FAST LEVEL SET ESTIMATION FROM PROJECTION MEASUREMENTS

Kalyani Krishnamurthy, Waheed U. Bajwa, Rebecca Willett, and Robert Calderbank

Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708

ABSTRACT

Estimation of the level set of a function (i.e., regions where the function exceeds some value) is an important problem with applications in digital elevation maps, medical imaging, and astronomy. In many applications, however, the function of interest is acquired through indirect measurements, such as tomographic projections, coded-aperture measurements, or pseudo-random projections associated with compressed sensing. This paper describes a new methodology and associated theoretical analysis for rapid and accurate estimation of the level set from such projection measurements. The proposed method estimates the level set from projection measurements without an intermediate function reconstruction step, thereby leading to significantly faster computation. In addition, the coherence of the projection operator and McDiarmid's inequality are used to characterize the estimator's performance.

***Index Terms*—** Compressed sensing, coherence, level sets, performance bounds, segmentation, thresholding

1. INTRODUCTION

Level set estimation is the process of using observations of a function f defined on a Hilbert space $\mathcal{X}$ to estimate the region(s) in $\mathcal{X}$ where f exceeds some critical value γ; i.e. $S^* \equiv \{x \in \mathcal{X} : f(x) \geq \gamma\}$. Accurate and efficient level set estimation plays a crucial role in a variety of scientific and engineering tasks, including the localization of "hot spots" signifying tumors in medical imaging, significant photon sources in astronomy, or strong reflectors in remote sensing. Previous work by one of the authors [1] explored the estimation of level sets of a function f from noisy observations of the form $\widetilde{y} = f + n$, where n denotes a vector of *independent*, zero-mean noise realizations. However, there are many contexts where direct observations of this form are not available; instead, we make observations of the form $y = Af + n$, where A is a linear projection operator that may not be invertible. For instance, y might correspond to tomographic projections in tomography, multiple blurred, low-resolution, dithered snapshots in astronomy, or pseudo-random projections in compressed sensing systems [2].

Our goal in this setting is to perform level set estimation *without* an intermediate step involving time-consuming reconstruction of f. There are two reasons for this: (1) Level set estimation without reconstruction of f would allow sequential measurement schemes to be performed on the fly. For instance, in tomography we would like to estimate S^* quickly from the observations so that additional data focused on S^* can be collected immediately, resulting in an overall low radiation dose. (2) "Plug-in" approaches that estimate f and threshold the estimate $\widehat{f}$ to extract S^* are notoriously difficult to characterize; the performance hinges upon the statistics of $\widehat{f} - f$, which for most reconstruction methods are unknown (with the possible exception of the first moment). More generally, reconstruction methods aim to minimize the total error, integrated or averaged spatially over the entire function. This does little to control the error at specific locations of interest, such as in the vicinity of the level set.

K. Krishnamurthy and R. Willett are supported by NSF Career Award No. CCF-06-43947, NSF Grant No. DMS-08-11062, and AFRL Grant No. FA8650-07-D-1221. W.U. Bajwa and R. Calderbank are supported by NSF grant DMS-0914892, ONR grant N00014-08-1-1110, and AFOSR grants FA9550-09-1-0643 and FA9550-09-1-0551.

1.1. Our contribution and relation with previous work

In this paper, we demonstrate that, subject to certain conditions on the (linear) projection operator A and the ℓ_1 norm of f, the level set S^* can be estimated quickly and accurately without first reconstructing f. Our method consists of constructing a set of *proxy observations* $z = f + n'$ from the actual observations y and applying a variation on the tree-based set estimation techniques established in [1] on z. The idea of constructing proxy observations z to deduce certain properties of the underlying f has been successfully employed in recent compressed sensing and statistics literature to solve the problem of support detection of a discrete f having no more than m non-zero entries; see, e.g., [3, 4, 5]. In fact, our level set estimation methodology is inspired by the empirical and theoretical success of using thresholded proxy observations for support detection of (discrete) sparse functions. However, despite the fact that both our method and the thresholded support detection of [3, 4, 5] make use of a function proxy, there are some key differences between the two lines of work that stem partly from different underlying assumptions. Specifically, it is established in [3] that the support of an m-sparse f can be reliably detected from appropriately thresholded proxy observations with an overwhelming probability as long as A satisfies a certain coherence property. On the other hand, the level set estimation analysis carried out in this paper does not impose any sparsity constraints on the underlying function f. Indeed, it is not difficult to convince oneself that directly thresholding the proxy observations for level set estimation in the case of a non-sparse f would lead to numerous false positives and false negatives (see, e.g., Figs. 1(c) and 1(e)). In contrast, our methodology relies on a novel two-step approach that enables us to work with proxy observations without requiring f to be strictly sparse. In our experiments, the proposed method performs an order of magnitude better than sparse support estimation methods in [3, 4, 5] with an appropriate threshold.

