Zero-error Source Coding with Maximum Distortion Criterion.

Ertem Tuncel, Prashant Koulgi, Shankar Regunathan and Kenneth Rose
Dept. of ECE, UC Santa Barbara, CA 93106, USA

Abstract

Let finite source and reproduction alphabets $\mathcal{X}$ and $\mathcal{Y}$ and a distortion measure $d : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty)$ be given. We study the minimum asymptotic rate required to describe a source distributed over $\mathcal{X}$ within a (given) distortion threshold D *at every sample*. The problem is hence a min-max problem, and the distortion measure is extended to vectors as follows: for $x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$, $d(x^n, y^n) = \max_i d(x_i, y_i)$.

In the graph-theoretic formulation we introduce, a code for the problem is a dominating set of an equivalent *distortion graph*. We introduce a linear programming lower bound for the minimum dominating set size of an arbitrary graph, and show that this bound is also the minimum asymptotic rate required for the corresponding source. Turning then to the optimality of scalar coding, we show that scalar codes are asymptotically optimal if the underlying graph is either an interval graph or a tree.

1 Introduction

Consider a signal compression scenario, where it is unacceptable to have occasional high distortion at the sample level, even if the expected distortion is kept small. The motivation for this more restrictive constraint is that high distortion at individual samples is often perceptually disturbing, and compromises the accuracy of subsequent analysis of the reconstructed signal. Examples include compression of medical images, where detail may be of paramount importance to ensure correct diagnoses. Hence conventional high rate lossy compression techniques which only promise a small *expected* distortion, cannot be used. JPEG-LS [11], a recent image coding standard, guarantees, at every pixel, reconstruction within a specified distortion. Our work is partly motivated by such applications, and studies the optimum theoretically achievable performance of such codes.

Let the finite sets $\mathcal{X}$ and $\mathcal{Y}$ denote the source and the reproduction alphabets, respectively, and let a distortion measure $d : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty)$ be given. Assume that the source distribution P satisfies $P(x^n) > 0$ for all $x^n \in \mathcal{X}^n$. We investigate both finite and asymptotic rate-distortion performance of source coders $\phi : \mathcal{X}^n \longrightarrow \mathcal{Y}^n$, which guarantee that *every* source sample is reproduced within a prespecified distortion level D. This problem is fundamentally different from the problem addressed by classical rate-distortion theory [2] in two ways:

i. *Maximum distortion criterion*: Let x_i and y_i denote the ith coordinate of x^n and y^n, respectively. Given a single-letter distortion criterion $d : \mathcal{X} \times \mathcal{Y} \longrightarrow [0, \infty)$, the distortion between source block $x^n \in \mathcal{X}^n$ and reproduction block $y^n \in \mathcal{Y}^n$ is measured by

$$d(x^n, y^n) = \max_i d(x_i, y_i) , \quad (1)$$

Proceedings of the DATA COMPRESSION CONFERENCE (DCC'02)
1068-0314/02 $17.00 © 2002 IEEE

instead of the averaging measure

$$d(x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) \ , \tag{2}$$

as considered in the classical theory.

ii. *Zero-error constraint*: In many derivations in classical information theory, a small but eventually vanishing error probability is allowed, i.e.,

$$\lim_{n \longrightarrow \infty} \Pr[d(X^n, Y^n) > D] = 0 \ . \tag{3}$$

However, we should enforce a *non-asymptotic* zero-error constraint, i.e., for all $n > 0$,

$$\Pr[d(X^n, Y^n) > D] = 0 \ . \tag{4}$$

The zero-error constraint is a natural complement to the maximum distortion criterion. For finite block coding ($n < \infty$), if the encoder is permitted to occasionally exceed distortion D, albeit with a small probability, then with probability 1, we will eventually observe unacceptable spikes at the reproduction, in contradiction to the main purpose of the maximum distortion criterion.

In this investigation we will focus our attention on fixed-length codes. Further, we assume no knowledge about the underlying source statistics. Fixed-length codes, with zero error for the maximum distortion criterion, yield a robust coding system: the rate of transmission depends only on the distortion measure, and system performance is insensitive to the often-encountered practical problem of changing source statistics. Interestingly, our results imply that this coding system is also asymptotically optimal in the following sense: under the assumption that the source statistics may vary arbitrarily with time, variable-length codes of smaller rate than the minimum asymptotic fixed-length coding rate do not exist.

Throughout, we further assume that $\mathcal{X} = \mathcal{Y}$, and $d(x, x) = 0$. Our results retain their validity when these assumptions are dropped; however, constraints of space force us to omit the corresponding extended proofs.

