

---

# Adaptive Hausdorff Estimation of Density Level Sets

---

Aarti Singh and Robert D. Nowak\*

University of Wisconsin - Madison, USA

singh@cae.wisc.edu, nowak@enr.wisc.edu

Clayton D. Scott

University of Michigan - Ann Arbor, USA

cscott@eecs.umich.edu

## Abstract

Hausdorff accurate estimation of density level sets is relevant in applications where a spatially uniform mode of convergence is desired to ensure that the estimated set is close to the target set at all points. The minimax optimal rate of error convergence for the Hausdorff metric is known to be  $(n/\log n)^{-1/(d+2\alpha)}$  for level sets with boundaries that have a Lipschitz functional form, and where the parameter  $\alpha$  characterizing the regularity of the density around the level of interest is known. Thus, all previous work is non-adaptive to the density regularity and assumes knowledge of the regularity parameter  $\alpha$ . Moreover, the estimators proposed in previous work achieve the minimax optimal rate for rather restricted classes of sets (for example, the boundary fragment and star-shaped sets) that effectively reduce the set estimation problem to a function estimation problem. This characterization precludes level sets with multiple connected components, which are fundamental to many applications. This paper presents a fully data-driven procedure that is adaptive to unknown local density regularity, and achieves minimax optimal Hausdorff error control for a class of level sets with very general shapes and multiple connected components.

## 1 Introduction

Density level sets provide useful summaries of a density function for many applications including clustering [Har75, Stu03], anomaly detection [SHS05, SN06, VV06], and data ranking [LPS99]. In practice, however, the density function itself is unknown a priori and only a finite number of observations from the density are available. Let  $X_1, \dots, X_n$  be independent, identically distributed observations drawn from an unknown probability measure  $P$ , having density  $f$  with respect to the Lebesgue measure, and defined on the domain  $\mathcal{X} \subseteq \mathbb{R}^d$ . Given a desired density level  $\gamma$ , consider the  $\gamma$ -level set of the density  $f$

$$G_\gamma^* := \{x \in \mathcal{X} : f(x) \geq \gamma\}.$$

---

\*This work was partially supported by the National Science Foundation grants CNS-0519824 and ECS-0529381.

The goal of the density level set estimation problem is to generate an estimate  $\hat{G}$  of the level set based on the  $n$  observations  $\{X_i\}_{i=1}^n$ , such that the error between the estimate  $\hat{G}$  and the target set  $G_\gamma^*$ , as assessed by some performance measure which gauges the closeness of the two sets, is small.

Most literature available on level set estimation [SHS05, SN06, SD07, WN07, KT93, Tsy97, Pol95, RV06] considers global error measures related to the symmetric set difference. However some applications may need a more local or spatially uniform error measure as provided by the Hausdorff metric, for example, to ensure robustness to outliers or preserve topological properties of the level set. The Hausdorff error metric is defined as follows between two non-empty sets:

$$d_\infty(G_1, G_2) = \max\left\{\sup_{x \in G_2} \rho(x, G_1), \sup_{x \in G_1} \rho(x, G_2)\right\}$$

where  $\rho(x, G) = \inf_{y \in G} \|x - y\|$ , the smallest Euclidean distance of a point in  $G$  to the point  $x$ . If  $G_1$  or  $G_2$  is empty, then let  $d_\infty(G_1, G_2)$  be defined as the largest distance between any two points in the domain. Control of this error measure provides a uniform mode of convergence as it implies control of the deviation of a single point from the desired set. A symmetric set difference based estimator may not provide such a uniform control as it is easy to see that a set estimate can have very small symmetric difference error but large Hausdorff error. Conversely, as long as the set boundary is not space-filling, small Hausdorff error implies small symmetric difference error.

Existing results pertaining to nonparametric level set estimation using the Hausdorff metric [KT93, Tsy97, Cav97] focus on rather restrictive classes of level sets (for example, the boundary fragment and star-shaped set classes). These restrictions, which effectively reduce the set estimation problem to a boundary function estimation problem (in rectangular or polar coordinates, respectively), are typically not met in practical applications. In particular, the characterization of level set estimation as a boundary function estimation problem precludes level sets with multiple connected components, which are fundamental to many applications. Moreover, the estimation techniques proposed in [KT93, Tsy97, Cav97] require precise knowledge of the local regularity of the distribution (quantified by the parameter  $\alpha$ , to be defined below) in the vicinity of the desired level set in order to achieve minimax optimal rates of convergence. Such prior knowledge is unavailable in most practical ap-

plications. Recently, a plug-in method based on sup-norm density estimation was put forth in [CMC06] that can handle more general classes than boundary fragments or star-shaped sets, however sup-norm based methods require global smoothness assumptions on the density to ensure that the density estimate is good everywhere. Also, the method only deals with a special case of the density regularity condition considered in this paper ( $\alpha = 1$ ), and is therefore not adaptive to unknown density regularity.

In this paper, we propose a plug-in procedure based on a regular histogram partition that can adaptively achieve minimax optimal rates of Hausdorff error convergence over a broad class of level sets with very general shapes and multiple connected components, without assuming *a priori* knowledge of the density regularity parameter  $\alpha$ . Adaptivity is achieved by a new data-driven procedure for selecting the histogram resolution. The procedure is specifically designed for the level set estimation problem and only requires local regularity of the density in the vicinity of the desired level.

## 2 Density assumptions

In this paper, we assume that the density  $f$  is supported on the unit hypercube in  $d$ -dimensions, that is  $\mathcal{X} = [0, 1]^d$ , and is bounded with range  $[0, f_{\max}]$ . Controlling the Hausdorff accuracy of level set estimates also requires some smoothness assumptions. The most crucial assumption is the first, which characterizes the relationship between distances and changes in density. The last two are topological assumptions on the level set and essentially generalize the notion of Lipschitz functions to closed hypersurfaces.

Here we define an  $\epsilon$ -ball centered at a point  $x$  as  $B(x, \epsilon) = \{y \in \mathcal{X} : \|x - y\| \leq \epsilon\}$ , where  $\|\cdot\|$  denotes the Euclidean distance. Also, an *inner*  $\epsilon$ -cover of a set  $G$  is defined as the union of all  $\epsilon$ -balls contained in  $G$ . Formally,  $\mathcal{I}_\epsilon(G) = \bigcup_{x: B(x, \epsilon) \subseteq G} B(x, \epsilon)$ .

**[A] Local density regularity:** The density is  $\alpha$ -regular around the  $\gamma$ -level set,  $0 < \alpha < \infty$  and  $\gamma < f_{\max}$ , if there exist constants  $C_2 > C_1 > 0$  and  $\delta_0, \delta_1 > 0$  such that

$$C_1 \rho(x, \partial G_\gamma^*)^\alpha \leq |f(x) - \gamma| \leq C_2 \rho(x, \partial G_\gamma^*)^\alpha$$

for all  $x \in \mathcal{X}$  with  $|f(x) - \gamma| \leq \delta_0$ , where  $\partial G_\gamma^*$  is the boundary of the true level set  $G_\gamma^*$ . And there exists  $y_0 \in \partial G_\gamma^*$  such that for all  $x \in B(y_0, \delta_1)$ ,  $|f(x) - \gamma| \leq \delta_0$ .

This assumption is similar to the one used in [Tsy97, Cav97] (we elaborate on the differences later on). The regularity parameter  $\alpha$  determines the rate of error convergence for level set estimation. Accurate estimation is more difficult at levels where the density is relatively flat (large  $\alpha$ ), as intuition would suggest. In this paper, we do not assume knowledge of  $\alpha$  unlike previous investigations into Hausdorff accurate level set estimation [KT93, Tsy97, Cav97, CMC06]. Therefore, here the assumption simply states that there is a relationship between distance and density level, but the precise nature of the relationship is unknown.

**[B] Level set regularity:** There exist constants  $\epsilon_o > 0$  and  $C_3 > 0$  such that for all  $\epsilon \leq \epsilon_o$ ,  $\mathcal{I}_\epsilon(G_\gamma^*) \neq \emptyset$  and  $\rho(x, \mathcal{I}_\epsilon(G_\gamma^*)) \leq C_3 \epsilon$  for all  $x \in \partial G_\gamma^*$ .

This assumption states that the level set is not arbitrarily narrow anywhere. It precludes features like cusps and arbitrarily thin ribbons, as well as connected components of arbitrarily small size. This condition is necessary since arbitrarily small features cannot be detected and resolved from a finite sample. However, from a practical perspective, if the assumption fails to hold then it simply means that it is not possible to theoretically guarantee that such small features will be recovered.

**[C] Level set boundary dimension:** There exists a constant  $C_4 > 0$  such that for all  $x \in \partial G_\gamma^*$  and all  $\epsilon, \delta$  such that  $0 < \delta \leq \epsilon$ , the minimum number of  $\delta$ -balls required to cover  $\partial G_\gamma^* \cap B(x, \epsilon)$  is  $\leq C_4 (\delta/\epsilon)^{-(d-1)}$ .

This assumption is related to the box-counting dimension [Fal90] of the boundary of the level set. It essentially says that, at any scale, the boundary behaves locally like a  $(d - 1)$ -dimensional surface in the  $d$ -dimensional domain and is not space-filling. This condition is not restrictive since the Hausdorff error itself is inappropriate for space-filling curves, and in fact it is not required if the density regularity parameter  $\alpha$  is known. However, the condition is needed to achieve adaptivity using the proposed method, as we shall discuss later.

Let  $\mathcal{F}_1^*(\alpha)$  denotes the class of densities satisfying assumptions **[A, B]**, and  $\mathcal{F}_2^*(\alpha)$  denotes the class of densities satisfying assumptions **[A, B, C]**. The dependence on other parameters is omitted as these do not influence the minimax optimal rate of convergence. The classes  $\mathcal{F}_1^*(\alpha), \mathcal{F}_2^*(\alpha)$  are a generalization of the Lipschitz boundary fragments or star-shaped sets considered in [KT93, Tsy97, Cav97] since assumptions **[B, C]** basically imply that the boundary looks locally like a Lipschitz function. In fact assumptions **[B, C]** are satisfied by a Lipschitz boundary fragment or star-shaped set; please refer to Section 5.3 for a formal proof. However, there is a slight difference between the upper bound of assumption **[A]** here and that employed in [Tsy97, Cav97]. The upper bound assumption in [Tsy97, Cav97] only requires that the set  $\{x : |f(x) - \gamma| \leq \delta_0\}$  be non-empty. So as long as there is at least one point on the boundary where the density regularity assumption **[A]** holds, this determines the complexity of the class. Our assumption requires the density regularity to hold for an open neighborhood about at least one point on the boundary. This is necessary for adaptivity since a procedure cannot sense the regularity as characterized by  $\alpha$  unless the regularity holds in a region with positive measure.

