Adaptive Hausdorff Estimation of Density Level Sets

Aarti Singh, Clayton Scott and Robert Nowak

Technical Report ECE-07-06
Department of Electrical and Computer Engineering
University of Wisconsin - Madison

November 26, 2007

Abstract

Consider the problem of estimating the γ-level set $G^*_\gamma = \{x : f(x) \geq \gamma\}$ of an unknown d-dimensional density function f based on n independent observations $X_1, \ldots, X_n$ from the density. This problem has been addressed under global error criteria related to the symmetric set difference. However, in certain applications such as anomaly detection and clustering, a more uniform mode of convergence is desirable to ensure that the estimated set is close to the target set everywhere. The Hausdorff error criterion provides this degree of uniformity and hence is more appropriate in such situations. It is known that the minimax optimal rate of convergence for the Hausdorff error is $(n/\log n)^{-1/(d+2\alpha)}$ for level sets with Lipschitz boundaries, where the parameter α characterizes the regularity of the density around the level of interest. However, the estimators proposed in previous work achieve this rate for very restricted classes of sets (e.g. the boundary fragment and star-shaped sets) that effectively reduce the set estimation problem to a function estimation problem. This characterization precludes the existence of multiple connected components, which is fundamental to many applications such as clustering. Also, all previous work assumes knowledge of the density regularity as characterized by the parameter α. In this paper, we present a procedure that is adaptive to unknown regularity conditions and achieves near minimax optimal rates of Hausdorff error convergence for a class of level sets with very general shapes and multiple connected components at arbitrary orientations.

1 Introduction

Level sets provide a useful summary of the function for many applications including clustering [1, 2], anomaly detection [3, 4, 5], functional neuroimaging [6, 7], bioinformatics [8], digital elevation mapping [9, 10], and environmental monitoring [11]. In practice, however, the function itself is unknown a priori and only a finite number of observations related to f are available. In this paper we focus on the density level set problem; extensions to general regression level set estimation should be possible using a similar approach, but are beyond the scope of this paper. Let $X_1, \ldots, X_n$ 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 γ, consider the γ-level set of the density f:

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

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

Most literature available on level set estimation methods [3, 4, 12, 9, 13, 14, 15, 16] considers error measures related to the symmetric set difference, $G_1 \Delta G_2 = (G_1 \setminus G_2) \cup (G_2 \setminus G_1)$. For example, in [3, 13, 14, 16] a probability measure of the symmetric set difference is considered, and in [12, 9, 16] a probability measure of weighted symmetric set difference is considered, the weight being proportional to the deviation of the function from the desired level. Both these measures are global measures of *average* closeness between two sets. However, many applications such as anomaly detection and clustering require a more local and uniform error measure. Controlling the symmetric difference error does not provide this kind of control and does not ensure accurate recovery of the topological features. To see this, consider a level set with two components as depicted in Figure 1. The figure also shows two candidate estimates, one estimate connects the two components by a "bridge" (resulting in a dumbbell shaped set), while the other preserves the (non)-connectivity. However, both candidate sets have the same symmetric difference, and hence a method that controls the symmetric difference may not favor the one that preserves topological properties over the other. Thus, a uniform measure of closeness between sets is necessary in such situations. We advocate the use of the Hausdorff error metric, defined as follows between two non-empty sets:

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

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. This error measure provides a uniform mode of convergence as it controls the deviation of even a single point from the desired set. In the dumbbell shaped set in Figure 1, the Hausdorff error is proportional to the distance between the clusters (i.e., the length of the bridge).

There are some existing results pertaining to nonparametric level set estimation using the Hausdorff error metric [13, 14, 17], but these works focus on very

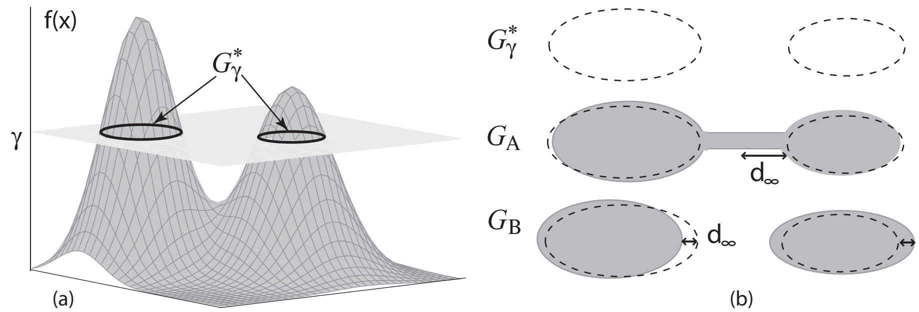

Figure 1: (a) The γ-level set G^*_γ of a density function $f(x)$, (b) Two candidate set estimates G_A and G_B with the same symmetric difference error ($G_A \Delta G^*_\gamma = G_B \Delta G^*_\gamma$), however G_A does not preserve the topological properties (non-connectivity) and has large Hausdorff error $d_\infty(G_A, G^*_\gamma)$, while G_B preserves non-connectivity and has small Hausdorff error $d_\infty(G_B, G^*_\gamma)$.

restrictive classes of level sets (e.g., the boundary fragment and star-shaped set classes). These restrictions, which effectively reduce the set estimation problem to a 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 function estimation problem precludes the existence of multiple connected components, which is fundamental to many applications. Moreover, the estimation techniques proposed in [13, 14, 17] require precise knowledge of the local regularity of the distribution (quantified by the parameter α, 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 applications. Recently, a plug-in method based on sup-norm density estimation was put forth in [18] that can handle more general classes than boundary fragments or star-shaped sets, however density estimation requires global smoothness assumptions. Also, the method only deals with a special case of the regularity condition considered here ($\alpha = 1$), and is therefore not adaptive to unknown density regularity.

The major contribution of this paper is the development of a novel theoretical framework for Hausdorff accurate level set estimation that can handle broad classes of level sets with very general shapes and multiple connected components at arbitrary orientations, and is also adaptive to unknown degrees of density regularity. It is thus applicable to clustering and other practical level set estimation problems. Further, the theoretical guarantees only require the density to be regular locally in the vicinity of the level of interest. The basic approach is illustrated through the use of histogram-based estimators, although extensions to more general partitioning schemes such as spatially adaptive partitions [19, 20, 21] are possible. The theory and method may also provide a useful starting point for future investigations into alternative schemes, such as kernel-based approaches [5], that may be better suited for higher dimensional settings.

To motivate the importance of Hausdorff accurate level set estimation, let us briefly discuss its relevance in some applications.

Anomaly detection - A common approach to anomaly detection is to learn a (high) density level set of the nominal data distribution [3, 4, 5]. Samples that fall outside the level set, in the low density region, are considered anomalies. Level set methods based on a symmetric difference error measure may produce estimates that veer greatly from the desired level set at certain places and potentially include regions of low density, since the symmetric difference is a global error. Anomalous distributions concentrated in such places would elude detection. On the other hand, level set estimators based on the Hausdorff metric are guaranteed to be uniformly close to the desired level set, and therefore are more robust to anomalies in such situations.

Clustering - Density levels set estimators are used by many data clustering procedures [1, 2, 22], and the correct identification of connected level set components (i.e., clusters) is crucial to their success. The Hausdorff criterion can be used to provide theoretical guarantees regarding clustering since the connected components of a level set estimate that is ϵ-accurate in the Hausdorff sense, characterize the true level set clusters (in number, shapes, and locations), provided the true clusters remain topologically distinct upon erosion or dilation by an ϵ-ball. The last statement holds since

$$d_\infty(G_1, G_2) \le \epsilon \qquad \Longrightarrow \qquad G_1 \subseteq G_2^\epsilon, \;\; G_2 \subseteq G_1^\epsilon,$$

where G^ϵ denotes the set obtained by dilation of set G by an ϵ-ball.

Data Ranking - Hausdorff accurate level set estimation is also relevant for ranking or ordering data using the notion of data-depth [23]. Density level sets correspond to likelihood-depth contours and Hausdorff distance is a robust measure of accuracy in estimating the data-depth as it is less susceptible to severe misranking as compared to symmetric set difference based measures.

Thus, Hausdorff accurate estimation of density level sets is an important problem with many potential applications. However, in all these applications there are other issues, for example, selection of the density levels of interest, that are beyond the scope of this paper.

The paper is organized as follows. Section 2 states the basic assumptions needed for Hausdorff accurate level set estimation. Section 3 discusses the issue with direct Hausdorff estimation and provides motivation for an alternate error measure. Section 4 proposes a histogram-based approach to Hausdorff accurate level set estimation, and Section 5 addresses adaptivity to unknown density regularity. Extensions and some concluding remarks are given in Section 6. Section

7 characterizes the Hausdorff error performance of the estimation procedure and presents proofs of the main results. The Appendix contains proofs of lemmas used in the theoretical analysis.

2 Density assumptions

In this paper, we assume that the domain of the density f is the unit hypercube in d-dimensions, i.e. $\mathcal{X} = [0, 1]^d$. Extensions to other compact domains are straightforward. Further, the density is assumed to be bounded with range $[0, f_{\max}]$. Controlling the Hausdorff accuracy of level set estimates also requires some smoothness assumptions on the density and the level set boundary, which are stated below. But before that we need some definitions:

- **ϵ-Ball:** An ϵ-ball centered at a point $x \in \mathcal{X}$ is defined as

$$B(x, \epsilon) = \{y \in \mathcal{X} : ||x - y|| \leq \epsilon\}.$$

  Here $|| \cdot ||$ denotes the Euclidean distance.

- **Inner ϵ-cover:** - An inner ϵ-cover of a set $G \subseteq \mathcal{X}$ is defined as the union of all ϵ-balls contained in G. Formally,

$$\mathcal{I}_\epsilon(G) = \bigcup_{x: B(x,\epsilon) \subseteq G} B(x, \epsilon)$$

We are now ready to state the assumptions. The most crucial one is the first, which characterizes the relationship between distances and changes in density. The last two are topological assumptions on the level set and are essentially a generalization of the notion of Lipschitz functions to closed hypersurfaces.

[A] *Local density regularity:* The density is α-regular around the γ-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 ∂G^*_γ is the boundary of the true level set G^*_γ.

This assumption is similar to the one employed in [14, 17]. The condition states that the deviation in density from the level of interest scales as the α-th power of distance from the level set boundary. The regularity parameter α determines the rate of error convergence for level set estimation. Accurate estimation is more difficult at levels where the density is relatively flat (large α), as intuition would suggest. It is important to point out that in this paper we do not assume knowledge of α unlike previous investigations into Hausdorff accurate level set estimation [13, 14, 17, 18]. 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. The case $\alpha = 0$, which corresponds to a jump in the density at level γ, can also be handled but requires a slightly modified approach. Hence, we present the main analysis restricting $\alpha > 0$ to keep the presentation simple, and later discuss extension of the method to $\alpha \geq 0$ (see Section 6.1). We also discuss a generalization of this two-sided assumption in Section 6.2 that allows the regularity parameter α to vary along the level set boundary.

[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.

For a fixed set of positive numbers C_1, C_2, C_3, ϵ_0, δ_0, $f_{\max}$, γ and α, we define the following class of densities:

Definition 1. $\mathcal{F}^*_1(\alpha)$ *denotes the class of densities satisfying assumptions **[A,B]**.*

The dependence on other parameters is omitted as these do not influence the minimax optimal rate of convergence. In the paper, we present a method that provides minimax optimal rates of convergence for this class of densities, given knowledge of the density regularity parameter α. We also extend the method to achieve adaptivity to α under the following additional assumption:

[C] *Level set boundary dimension:* There exists a constant $C_4 > 0$ such that for all $x \in \partial G^*_\gamma$ and all ϵ, δ 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 [24] 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 α is known. However, the condition is needed to achieve adaptivity using the proposed method, as we shall discuss later.

