \right),$$

where $\mathcal{L}(r)$ is the following set of vectors $m$ with integer coordinates:

$$\mathcal{L}(r) = \{ m : tr < \|m - m_i^{(0)}\| \leq t(r + 1) \} \cap \{ m : \theta(m) \in \Theta_n \}, \quad r = 1, 2, \ldots,$$

and $m_i^{(0)}$ is introduced in Definition 3.

Assume that $t > 2c_3/c_2$. Consequently, $c_2\|m - m_i^{(0)}\| - c_3 \geq c_3\|m - m_i^{(0)}\|/2$, for $m \in \mathcal{L}(r)$, $r \geq 1$. Note that, for $l$ fixed, $\xi_{il}$ are i.i.d. bounded random variables $|\xi_{il}| \leq 2(1 + \pi R^2) = c_5$. Using the Bernstein inequality and Lemma 3, we find

$$P_f\left( \frac{1}{n} \sum_{i=1}^{n} \xi_{il} > 0 \right)$$

$$\leq P_f\left( \frac{1}{n} \sum_{i=1}^{n} (\xi_{il} - E_f(\xi_{il})) > \delta_0 \delta_n^{\gamma(\alpha + 1) + 1} c_2 \|m - m_i^{(0)}\|/2 \right)$$

$$\leq 2 \exp\left( -\frac{n [\delta_0 \delta_n^{\gamma(\alpha + 1) + 1} c_2 \|m - m_i^{(0)}\|/2]^2}{2 \text{Var}_f(\xi_{il}) + \frac{1}{3} c_5 \delta_n^{\gamma(\alpha + 1) + 1} c_2 \|m - m_i^{(0)}\|} \right)$$

$$\leq 2 \exp\left( -\frac{n \delta_n^{2\gamma(\alpha + 1) + 2} \delta_0 \delta_n^{\gamma(\alpha + 1) + 1} c_2 \|m - m_i^{(0)}\|/2}{2 c_4 \delta_n^{\gamma + 1} + \frac{1}{3} c_5 \delta_n^{\gamma(\alpha + 1) + 1} c_2} \right)$$

$$\leq 2 \exp\left( -c_6 n \delta_n^{2(\alpha + 1) + 1} \right),$$

for all $m \in \mathcal{L}(r)$, where $c_6 > 0$ is a constant.

The definition of $\mathcal{L}(r)$ implies card $\mathcal{L}(r) \leq c_7 t^{k+1} r^k$, where $c_7 > 0$ depends only on $k$. Using this and (16), (17), we obtain

$$P_f(\|\hat{m}_i - m_i^{(0)}\| > t) \leq 2 c_7 \sum_{r=1}^{\infty} t^{k+1} r^k \exp(-c_6 n \delta_n^{2(\alpha + 1) + 1} t)$$

$$\leq c_8 \exp(-c_9 n \delta_n^{2(\alpha + 1) + 1} t) \quad \text{for } t > 2c_3/c_2,$$

where the constants $c_8, c_9 > 0$ do not depend on $f, l$ and $n$.

If $d = d_1$, then $\delta_n = n^{-1/(\gamma(\alpha + 1) + 1)}$, and (18) reads as

$$P_f(\xi_l > t) \leq c_8 \exp(-c_9 t), \quad t > 2c_3/c_2,$$

where $\xi_l = \|\hat{m}_i - m_i^{(0)}\|$. Next, if $d = d_2$, then $\delta_n = (n/\log n)^{-1/(\gamma(\alpha + 1) + 1)}$, and (18) reads as

$$P_f(\xi_l > t) \leq c_8 n^{-c_9 t}, \quad t > 2c_3/c_2.$
Using (19) and (20) and arguing in a standard way [see, e.g., pages 115–117 of Korostelev and Tsybakov (1993b)], one gets, both for $d = d_1$ and $d = d_\alpha$,
\begin{equation}
\sup_{f \in \mathcal{F}_{\gamma,\alpha}} E_f([\delta_n^\ast d(G, G_n^\ast)]^q) < \infty \quad \forall \ q > 0,
\end{equation}
which proves the upper bound (1) in Theorem 1.

**Proof of the Lower Bound of Theorem 1.** Consider only the case where $\alpha > 0$, since for $\alpha = 0$ the lower bound is proved in the same way as in the problem of estimating the support of a uniform density [Korostelev and Tsybakov (1993b), Chapter 7].

