Tree Partitioning Compression of Piecewise Polynomial Functions

Arian Maleki
Department of Electrical Engineering
Stanford University
arianm@stanford.edu

Morteza Shahram
Department of Statistics
Stanford University
mshahram@stanford.edu

***Abstract*— In this paper compression of piecewise polynomial functions is considered. The importance of this class of functions is due to the fact that most of the signals observed in the nature are piecewise smooth and piecewise polynomials provide an efficient model for representing piecewise smooth signals. We compute the ϵ-entropy of the class of piecewise polynomials and show that the rate-distortion behavior of the traditional compression algorithms is far from optimum. Then we will discuss the binary partitioning algorithm, and will demonstrate that with a simple modification of this algorithm, it is possible to get close to the optimal performance. Moreover, this modification will retain all the advantages of the binary tree algorithms such as simplicity, automatic parameter selection and extension to higher dimensions.**

I. Introduction

Most of the signals observed in the nature are piecewise smooth. One of the simple and yet efficient models for representing smooth signals is the class of piecewise polynomials [1], [2], [3]. For example this class of functions can be used to model the edges in natural and cartoon images [4], [5] . In this paper, this class of functions is studied. The main goal is to find a simple and practical compression scheme with close to optimal performance. In order to compare different compression schemes we will use the concept of rate-distortion. Intuitively speaking, the goal of rate-distortion is to find the minimum achievable bit rate at a given distortion level. There are two closely related rate-distortion theories: Shannon's rate-distortion [6] and Kolmogorov's ϵ-entropy [7]. Shannon's theory deals with stochastic sources. It ignores the processes that have very small probabilities and codes the rest. Kolmogorov's ϵ-entorpy, on the other hand, deals with deterministic sources and enforces the algorithm to code the signals such that the distortion of any signal after compression is less than the desired level. Since we do not have any probabilistic model for the space of piecewise polynomials and all the functions are equally important in our analysis, we will use Kolmogorov's ϵ-entropy instead of Shannon's well known rate-distortion theory. It should also be mentioned that there is an interesting connection between these two theories which is explained in [8].

There exists some related work in the literature as well. Prandoni et al. [1] proved that the traditional bases such as Fourier and wavelet [9] are not optimal for compressing piecewise polynomial functions. The proof is based on the fact that there exists an oracle method with better rate-distortion behavior. Their algorithm is based on an exhaustive search in the space of piecewise polynomials and it cannot be used in practice and therefore they have proposed a dynamic programming algorithm to implement their method. However, despite using dynamic programming, the computational complexity is still very high. Besides, their algorithm can not be extended to higher dimensional signals. In order to address these problems binary tree partitioning (or in two dimensions quad-tree partitioning [4]) algorithm with lagrangian cost function has been proposed [2]. In spite of all the advantages of binary trees, such as simplicity, easy extension to higher dimensions [4], [5] and automatic parameter selection [2], [10], it will be shown that the performance of binary partitioning (derived in [2]) is far from the ϵ-entropy bound that we will derive in this paper. Shukla et al. [2] improved the performance of binary trees by adding a join step to the algorithm. In order to have all the advantages of binary partitioning it is more interesting to modify binary partitioning algorithm, instead of adding a second stage to it. It will be shown that binary trees are able to perform closely to ϵ-entropy bounds by just a few small modifications.

The organization of this paper is as follows: in section II, some properties of the space of piecewise polynomials, such as compactness, will be proved. In section III, some bounds for the ϵ-entropy of piecewise polynomial functions will be derived. In section IV, the binary partitioning algorithm will be discussed in a general framework and finally in the last section we propose a way to improve the performance of binary trees.

II. Characteristics of The Space of Piecewise Polynomials

Consider the space of piecewise polynomial functions on an interval I to be represented by $\mathcal{P}_N^Q(I)$ in which N is the maximum degree of the polynomials on the interval and Q is the maximum number of singularities. A singularity is a point at which the function is not infinitely differentiable. Let $\mathcal{BP}_N^Q(I,A)$ denote the set of all piecewise polynomials of maximum degree N over I with Q singularities which are bounded by A, i.e. ,

$$\mathcal{BP}_N^Q(I,A) = \{f \in \mathcal{P}_N^Q(I) \mid \sup_{x\in I}|f| \leq A\}. \tag{1}$$

In this section we will prove that $\mathcal{BP}_N^Q(I,A)$ is compact and so ϵ-entropy can be defined for it.

Theorem 2.1: Consider a sequence of functions $\{f_n(x)\}_{n=1}^{\infty} \subset \mathcal{BP}_N^0(I,A)$ and assume that this sequence is convergent in $L^2(I)$ sense to a function f. Then $f \in \mathcal{BP}_N^0(I,A)$; In other words $\mathcal{BP}_N^0(I,A)$ is closed.