In [Tsy97], Tsybakov established a minimax lower bound of  $(n/\log n)^{-\frac{1}{d+2\alpha}}$  for the class of Lipschitz star-shaped sets, which satisfy our assumptions **[B, C]** (see Section 5.3) and the slightly modified version of assumption **[A]**, as discussed above. His proof uses Fano's lemma to derive the lower bound for a discrete subset of densities from this class. It is easy to see that the discrete subset of densities

used in his construction also satisfy our form of assumption **[A]**. Hence, the same lower bound holds for the classes  $\mathcal{F}_1^*(\alpha)$  and  $\mathcal{F}_2^*(\alpha)$  under consideration as well and we have the following proposition. Here  $\mathbb{E}$  denotes expectation with respect to the random data sample.

**Proposition 1** *There exists  $c > 0$  such that*

$$\begin{aligned} \inf_{\widehat{G}_n} \sup_{f \in \mathcal{F}_1^*(\alpha)} \mathbb{E}[d_\infty(\widehat{G}_n, G_\gamma^*)] &\geq \inf_{\widehat{G}_n} \sup_{f \in \mathcal{F}_2^*(\alpha)} \mathbb{E}[d_\infty(\widehat{G}_n, G_\gamma^*)] \\ &\geq c \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}. \end{aligned}$$

Here the inf is taken over all possible set estimators  $\widehat{G}_n$ .

In the paper, we present a method that achieves this minimax lower bound for the class  $\mathcal{F}_1^*(\alpha)$ , given knowledge of the density regularity parameter  $\alpha$ . We also extend the method to achieve adaptivity to  $\alpha$  for the class  $\mathcal{F}_2^*(\alpha)$  under the additional assumption **[C]**, while preserving the minimax optimal rate of convergence.

### 3 Proposed method

In this section, we propose a plug-in level set estimator that is based on a regular histogram. The histogram resolution is adaptively selected in a purely data-driven way without assuming knowledge of the local density regularity.

Let  $\mathcal{A}_j$  denote the collection of cells in a regular partition of  $[0, 1]^d$  into hypercubes of dyadic sidelength  $2^{-j}$ , where  $j$  is a non-negative integer. The estimator at this resolution is given as

$$\widehat{G}_j = \{A \in \mathcal{A}_j : \widehat{f}(A) \geq \gamma\}. \quad (1)$$

Here  $\widehat{f}(A) = \widehat{P}(A)/\mu(A)$ , where  $\widehat{P}(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \in A\}}$  denotes the empirical probability of an observation occurring in  $A$  and  $\mu$  is the Lebesgue measure.

Our first result shows that, if the density regularity parameter  $\alpha$  is known, then the correct resolution can be chosen (as in [Tsy97, Cav97]), and the corresponding estimator achieves near minimax optimal rate over the class of densities given by  $\mathcal{F}_1^*(\alpha)$ . Here  $\mathbb{E}$  denotes expectation with respect to the random data sample. We introduce the notation  $a_n \asymp b_n$  to denote that  $a_n = O(b_n)$  and  $b_n = O(a_n)$ .

**Theorem 1** *Assume that the local density regularity  $\alpha$  is known. Pick resolution  $j$  such that  $2^{-j} \asymp s_n(n/\log n)^{-\frac{1}{(d+2\alpha)}}$ , where  $s_n$  is a monotone diverging sequence. Then*

$$\sup_{f \in \mathcal{F}_1^*(\alpha)} \mathbb{E}[d_\infty(\widehat{G}_j, G_\gamma^*)] \leq C s_n \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}$$

for all  $n$ , where  $C \equiv C(C_1, C_3, \epsilon_o, f_{\max}, \delta_0, d, \alpha) > 0$  is a constant.

The proof is given in Section 5.1.

Theorem 1 provides an upper bound on the Hausdorff error of our estimate. If  $s_n$  is slowly diverging, for example if  $s_n = (\log n)^\epsilon$  where  $\epsilon > 0$ , this upper bound agrees with the minimax lower bound of Proposition 1 up to a  $(\log n)^\epsilon$

factor. Hence the proposed estimator can achieve near minimax optimal rates, given knowledge of the density regularity. We would like to point out that if the parameter  $\delta_0$  characterizing assumption **[A]** and the density bound  $f_{\max}$  are also known, then the appropriate resolution can be chosen as  $j = \lfloor \log_2(c^{-1}(n/\log n)^{1/(d+2\alpha)}) \rfloor$ , where the constant  $c \equiv c(\delta_0, f_{\max})$ . With this choice, the optimal sidelength scales as  $2^{-j} \asymp (n/\log n)^{-1/(d+2\alpha)}$ , and the estimator  $\widehat{G}_j$  exactly achieves the minimax optimal rate.

#### 3.1 Adaptivity to unknown local density regularity

In this section we present a procedure that automatically selects the appropriate resolution  $j$  without prior knowledge of  $\alpha$ . The selected resolution needs to be adapted to the local regularity of the density around the level of interest. To achieve this, we propose the following *vernier*:

$$\mathcal{V}_{\gamma,j} = \min_{A \in \mathcal{A}_j} \max_{A' \in \mathcal{A}_{j'} \cap A} |\gamma - \bar{f}(A')|.$$

Here  $\bar{f}(A) = P(A)/\mu(A)$ , and  $j' = \lfloor j + \log_2 s_n \rfloor$ , where  $s_n$  is a slowly diverging monotone sequence, for example  $\log n$ ,  $\log \log n$ , etc. Hence  $\mathcal{A}_{j'} \cap A$  denotes the collection of subcells with sidelength  $2^{-j'} \in [2^{-j}/s_n, 2^{-j+1}/s_n)$  within the cell  $A$ . The vernier focuses on cells  $A$  at resolution  $j$  that intersect the boundary (have smallest density deviation from the desired level  $\gamma$ ), and then evaluates the deviation in average density within subcells of  $A$  to judge whether or not the density is uniformly close to  $\gamma$  over the cell. Thus, the vernier is sensitive to the local density regularity in the vicinity of the desired level and in fact minimizing the vernier leads to selection of the appropriate resolution adapted to the unknown density regularity parameter  $\alpha$ . By choosing  $s_n$  with arbitrarily slow divergence, it is possible to get arbitrarily close to the optimal rate of convergence in the Hausdorff sense. However, note that the vernier may not function properly if the boundary of  $G_\gamma^*$  passes through every subcell of  $A$  (since then the subcell averages may be arbitrarily close to  $\gamma$  irrespective of the density regularity). Assumption **[C]** precludes this possibility at sufficiently high resolutions.

Since  $\mathcal{V}_{\gamma,j}$  requires knowledge of the unknown probability measure, we must work with the empirical version, defined analogously as:

$$\widehat{\mathcal{V}}_{\gamma,j} = \min_{A \in \mathcal{A}_j} \max_{A' \in \mathcal{A}_{j'} \cap A} |\gamma - \widehat{f}(A')|.$$

We propose a complexity regularization scheme wherein the empirical vernier  $\widehat{\mathcal{V}}_{\gamma,j}$  is balanced by a penalty term:

$$\Psi_{j'} := \max_{A \in \mathcal{A}_{j'}} \sqrt{8 \frac{\log(2^{j'(d+1)} 16)}{\delta} \max \left( \widehat{f}(A), 8 \frac{\log(2^{j'(d+1)} 16)}{\delta} \right) n \mu(A)}$$

where  $0 < \delta < 1$  is a confidence parameter, and  $\mu(A) = 2^{-j'd}$ . Notice that the penalty is computable from the given observations. The precise form of  $\Psi$  is chosen so that minimizing the empirical vernier plus penalty provides control over the true vernier (refer to Section 5.2 for a formal proof). The final level set estimate is given by

$$\widehat{G} = \widehat{G}_{\widehat{j}} \quad (2)$$

where

$$\hat{j} = \arg \min_{0 \leq j \leq J} \left\{ \widehat{\mathcal{V}}_{\gamma, j} + \Psi_{j'} \right\} \quad (3)$$

Thus the search is focused on regular partitions of dyadic sidelength  $2^{-j}$ ,  $j \in \{0, 1, \dots, J\}$ . The choice of  $J$  will be specified below. Observe that the value of the empirical vernier decreases with increasing resolution as better approximations to the true level are available. On the other hand, the penalty is designed to increase with resolution to penalize high complexity estimates that might overfit the given sample of data. Thus, the above procedure chooses the appropriate resolution automatically by balancing these two terms.

We now establish that our complexity penalized procedure leads to minimax optimal rates of convergence without requiring prior knowledge of any parameters.

**Theorem 2** *Pick  $J \equiv J(n)$  such that  $2^{-J} \asymp s_n(n/\log n)^{-\frac{1}{d}}$ , where  $s_n$  is a monotone diverging sequence. Let  $\hat{j}$  denote the resolution chosen by the complexity penalized method as given by Eq. (3), and  $\widehat{G}$  denote the final estimate of Eq. (2). Then with probability at least  $1 - 3/n$ , for all densities in the class  $\mathcal{F}_2^*(\alpha)$ ,*

$$c_1 s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}} \leq 2^{-\hat{j}} \leq c_2 s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}$$

for  $n$  large enough (so that  $s_n > c(C_4, d)$ ), where  $c_1, c_2 > 0$  are constants. In addition,

$$\sup_{f \in \mathcal{F}_2^*(\alpha)} \mathbb{E}[d_\infty(\widehat{G}, G_\gamma^*)] \leq C s_n^2 \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}$$

for all  $n$ , where  $C \equiv C(C_1, C_2, C_3, C_4, \epsilon_0, f_{\max}, \delta_0, \delta_1, d, \alpha) > 0$  is a constant.

The proof is given in Section 5.2.

The maximum resolution  $2^J \asymp s_n^{-1}(n/\log n)^{\frac{1}{d}}$  can be easily chosen, based only on  $n$ , and allows the optimal resolution for any  $\alpha$  to lie in the search space. Observe that by appropriate choice of  $s_n$ , for example  $s_n = (\log n)^{\epsilon/2}$  with  $\epsilon$  a small number  $> 0$ , the bound of Theorem 2 matches the minimax lower bound of Proposition 1, except for an additional  $(\log n)^\epsilon$  factor. Hence our method *adaptively* achieves near minimax optimal rates of convergence for the class  $\mathcal{F}_2^*(\alpha)$ .