We will estimate from below the maximal risk over the class of densities $\mathcal{F}_{\gamma,\alpha}$ by the maximal risk over a finite family $\mathcal{N} \subset \mathcal{F}_{\gamma,\alpha}$, and then apply the Fano lemma. This approach is due to Ibragimov and Khasminskii (1981) [see also Bigré (1983)]. To derive the results in the shortest way we use Lemma A.1 of Tsybakov (1982) which we state below for convenience, in the form adapted to our purposes.

**Lemma 4.** Let $d$ be a pseudometric. Assume that there exists a subset $\mathcal{N} \subset \mathcal{F}$ and numbers $\varepsilon > 0, 0 < \theta < 1$, such that
\begin{equation}
2 \leq \text{card}(\mathcal{N}) = s < \infty, \quad d(G_f(\lambda), G_h(\lambda)) \geq \varepsilon \quad \forall f, h \in \mathcal{N}, f \neq h,
\end{equation}
and
\begin{equation}
\max_{f, h \in \mathcal{N}} I(P_f, P_h)/(\log s) \leq 1/2,
\end{equation}
where $I(P_f, P_h)$ is the Kullback information of $P_f$ with respect to $P_h$. Then, for any loss function $w$, we have
\begin{equation}
\sup_{\mathcal{F}} \inf_{G_n} E_f\left(w \left(\frac{ud(G_f(\lambda), G_n^\ast)}{\varepsilon}\right)\right) \geq \nu w\left(\frac{u}{2}\right) \quad \forall \ u > 0,
\end{equation}
where $\nu > 0$ is an absolute constant.

Since $w$ is a loss function, there exists $u > 0$ such that $w(u/2) > 0$. Thus, to prove the lower bound we need to construct a family $\mathcal{N} \subset \mathcal{F}_{\gamma,\alpha}$, such that (22) and (23) are satisfied, with $\varepsilon \sim \psi_n(\mathcal{F}_{\gamma,\alpha}, d), \text{ for } n \text{ large enough}.$

**Construction of the Family $\mathcal{N}$.** Let $\eta(t)$ be an infinitely many times differentiable function on $\mathbb{R}^1$, such that supp $\eta = (-\pi, \pi)$, $\eta(t) \geq 0$, and max, $\eta(t) = 1$. Let $M$ be an integer, and let $0 < \tau < 1$,
\begin{equation}
r_0 = \left(\frac{2}{\pi(f_{\text{max}} + \lambda)}\right)^{1/2}, \quad \Delta = \left(\frac{2\delta_0}{b_1 + b_2}\right)^{1/\alpha}.
\end{equation}
Note that $\lambda < (\pi r_0^2)^{-1} < f_{\text{max}}$.

Consider the function
\begin{equation}
f_0(x) = \lambda_{\text{max}} I[r \leq r_0 - \Delta] + \left(\lambda - \frac{b_1 + b_2}{2}|r - r_0|^\alpha \text{sign}(r - r_0)\right)
\times I[-\Delta < r - r_0 \leq \Delta], \quad x = (r, \varphi),
\end{equation}
where \( \lambda_{\text{max}} \) is a positive constant chosen so that \( f_0 \) is a probability density and \( f_0 \in \mathcal{G}_{\gamma, \alpha} \). This is possible if \( \delta_0 \) is small enough. In fact, the equality \( \int f_0(x) \, dx = 1 \) reads as \( \lambda_{\text{max}} = (\pi \tau_2^2)^{-1} + O(\Delta) \), and thus \( \lambda < \lambda_{\text{max}} < f_{\text{max}} \) for \( \delta_0 \) small enough. Therefore, \( G_{f_0}(\lambda) = B(0, r_0) \), \( g_\alpha(\varphi) \equiv r_0 \), and, if \( |f_0(x) - \lambda| \leq \delta_0 \), we have

\[
|f_0(x) - \lambda| = \frac{b_1 + b_2}{2} |r - r_0|^\alpha = \frac{b_1 + b_2}{2} |r - g_\alpha(\varphi)|^\alpha.
\]