We begin by reformulating the problem in a graph-theoretic framework. In Section 2, we define a characteristic graph $\mathcal{G}$ for the triplet $(\mathcal{X}, d, D)$. A valid code is then a *dominating set* of $\mathcal{G}$, i.e., a set of vertices the union of whose neighborhoods cover the entire graph [7]. Hence, the domination number (the size of the smallest dominating set) of $\mathcal{G}^n$ (the n-fold AND product of $\mathcal{G}$), characterizes the minimum rate required for block length n. We are interested in the minimum asymptotic rate $R(\mathcal{G})$, given by the limit of the normalized domination numbers of $\mathcal{G}^n$.

In [8], Lovász showed that $R(\mathcal{G}) = r^*(\mathcal{G})$, where $r^*(\mathcal{G})$ is the solution of a linear program obtained by relaxing the integer program defining the domination number. In Section 3, we obtain Lovász's result in the context of a rate-distortion theoretic approach. We introduce a transformation from the maximum distortion measure $d(x^n, y^n)$ to an appropriately defined averaging measure $d'(x^n, y^n)$. Specifically, the averaging measure d' (in the sense of (2)) is such that $d'(x^n, y^n) = 0$ if and only if $d(x^n, y^n) \le D$. Then we show that $R(\mathcal{G})$ is in fact the rate-distortion function of an arbitrarily varying source (considered earlier by Csiszár and Körner in [5]) with this new distortion measure d'. The resultant formula could be simplified to $r^*(\mathcal{G})$.

We next consider conditions for optimality of scalar quantizers. The complexity vs. performance trade-off of vector quantizers is well-known. The simplicity of scalar coding makes it an attractive option in practice. An important question is then the following: Are there cases where scalar coding is optimal, i.e., where scalar coding achieves the minimum *asymptotic* rate? For the maximum-distortion problem under consideration, the graph-theoretic formulation provides a natural setting to study this question: Applying the duality theorem [9] to the linear program defining $r^*(\mathcal{G})$, and then reinstalling the integer constraints, one obtains a lower bound on the domination number, which we call the *tiling number* of the graph. A sufficient condition for optimality of scalar coding is the equality of the domination and tiling numbers of the characteristic graph. Pursuing this approach, in Section 4, we exhibit two infinite families of graphs - interval graphs and forests [1] - for which scalar coding is optimal. The result involving interval graphs is especially interesting, since for the case where $\mathcal{X}$ consists of points on the real line, the most popular distortion measure $d(x, y) = |x - y|^m$ yields an interval graph. This result can be generalized: For a distortion measure such that each source symbol x is allowed to be coded by any reproduction symbol y in the prespecified interval I_x, we prove the optimality of scalar coding. The presented proofs are constructive, and provide *linear-time* algorithms to design optimal codes.

2 Problem Statement, Examples and Observations

Given a source distributed over the alphabet $\mathcal{X}$, the reproduction alphabet $\mathcal{Y}$, and the distortion measure d between these alphabets, we are interested in the minimum rate of a code that represents the source without exceeding a *maximum* distortion of D per input symbol. In other words, for all block lengths n, we enforce

$$d(x^n, y^n) = \max_i d(x_i, y_i) \leq D \ ,$$

for all $x^n \in \mathcal{X}^n$, where y^n denotes the quantized version of x^n.

We will make heavy use of graph theory to attack this problem; we briefly summarize here simple and well-known definitions from graph theory [1]: A graph is denoted by $\mathcal{G} = (V, E)$, where V and E are its vertices and the edges, respectively. Two vertices connected by an edge are called adjacent or neighbor vertices. The complement of a graph $\mathcal{G}$ is denoted by $\overline{\mathcal{G}}$, and consists of the same vertices for which the adjacency relation is complemented. The *neighborhood* $\mathcal{N}_v$ of a vertex $v \in V$ is the set consisting of v and all vertices adjacent to v. A *dominating set* $\mathcal{D}$ is a subset of vertices where any vertex in the graph is in the neighborhood of some vertex in $\mathcal{D}$. The domination number, which we denote by $r(\mathcal{G})$, is defined as the size of the smallest dominating set. If $V' \subset V$ and $E' = \{(v_1, v_2) : v_1, v_2 \in V', (v_1, v_2) \in E\}$, then the subgraph $\mathcal{G}' = (V', E')$ is said to be *induced* by V'.