Proof: Let $\{\phi_i(x)\}_{i=0}^N$ be an orthonormal basis for the space $\mathcal{P}_N^0(I)$. Each element of the sequence can be written as

$$f_m(x) = \sum_{i=0}^{N} a_i^{(m)}\phi_i(x).$$

Since f_m is convergent, it will be a Cauchy sequence, i.e., for every ϵ there exist an M such that $\forall\ m,n > M$, $\|f_m - f_n\| \leq \epsilon$; Also,

$$\|f_m - f_n\|^2 = \sum_{i=0}^{M} |a_i^{(m)} - a_i^{(n)}|^2.$$

Therefore, each $a_i^{(m)}$ is a Cauchy sequence in $\mathbb{R}$ and converges to some number a_i. It can be confirmed that $\lim_{m\to\infty} f_m \ =$

$\sum a_i\phi_i(x)$, since $\|f_m - f\|^2 = \sum_{i=0}^{M} |a_i^{(m)} - a_i|^2$. The bound on the polynomials also holds and the proof is complete. ■

It should be noted that since this space is finite dimensional, the convergent sequence under any norm is convergent under L^2 norm and therefore this space is closed under all the norms that can be defined for it.

Theorem 2.2: The space $\mathcal{BP}_N^Q(I,A)$ as a subset of $L^2(I)$ is compact.

Proof: A subset of a metric space is compact if and only if, it is totally bounded and complete. In order to prove that $\mathcal{BP}_N^Q(I,A)$ is totally bounded, we first divide the whole interval into $M_1 = \frac{1}{\epsilon^2}16|I|(Q+1)^2A^2$ equispaced subintervals and the functions that have their Q singularities on these points are considered. Now Assume that $d_1, d_2, \dots d_Q$ are chosen among these points; The functions from $\mathcal{BP}_N^Q(I,A)$ that have singularities at $d_1, d_2, \dots d_Q$ can be written as

$$f(x) = \sum_{j=0}^{Q}\sum_{i=0}^{N} a_i^j \phi_i^j(x)\mathbf{1}_{[d_j,d_{j+1})}, \tag{2}$$

where $\{\phi_i^j(x)\}_{i=0}^N$ is an orthonormal basis for the polynomials of degree up to N on the interval $[d_j, d_{j+1})$ and $\mathbf{1}$ is the indicator function. $|a_i^j| \le A\sqrt{|I|}$ since the L_2 norm of the function is less than $A\sqrt{|I|}$. For each coefficient a_i^j, we consider $M_2 = \frac{1}{\epsilon}4A(Q+1)\sqrt{(N+1)}$ discrete uniformly distributed values. Therefore, for any choice of $d_1, d_2, \dots d_Q$, $M_2^{(N+1)(Q+1)}$ functions are chosen from this space. In other words, for any ϵ we are considering $\binom{M_1}{Q}M_2^{(N+1)(Q+1)}$ functions in the space $\mathcal{BP}_N^Q(I,A)$. The set of all these functions is called S. Now the claim is

$$\sup_{f\in\mathcal{BP}_N^Q(I,A)} \min_{\hat{f}\in S} \|f-\hat{f}\| \le \epsilon.$$

Let f be an arbitrary function in $\mathcal{BP}_N^Q(I,A)$. Assume that the singularity points of f are $d_1, d_2, \dots d_Q$ and find Q singularity points $\hat{d}_1, \hat{d}_2, \dots, \hat{d}_Q$ on the chosen grid such that $\hat{d}_i \le d_i$ and among the grid points that satisfy this property $\hat{d}_i$ is the closest to d_i (if the number of singularities is less than Q, we set some of them to zero). On each interval $[d_{i-1}, \hat{d}_i)$ we find the polynomial which is the closest to the polynomial of f and call it $\hat{P}_i$. $\hat{f}$ is a piecewise polynomial function generated from these discretized singularities and polynomials, i.e.,

$$\hat{f}(x) = \sum_{i=1}^{Q+1} \hat{P}_i(x)\mathbf{1}_{[\hat{d}_{i-1},\hat{d}_i)}, \tag{3}$$

where $d_0 = \hat{d}_0$ is defined as the starting point of the interval and $d_{Q+1} = \hat{d}_{Q+1}$ is the end point of the interval I. Therefore,

$$\begin{aligned}&\|f(x)-\hat{f}(x)\|_{L^2(I)} \le \\ &\sum_{i=1}^{Q+1}\|f(x)-\hat{f}(x)\|_{L^2[\hat{d}_{i-1},d_{i-1}]} \\ &+\sum_{i=1}^{Q+1}\|f(x)-\hat{f}(x)\|_{L^2[d_{i-1},\hat{d}_i]},\end{aligned} \tag{4}$$

where $\|f\|_{L^2(I)} = (\int_I f^2)^{(\frac{1}{2})}$.