## 4 Concluding Remarks

In this paper, we propose a Hausdorff accurate level set estimation method that is adaptive to unknown local density regularity and achieves minimax optimal rates of error convergence over a very general classes of level sets. The analysis in this paper assumes  $\alpha > 0$ , however the case  $\alpha \geq 0$  that allows jumps in the density can also be handled (see [SSN07]), but is omitted here to keep the presentation and proofs simpler. Also, this paper considers locally Lipschitz boundaries, however extensions to additional boundary smoothness (for example, Hölder regularity  $> 1$ ) may be possible in the proposed framework using techniques such as wedgelets [Don99] or curvelets [CD99]. The earlier work on Hausdorff accurate level set estimation [KT93, Tsy97, Cav97] does address higher smoothness of the boundary but that follows as a

straightforward consequence of assuming a functional form for the boundary. We would also like to comment that while we only addressed the density level set problem in this paper, extensions to general regression level set estimation should be possible using a similar approach.

The complexity regularization approach (Eq. 3) based on the vernier is similar in spirit to the so-called Lepski methods (for example, [LMS97]) for function estimation which are spatially adaptive bandwidth selectors, however the vernier focuses on cells close to the desired level and thus is specifically tailored to the level set problem. The vernier provides the key to achieve adaptivity while requiring only local regularity of the density in the vicinity of the desired level.

In this paper, we assume that the density regularity is the same everywhere along the level set. This might be somewhat restrictive, particularly if the level set consists of multiple components. Adaptivity to spatial variations in the density regularity can be achieved using a spatially adapted partition instead of a regular histogram partition. This might be possible by developing a tree-based approach or a modified Lepski method, and is the subject of current research.

## 5 Proofs

Before proceeding to the proofs, we establish two lemmas that are used throughout. The first one establishes a bound on the deviation of true and empirical density averages.

**Lemma 1** *Consider  $0 < \delta < 1$ . With probability at least  $1 - \delta$ , the following is true for all dyadic resolutions  $j$ :*

$$\max_{A \in \mathcal{A}_j} |\bar{f}(A) - \widehat{f}(A)| \leq \Psi_j.$$

**Proof:** The proof relies on a pair of VC inequalities (See [DL01] Chapter 3) that bound the *relative* deviation of true and empirical probabilities. For the collection  $\mathcal{A}_j$  with shatter coefficient bounded by  $2^{jd}$ , the relative VC inequalities state that for any  $\epsilon > 0$

$$P \left( \sup_{A \in \mathcal{A}_j} \frac{P(A) - \widehat{P}(A)}{\sqrt{P(A)}} > \epsilon \right) \leq 4 \cdot 2^{jd} e^{-n\epsilon^2/4}$$

and

$$P \left( \sup_{A \in \mathcal{A}_j} \frac{\widehat{P}(A) - P(A)}{\sqrt{\widehat{P}(A)}} > \epsilon \right) \leq 4 \cdot 2^{jd} e^{-n\epsilon^2/4}.$$

Also observe that

$$\widehat{P}(A) \leq P(A) + \epsilon \sqrt{\widehat{P}(A)} \implies \widehat{P}(A) \leq 2 \max(P(A), 2\epsilon^2)$$

$$P(A) \leq \widehat{P}(A) + \epsilon \sqrt{P(A)} \implies P(A) \leq 2 \max(\widehat{P}(A), 2\epsilon^2).$$

To see the first statement, consider 1)  $\widehat{P}(A) \leq 4\epsilon^2$  - The statement is obvious. 2)  $\widehat{P}(A) > 4\epsilon^2$  - This gives a bound on  $\epsilon$ , which implies  $\widehat{P}(A) \leq P(A) + \widehat{P}(A)/2$  and hence  $\widehat{P}(A) \leq 2P(A)$ . The second statement follows similarly.

Using these statements and the relative VC inequalities for the collection  $\mathcal{A}_j$ , we have: With probability  $> 1 - 8 \cdot 2^{jd} e^{-n\epsilon^2/4}$ ,  $\forall A \in \mathcal{A}_j$  both

$$P(A) - \widehat{P}(A) \leq \epsilon \sqrt{P(A)} \leq \epsilon \sqrt{2 \max(\widehat{P}(A), 2\epsilon^2)}$$

and

$$\widehat{P}(A) - P(A) \leq \epsilon \sqrt{\widehat{P}(A)} \leq \epsilon \sqrt{2 \max(\widehat{P}(A), 2\epsilon^2)}$$

Setting  $\epsilon = \sqrt{4 \log(2^{jd} 8 / \delta_j) / n}$ , we have with probability  $> 1 - \delta_j, \forall A \in \mathcal{A}_j$

$$\begin{aligned} |P(A) - \widehat{P}(A)| \\ \leq \sqrt{8 \frac{\log(2^{jd} 8 / \delta_j)}{n} \max\left(\widehat{P}(A), 8 \frac{\log(2^{jd} 8 / \delta_j)}{n}\right)} \end{aligned}$$

The result follows by dividing the result by  $\mu(A)$ , setting  $\delta_j = \delta 2^{-(j+1)}$  and taking union bound.  $\blacksquare$

The next lemma states how the density deviation bound or penalty  $\Psi_j$  scales with resolution. This will be used to derive rates of convergence.

**Lemma 2** *For all resolutions such that  $2^j = O((n/\log n)^{\frac{1}{\alpha}})$ , there exist constants  $c_3, c_4 \equiv c_4(f_{\max}, d) > 0$  such that for all  $n$ , with probability at least  $1 - 1/n$ ,*

$$c_3 \sqrt{2^{jd} \frac{\log n}{n}} \leq \Psi_j \leq c_4 \sqrt{2^{jd} \frac{\log n}{n}}$$

**Proof:** The lower bound follows by observing that

$$1 = \sum_{A \in \mathcal{A}_j} \widehat{P}(A) \leq \max_{A \in \mathcal{A}_j} \widehat{P}(A) \times |\mathcal{A}_j| = \max_{A \in \mathcal{A}_j} \frac{\widehat{P}(A)}{\mu(A)} = \max_{A \in \mathcal{A}_j} \widehat{f}(A)$$

and using  $\delta = 1/n, j \geq 0$  and  $\mu(A) = 2^{-jd}$ .

To get an upper bound, using the same arguments as in proof of Lemma 1 based on the relative VC inequality it follows that [SSN07] with probability  $> 1 - \delta_j$ , for all  $A \in \mathcal{A}_j$

$$\widehat{P}(A) \leq 2 \max\left(P(A), 8 \frac{\log(2^{jd} 8 / \delta_j)}{n}\right).$$

Dividing by  $\mu(A) = 2^{-jd}$ , using density bound  $f_{\max}$ , setting  $\delta_j = \delta 2^{-(j+1)}$  and taking union bound, we have with probability  $> 1 - \delta$ , for all dyadic resolutions  $j$

$$\max_{A \in \mathcal{A}_j} \widehat{f}(A) \leq 2 \max\left(f_{\max}, 2^{jd} 8 \frac{\log(2^{j(d+1)} 16 / \delta)}{n}\right).$$

This implies the upper bound using  $\delta = 1/n$  and  $2^j = O((n/\log n)^{1/d})$ .  $\blacksquare$

## 5.1 Proof of Theorem 1

The proof relies on the following lemma that will also be used in the proof of Theorem 2.

**Lemma 3** *Consider densities satisfying assumptions [A] and [B]. Then for all resolutions such that  $2^j = O(s_n^{-1}(n/\log n)^{\frac{1}{\alpha}})$ , where  $s_n$  is a monotone diverging sequence, and  $n \geq n_0 \equiv n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$  with probability at least  $1 - 3/n$*

$$d_\infty(\widehat{G}_j, G_\gamma^*) \leq \max(2C_3 + 3, 8\sqrt{d}\epsilon_o^{-1}) \left[ \left( \frac{\Psi_j}{C_1} \right)^{\frac{1}{\alpha}} + \sqrt{d}2^{-j} \right].$$

**Proof:** Let  $J_0 = \lceil \log_2 4\sqrt{d}/\epsilon_o \rceil$ , where  $\epsilon_o$  is as defined in assumption [B]. Also define

$$\epsilon_j := \left[ \left( \frac{\Psi_j}{C_1} \right)^{\frac{1}{\alpha}} + \sqrt{d}2^{-j} \right].$$

Consider two cases:

I.  $j < J_0$ .

For this case, since the domain  $\mathcal{X} = [0, 1]^d$ , we use the trivial bound

$$d_\infty(\widehat{G}_j, G_\gamma^*) \leq \sqrt{d} \leq 2^{J_0} (\sqrt{d}2^{-j}) \leq 8\sqrt{d}\epsilon_o^{-1}\epsilon_j.$$

The last step follows by choice of  $J_0$  and since  $\Psi_j, C_1 > 0$ .

II.  $j \geq J_0$ .

Observe that assumption [B] implies that  $G_\gamma^*$  is not empty since  $G_\gamma^* \supseteq \mathcal{I}_\epsilon(G_\gamma^*) \neq \emptyset$  for  $\epsilon \leq \epsilon_o$ . We will show that for large enough  $n$ , with high probability,  $\widehat{G}_j \cap G_\gamma^* \neq \emptyset$  for  $j \geq J_0$  and hence  $\widehat{G}_j$  is not empty. Thus the Hausdorff error is given as

$$d_\infty(\widehat{G}_j, G_\gamma^*) = \max\left\{ \sup_{x \in \widehat{G}_j} \rho(x, \widehat{G}_j), \sup_{x \in G_\gamma^*} \rho(x, G_\gamma^*) \right\}, \quad (4)$$

and we need bounds on the two terms in the right hand side.