Let us first consider a scalar coder. We define the characteristic graph $\mathcal{G}$ for the triplet $(\mathcal{X}, d, D)$ as follows: $V = \mathcal{X}$, and two distinct vertices x and y are connected if and only if $d(x, y) \leq D$. A one-to-one correspondence exists between dominating sets of $\mathcal{G}$ and codebooks satisfying the maximum distortion constraint. The minimum codebook size is given by $r(\mathcal{G})$.

For block coding, two distinct vectors x^n and y^n can represent each other if and only if $d(x_i, y_i) \leq D$ for all $0 < i \leq n$. In the corresponding characteristic graph for the triplet $(\mathcal{X}^n, d, D)$, two distinct vertices $(v_1, v_2, \ldots, v_n) \in V^n$ and $(v'_1, v'_2, \ldots, v'_n) \in V^n$ are connected if and only if v_i is adjacent to v'_i in $\mathcal{G}$ for all $0 < i \leq n$ such that $v_i \neq v'_i$. But, this is precisely the definition of the n-fold AND product of the graph $\mathcal{G}$ with itself, denoted by $\mathcal{G}^n$. The number $r(\mathcal{G}^n)$ therefore defines the minimum number of codevectors needed, for block length n.

It is obvious that r is sub-multiplicative over AND products: $r(G^{m+n}) \le r(G^m)r(G^n)$. Thus, by Fekete's lemma, the limit

$$R(\mathcal{G}) = \lim_{n \longrightarrow \infty} \sqrt[n]{r(\mathcal{G}^n)} \tag{5}$$

exists, and $\log R(\mathcal{G})$ is the minimum asymptotic zero-error rate for the source P with maximum distortion constraint D.

In the following examples with possible practical distortion measures, we investigate how $r(G^n)$ changes with increasing n.

Example 1: Consider $\mathcal{X} = \{0, 1, \ldots, K-1\}$, and $d(x, y) = |x - y|$. This setting is useful in various signal compression applications which require a guarantee of no more than D units of distortion at any sample. For instance, the signal may be a medical image, where a large absolute error in the value of a few pixels may lead to incorrect diagnosis.

Let $\mathcal{G}$ be the characteristic graph corresponding to the case $K = 6$ and $D = 2$. It is obvious that $r(\mathcal{G}) = 2$, e.g., $\{2, 3\}$ is one of the smallest dominating sets in $\mathcal{G}$. For evaluating $r(\mathcal{G}^2)$, we make the following observation: On $\mathcal{G}^2$, all "corner" points $\{(0, 0), (0, 5), (5, 0), (5, 5)\}$ have to be adjacent to at least one vertex in the minimal dominating set, by definition. But no vertex can be adjacent to more than one corner vertex, and hence $r(\mathcal{G}^2) \ge 4$. Since $r(\mathcal{G}^2) \le r(\mathcal{G})^2 = 4$, we have $r(\mathcal{G}^2) = 4$. This argument can be generalized for all n, so that $R(\mathcal{G}) = r(\mathcal{G}) = 2$. Thus, scalar coding is in fact optimal. Optimality of scalar coding is a very useful property of a graph, and it is of interest to identify classes of graphs that possess it. The above graph belongs to the special class of graphs called *interval graphs*, for which we will show later that $R(\mathcal{G}) = r(\mathcal{G})$. $\diamond$

Example 2: Let $\mathcal{X} = \{0, 1, \ldots, K-1\}$, and $d(x, y) = \min\{(x - y) \bmod K, (y - x) \bmod K\}$. Thus, if the vertices are considered as equally spaced points on a circle, the distortion between two points is proportional to the length of the shortest path connecting them. An example for this scenario is *phase* quantization, commonly encountered in speech coding. The characteristic graph for the case $K = 4$, and $D = 1$ is the familiar four-cycle, denoted by $\mathcal{C}_4$. Obviously, $r(\mathcal{C}_4) = 2$ (choose the dominating set $\{0, 2\}$). However, $r(\mathcal{C}_4^2) = 3$, e.g., $\{(1, 1), (2, 3), (3, 2)\}$ is a minimal dominating set for $\mathcal{C}_4^2$. Further increasing the block size, we get $r(\mathcal{C}_4^3) = 5$, and $r(\mathcal{C}_4^4) = 8$. In fact, it will follow from the results of Section 4 that $R(\mathcal{G}) = 4/3$. Since $(4/3)^n$ is not an integer for any finite n, the minimum rate $\log R(\mathcal{G})$ is not achieved by vector quantization for any finite dimension n. $\diamond$