There are two key challenges we address in our analysis to specify the proposed estimator's performance. First, we must characterize n', which can be considered a combination of noise and interference caused by calculations of the proxy observations; the interference plays a crucial role in our estimation error and scales with the *worst-case coherence* of the projection operator A and the ℓ_1 norm of f. Second, the original analysis in [1] considered independent noise realizations, allowing for the application of Hoeffding's inequality to analyze estimator's performance. However, n' in our proxy observations contains statistical dependencies which we consider in our revised analysis. Our theoretical analysis is validated via a medical imaging simulation.

2. FAST LEVEL SET ESTIMATION FROM PROJECTION MEASUREMENTS

Let us consider a function $f \in \mathbb{R}^N$ and let f_i denote the i^{th} element of f. We are interested in estimating a γ level set $S^* = \{i : f_i > \gamma\}$ of f from projection measurements of the form $y = Af + n \in \mathbb{R}^K$ for $K \leq N$, where $A \in \mathbb{R}^{K \times N}$ is a projection operator that is assumed to be known, and n is bounded, independent and zero-mean noise [1]. The main goal of this paper is to estimate S^* from y *without reconstructing* f.

For $A = I$, [1] provides a minimax optimal level set estimation strategy that estimates S^* from noisy observations $\widetilde{y} = f + n \in \mathbb{R}^N$ without estimating f. Though a direct application of the estimation strategy in [1] to our problem is nontrivial because $A \neq I$, we will construct proxy observations $z = f + n'$ from y and draw on some of the key insights from [1] to address our problem. Before we present our estimation method, we briefly discuss the main ideas of [1] below.

2.1. Previous work on level set estimation

The basic idea in [1] was to design an estimator of the form $\widehat{S} = \arg\min_{S \in \mathcal{S}} \widehat{R}(S) + \Phi(S)$, where $\mathcal{S}$ is a class of candidate estimates, $\widehat{R}$ is an empirical measure of the estimator risk based on N noisy observations of the function f, and Φ is a regularization term which penalizes improbable level sets. They described choices for $\widehat{R}$, Φ, and $\mathcal{S}$ which made $\widehat{S}$ rapidly computable and minimax optimal for a large class of level set problems. In particular, [1] proposed a novel error metric between S^* and a candidate estimate S that was ideally suited to the problem at hand. That error metric in our case can be written as

$$\frac{1}{N} \sum_{i \in \Delta(S^*, S)} |\gamma - f_i| \tag{1}$$

where $\Delta(S^*, S) \equiv \{i \in (S^* \cap \overline{S}) \cup (\overline{S^*} \cap S)\}$ denotes the symmetric difference, and $\overline{S}$ is the complement of S. While the expression in (1) is not directly computable (since S^* is unknown), one can nevertheless minimize it by defining the *risk*

$$R(S) \equiv \frac{1}{N} \sum_i \underbrace{(\gamma - f_i) \left[\mathbb{I}_{\{i \in S\}} - \mathbb{I}_{\{i \notin S\}} \right]}_{\ell_i(S)} \tag{2}$$

where $\mathbb{I}_{\{E\}} = 1$ if event E is true and 0 otherwise. The loss function $\ell_i(S)$ measures the distance between the function at location i, f_i, and the threshold, γ, and weights this distance by ± 1 according to whether $i \in S$ or not. Note that $R(S) - R(S^*)$ is equivalent to the error metric defined in (1), and that it is possible to estimate $R(S)$ from $\widetilde{y}$ in a straightforward manner.