We now prove that  $\widehat{G}_j$  is not empty and obtain bounds on the two terms in the Hausdorff error. Towards this end, we establish two propositions. The first proposition proves that for large enough  $n$ , with high probability, the distance of all points that are erroneously excluded or included in the level set estimate, from the true set boundary is bounded by  $\epsilon_j$ . Notice that, if  $\widehat{G}_j$  is non-empty, this provides an upper bound on the second term of the Hausdorff error (Eq. 4). The second proposition establishes that, for large enough  $n$  and  $j \geq J_0$ ,  $2\epsilon_j \leq \epsilon_o$  and hence the inner cover  $\mathcal{I}_{2\epsilon_j}(G_\gamma^*)$  is not empty. And using the first proposition, with high probability,  $\mathcal{I}_{2\epsilon_j}(G_\gamma^*)$  contains points that are correctly included in the level set estimate and lie in  $\widehat{G}_j \cap G_\gamma^*$ . Thus  $\widehat{G}_j$  is not empty. Further, along with assumption [B], this provides a bound of  $\epsilon_j$  on the distance of any point in  $G_\gamma^*$  from the estimate  $\widehat{G}_j$ , thus bounding the first term of the Hausdorff error (Eq. 4).

We end the proof of the two propositions with a white box  $\square$  to indicate that these propositions are included within the proof of Lemma 3, and do not signify end of the proof of Lemma 3.

**Proposition 2** *If  $\widehat{G}_j \Delta G_\gamma^* \neq \emptyset$ , then for resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{1/d})$  and  $n \geq n_1(f_{\max}, d, \delta_0)$  with probability at least  $1 - 2/n$*

$$\sup_{x \in \widehat{G}_j \Delta G_\gamma^*} \rho(x, \partial G_\gamma^*) \leq \left( \frac{\Psi_j}{C_1} \right)^{1/\alpha} + \sqrt{d}2^{-j} = \epsilon_j.$$

**Proof:** Since by assumption  $\widehat{G}_j \Delta G_\gamma^* \neq \emptyset$ , consider  $x \in \widehat{G}_j \Delta G_\gamma^*$ . Let  $A_x \in \mathcal{A}_j$  denote the cell containing  $x$  at resolution  $j$ . Consider two cases:

(i)  $A_x \cap \partial G_\gamma^* \neq \emptyset$ . This implies that

$$\rho(x, \partial G_\gamma^*) \leq \sqrt{d}2^{-j}.$$

(ii)  $A_x \cap \partial G_\gamma^* = \emptyset$ . Since  $x \in \widehat{G}_j \Delta G_\gamma^*$ , it is erroneously included or excluded from the level set estimate  $\widehat{G}_j$ . Therefore, if  $\bar{f}(A_x) \geq \gamma$ , then  $\widehat{f}(A_x) < \gamma$  otherwise if  $\bar{f}(A_x) < \gamma$ , then  $\widehat{f}(A_x) \geq \gamma$ . This implies that  $|\gamma - \bar{f}(A_x)| \leq |\bar{f}(A_x) - \widehat{f}(A_x)|$ . Using Lemma 1, we get  $|\gamma - \bar{f}(A_x)| \leq \Psi_j$  with probability at least  $1 - \delta$ .

Now let  $x_0$  be any point in  $A_x$  such that  $|\gamma - f(x_0)| \leq |\gamma - \bar{f}(A_x)|$  (Notice that at least one such point must exist in  $A_x$  since this cell does not intersect the boundary). As argued above,  $|\gamma - \bar{f}(A_x)| \leq \Psi_j$  with probability at least  $1 - 1/n$  (for  $\delta = 1/n$ ) and using Lemma 2,  $\Psi_j$  decreases with  $n$  for resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{1/d})$  with probability at least  $1 - 1/n$ . So for large enough  $n \geq n_1(f_{\max}, d, \delta_0)$ ,  $\Psi_j \leq \delta_0$  and hence  $|\gamma - f(x_0)| \leq \delta_0$ . Thus, the density regularity assumption **[A]** holds at  $x_0$  with probability  $> 1 - 2/n$  and we have

$$\begin{aligned} \rho(x_0, \partial G_\gamma^*) &\leq \left( \frac{|\gamma - f(x_0)|}{C_1} \right)^{\frac{1}{\alpha}} \\ &\leq \left( \frac{|\gamma - \bar{f}(A_x)|}{C_1} \right)^{\frac{1}{\alpha}} \leq \left( \frac{\Psi_j}{C_1} \right)^{\frac{1}{\alpha}}. \end{aligned}$$

Since  $x, x_0 \in A_x$ ,  $\rho(x, \partial G_\gamma^*) \leq \rho(x_0, \partial G_\gamma^*) + \sqrt{d}2^{-j}$ . Therefore,

$$\rho(x, \partial G_\gamma^*) \leq \left( \frac{\Psi_j}{C_1} \right)^{1/\alpha} + \sqrt{d}2^{-j}.$$

So for both cases, we can say that for resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{1/d})$  and  $n \geq n_1(f_{\max}, d, \delta_0)$  with probability at least  $1 - 2/n$ ,  $\forall x \in \widehat{G}_j \Delta G_\gamma^*$

$$\rho(x, \partial G_\gamma^*) \leq \left( \frac{\Psi_j}{C_1} \right)^{1/\alpha} + \sqrt{d}2^{-j} = \epsilon_j. \quad \square$$

**Proposition 3** Recall assumption **[B]** and denote the inner cover of  $G_\gamma^*$  with  $2\epsilon_j$ -balls,  $\mathcal{I}_{2\epsilon_j}(G_\gamma^*) \equiv \mathcal{I}_{2\epsilon_j}$ . For resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{1/d})$ ,  $j \geq J_0$  and  $n \geq n_0 \equiv n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$ , with probability at least  $1 - 3/n$ ,

$$\widehat{G}_j \neq \emptyset \quad \text{and} \quad \sup_{x \in \mathcal{I}_{2\epsilon_j}} \rho(x, \widehat{G}_j \cap G_\gamma^*) \leq \epsilon_j.$$

**Proof:** Observe that for  $j \geq J_0$ ,  $2\sqrt{d}2^{-j} \leq 2\sqrt{d}2^{-J_0} \leq \epsilon_o/2$ . And using Lemma 2 for large enough  $n \geq n_2 \equiv n_2(\epsilon_o, f_{\max}, C_1, \alpha)$ ,

$2(\Psi_j/C_1)^{1/\alpha} \leq \epsilon_o/2$  with probability at least  $1 - 1/n$ . Therefore for all  $j \geq J_0$  and  $n \geq n_2$ ,  $2\epsilon_j \leq \epsilon_o$  with probability at least  $1 - 1/n$  and hence  $\mathcal{I}_{2\epsilon_j} \neq \emptyset$ . Now consider any  $2\epsilon_j$ -ball in  $\mathcal{I}_{2\epsilon_j}$ . Then the distance of all points in the interior of the concentric  $\epsilon_j$ -ball from the boundary of  $\mathcal{I}_{2\epsilon_j}$ , and hence from the boundary of  $G_\gamma^*$  is greater than  $\epsilon_j$ . As per Proposition 2 for  $n \geq n_0 = \max(n_1, n_2)$ , with probability  $> 1 - 3/n$ , none of these points can lie in  $\widehat{G}_j \Delta G_\gamma^*$ , and hence must lie in  $\widehat{G}_j \cap G_\gamma^*$  since they are in  $\mathcal{I}_{2\epsilon_j} \subseteq G_\gamma^*$ . Therefore,

$$\widehat{G}_j \neq \emptyset \quad \text{and} \quad \sup_{x \in \mathcal{I}_{2\epsilon_j}} \rho(x, \widehat{G}_j \cap G_\gamma^*) \leq \epsilon_j. \quad \square$$

Now since  $G_\gamma^*$  and  $\widehat{G}_j$  are non-empty sets, we bound the two terms that contribute to the Hausdorff error

$$\sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j) \quad \text{and} \quad \sup_{x \in \widehat{G}_j} \rho(x, G_\gamma^*).$$

For this, we will use Propositions 2 and 3, hence all the following statements will hold for resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{1/d})$ ,  $j \geq J_0$  and  $n \geq n_0 \equiv n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$ , with probability at least  $1 - 3/n$ .

To bound the second term, observe that

- (i) If  $\widehat{G}_j \setminus G_\gamma^* = \emptyset$ , then  $\sup_{x \in \widehat{G}_j} \rho(x, G_\gamma^*) = 0$ .
- (ii) If  $\widehat{G}_j \setminus G_\gamma^* \neq \emptyset$ , it implies that  $\widehat{G}_j \Delta G_\gamma^* \neq \emptyset$ . Hence, using Proposition 2, we have

$$\begin{aligned} \sup_{x \in \widehat{G}_j} \rho(x, G_\gamma^*) &= \sup_{x \in \widehat{G}_j \setminus G_\gamma^*} \rho(x, G_\gamma^*) \\ &= \sup_{x \in \widehat{G}_j \setminus G_\gamma^*} \rho(x, \partial G_\gamma^*) \\ &\leq \sup_{x \in \widehat{G}_j \Delta G_\gamma^*} \rho(x, \partial G_\gamma^*) \leq \epsilon_j. \end{aligned}$$

Thus, for either case

$$\sup_{x \in \widehat{G}_j} \rho(x, G_\gamma^*) \leq \epsilon_j. \quad (5)$$

To bound the first term, observe that

- (i) If  $G_\gamma^* \setminus \widehat{G}_j = \emptyset$ , then  $\sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j) = 0$ .
- (ii) If  $G_\gamma^* \setminus \widehat{G}_j \neq \emptyset$ , we proceed as follows:

$$\begin{aligned} \sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j) &\leq \sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j \cap G_\gamma^*) \\ &= \max\{ \sup_{x \in \mathcal{I}_{2\epsilon_j}} \rho(x, \widehat{G}_j \cap G_\gamma^*), \\ &\quad \sup_{x \in G_\gamma^* \setminus \mathcal{I}_{2\epsilon_j}} \rho(x, \widehat{G}_j \cap G_\gamma^*) \} \\ &\leq \max\{ \epsilon_j, \sup_{x \in G_\gamma^* \setminus \mathcal{I}_{2\epsilon_j}} \rho(x, \widehat{G}_j \cap G_\gamma^*) \}. \end{aligned}$$

The last step follows using Proposition 3.