We proceed by reformulating $r(\mathcal{G})$ as the solution to an integer program:

$$r(\mathcal{G}) = \min_{\substack{x_i \in \{0,1\} \\ \sum_{i \in \mathcal{N}_j} x_i \ge 1 \ \ \forall j \in V}} \sum_{i=1}^{|V|} x_i \ , \tag{6}$$

where $\mathcal{N}_j$ denotes the neighborhood of vertex j. The minimization is obviously over all dominating sets of $\mathcal{G}$, hence the achieved minimum is $r(\mathcal{G})$. When the integer constraints on x_i are relaxed to only impose non-negativity, we obtain the linear program

$$r^*(\mathcal{G}) \stackrel{\triangle}{=} \min_{\substack{x_i \ge 0 \\ \sum_{i \in \mathcal{N}_j} x_i \ge 1 \ \ \forall j \in V}} \sum_{i=1}^{|V|} x_i \ . \tag{7}$$

Applying the duality principle in linear programming [9], we obtain an alternative definition for $r^*(\mathcal{G})$:

$$r^*(\mathcal{G}) = \max_{\substack{w_j \geq 0 \\ \sum_{j \in \mathcal{N}_i} w_j \leq 1 \ \forall i \in V}} \sum_{j=1}^{|V|} w_j \ . \tag{8}$$

The nonnegative weights w_j are in fact assigned to neighborhoods. The above formula follows because there is a one-to-one correspondence between vertices and neighborhoods. Reinstalling the integer constraints in (8) yields a smaller quantity given by

$$t(\mathcal{G}) \stackrel{\triangle}{=} \max_{\substack{w_j \in \{0,1\} \\ \sum_{j \in \mathcal{N}_i} w_j \leq 1 \ \forall i \in V}} \sum_{j=1}^{|V|} w_j \ .$$

This time, the maximization is over all subsets of V such that no two vertices in the subset have a common neighbor, nor are they adjacent. In other words, the neighborhoods of vertices in the subset shall be pairwise disjoint. We define such sets as *tiling sets* and call the size of the largest tiling set, $t(\mathcal{G})$, the *tiling number*. The name naturally follows after observing the correspondence between choosing vertices and packing tiles (neighborhoods) in the graph.

It easily follows from the discussion above that $r(\mathcal{G}) \geq r^*(\mathcal{G}) \geq t(\mathcal{G})$. Lovász [8], after introducing the above relaxation, proved using combinatorial arguments that $R(\mathcal{G}) = r^*(\mathcal{G})$. It can be easily shown that the sequence $\sqrt[n]{t(\mathcal{G}^n)}$ is super-multiplicative, so (again by Fekete's lemma) the limit $T(\mathcal{G}) \stackrel{\triangle}{=} \lim_{n\to\infty} \sqrt[n]{t(\mathcal{G}^n)}$ exists. Thus, for every $n \geq 1$,

$$r(\mathcal{G}) \geq \sqrt[n]{r(\mathcal{G}^n)} \geq R(\mathcal{G}) = r^*(\mathcal{G}) \geq T(\mathcal{G}) \geq \sqrt[n]{t(\mathcal{G}^n)} \geq t(\mathcal{G}) \ .$$

The significance of this chain of inequalities is that if $r(\mathcal{G}^n) = t(\mathcal{G}^n)$, for some $0 < n < \infty$, then $\sqrt[n]{r}(\mathcal{G}^n) = R(\mathcal{G}) = T(\mathcal{G}) = \sqrt[n]{t}(\mathcal{G}^n)$, i.e., $R(\mathcal{G})$ is achieved by a coding scheme using blocks of length n. The most interesting case, of course, is $n = 1$, which means that scalar coding is optimal, as is the case for Example 1. In Section 4, we will investigate examples of this phenomenon further. In fact, we will show that for certain classes of graphs, $r(\mathcal{G}) = t(\mathcal{G})$.

3 A Single-letter Formula for $R(\mathcal{G})$

We prove the result of Lovász [8] with a purely rate-distortion theoretic approach. We begin by converting the maximum distortion constraint into the more familiar *average* distortion constraint via a transformation of the distortion measure. The transformation is given by:

$$d'(x,y) = \begin{cases} 0 & d(x,y) \leq D \\ 1 & d(x,y) > D \end{cases} \ . \tag{9}$$

Defining

$$d'(x^n, y^n) \stackrel{\triangle}{=} \frac{1}{n} \sum_i d'(x_i, y_i) \ ,$$

it easily follows that $d'(x^n, y^n) = 0$ if and only if $d(x^n, y^n) = \max_i d(x_i, y_i) \leq D$.