Using Hoeffding's inequality, [1] derived a regularization term Φ and developed a dyadic tree-based framework which can be used to obtain $\widehat{S}$. Trees were utilized for a couple of reasons. First, they both restricted and structured the space of potential estimators in a way that allowed the global optimum to be both rapidly computable and very close to the best possible (not necessarily tree-based) estimator. Second, they allowed the estimator selection criterion to be spatially adaptive, which appeared to be critical for the formation of provably optimal estimators.

[1] This assumption is reasonable, since, in practice, the noise is always bounded due to hardware limitations in physical sensors.

2.2. Our method

In order to extract the γ- level set of f from y, we propose a novel two-step procedure. First, we construct a proxy of f as follows:

$$z = A^T y = f + \underbrace{\left(A^T A - I \right) f + A^T n}_{n'}. \tag{3}$$

This allows us to arrive at the canonical signal plus noise observation model. Next, we perform level set estimation on the proxy observations z, rather than on y, by relying on the insights of [1]. Note that for any unitary A, z in (3) reduces to $\widetilde{y}$. However, for non-unitary A, the proxy defined in (3) creates a signal-based interference term $\left(A^T A - I\right) f$ and a zero-mean *correlated* noise term $A^T n$. This interference term n' makes a direct extension of the level set approach discussed in [1] to our problem nontrivial. In our work, we theoretically analyze the impact of n' and use the theoretical results to develop a dyadic, tree-based level set estimation approach that adapts to the interference term.

Given z, our goal is to find a level set estimate $\widetilde{S} = \arg\min_{S \in \mathcal{S}} R(S) - R(S^*)$ where $\mathcal{S}$ is a family of candidate level set estimates and $R(\cdot)$ is defined in (2). (Note that $\widetilde{S} = S^*$ if $S^* \in \mathcal{S}$.) Since f is unknown, we consider an empirical risk of the form

$$\widehat{R}(S) = \frac{1}{N} \sum_{i=1}^{N} (\gamma - z_i) \left[\mathbb{I}_{\{i \in S\}} - \mathbb{I}_{\{i \notin S\}} \right] \tag{4}$$

and show that finding an estimate $\widehat{S} = \arg\min_{S \in \mathcal{S}} \widehat{R}(S) + \Phi(S)$, where $\Phi(S)$ is an interference-dependent penalty term, yields $\left| R(\widehat{S}) - R(\widetilde{S}) \right| \xrightarrow{N} 0$. The penalty term plays a major role in our estimation strategy and is crucial in finding estimates that hone in on the boundary of the level set S^*. We thus focus on designing a spatially adaptive penalty $\Phi(S)$ that promotes well-localized level sets with potentially non-smooth boundaries. Following the analysis in [1] we let $\mathcal{S}$ be a family of level set estimates defined on recursive dyadic partitions of the domain of f; e.g., an image could be partitioned into patches of varying side-lengths using a quad-tree, so that each leaf of the tree corresponded to one patch. Each leaf in the partition would be estimated to be in or out of the level set of interest. Let $\pi(S)$ be the partition induced by an estimate $S \in \mathcal{S}$. Then, the risk of S in each of its leaf $L \in \pi(S)$ is given by $R(L) = \frac{1}{N} \sum_{i=1}^{N} (\gamma - f_i) \left[\mathbb{I}_{\{\ell(L)=1\}} - \mathbb{I}_{\{\ell(L)=0\}} \right] \mathbb{I}_{\{i \in L\}}$ where $\ell(L) = 1$ if $i \in L$ and 0 otherwise. We design a spatially adaptive penalty term by analyzing $R(L) - \widehat{R}(L)$ at each leaf separately. Note that $R(S) = \sum_{L \in \pi(S)} R(L)$. To facilitate our analysis, let us define $\widetilde{R}(L) = \frac{1}{N} \sum_{i=1}^{N} (\gamma - \mathbb{E}[z_i]) \left[\mathbb{I}_{\{\ell(L)=1\}} - \mathbb{I}_{\{\ell(L)=0\}} \right] \mathbb{I}_{\{i \in L\}}$. Then