Now consider any  $x \in G_\gamma^* \setminus \mathcal{I}_{2\epsilon_j}$ . Then using triangle inequality,  $\forall y \in \partial G_\gamma^*$  and  $\forall z \in \mathcal{I}_{2\epsilon_j}$ ,

$$\begin{aligned} \rho(x, \widehat{G}_j \cap G_\gamma^*) &\leq \rho(x, y) + \rho(y, z) + \rho(z, \widehat{G}_j \cap G_\gamma^*) \\ &\leq \rho(x, y) + \rho(y, z) + \\ &\quad \sup_{z \in \mathcal{I}_{2\epsilon_j}} \rho(z, \widehat{G}_j \cap G_\gamma^*) \\ &\leq \rho(x, y) + \rho(y, z) + \epsilon_j. \end{aligned}$$

The last step follows using Proposition 3. This implies that  $\forall y \in \partial G_\gamma^*$ ,

$$\begin{aligned} \rho(x, \widehat{G}_j \cap G_\gamma^*) &\leq \rho(x, y) + \inf_{z \in \mathcal{I}_{2\epsilon_j}} \rho(y, z) + \epsilon_j \\ &= \rho(x, y) + \rho(y, \mathcal{I}_{2\epsilon_j}) + \epsilon_j \\ &\leq \rho(x, y) + \sup_{y \in \partial G_\gamma^*} \rho(y, \mathcal{I}_{2\epsilon_j}) + \epsilon_j \\ &\leq \rho(x, y) + 2C_3\epsilon_j + \epsilon_j. \end{aligned}$$

Here the last step invokes assumption **[B]**. This in turn implies that

$$\begin{aligned} \rho(x, \widehat{G}_j \cap G_\gamma^*) &\leq \inf_{y \in \partial G_\gamma^*} \rho(x, y) + (2C_3 + 1)\epsilon_j \\ &\leq 2\epsilon_j + (2C_3 + 1)\epsilon_j. \end{aligned}$$

The second step is true for  $x \in G_\gamma^* \setminus \mathcal{I}_{2\epsilon_j}$ . If it was not true, then  $\forall y \in \partial G_\gamma^*$ ,  $\rho(x, y) > 2\epsilon_j$  and hence there exists a closed  $2\epsilon_j$ -ball around  $x$  that is in  $G_\gamma^*$ . This contradicts the fact that  $x \notin \mathcal{I}_{2\epsilon_j}$ . Therefore, we have:

$$\sup_{x \in G_\gamma^* \setminus \mathcal{I}_{2\epsilon_j}} \rho(x, \widehat{G}_j \cap G_\gamma^*) \leq (2C_3 + 3)\epsilon_j.$$

And going back to the start of case (ii) we get:

$$\sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j) \leq (2C_3 + 3)\epsilon_j.$$

Therefore, for either case we have

$$\sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j) \leq (2C_3 + 3)\epsilon_j. \quad (6)$$

From Eq. (5) and (6), we have that for all densities satisfying assumptions **[A, B]**, for resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{1/d})$ ,  $j \geq J_0$  and  $n \geq n_0 \equiv n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$ , with probability  $> 1 - 3/n$ ,

$$\begin{aligned} d_\infty(\widehat{G}_j, G_\gamma^*) &= \max\left\{ \sup_{x \in G_\gamma^*} \rho(x, \widehat{G}_j), \sup_{x \in \widehat{G}_j} \rho(x, G_\gamma^*) \right\} \\ &\leq (2C_3 + 3)\epsilon_j. \end{aligned}$$

And addressing both Case I ( $j < J_0$ ) and Case II ( $j \geq J_0$ ), we finally have that for all densities satisfying assumptions **[A, B]**, for resolutions satisfying  $2^j = O(s_n^{-1}(n/\log n)^{\frac{1}{d}})$  and  $n \geq n_0 \equiv n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$ , with probability  $> 1 - 3/n$ ,

$$d_\infty(\widehat{G}_j, G_\gamma^*) \leq \max(2C_3 + 3, 8\sqrt{d}\epsilon_o^{-1})\epsilon_j. \quad \blacksquare$$

Since the chosen resolution  $2^{-j} \asymp s_n(n/\log n)^{-\frac{1}{(d+2\alpha)}}$  satisfies conditions of Lemma 3, proof of Theorem 1 now follows using the bound on  $\Psi_j$  from Lemma 2. Let  $\Omega$  denote the event such that the bounds of Lemma 2 and Lemma 3 hold. Then for  $n \geq n_0$ ,  $P(\bar{\Omega}) \leq 4/n$ . Hence for all  $n$ ,  $P(\bar{\Omega}) \leq \max(4, n_0)/n$ . So  $\forall f \in \mathcal{F}_1^*(\alpha)$ : (Here  $C$  may denote a different constant from line to line. Explanation for each step is provided after the equations.)

$$\begin{aligned} \mathbb{E}[d_\infty(\widehat{G}_j, G_\gamma^*)] &= P(\Omega)\mathbb{E}[d_\infty(\widehat{G}_j, G_\gamma^*)|\Omega] + P(\bar{\Omega})\mathbb{E}[d_\infty(\widehat{G}_j, G_\gamma^*)|\bar{\Omega}] \\ &\leq \mathbb{E}[d_\infty(\widehat{G}_j, G_\gamma^*)|\Omega] + P(\bar{\Omega})\sqrt{d} \\ &\leq C \left[ \left( \frac{\Psi_j}{C_1} \right)^{1/\alpha} + \sqrt{d}2^{-j} + \frac{\sqrt{d}}{n} \right] \\ &\leq C \max \left\{ \left( 2^{jd} \frac{\log n}{n} \right)^{\frac{1}{2\alpha}}, 2^{-j}, \frac{1}{n} \right\} \\ &\leq C(C_1, C_3, \epsilon_o, f_{\max}, \delta_0, d, \alpha) s_n \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}. \end{aligned}$$

The second step follows using the trivial bounds  $P(\Omega) \leq 1$  and since the domain  $\mathcal{X} = [0, 1]^d$ ,  $\mathbb{E}[d_\infty(\widehat{G}_j, G_\gamma^*)|\bar{\Omega}] \leq \sqrt{d}$ . The third step follows from Lemma 3 and the fourth one using Lemma 2. The last step follows since the chosen resolution  $2^{-j} \asymp s_n(n/\log n)^{-\frac{1}{(d+2\alpha)}}$ .

## 5.2 Proof of Theorem 2

To analyze the resolution chosen by the complexity penalized procedure of Eq. (3) based on the vernier, we first establish two results regarding the vernier. Using Lemma 1, we have the following corollary that bounds the deviation of true and empirical vernier.

**Corollary 1** Consider  $0 < \delta < 1$ . With probability at least  $1 - \delta$ , the following is true for all dyadic resolutions  $j$ :

$$|\mathcal{V}_{\gamma, j} - \widehat{\mathcal{V}}_{\gamma, j}| \leq \Psi_{j'}.$$

**Proof:** Let  $A_0 \in \mathcal{A}_j$  denote the cell achieving the min defining  $\mathcal{V}_{\gamma, j}$  and  $A_1 \in \mathcal{A}_j$  denote the cell achieving the min defining  $\widehat{\mathcal{V}}_{\gamma, j}$ . Also let  $A'_0$  and  $A'_1$  denote the subcells at resolution  $j'$  within  $A_0$  and  $A_1$ , respectively, that have maximum average density deviation from  $\gamma$ . Similarly, let  $\widehat{A}'_0$  and  $\widehat{A}'_1$  denote the subcells at resolution  $j'$  within  $A_0$  and  $A_1$ , respectively, that have maximum empirical density deviation from  $\gamma$ . Then we have: (Explanation for the steps are given after the equations.)

$$\begin{aligned} \mathcal{V}_{\gamma, j} - \widehat{\mathcal{V}}_{\gamma, j} &= |\gamma - \bar{f}(A'_0)| - |\gamma - \widehat{f}(\widehat{A}'_1)| \\ &\leq |\gamma - \bar{f}(A'_1)| - |\gamma - \widehat{f}(\widehat{A}'_1)| \\ &\leq |\bar{f}(A'_1) - \widehat{f}(\widehat{A}'_1)| \\ &= \max\{\bar{f}(A'_1) - \widehat{f}(\widehat{A}'_1), \widehat{f}(\widehat{A}'_1) - \bar{f}(A'_1)\} \\ &\leq \max\{\bar{f}(A'_1) - \widehat{f}(\widehat{A}'_1), \widehat{f}(\widehat{A}'_1) - \bar{f}(\widehat{A}'_1)\} \\ &\leq \max_{A \in \mathcal{A}_{j'}} |\bar{f}(A) - \widehat{f}(A)| \\ &\leq \Psi_{j'} \end{aligned}$$

The first inequality invokes definition of  $A_0$ , the third inequality invokes definitions of the subcells  $A'_1$ ,  $\widehat{A}'_1$ , and the last one follows from Lemma 1. Similarly,

$$\begin{aligned}\widehat{\mathcal{V}}_{\gamma,j} - \mathcal{V}_{\gamma,j} &= |\gamma - \widehat{f}(\widehat{A}'_1)| - |\gamma - \bar{f}(A'_0)| \\ &\leq |\gamma - \widehat{f}(\widehat{A}'_0)| - |\gamma - \bar{f}(A'_0)| \\ &\leq |\bar{f}(A'_0) - \widehat{f}(\widehat{A}'_0)|\end{aligned}$$

Here the first inequality invokes definition of  $A_1$ . The rest follows as above, considering cell  $A_0$  instead of  $A_1$ . ■

The second result establishes that the vernier is sensitive to the resolution and density regularity.

**Lemma 4** *Consider densities satisfying assumptions [A] and [C]. Recall that  $j' = \lfloor j + \log_2 s_n \rfloor$ , where  $s_n$  is a monotone diverging sequence. Then for all dyadic resolutions  $j$*

$$\min(\delta_0, C_1)2^{-j'\alpha} \leq \mathcal{V}_{\gamma,j} \leq C(\sqrt{d}2^{-j})^\alpha$$

holds for  $n$  large enough such that  $s_n > 4C_46^d$ . Here  $C \equiv C(C_2, f_{\max}, \delta_1, \alpha) > 0$ .

**Proof:** We first establish the upper bound. Recall assumption [A] and consider the cell  $A \in \mathcal{A}_j$  that contains the point  $y_0$ . Then  $A \cap \partial G_\gamma^* \neq \emptyset$ . Let  $A'$  denote the subcell at resolution  $j'$  within  $A$  that has maximum average density deviation from  $\gamma$ . Consider two cases:

- (i) If the resolution is large enough so that  $\sqrt{d}2^{-j} \leq \delta_1$ , then the density regularity assumption [A] holds  $\forall x \in A$  since  $A \subset B(y_0, \delta_1)$ , the  $\delta_1$ -ball around  $y_0$ . The same holds also for the subcell  $A'$ . Hence

$$|\gamma - \bar{f}(A')| \leq C_2(\sqrt{d}2^{-j})^\alpha$$

- (ii) If the resolution is not large enough and  $\sqrt{d}2^{-j} > \delta_1$ , the following trivial bound holds:

$$|\gamma - \bar{f}(A')| \leq f_{\max} \leq \frac{f_{\max}}{\delta_1^\alpha}(\sqrt{d}2^{-j})^\alpha$$

The last step holds since  $\sqrt{d}2^{-j} > \delta_1$ .