Since the solution to the regular rate-distortion problem is for the case where an *eventually* vanishing $\Pr[d'(X^n, Y^n) > 0]$ is sufficient, it cannot be directly applied here. Specifically, we are interested in the asymptotical rate when the error constraint is

$$\Pr[d(X^n, Y^n) > D] = \Pr[d'(X^n, Y^n) > 0] = 0 \ \ \forall n > 0 \ .$$

Towards this end, we define the "D'-ball" around a reproduction point y^n, with respect to the single-letter distortion measure d', as $\mathcal{B}_{d'}(y^n, D') = \{x^n \in \mathcal{X}^n : \frac{1}{n}\sum_i d'(x_i, y_i) \le D'\}$. Hence, we need to "cover" $\mathcal{X}^n$ with as few 0-balls as possible, i.e., we are interested in finding the smallest cardinality of a set $\mathcal{C} \subset \mathcal{Y}^n$ such that $\bigcup_{y^n \in \mathcal{C}} \mathcal{B}_{d'}(y^n, 0) = \mathcal{X}^n$. Recall that the type of a sequence $x^n \in \mathcal{X}^n$ is the distribution P_{x^n} defined by $P_{x^n}(a) = \frac{1}{n}N(a|x^n) \;\; \forall a \in \mathcal{X}$, where $N(a|x^n)$ denotes the number of occurences of a in the sequence x^n. T_P^n denotes the set of all $x^n \in \mathcal{X}^n$ having type P.

Fix $\delta > 0$. The type covering lemma of Csiszár and Körner [5, Lemma 2.4.1] guarantees that, for any distortion measure d' on $\mathcal{X} \times \mathcal{Y}$, and for $n \ge n_0(d', \delta) > 0$, T_P^n for any type P can be covered with not more than $2^{nR_{P,d'}(0)+n\delta}$ 0-balls $\mathcal{B}_{d'}(y^n, 0)$, where $y^n \in \mathcal{Y}^n$. Here $R_{P,d'}(D')$ is the rate-distortion function for the memoryless source P and single-letter distortion d', given by

$$R_{P,d'}(D') = \min_{Q(y|x): E\{d'(X,Y)\} \le D'} I(X;Y) \; .$$

Only the case $D' = 0$ interests us here. Since $\mathcal{X}^n = \bigcup_P T_P^n$, the type covering lemma now assures us that as $n \longrightarrow \infty$, the sufficient number of 0-balls to cover the space $\mathcal{X}^n$ is given by

$$\begin{aligned} \sum_P 2^{nR_{P,d'}(0)} &\le (n+1)^{|\mathcal{X}|} 2^{n \max_P R_{P,d'}(0)} \\ &= 2^{n\left[\max_P R_{P,d'}(0) + \frac{|\mathcal{X}|}{n}\log(n+1)\right]} \; , \end{aligned}$$

where for the inequality, we use the polynomial bound on the number of types. Hence $R(\mathcal{G}) \le 2^{\max_P R_{P,d'}(0)}$. However, according to Marton's result [10], it is not possible to achieve an infinite error probability exponent, and hence a zero error probability, with a coding rate less than $\max_P R_{P,d'}(0)$. Therefore,

$$R(\mathcal{G}) = 2^{\max_P R_{P,d'}(0)} \; . \tag{10}$$

Note that, by definition, $R(\mathcal{G})$ is the minimum asymptotic rate for *fixed-length* codes. Thus, the above discussion shows that fixed-length codes are asymptotically optimal for the maximum distortion criterion, for sources whose statistics may vary arbitrarily with time.

To expand and simplify the formula (10), we first repeat the variational formula given in [2] for the rate-distortion function for arbitrary $D' \ge 0$:

$$R_{P,d'}(D') = \max_{\beta \ge 0} \left\{ -\beta D' + \min_Q \left(-\sum_{x \in \mathcal{X}} P(x) \log \sum_{y \in \mathcal{X}} Q(y) 2^{-\beta d'(x,y)} \right) \right\} \; . \tag{11}$$