The first term on the righthand side of the inequality can be bounded by $2A\sqrt{|d_i - \hat{d}_i|}$ which is less than $\frac{\epsilon}{2Q+2}$. The second term include the norm of two polynomials subtracted from each other. As a result, this norm is also bounded by $\frac{\epsilon}{2Q+2}$ and the summation would be less than ϵ.

For proving the completeness, since $\mathcal{BP}_N^Q(I,A)$ is a subset of a Banach space it is sufficient to prove that it is closed [11]. To prove the closedness let $\{f_n(x)\}_{n=1}^\infty$ be a sequence in this space that converges to $f(x)$. Let the sequence of the first singularity points be $\{d_m^1\}_{m=1}^\infty$. This sequence is chosen in a compact subset of $\mathbb{R}$. Therefore, it has a convergent subsequence [11] that converges to some point called d_1. We take the subsequence $\{f_{n_k}^1\}_{k=1}^\infty$ of $\{f_n\}_{n=1}^\infty$ such that their first singularity point converges to d^1 and look at their second singularity points. Again there exist a subsequence of this new sequence whose second singularity point converge to a point in the interval called d^2. This sequence is called $\{f_{n_k}^{(2)}\}_{k=1}^\infty$. By repeating the same algorithm on all the singularity points one can get a subsequence of the main sequence in which $(d_m^1, d_m^2, \dots d_m^Q)$ converge to a Q dimensional vector called $(d^1, d^2, \dots, d^Q)$in the sense of some norm in $\mathbb{R}^Q$(regardless of the norm used). Let this subsequence be $\{g_k\}_{k=1}^\infty$. The claim is that $\lim_{k\to\infty} g_k$ is in $\mathcal{BP}_N^Q(I,A)$. Now, consider the interval I_i defined as: $[d_{i-1}+\epsilon, d_i - \epsilon]$. Here, ϵ is a small positive constant. For every i there exist a finite integer M_i such that for $k > M_i$, there is no singularity in I_i. Therefore, in this interval some polynomials of degree at most N exist that converge to a function which according to theorem 2.1 is a polynomial itself and is called $P_i(x)$. Let $P(x)$ be

$$P(x) = \sum_{i=1}^{Q+1} P_i(x)\mathbf{1}_{[d_{i-1},d_i)}. \tag{5}$$

$P(x)$ is in fact the limit of $g_k(x)$, since

$$\begin{aligned}&\|g_k(x)-P(x)\|_{L^2(I)} \le \\ &\sum\|g_k(x)-P(x)\|_{L^2[\min(d_{i-1},d_{i-1}^k),\max(d_{i-1},d_{i-1}^k)]} \\ &+\sum\|g_k(x)-P(x)\|_{L^2[\max(d_{i-1},d_{i-1}^k),\min(d_i,d_i^k)]}\end{aligned} \tag{6}$$

and by choosing large enough k, each of the terms on the right hand side of equation (6) can be chosen arbitrarily small and so $\lim_{k\to\infty}\|g_k(x)-P(x)\|_{L^2(I)} = 0$. On the other hand $\lim_{k\to\infty} g_k = f$ and according to the uniqueness of the limit, $P(x) = f(x)$ and $f(x)$ is a piecewise polynomial function. ■

Now that the compactness of this space is proved we can use this fact to define Kolmogorov's ϵ-entropy for this space and derive some bounds for this entropy to see how much an encoder can compress this class of functions at distortion ϵ.

III. Kolmogorov's ϵ-Entropy of the Class $\mathcal{BP}_N^Q(I,A)$

Let $F \subset L^2([0,1])$ be a compact class of functions. The goal is to compress the functions in this space such that the distortion is less than ϵ and the question is: what is the least number of bits needed to code the functions in this space? The answer to this question is given by Kolmogorov [7], [12]. The *ϵ-entropy* of a compact set F is shown with $H_\epsilon(F)$ and is defined as

$$H_\epsilon(F) = \log_2 N_\epsilon(F), \tag{7}$$

where $N_\epsilon(F)$ is the minimal number of points in an ϵ-covering of F [7], [11]. In other words, for coding an element of F with the distortion at most equal to ϵ the best achievable rate is $H_\epsilon(F)$.

In the previous section we showed that the space $\mathcal{BP}_N^Q(I,A)$ is compact and therefore, ϵ-entropy can be defined for this space. By considering the L_2 norm as the metric, the following theorems can be proved.

Theorem 3.1: There exist two positive constants B_1 and B_2 such that the ϵ-entropy of the space $\mathcal{BP}_N^0([0,1],A)$ satisfies the following inequalities

$$B_1 + \log(\frac{1}{\epsilon}) \leq \frac{1}{N+1} H_\epsilon(\mathcal{BP}_N^0([0,1], A)) \leq B_2 + \log(\frac{1}{\epsilon}),$$

where $B_2 = \log\left(\frac{A\sqrt{N+1}}{2}\right)$ and $B_1 = \log\left(\frac{A}{C}\right)$, in which C is the norm equivalence constant appearing in (9).