Hence we can say for all  $j$  there exists  $A \in \mathcal{A}_j$  such that

$$|\gamma - \bar{f}(A')| \leq \max\left(C_2, \frac{f_{\max}}{\delta_1^\alpha}\right)(\sqrt{d}2^{-j})^\alpha$$

This yields the upper bound on the vernier:

$$\mathcal{V}_{\gamma,j} \leq \max\left(C_2, \frac{f_{\max}}{\delta_1^\alpha}\right)(\sqrt{d}2^{-j})^\alpha := C(\sqrt{d}2^{-j})^\alpha$$

where  $C \equiv C(C_2, f_{\max}, \delta_1, \alpha)$ .

For the lower bound, consider a cell  $A \in \mathcal{A}_j$ . We will show that assumption [C] on the level set boundary dimension basically implies that the boundary does not intersect all subcells at resolution  $j'$  within the cell  $A$  at resolution  $j$ . And in fact for large enough  $n$  (so that  $2^{-j'}$  is small enough, recall that  $j' = \lfloor j + \log_2 s_n \rfloor$  where  $s_n$  is a monotone diverging sequence), there exists at least one subcell  $A'_o \in A \cap \mathcal{A}_{j'}$  such that  $\forall x \in A'_o$ ,

$$\rho(x, \partial G_\gamma^*) \geq 2^{-j'}.$$

We establish this statement formally later on, but for now assume that it holds. The local density regularity condition [A] now gives that for all  $x \in A'_o$ ,  $|\gamma - f(x)| \geq \min(\delta_0, C_1)2^{-j'\alpha}$ . So we have:

$$\max_{A' \in A \cap \mathcal{A}_{j'}} |\gamma - \bar{f}(A')| \geq |\gamma - \bar{f}(A'_o)| \geq \min(\delta_0, C_1)2^{-j'\alpha}$$

Since this is true for any  $A \in \mathcal{A}_j$ , in particular, this is true for the cell achieving the min defining  $\mathcal{V}_{\gamma,j}$ . Hence, the lower bound on the vernier  $\mathcal{V}_{\gamma,j}$  follows.

We now formally prove that assumption [C] on the level set boundary dimension implies that for large enough  $n$  (so that  $s_n > 4C_46^d$ ),  $\exists A'_o \in A \cap \mathcal{A}_{j'}$  s.t.  $\forall x \in A'_o$ ,

$$\rho(x, \partial G_\gamma^*) \geq 2^{-j'}.$$

Observe that it suffices to show that for large enough  $n$ ,  $\exists A'' \in A \cap \mathcal{A}_{j'-2}$  s.t.  $A'' \cap \partial G_\gamma^* = \emptyset$ . To prove this last statement, consider two cases:

- (i)  $A \cap \partial G_\gamma^* = \emptyset$ . For  $s_n \geq 8$ ,  $j' - 2 \geq j$  (recall definition of  $j'$ ), and since  $A$  does not intersect the boundary, clearly  $\exists A'' \in A \cap \mathcal{A}_{j'-2}$  s.t.  $A'' \cap \partial G_\gamma^* = \emptyset$ .

- (ii)  $A \cap \partial G_\gamma^* \neq \emptyset$ . Let  $x \in A \cap \partial G_\gamma^*$ . Consider  $\epsilon = \sqrt{d}2^{-j}$  (the diagonal length of a cell), then  $A \subseteq B(x, \epsilon)$ . Also let  $\delta = \sqrt{d}2^{-(j'-2)}/2$  (the choice will be justified below). For  $s_n \geq 4$ ,  $0 < \delta \leq \epsilon$  and using assumption [C], the minimum number of  $\delta$ -balls required to cover  $\partial G_\gamma^* \cap B(x, \epsilon)$  is  $\leq C_4(\delta/\epsilon)^{-(d-1)}$ . Since  $A \subseteq B(x, \epsilon)$ , the minimum number of  $\delta$ -balls required to cover  $\partial G_\gamma^* \cap A$  is also  $\leq C_4(\delta/\epsilon)^{-(d-1)}$ . Now consider a uniform partition of the cell  $A$  into subcells of sidelength  $2\delta/\sqrt{d} = 2^{-(j'-2)}$ . Since the diagonal length of a subcell  $\sqrt{d}2^{-(j'-2)} = 2\delta$ , this choice of  $\delta$  implies that a subcell at resolution  $2^{-(j'-2)}$  is inscribed within an aligned  $\delta$ -ball. Observe that at this resolution, in  $d$ -dim, an unaligned  $\delta$ -ball can intersect up to  $3^d - 1$  subcells (number of neighbors of any hypercube). Therefore, the number of subcells in  $A \cap \mathcal{A}_{j'-2}$  that intersect the boundary can be no more than

$$\begin{aligned}3^d C_4 (\delta/\epsilon)^{-(d-1)} &= 3^d C_4 \left( \frac{\sqrt{d}2^{-(j'-2)}}{2\sqrt{d}2^{-j}} \right)^{-(d-1)} \\ &= \frac{C_4 6^d}{2} 2^{(j'-2-j)d} 2^{-(j'-2-j)} \\ &< \frac{4C_4 6^d}{s_n} 2^{(j'-2-j)d}\end{aligned}$$

where the last step uses the fact  $2^{-j'} < 2^{-j+1}/s_n$ . For  $s_n > 4C_4 6^d$ , the number of subcells within  $A$  at resolution  $j' - 2$  that intersect the boundary is less than the total number of subcells within  $A$  at that resolution. Therefore,  $\exists A'' \in A \cap \mathcal{A}_{j'-2}$  s.t.  $A'' \cap \partial G_\gamma^* = \emptyset$ .

This in turn implies that for  $n$  large enough (so that  $s_n > 4C_4 6^d$ ),  $\exists A'_o \in A \cap \mathcal{A}_{j'}$  such that  $\forall x \in A'_o$ ,  $\rho(x, \partial G_\gamma^*) \geq 2^{-j'}$ . ■

We are now ready to prove Theorem 2. Observe that Lemmas 2, 3 and Corollary 1 hold together with probability at least  $1 - 5/n$  (taking  $\delta = 1/n$ ). Using these lemmas, we will show that for the resolution  $\hat{j}$  chosen by Eq. (3), both  $\mathcal{V}_{\gamma, \hat{j}}$  and  $\Psi_{\hat{j}}$  are upper bounded by  $C s_n^{\frac{d\alpha}{d+2\alpha}} (n/\log n)^{-\frac{\alpha}{d+2\alpha}}$ , where  $C \equiv C(C_2, f_{\max}, \delta_1, d, \alpha) > 0$ . If this holds, then using Lemma 4 and the definition of  $j'$ , we have the following upper bound on the sidelength: For  $s_n > 4C_4 6^d$

$$\begin{aligned} 2^{-\hat{j}} \leq s_n 2^{-\hat{j}'} &\leq s_n \left( \frac{\mathcal{V}_{\gamma, \hat{j}}}{\min(\delta_0, C_1)} \right)^{\frac{1}{\alpha}} \\ &\leq c_2 s_n s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}, \end{aligned}$$

where  $c_2 \equiv c_2(C_1, C_2, f_{\max}, \delta_0, \delta_1, d, \alpha) > 0$ . Also notice that since  $2^J \asymp s_n^{-1} (n/\log n)^{1/d}$ , we have  $2^{j'} \leq 2^{J'}$   $\leq s_n 2^J \asymp (n/\log n)^{1/d}$ , and hence Lemma 2 can be used to provide a lower bound on the sidelength:

$$\begin{aligned} 2^{-\hat{j}} > \frac{s_n}{2} 2^{-\hat{j}'} &\geq \frac{s_n}{2} \left( \frac{\Psi_{\hat{j}'}^2}{c_3^2 \log n} n \right)^{-\frac{1}{d}} \\ &\geq c_1 s_n \left( s_n^{\frac{2d\alpha}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{2\alpha}{d+2\alpha}} \frac{n}{\log n} \right)^{-\frac{1}{d}} \\ &= c_1 s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{\frac{-1}{d+2\alpha}}, \end{aligned}$$

where  $c_1 \equiv c_1(C_2, f_{\max}, \delta_1, d, \alpha) > 0$ . So we have for  $s_n > 4C_4 6^d$ , with probability at least  $1 - 5/n$ ,

$$c_1 s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}} \leq 2^{-\hat{j}} \leq c_2 s_n s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}. \quad (7)$$

Hence the automatically chosen resolution behaves as desired.

Let us now derive the claimed bounds on  $\mathcal{V}_{\gamma, \hat{j}}$  and  $\Psi_{\hat{j}}$ . Using Corollary 1 and Eq. (3), we have the following oracle inequality:

$$\begin{aligned} \mathcal{V}_{\gamma, \hat{j}} &\leq \hat{\mathcal{V}}_{\gamma, \hat{j}} + \Psi_{\hat{j}} \\ &= \min_{0 \leq j \leq J} \left\{ \hat{\mathcal{V}}_{\gamma, j} + \Psi_{j'} \right\} \leq \min_{0 \leq j \leq J} \left\{ \mathcal{V}_{\gamma, j} + 2\Psi_{j'} \right\} \end{aligned}$$

Lemma 4 provides an upper bound on the vernier  $\mathcal{V}_{\gamma, j}$ , and Lemma 2 provides an upper bound on the penalty  $\Psi_{j'}$ . We now plug these bounds into the oracle inequality. Here  $C$  may denote a different constant from line to line.

$$\begin{aligned} \mathcal{V}_{\gamma, \hat{j}} &\leq \hat{\mathcal{V}}_{\gamma, \hat{j}} + \Psi_{\hat{j}} \leq C \min_{0 \leq j \leq J} \left\{ 2^{-j\alpha} + \sqrt{2^{j'd} \frac{\log n}{n}} \right\} \\ &\leq C \min_{0 \leq j \leq J} \left\{ \max \left( 2^{-j\alpha}, \sqrt{2^{j'd} s_n^d \frac{\log n}{n}} \right) \right\} \\ &\leq C s_n^{\frac{d\alpha}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{\alpha}{d+2\alpha}}. \end{aligned}$$

Here  $C \equiv C(C_2, f_{\max}, \delta_1, d, \alpha)$ . The second step uses the definition of  $j'$  and the last step follows by balancing the two

terms for optimal resolution  $2^{-j^*} \asymp s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}$ . This establishes the desired bounds on  $\mathcal{V}_{\gamma, \hat{j}}$  and  $\Psi_{\hat{j}}$ .

