# LEVEL SET ESTIMATION IN MEDICAL IMAGING

*Rebecca Willett and Robert Nowak*

Department of Electrical and Computer Engineering
University of Wisconsin, 1415 Engineering Drive, Madison, WI 53706, USA

## ABSTRACT

Rapid and accurate extraction of level sets and isoconcentration surfaces from noisy medical images is a common problem arising in a variety of contexts, such as estimating regions in which uptake of a pharmaceutical has exceeded some critical value or identifying areas of brain activity in neuroimaging. In general, a level set is the set $S$ on which a function $f$ exceeds a critical value (*e.g.* $S = \{x : f(x) > \gamma\}$). Boundaries of level sets and isoconcentration surfaces typically constitute manifolds embedded in the high-dimensional observation space. The tree structures underlying our method are constructed by minimizing a complexity regularized data-fitting term over a family of dyadic partitions. Our method specifically aims to minimize an error metric sensitive to both deviations in the location of the level set and the rate of change of the surface intensity or activity level statistic in the vicinity of the level set. Explicit extraction of level sets using multiresolution trees can be implemented in near linear time; simulations demonstrate that explicit level set extraction methods can achieve significantly higher accuracy in neuroimaging applications than more indirect approaches.

## 1. LEVEL SET ESTIMATION

In this paper, we address the problem of recovering level sets from multi-dimensional observations. Level set estimation may be more desirable than complete signal reconstruction for a variety of reasons. In many medical imaging applications, for example, the location of boundaries or level sets are of principal importance, while the amplitude of the function away from any boundary is secondary, if not irrelevant. In particular, doctors may wish to identify regions where uptake of a pharmaceutical exceeds some critical level using nuclear medicine data or want to localize regions of brain activity using functional magnetic resonance imaging (fMRI) data. Because set estimation is intrinsically simpler than field estimation, explicit level set extraction methods can potentially achieve higher accuracy than more indirect approaches (such as estimating a set from an estimate of the function).

This distinction manifests itself in the selection of tree-pruning criteria. Tree-based signal estimation frameworks (*e.g.* [1; 2; 3]) determine whether nodes should be pruned from the tree based on the size of the tree. Recent work in tree-based methods for binary classification [4], however, has revealed that optimal pruning decisions must exhibit a spatial adaptivity not possible using traditional, tree size-based criteria. This difference implies that one critical aspect of tree-based set estimation must be the selection of the tree-pruning rule most appropriate for set estimation.

The level set recovery problem can be formulated as follows: Let $f : [0,1]^d \longrightarrow [C_\ell, C_u]$ be a field of bounded amplitude, and let

$$S \equiv \{x \in [0,1]^d : f(x) \geq \gamma\}$$

for some $\gamma \in [C_\ell, C_u]$. Given $n$ noisy observations $(x_i, y_i) \in [0,1]^d \times [C_\ell, C_u]$, $i = 1, \ldots, n$, where $\mathbb{E}[y_i] = f(x_i)$, our goal is to learn an estimate of $S$. Let $\mathbb{P}_X$ be the probability measure for $X$, which determines the marginal distribution of observations on the support of $f$.

### 1.1. Relationship to previous work

One approach to this problem might be to first compute an estimate of $f$, denoted $\widehat{f}$, and then let $\widehat{S}_{\widehat{f}} = \{x \in [0,1]^d : \widehat{f}(x) \geq \gamma\}$; this is frequently called a "plug-in" estimator. To see why this approach may not be as effective as a more direct approach, consider the following sketch. Assume we wished to extract the level set displayed in Figure 1(b) from noisy observations of the function in Figure 1(a). Partition-based multiresolution estimators such as those described in [3] form a partition adapted to the observations and fit a constant to each cell in the partition. As described in [3], this approach has many advantages in the context of signal estimation, including near minimax optimal rates of convergence for broad classes of functions. However, these estimators attempt to minimize a global error metric and are unable to hone in on the boundary of a level set when it does not correspond to a boundary or edge in the function itself. In particular, in this example of a very smooth function devoid of sharp edges and boundaries, the estimator would form the uniform partition displayed in Figure 1(c), resulting in the level set estimate displayed in Figure 1(d). This paper, in contrast, describes a methodology which will partition $f$ in a way that hones in on the boundary of the level set, as displayed in Figure 1(e), to get more accurate set estimates, as displayed in Figure 1(f).

An alternative approach might consist of thresholding all the observations ($z_i = \mathbb{I}_{\{y_i \geq \gamma\}}$) and then applying a binary classification method to the result. However, this approach immediately disregards important information about the distance of each observation from $\gamma$, the level of interest. In a sense, this distance is an indication of our confidence that the observation is in $S$. That said, we have demonstrated that the problem of level set estimation is very closely related to classification and this allows us to use results from that domain for this problem. In particular, trees have been studied very carefully in the context of classification [4] and we will be using a number of analysis techniques and results from that domain for this problem.