Proof: Let $\{\phi_i(x)\}_{i=0}^N$ be an orthonormal basis for $\mathcal{P}_N^0([0,1])$. Consider the following space:

$$\mathcal{SP}([0,1], A) = \{f \mid f = \sum_{i=0}^N a_i \phi_i(x) \ , \sum_{i=0}^N |a_i|^2 \leq A^2\}. \quad (8)$$

It can be confirmed that $\mathcal{BP}_N^0([0,1], A) \subset \mathcal{SP}([0,1], A)$. Therefore, an ϵ-covering of this new space will also serve as an ϵ-covering of $\mathcal{BP}_N^0([0,1], A)$. $\mathcal{SP}([0,1], A)$ is isometric with $B^{N+1}(\mathbf{0}, A)$ ball in $\mathbb{R}^{N+1}$,

$$B^{N+1}(\mathbf{x_o}, A) = \{\mathbf{x} \in \mathbb{R}^{N+1} \mid \|\mathbf{x} - \mathbf{x_o}\|_{\ell_2} \leq A\},$$

where the norm of $\mathbf{x} = (x_0, \ldots, x_N)$ is defined as $(\sum |x_i|^2)^{1/2}$. Because of the bound on the norm of $\mathbf{x}$, $-A \leq x_i \leq A$. Consider dividing [-A,A] interval into $M = \frac{2A\sqrt{N+1}}{\epsilon}$ equispaced subintervals for each x_i. Balls of radius ϵ with centers on the grid which are constructed by these partitions are an ϵ-covering of the space and for each point in the $B^{N+1}(\mathbf{0}, A)$ there exist at least one point on the grid with less than ϵ distance from the given point. The number of balls which are considered, is $M^{(N+1)}$ which is equal to $B_2(\frac{1}{\epsilon})^N$. In order to find a lower bound we first note that since $\mathcal{SP}([0,1], A)$ is larger than $\mathcal{SP}_N^0([0,1], A)$, it cannot be used for deriving a lower bound. Nevertheless since the space of polynomials of degree at most N is finite dimensional, all the norms are equivalent. In this special case it means that there exist two constants c and C such that:

$$c\|f\|_2 \leq \|f\|_\infty \leq C\|f\|_2. \quad (9)$$

In other words, $\mathcal{SP}([0,1], A/C) \subset \mathcal{BP}_N^0([0,1], A)$ and $\mathcal{SP}([0,1], A/C)$ is isometric to $B^{N+1}(\mathbf{0}, A/C)$. In order to find the minimum number of balls to cover $B^{N+1}(\mathbf{0}, A/C)$ we use the fact that the total volume of ϵ-covering of this ball is greater than the volume of the ball itself. Therefore, if there are M balls of radius ϵ in the cover,

$$M \frac{\pi^{N/2}}{\Gamma(N/2+1)} \epsilon^{N+1} \geq \frac{\pi^{N/2}}{\Gamma(N/2+1)} \left(\frac{A}{C}\right)^{N+1},$$

where Γ is the gamma function. According to this equation, $M \geq (\frac{A}{C})^{N+1}(\frac{1}{\epsilon})^{N+1}$, and the proof is complete. ■

Theorem 3.2: The ϵ-entropy of the space $\mathcal{BP}_N^Q([0,1], A)$ satisfies the following two constraints

$$C_1 + (N+1)(Q+1)\log(\frac{1}{\epsilon}) \leq H_\epsilon(\mathcal{BP}_N^Q([0,1], A)) \leq C_2 + (N+3)(Q+1)\log(\frac{1}{\epsilon}), \quad (10)$$

where

$$C_1 = (N+1)(Q+1)\log\left(\frac{A}{C'}\right),$$

$$C_2 = (N+3)(Q+1)\log\left(\frac{A(Q+1)(N+3)}{(2\sqrt{N+1})^{(N+1)/(N+3)}}\right)$$

and C' is the L_2-norm and L_∞-norm equivalence constant.