$$\begin{aligned} R(L) - \widehat{R}(L) &= R(L) - \widetilde{R}(L) + \widetilde{R}(L) - \widehat{R}(L) \\ &\equiv \underbrace{\frac{1}{N} \sum_{i=1}^{N} (\mathbb{E}[z_i] - f_i) \left[\mathbb{I}_{\{\ell(L)=1\}} - \mathbb{I}_{\{\ell(L)=0\}} \right] \mathbb{I}_{\{i \in L\}}}_{T_1} \\ &+ \underbrace{\frac{1}{N} \sum_{i=1}^{N} (z_i - \mathbb{E}[z_i]) \left[\mathbb{I}_{\{\ell(L)=1\}} - \mathbb{I}_{\{\ell(L)=0\}} \right] \mathbb{I}_{\{i \in L\}}}_{T_2}. \end{aligned} \tag{5}$$

Note that while T_1 is a measure of the biases of $\{z_i\}$, T_2 is a measure of the concentration of $\{z_i\}$ around their means. Let us consider the statistics of z to further understand T_1 and T_2 respectively. Assuming without loss of generality that the columns of A have unit ℓ_2 norms, one can easily see from (3) that $z_i = f_i + \sum_{j=1,j\neq i}^{N} f_j \langle A^{(i)}, A^{(j)}\rangle + \langle A^{(i)}, n\rangle$ where $A^{(i)}$ denotes the i^{th} column of A. Since A is given, and n is zero mean, the term $\mathbb{E}[z_i] - f_i = \sum_{j=1,j\neq i}^{N} f_j \langle A^{(i)}, A^{(j)}\rangle$ in T_1 is the signal-based interference term at the i^{th} location due to the signal energies at other locations. We upper bound T_1 by the ℓ_1 norm of f and the worst-case coherence of A, and bound T_2 using McDiarmid's inequality [6] and sum the risk in each leaf of the estimate S to arrive at our main result stated below:

Theorem 1. *Suppose that the entries of noise n are bounded between $[c_\ell, c_u]$. Then, for $\delta \in [0, 1/2]$, with probability at least $1-\delta$, the following holds for all $S \in \mathcal{S}$:*

$$R(S) \le \widehat{R}(S) + \left(\frac{N-1}{N}\right)\mu(A)\|f\|_1 + \sum_{L\in\pi(S)} \sqrt{\frac{[\log(2/\delta) + [\![L]\!]\log 2]\, c^2 \widehat{p}_L\, (\mu(A)\,[N\widehat{p}_L - 1] + 1)}{2N}}. \tag{6}$$

where $c = |c_u - c_\ell|$, $\mu(A)$ is the worst-case coherence of A given by $\mu(A) = \max_{i,j\in\{1,\ldots,N\}, i\neq j} \left|\langle A^{(i)}, A^{(j)}\rangle\right|$, $\widehat{p}_L = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{\{i\in L\}}$ and $[\![L]\!]$ is the number of bits in a prefix code for L.

A sketch of the proof of this theorem is provided in the appendix. Note that the bound in (6) depends on (a) the signal-based interference term in (3) through $\|f\|_1$, (b) the noise through $|c_u - c_\ell|$, (c) the choice of A through $\mu(A)$, (d) the size and depth of each leaf through $\widehat{p}_L$ and $[\![L]\!]$, respectively, and (e) the parameter δ. If L is in level j of a dyadic, tree-based estimate S, then it has been shown in [1] that $[\![L]\!] \asymp j$ and $\widehat{p}_L \asymp 2^{-j}$ for $j \in 0, \ldots, \log_2 N$. As a result, searching for an estimate $\widehat{S} \in \mathcal{S}$ that minimizes the upper bound of $R(S)$ in (6) will favor estimates with few, deep leaves that hone in on the boundary of the level set. Though the theoretical analysis of our method is significantly different from the analysis in [1] because of the interference term, it only changes the way the penalty is defined in our setup. As a result, we can adapt the computational techniques discussed in [1] to compute our estimator in an efficient way. Our method is computationally efficient since the proxy computation needs at most $O(KN)$ operations (fewer if A contains certain structure; e.g., A is a Toeplitz matrix) and the level set estimation method needs $O(N \log N)$ operations.