Here, β denotes the negative slope of the rate-distortion function, and Q denotes the reproduction distribution. When $D' > 0$, the maximum over β is attained by the actual slope at distortion point D'. When $D' = 0$, however, the maximum is guaranteed to be achieved by $\beta \longrightarrow \infty$, for all sources P. In this case, turning back to the graph interpretation, for every vertex $i \in V$, only the vertex pairs $i, j \in V$ with $d'(i,j) = 0$ (i.e., only the connected pairs of vertices) contribute to the sum above. Hence the formula for $R(\mathcal{G})$ simplifies to

$$\log R(\mathcal{G}) = \max_P \min_Q \left\{ -\sum_{i \in V} P(i) \log \left[\sum_{j \in \mathcal{N}_i} Q(j) \right] \right\} \; .$$

The argument of the above min-max problem is concave (in fact linear) in P and convex in Q. Therefore, from von Neumann's minmax theorem, the order of minimization and maximization may be reversed:

$$\log R(\mathcal{G}) = \min_Q \max_P \left\{ -\sum_{i \in V} P(i) \log \left[\sum_{j \in \mathcal{N}_i} Q(j) \right] \right\} . \qquad (12)$$

Since the inner maximization is over a linear functional of P, which is constrained on a simplex, the maximum is attained when

$$P(i) = \begin{cases} 1 & i = \arg\min_{k \in V} \log \left[\sum_{j \in \mathcal{N}_k} Q(j) \right] \\ 0 & \text{otherwise} \end{cases} .$$

Substituting this in (12), we obtain

$$R(\mathcal{G}) = \min_Q \frac{1}{\min_{i \in V} \sum_{j \in \mathcal{N}_i} Q(j)} . \qquad (13)$$

In the next lemma, we prove that the above formula for $R(\mathcal{G})$ simplifies to $r^*(\mathcal{G})$.

Lemma 1

$$R(\mathcal{G}) = r^*(\mathcal{G}) \qquad (14)$$

$\diamond$

Proof:
$R(\mathcal{G}) \leq r^*(\mathcal{G})$: Let $\{x_1, x_2, \ldots, x_{|V|}\}$ be the solution to the minimization in (7), i.e., $r^*(\mathcal{G}) = \sum_{j=1}^{|V|} x_j$. There must exist at least one i such that $\sum_{j \in \mathcal{N}_i} x_j = 1$, otherwise the objective could be decreased further. Now, for $j \in V$, set $Q(j) = \frac{x_j}{r^*(\mathcal{G})}$. Obviously Q is a legitimate probability distribution and

$$\min_{i \in V} \sum_{j \in \mathcal{N}_i} Q(j) = \frac{1}{r^*(\mathcal{G})} \min_{i \in V} \sum_{j \in \mathcal{N}_i} x_j = \frac{1}{r^*(\mathcal{G})} . \qquad (15)$$

From (13), $R(\mathcal{G}) \leq r^*(\mathcal{G})$ follows.

$R(\mathcal{G}) \geq r^*(\mathcal{G})$: Let Q achieve the minimum in (13), i.e., $\min_{i \in V} \sum_{j \in \mathcal{N}_i} Q(j) = \frac{1}{R(\mathcal{G})}$. Choose $x_j = R(\mathcal{G})Q(j)$ for all $j \in V$. Observing that $x_j \geq 0$ and $\min_{i \in V} \sum_{j \in \mathcal{N}_i} x_j = 1$, or in other words, $\sum_{j \in \mathcal{N}_i} x_j \geq 1$ for all i, it easily follows that $\sum_{j \in V} x_j = R(\mathcal{G}) \geq r^*(\mathcal{G})$. $\square$

4 Optimality of Scalar Coding for Special Classes of Graphs

In this section, we show that for some classes of graphs, scalar coding is optimal. This means $r(\mathcal{G}) = r^*(\mathcal{G})$ for all graphs in those particular classes. To this end, we will prove the sufficient condition $r(\mathcal{G}) = t(\mathcal{G})$.

For the following discussion, we need to define generalized versions of tiling and dominating sets and corresponding numbers. Let a graph with *free* vertices be denoted as $\mathcal{G} = (V, E, F)$, where V and E are the vertices and the edges as usual, and $F \subseteq V$ are called free vertices. A dominating set $\mathcal{D}_f$ for $\mathcal{G}$ is defined as a subset of vertices such that all non-free vertices are either in $\mathcal{D}_f$, or adjacent to a vertex in $\mathcal{D}_f$. A tiling set $\mathcal{T}_f$ for $\mathcal{G}$ is defined as a subset of non-free vertices whose neighborhoods are pairwise disjoint. Let $r_f(\mathcal{G})$ and $t_f(\mathcal{G})$ denote the size of the smallest dominating set and the largest tiling set, respectively. It is easy to check that for graphs with free nodes, $r_f(\mathcal{G}) \geq t_f(\mathcal{G})$.