Related work was conducted by Mammen and Tsybakov in

Supported by the National Science Foundation, grants CCR-0310889 and ANI-0099148, and the Office of Naval Research, grant N00014-00-1-0390

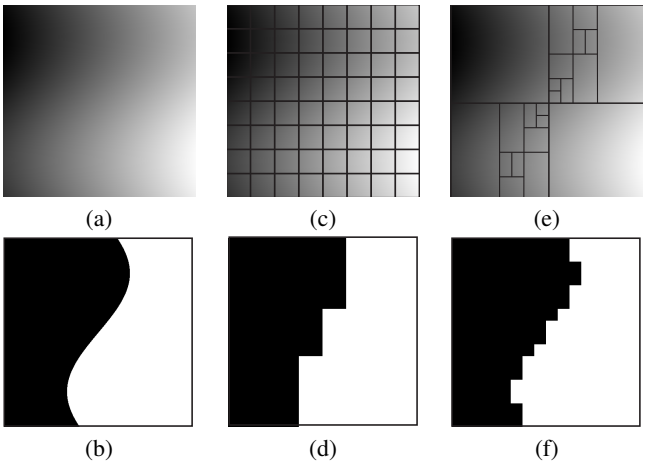

**Fig. 1**. Unbalanced partitions for level set estimation. (a) Smooth function. (b) Level set of function. (c) Partition induced on function by denoising. (d) Thresholding denoised image results in a large error. (e) Partition induced on function by set estimation method. (f) Direct set estimation can result in a smaller error.

[5], but their work focused on estimation of a boundary between a black and a white region from binary observations, an edge detection problem which is a special case of the more general level set estimation problem presented in this paper. The advantage of the method proposed in this paper is that it is capable of utilizing additional information available from non-binary observations. Cavalier [6] examined a problem similar to the one discussed here and based on the work of Tsybakov [7] for density level set estimation using piecewise polynomials. The estimators proposed in these works, however, are not computable and place stronger assumptions on $S$.

### 1.2. Error metrics for set estimation

Careful selection of an error metric is the first step in designing a level set estimator. The goal of signal estimation is typically to minimize the mean squared error between the true signal and the estimate, and the goal of binary classification is usually to minimize the probability of misclassification, which leads to minimizing the symmetric difference between the decision sets of the Bayes' and the learned classifiers. In level set estimation, however, it is more appropriate to minimize the symmetric difference between the true set of interest and its estimate weighted by severity of the error over the symmetric difference.

An appropriate error metric can be designed as follows. For a given set $F$, we define the loss function to be

$$e_F(x) = \frac{\gamma - f(x)}{2(C_u - C_\ell)} \left[\mathbb{I}_{\{x \in F\}} - \mathbb{I}_{\{x \in F^c\}}\right] + \frac{1}{2}$$

where $\mathbb{I}$ is the indicator function. The normalization ensures that $e_F \in [0, 1]$ and does not impact the result. From here, we define risk function as

$$\mathcal{R}(F) = \int e_F(x) d\mathbb{P}_X;$$

this measures the distance between the signal, $f$, and the threshold, $\gamma$, and weights the distance at each location $x$ by plus or minus one according to whether $x \in F$. Thus regions where $x \in F$ but $f(x) < \gamma$ (that is, $x \in S^c$) will contribute positively to the risk function. Given the risk function, we can define the "excess risk" as $\mathcal{R}(F) - \mathcal{R}(S)$, which measures the difference between the risk of an estimate and the risk of the true level set, $S$. Using the definition of the risk, the excess risk can be written as

$$\mathcal{R}(F) - \mathcal{R}(S) = \frac{1}{C_u - C_\ell} \int_{\Delta(S,F)} |\gamma - f(x)| d\mathbb{P}_X, \quad (1)$$

where

$$\Delta(S, F) \equiv \{x : x \in (S \cap F^c) \cup (F \cap S^c)\}$$

denotes the symmetric difference, where $S^c$ is the complement of $S$. The excess risk gives a weighted measure of the symmetric difference between $S$ and $F$, as desired. Note that minimizing (1) is equivalent to minimizing $\mathcal{R}(F)$ since $\mathcal{R}(S)$ is a constant.

The effect of such a metric is demonstrated in Figure 2. On the left is drawn a contour outlining the true level set $S$. The center and rightmost figures show the boundary of two different candidate level set estimates. There is only a small symmetric difference between the set in the center image and the truth, but the distance of the function from the level $\gamma$ is large in this region. In contrast, there is a large symmetric difference between the set in the rightmost image and the truth, but the distance of the function from the level $\gamma$ is relatively small in that region.