Recent work on sparse support estimation in [3] suggests that the second term (proportional to $\mu(A)\|f\|_1$) in the right hand side of (6) may be tightened significantly if we make some additional assumptions. In particular, if f is assumed to be approximately sparse, so that the ℓ_1 norm of f restricted to the set $\overline{S^*}$ is sufficiently small, then the bound on T_1 may be tightened by using the *average coherence* $\nu(A) = \frac{1}{N-1}\left\|(A^T A - I)\mathbb{1}\right\|_\infty$ of the projection operator A [3]. A comprehensive analysis that utilizes both $\mu(A)$ and $\nu(A)$ to tighten the bounds in Theorem 1 will be explored in a sequel to this work.

3. EXPERIMENTAL RESULTS

We demonstrate the effectiveness of the proposed approach by performing level set estimation of the medical image $f \in \mathbb{R}^{128\times128}$ shown in Fig. 1(a) from its projection measurements $y = Af + n \in \mathbb{R}^K$ for $K \le N$, N = 128 × 128, where the entries of A are drawn from $\mathcal{N}(0, 1/K)$ and $n \sim \mathcal{N}(0, I)$. Note that we consider a Gaussian noise model here, which is unbounded. However, it is bounded with high probability and our theory can also be extended to this case. This input image is a cropped portion of a magnetic resonance angiography image of the brain (http://en.wikipedia.org/wiki/File:Mra-mip.jpg) where one is interested in identifying the blood vessels that react to an injected contrast agent. The goal of this experiment is to find the γ-level set of the image in Fig. 1(a) for $\gamma = 420$ from y without reconstructing f. Note that the input image is not even approximately sparse, but the locations corresponding to the level set occupy only a small portion of the entire image scene as shown in Fig. 1(b). Specifically, the cardinality of the true level set is $|S^*| = 1452$, which is approximately 9% of N. Moreover, the mean intensity at the locations in the level set is about 646.8 and that outside the level set is 185.4.

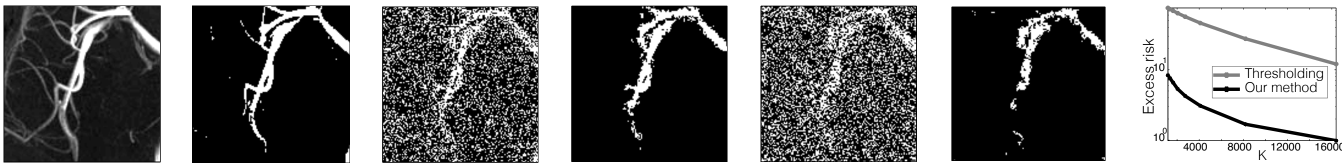
We compare the estimates obtained using our method with the ones obtained by simply thresholding the proxy observations at level γ for different values of $K \le N$. Estimates are computed by minimizing (6) with a scaling factor in front of the sum over leafs. This scaling factor is chosen to minimize the excess risk in (1). Figs. 1(c) and 1(e) show the results obtained by the thresholding method for $K \approx N/2$ and $K \approx N/4$ respectively. These pictures illustrate the effect of interference on the proxy observations z. Since $z_i = f_i + n'_i$, where $n'_i = \sum_{j:j\neq i} f_j \langle A^{(i)}, A^{(j)}\rangle + \langle A^{(i)}, n\rangle$, a thresholding operation on z will result in several false positives and misses at locations where f_i is comparable to n'_i, as shown in Figs. 1(c) and 1(e). The interference increases as K decreases since the worst-case coherence of A increases with a decrease in K. Figs. 1(d) and 1(f) show the results obtained by our method for $K \approx N/2$ and $K \approx N/4$ respectively, which offers an order of magnitude improvement in the excess risk over thresholding methods associated with sparse support estimation. Fig. 1(g) shows how the excess risk in both these approaches vary as a function of K. These plots are obtained by averaging the results obtained over 500 different noise and projection operator realizations. Estimates obtained using our method outperform the ones obtained by thresholding the proxy observations since our method relies on the spatial piecewise homogeneity of the underlying function f to eliminate false positives and misses. This is also the reason why our method misses some of the isolated pixels that correspond to the true level set as shown in Figs. 1(d) and 1(f).