The corresponding class of densities is defined as:

Definition 2. $\mathcal{F}^*_2(\alpha)$ *denotes the class of densities satisfying assumptions **[A,B, C]**.*

Assumptions [**B,C**] are essentially a generalization of the notion of Lipschitz functions to closed hypersurfaces, and allow for level sets with multiple connected components and arbitrary orientations. They basically imply that the boundary looks locally like a Lipschitz function. Thus these restrictions on the shape of the level sets are quite mild and less restrictive than those considered in the previous literature on Hausdorff level set estimation. In fact [**B,C**] are satisfied by a Lipschitz boundary fragment or star-shaped set as considered in [13, 14, 17]; please refer to Section 7.4 for a formal proof. Since the class of densities with star-shaped Lipschitz level set boundaries considered by Tsybakov [14] satisfy assumptions [**A,B,C**], it is a subset of $\mathcal{F}_2^*(\alpha) \subset \mathcal{F}_1^*(\alpha)$. This implies that the minimax lower bound for Lipschitz star-shaped sets established by Tsybakov (Theorem 4 in [14]) holds for the classes $\mathcal{F}_1^*(\alpha)$ and $\mathcal{F}_2^*(\alpha)$ under consideration as well. Hence, we have the following proposition. Let $\mathbb{E}$ denote expectation with respect to the random data sample.

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

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

and

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

Here the inf *is taken over all possible set estimators $\widehat{G}_n$, i.e. all measurable sets in $\mathcal{X}$.*

Remark: Extensions to additional smoothness conditions on the boundary ∂G_γ^* (e.g., Hölder regularity > 1) may be possible, but are beyond the scope of this paper. The earlier work of [13, 14, 17] does address higher smoothness of the boundary, but that is possible only because for the classes under consideration, level set estimation reduces to a function estimation problem.

3 Motivating an Error Measure for Hausdorff control

Direct Hausdorff estimation is challenging as there exists no natural empirical measure that can be used to gauge the Hausdorff error of an estimate. In this section, we investigate how Hausdorff control can be obtained indirectly using an alternate error measure that is based on density deviation error rather than distance deviation. While the alternate error measure we introduce in this section is easily motivated and arises naturally, it requires global smoothness assumptions on the density, whereas only local smoothness in the vicinity of the level set is required for accurate level set estimation. Hence, in the next section, we present our final alternate measure that is based on the insights gained here.

Notice that the density regularity condition [**A**] suggests that control over the deviation of any point in the estimate from the true level set boundary $\rho(x, \partial G^*_\gamma)$ can be obtained by controlling the deviation from the desired density level. In other words, a change in density level reflects change in distance. Moreover, in order to obtain a sense of distance from an estimate of density variation based on a small sample, the level set boundary cannot vary too irregularly. Specifically, the boundary should not have arbitrarily small features (e.g., cusps) that cannot be reliably detected from a small sample. Such features are ruled-out by assumption [**B**]. Thus, under regularity conditions on the function and level set boundary, the deviation of the density function from the desired level can be used as a surrogate for Hausdorff error. Consider the following error measure:

$$\begin{aligned} \mathcal{E}(G) &= \max\{\sup_{x \in G^*_\gamma \setminus G} (f(x) - \gamma), \sup_{x \in G \setminus G^*_\gamma} (\gamma - f(x))\} && (1) \\ &= \sup_{x \in \mathcal{X}} (\gamma - f(x))[\mathbb{I}_{x \in G} - \mathbb{I}_{x \notin G}] && (2) \end{aligned}$$

where $\mathbb{I}$ denotes the indicator function and by convention $\sup_{x \in \emptyset} g(x) = 0$ for any non-negative function $g(\cdot)$. The error measure $\mathcal{E}(G)$ has a natural empirical counterpart, $\widehat{\mathcal{E}}(G)$, obtained by simply replacing $f(x)$ by a density estimator $\widehat{f}(x)$. Notice that the set $\widehat{G}$ minimizing the empirical error corresponds to a plug-in level set estimator. It can be shown (see Appendix A) that, if $\widehat{f}(x)$ is consistent in the sup-norm, then under assumptions [**A**] and [**B**]

$$d_\infty(\widehat{G}, G^*_\gamma) \leq C\, \mathcal{E}(\widehat{G})^{1/\alpha},$$

where $C > 0$ is a constant. Since the difference between the true and empirical errors can be bounded as

$$|\mathcal{E}(G) - \widehat{\mathcal{E}}(G)| \leq \sup_{x \in \mathcal{X}} |f(x) - \widehat{f}(x)|,$$

we have:

$$d_\infty(\widehat{G}, G^*_\gamma) = O(||f(x) - \widehat{f}(x)||_\infty^{1/\alpha}).$$

This shows that the sup-norm error of a density estimate gives an upper bound on the Hausdorff error of a plug-in level set estimate, which agrees with Cuevas' result [18] for $\alpha = 1$. However, Hausdorff accuracy of a level set estimate only depends on the accuracy of the density estimate around the level of interest. Arbitrarily rough and complicated behavior of the density away from the level of interest can cause a large sup-norm density error, leading to large Hausdorff error of the plug-in level set estimate. This reflects a major drawback of the plug-in approach. Therefore, we follow Vapnik's maxim: *When solving a given problem, try to avoid solving a more general problem as an intermediate step* [25], and instead of solving the harder intermediate problem of sup-norm density estimation (which depends on the global smoothness of the density), we approach the set estimation problem directly.

4 Hausdorff accurate Level Set Estimation using Histograms

We consider a modified version of the error measure introduced above. This alternative will form the basis for our theory and methodology. Let Π denote a partition of $[0,1]^d$ and let G be any set defined in terms of this partition (i.e., the union of any collection of cells of the partition). We will consider a hierarchy of partitions with increasing complexity and the sets G, defined in terms of the partitions, form candidate representations of the γ level set of the density f. The partition could, for example, correspond to a decision tree or regular histogram. In this paper, we will focus on the regular histogram. Define the error of G as

$$\mathcal{E}_\gamma(G) = \sup_{A \in \Pi(G)} (\gamma - \bar{f}(A))[\mathbb{I}_{A \subseteq G} - \mathbb{I}_{A \nsubseteq G}].$$

Here $\Pi(G)$ denotes the partition associated with set G and $\bar{f}(A) = P(A)/\mu(A)$ denotes average of the density function on the cell A, where P is the unknown probability measure and μ is the Lebesgue measure. Note the analogy between this error and that defined in (1). We would like to point out that even though this error depends on the class of candidate sets being considered, we will use it to establish control over the Hausdorff error which is independent of the candidate class. This performance measure evaluates a set based on the maximum deviation of the average density in a cell of the partition from the γ level. Note that $\left(\gamma - \bar{f}(A)\right)[\mathbb{I}_{A \subseteq G} - \mathbb{I}_{A \nsubseteq G}] > 0$ whenever a cell with average density $\bar{f}(A) < \gamma$ is included in the set G or a cell with $\bar{f}(A) > \gamma$ is excluded. We establish that control over $\mathcal{E}_\gamma(G)$, along with appropriate choice of the partition resolution, is sufficient for Hausdorff control. The appropriate resolution depends on the density regularity α near the level of interest. If the density varies sharply (small α) near the level of interest, then accurate estimation is easier and a fine resolution suffices. Level set estimation is more difficult if the density is very flat (large α) and hence a lower resolution (more averaging) is required. We develop a method that automatically selects an appropriate resolution without requiring prior knowledge of the density regularity. Notice that even though the method is based on density averages over some partition, it is not a plug-in approach as the partition is not optimized for density estimation.

We propose to select a density level estimate based on regular histograms. 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. Then the family of candidate sets, denoted $\mathcal{G}_j$, is comprised of sets G formed by taking the union of any collection of cells in $\mathcal{A}_j$. Let $j(G)$ denote the smallest j such that $G \in \mathcal{G}_j$. Then we may write $\Pi(G) = \mathcal{A}_{j(G)}$. Therefore, for histogram based sets, the error measure can be written as:

$$\mathcal{E}_\gamma(G) = \max_{A \in \mathcal{A}_{j(G)}} \left(\gamma - \bar{f}(A)\right)[\mathbb{I}_{A \subseteq G} - \mathbb{I}_{A \nsubseteq G}]$$

A natural empirical error, $\widehat{\mathcal{E}}(G)$, is obtained by replacing $\bar{f}(A)$ with its empirical counterpart.

$$\widehat{\mathcal{E}}_\gamma(G) = \max_{A \in \mathcal{A}_{j(G)}} \left(\gamma - \widehat{f}(A)\right) [\mathbb{I}_{A \subseteq G} - \mathbb{I}_{A \nsubseteq G}]$$

Here $\widehat{f}(A) = \frac{\widehat{P}(A)}{\mu(A)}$, 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.

Among all sets at a fixed resolution, the one minimizing the empirical error $\widehat{\mathcal{E}}_\gamma$ is a natural candidate:

$$\widehat{G}_j = \arg\min_{G \in \mathcal{G}_j} \widehat{\mathcal{E}}_\gamma(G) \qquad\qquad j = 0, 1, \ldots, J \tag{3}$$

This rule selects the set that includes all cells with empirical density $\widehat{f}(A) > \gamma$ and excludes all cells with $\widehat{f}(A) < \gamma$. The empirical error minimization procedure is sufficient to guarantee minimax optimal Hausdorff estimation if the appropriate resolution can be determined. If the regularity parameter α is known, then the correct resolution can be chosen (as in [14, 17]). In this case, the empirical error minimization procedure of Eq. (3) achieves the minimax rate over the class of densities given by $\mathcal{F}_1^*(\alpha)$.

We introduce the notation $a_n \asymp b_n$ to denote that $a_n = O(b_n)$ and $b_n = O(a_n)$. Also, let $\mathbb{E}$ denote expectation with respect to the random data sample.

Theorem 1. *Assume that the local density regularity parameter α is known. Pick j such that $2^{-j} \asymp s_n(n/\log n)^{-1/(d+2\alpha)}$, where s_n is a monotone diverging sequence. Let $\widehat{G}_j$ be the estimate generated by empirical error minimization as per Eq.(3). 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 7.

Theorem 1 provides an upper bound on the Hausdorff error of our estimate. If s_n is slowly diverging, e.g. 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 empirical error minimization method can achieve near minimax optimal rates, given knowledge of the density regularity. We would like to point out that if the parameter δ_0 characterizing the locality of assumption **[A]** and the density bound $f_{\max}$ are also known, then the appropriate resolution can be chosen as $j = \lfloor \log_2 \left(c^{-1}(n/\log n)^{1/(d+2\alpha)}\right) \rfloor$, 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 empirical error minimization procedure exactly achieves the minimax optimal rate.

5 Adaptivity to unknown density regularity

Without prior knowledge of α, determining the proper resolution is a delicate matter. This is because the Hausdorff error of the set minimizing $\mathcal{E}_\gamma$ is determined by one of the cells intersecting the boundary of G^*_γ. The average density in a boundary cell could be arbitrarily close to the level γ (error $\mathcal{E}_\gamma$ can be arbitrarily small), irrespective of the density regularity and the resolution. As an extreme example, even if the density varies significantly over $[0,1]^d$ it is possible that the average density on the unit hypercube is γ, and clearly this single-cell partition would provide a very poor level set estimate in general.