For \( j = 0, 1, \ldots, M \), denote \( \varphi_j = 2\pi j/M \),

\[
\eta_j(\varphi) = \tau M^{-\gamma} \eta(M(\varphi - (2j - 1)\pi/M)),
\]

\[
r_{*j}(\varphi) = r_0 - \eta_j(\varphi) \frac{(2b_2)^{1/\alpha}}{(2b_2)^{1/\alpha} - (b_1 + b_2)^{1/\alpha}},
\]

\[
r_j^*(\varphi) = r_0 + \eta_j(\varphi) \frac{(2b_1)^{1/\alpha}}{(b_1 + b_2)^{1/\alpha} - (2b_1)^{1/\alpha}}.
\]

We assume that \( \tau \) is small enough to guarantee that \( r_0 - \Delta < r_{*j}(\varphi), r_j^*(\varphi) < r_0 + \Delta \). Let \( \omega = (\omega_1, \ldots, \omega_M) \) be a vector with binary entries \( \omega_j \in \{0, 1\} \). Define the function

\[
f_1(x, \omega) = \sum_{j=1}^M \omega_j \left( \left( b_2 |r - r_0 + \eta_j(\varphi)|^{\alpha} - \frac{b_1 + b_2}{2} |r - r_0|^\alpha \right) \right.
\]

\[
\times I\{r_{*j}(\varphi) < r \leq r_0 - \eta_j(\varphi), \varphi_{j-1} \leq \varphi < \varphi_j \}
\]

\[
+ \left( \frac{b_1 + b_2}{2} |r - r_0|^\alpha \text{sign}(r - r_0) - b_1 |r - r_0 + \eta_j(\varphi)|^{\alpha} \right)
\]

\[
\times I\{r_0 - \eta_j(\varphi) < r \leq r_j^*(\varphi), \varphi_{j-1} \leq \varphi < \varphi_j \},
\]

where \( x = (r, \varphi) \). Clearly, \( f_1(x, \omega) \leq 0, \forall x, \omega \). Let

\[
\Delta_j = \{x = (r, \varphi): r_{*j}(\varphi) < r \leq r_j^*(\varphi), \varphi_{j-1} \leq \varphi < \varphi_j \}.
\]

**Lemma 5.** The function \( f_1(x, \omega) \) and the set \( \Delta_j \) satisfy

\[
\max_{x, \omega} |f_1(x, \omega)| = \max_{\omega} \max_x |f_1(x, \omega)| = O(\tau^{\alpha}M^{-\gamma}),
\]

\[
\mes(\Delta_j) = \mes(\Delta_1) = O(\tau M^{-\gamma - 1}),
\]

\[
\int f_1(x, \omega) \, dx = -C_\eta \tau^{\alpha+1} \sum_{j=1}^M \omega_j M^{-(\alpha+1)\gamma - 1},
\]

where the constant \( C_\eta > 0 \) depends only on \( \eta(\cdot) \).
PROOF. The definition of \( r_*(\varphi) \) and \( r^*(\varphi) \) implies

\[
\max_{\omega} \max_{x \in \Omega_j} |f_j(x, \omega)| \\
\leq \frac{b_1 + b_2}{2} \max_{\varphi_j - 1 \leq \varphi < \varphi_j} \left( |r_*(\varphi) - r_0|^\alpha + |r^*(\varphi) - r_0 + \eta(\varphi)|^\alpha \right) \\
= O\left( \max_{\varphi} |\eta(\varphi)|^\alpha \right) = O(\tau^\alpha M^{-\alpha \gamma}).
\]

Next, \( \text{mes}(\Delta_j) \leq 2\pi M^{-1} \max_{\varphi} \eta(\varphi) = O(\tau M^{-\gamma - 1}) \). Combining (24) and (25), we get \( \int f_1(x, \omega) \, dx = O(\tau^{\alpha+1} \sum_{j=1}^M \omega_j M^{-(\alpha+1)\gamma-1}) \), and it is easy to show that the constant in the \( O(\cdot) \) depends on \( \eta(\cdot) \) only. \( \square \)

For any vector \( \omega \) consider the function \( f(x, \omega) = f_0(x) + f_1(x, \omega) + f_2(x, \omega) \), where

\[
f_2(x, \omega) = \frac{1}{\pi(r_0 - \Delta)^2} C_{\eta} \tau^{\alpha+1} \sum_{j=1}^M \omega_j M^{-(\alpha+1)\gamma-1} I\{r \leq r_0 - \Delta\}.
\]