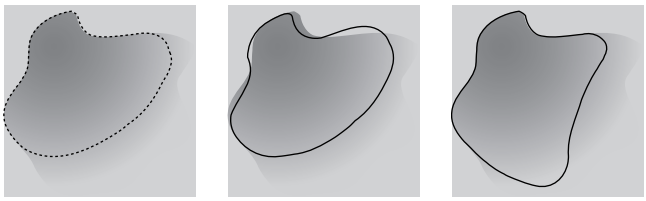

**Fig. 2**. Behavior of level set error metric. (a) Field $f$ and true level set $S$. (b) Level set estimate (solid line) with a small symmetric difference but large errors within the symmetric difference region. (c) Second level set estimate (solid line) with same error as estimate in (b); this estimate has a large symmetric difference but small errors within the symmetric difference region. Despite these differences, these two set estimates could have the same weighted symmetric difference risk.

An additional advantage of the proposed metric is that it is simple to define an empirical loss metric for a candidate level set estimate $F$ as

$$\widehat{e}_F(x_i) = \frac{\gamma - y_i}{2(C_u - C_\ell)} \left[\mathbb{I}_{\{x_i \in F\}} - \mathbb{I}_{\{x_i \in F^c\}}\right] + \frac{1}{2}$$

resulting in the empirical risk function

$$\widehat{\mathcal{R}}_n(F) = \frac{1}{n} \sum_{i=1}^{n} \widehat{e}_F(x_i),$$

which is both computable and constructed so that $\mathbb{E}[\widehat{\mathcal{R}}_n(F) - \mathcal{R}(S)] = \mathcal{R}(F) - \mathcal{R}(S)$. This metric is distinctly different from the global $L_p$ norms typically encountered in signal estimation or the "unweighted" symmetric difference metrics arising in classifi-

cation.

## 2. LEVEL SET ESTIMATION PROCEDURE

We propose to estimate the level set of a function from noisy observations by using a tree-pruning method akin to CART [1] or dyadic decision trees [4]. We will focus on binary trees, where each internal node in the tree is assigned a label corresponding to the coordinate direction in which the dyadic split takes place. Let $\pi(T)$ denote the partition induced on $[0,1]^d$ by the binary tree $T$. That is, if $T$ has $k$ leaf nodes ($|T| = k$), then $\pi(T) = \{A_1, A_2, \ldots, A_k\}$ so that each $A_i \in \pi(T)$ is a hyperrectangular subset of $[0,1]^d$ determined by the sequence of coordinate directions assigned to the ancestors of the corresponding leaf node $v_i$ in $T$. Note that $\bigcup_k A_k = [0,1]^d$, and $A_i \cap A_j = \emptyset$ for $i \neq j$. A zero or one is assigned to each leaf node of $T$ (equivalently, to each $A \in \pi(T)$) to indicate whether that cell of the partition is estimated to be in $\widehat{S}$. A sample tree and corresponding partition in displayed in Figure 3. We will consider the collection of all binary trees with at least $M = 2^J$ leaf nodes, for some nonnegative integer $J$, and denote this collection $\mathcal{T}_M$. Note that the sidelength of each cell must then necessarily be no less than $2^{-J}$.

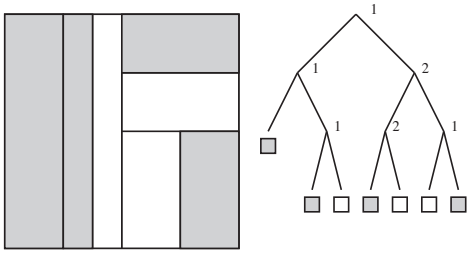


**Fig. 3**. Sample partition and corresponding tree.

Our goal is to find a tree based on noisy observations, $\widehat{T}_n$, so that $\mathcal{R}(\widehat{T}_n)$ is as small as possible. We will choose an estimate according to

$$\widehat{T}_n = \arg\min_{T \in \mathcal{T}_M} \widehat{\mathcal{R}}_n(T) + \Phi_n(T) \tag{2}$$

and explore how the choice of the penalty term $\Phi_n(T)$ impacts $\mathcal{R}(\widehat{T}_n)$. To avoid the suboptimal rates described in the preceding subsection, we must use a spatially adaptive penalty similar to that described in [4]. First, note that it is possible to express the difference between the true risk and the empirical risk as

$$\mathcal{R}(T) - \widehat{\mathcal{R}}_n(T) = \sum_{A \in \pi(T)} \mathcal{R}(T, A) - \widehat{\mathcal{R}}_n(T, A),$$

where $\mathcal{R}(T, A) = \int_A e_T(x) d\mathbb{P}_X$ and $\widehat{\mathcal{R}}_n(T, A) = \frac{1}{n} \sum_{i=1}^n \widehat{e}_T(x_i) \mathbb{I}_{\{x_i \in A\}}$.