4. CONCLUSION

This work proposes a theoretically tractable and computationally efficient tree-based approach to extract level sets of a function from projection measurements without reconstructing the underlying function. Simulation experiments demonstrate the applicability of our method to medical imaging applications. One of the key advantages of our approach is that we can parallelize the level set estimation problem when the domain of the function of interest is very large. In such cases, we can partition the data into different patches, run our estimation algorithm on each patch separately and merge the results to identify the regions that correspond to the level set. In applications such as medical imaging, the time saved by collecting fewer projective measurements and parallelization can be significant and crucial.

5. APPENDIX

Let us begin by bounding T_1 and T_2 in (5) separately. Let $\widehat{p}_L = \sum_{i\in L}\frac{1}{N}$. From the statistics of z, we can bound $T_1 =$



(a) (b) (c) (d) (e) (f) (g)

Fig. 1. Simulation results. (a) True function $f \in \mathbb{R}^{128\times128}$ such that $f_i \in [87, 765]$. We measure K Gaussian random projections of this image. (b) Level set $S^* = \{i : f_i > 420\}$ such that $|S^*| \approx 0.088N$ where $N = 128 \times 128$. (c) Level set obtained by thresholding the proxy observations z when $K \approx N/2$; excess risk = 59.35. (d) Level set obtained by the proposed approach when $K \approx N/2$; excess risk = 3.746. (e) Level set obtained by thresholding the proxy observations z when $K \approx N/4$; excess risk = 73.40. (d) Level set obtained by the proposed approach when $K \approx N/4$; excess risk = 7.407. (g) Performance comparison of our method and the thresholding approach for different $K \le N = 16384$.

$\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}\left[z_i\right]-f_i\right)\left[\mathbb{I}_{\{\ell(L)=1\}}-\mathbb{I}_{\{\ell(L)=0\}}\right]\mathbb{I}_{\{i\in L\}}$ as follows:

$$
\begin{aligned}
T_1 &\le \frac{1}{N}\sum_{i,j:j\neq i}|f_j|\left|\langle A^{(i)}, A^{(j)}\rangle\right|\left|\left[\mathbb{I}_{\{\ell(L)=1\}}-\mathbb{I}_{\{\ell(L)=0\}}\right]\right|\mathbb{I}_{\{i\in L\}} \\
&\le \frac{\mu(A)}{N}\sum_{i\in L}\sum_{j=1:j\neq i}^{N}|f_j| = \frac{\mu(A)}{N}\sum_{i\in L}\left(\sum_{j=1}^{N}|f_j| - |f_i|\right) \\
&\le \mu(A)\widehat{p}_L\|f\|_1 - \frac{\mu(A)}{N}\sum_{i\in L}|f_i| \qquad (7)
\end{aligned}
$$

where the third inequality is due to the fact that $\left|\left[\mathbb{I}_{\{\ell(L)=1\}}-\mathbb{I}_{\{\ell(L)=0\}}\right]\right| = 1$ and $\left|\langle A^{(i)}, A^{(j)}\rangle\right| \le \mu(A)$ for all $j \neq i$.

We can bound T_2, which is a function of z, using McDiarmid's inequality by observing that $z_i = \sum_{j=1}^{K} a_{j,i}y_j$ and thus T_2 can be written as a function of the independent random variables $\{y_i\}$. Specifically,

$$
\begin{aligned}
T_2 &= \frac{1}{N}\sum_{i\in L}\left(z_i - \mathbb{E}\left[z_i\right]\right)\left[\mathbb{I}_{\{\ell(L)=1\}}-\mathbb{I}_{\{\ell(L)=0\}}\right] \\
&\equiv \underbrace{\sum_{j=1}^{K} y_j w_j(L)}_{g(y_1,\ldots,y_K)} - \underbrace{\mathbb{E}\left[\sum_{j=1}^{K} y_j w_j(L)\right]}_{\mathbb{E}[g(y_1,\ldots,y_K)]}
\end{aligned}
$$