Since \( \int (f_1 + f_2) = 0, \forall \omega \), the function \( f(x, \omega) \) is a probability density. Moreover, for \( M \) large enough, \( \max_x f(x) < f_{\text{max}} \), and the \( \lambda \)-level set of \( f \) is that of \( f_0 + f_1 \), that is, the set

\[
G_{f(\cdot, \omega)}(\lambda) = \left\{ x = (r, \varphi) : 0 \leq r \leq r_0 - \sum_{j=1}^M \omega_j \eta(\varphi) \right\}.
\]

For \( \tau \) small enough the shape function \( g(\varphi) = r_0 - \sum_{j=1}^M \omega_j \eta(\varphi) \in \Sigma(\gamma, L) \), and hence \( G_{f(\cdot, \omega)}(\lambda) \in \mathcal{S}(\gamma, L, h) \). The definition of \( f_0 \) and \( f_1 \) implies that \( f_0 + f_1 \) is \( \alpha \)-regular around the level \( \lambda \), and thus \( f \) is \( \alpha \)-regular.

We conclude that \( f(\cdot, \omega) \in \mathcal{S}_{\gamma, \alpha} \) for any \( \omega \). Define

\[
\mathcal{N} = \{ f(x, \omega) : \omega \in \Omega \},
\]

where \( \Omega \) is an appropriately chosen set of binary vectors. The definition of this set is different for \( d = d_1 \) and \( d = d_2 \).

PROOF OF THE LOWER BOUND FOR THE CASE \( d = d_1 \). Set \( M = \lceil n^{1/(\gamma(2\alpha + 1) + 1)} \rceil + 1 \). A well-known combinatorial result [see, e.g., Korostelev and Tsybakov (1993b), Lemma 2.7.4] guarantees the existence of at least \( s = \lceil 2^M / 5 \rceil \) binary vectors \( \omega^1, \ldots, \omega^s \) of length \( M \) such that

\[
\sum_{j=1}^M |\omega^l_j - \omega^k_j| \geq M/16, \quad 1 \leq l, k \leq s, k \neq l,
\]

where \( \omega^l = (\omega^l_1, \ldots, \omega^l_M) \), that is, \( \omega^l_j \) is the \( j \)th component of \( \omega^l \).

Choose \( \mathcal{N} \) as the set of densities \( f(x, \omega) \), with \( \Omega = \{ \omega^1, \ldots, \omega^s \} \) such that (27) holds. The condition (22) for the set \( \mathcal{N} \) is satisfied with \( \varepsilon = \)
(r_0\tau/\eta/32)M^{-\gamma}, \text{ since, in view of (27),}
\begin{align*}
d_1(G_{f^{(\cdot, \omega^j)}(\lambda)}, G_{f^{(\cdot, \omega^k)}(\lambda)}) &= \sum_{j=1}^{M} |\omega^j - \omega^k| \int_{\phi_{j-1}}^{\phi_j} d\varphi \int_{r_0 - \eta(\varphi)}^{r_0} r \, dr \\
&\geq (M/16) \left( r_0 \tau M^{-\gamma - 1} \int \eta(u) \, du + O(M^{-2\gamma - 1}) \right) \\
&\geq \left( r_0 \tau \int \eta \sqrt[32]{\eta} \right) M^{-\gamma},
\end{align*}
for M large enough.