Now let

$$\widehat{p}'_A(\delta) \equiv 4 \max\left(\widehat{p}_A \,,\, \frac{[\![A]\!] \log 2 + \log(1/\delta)}{n}\right)$$

for some $\delta \in [0,1]$, where $p_A \equiv \int_A d\mathbb{P}_X$ and $[\![A]\!]$ denotes the number of bits required to encode the position of $A$. Specifically, consider the prefix code proposed in [4] for $A \in \pi(T)$. If $A$ is at level $j$ in the binary tree $T$, then $j+1$ bits must be used to describe the depth of $A$, $j$ bits must be used to describe whether each branch is a left or right branch, and $j \log_2 d$ bits must be used to describe the coordinate direction of each of the $j$ branches. This results in a total of $j(\log_2 d + 2) + 1$ bits, and this expression is denoted as $[\![A]\!]$.

From here we can define the penalty term to be

$$\Phi_n(T) = \sum_{A \in \pi(T)} \sqrt{\frac{2\widehat{p}'_A(\delta)}{n} (\log(2/\delta) + [\![A]\!] \log 2)}. \tag{3}$$

Intuitively, this penalty is designed to favor unbalanced trees which hone in on the location of the manifold defining the boundary of the level set; the same reasoning explains why spatially adaptive penalties are so effective for binary classification problems using dyadic decision trees [4]. To see this, note that $[\![A]\!] \asymp j$, while $\widehat{p}'_A(\delta) \asymp 2^{-j}$. This implies that deep nodes contribute less to $\Phi_n(T)$ than shallow nodes, and so, for two trees with the same number of leafs, $\Phi_n(T)$ will be smaller for the more unbalanced tree, similar to the partition displayed in Figure 1(e).

As shown in the following section, the estimator defined by (2) and (3) is nearly minimax optimal. Furthermore, the estimator is rapidly computable; fast methods for binary classification using dyadic decision trees are described in [4; 8] and can easily be extended to the level set estimation problem. Specifically, let $L = \log_2 M$ be the maximum number of dyadic refinements along any coordinate used to form a tree $T$. Then $\widehat{T}_n$ can be computed in $O(ndL^d \log(nL^d))$ operations using a dynamic programming algorithm. To counteract grid effects which results from the use of binary trees, it is possible to perform a "voting over shifts" routine similar to the "averaging over shifts" or "cycle spinning" methods commonly used in signal estimation. Rather than averaging the results of shifted estimators, we determine whether each of the $M$ initial cells is more often in $\widehat{T}_n$ or $\widehat{T}_n^c$ and make the corresponding set assignment. A naive implementation of this approach would increase the complexity by a factor of $n$, but clever algorithms such as the one described in [9] only increase the complexity by a factor of $\log n$.

## 3. PERFORMANCE ANALYSIS

Many of the key points in the following theoretical analysis were derived using the error bounding techniques developed by Scott and Nowak [4] in the context of binary classification. The proposed level set estimation method hinges on the following main result:

**Theorem 1.** *Let $\Phi_n(T)$ be defined as in (3). Then with probability at least $1 - 2\delta$,*

$$\mathcal{R}(T) \leq \widehat{\mathcal{R}}_n(T) + \Phi_n(T) \tag{4}$$

*for all $T \in \mathcal{T}_M$.*

This is proved in [3]. Not only does this framework give us a principled way to choose a good level set estimator, but it also allows us to bound the expected risk for a collection of $n$ observations. In particular, we have the following theorem:

**Theorem 2.** *Let $\widehat{T}_n$ be as in (2) with $\Phi_n(T)$ as in (3), with $\delta =$*

$1/n$. *Then*

$$\mathbb{E}\left[\mathcal{R}(\widehat{T}_n) - \mathcal{R}(S)\right] \leq \min_{T \in \mathcal{T}_M} \left(\mathcal{R}(T) - \mathcal{R}(S) + 2\Phi(T)\right) + \frac{4}{n}.$$

This too is proved in [3]. This bound on the expected error in Theorem 2 allows us to analyze the proposed method in terms of rates of error convergence. In particular, because this problem has been posed as a generalization of the binary classification problem, we may draw extensively from the results of [4] to highlight several advantageous features of the error convergence of the proposed method. In this analysis, for sequences $a_n$ and $b_n$ let the notation $a_n \preceq b_n$ imply there exists some $C > 0$ such that $a_n \leq Cb_n$ for all $n$. For example, the following theorem shows that for a broad class of functions $f$ and distributions $\mathbb{P}_X$, the proposed method adapts to the regularity of the curve $\partial S$.