where $w_j(L) = \sum_{i\in L}\left[\mathbb{I}_{\{\ell(L)=1\}}-\mathbb{I}_{\{\ell(L)=0\}}\right]\frac{a_{j,i}}{N}$. Since the entries of noise n are bounded, it can be shown that the function $g(y_1,\ldots,y_K)$ satisfies the bounded differences property [6], which lets us use McDiarmid's inequality to arrive at the following result:

$$
\mathbb{P}\left(T_2 \ge \epsilon\right) \le \exp\left(\frac{-2\epsilon^2}{\sum_{p=1}^{K}|w_p(L)|^2|c_u - c_\ell|^2}\right). \qquad (8)
$$

We can easily bound $\sum_{p=1}^{K}|w_p(L)|^2$ in (8) similar to the bounding techniques used in bounding T_1 and show that

$$
\sum_{p=1}^{K}|w_p(L)|^2 \le \frac{\widehat{p}_L}{N}\left(\mu(A)\left[N\widehat{p}_L - 1\right] + 1\right). \qquad (9)
$$

By substituting (9) in (8) and by equating the right hand side of (8) to $\delta_L \in (0, 1/2)$ and solving for ϵ, we can show that, with probability at least $1 - \delta_L$,

$$
T_2 \le \sqrt{\frac{\log(1/\delta_L)|c_u - c_\ell|^2\widehat{p}_L\left(\mu(A)\left[N\widehat{p}_L - 1\right] + 1\right)}{2N}}. \qquad (10)
$$

Applying the bounds in (7) and (10) to (5) we can see that with probability at least $1 - \delta_L$, the following holds:

$$
\begin{aligned}
R(L) - \widehat{R}(L) &\le \left(\mu(A)\widehat{p}_L\|f\|_1 - \frac{\mu(A)}{N}\sum_{i\in L}|f_i|\right) \\
&+ \sqrt{\frac{\log(1/\delta_L)|c_u - c_\ell|^2\widehat{p}_L\left(\mu(A)\left[N\widehat{p}_L - 1\right] + 1\right)}{2N}}.
\end{aligned}
$$

Thus for a given $S \in \mathcal{S}_M$, the risk difference $R(S) - \widehat{R}(S)$ is upper bounded by summing the bound corresponding to each leaf separately. Since $\sum_{L\in\pi(S)}\widehat{p}_L = 1$ and $\sum_{L\in\pi(S)}\sum_{i\in L}|f_i| = \|f\|_1$ we have

$$
\begin{aligned}
R(S) - \widehat{R}(S) &\le \mu(A)\left(\frac{N-1}{N}\right)\|f\|_1 \\
&+ \sum_{L\in\pi(S)}\sqrt{\frac{\log(1/\delta_L)|c_u - c_\ell|^2\widehat{p}_L\left(\mu(A)\left[N\widehat{p}_L - 1\right] + 1\right)}{2N}}.
\end{aligned}
$$

with high probability. If we let $\delta_L = \delta 2^{-(\llbracket L\rrbracket+1)}$ where $\llbracket L\rrbracket$ is the number of bits required to uniquely encode the position of leaf L, then we can follow the proof of Lemma 2 in [1] to show that the bound above holds for every $S \in \mathcal{S}$, which leads to the result of Theorem 1.

6. REFERENCES

[1] R. Willett and R. Nowak, "Minimax optimal level set estimation," *IEEE Trans. Image Proc.*, pp. 2965–2979, 2007.

[2] *IEEE Signal Processing Mag., Special Issue on Compressive Sampling*, vol. 25, no. 2, Mar. 2008.

[3] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.*, pp. 289–307, Aug. 2010.

[4] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inform. Theory*, pp. 5758–5772, Dec. 2009.

[5] C. Genovese, J. Jin, and L. Wasserman, "Revisiting marginal regression," submitted. [Online]. Available: arXiv:0911.4080v1

[6] C. McDiarmid, "On the method of bounded differences," in *Surveys in Combinatorics*, 1989, pp. 148–188.