4.1 Interval Graphs and Forests

A graph is called an *interval graph*, [1], if there exists a one-to-one correspondence between the vertices of the graph and intervals on the real line, such that two vertices are connected if and only if the corresponding intervals overlap. Thus, in the subsequent discussion, we will use the words "vertex" and "interval" interchangeably. An example is shown in Figure 1. The graph of Example 1, Section 2 was also an interval graph.

A connected graph is a *tree*, [1], if it does not contain any circuits of length more than 2. A *forest* is a collection of disconnected trees. The vertices with only one neighbor are called *leaves*. Note that a forest has at least two leaves unless it is composed only of isolated vertices.

Theorem 1 *For all interval graphs $\mathcal{G}$, $r(\mathcal{G}) = t(\mathcal{G})$. Similarly, for all forests $\mathcal{G}$, $r(\mathcal{G}) = t(\mathcal{G})$.* $\diamond$

Remark: The significance of the result for interval graphs is immediately seen once the following observation is made: whenever $\mathcal{X}$ consists of a finite set of points on the real line, and a difference distortion measure in the form $d(x, y) = |x - y|^m$ is employed, the corresponding graph is an interval graph. This is demonstrated by mapping the symbol x to the interval $I_x = [x - \sqrt[m]{D}/2, x + \sqrt[m]{D}/2]$. (The source and the distortion measure in Example 1, Section 2 correspond to $m = 1$, and $D = 2$.) The theorem establishes that for such distortion measures, which are the most widely used, scalar coding is optimal. The theorem can be generalized to the case where the finite set of reproduction points $\mathcal{Y} \neq \mathcal{X}$ is pre-specified, the intervals I_x are of different widths, and each $x \in \mathcal{X}$ is constrained to be reproduced by some $y \in \mathcal{Y} \cap I_x$.

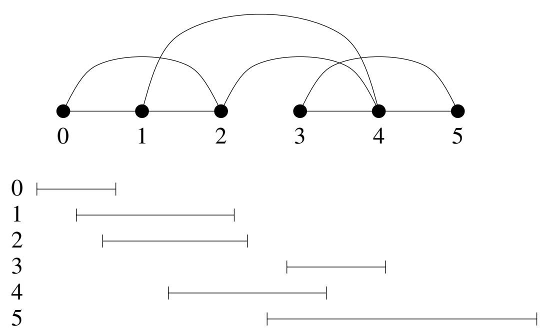

Figure 1: Intervals and the corresponding graph.

Remark: Due to lack of space we will omit the proof for the case of forests, which is essentially similar. Note that both the proofs are constructive, in that they also yield the smallest dominating set for the corresponding graph.

Proof:
We prove by induction the more general statement that for all interval graphs $\mathcal{G}$ with free vertices, $r_f(\mathcal{G}) = t_f(\mathcal{G})$. The base case for the induction is when the intervals are pairwise disjoint, in which case $r_f(\mathcal{G}) = t_f(\mathcal{G}) = |V| - |F|$.

Now, for a general interval graph, sort the intervals in increasing order of their right end points. Take the first interval v. Denote by w the neighbor of v with the highest order (rightmost right end). Consider the two possibilities.

i. v is a free interval: Removing v does not change $r_f(\mathcal{G})$ or $t_f(\mathcal{G})$. To see this, note that $\mathcal{N}_v \subseteq \mathcal{N}_w$. Hence v and w cannot both be in the smallest dominating set. Also, if v is in the smallest dominating set, we can replace it with w, and obtain another dominating set of the same size. This proves that $r_f(\mathcal{G})$ does not change by removing v. Now, since v is free, it cannot be in any tiling set. However, if v is the only common neighbor of two other intervals, then removing it may increase the size of the largest tiling set. But since $\mathcal{N}_v \subseteq \mathcal{N}_w$, any two intervals connected to v are either also connected to w, or (if w is already one of the two intervals) connected to each other. Hence removing v does not change $t_f(\mathcal{G})$.