In particular, define the boundary fragment class $\mathcal{D}_{\mathrm{BF}}(\alpha, \gamma, c_0, c_1)$ for $\alpha < 1$ to be the set of all $(f, \mathbb{P}_X)$ such that

1. $\mathbb{P}_X(A) \leq c_0 \lambda(A)$ for all measurable $A \subseteq [0,1]^d$; and
2. one coordinate of $\partial S$ is a function of the others, where the function has Hölder smoothness $\alpha < 1$ and constant $c_1$.

The analysis here would also apply to compositions of members of $\mathcal{D}_{\mathrm{BF}}$, even if the "orientation" (which coordinate is a function of the others) varies member to member. The orientation does not need to be known. For problems in this class, we have the following theorem:

**Theorem 3.** *Choose $M$ such that $M \succeq (n/\log n)^{1/(d-1)}$. For $d \geq 2$ and $\alpha < 1$, we have*

$$\begin{aligned}
&\sup_{\mathcal{D}_{\mathrm{BF}}(\alpha,\gamma,c_0,c_1)} \left\{\mathbb{E}\left[\mathcal{R}(\widehat{T}_n) - \mathcal{R}(S)\right]\right\} \\
&\quad \preceq \quad \min_m \left\{ m^{-\alpha} + m^{(d-\alpha-1)/2}\sqrt{\frac{\log n}{n}} \right\} \\
&\quad \preceq \quad \left(\frac{\log n}{n}\right)^{\frac{\alpha}{\alpha+d-1}}.
\end{aligned}$$

This is proved in [3]. An examination of the lower bounds derived in [4] indicates that this rate is within a log factor of the minimax optimal rate. Note that $\alpha < 1$ implies fractal-like boundaries, which are important for a variety of medical imaging applications and similar to the ones displayed in the simulation section later in this paper.

The above analysis assumes that the locations of the observations are random and distributed according to $\mathbb{P}_X$. In many biomedical domains, however, the locations of the observations are dictated by the measurement device and lay on a grid (*e.g.* pixels in an image or voxels in a volume). In this case, the analysis above must be modified slightly, as described in [3].

## 4. EXTENSIONS TO NUCLEAR MEDICINE

The above framework can also be used in the context of nuclear medicine, in which observations do not fit the signal plus noise model, but rather indicate where and when photons hit a detector. Specifically, given $n$ iid observations of some density $f : [0,1]^d \longrightarrow [C_\ell, C_u]$, where $C_\ell \geq 0$, we wish to identify the set

$$S \equiv \{x \in [0,1]^d : f(x) \geq \gamma\}$$

as before. This problem has been explored in other contexts, notably [7]. Tsybakov's work exhibits similar rates of error convergence for estimation of density level sets, and in fact can nearly optimally estimate much smoother boundaries of level sets than those proposed here by fitting polynomials to the boundary. However, his problem is formulated solely in the plane ($d = 2$), he assumes that the boundary of the level set is "star-shaped" about the origin, and his approach does not admit a computationally tractable estimator.

The formulation of this problem is slightly different from the one described earlier in this paper because we do not make noisy observations of the amplitude of $f$. However, we will see that the above method and analysis hold after a change of loss metric. Let $\lambda(F)$ denote the Lebesgue measure of $F$, and let

$$\widehat{e}_F(x_i) = \frac{\left[\mathbb{I}_{\{x_i \in F^c\}} - \mathbb{I}_{\{x_i \in F\}}\right]}{2(C_u + 1)} + \frac{1}{2} + \frac{\gamma\left(\lambda(F) - \lambda(F^c)\right)}{2(C_u + 1)}.$$

This definition of empirical loss both (a) leads to an excess risk proportional to a weighted symmetric difference function as above and (b) allows us to apply the same analysis techniques as above since $\widehat{e}_F(x_i)$ is normalized to be in $[0,1]$. Specifically, since $\gamma \in [C_\ell, C_u]$, and $\lambda(F) \leq 1$, and $\lambda(F^c) = 1 - \lambda(F)$, we have $\widehat{e}_F(x_i) \in [0,1]$. Also, if

$$\widehat{\mathcal{R}}_n(F) = \frac{1}{n}\sum_{i=1}^n \widehat{e}_F(x_i),$$

then

$$\begin{aligned}
\mathcal{R}(F) &= \mathbb{E}\left[\widehat{\mathcal{R}}_n(F)\right] \\
&= \int \frac{\gamma - f(x)}{2(C_u+1)}\left[\mathbb{I}_{\{x\in F\}} - \mathbb{I}_{\{x \in F^c\}}\right] dx + \frac{1}{2}
\end{aligned}$$

and the excess risk is

$$\mathcal{R}(F) - \mathcal{R}(S) = \frac{1}{C_u + 1}\int_{\Delta(F,S)} |\gamma - f(x)| dx.$$

From here it is possible to apply the above analysis to derive an analogous objective function, method, and performance characterizations to those described for the alternative formulation. See [3] for details.