Now we can invoke Lemma 3 to derive the rate of convergence for the Hausdorff error. Consider large enough  $n \geq n_1(C_4, d)$  so that  $s_n > 4C_4 6^d$ . Also, recall that the condition of Lemma 3 requires that  $n \geq n_0(f_{\max}, d, \delta_0, \epsilon_0, C_1, \alpha)$ . Pick  $n \geq \max(n_0, n_1)$  and let  $\Omega$  denote the event such that the bounds of Lemma 2, Lemma 3 and Corollary 1 hold with  $\delta = 1/n$ . Then, we have  $P(\bar{\Omega}) \leq 5/n$  for  $n \geq \max(n_0, n_1)$ , or for all  $n$ ,  $P(\bar{\Omega}) \leq \max(5, n_0, n_1)/n$ . So  $\forall f \in \mathcal{F}_2^*(\alpha)$ , we have: (Here  $C$  may denote a different constant from line to line. Explanation for each step is provided after the equations.)

$$\begin{aligned} \mathbb{E}[d_\infty(\hat{G}, G_\gamma^*)] &= P(\Omega) \mathbb{E}[d_\infty(\hat{G}, G_\gamma^*) | \Omega] + P(\bar{\Omega}) \mathbb{E}[d_\infty(\hat{G}, G_\gamma^*) | \bar{\Omega}] \\ &\leq \mathbb{E}[d_\infty(\hat{G}, G_\gamma^*) | \Omega] + P(\bar{\Omega}) \sqrt{d} \\ &\leq C \left[ \left( \frac{\Psi_{\hat{j}}}{C_1} \right)^{1/\alpha} + \sqrt{d} 2^{-\hat{j}} + \frac{\sqrt{d}}{n} \right] \\ &\leq C \max \left\{ \left( 2^{\hat{j}d} \frac{\log n}{n} \right)^{\frac{1}{2\alpha}}, 2^{-\hat{j}}, \frac{1}{n} \right\} \\ &\leq C s_n s_n^{\frac{d}{d+2\alpha}} \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}} \leq C s_n^2 \left( \frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}. \end{aligned}$$

Here  $C \equiv C(C_1, C_2, C_3, C_4, \epsilon_0, f_{\max}, \delta_0, \delta_1, d, \alpha)$ . The second step follows by observing the trivial bounds  $P(\Omega) \leq 1$  and since the domain  $\mathcal{X} = [0, 1]^d$ ,  $\mathbb{E}[d_\infty(\hat{G}, G_\gamma^*) | \bar{\Omega}] \leq \sqrt{d}$ . The third step follows from Lemma 3 and the fourth one from Lemma 2. The last step follows using the upper and lower bounds established on  $2^{-\hat{j}}$  in Eq. (7).

### 5.3 Star-shaped sets satisfy assumptions [B] and [C]

Recall the definition of  $\mathcal{F}_{SL}$  as defined in [Tsy97]. The class corresponds to densities bounded above by  $f_{\max}$ , satisfying a slightly modified form of the local density regularity assumption [A]:

[A'] *Local density regularity:* The density is  $\alpha$ -regular around the  $\gamma$ -level set,  $0 < \alpha < \infty$  and  $\gamma < f_{\max}$ , if there exist constants  $C_2 > C_1 > 0$  and  $\delta_0 > 0$  such that

$$C_1 \rho(x, \partial G_\gamma^*)^\alpha \leq |f(x) - \gamma| \leq C_2 \rho(x, \partial G_\gamma^*)^\alpha$$

for all  $x \in \mathcal{X}$  with  $|f(x) - \gamma| \leq \delta_0$ , where  $\partial G_\gamma^*$  is the boundary of the true level set  $G_\gamma^*$ , and the set  $\{x : |f(x) - \gamma| \leq \delta_0\}$  is non-empty.

and the densities have  $\gamma$  level sets of the form

$$G_\gamma^* = \{(r, \phi); \phi \in [0, \pi)^{d-2} \times [0, 2\pi), 0 \leq r \leq g(\phi) \leq R\},$$

where  $(r, \phi)$  denote the polar/hyperspherical coordinates and  $R > 0$  is a constant.  $g$  is a periodic Lipschitz function that satisfies  $g(\phi) \geq h$ , where  $h > 0$  is a constant, and

$$|g(\phi) - g(\theta)| \leq L \|\phi - \theta\|_1, \quad \forall \phi, \theta \in [0, \pi)^{d-2} \times [0, 2\pi).$$

Here  $L > 0$  is the Lipschitz constant, and  $\|\cdot\|_1$  denotes the  $\ell_1$  norm.

We set  $R = 1/2$  in the definition of the star-shaped set so that the domain is a subset of  $[-1/2, 1/2]^d$ . With this domain, the following lemma shows that the level set  $G_\gamma^*$  of a density  $f \in \mathcal{F}_{SL}$  satisfies **[B]** and **[C]**.

**Lemma 5** *Consider the  $\gamma$  level set  $G_\gamma^*$  of a density  $f \in \mathcal{F}_{SL}$ . Then  $G_\gamma^*$  satisfies the assumptions **[B]** and **[C]** on the level set regularity and the level set boundary dimension, respectively.*

**Proof:** We first present a sketch of the main ideas, and then provide a detailed proof. Consider the  $\gamma$ -level set  $G_\gamma^*$  of a density  $f \in \mathcal{F}_{SL}$ . To see that it satisfies **[B]**, divide the star-shaped set  $G_\gamma^*$  into sectors of width  $\asymp \epsilon$  so that each sector contains at least one  $\epsilon$ -ball and the inner cover  $\mathcal{I}_\epsilon(G_\gamma^*)$  touches the boundary at some point(s) in each sector. Now one can argue that, in each sector, all other points on the boundary are  $O(\epsilon)$  from the inner cover since the boundary is Lipschitz. Since this is true for each sector, we have  $\forall x \in \partial G_\gamma^*, \rho(x, \mathcal{I}_\epsilon(G_\gamma^*)) = O(\epsilon)$ . To see that  $G_\gamma^*$  satisfies **[C]**, consider any sector of width  $\asymp \epsilon$  and divide it into sub-sectors of width  $O(\delta)$ ,  $0 < \delta \leq \epsilon$ . Since the boundary is Lipschitz, a constant number of  $\delta$ -balls can cover the boundary in each sub-sector. Thus, the minimum number of  $\delta$ -balls needed to cover the boundary in all sub-sectors is of the order of the minimum number of sub-sectors, that is,  $O((\epsilon/\delta)^{d-1})$ . Hence, the result follows. We now present the proof in detail.

To see that  $G_\gamma^*$  satisfies **[B]**, fix  $\epsilon_o \leq h/3$ . Then for all  $\epsilon \leq \epsilon_o$ ,  $B(0, \epsilon) \subseteq G_\gamma^*$  (since  $g(\phi) \geq h > \epsilon_o$ ), and hence  $\mathcal{I}_\epsilon(G_\gamma^*) \neq \emptyset$ . We also need to show that  $\exists C_3 > 0$  such that for all  $x \in \partial G_\gamma^*, \rho(x, \mathcal{I}_\epsilon(G_\gamma^*)) \leq C_3\epsilon$ . For this, divide  $G_\gamma^*$  into  $M^{d-1}$  sectors indexed by  $\mathbf{m} = (m_1, m_2, \dots, m_{d-1}) \in \{1, \dots, M\}^{d-1}$

$$S_{\mathbf{m}} = \left\{ (r, \phi) : 0 \leq r \leq g(\phi), \right.$$

$$\left. \begin{aligned} \frac{\pi(m_i - 1)}{M} \leq \phi_i < \frac{\pi m_i}{M}, i = 1, \dots, d-2, \\ \frac{2\pi(m_{d-1} - 1)}{M} \leq \phi_{d-1} < \frac{2\pi m_{d-1}}{M} \end{aligned} \right\},$$

where  $\phi = (\phi_1, \phi_2, \dots, \phi_{d-1})$ . Let

$$M = \left\lceil \frac{\pi}{2 \sin^{-1} \frac{\epsilon}{h - \epsilon_o}} \right\rceil$$

This choice of  $M$  implies that:

- (i) There exists an  $\epsilon$ -ball within  $S_{\mathbf{m}} \cap B(0, h)$  for every  $\mathbf{m} \in \{1, \dots, M\}^{d-1}$ , and hence within each sector  $S_{\mathbf{m}}$ . This follows because the minimum angular width of a sector with radius  $h$  required to fit an  $\epsilon$ -ball within is

$$2 \sin^{-1} \frac{\epsilon}{h - \epsilon} \leq 2 \sin^{-1} \frac{\epsilon}{h - \epsilon_o} \leq \frac{\pi}{M}.$$

- (ii) The angular-width of the sectors scales as  $O(\epsilon)$ .

$$\begin{aligned} \frac{\pi}{M} &< \frac{\pi}{2 \sin^{-1} \frac{\pi}{h - \epsilon_o} - 1} = \frac{1}{2 \sin^{-1} \frac{1}{h - \epsilon_o} - \frac{1}{\pi}} \\ &\leq 3 \sin^{-1} \frac{\epsilon}{h - \epsilon_o} \leq 6 \frac{\epsilon}{h - \epsilon_o} \leq \frac{9}{h} \epsilon \end{aligned}$$

The second inequality follows as

$$\frac{1}{\pi} \leq \frac{1}{6 \sin^{-1} \frac{\epsilon}{h - \epsilon_o}}$$

since  $\frac{\epsilon}{h - \epsilon_o} \leq \frac{\epsilon_o}{h - \epsilon_o} \leq \frac{1}{2}$  by choice of  $\epsilon_o \leq h/3$ . The third inequality is true since  $\sin^{-1}(z/2) \leq z$  for  $0 \leq z \leq \pi/2$ . The last step follows by choice of  $\epsilon_o \leq h/3$ .