ii. v is a non-free interval: We define two new graphs: The first graph $\mathcal{G}_1$ is induced by $\mathcal{N}_w$, while preserving the freeness of intervals. The second graph, $\mathcal{G}_2$, is induced by $V - \{v, w\}$, preserving the freeness of intervals in $V - \mathcal{N}_w$, but declaring all intervals in $\mathcal{N}_w - \{v, w\}$ free. See Figure 2 for an example on the interval graph in Figure 1. Now, it is clear that $r_f(\mathcal{G}_1) = t_f(\mathcal{G}_1) = 1$, since in $\mathcal{G}_1$, $\{w\}$ and $\{v\}$ are the smallest dominating and the largest tiling sets, respectively. Moreover, clearly if $\mathcal{D}_f$ is a dominating set in $\mathcal{G}_2$, then $\mathcal{D}_f \cup \{w\}$ is a dominating set for $\mathcal{G}$. Hence, $1 + r_f(\mathcal{G}_2) \geq r_f(\mathcal{G})$. Let $\mathcal{T}_f$ be a tiling set in $\mathcal{G}_2$. An interval in $\mathcal{G}_2$ that overlaps v or w is declared free, and hence cannot be picked in $\mathcal{T}_f$. Since v is non-free, it follows that $\mathcal{T}_f \cup \{v\}$ is a tiling set for $\mathcal{G}$. So, we have

$$1 + r_f(\mathcal{G}_2) \geq r_f(\mathcal{G}) \geq t_f(\mathcal{G}) \geq 1 + t_f(\mathcal{G}_2) \ .$$

The proof is complete after observing that $\mathcal{G}_2$ is also an interval graph, which contains fewer intervals than $\mathcal{G}$. The above recursion can be applied to $\mathcal{G}$, until $\mathcal{G}$ becomes a graph with nonoverlapping intervals (or an empty graph). □

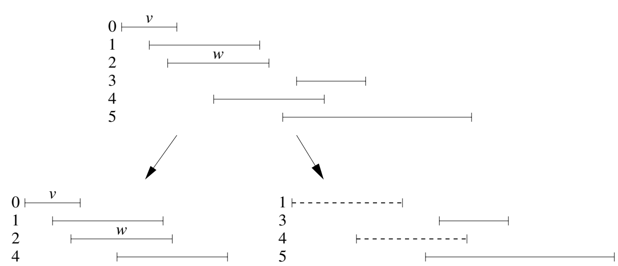

Figure 2: Original intervals (top) and intervals corresponding to $\mathcal{G}_1$ (left) and $\mathcal{G}_2$ (right). Free intervals are shown as dashed lines.

4.2 On Codebook Design Algorithms

Finally, let us briefly consider here the question of codebook design algorithms for the maximum distortion criterion. This, as noted before, is equivalent to finding dominating sets in graphs. But it is known, [7], that finding the smallest dominating set in an arbitrary graph is NP-Hard. In fact, even polynomial-time algorithms guaranteed to finding a dominating set of size within a factor $O(\log|V|)$ of the minimum are unlikely to exist, unless $P = NP$, [6].

Thus fast design of even near-optimal scalar quantizers appears to be infeasible. This shifts the attention to algorithms developed for particular classes of graphs. The practical relevance of interval graphs was pointed out above. For the class of interval graphs (and of forests), the proof of Theorem 1 provides an algorithm with worst-case running time $O(|V|)$ which designs the *optimal* (scalar) codebook. Note that design algorithms with shorter worst-case running times cannot exist, since even listing the neighbors of all the nodes of a graph takes $O(|V|)$ time.

Algorithms to find the smallest dominating sets of particular classes of graphs had been considered in the computer science community previously, and $O(|V|)$ algorithms for interval graphs and forests were previously proposed in [4] and [3] respectively. The algorithms for interval graphs and forests derived via our common approach are close relatives of these earlier independent algorithms, and our approach explains why finding the smallest dominating sets for these classes of graphs is so easy.

References

[1] C. Berge. *Graphs and Hypergraphs.* North Holland Publishing Co., Amsterdam, Holland, 1973.

[2] T. Berger. *Rate Distortion Theory. A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[3] K. S. Booth and G. S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *Journal of Computer and System Sciences*, 13(3):335–379, December 1976.

[4] E. J. Cockayne, S. Goodman, and S. T. Hedetniemi. A linear algorithm for the domination number of a tree. *Information Processing Letters*, 4(2):41–44, 1975.

[5] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic, New York, 1982.

[6] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, July 1991.

[7] S. T. Hedetniemi and R. C. Laskar. *Topics on Domination.* Elsevier Science Publishers, New York, NY, 1991.

[8] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383–390, December 1975.

[9] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley Publishing Company, Reading, MA, 1984.

[10] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Trans. on Information Theory*, 20(2):197–199, March 1974.

[11] M. J. Weinberger, G. Seroussi, and G. Sapiro. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Trans. on Image Processing*, 9(8):1309–1324, August 2000.