## 5. APPLICATIONS AND SIMULATION RESULTS

To test the practical effectiveness of the proposed method, we simulated estimating regions of brain activity from an fMRI study. An fMRI data set typically is composed of a time series of MRI brain images which are collected while the subject is performing a designated task (such as watching a video sequence or tapping a finger). For each pixel or voxel location $i$ in the brain, then, we measure $\boldsymbol{z}_i = \boldsymbol{s}_i + \boldsymbol{n}_i$, where $\boldsymbol{s}_i$ is a signal corresponding to the brain's activity in location $i$, and $\boldsymbol{n}_i$ is iid additive noise. (There may also exist additional distortions to the signal $\boldsymbol{s}_i$, such as a baseline shift, but there exist several techniques to eliminate such artifacts so we assume it is not present in this study. Allowing for these distortions within the proposed framework is an important avenue of future research.) In many such studies, we also have access to a reference signal, $\boldsymbol{r}$, which an expert has identified as a typical signal in an active region of the brain. Armed with the reference

signal, we can compute a cross-correlation map using the formula [10]

$$CC_i = \frac{\boldsymbol{z}_i^T \boldsymbol{r}}{\sqrt{\boldsymbol{z}_i^T \boldsymbol{z}_i \boldsymbol{r}^T \boldsymbol{r}}}.$$

(This assumes without loss of generality that $\boldsymbol{z}_i$ and $\boldsymbol{r}$ both have zero mean.) These cross-correlations can then be used to compute $t$-statistics and the associated statistical parametric map (SPM). The final step is then to assign a label (*e.g.* active or inactive) to each location based on the SPM. Labels can be assigned by thresholding the SPM, and thresholds can be chosen to minimize the probability of false alarm or family wise error rate [11]. Such techniques are statistically sound, they do not take into account the spatial information associated with SPMs. Other researchers have approached this problem by using Markov random fields [12] or by modeling hemodynamic response function [13].

The level set estimation method described in this paper can be used to label active and inactive regions of the brain from an SPM in a statistically robust fashion while implicitly accounting for spatial variations. Consider the brain activation map displayed in Figure 4(a). The intensity of this simulated image is proportional to the correlation between the corresponding pixel's signal and the reference signal. We wish to identify the region of the brain where the correlation exceeds some critical level and indicates activity, as displayed using white pixels in Figure 4(b). (The darker brain image in the backgrounds of Figures 4(b), (d), (e), and (f) is for visualization purposes and was not produced by the proposed method.) However, because the measured signals $\boldsymbol{z}_i$'s are noisy, we observe the correlation map displayed in Figure 4(c). This simulated data was generated by adding zero-mean uniform noise to the map displayed in Figure 4(a).

If we were to simply threshold the correlation map, we would estimate that the white pixels in Figure 4(d) corresponded to active regions; this clearly leads to many undesirable false alarms and does not exhibit the amount of structure we expect due to the physiology of the brain. The estimated generated by the proposed method, in contrast, exhibits both fewer false alarms and more inherent structure, as displayed in Figure 4(e). Because the number of dimensions in this example (two) is small relative to the number of observations, we employ a quad-tree structure instead of the binary tree described above. While this would not be feasible for much larger dimensions, in two dimensions it is both easy to implement and subject to all the performance characterizations derived earlier in this paper. This estimate was formed by weighting the penalty $\Phi_n(T)$ to minimize the average empirical risk and objectively highlight the difference between the proposed approach and a wavelet-based approach (which was given the same advantage, as described below). Furthermore, we employed the "voting over shifts" technique described earlier.