Thus, an additional criterion is required to select the appropriate resolution, one that gauges the uniformity of the density (as governed by α) within the boundary cells. Thus, we introduce the following auxiliary device or *vernier*:

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

Here $j' = \lfloor j + \log_2 s_n \rfloor$, where s_n is a slowly diverging monotone sequence, e.g. $\log n$, $\log\log n$, 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. Even though the average density on a cell A may be close to γ, contrasting the average density within subcells of A indicates whether or not the density is uniformly close to γ over the cell. Thus, because the vernier gauges whether or not the histogram values are reasonable surrogates for the density function, it indicates if the histogram resolution is fine enough. We will show that in fact the vernier allows our procedure to adapt to the unknown regularity parameter α. And 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^*_γ passes through every subcell of A (since then the subcell averages too may be deceptively close to γ). 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:

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

Here $\widehat{f}(A) = \frac{\widehat{P}(A)}{\mu(A)}$, 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. We propose a complexity regularization scheme wherein the empirical vernier $\widehat{\mathcal{V}}_{\gamma,j}$ is balanced by a penalty term:

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

where $0 < \delta < 1$ is a confidence parameter. Notice that the penalty is computable from the given observations. The precise form of Ψ is chosen so that minimizing the empirical vernier plus penalty provides control over the true vernier (refer to Section 7.2). The vernier control leads to selection of the appropriate resolution. The final level set estimate is given by

$$\widehat{G} = \widehat{G}_{\widehat{j}} \tag{4}$$

where

$$\widehat{j} = \arg\min_{0 \leq j \leq J} \left\{ \widehat{\mathcal{V}}_{\gamma,j} + \Psi_{j'} \right\} \tag{5}$$

Recall that $j' = \lfloor j + \log_2 s_n \rfloor$. Observe that the value of the empirical vernier decreases with increasing resolution j as better approximations to the true level are available. On the other hand, the penalty is designed to increase with j 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 and use of the vernier $\mathcal{V}_{\gamma,j}$, that is sensitive to the resolution and unknown regularity parameter α, ensures optimal choice of the resolution. Hence, the proposed method leads to minimax optimal rates of convergence without requiring prior knowledge of any parameters.

Recall the notation $a_n \asymp b_n$ to denote $a_n = O(b_n)$ and $b_n = O(a_n)$. Also, let $\mathbb{E}$ denote expectation with respect to the random data sample.

Theorem 2. *Pick $J = J(n)$ such that $2^{-J} \asymp s_n(n/\log n)^{-1/d}$, where s_n is a monotone diverging sequence. Let $\widehat{j}$ denote the resolution chosen by the complexity penalized method as given by Eq. (5), and $\widehat{G}$ denote the final estimate of Eq. (4). With probability at least $1 - 3/n$, for all densities in the class $\mathcal{F}_2^*(\alpha)$,*

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

for n large enough (so that $s_n > 4C_4 6^d$), where $c_1, c_2 > 0$ are constants. And

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

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

The proof is given in Section 7.

The maximum resolution 2^J can be easily chosen, based only on n, and allows the optimal resolution for any α to lie in the search space. Observe that by appropriate choice of s_n, for example $s_n = (\log n)^{\epsilon/2}$ with ϵ 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.

Remark 1: To prove the results of Theorems 1 and 2, we do not need to assume

an exact form for s_n except that it is a monotone diverging sequence. However, s_n needs to be slowly diverging for the derived rates to be near minimax optimal.
Remark 2: The factor of 6^d in the condition $s_n > 4C_4 6^d$ can be eliminated if assumption **[C]** is stated as a box-counting dimension [24] assumption, i.e. in terms of cells of a uniform partition of the domain rather than ϵ and δ balls, and hence the number of samples required only depends on the constant C_4.
Remark 3: We would like to point out that even though we state the convergence results in expectation, the proofs also establish convergence in probability. Thus, convergence actually holds in a stronger sense.

6 Conclusions and Extensions

In this paper, we developed a Hausdorff accurate level set estimation method that is adaptive to unknown density parameters and handles very general classes of level sets, while achieving minimax optimal rates. We now discuss extensions of this framework to handle jump in the density ($\alpha = 0$) around the level of interest, and generalize the two-sided regularity assumption to allow the regularity α to vary along the level set boundary. We also extend the results to multiple level set estimation and discuss adaptive partition based estimators.

6.1 Allowing jumps in the density

The case $\alpha = 0$ corresponds to a jump in the density at the level of interest, and can be handled by a slight modification of the earlier framework. Notice that the current form of the vernier may fail to select an appropriate resolution in the jump case; for example, if the density is piecewise constant on either side of the jump, the vernier output is the same irrespective of the resolution. A slight modification of the vernier as follows

$$\mathcal{V}_{\gamma,j} = 2^{-j'/2} \min_{A \in \mathcal{A}_j} \left\{ |\gamma - \max_{A' \in \mathcal{A}_{j'} \cap A} \bar{f}(A')| + |\gamma - \min_{A'' \in \mathcal{A}_{j'} \cap A} \bar{f}(A'')| \right\},$$

makes the vernier sensitive to the resolution even for the jump case and biases a vernier minimizer towards finer resolutions. A fine resolution is needed for the jump case to approximate the density well (notice that a fine resolution implies less averaging, however the resulting instability in the estimate can be tolerated as there is a jump in the density). Since the penalty is designed to control the deviation of empirical and true vernier, it also needs to be scaled accordingly:

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

This ensures that balancing the vernier and penalty leads to the appropriate resolution for the whole range of the regularity parameter, $0 \leq \alpha < \infty$.

While it is clear why the modification is needed, the exact form of the modifying factor $2^{-j'/2}$ arises from technical considerations and is somewhat non-intuitive. Hence, we omitted the jump case in our earlier analysis to keep the presentation simple. Also, the proof arguments are slightly different and rely on the fact that for large enough n, only the cells intersecting the boundary can be in error with high probability. Please refer to Appendix B for a detailed proof.

6.2 Allowing regularity variation along the boundary

The two-sided regularity assumption **[A]** used in this paper and in earlier work [13, 14, 17] is necessary for adaptive Hausdorff accurate level set estimation. The lower bound is necessary to translate control over density level deviation to control over distance deviation. In fact, a similar lower bound is also needed when one is interested in symmetric difference control [14], and not density deviation weighted symmetric difference. The upper bound is needed for adaptivity so that balancing approximation and estimation errors for density deviation imply balancing the same for distance deviation. However, the two-sided nature of this assumption might be stringent in practice, and hence we consider the following generalization of this condition:

[A'] The density is $[\alpha_2, \alpha_1]$-regular around the γ-level set, $0 < \alpha_2 \leq \alpha_1 < \infty$, if there exist constants $C_1, C_2 > 0$ and $\delta_0 > 0$ such that

$$C_1 \rho(x, \partial G^*_\gamma)^{\alpha_1} \leq |f(x) - \gamma| \leq C_2 \rho(x, \partial G^*_\gamma)^{\alpha_2}$$

for all $x \in \mathcal{X}$ with $|f(x) - \gamma| \leq \delta_0$, where ∂G^*_γ is the boundary of the true level set G^*_γ.

Now let us consider the class of densities $\mathcal{F'}^*_2(\alpha_1, \alpha_2)$ satisfying assumptions **[A', B]** and **[C]**. Observe that $\mathcal{F}^*_2(\alpha_1) \subseteq \mathcal{F'}^*_2(\alpha_1, \alpha_2)$, hence we have a lower bound that scales as $(n/\log n)^{-\frac{1}{d+2\alpha_1}}$. If α_1 is known, the resolution can be chosen as $2^{-j} \asymp s_n (n/\log n)^{-1/(d+2\alpha_1)}$, and we obtain an upper bound (using Lemma 3 in the proof of Theorem 1) that matches the lower bound up to an s_n factor. However, if the regularity is not known, we obtain the following generalization of Theorem 2:

Corollary 1. *Pick $J = J(n)$ such that $2^{-J} \asymp s_n (n/\log n)^{-1/d}$, where s_n is a monotone diverging sequence. Let $\widehat{G}_\gamma$ denote the estimate generated using the complexity penalized procedure of Eq. (4). Then*

$$\sup_{f \in \mathcal{F'}^*_2(\alpha_1, \alpha_2)} \mathbb{E}[d_\infty(\widehat{G}_\gamma, G^*_\gamma)] \leq C s_n^2 \left(\frac{n}{\log n}\right)^{-\frac{\alpha_2/\alpha_1}{d+2\alpha_2}}$$

for all n, here $C \equiv C(C_1, C_2, C_3, C_4, \epsilon_o, f_{\max}, \delta_0, d, \alpha_1, \alpha_2) > 0$.

Notice that the upper bound is not minimax optimal. This is because the adaptive resolution needs to be picked based on some form of density estimate, in our case the vernier, which results in a choice of $2^{-j} \asymp s_n (n/\log n)^{-1/(d+2\alpha_2)}$ for the resolution under assumption **[A']** (see proof of Theorem 2), which is not optimal unless $\alpha_1 = \alpha_2$.

6.3 Multiple level set estimation

The proposed framework can easily be extended to simultaneous estimation of level sets at multiple levels $\Gamma = \{\gamma_k\}_{k=1}^K$ $(K < \infty)$. Assuming the density regularity condition [**A**] holds with parameter α_k for the γ_k level, we have:

Corollary 2. *Pick $J = J(n)$ such that $2^{-J} \asymp s_n(n/\log n)^{-1/d}$, where s_n is a monotone diverging sequence. Let $\widehat{G}_{\gamma_k}$ denote the estimate generated using the complexity penalized procedure of Eq. (4) for level γ_k. Then*

$$\max_{1\le k\le K} \sup_{f\in\mathcal{F}_2^*(\alpha_k)} \mathbb{E}[d_\infty(\widehat{G}_{\gamma_k}, G^*_{\gamma_k})] \le C s_n^2 \left(\frac{n}{\log n}\right)^{-1/(d+2\max_k \alpha_k)}$$

for all n, here $C \equiv C(C_1, C_2, C_3, C_4, \epsilon_o, f_{\max}, \delta_0, d, \{\alpha_k\}_{k=1}^K) > 0$.

Notice that, while the estimate $\widehat{G}_{\gamma_k}$ at each level is adaptive to the local density regularity as determined by α_k, the overall convergence rate is determined by the level where the density is most flat (largest α_k).

Another issue that comes up in multiple level set estimation is nestedness. If the density levels of interest Γ are sorted, $\gamma_1 \le \gamma_2 \le \ldots \le \gamma_K$, then the true level sets will be nested $G^*_{\gamma_1} \supseteq G^*_{\gamma_2} \supseteq \ldots \supseteq G^*_{\gamma_K}$. However, the estimates $\{\widehat{G}_{\gamma_k}\}_{k=1}^K$ may not be nested as the resolution at each level is determined by the local density regularity (α_k). For some applications, for example hierarchical clustering, nestedness of the estimates may be desirable. We can enforce nestedness by choosing the same resolution, corresponding to the largest α_k, at all levels. Since the largest α_k corresponds to smallest vernier $\mathcal{V}_{\gamma_k,j}$ (see Lemma 5), this is accomplished by selecting the resolution according to

$$\widehat{j} = \arg\min_{0\le j\le J} \left\{ \min_{1\le k\le K} \widehat{\mathcal{V}}_{\gamma_k,j} + \Psi_{j'} \right\}.$$

This does not change the rate of convergence, however if the density is flat at one level of interest, this forces large Hausdorff error at all levels, even if the density at those levels is well-behaved (varies sharply near the level of interest).

6.4 Adaptive vs. non-adaptive partitions

It is well known that spatially adaptive partitions such as recursive dyadic partitions (RDPs) [19, 20, 21] may provide significant improvements over non-adaptive partitions like histograms for many set learning problems involving a weighted symmetric difference error measure, including classification [21], minimum volume set estimation [4] and level set estimation [9]. In fact, for many function classes, estimators based on adaptive, non-uniform partitions can achieve minimax optimal rates that cannot be achieved by estimators based on non-adaptive partitions. However, the results of this paper establish that this is not the case for the Hausdorff metric. This is a consequence of the fact that symmetric difference based errors are global, whereas the Hausdorff error is sensitive to local errors and depends on the worst case error at any point. Having non-uniform cells adapted to the regularity along the boundary can lead to faster convergence rates under global measures, whereas the Hausdorff error being dominated by the worst case error does not benefit from adaptivity of the partition.

While spatially adaptive, non-uniform partitions do not provide an improvement in convergence rates under the Hausdorff error metric, it should be noted that in practice these may have certain advantages. For example, if the connected components of a level set have different density regularities, non-uniform partitions are capable of adapting to the local smoothness around each component and this may generate better estimates in practice.

7 Proofs

In this section, we analyze the performance of the estimation procedure and establish Theorems 1 and 2. Also, we establish the lower bound of Proposition 1. The reader is referred to the Appendix for detailed proofs of the associated lemmas.