Now, for every pair of probability measures \( P^l = P_{f^{(\cdot, \omega^l)}}, P^k = P_{f^{(\cdot, \omega^k)}}, \)
\begin{equation}
I(P^l, P^k) = \int \log \frac{dP^l}{dP^k} dP^l = n \int \log \frac{f(x, \omega^l)}{f(x, \omega^k)} f(x, \omega^l) \, dx
\end{equation}
(28)
\begin{align*}
&= n \int \log \frac{\tilde{f}_l(x) + \tilde{\tilde{f}}_l(x)}{f_0(x) + \tilde{f}_k(x)} \left( f_0(x) + \tilde{\tilde{f}}_l(x) \right) \, dx,
\end{align*}
where \( \tilde{f}_l(x) = f_1(x, \omega^l) + f_2(x, \omega^l), \) for brevity. Since \( f_0 + \tilde{f}_l, f_0 + \tilde{f}_k \) are uniformly bounded from below by \( \lambda - \delta_0 > 0 \) on their support, it is clear that
\begin{align*}
\left| \log \frac{f_0 + \tilde{f}_l}{f_0 + \tilde{f}_k} - \frac{1}{f_0 + \tilde{f}_k} (\tilde{f}_l - \tilde{f}_k) \right| &\leq c_{13} (\tilde{f}_l - \tilde{f}_k)^2,
\end{align*}
where \( c_{13} > 0. \) Since also \( \int \tilde{f}_l(x) \, dx = \int \tilde{f}_k(x) \, dx = 0, \) we get, with some \( c_{14} > 0, \)
\begin{equation}
I(P^l, P^k) \leq c_{14} n \int \left( \tilde{f}_l(x) - \tilde{f}_k(x) \right)^2 \, dx
\end{equation}
(29)
\begin{align*}
&\leq c_{14} n \left( \sum_{j=1}^{M} |\omega^j - \omega^k| \max_{x, \omega} \left| f_1(x, \omega) \right|^2 \\
&\quad + C_{15}^2 \tau^2 \frac{1}{\pi(r_0 - \Delta)^2} \left( \sum_{j=1}^{M} |\omega^j - \omega^k| \right)^2 M^{-2(\alpha + 1)\gamma - 2} \right)
\end{align*}
\begin{align*}
&= O(n(\tau M^{-\gamma})^{2(\alpha + 1)}),
\end{align*}
where we used Lemma 5. Since \( s = [2^{M/8}], \) we find
\begin{align*}
I(P^l, P^k)/(\log s) &= O(n\tau^{2\alpha + 1} M^{-\gamma(2\alpha + 1) - 1}) \leq c_{15} \tau^{2\alpha + 1}.
\end{align*}
Here \( c_{15} > 0 \) does not depend on \( k \) and \( l, \) and as \( \tau \) can be chosen arbitrarily small, (23) is satisfied. This proves the lower bound if \( d = d_1. \) \( \square \)

PROOF OF THE LOWER BOUND FOR THE CASE \( d = d_\infty. \) Set \( M = [(n/\log n)^{1/(\gamma(2\alpha + 1) + 1)}] + 1. \) We use again Lemma 4, but we choose a different set \( \Omega \) in the definition of the family \( \mathcal{Y}: \Omega = \{\omega^{(1)}, \ldots, \omega^{(M)}\}, \) where \( \omega^{(j)} \) is the binary vector of length \( M \) whose \( j \)th component is 1 and other components
are 0. The condition (22) for this family $\mathcal{N}$ is satisfied with $\varepsilon = \tau M^{-\gamma}$, since $d_\alpha(G_{f^{(i)}}, G_{\hat{f}^{(i)}}, \mathcal{F}_{\varepsilon}^{(i)}); \tau = \tau M^{-\gamma}$ max, $\eta_\tau(t) = \tau M^{-\gamma}$. To check (23) we act similarly to the case of $d = d_1$. Denote $P_{\hat{f}^{(i)}} = P_{\hat{f}^{(i)}}, \tau$, $\tau = \tau M^{-\gamma}$ max, $\eta_\tau(t) = \tau M^{-\gamma}$. Note that $f_2(x, \omega^{(k)}) = f_2(x, \omega^{(l)})$. As in (29),

$$I(P^{(l)}, P^{(k)}) = c_{14} n \int \left( \tilde{f}_1(x) - \tilde{f}_k(x) \right)^2 \, dx$$

$$= c_{14} n \int \left( f_1^2(x, \omega^{(l)}) + f_2^2(x, \omega^{(k)}) \right) \, dx$$

$$\leq 2c_{14} n \max \{ f_1(x, \omega^{(l)}) \}^2$$

$$= O(n^{(2\alpha+1)}M^{-\gamma(2\alpha+1)-1}) = O(\tau^{2\alpha+1} \log n).$$

Since $s = M$, (23) is satisfied for $\tau$ small enough. □

**Proof of Theorem 2.** As noted in Section 4, we need to show only the lower bound. Also, it suffices to consider the case $\alpha > 0$, since for $\alpha = 0$ the proof of the lower bound follows the same lines as in Mammen and Tsybakov (1995).