Compare this result with the result of a more indirect approach: namely, performing wavelet denoising and thresholding the denoised image to obtain a level set estimate, $\widehat{S}_{wavelet}$, to produce the estimate displayed in Figure 4(f). We used undecimated Haar wavelet denoising, and set the hard threshold to minimize the average empirical risk. After empirically selecting an appropriate weight on $\Phi_n(T)$ and the wavelet threshold, we observed the following mean risks over one hundred noise realizations:

$$\begin{aligned} \mathcal{R}(\widehat{S}_{thresh}) - \mathcal{R}(S) &= 90.66 \\ \mathcal{R}(\widehat{S}_{wavelet}) - \mathcal{R}(S) &= 8.95 \\ \mathcal{R}(\widehat{T}_n) - \mathcal{R}(S) &= 8.14. \end{aligned}$$

As predicted by our analysis, this plug-in estimator does not exhibit sufficient spatial adaptivity to adequately hone in on the boundary of the level set, resulting in a less accurate estimate.

Roughly speaking, wavelet denoising is analogous to choosing a partition with a penalty proportional to the size of the tree or partition, as opposed to the spatially adaptive penalty employed in this method. This example demonstrates that, as expected, the spatially adaptive penalty results in a partition which drills down on the location of the boundary; the wavelet-based approach, in contrast, appears to oversmooth the boundary. Furthermore, since the level set of interest does not correspond to an edge in the SPM, we would not expect curvelets or platelets to significantly outperform wavelets in this context.

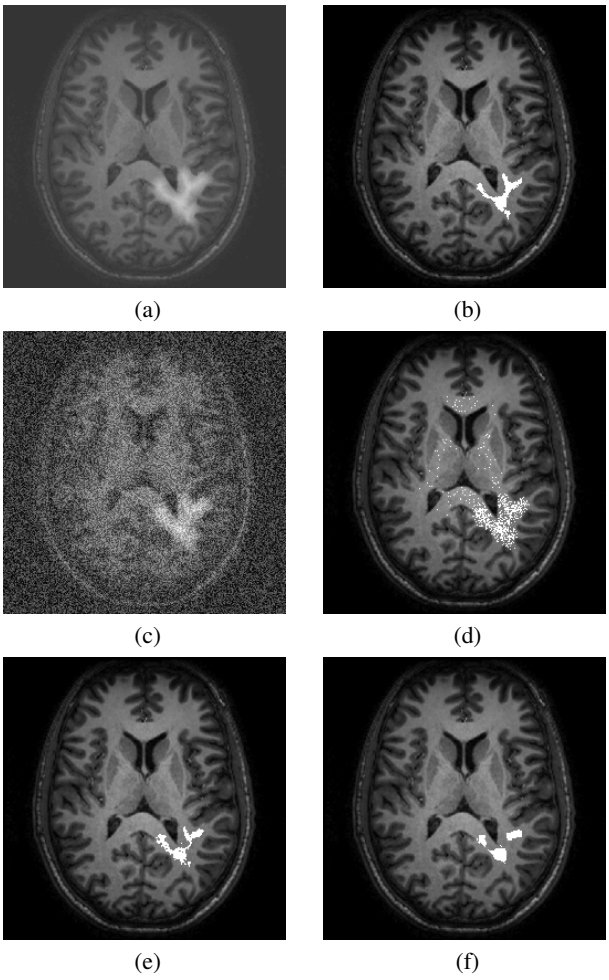


**Fig. 4**. FMRI simulation results. (a) True function $f : [0,1]^2 \rightarrow [0,255]$. (b) True level set $S = \{x \in [0,1]^2 : f(x) > 120\}$. (c) Noisy observations, $y_i \in [-95, 350]$, $i = 1, \ldots, 256^2$. (d) Level set of observations $\widehat{S}_{thresh} = \{x_i : y_i > 210\}$. $\mathcal{R}(\widehat{S}_{thresh}) - \mathcal{R}(S) = 94.28$. (e) Level set estimated with the proposed method. $\mathcal{R}(\widehat{T}_n) - \mathcal{R}(S) = 9.70$. (f) Level set estimated by TI Haar wavelet denoising followed by thresholding. $\mathcal{R}(\widehat{S}_{wavelet}) - \mathcal{R}(S) = 11.93$.

The results in this paper suggest that higher resolution fMRI

systems may be possible using sophisticated signal processing techniques – like the one presented here – which do not rely on independent voxel-wise testing. Because there are very few high-resolution fMRI systems available today, we demonstrate the effectiveness of our method at high resolutions in the context of digital elevation maps; we feel that these results indicate that higher-resolution fMRI systems would be quite advantageous in neuroimaging. In particular, we simulated observations of the elevation of St. Louis, where the true elevations, normalized to lie between zero and $255$, were obtained from the U.S. Geological Survey website and are displayed in Figure 5(a). Organizations such as the USGS are often interested in identifying flood plains, which shift as a result of plate tectonics and erosion. The true flood plain (the level set of interest) is displayed in Figure 5(b); note that it encompasses low-lying regions outside the river, distinguishing this problem from an edge detection problem. Our goal is to extract the flood plain from the set of noisy observations displayed in Figure 5(c). The noisy observations were obtained by adding zero-mean uniform noise with a variance of three thousand to the true image. (Note that this implies that the $x_i$'s are deterministic.) As shown in Figure 5(d), simply thresholding the observations to obtain the level set $\widehat{S}_{thresh}$ is highly insufficient in the presence of noise. In contrast, the application of the proposed method to this data results in an accurate estimate of the level set, $\widehat{T}_n$, as displayed in Figure 5(e). The plug-in estimator described for the fMRI data was used to produce the image in Figure 5(f).