Before proceeding to the proofs, we establish two lemmas that are used throughout the proofs. Define

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

where $0 < \delta < 1$ is a confidence parameter. Notice that $\Phi_j = \Psi_j/2$ and both are empirical quantities, though we suppress the dependence on n in the notation. The choice of penalty Ψ_j is motivated by this relation since Φ_j bounds the deviation of true and empirical density averages, as the following lemma shows.

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

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

The lemma is proved in the Appendix and hinges on a pair of *relative* Vapnik-Chervonenkis (VC) inequalities ([26], Chapter 3).

The next lemma states how the density deviation bound Φ_j scales with resolution. This will be used to derive rates of convergence. For this purpose, we set $\delta = 1/n$.

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

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

It essentially reflects the fact that at finer resolutions, the amount of data per cell decreases leading to larger estimation error. The proof is given in the Appendix.

We are now ready to prove the main results. We would like to point out that some arguments in the proofs hold for s_n large enough. This implies that some of the constants in our proofs will depend on $\{s_i\}_{i=1}^{\infty}$, the exact form that the sequence s_n takes (but not on n). However, we omit this dependence for simplicity.

7.1 Proof of Theorem 1

We analyze the performance of the empirical error minimization procedure of Eq. (3), and first establish the following bound on the Hausdorff error.

Lemma 3. *Consider densities satisfying assumptions [**A**] and [**B**]. Let $\widehat{G}_j$ denote the set at resolution j that minimizes the empirical error $\widehat{\mathcal{E}}_\gamma$ as per Eq. (3). Then for all resolutions $j \equiv j(n)$ such that $2^j = O(s_n^{-1}(n/\log n)^{1/d})$, where s_n is a monotone diverging sequence, and $n \geq n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$ with probability at least $1 - 3/n$*

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

A detailed proof is given in the Appendix. The first term denotes the estimation error while the second term that is proportional to the sidelength of a cell (2^{-j}) reflects the approximation error. Recall that the estimation error for average density estimation in a cell is bounded by Φ_j as per Lemma 1. Under regularity assumption [**A**] this translates to an estimation error bound of $(\Phi_j/C_1)^{1/\alpha}$ for deviation in distance of the estimated set from the true boundary $\rho(x, \partial G^*_\gamma)$. Under assumption [**B**], the true boundary itself is well-behaved and hence does not deviate too much from the estimated boundary; thus $\rho(x, \partial\widehat{G}_j)$ scales similarly.

We now establish the result of Theorem 1. Since the regularity parameter α is known, the appropriate resolution can be chosen as $2^{-j} \asymp s_n(n/\log n)^{-\frac{1}{(d+2\alpha)}}$. Let Ω denote the event such that the bounds of Lemma 2 (with $\delta = 1/n$) and Lemma 3 hold. Then for $n \geq n_0$, $P(\bar{\Omega}) \leq 4/n$. For $n < n_0$, we can use the trivial inequality $P(\bar{\Omega}) \leq 1$. So we have, for all n

$$P(\bar{\Omega}) \leq \max(4, n_0)\frac{1}{n} =: C'\frac{1}{n}$$

Here $C' \equiv C'(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$.

So $\forall f \in \mathcal{F}^*_1(\alpha)$, we have:

$$\mathbb{E}[d_\infty(\widehat{G}_j, G^*_\gamma)] = P(\Omega)\mathbb{E}[d_\infty(\widehat{G}_j, G^*_\gamma)|\Omega] + P(\bar{\Omega})\mathbb{E}[d_\infty(\widehat{G}_j, G^*_\gamma)|\bar{\Omega}]$$

$$
\begin{aligned}
&\leq \quad \mathbb{E}[d_\infty(\widehat{G}_j, G^*_\gamma)|\Omega] + P(\bar{\Omega})\sqrt{d} \\
&\leq \quad \max(2C_3+1, 8\sqrt{d}\epsilon_o^{-1}) \left[\left(\frac{\Phi_j}{C_1}\right)^{1/\alpha} + \sqrt{d}2^{-j} \right] + C' \frac{\sqrt{d}}{n} \\
&\leq C(C_1, C_3, \epsilon_o, f_{\max}, \delta_0, d, \alpha) \max \left\{ \left(2^{jd} \frac{\log n}{n} \right)^{\frac{1}{2\alpha}}, 2^{-j}, \frac{1}{n} \right\} \\
&\leq C \max \left\{ s_n^{-d/2\alpha} \left(\frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}, s_n \left(\frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}, \frac{1}{n} \right\} \\
&\leq C(C_1, C_3, \epsilon_o, f_{\max}, \delta_0, d, \alpha) s_n \left(\frac{n}{\log n} \right)^{-\frac{1}{d+2\alpha}}
\end{aligned}
$$

The second step follows by observing 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.

□

7.2 Penalty Design

Careful selection of the penalty term is crucial to the performance of the proposed adaptive method. Ideally the appropriate resolution is obtained by minimizing the vernier $\mathcal{V}_{\gamma,j}$, and hence if the penalty term is chosen such that $\mathcal{V}_{\gamma,j} \leq \widehat{\mathcal{V}}_{\gamma,j} + \Psi_{j'}$ with high probability for all resolutions j, then the chosen resolution will be the one for which an upper bound on the vernier $\mathcal{V}_{\gamma,j}$ is minimized. If this upper bound is tight, it will lead to an effective estimator and the correct rate of convergence. This leads to the choice of the penalty

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

where $0 < \delta < 1$ is a confidence parameter. Notice that the penalty increases with decreasing cell size (i.e. increasing j) and thus penalizes high resolution sets as desired. With this penalty design, we have the following desired vernier control.

Lemma 4. *Consider $0 < \delta < 1$. With probability at least $1 - \delta$ with respect to the draw of the data, the following is true for all resolutions j:*

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

A detailed proof is given in the Appendix. The proof follows by observing that $|\mathcal{V}_{\gamma,j} - \widehat{\mathcal{V}}_{\gamma,j}| \leq 2\max_{A \in \mathcal{A}_{j'}} |\bar{f}(A) - \widehat{f}(A)| \leq \Psi_{j'}$.

7.3 Proof of Theorem 2

First we prove that the complexity penalized procedure automatically picks the right resolution for level sets with densities $f \in \mathcal{F}_2^*(\alpha)$, i.e. for densities satisfying assumptions [**A-C**]. To analyze the resolution chosen by the estimation procedure, we need the following lemma. It establishes that the vernier is sensitive to the resolution and density regularity.

Lemma 5. *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 resolutions j*

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

holds for n large enough such that $s_n > 4C_4 6^d$. Here $C \equiv C(C_2, f_{\max}, \delta_0) > 0$.

Please refer to the Appendix for proof.

Now observe that Lemmas 2, 3 and 4 hold together with probability at least $1 - 5/n$ (taking $\delta = 1/n$). Using these lemmas, we will show that for the resolution $\widehat{j}$ chosen by Eq. (5), both $\mathcal{V}_{\gamma,\widehat{j}}$ and $\Psi_{\widehat{j}'}$ are upper bounded by $C s_n^{\frac{d\alpha}{d+2\alpha}} (n/\log n)^{-\frac{\alpha}{d+2\alpha}}$, where $C \equiv C(C_2, f_{\max}, \delta_0, d, \alpha) > 0$. If this holds, then using Lemma 5, we have the following upper bound on the sidelength: For $s_n > 4C_4 6^d$,

$$2^{-\widehat{j}} \le s_n \left(\frac{\mathcal{V}_{\gamma,\widehat{j}}}{\min(\delta_0, C_1)}\right)^{1/\alpha} \quad \le \quad c_2 s_n s_n^{\frac{d}{d+2\alpha}} \left(\frac{n}{\log n}\right)^{-\frac{1}{d+2\alpha}},$$

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

$$\begin{aligned}
2^{-\widehat{j}} > \frac{s_n}{2}\left(\frac{\Psi^2_{\widehat{j}'}}{4c_3^2}\frac{n}{\log n}\right)^{-1/d} \quad &\ge \quad c_1 s_n \left(s_n^{\frac{2d\alpha}{d+2\alpha}}\left(\frac{n}{\log n}\right)^{-\frac{2\alpha}{d+2\alpha}}\frac{n}{\log n}\right)^{-1/d} \\
&= \quad c_1 s_n s_n^{\frac{-2\alpha}{d+2\alpha}}\left(\frac{n}{\log n}\right)^{\frac{-1}{d+2\alpha}} \\
&= \quad c_1 s_n^{\frac{d}{d+2\alpha}}\left(\frac{n}{\log n}\right)^{\frac{-1}{d+2\alpha}},
\end{aligned}$$

where $c_1 \equiv c_1(C_2, f_{\max}, \delta_0, d, \alpha) > 0$. So we have for $s_n > 4C_4 6^d$, with probability at least $1 - 5/n$,

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

where $c_1 \equiv c_1(C_2, f_{\max}, \delta_0, d, \alpha) > 0$ and $c_2 \equiv c_2(C_1, C_2, f_{\max}, \delta_0, d, \alpha) > 0$. Hence the automatically chosen resolution behaves as desired.

Let us now derive the claimed bounds on $\mathcal{V}_{\gamma,\widehat{j}}$ and $\Psi_{\widehat{j}'}$. Using Lemma 4 and Eq. (5), we have the following oracle inequality:

$$\mathcal{V}_{\gamma,\widehat{j}} \leq \widehat{\mathcal{V}}_{\gamma,\widehat{j}} + \Psi_{\widehat{j}'} = \min_{0\leq j\leq J}\left\{\widehat{\mathcal{V}}_{\gamma,j} + \Psi_{j'}\right\} \leq \min_{0\leq j\leq J}\left\{\mathcal{V}_{\gamma,j} + 2\Psi_{j'}\right\}$$

Lemma 5 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.

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

Here $C \equiv C(C_2, f_{\max}, \delta_0, d, \alpha)$. The second step uses the definition of j', and the last step follows by balancing the two terms for optimal resolution j^* given by $2^{-j^*} \asymp s_n^{\frac{d}{d+2\alpha}}\left(\frac{n}{\log n}\right)^{-\frac{1}{d+2\alpha}}$. This establishes the desired bounds on $\mathcal{V}_{\gamma,\widehat{j}}$ and $\Psi_{\widehat{j}'}$.

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 $n \geq n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$. Pick $n \geq \max(n_0, n_1)$ and let Ω denote the event such that the bounds of Lemma 2, Lemma 3 and Lemma 4 hold with $\delta = 1/n$. Then, we have $P(\bar{\Omega}) \leq 5/n$. For $n < \max(n_0, n_1)$, we can use the trivial inequality $P(\bar{\Omega}) \leq 1$. So we have, for all n

$$P(\bar{\Omega}) \leq \max(5, \max(n_0, n_1))\frac{1}{n} := C\frac{1}{n}$$

Here $C \equiv C(C_1, C_4, \epsilon_o, f_{\max}, \delta_0, d, \alpha)$.

So $\forall f \in \mathcal{F}_2^*(\alpha)$, we have: (Here C may denote a different constant from line to line.)

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

$$\leq C \max\left\{ s_n^{\frac{-d^2/2\alpha}{d+2\alpha}} \left(\frac{n}{\log n}\right)^{-\frac{1}{d+2\alpha}}, s_n s_n^{\frac{d}{d+2\alpha}} \left(\frac{n}{\log n}\right)^{-\frac{1}{d+2\alpha}}, \frac{1}{n} \right\}$$
$$\leq C s_n s_n^{\frac{d}{d+2\alpha}} \left(\frac{n}{\log n}\right)^{-\frac{1}{d+2\alpha}}$$
$$\leq C s_n^2 \left(\frac{n}{\log n}\right)^{-\frac{1}{d+2\alpha}}$$

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

□

7.4 Proof of Proposition 1

Notice that since $\mathcal{F}^*_2(\alpha) \subset \mathcal{F}^*_1(\alpha)$, we have

$$\inf_{\widehat{G}_n} \sup_{f \in \mathcal{F}^*_1(\alpha)} \mathbb{E}[d_\infty(\widehat{G}_n, G^*_\gamma)] \geq \inf_{\widehat{G}_n} \sup_{f \in \mathcal{F}^*_2(\alpha)} \mathbb{E}[d_\infty(\widehat{G}_n, G^*_\gamma)]$$

Therefore, it suffices to establish a lower bound for the class of densities given by $\mathcal{F}^*_2(\alpha)$.