To prove the lower bound of Theorem 2, we use the argument of Theorem 1 (case $d = d_1$). The only difference is in the choice of the functions $\eta_j$. We choose them so that the level sets are convex, namely

$$\eta_j(\phi) = r_0 \left( 1 - \frac{\cos(\pi/M)}{\cos(\phi - (2j-1)\pi/M)} \right), \quad \phi_{j-1} \leq \phi < \phi_j.$$

Then the shape function $g(\phi) = r_0 - \sum_{j=1}^M \omega_j \eta_j(\phi)$ bounds a convex set, and the function $f(x, \omega)$ belongs to $\mathcal{F}_{\text{conv}, \alpha}$ for every $\omega \in \Omega$. It is easy to show that, as $M \to \infty$, one has

$$\eta_j(\phi) = M^{-2} \eta(M(\phi - (2j-1)\pi/M)) + o(M^{-2}),$$

where $\eta(u) = r_0(\pi^2 - u^2)I(\{u \leq \pi\})$, and $o(\cdot)$ is uniform in $\phi$. Thus, all the calculations are carried out similarly to the proof of the lower bound in Theorem 1, with $\gamma = 2$. However, since now $\tau = 1$, we have to choose $M$ in a slightly different form: $M = [\tau r_0^{1/(4\alpha+3)}] + 1$, where the factor $\tau > 0$ is large enough to guarantee that the left-hand side of (29) is less than $1/2$. □

**Proof of Theorem 3.** We use Lemma 4, with $s = 2$. Thus, it suffices to define two densities, $\tilde{f}_0$ and $\tilde{f}_1$, belonging to $\mathcal{F}_{\text{conv}, \alpha}$, such that

$$d_\alpha(G_{\tilde{f}_0}(\lambda), G_{\tilde{f}_1}(\lambda)) \geq \varepsilon,$$

$$I(P_{\tilde{f}_0}, P_{\tilde{f}_1}) \leq (\log 2)/2,$$

where $\varepsilon \sim n^{-1/(2\alpha+2)}$.

Choose $r_0 = (2/\pi(\lambda + 1/\mu_0))^{1/2}$. Then $\pi r_0^2 > \mu_0$. Let $\Delta > 0$ be small enough that $\pi(r_0 - \Delta)^2 > \mu_0$. Define $\vartheta = 2 \arccos(r_0/(r_0 + \Delta))$, and define
the shape function
\[
\bar{g}(\varphi) = \begin{cases} 
    r_0[\cos(\varphi/2 - |\varphi/2 - \varphi|)]^{-1}, & 0 \leq \varphi < \vartheta, \\
    r_0, & \vartheta \leq \varphi \leq 2\pi.
\end{cases}
\]
Let
\[
\tilde{f}_0(x) = \lambda_{\max} I[r \leq \bar{g}(\varphi) - \Delta] + \left(\lambda - \frac{b_1 + b_2}{2} |r - \bar{g}(\varphi)|^\alpha \text{sign}(r - \bar{g}(\varphi))\right) \times I[-\Delta < r - \bar{g}(\varphi) \leq \Delta],
\]
where \(x = (r, \varphi)\) and \(\lambda_{\max}\) is a positive constant chosen so that \(\tilde{f}_0\) is a probability density, \(\tilde{f}_0\) is \(\alpha\)-regular around the level \(\lambda\) and \(\tilde{f}_0(x) \leq f_{\max}\). Arguing as in the case of \(f_0\) (see the proof of lower bounds in Theorem 1), one easily shows that this is possible if \(\delta_0\) and \(\Delta\) are small enough (but fixed). Now, the level set \(G_\lambda(f_0)\) is convex since it is defined by the shape function \(\bar{g}\). Also, \(\text{mes}(G_\lambda(f_0)) > \pi r_0^2 > \mu_0\). Hence \(\tilde{f}_0 \in \mathcal{F}_{\text{conv}, \alpha}\).