Proof: In order to prove the upper bound, assume that the $[0,1]$ is partitioned into n_1 equispaced subintervals. We choose Q points out of these n_1 points in $\binom{n_1}{Q}$ different ways. For each chosen point, the point next to it will also be considered. Let's call these points $0 = \hat{d}_0, \hat{d}_1, \ldots, \hat{d}_{2Q+1} = 1$. For each interval $[\hat{d}_{2i}, \hat{d}_{2i+1})$, n_2 polynomials will be picked. Therefore, the total number of balls is: $\binom{n_1}{Q}\ {n_2}^{Q+1}$. Let $\mathcal{D}$ be the set of all these functions and f be an arbitrary function in the space $\mathcal{BP}_N^Q([0,1], A)$. The distortion in approximating this function is $\|f - \hat{f}\|$ where $\hat{f} \in \mathcal{D}$. The distortion can be upper bounded by

$$\|f(x) - \hat{f}(x)\|_{L^2(I)} \leq \sum_i \|f(x) - \hat{f}(x)\|_{L^2[\hat{d}_{2i+1}, \hat{d}_{2i+2}]} + \sum_i \|f(x) - \hat{f}(x)\|_{L^2[\hat{d}_{2i}, \hat{d}_{2i+1}]}. \quad (11)$$

Let us assume that $\hat{f}$ is chosen from $\mathcal{D}$ with the following method: First the decision is made on the $\hat{d}_{2i-1}$. $\hat{d}_{2i-1}$ is chosen such that it is less than or equal to d_i and among all $\hat{d}$ that satisfy this condition is the closest to the point d_i. Therefore we have: $|d_i - \hat{d}_{2i-1}| \leq 1/n_1$. Then in the intervals $[\hat{d}_{2i}, \hat{d}_{2i+1}]$ we find the closest polynomial in that part which belongs to $\mathcal{D}$. Hence each term on the second line of (11) is less than or equal to $2A\sqrt{1/n_1}$. Also according to the previous theorem the approximation error could be made less than $\frac{A}{2}\sqrt{N+1}\left(\frac{1}{n_2}\right)^{\frac{1}{N+1}}$. Therefore, the total error is bounded by

$$\|f - \hat{f}\| \leq (Q+1)A\left[\frac{\sqrt{N+1}}{2}\left(\frac{1}{n_2}\right)^{\frac{1}{N+1}} + 2\left(\frac{1}{n_1}\right)^{\frac{1}{2}}\right] \leq \epsilon. \quad (12)$$

The last inequality is imposed in order to keep the distortion less than ϵ. Now the problem is how to choose n_1 and n_2. In other words, n_1 and n_2 should be chosen such that they satisfy (12) and minimize $\binom{n_1}{Q}\ {n_2}^{Q+1}$. Since the goal is to find an upper bound for the number of balls this expression will be replaaced by ${n_1}^{Q+1}{n_2}^{Q+1}$ and the problem is simplified to

$$\begin{aligned} \text{minimize} \quad & n_1^{Q+1} n_2^{Q+1} \\ \text{s.t.} \quad & (Q+1)A\left[\frac{1}{2}\sqrt{N+1}\left(\frac{1}{n_2}\right)^{\frac{1}{N+1}} + \frac{2}{\sqrt{n_1}}\right] \leq \epsilon. \end{aligned} \quad (13)$$

Instead of solving this problem for integer numbers, we assume that n_1 and n_2 are not necessarily integers. Then the results will be rounded to the closest greater integer. The new problem is geometric programming [13]. By some calculations the upper bound can be found.

For proving the lower bound consider a subset of $\mathcal{BP}_N^Q([0,1], A)$ consisting of all functions in $\mathcal{BP}_N^Q([0,1], A)$ that have singularities at $\{\frac{1}{Q+1}, \frac{2}{Q+1}, \ldots \frac{Q}{Q+1}\}$. Since this is a $(N+1)(Q+1)$ dimensional space, according to the equivalence of L_∞ and L_2 norms of any finite dimensional space, i.e. ,

$$c'\|f\|_{L2} \leq \|f\|_{L_\infty} \leq C'\|f\|_{L2} \quad (14)$$

and the fact that the L_∞-norm of these signals is bounded, one can find an L_2 ball in this space. Using the same argument used in theorem 3.1 at least $2^{C_1}(\frac{1}{\epsilon})^{(N+1)(Q+1)}$ balls will be necessary to code this ball, where

$$C_1 = (N+1)(Q+1)\log\left(\frac{A}{C'}\right) \quad (15)$$

and the proof of the theorem is complete. ■

In the last two theorems, ϵ is playing the role of maximum distortion and the number of balls is related to the number of bits that should be used for compressing the signal at the ϵ distortion level. But it is also possible to look at this problem from a different point of view. Assume that the number of bits that can be used for representing the signals is fixed at some bit rate R and the goal is to minimize the distortion.

Theorem 3.3: If the number of bits for representing the signals in $\mathcal{BP}_N^0([0,1],A)$ is fixed at R, the best achievable L^2 norm distortion satisfies the following inequalities.

$$c_1 2^{\frac{-R}{N+1}} \leq D(R) \leq c_2 2^{\frac{-R}{N+1}}, \tag{16}$$

where $c_2 = \frac{A\sqrt{N+1}}{2}$, and $c_1 = \frac{A}{C}$.