We consider the class of densities $\mathcal{F}_{SL}$ with star-shaped levels sets having Lipschitz boundaries, as defined in [14]. We establish that $\mathcal{F}_{SL}$ (re-defined so that densities have domain $[0,1]^d$) is a subset of $\mathcal{F}^*_2(\alpha)$, i.e. densities in $\mathcal{F}_{SL}$ satisfy assumptions **[A,B,C]**. If this holds, then

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

The last step follows from Theorem 4 in [14]. Thus the result of the proposition holds.

We proceed by recalling the definition of $\mathcal{F}_{SL}$ as defined in [14]. The class corresponds to densities bounded above by $f_{\max}$, satisfying the local density regularity assumption **[A]**, and with γ level sets of the form

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

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

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

Here $L > 0$ is the Lipschitz constant, and $||\cdot||_1$ denotes the ℓ_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^*_γ of a density $f \in \mathcal{F}_{SL}$ satisfies **[B]** and **[C]**; assumption **[A]** is satisfied as per the definition of $\mathcal{F}_{SL}$. The same result holds for star-shaped sets defined on the shifted domain $[0, 1]^d$, and hence $\mathcal{F}_{SL}$ (defined with $R = 1/2$ and origin shifted to the center of $[0, 1]^d$) is a subset of $\mathcal{F}^*_2(\alpha)$.

Lemma 6. *Consider the γ level set G^*_γ of a density $f \in \mathcal{F}_{SL}$. Then G^*_γ satisfies the assumptions **[B]** and **[C]** on the level set regularity and the level set boundary dimension, respectively.*

The proof is given in the Appendix.

□

Appendix

Here we present the proofs of various lemmas used in Section 7.

Proof of Lemma 1: The proof relies on the following pair of VC inequalities (See [26] Chapter 3) that bound the *relative* deviation of true and empirical probabilities.

Let $Y_1, \ldots, Y_n$ be iid random variables taking values in the $\mathbb{R}^d$ with common distribution

$$P(B) = P\{Y_1 \in B\} \qquad (B \subset \mathbb{R}^d).$$

Define the empirical distribution

$$\widehat{P}(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{Y_i \in B\}} \qquad (B \subset \mathbb{R}^d).$$

Consider a class $\mathcal{B}$ of subsets of $\mathbb{R}^d$. Then for any $\epsilon > 0$,

$$P\left(\sup_{B \in \mathcal{B}} \frac{P(B) - \widehat{P}(B)}{\sqrt{P(B)}} > \epsilon\right) \leq 4\mathbb{S}_{\mathcal{B}}(2n) e^{-n\epsilon^2/4}$$

and

$$P\left(\sup_{B \in \mathcal{B}} \frac{\widehat{P}(B) - P(B)}{\sqrt{\widehat{P}(B)}} > \epsilon\right) \leq 4\mathbb{S}_{\mathcal{B}}(2n) e^{-n\epsilon^2/4}.$$

Here $\mathbb{S}_{\mathcal{B}}(n)$ is the n^{th} VC shatter-coefficient of $\mathcal{B}$, defined as

$$\mathbb{S}_{\mathcal{B}}(n) = \max_{z_1, \ldots, z_n \in \mathbb{R}^d} |\{\{z_1, \ldots, z_n\} \cap B : B \in \mathcal{B}\}|.$$

In words, $\mathbb{S}_{\mathcal{B}}(n)$ is the maximal number of different subsets of a set of n points which can be obtained by intersecting it with the elements of $\mathcal{B}$.

To establish bounds on the deviation of true and empirical average density, first observe that

$$\widehat{P}(A) \le P(A) + \epsilon\sqrt{\widehat{P}(A)} \Longrightarrow \widehat{P}(A) \le 2\max(P(A), 2\epsilon^2) \tag{6}$$

and

$$P(A) \le \widehat{P}(A) + \epsilon\sqrt{P(A)} \Longrightarrow P(A) \le 2\max(\widehat{P}(A), 2\epsilon^2). \tag{7}$$

To see the first statement, consider two cases:

1) $\widehat{P}(A) \le 4\epsilon^2$. The statement is obvious.

2) $\widehat{P}(A) > 4\epsilon^2$. This gives a bound on epsilon, which implies

$$\widehat{P}(A) \le P(A) + \widehat{P}(A)/2 \Longrightarrow \widehat{P}(A) \le 2P(A).$$

The second statement follows similarly.

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

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

and

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

In other words, with probability $> 1 - 8 \cdot 2^{jd} e^{-n\epsilon^2/4}$, $\forall A \in \mathcal{A}_j$

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

Define $\delta_j = 8 \cdot 2^{jd} e^{-n\epsilon^2/4}$. Then we have, with probability $> 1 - \delta_j$, $\forall A \in \mathcal{A}_j$

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

Setting $\delta_j = \delta 2^{-(j+1)}$ and taking union bound, we have with probability $> 1-\delta$, for all resolutions j and all cells $A \in \mathcal{A}_j$

$$|P(A) - \widehat{P}(A)| \le \sqrt{8\frac{\log(2^{j(d+1)}16/\delta)}{n}\max\left(\widehat{P}(A), 8\frac{\log(2^{j(d+1)}16/\delta)}{n}\right)}$$

The result follows after dividing both sides by $\mu(A)$.

□

Proof of Lemma 2: Recall the definition of Φ_j

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

We first derive a lower bound. Observe that since the total empirical probability mass is 1, we have:

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

Use this along with $\delta = 1/n$, $j \geq 0$ and $\mu(A) = 2^{-jd}$ to get:

$$\Phi_j \geq \sqrt{2^{jd} 8 \frac{\log 16n}{n}}$$

To get an upper bound, using Eq. (6) from the proof of Lemma 1 and dividing by $\mu(A) = 2^{-jd}$ we have with probability $> 1 - 8 \cdot 2^{jd} e^{-n\epsilon^2/4}$, for all $A \in \mathcal{A}_j$

$$\widehat{f}(A) \leq 2\max(\bar{f}(A), 2^{jd+1}\epsilon^2) \leq 2\max(f_{\max}, 2^{jd+1}\epsilon^2).$$

Define $\delta_j = 8 \cdot 2^{jd} e^{-n\epsilon^2/4}$. Then we have with probability $> 1 - \delta_j$, for all $A \in \mathcal{A}_j$

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

Setting $\delta_j = \delta 2^{-(j+1)}$ and taking union bound, we have with probability $> 1 - \delta$, for all resolutions j and all cells $A \in \mathcal{A}_j$

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

This implies

$$\Phi_j \leq \sqrt{2^{jd} 8 \frac{\log(2^{j(d+1)} 16/\delta)}{n} \cdot 2\max\left(f_{\max}, 2^{jd} 8 \frac{\log(2^{j(d+1)} 16/\delta)}{n}\right)}$$

Using $\delta = 1/n$ and $2^j = O((n/\log n)^{1/d})$, we get:

$$\Phi_j \leq c_4(f_{\max}, d) \sqrt{2^{jd} \frac{\log n}{n}}$$

□

Proof of Lemma 3: Let $J_0 = \lceil \log_2 4\sqrt{d}/\epsilon_o \rceil$, where ϵ_o is as defined in assumption **[B]**. Also define

$$\epsilon_j := \left\lceil \left(\frac{\Phi_j}{C_1}\right)^{1/\alpha} + \sqrt{d} 2^{-j} \right\rceil.$$

Consider two cases:

I. $j < J_0$.

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

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

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

II. $j \geq J_0$.

For this case, we argue that for large enough n, with high probability, $\widehat{G}_j \cap G^*_\gamma \neq \emptyset$ and hence $\widehat{G}_j$ is not empty. Also, observe that assumption **[B]** implies that G^*_γ is not empty since $G^*_\gamma \supseteq \mathcal{I}_\epsilon(G^*_\gamma) \neq \emptyset$ for $\epsilon \leq \epsilon_o$. Thus the Hausdorff error is given as

$$d_\infty(\widehat{G}_j, G^*_\gamma) = \max\{\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}_j), \sup_{x \in \widehat{G}_j} \rho(x, G^*_\gamma)\}, \tag{8}$$

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 ϵ_j. Notice that, if $\widehat{G}_j$ is non-empty, this provides an upper bound on the second term of the Hausdorff error (Eq. 8). Building on this proposition, the second one establishes that, for large enough n and $j \geq J_0$, with high probability, every ball in the inner cover $\mathcal{I}_{2\epsilon_j}(G^*_\gamma)$ contains points that are correctly included in the level set estimate, and hence lie in $\widehat{G}_j \cap G^*_\gamma$. Thus $\widehat{G}_j$ is not empty. And along with assumption **[B]**, this provides a bound on the distance of any point in G^*_γ from the estimate $\widehat{G}_j$, thus bounding the first term of the Hausdorff error (Eq. 8).

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 $1-2/n$*

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

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) $A_x \cap \partial G^*_\gamma = \emptyset$. This implies that $A_x \subseteq \widehat{G}_j \Delta G^*_\gamma$ and hence it is erroneously included or excluded from the level set estimate $\widehat{G}_j$. Notice that A_x is included in the true set G^*_γ if $\bar{f}(A_x) > \gamma$ and excluded if $\bar{f}(A_x) < \gamma$. Since $\widehat{G}_j$ includes all cells with $\widehat{f}(A) > \gamma$ and excludes all cells with $\widehat{f}(A) < \gamma$, this implies that $|\gamma - \bar{f}(A_x)| \leq |\bar{f}(A_x) - \widehat{f}(A_x)|$. Using Lemma 1, we get $|\gamma - \bar{f}(A_x)| \leq \Phi_j$ 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, $|\gamma - \bar{f}(A_x)| \leq \Phi_j$ with probability at least $1 - 1/n$ (for $\delta = 1/n$) and using Lemma 2, Φ_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 $n \geq n_1(f_{\max}, d, \delta_0)$, $\Phi_j \leq \delta_0$ 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

$$\rho(x_0, \partial G^*_\gamma) \leq \left(\frac{|\gamma - f(x_0)|}{C_1}\right)^{1/\alpha} \leq \left(\frac{|\gamma - \bar{f}(A_x)|}{C_1}\right)^{1/\alpha} \leq \left(\frac{\Phi_j}{C_1}\right)^{1/\alpha}.$$

Since $x, x_0 \in A_x$,

$$\rho(x, \partial G^*_\gamma) \leq \rho(x_0, \partial G^*_\gamma) + \sqrt{d}2^{-j} \leq \left(\frac{\Phi_j}{C_1}\right)^{1/\alpha} + \sqrt{d}2^{-j}.$$

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 $1 - 2/n$ $\forall x \in \widehat{G}_j \Delta G^*_\gamma$

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

□

Proposition 3. *Recall assumption [**B**] and denote the inner cover of G^*_γ with $2\epsilon_j$-balls, $\mathcal{I}_{2\epsilon_j}(G^*_\gamma) \equiv \mathcal{I}_{2\epsilon_j}$ for simplicity. 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 $1 - 3/n$, $\mathcal{I}_{2\epsilon_j} \neq \emptyset$ and for every $2\epsilon_j$-ball in $\mathcal{I}_{2\epsilon_j}$, all points in the interior of the concentric ϵ_j-ball are in $\widehat{G}_j \cap G^*_\gamma$.*

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 $n \geq n_2(\epsilon_o, f_{\max}, C_1, \alpha)$, $2(\Phi_j/C_1)^{1/\alpha} \leq \epsilon_o/2$ 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 ϵ_j-ball from the boundary of $\mathcal{I}_{2\epsilon_j}$, and hence from the boundary of G^*_γ is greater than ϵ_j. As per Proposition 1 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 $\mathcal{I}_{2\epsilon_j} \subseteq G^*_\gamma$. $\square$