Let us define \(\tilde{f}_1\). Denote \(M = [\tau_0 n^{1/(2\alpha + 2)}] + 1\), where \(\tau_0 > 0\). Introduce the function
\[
\bar{\eta}(\varphi) = r_0 \left( \left[ \cos(\varphi/2 - |\varphi/2 - \varphi|) \right]^{-1} - \left[ \cos(\varphi/2 - \pi/M) \right]^{-1} \right) \times I[|\varphi/2 - \varphi| \leq \pi/M].
\]
Clearly, \(\bar{g}(\varphi) - \bar{\eta}(\varphi)\) is the shape function of the convex set \(G_{f_0}(\lambda) \cap B(0, \bar{r})\), where \(\bar{r} = r_0[\cos(\varphi/2 - \pi/M)]^{-1}\), whose Lebesgue measure is greater than \(\mu_0\) for \(M\) large. As \(M \to \infty\), one has
\[
\bar{\eta}(\varphi) = M^{-1} \eta(M(\varphi - \vartheta/2)) + o(M^{-1}),
\]
where \(\eta(u) = c_{16}(\pi - |u|)I[|u| \leq \pi], c_{16} > 0\) and \(o(\cdot)\) is uniform in \(\varphi\). Now, one can define \(\tilde{f}_1\) similarly to \(f_1(x, \omega)\). Set
\[
\begin{align*}
    r_*(\varphi) &= \bar{g}(\varphi) - \bar{\eta}(\varphi) = \frac{(2b_2)^{1/\alpha}}{(2b_2)^{1/\alpha} - (b_1 + b_2)^{1/\alpha}}, \\
    r^*(\varphi) &= \bar{g}(\varphi) + \bar{\eta}(\varphi) = \frac{(2b_1)^{1/\alpha}}{(b_1 + b_2)^{1/\alpha} - (2b_2)^{1/\alpha}},
\end{align*}
\]
assuming that \(M\) is large enough to guarantee that \(\bar{g}(\varphi) - \Delta < r_*(\varphi)\), \(r^*(\varphi) < \bar{g}(\varphi) + \Delta\), and let
\[
\tilde{f}_{1,1}(x) = \left(b_2 |r - \bar{g}(\varphi) + \bar{\eta}(\varphi)|^\alpha - \frac{b_1 + b_2}{2} |r - \bar{g}(\varphi)|^\alpha\right) \times I[r_*(\varphi) < r \leq \bar{g}(\varphi) - \bar{\eta}(\varphi)] + \left(\frac{b_1 + b_2}{2} |r - \bar{g}(\varphi)|^\alpha - b_1 |r - \bar{g}(\varphi) + \bar{\eta}(\varphi)|^\alpha\right) \times I[\bar{g}(\varphi) - \bar{\eta}(\varphi) < r \leq r^*(\varphi)], \quad x = (r, \varphi). 
\]
As for $f_1(x, \omega)$, one proves that $\tilde{f}_{1,1}$ is $\alpha$-regular around the level $\lambda$. Also, define the function $\tilde{f}_{1,2}$ proportional to $I[r \leq g(\varphi) - \Delta]$ and such that $\tilde{f}_{1,1} + \tilde{f}_{1,2}$ integrates to 0. In particular, $\max_x \tilde{f}_{1,2}(x) = O(M^{-\alpha-2})$. Finally, put $\tilde{f}_1 = \tilde{f}_0 + \tilde{f}_{1,1} + \tilde{f}_{1,2}$. Note that $\tilde{f}_1$ is $\alpha$-regular around the level $\lambda$ and $G_{\tilde{f}}(\lambda) = G_f(\lambda) \cap B(0, \tilde{r})$ for $M$ large enough. Thus, $\tilde{f}_1 \in \mathcal{F}_{\text{conv}, \alpha}$ for $M$ large enough.

By the construction, $d_{\alpha}(G_f(\lambda), G_{\tilde{f}}(\lambda)) = M^{-1} \tilde{\eta}(0) = c_{16}\pi M^{-1}(1 + o(1))$, which gives the desired value of $\varepsilon$ in (30). Next, as in (29),

$$I(P_0, P_1)$$

$$\leq c_{14} n \int \tilde{f}_1^2(x) \, dx = c_{14} n \int \left( \tilde{f}_{1,1}^2(x) + \tilde{f}_{1,2}^2(x) \right) \, dx$$

$$\leq 2c_{14} n \left( \text{mes}\{(r, \varphi) : r_*(\varphi) < r \leq r^*(\varphi)\} \max_x |\tilde{f}_{1,1}(x)|^2 + O(M^{-2(\alpha+2)}) \right)$$

$$= O(nM^{-2\alpha-2}),$$

and choosing $\tau_0$ large enough, we get (31). □

REFERENCES