Proof: this will be a direct consequence of theorem 3.1. ∎

Theorem 3.4: For a fixed number of bits R the least achievable distortion satisfies the following inequalities

$$k_1 2^{\frac{-R}{(N+1)(Q+1)}} \leq D(R) \leq k_2 2^{\frac{-R}{(N+3)(Q+1)}}, \tag{17}$$

where $k_2 = A(Q+1)(N+3)$ and $k_1 = \frac{A}{C'}$.

Proof: It follows directly from theorem 3.2, by noting that $\left(2\sqrt{N+1}\right)^{(N+1)/(N+3)} \geq 1$. ∎

The goal of the next two sections is to find a near-optimal method for compressing signals from this class of functions.

IV. Binary Partitioning Algorithm

As explained earlier, wavelet and the other traditional algorithms cannot perform well on this class of functions. In order to solve the problem Shukla et al. [2] have proposed an interesting algorithm which is called Binary Tree Partitioning. The idea of this algorithm is somewhat similar to that of wedgelet [4]. By using binary partitioning or (in 2-D quad Partitioning) these two algorithms partition the signal into some segments and then code each segment with a polynomial(or wedge like patches in wedgelet). The algorithm works as following: Let $P^j = \{I_1^j, I_2^j, \dots I_L^j\}$ be a partition of the interval $[0,T)$ at level j with the following characteristics: first, each $I_i^j \subset [0,T)$ is a half open interval e.g., $[d_{i-1}, d_i)$. Second,

$$\bigcup_k I_k^j = [0,T) \ , \ I_k^j \bigcap I_i^j = \emptyset \ \forall k \neq i. \tag{18}$$

For each interval I_k^j a cost function is defined as:

$$c_k^j = \left\| [x(t) - P_o(t)] \mathbf{1}_{I_k^j}(t) \right\|_{L_2([0,T])} + \lambda^2 R_k^j, \tag{19}$$

where $P_o(t)$ is the best polynomial of degree N that minimizes the first term of the cost function, R_k^j is the required rate for coding that part of the tree and λ is a real number. At level j, we divide each I_k^j with $|I_k^j| = T2^{-j}$ at its midpoint to two smaller intervals I_{2k-1}^{j+1} and I_{2k}^{j+1} if:

$$c_k^j > c_{2k-1}^{j+1} + c_{2k}^{j+1}. \tag{20}$$

By starting from $P^0 = \{[0,T)\}$ and applying the same algorithm to the partitions iteratively one will get to the point that none of the partitions can be divided more. The final partition is the one which is used in coding. The intervals that are divided into two subintervals are called internal nodes and the intervals that cannot be divided are called terminal nodes or leaves. This terminology will be used during this paper. Let us summarize the binary tree partitioning algorithm in a different way that will lead us to a more general framework that can be used later for improvement of binary trees. Assume that at each scale j, a set of signals is given $\Phi^j = \{\phi_k^j\}_{k=1}^{K_j}$. At each level one can replace a signal in one of the intervals with one of the signals in the corresponding set in that scale. In the case of binary partitioning algorithm explained earlier each dictionary includes a discrete version of the polynomials of degree N that are defined in the interval of size $T2^{-j}$. It can be proved that for this algorithm $D_{BT}(R) \geq \alpha_0 \sqrt{R} 2^{-\alpha_1 \sqrt{R}}$ (α_0 and α_1 are just dependent on N and Q and size of the interval) which is far from the distortion rate bounds. Since the proof of this theorem is very similar to the proof given in [2] we ignore this proof. The question that will be answered in section V, is to find a better set Φ^j to code the signals in $BPoly_N^Q(I,A)$.

V. Enriching the Dictionaries

As mentioned in section IV, the performance of binary tree algorithm with polynomial sets is far from optimal distortion rate performance. The reason is that no matter how deep the algorithm goes in the tree the singularity points are not presented very well; In other words, singularities cannot be represented with polynomials. Therefore, in order to improve the performance of the binary tree algorithm the sets Φ^j should be enriched with some notion of singularity. In this section the polynomial set is replaced by this new set:

$$\begin{aligned}\Phi_c^j =& \{f : [0, T2^{-j}] \to \mathbb{R} \mid \exists\, t_o, P_1(t), P_2(t);\\ \text{s.t.} \quad & f(t) = P_1(t) 1_{(0<t<t_o)} + P_2(t) 1_{(t_0<t<T2^{-j})}\},\end{aligned} \tag{21}$$

in which t_o is a point in the interval and P_1, P_2 are two polynomials of degree N. As it can be seen the indicator function introduces the singularity to the set. The main set is the quantized version of this set in which t_o and the projections of P_1 and P_2 on the Legendre polynomials are quantized with uniform scaler quantizer.

Definition: *Homogeneity Coefficient* is defined as:

$$\rho = \frac{\min_{0\leq i \leq Q} |d_{i+1} - d_i|(Q+1)}{T}, \tag{22}$$

where d_0 is assumed to be the starting point of the interval and $d_{(Q+1)}$ the end point of the interval.