Now since G^*_γ and $\widehat{G}_j$ are non-empty sets, we use Propositions 2, 3 to bound the two terms that contribute to the Hausdorff error:

$$\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}_j) \qquad \text{and} \qquad \sup_{x \in \widehat{G}_j} \rho(x, G^*_\gamma)$$

The following statements 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

$$\sup_{x \in \widehat{G}_j} \rho(x, G^*_\gamma) = \sup_{x \in \widehat{G}_j \setminus G^*_\gamma} \rho(x, \partial G^*_\gamma) \quad \leq \quad \sup_{x \in \widehat{G}_j \Delta G^*_\gamma} \rho(x, \partial G^*_\gamma) \leq \epsilon_j.$$

Thus, for either case

$$\sup_{x \in \widehat{G}_j} \rho(x, G^*_\gamma) \leq \epsilon_j. \tag{9}$$

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:

$$\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}_j) \leq \sup_{x \in G^*_\gamma} \rho(x, \widehat{G}_j \cap G^*_\gamma) \quad = \quad \sup_{x \in \partial G^*_\gamma} \rho(x, \widehat{G}_j \cap G^*_\gamma).$$

Now consider any $x \in \partial G^*_\gamma$. Then

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

The first step follows using triangle inequality since $\forall y \in \widehat{G}_j \cap G^*_\gamma$, $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$, which implies $\inf_{y \in \widehat{G}_j \cap G^*_\gamma} \rho(x, y) \leq \rho(x, z) + \inf_{y \in \widehat{G}_j \cap G^*_\gamma} \rho(z, y)$. The second step follows by considering a $2\epsilon_j$-ball in $\mathcal{I}_{2\epsilon_j}$ that contains z and using Proposition 3. Thus we get

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

Here the second step invokes assumption [**B**].

Therefore, for either case we have

$$\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}_j) \leq (2C_3 + 1)\epsilon_j. \qquad (10)$$

From Eq. (9) and (10), 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 $> 1 - 3/n$,

$$d_\infty(\widehat{G}_j, G^*_\gamma) \;=\; \max\{\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}_j), \sup_{x \in \widehat{G}_j} \rho(x, G^*_\gamma)\} \leq (2C_3 + 1)\epsilon_j.$$

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

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

□

Proof of Lemma 4: 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, and let A''_0 and A''_1 denote the subcells at resolution j' within A_0 and A_1, respectively, that have minimum average density. 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, and let $\widehat{A}''_0$ and $\widehat{A}''_1$ denote the subcells at resolution j' within A_0 and A_1, respectively, that have minimum empirical density. Then we have:

$$\begin{aligned}
\mathcal{V}_{\gamma,j} - \widehat{\mathcal{V}}_{\gamma,j} &= |\gamma - \bar{f}(A'_0)| + |\gamma - \bar{f}(A''_0)| - |\gamma - \widehat{f}(\widehat{A}'_1)| - |\gamma - \widehat{f}(\widehat{A}''_1)| \\
&\leq |\gamma - \bar{f}(A'_1)| + |\gamma - \bar{f}(A''_1)| - |\gamma - \widehat{f}(\widehat{A}'_1)| - |\gamma - \widehat{f}(\widehat{A}''_1)| \\
&\leq |\bar{f}(A'_1) - \widehat{f}(\widehat{A}'_1)| + |\bar{f}(A''_1) - \widehat{f}(\widehat{A}''_1)| \\
&\leq \max\{\bar{f}(A'_1) - \widehat{f}(\widehat{A}'_1), \widehat{f}(\widehat{A}'_1) - \bar{f}(A'_1)\} \\
&\qquad + \max\{\bar{f}(A''_1) - \widehat{f}(\widehat{A}''_1), \widehat{f}(\widehat{A}''_1) - \bar{f}(A''_1)\} \\
&\leq \max\{\bar{f}(A'_1) - \widehat{f}(A'_1), \widehat{f}(\widehat{A}'_1) - \bar{f}(\widehat{A}'_1)\} \\
&\qquad + \max\{\bar{f}(\widehat{A}''_1) - \widehat{f}(\widehat{A}''_1), \widehat{f}(A''_1) - \bar{f}(A''_1)\} \\
&\leq 2 \max_{A \in \mathcal{A}_{j'}} |\bar{f}(A) - \widehat{f}(A)| \\
&\leq \Psi_{j'}
\end{aligned}$$

The first inequality invokes definition of A_0, the fourth inequality invokes definitions of the subcells A'_1, A''_1, $\widehat{A}'_1$ and $\widehat{A}''_1$, and the last one follows from the uniform density deviation control established in Lemma 1. Similarly,

$$\begin{aligned}\widehat{\mathcal{V}}_{\gamma,j} - \mathcal{V}_{\gamma,j} &= |\gamma - \widehat{f}(\widehat{A}_1')| + |\gamma - \widehat{f}(\widehat{A}_1'')| - |\gamma - \bar{f}(A_0')| - |\gamma - \bar{f}(A_0'')| \\ &\leq |\gamma - \widehat{f}(\widehat{A}_0')| + |\gamma - \widehat{f}(\widehat{A}_0'')| - |\gamma - \bar{f}(A_0')| - |\gamma - \bar{f}(A_0'')| \\ &\leq |\bar{f}(A_0') - \widehat{f}(\widehat{A}_0')| + |\bar{f}(A_0'') - \widehat{f}(\widehat{A}_0'')|\end{aligned}$$

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

□

Proof of Lemma 5: We first establish the upper bound. Consider a cell $A \in \mathcal{A}_j$ s.t. $A \cap \partial G_\gamma^* \neq \emptyset$. Let A' and A'' denote subcells at resolution j' within A that have maximum and minimum average density respectively. Consider two cases:

(i) If $(\sqrt{d}2^{-j})^\alpha \leq \delta_0/C_2$, then regularity holds $\forall x \in A$ since $|f(x) - \gamma| \leq C_2(\sqrt{d}2^{-j})^\alpha \leq \delta_0$. The same holds also for subcells A' and A''. Hence

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

(ii) If $(\sqrt{d}2^{-j})^\alpha > \delta_0/C_2$, the following trivial bound holds:

$$|\gamma - \bar{f}(A')| + |\gamma - \bar{f}(A'')| \leq 2f_{\max}$$

Notice that the bound for the first case $2C_2(\sqrt{d}2^{-j})^\alpha \leq \frac{2C_2 f_{\max}}{\delta_0}(\sqrt{d}2^{-j})^\alpha$ since $\delta_0 \leq f_{\max}$. And the bound for the second case, $2f_{\max} \leq \frac{2C_2 f_{\max}}{\delta_0}(\sqrt{d}2^{-j})^\alpha$ since $(\sqrt{d}2^{-j})^\alpha > \delta_0/C_2$. Hence we can say for all j there exists $A \in \mathcal{A}_j$ such that

$$|\gamma - \bar{f}(A')| + |\gamma - \bar{f}(A'')| \leq \frac{2C_2 f_{\max}}{\delta_0}(\sqrt{d}2^{-j})^\alpha$$

This yields the upper bound on the vernier:

$$V_{\gamma,j} \leq \frac{2C_2 f_{\max}}{\delta_0}(\sqrt{d}2^{-j})^\alpha := C(\sqrt{d}2^{-j})^\alpha$$

where $C \equiv C(C_2, f_{\max}, \delta_0)$.

For the lower bound, consider a cell $A \in \mathcal{A}_j$. Note 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 $A_0' \in A \cap \mathcal{A}_{j'}$ such that $\forall x \in A_0'$,

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

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 $x \in A'_0$, $|\gamma - f(x)| \geq \min(\delta_0, C_1 2^{-j'\alpha}) \geq \min(\delta_0, C_1) 2^{-j'\alpha}$. So we have:

$$|\gamma - \bar{f}(A'_0)| \geq \min(\delta_0, C_1) 2^{-j'\alpha}$$

Let A' and A'' denote the subcells within A that have maximum and minimum average density respectively. Then we have $\bar{f}(A') \geq \bar{f}(A'_0) \geq \bar{f}(A'')$. Therefore, we get

$$\begin{aligned} |\gamma - \bar{f}(A')| + |\gamma - \bar{f}(A'')| &\geq \max\{|\gamma - \bar{f}(A')|, |\gamma - \bar{f}(A'')|\} \\ &\geq |\gamma - \bar{f}(A'_0)| \\ &\geq \min(\delta_0, C_1) 2^{-j'\alpha} \end{aligned}$$

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 $s_n > 4C_4 6^d$), $\exists A'_0 \in A \cap \mathcal{A}_{j'}$ s.t. $\forall x \in A'_0$,

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

Observe that it suffices to show that for large enough n, $\exists A'_1 \in A \cap \mathcal{A}_{j'-2}$ s.t. $A'_1 \cap \partial G^*_\gamma = \emptyset$. To prove this, 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 $\exists A'_1 \in A \cap \mathcal{A}_{j'-2}$ s.t. $A'_1 \cap \partial G^*_\gamma = \emptyset$.

(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 $\leq C_4 (\delta/\epsilon)^{-(d-1)}$. Consider a uniform partition of the cell A into subcells of sidelength $2\delta/\sqrt{d} = 2^{-(j'-2)}$. With the choice of δ, this implies that a subcell at resolution $2^{-(j'-2)}$ is inscribed within an aligned δ-ball. Observe that at this resolution, in d-dim, an unaligned δ-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

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

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, $\exists A'_1 \in A \cap \mathcal{A}_{j'-2}$ s.t. $A'_1 \cap \partial G^*_\gamma = \emptyset$.

This in turn implies that for n large enough (so that $s_n > 4C_4 6^d$), $\exists A'_0 \in A \cap \mathcal{A}_{j'}$ such that $\forall x \in A'_0$,

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

□

Proof of Lemma 6: We first present a sketch of the main ideas, and then provide a detailed proof. Consider the γ-level set G^*_γ of a density $f \in \mathcal{F}_{SL}$. To see that it satisfies **[B]**, divide the star-shaped set G^*_γ into sectors of width $\asymp \epsilon$ so that each sector contains at least one ϵ-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^*_γ 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 δ-balls can cover the boundary in each sub-sector. Thus, the minimum number of δ-balls needed to cover the boundary in all sub-sectors is of the order of the minimum number of sub-sectors, i.e. $O((\epsilon/\delta)^{d-1})$. Hence, the result follows. We now present the proof in detail.