Now from (i) above, each sector contains at least one  $\epsilon$ -ball. Consider any  $\mathbf{m} \in \{1, \dots, M\}^{d-1}$ . We claim that there exists a point  $x_{\mathbf{m}} \in \partial G_\gamma^* \cap S_{\mathbf{m}}$ ,  $x_{\mathbf{m}} = (g(\theta), \theta)$  for some  $\theta \in [0, \pi)^{d-2} \times [0, 2\pi)$ , such that  $\rho(x_{\mathbf{m}}, \mathcal{I}_\epsilon(G_\gamma^*)) = 0$ . Suppose not. Then one can slide the  $\epsilon$ -ball within the sector towards the periphery and never touch the boundary, implying that the set  $G_\gamma^*$  is unbounded. This is a contradiction by the definition of the class  $\mathcal{F}_{SL}$ . So now we have,  $\forall y \in \partial G_\gamma^* \cap S_{\mathbf{m}}, y = (g(\phi), \phi)$

$$\begin{aligned} \rho(y, \mathcal{I}_\epsilon(G_\gamma^*)) &\leq \rho(y, x_{\mathbf{m}}) = \|y - x_{\mathbf{m}}\| \\ &= \|(g(\phi), \phi) - (g(\theta), \theta)\| \\ &\leq |g(\phi) - g(\theta)| + 2\sqrt{g(\phi)g(\theta)} \cdot \\ &\quad \sum_{i=1}^{d-1} \left| \sin \frac{\phi_i - \theta_i}{2} \right| \\ &\leq L \|\phi - \theta\|_1 + \sum_{i=1}^{d-1} \frac{|\phi_i - \theta_i|}{2} \\ &= (L + 1/2) \sum_{i=1}^{d-1} |\phi_i - \theta_i| \\ &\leq (L + 1/2) d \frac{\pi}{M} \\ &\leq \frac{9d(L + 1/2)}{h} \epsilon := C_3\epsilon \end{aligned}$$

The third step follows using simple algebra (see [SSN07]), the fourth step follows by the Lipschitz condition on  $g(\cdot)$ ,  $g(\cdot) \leq R = 1/2$  and since  $|\sin(z)| \leq |z|$ . The sixth step follows since  $x, y \in S_{\mathbf{m}}$  and hence  $|\phi_i - \theta_i| \leq \pi/M$  for  $i = 1, \dots, d-2$  and  $|\phi_{d-1} - \theta_{d-1}| \leq 2\pi/M$ . The last step invokes (ii) above. Therefore, we have for all  $y \in \partial G_\gamma^* \cap S_{\mathbf{m}}, \rho(y, \mathcal{I}_\epsilon(G_\gamma^*)) \leq C_3\epsilon$ . And since the result is true for any sector, condition **[B]** is satisfied by any level set  $G_\gamma^*$  with density  $f \in \mathcal{F}_{SL}$ .

To see that  $G_\gamma^*$  satisfies **[C]**, consider  $x \in \partial G_\gamma^*$ . Let  $x = (g(\phi_0), \phi_0)$ . Also let  $\phi_i^{(1)} = \min\{\phi_i : (g(\phi), \phi) \in B(x, \epsilon)\}$  and  $\phi_i^{(2)} = \max\{\phi_i : (g(\phi), \phi) \in B(x, \epsilon)\}$ . Define the sector

$$S_\epsilon^x = \left\{ (r, \phi) : 0 \leq r \leq g(\phi), \right. \\ \left. \phi_i^{(1)} \leq \phi_i \leq \phi_i^{(2)}, \forall i = 1, \dots, d-1 \right\}$$

Observe that if  $\epsilon \leq \pi h/4 < h$ , the width of  $S_\epsilon^x$  in the  $i^{\text{th}}$  coordinate,  $\Delta\phi_i = \phi_i^{(2)} - \phi_i^{(1)} \leq 2 \sin^{-1} \frac{\epsilon}{g(\phi_0)}$  by construction. Since  $g(\cdot) \geq h$ , we have  $\Delta\phi_i \leq 2 \sin^{-1} \frac{\epsilon}{h} \leq 4\epsilon/h$ , where the last step follows since for  $0 \leq z \leq \pi/2$ ,  $\sin^{-1}(z/2) \leq z$ . If  $\epsilon > \pi h/4$ , then use the trivial bound

$\Delta\phi_i \leq 2\pi \leq 8\epsilon/h$ . Equivalently, we can say for all  $\epsilon$  and all  $i$ ,

$$\Delta\phi_i \leq 8\epsilon/h. \quad (8)$$

Further subdivide  $S_\epsilon^x$  into  $M^{d-1}$  sub-sectors indexed by  $\mathbf{m} = (m_1, \dots, m_{d-1})$

$$S_{\mathbf{m}} = \left\{ (r, \phi) : 0 \leq r \leq g(\phi), \phi_i^{(1)} + \frac{(m_i - 1)\Delta\phi_i}{M} \leq \phi_i < \phi_i^{(1)} + \frac{m_i\Delta\phi_i}{M}, \forall i = 1, \dots, d-1 \right\}$$

Pick  $M$  such that for all coordinates, the sub-sector width  $\frac{\Delta\phi_i}{M} \leq \frac{2\delta}{(d-1)(L+1/2)}$ , where  $0 < \delta \leq \epsilon$ . With this choice of sub-sector width,  $S_{\mathbf{m}} \cap \partial G_\gamma^*$  can be covered by a  $\delta$ -ball. To see this, consider two points in  $S_{\mathbf{m}} \cap \partial G_\gamma^*$  -  $(g(\phi), \phi)$  and  $(g(\theta), \theta)$ . Proceeding as before, we have:

$$\begin{aligned} \|(g(\phi), \phi) - (g(\theta), \theta)\| &\leq (L+1/2) \sum_{i=1}^{d-1} |\phi_i - \theta_i| \\ &\leq (L+1/2) \sum_{i=1}^{d-1} \frac{\Delta\phi_i}{M} \leq 2\delta. \end{aligned}$$

Since each sub-sector can be covered by a  $\delta$ -ball, the minimum number of  $\delta$ -balls needed to cover  $B(x, \epsilon) \cap \partial G_\gamma^*$  is equal to the minimum number of sub-sectors needed ( $M^{d-1}$ ). This corresponds to the smallest  $M$  such that  $\max_i \frac{\Delta\phi_i}{M} \leq \frac{2\delta}{(d-1)(L+1/2)}$ . Therefore, minimum number of  $\delta$ -balls needed to cover  $B(x, \epsilon) \cap \partial G_\gamma^*$  is equal to

$$\begin{aligned} &\left( \left\lceil \frac{(d-1)(L+1/2) \max_i \Delta\phi_i}{2\delta} \right\rceil \right)^{d-1} \\ &\leq \left( \frac{(d-1)(L+1/2) \max_i \Delta\phi_i}{2\delta} + 1 \right)^{d-1} \\ &\leq \left( \frac{2(d-1)(2L+1)\epsilon}{h} \frac{\epsilon}{\delta} + \frac{\epsilon}{\delta} \right)^{d-1} \\ &\leq \left( \frac{2(d-1)(2L+1)}{h} + 1 \right)^{d-1} \left( \frac{\epsilon}{\delta} \right)^{d-1} \\ &:= C_4 \left( \frac{\epsilon}{\delta} \right)^{d-1} \end{aligned}$$

The second inequality follows since from Eq. (8),  $\Delta\phi_i \leq \frac{8\epsilon}{h}$  for all  $i$ , and since  $\delta \leq \epsilon$ . Therefore, any level set  $G_\gamma^*$  with density  $f \in \mathcal{F}_{SL}$  also satisfies [C]. ■

## Acknowledgements

The authors would like to thank Rui Castro for helpful discussions and carefully reviewing the paper.

## References

- [Cav97] L. Cavalier. Nonparametric estimation of regression level sets. *Statistics*, 29:131–160, 1997.  
[CD99] E. Candés and D. L. Donoho. Curvelets: A surprisingly effective nonadaptive representation for

objects with edges. *Curves and Surfaces*, Larry Schumaker et al., Ed. Vanderbilt University Press, Nashville, TN, 1999.

- [CMC06] A. Cuevas, W. G. Manteiga, and A. R. Casal. Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.*, 48(1):7–19, 2006.  
[DL01] L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation*. Springer, NY, 2001.  
[Don99] D. L. Donoho. Wedgelets: Nearly-minimax estimation of edges. *Ann. Statist.*, 27:859–897, 1999.  
[Fal90] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, West Sussex, England, 1990.  
[Har75] J. A. Hartigan. *Clustering Algorithms*. Wiley, NY, 1975.  
[KT93] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*. Springer, NY, 1993.  
[LMS97] O. V. Lepski, E. Mammen, and V. G. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25(3):929–947, 1997.  
[LPS99] R. Y. Liu, J. M. Parelus, and K. Singh. Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.*, 27(3):783–858, 1999.  
[Pol95] W. Polonik. Measuring mass concentrations and estimating density contour cluster-an excess mass approach. *Ann. Statist.*, 23(3):855–881, 1995.  
[RV06] Philippe Rigollet and Regis Vert. Fast rates for plug-in estimators of density level sets, available at <http://www.citebase.org/abstract?id=oai:arxiv.org:math/0611473>, 2006.  
[SD07] C. Scott and M. Davenport. Regression level set estimation via cost-sensitive classification. *IEEE Trans. Signal Process.*, 55(6):2752–2757, 2007.  
[SHS05] I. Steinwart, D. Hush, and C. Scovel. A classification framework for anomaly detection. *J. Mach. Learn. Res.*, 6:211–232, 2005.  
[SN06] C. Scott and R. Nowak. Learning minimum volume sets. *J. Mach. Learn. Res.*, 7:665–704, 2006.  
[SSN07] A. Singh, C. Scott, and R. D. Nowak. Adaptive hausdorff estimation of density level sets. Technical Report ECE-07-06, University of Wisconsin - Madison, ECE Dept., available at [www.cae.wisc.edu/~singh/TR\\_Hausdorff.pdf](http://www.cae.wisc.edu/~singh/TR_Hausdorff.pdf), 2007.  
[Stu03] W. Stuetzle. Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample. *J. Classification*, 20(5):25–47, 2003.  
[Tsy97] A. B. Tsybakov. On nonparametric estimation of density level sets. *Ann. Statist.*, 25:948–969, 1997.  
[VV06] R. Vert and J.-P. Vert. Consistency and convergence rates of one-class svms and related algorithms. *J. Mach. Learn. Res.*, 7:817–854, 2006.  
[WN07] R. Willett and R. Nowak. Minimax optimal level set estimation. *IEEE Trans. Image Proc.*, 16(12):2965–2979, 2007.