Lemma 5.1: $0 \leq \rho \leq 1$ and it is equal to 1 if and only if all the discontinuity points are equispaced.

Proof: Let $\Delta_i = |x_{(i+1)} - x_i|$. The problem of maximizing ρ is equivalent to:

$$\begin{aligned}& \max \min \Delta_i \\ s.t. & \sum \Delta_i = T.\end{aligned} \tag{23}$$

If $\min \Delta_i = \Delta$ then, $T = \sum \Delta_i \geq (Q+1)\Delta$ therefore, $\Delta \leq T/(Q+1)$. The equality can hold if all the Δ_is are equal to each other and this is the optimal solution. ∎

Theorem 5.2: Assume that $f(t) : [0,T] \to \mathbb{R}$ is in $\mathcal{BP}^1([0,T],A)$ and that Φ^j is the quantized version of Φ_c^0 at Rate R (quantization is uniform for both the singularity point and the two polynomials. But the rates assigned to each of these three uniform quantizers is chosen optimally). The distortion is:

$$D(R) \leq 2A\sqrt{T}(N+3)2^{-(\frac{1}{N+3})(\frac{R}{2})}. \tag{24}$$

Proof: The proof of this theorem is very similar to the proof of (17). ∎

In the rest of our calculations we consider a subclass of $\mathcal{BP}_N^Q([0,T],A)$ in which the signals have the property $\rho_{min} \leq \rho \leq 1$. This subclass will be shown by $\mathcal{BP}_N^Q([0,T],A,\rho_{min})$. It is not difficult to prove that this subclass is also a compact subset of $L^2([0,T])$ and the same entropy bounds hold for the ϵ-entropy of this class of functions. It should also be mentioned that for natural signals the assumption we imposed on the homogeneity coefficients is a realistic assumption.

Lemma 5.3: At each level of the tree at most Q of the nodes will be partitioned and the maximum depth of the tree is $\lceil \log_2 \frac{Q+1}{\rho_{min}} \rceil$.

Proof: Only the intervals that have at least two singularity points can be divided. Since there are at most Q of those intervals in each partition, at most Q of the nodes will be divided more. At depth $\lceil \log_2 \frac{Q+1}{\rho_{min}} \rceil$ size of the smallest interval is less than $\frac{T\rho_{min}}{Q+1}$ which is less than the distance of any two singularity points. Therefore none of these intervals have two or more singularities and they will not be divided any more. ∎

Theorem 5.4: The binary partitioning algorithm with set Φ_c^j at high bit rate R can achieve the distortion:

$$D(R) \leq c_1 2^{-(\frac{1}{N+3})(\frac{R}{2(-\log_2 \rho_{min}+1)(Q+1)})} \tag{25}$$

on the class of functions $\mathcal{BP}_N^Q([0,T], A, \rho_{min})$.

Proof: We will divide the leaves of the tree into two sets. Some of the leaves are those that correspond to the scale $j < J_{min} = \lceil \log_2 Q + 1 \rceil$ and the rest are those with scale between J_{min} and J_{max}. The first set is called $S_{J_{min}}$. $|S_{J_{min}}| \leq 2^{J_{min}}$ and the distortion of each one satisfies:

$$D(R) \leq 2A\sqrt{T}(N+3)2^{-(\frac{1}{N+3})(\frac{R}{2})}. \tag{26}$$

On the other hand at each scale $j > J_{min}$ according to lemma 5.3, there are Q leaves and the rate and distortion of each leave satisfy:

$$D(R) \leq 2A\sqrt{\frac{T}{2^j}}(N+3)2^{-(\frac{1}{N+3})(\frac{R}{2})}. \tag{27}$$

Therefore, by assigning $\{R_{0_i}\}_{i=1}^{2^{J_{min}}}$ to the leaves in $S_{J_{min}}$ and $\{R_{j_k}\}_{j=J_{min}}^{J_{max}}$ $(1 \leq k \leq Q$)to the other leaves, the distortion would be equal to:

$$\begin{aligned} D(\mathbf{R}) = \sum_{k=1}^{2^{J_{min}}} 2A\sqrt{T}(N+3)2^{-(\frac{1}{N+3})(\frac{R_{0_k}}{2})} + \\ \sum_{j=J_{min}}^{J_{max}} \sum_{k=1}^{Q} 2A\sqrt{\frac{T}{2^k}}(N+3)2^{-(\frac{1}{N+3})(\frac{R_{j_k}}{2})}, \end{aligned} \tag{28}$$

in which $\mathbf{R}$ is the vector that includes all the rates. There is also a constraint on the bit rate:

$$\sum_{j=J_{min}}^{J_{max}} \sum_{k=1}^{Q} R_{j_k} + \sum_{i=1}^{2^{J_{min}}} R_{0_k} \leq R. \tag{29}$$