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

$$S_{\boldsymbol{m}} = \Big\{ (r, \boldsymbol{\phi}) : 0 \leq r \leq g(\boldsymbol{\phi}), \frac{2\pi(m_{d-1}-1)}{M} \leq \phi_{d-1} < \frac{2\pi m_{d-1}}{M} \\ \frac{\pi(m_i - 1)}{M} \leq \phi_i < \frac{\pi m_i}{M} \qquad i = 1, \ldots, d-2 \Big\},$$

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

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

This choice of M implies that:

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

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

The second inequality follows since

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

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

$$\rho(y, \mathcal{I}_\epsilon(G^*_\gamma)) \leq \rho(y, x_{\boldsymbol{m}}) = ||y - x_{\boldsymbol{m}}||$$

Now recall that if $y = (y_1, \ldots, y_d) \equiv (r, \phi_1, \ldots, \phi_{d-1}) = (g(\boldsymbol{\phi}), \boldsymbol{\phi})$, then the relation between the Cartesian and hypershperical coordinates is given as:

$$\begin{aligned}y_1 &= r\cos\phi_1 \\ y_2 &= r\sin\phi_1\cos\phi_2 \\ y_3 &= r\sin\phi_1\sin\phi_2\cos\phi_3 \\ &\vdots \\ y_{d-1} &= r\sin\phi_1\ldots\sin\phi_{d-2}\cos\phi_{d-1} \\ y_d &= r\sin\phi_1\ldots\sin\phi_{d-2}\sin\phi_{d-1}\end{aligned}$$

Now since $||y-x|| = \sum_{i=1}^d (y_i - x_i)^2$, using the above transformation and simple algebra, we can show that:

$$\begin{aligned}||y - x_{\boldsymbol{m}}||^2 &= ||(g(\boldsymbol{\phi}), \boldsymbol{\phi}) - (g(\boldsymbol{\theta}), \boldsymbol{\theta})||^2 \\ &= (g(\boldsymbol{\phi}) - g(\boldsymbol{\theta}))^2 \\ &\quad + 4g(\boldsymbol{\phi})g(\boldsymbol{\theta})\sum_{i=1}^{d-1}\sin\phi_1\ldots\sin\phi_{i-1}\sin\theta_1\ldots\sin\theta_{i-1}\sin^2\frac{\phi_i - \theta_i}{2} \\ &\leq (g(\boldsymbol{\phi}) - g(\boldsymbol{\theta}))^2 + 4g(\boldsymbol{\phi})g(\boldsymbol{\theta})\sum_{i=1}^{d-1}\sin^2\frac{\phi_i - \theta_i}{2}\end{aligned}$$

Using this, we have $\forall y \in \partial G^*_\gamma \cap S_{\boldsymbol{m}}$

$$\begin{aligned}
\rho(y, \mathcal{I}_\epsilon(G^*_\gamma)) &\leq \sqrt{(g(\boldsymbol{\phi}) - g(\boldsymbol{\theta}))^2 + 4g(\boldsymbol{\phi})g(\boldsymbol{\theta})\sum_{i=1}^{d-1}\sin^2\frac{\phi_i - \theta_i}{2}} \\
&\leq |g(\boldsymbol{\phi}) - g(\boldsymbol{\theta})| + 2\sqrt{g(\boldsymbol{\phi})g(\boldsymbol{\theta})}\sum_{i=1}^{d-1}\left|\sin\frac{\phi_i - \theta_i}{2}\right| \\
&\leq L||\boldsymbol{\phi} - \boldsymbol{\theta}||_1 + \sum_{i=1}^{d-1}\frac{|\phi_i - \theta_i|}{2} \\
&= (L + 1/2)\sum_{i=1}^{d-1}|\phi_i - \theta_i| \\
&\leq (L + 1/2)d\frac{\pi}{M} \\
&\leq \frac{9d(L + 1/2)}{h}\epsilon := C_3\epsilon
\end{aligned}$$

where the third step follows by using the Lipschitz condition on $g(\cdot)$, $g(\cdot) \leq R = 1/2$ and since $|\sin(z)| \leq |z|$. The fifth step follows since $x, y \in S_{\boldsymbol{m}}$ and hence $|\phi_i - \theta_i| \leq \pi/M$ for $i = 1, \ldots, d-2$ and $|\phi_{d-1} - \theta_{d-1}| \leq 2\pi/M$. The sixth step invokes (ii) above.

Therefore, we have for all $y \in \partial G^*_\gamma \cap S_{\boldsymbol{m}}$ $\rho(y, \mathcal{I}_\epsilon(G^*_\gamma)) \leq C_3\epsilon$. And since the result is true for any sector, condition **[B]** is satisfied by any level set G^*_γ with density $f \in \mathcal{F}_{SL}$.

To see that G^*_γ satisfies **[C]**, consider $x \in \partial G^*_\gamma$. Let $x = (g(\boldsymbol{\phi}_0), \boldsymbol{\phi}_0)$. Also let $\phi_i^{(1)} = \min\{\phi_i : (g(\boldsymbol{\phi}), \boldsymbol{\phi}) \in B(x, \epsilon)\}$ and $\phi_i^{(2)} = \max\{\phi_i : (g(\boldsymbol{\phi}), \boldsymbol{\phi}) \in B(x, \epsilon)\}$. Define the sector

$$S^x_\epsilon = \left\{(r, \boldsymbol{\phi}) : 0 \leq r \leq g(\boldsymbol{\phi}), \phi_i^{(1)} \leq \phi_i \leq \phi_i^{(2)} \quad \forall i = 1, \ldots, d-1\right\}$$

Observe that the width of S^x_ϵ in the i^{th} coordinate, $\Delta\phi_i = \phi_i^{(2)} - \phi_i^{(1)} \leq 2\sin^{-1}\frac{\epsilon}{g(\boldsymbol{\phi}_0)}$ by construction and 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 $0 \leq z \leq \pi/2$, $\sin^{-1}(z/2) \leq z$. Further subdivide S^x_ϵ into M^{d-1} sub-sectors indexed by $\boldsymbol{m} = (m_1, \ldots, m_{d-1})$

$$\begin{aligned}
S_{\boldsymbol{m}} = \Big\{(r, \boldsymbol{\phi}) : 0 \leq r \leq g(\boldsymbol{\phi}),& \\
\phi_i^{(1)} + \frac{(m_i - 1)\Delta\phi_i}{M} \leq \phi_i < \phi_i^{(1)} + \frac{m_i\Delta\phi_i}{M} &\quad \forall i = 1, \ldots, d-1\Big\}
\end{aligned}$$

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, $S_{\boldsymbol{m}} \cap \partial G^*_\gamma$ can be covered by a δ-ball. To see this, consider two points in $S_{\boldsymbol{m}} \cap \partial G^*_\gamma$ - $(g(\boldsymbol{\phi}), \boldsymbol{\phi})$ and $(g(\boldsymbol{\theta}), \boldsymbol{\theta})$. Proceeding as before, we have:

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

Since each sub-sector can be covered by a δ-ball, the minimum number of δ-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 δ-balls needed to cover $B(x, \epsilon) \cap \partial G^*_\gamma$ is equal to

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

The second inequality follows since $\Delta\phi_i \leq 4\frac{\epsilon}{h}$ for all i, and since $\delta \leq \epsilon$. Therefore, any level set G^*_γ with density $f \in \mathcal{F}_{SL}$ also satisfies **[C]**.

□

Appendix A

Here we sketch a proof of the claim that $d_\infty(\widehat{G}, G^*_\gamma) \leq C(\mathcal{E}(\widehat{G}))^{1/\alpha}$, where $\widehat{G}$ is the plug-in level set estimator obtained using a sup-norm consistent density estimate $\widehat{f}(x)$. Observe that assumption **[B]** implies that G^*_γ is not empty since $G^*_\gamma \supseteq \mathcal{I}_\epsilon(G^*_\gamma) \neq \emptyset$ for $\epsilon \leq \epsilon_o$. Hence, for large enough n, the plug-in level set estimate $\widehat{G}$ is also non-empty since $\widehat{f}(x)$ is consistent in the sup-norm. Now recall that for non-empty sets

$$d_\infty(\widehat{G}, G^*_\gamma) = \max\{ \sup_{x \in G^*_\gamma} \rho(x, \widehat{G}), \sup_{x \in \widehat{G}} \rho(x, G^*_\gamma) \}.$$

We now derive upper bounds on the two terms that control the Hausdorff error.

First, observe that for all incorrectly classified points $x \in \widehat{G}_j \Delta G^*_\gamma$, $|f(x) - \gamma| \leq |f(x) - \widehat{f}(x)|$, hence regularity condition **[A]** holds at x since for large enough n, $|f(x) - \widehat{f}(x)| \leq \delta_0$. So we have:

$$\sup_{x \in \widehat{G}\Delta G^*_\gamma} \rho(x, \partial G^*_\gamma) \leq \sup_{x \in \widehat{G}\Delta G^*_\gamma} \left(\frac{|f(x)-\gamma|}{C_1}\right)^{1/\alpha} \leq \left(\frac{\mathcal{E}(\widehat{G})}{C_1}\right)^{1/\alpha} := \epsilon. \tag{11}$$

The last inequality follows since $\forall x \in \widehat{G}\Delta G^*_\gamma$, $|f(x) - \gamma| \leq \mathcal{E}(\widehat{G})$. Equation (11) gives a bound on the second term of the Hausdorff error:

$$\sup_{x \in \widehat{G}} \rho(x, G^*_\gamma) = \sup_{x \in \widehat{G}\setminus G^*_\gamma} \rho(x, \partial G^*_\gamma) \leq \sup_{x \in \widehat{G}\Delta G^*_\gamma} \rho(x, \partial G^*_\gamma) \leq \left(\frac{\mathcal{E}(\widehat{G})}{C_1}\right)^{1/\alpha}.$$

To bound the first term of the Hausdorff error, we proceed as follows:

$$\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}) \leq \sup_{x \in G^*_\gamma} \rho(x, \widehat{G} \cap G^*_\gamma) = \sup_{x \in \partial G^*_\gamma} \rho(x, \widehat{G} \cap G^*_\gamma).$$

Now consider any $x \in \partial G^*_\gamma$. Then

$$\begin{aligned} \rho(x, \widehat{G} \cap G^*_\gamma) &\leq \rho(x, z) + \rho(z, \widehat{G} \cap G^*_\gamma) && \forall z \in \mathcal{I}_{2\epsilon}(G^*_\gamma) \\ &\leq \rho(x, z) + \epsilon && \forall z \in \mathcal{I}_{2\epsilon}(G^*_\gamma). \end{aligned}$$

The first step follows using triangle inequality since $\forall y \in \widehat{G}_j \cap G^*_\gamma$, $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$, which implies $\inf_{y \in \widehat{G}_j \cap G^*_\gamma} \rho(x, y) \leq \rho(x, z) + \inf_{y \in \widehat{G}_j \cap G^*_\gamma} \rho(z, y)$. The second step follows since for all 2ϵ-balls in $\mathcal{I}_{2\epsilon}(G^*_\gamma)$, all points in the interior of the concentric ϵ-balls are greater than ϵ away from the boundary of the true set ∂G^*_γ, and hence cannot lie in $\widehat{G}\Delta G^*_\gamma$ as per Eq. (11). In fact these points must lie in $\widehat{G} \cap G^*_\gamma$ since they lie in $\mathcal{I}_{2\epsilon}(G^*_\gamma) \subseteq G^*_\gamma$. So we have

$$\rho(x, \widehat{G} \cap G^*_\gamma) \quad \leq \quad \rho(x, \mathcal{I}_{2\epsilon}(G^*_\gamma)) + \epsilon \leq 2C_3\epsilon + \epsilon,$$

where the second step invokes assumption **[B]**. So we have the following bound on the first term of the Hausdorff error:

$$\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}) \leq (2C_3 + 1)\epsilon = (2C_3 + 1)\left(\frac{\mathcal{E}(\widehat{G})}{C_1}\right)^{1/\alpha}.$$

Putting together the bounds for both terms in the Hausdorff error, we get:

$$d_\infty(\widehat{G}, G^*_\gamma) = \max\{\sup_{x \in G^*_\gamma} \rho(x, \widehat{G}), \sup_{x \in \widehat{G}} \rho(x, G^*_\gamma)\} \leq (2C_3 + 1)\left(\frac{\mathcal{E}(\widehat{G})}{C_1}\right)^{1/\alpha}.$$

This concludes the proof.

Appendix B

Adaptivity for $\alpha \geq 0$: We prove that adaptivity can be achieved, and hence Theorem 2 holds, for the whole range $\alpha \geq 0$ using the modified vernier and penalty proposed in Section 6.1. But before that, we establish that Theorem 1 holds for the jump case i.e. when α is known to be zero, the empirical error minimization procedure of Eq. (3), along with choice of the resolution as $2^{-j} \asymp s_n(n/\log n)^{-1/d}$ achieves the minimax rate of Hausdorff error convergence for the class of densities given by $\mathcal{F}_1^*(0)$. Notice that for Theorem 1 to hold no change is required in the estimation procedure, however the proof technique presented earlier does not work for the jump case ($\alpha = 0$). So, here we provide a sketch of the arguments needed to show that Theorem 1 holds for the jump case as well.