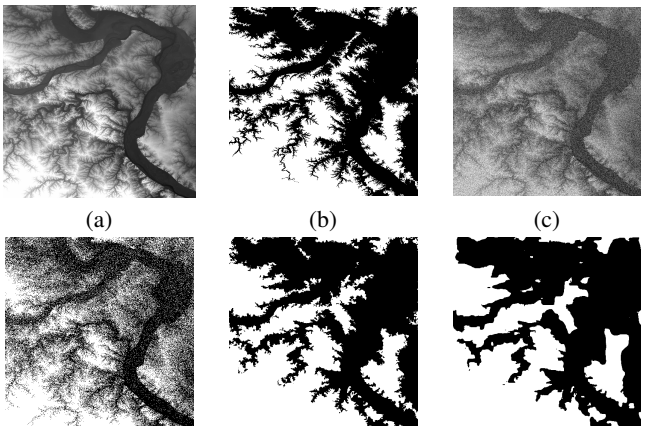


**Fig. 5**. Simulation results. (a) True function $f : [0,1]^2 \rightarrow [0,255]$. (b) True level set $S = \{x \in [0,1]^2 : f(x) > 120\}$. (c) Noisy observations, $y_i \in [-95, 350]$, $i = 1, \ldots, 512^2$. (d) Level set of observations $\widehat{S}_{thresh} = \{x_i : y_i > 120\}$. $\mathcal{R}(\widehat{S}_{thresh}) - \mathcal{R}(S) = 8,598$. (e) Level set estimated with the proposed method. $\mathcal{R}(\widehat{T}_n) - \mathcal{R}(S) = 1,104$. (f) Level set estimated by TI Haar wavelet denoising followed by thresholding. $\mathcal{R}(\widehat{S}_{wavelet}) - \mathcal{R}(S) = 1,472$.

### 6. CONCLUSIONS

We have demonstrated that tree-pruning based approaches to level set estimation result in nearly optimal estimates and can be computed rapidly to produce effective practical estimates. The introduction of a new error metric allows us to bound the weighted symmetric difference between the true level set and the estimate and show that it decays at nearly the minimax optimal rate.

## References


[1] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1983.

[2] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, 2003.

[3] R. Willett, *Multiresolution Methods for Recovering Signals and Sets from Noisy Observations*, Ph.D. thesis, Rice University, 2005.

[4] C. Scott, *Dyadic Decision Trees*, Ph.D. thesis, Rice University, 2004.

[5] E. Mammem and A. Tsybakov, "Asymptotical minimax recovery of sets with smooth boundaries," *Annals of Statistics*, vol. 23, no. 2, pp. 502–524, 1995.

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

[7] A. Tsybakov, "On nonparametric estimation of density level sets," *Annals of Statistics*, vol. 25, no. 3, pp. 948–969, 1997.

[8] G. Blanchard, C. Schäfer, and Y. Rozenholc, "Oracle bounds and exact algorithm for dyadic classification trees," in *Proceedings of COLT: The Annual Workshop on Learning Theory*, Banff, Canada, 2004, pp. 378–392.

[9] R. Willett and R. Nowak, "Fast multiresolution photon-limited image reconstruction," in *Proc. IEEE Int. Sym. Biomedical Imaging — ISBI '04*, 15-18 April, Arlington, VA, USA, 2004.

[10] H. Soltanian-Zadeh, D. Peck, D. Hearshen, and R. Lajiness-O'Neill, "Model-independent method for fmri analysis," *IEEE Transactions on Medical Imaging*, vol. 23, no. 3, pp. 285–296, 2004.

[11] T. Nichols and S. Hayasaka, "Controlling the familywise error rate in functional neuroimaging: a comparative review," *StatisticalMethods in Medical Research*, vol. 12, pp. 419–446, 2003.

[12] X. Descombes, F. Kruggel, and D. von Cramon, "Spatio-temporal fmri analysis using markov random fields," *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 1028–1039, 1998.

[13] P. Ciuciu, J. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali, "Unsupervised robust nonparametric estimation of the hemodynamic response function for any fmri experiment," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1235–1251, 2003.