By using Karush-Kuhn-Tucker (KKT) conditions [13], we get $R_{0_k} = R_{0_l}$ and $R_{j_k} = R_{j_l}$ for every value of k, l, j and therefore, by defining $R_{0_k} = R_0$ and $R_{j_l} = R_j$ (28) and (29) can be written as:

$$\begin{aligned} D(\mathbf{R}) = & 2^{J_{min}} 2A\sqrt{T}(N+3)2^{-(\frac{1}{N+3})(\frac{R_0}{2})} + \\ & (Q+1) \sum_{j=J_{min}}^{J_{max}} 2A\sqrt{\frac{T}{2^k}}(N+3)2^{-(\frac{1}{N+3})(\frac{R_j}{2})} \end{aligned} \tag{30}$$

and

$$\sum_{j=J_{min}}^{J_{max}} QR_j + 2^{J_{min}} R_0 \leq R. \tag{31}$$

KKT conditions for this simplified problem will result in (for high bit rate regime):

$$\begin{aligned} R_0 - R_{J_{min}} &= \frac{2J_{min}}{\frac{2}{N+3}} \\ R_k - R_{J_{min}} &= -\frac{2(k - J_{min})}{\frac{2}{N+3}}. \end{aligned} \tag{32}$$

By combining (31) and (32) we get:

$$R_{J_{min}} = \frac{R}{(Q+1)(-\log_2 \rho + 1)} + \alpha, \tag{33}$$

where α is a constant without dependence on R and D. By replacing (33) in (28) the final rate-distortion of this algorithm will be found. ∎

We see that adding the singularity to the set helps improve the performance from square root to linear (in terms of rate) for the subclass $\mathcal{BP}_N^Q([0,T], A, \rho_{min})$.

VI. Conclusion

In this paper we study the tree partitioning framework for compressing piecewise polynomial functions and propose a new approach to enrich the underlying dictionary by introducing the notion of singularity. In this framework, we allow existence of a single singularity per subintervals. We prove that the obtained performance is close to the ϵ-entropy bounds for the class of $\mathcal{BP}_N^Q([0,T], A, \rho_{min})$. One can ask about benefits of accommodating multiple singularities in such a tree partitioning framework. It is straightforward to show that by adding more singularities, we achieve linear decay rate for larger subclasses of piecewise polynomial functions. Although, on the other hand, the computational complexity of the algorithm is increased since it has to search through a larger set.

Extension to the 2-D case is relatively easy. Each function is the representative of an edge in the horizon model [4] and each element of the sets we introduced can be extended to 2-D by simply adding the notion of orientation to its pieces. Since the edges in natural images are usually continuous we can impose this continuity to the sets we have introduced. Therefore we obtain some wedge-like patches with a breakpoint as an extension to the wedgelet patches.

References

[1] P. Prandoni, M. Vetterli, "Approximation and compression of piecewise smooth functions," *Phil. Trans. Roy. Soc. London*, vol. 357, no. 1760, pp. 2573-2591, Sep. 1999.

[2] R. Shukla, P. L. Dragotti, M. N. Do, M. Vetterli, "Rate-distortion optimized tree-structured compression algorithms for piecewise polynomial images," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 1-17, March 2005.

[3] M. N. Do, P. L. Dragotti, R. Shukla, M. Vetterli, "On the compression of two-dimensional piecewise smooth functions," *Proceedings of IEEE International Conference on Image Processing*, Thessaloniki, Greece, Oct. 2001, pp. 14-17.

[4] D. L. Donoho, "Wedgelets: nearly minimax estimation of edges," *Annals of Stat.*, vol. 27, pp. 859-897, 1999.

[5] V. Chandrasekaran, M. B. Wakin, D. Baron, R.G. Baraniuk, "Representation and compression of multi-dimensional piecewise functions using surflets", submitted to IEEE Transaction on Information Theory, 2006.

[6] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[7] A. N. Kolmogorov, V. M. Tihomirov, "ϵ-entropy and ϵ-capacity of sets in functional spaces," *American Math. Society Translation* (Ser. 2), vol. 17, pp. 277-364, 1961.

[8] D. L. Donoho, "Wald lecture 1: counting bits with Shannon and Kolmogorov", Technical Report, Stanford University, 2000.

[9] M. Vetterli, "Wavelets, approximation and compression", IEEE Signal Processing Magazine, pp. 59-73, Sept. 2001.

[10] K. Ramchandran, M. Vetterli, "Best wavelet packet bases in a rate-distortion sense", *IEEE Transaction on Image Processing*, vol. 2, pp. 160-175, 1993.

[11] H. L. Royden, *Real Analysis*, Second Edition, New york:Macmillan, 1968.

[12] A. G. Vitushkin, *Theory of The Transmition and Processing of Information*, Pergamon Press, 1961.

[13] Stephan Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, First edition, 2004.