For the jump case, Theorem 1 basically follow similar to the non-jump case except that Lemma 3 is replaced by the following:

Lemma 7. *Consider densities satisfying assumptions [**A**] and [**B**]. Let $\widehat{G}_j$ denote the set at resolution j that minimizes the empirical error $\widehat{\mathcal{E}}_\gamma$ as per Eq. (3). Then for all resolutions $j \equiv j(n)$ such that $2^j = O(s_n^{-1}(n/\log n)^{1/d})$, where s_n is a diverging sequence, and $n \geq n_0(f_{\max}, d, \delta_0, \epsilon_o, C_1, \alpha)$ with probability at least $1 - 3/n$*

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

The proof follows similar to the proof of Lemma 3, except that we define $\epsilon_j = 2\sqrt{d}2^{-j}$, and we have the following result analogous to Proposition 2. The rest of the arguments in the proof of Lemma 3 follow through.

Proposition 4. *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 $n \geq n_1(f_{\max}, d, C_1)$ with probability at least $1 - 2/n$*

$$\sup_{x \in \widehat{G}_j \Delta G^*_\gamma} \rho(x, \partial G^*_\gamma) \leq 2\sqrt{d}2^{-j} =: \epsilon_j.$$

Proof. Define $\mathcal{B}_j$ to be the collection of cells in $\mathcal{A}_j$ that intersect the boundary ∂G^*_γ, or have one or more adjacent cells intersecting the boundary ∂G^*_γ. Consider $A \in \mathcal{B}_j$. Then $\forall x \in A$, $\rho(x, \partial G^*_\gamma) \leq 2\sqrt{d}2^{-j}$ (twice the diagonal length of a cell). Consider $A \notin \mathcal{B}_j$. Then $\forall x \in A$, $\rho(x, \partial G^*_\gamma) \geq 2^{-j}$ (the sidelength of a cell).

We prove that with high probability, all cells $A \notin \mathcal{B}_j$ are correctly labeled (not erroneously included or excluded from the level set estimate $\widehat{G}_j$). Notice that a cell $A \notin \mathcal{B}_j$ is included in the true level set G^*_γ if $\bar{f}(A) > \gamma$ and excluded if $\bar{f}(A) < \gamma$. Also recall that $\widehat{G}_j$ includes all cells with $\widehat{f}(A) > \gamma$ and excludes all cells with $\widehat{f}(A) < \gamma$. Therefore, a cell $A \notin \mathcal{B}_j$ is correctly labeled if $|\bar{f}(A) - \widehat{f}(A)| \leq |\gamma - \bar{f}(A)|$. We show that this is indeed the case.

Using Lemma 1, $\max_{A \notin \mathcal{B}_j} |\bar{f}(A) - \widehat{f}(A)| \leq \Phi_j$ with probability at least $1-1/n$ (with $\delta = 1/n$). And using Lemma 2, Φ_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 $n \geq n_1(f_{\max}, d, C_1)$, $\Phi_j \leq C_1$. Thus, with probability $> 1-2/n$,

$$\max_{A \notin \mathcal{B}_j} |\bar{f}(A) - \widehat{f}(A)| \leq \Phi_j \leq C_1 \leq |\gamma - \bar{f}(A)|.$$

The last step follows since under assumption **[A]** for $\alpha = 0$, $\forall x \in \mathcal{X}$, $|\gamma - f(x)| \geq C_1$. For $A \notin \mathcal{B}_j$, this implies that $|\gamma - \bar{f}(A)| \geq C_1$.

Therefore, for resolutions satisfying $2^j = O(s_n^{-1}(n/\log n)^{1/d})$ and $n \geq n_1$ $(f_{\max}, d, C_1)$, with probability at least $1 - 2/n$, all cells $A \notin \mathcal{B}_j$ are correctly labeled. Hence, $G^*_\gamma \Delta \widehat{G}_j \subseteq \mathcal{B}_j$. It follows that:

$$\sup_{x \in G^*_\gamma \Delta \widehat{G}_j} \rho(x, \partial G^*_\gamma) \leq \sup_{x \in \mathcal{B}_j} \rho(x, \partial G^*_\gamma) \leq 2\sqrt{d}2^{-j}, \tag{12}$$

where the last step follows from the definition of $\mathcal{B}_j$ since $2\sqrt{d}2^{-j}$ is twice the diagonal length of a cell. □

Now, to establish adaptivity to the whole range $\alpha \geq 0$, we re-sketch the proof of Theorem 2. First, notice that Lemma 4 still holds for the modified vernier and modified penalty since $\mathcal{V}_{\gamma,j}, \widehat{\mathcal{V}}_{\gamma,j}$ as well as $\Psi_{j'}$ are all scaled by the same factor of $2^{-j'/2}$. And we have the following analogue of Lemma 5 using the modified vernier:

Lemma 8. *Consider densities satisfying assumption **[A]** for $\alpha \geq 0$ and assumption **[C]**. Recall that $j' = \lfloor j + \log_2 s_n \rfloor$, where s_n is a diverging sequence. Then for all resolutions j*

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

holds for n large enough such that $s_n > 4C_4 6^d$. Here $C \equiv C(C_2, f_{\max}, \delta_0) > 0$.

Following the proof of Theorem 2, we derive upper bounds on $\mathcal{V}_{\gamma,\widehat{j}}$ and $\Psi_{\widehat{j}'}$ using the oracle inequality. Since both the modified vernier and penalty are scaled by the same factor, the two terms in the oracle inequality are still balanced for the same optimal resolution j^* given by $2^{-j^*} \asymp s_n^{\frac{d}{d+2\alpha}} (n/\log n)^{-\frac{1}{d+2\alpha}}$. Hence we get:

$$\begin{aligned} \mathcal{V}_{\gamma,\widehat{j}} \leq \widehat{\mathcal{V}}_{\gamma,\widehat{j}} + \Psi_{\widehat{j}'} &\leq C 2^{-{j^*}'/2} 2^{-j^*\alpha} \\ &\leq C s_n^{-1/2} s_n^{\frac{d(\alpha+1/2)}{d+2\alpha}} \left(\frac{n}{\log n}\right)^{-\frac{(\alpha+1/2)}{d+2\alpha}}. \end{aligned}$$

Using this upper bound on $\mathcal{V}_{\gamma,\widehat{j}}$ and $\Psi_{\widehat{j}'}$, we derive upper and lower bounds on the chosen resolution $\widehat{j}$ as in the proof of Theorem 2. Using Lemma 8, we have the following upper bound on the sidelength: For $s_n > 4C_4 6^d$,

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

And using Lemma 2 for the modified penalty, we have:

$$c_3 2^{-j'/2} \sqrt{2^{j'd} \frac{\log n}{n}} \leq \Psi_{j'}$$

This provides a lower bound on the sidelength:

$$\begin{aligned} 2^{-\hat{j}} > \frac{s_n}{2} \left(\frac{\Psi^2_{\hat{j}'}}{4c_3^2} \frac{n}{\log n} \right)^{-\frac{1}{(d-1)}} &\geq c_1 s_n \left(s_n^{-1} s_n^{\frac{2d(\alpha+1/2)}{d+2\alpha}} \left(\frac{n}{\log n} \right)^{-\frac{2(\alpha+1/2)}{d+2\alpha}} \frac{n}{\log n} \right)^{-\frac{1}{(d-1)}} \\ &= c_1 s_n s_n^{\frac{1}{(d-1)}} s_n^{\frac{-2d(\alpha+1/2)}{(d-1)(d+2\alpha)}} \left(\frac{n}{\log n} \right)^{\frac{-1}{d+2\alpha}} \\ &= c_1 s_n^{\frac{d}{d+2\alpha}} \left(\frac{n}{\log n} \right)^{\frac{-1}{d+2\alpha}}. \end{aligned}$$

So as before we have for $s_n > 4C_4 6^d$, with probability at least $1 - 5/n$,

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

where $c_1 \equiv c_1(C_2, f_{\max}, \delta_0, d, \alpha) > 0$ and $c_2 \equiv c_2(C_1, C_2, f_{\max}, \delta_0, d, \alpha) > 0$. Hence the automatically chosen resolution behaves as desired for $\alpha \geq 0$.

To arrive at the result of Theorem 2, follow the same arguments as before but using Lemma 3 to bound the Hausdorff error for $\alpha > 0$ and Lemma 7 to bound the Hausdorff error for $\alpha = 0$. Thus, Theorem 2 holds and the proposed method is adaptive for all $\alpha \geq 0$ (including the jump case), using the modified vernier and penalty.

Acknowledgements

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

References

[1] J. A. Hartigan, *Clustering Algorithms*. NY: Wiley, 1975.

[2] W. Stuetzle, "Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample," *J. Classification*, vol. 20(5), pp. 25–47, 2003.

[3] I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," *J. Mach. Learn. Res.*, vol. 6, pp. 211–232, 2005.

[4] C. Scott and R. Nowak, "Learning minimum volume sets," *J. Mach. Learn. Res.*, vol. 7, pp. 665–704, 2006.

[5] R. Vert and J.-P. Vert, "Consistency and convergence rates of one-class svms and related algorithms," *J. Mach. Learn. Res.*, vol. 7, pp. 817–854, 2006.

[6] M. Pacifico, C. Genovese, I. Verdinelli, and L. Wasserman, "False discovery control for random fields," *J. Amer. Statist. Assoc.*, vol. 99(468), pp. 1002–1014, 2004.

[7] R. Willett and R. Nowak, "Level set estimation in medical imaging," in *IEEE SSP Workshop*, 2005.

[8] Y. H. Yang, M. Buckley, S. Dudoit, and T. Speed, "Comparision of methods for image analysis on cdna microarray data," *J. Comput. Graph. Statist.*, vol. 11, pp. 108–136, 2002.

[9] R. Willett and R. Nowak, "Minimax optimal level set estimation," *IEEE Trans. Inform. Theory*, 2007, accepted for publication. [Online]. Available: http://www.ee.duke.edu/~willett/papers/WillettLevelSetsSubmitted.pdf

[10] A. Sole, V. Caselles, G. Sapiro, and F. Arandiga, "Morse description and geometric encoding of digital elevation maps," *IEEE Trans. Image Process.*, vol. 13(9), pp. 1245–1262, 2004.

[11] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, "Habitat monitoring with sensor networks," *Communications of the ACM*, vol. 47(6), pp. 34–40, 2004.

[12] C. Scott and M. Davenport, "Regression level set estimation via cost-sensitive classification," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2752–2757, 2007.

[13] A. P. Korostelev and A. B. Tsybakov, *Minimax Theory of Image Reconstruction.* NY: Springer, 1993.

[14] A. B. Tsybakov, "On nonparametric estimation of density level sets," *Ann. Statist.*, vol. 25, pp. 948–969, 1997.

[15] W. Polonik, "Measuring mass concentrations and estimating density contour cluster-an excess mass approach," *Ann. Statist.*, vol. 23, no. 3, pp. 855–881, 1995.

[16] P. Rigollet and R. Vert, "Fast rates for plug-in estimators of density level sets," 2006. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:math/0611473

[17] L. Cavalier, "Nonparametric estimation of regression level sets," *Statistics*, vol. 29, pp. 131–160, 1997.

[18] A. Cuevas, W. G. Manteiga, and A. R. Casal, "Plug-in estimation of general level sets," *Aust. N. Z. J. Stat.*, vol. 48(1), pp. 7–19, 2006.

[19] D. L. Donoho, "Cart and best-ortho-basis: A connection," *Ann. Statist.*, vol. 25, pp. 1870–1911, 1997.

[20] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Ann. Statist.*, vol. 32, no. 2, pp. 500–527, 2004.

[21] C. Scott and R. Nowak, "Minimax-optimal classification with dyadic decision trees," *IEEE Trans. Inform. Theory*, vol. 52(4), pp. 1335–1353, 2006.

[22] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 1996.

[23] R. Y. Liu, J. M. Parelius, and K. Singh, "Multivariate analysis by data depth: Descriptive statistics, graphics and inference," *Ann. Statist.*, vol. 27(3), pp. 783–858, 1999.

[24] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications.* West Sussex, England: Wiley, 1990.

[25] V. Vapnik, *The Nature of Statistical Learning Theory.* NY: Springer, 1995.

[26] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation.* NY: Springer, 2001.