On Entropy Minimization for Near-Lossless Differential Coding

Hannes Hartenstein and Dietmar Saupe

Abstract— **In this letter we investigate near-lossless signal coding for which the reconstruction signal is required to have an absolute error in each component bounded by a tolerance τ. For differential coding no practical algorithm is known that computes an optimal reconstruction, i.e., one for which the sequence of consecutive differences has minimal entropy. In this letter we give the optimal lower bound for the entropy of the differences of the reconstruction and, in addition, present tighter bounds for some special cases.**

Index Terms— **Entropy minimization, near-lossless compression, signal coding.**

I. Introduction

In many applications, for example, medical imagery or numerical weather simulations, the large amount of data to be stored or transmitted demands for data compression. Since lossless coding usually yields only a rather small compression ratio, lossy coding methods have to be employed when higher compression ratios are needed. Most lossy compression schemes operate by minimizing some *average* error measure such as the root mean square error. However, in error critical applications such average error measures may be inappropriate. Instead, it is sometimes desirable that each component not be distorted by more than a certain tolerance.

In this letter, we restrict ourselves to one-dimensional signals, e.g., signals that stem from time series data or one-dimensional image scans. Therefore, the code for a finite one-dimensional signal $s \in \mathbb{Z}^n$ has to represent a reconstruction signal $\hat{s} \in \mathcal{N}_\tau(s)$, where

$$\mathcal{N}_\tau(s) := \{t \in \mathbb{Z}^n | \, \|t - s\|_\infty \le \tau\}$$

with $\|t-s\|_\infty = \max_{i \in \{1 \cdots n\}} |t_i - s_i|$. For $\tau = 0$ this leads to lossless coding. If τ is small, the term *near-lossless coding* appears to be justified. $\mathcal{N}_\tau(s)$ can be seen as the set of all left-to-right paths in a trellis as depicted in Fig. 1.

Let D be the difference operator

$$D: \mathbb{Z}^n \to \mathbb{Z}^{n-1}$$
$$s = (s_1, \cdots, s_n) \mapsto (s_2 - s_1, \cdots, s_n - s_{n-1}).$$

The distribution $P(s)$ of a signal $s \in \mathbb{Z}^n$ is given by $p_j(s) = 1/n|\{i | s_i = j\}|$, $j \in \mathbb{Z}$, and the empirical entropy of s is defined via $H(s) := H[P(s)] = -\sum_{j\in\mathbb{Z}} p_j(s) \log_2 [p_j(s)]$.

Manuscript received August 25, 1997. The associate editor coordinating the review of this letter and approving it for publication was Prof. C. Georghiades.

The authors are with the Institut für Informatik, Universität Freiburg, 79110 Freiburg, Germany (e-mail: hartenst,saupe@informatik.uni-freiburg.de).

Publisher Item Identifier S 1089-7798(98)03275-X.

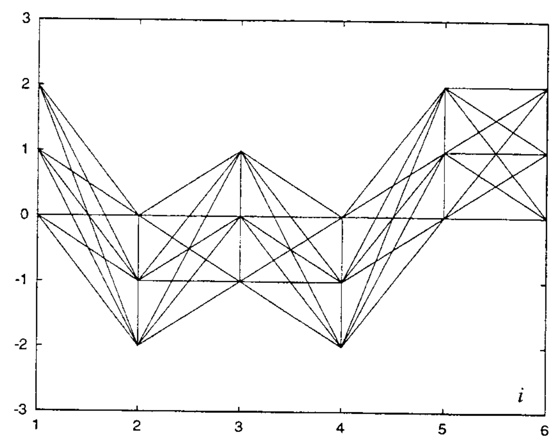

Fig. 1. Each left-to-right path corresponds to an element of $\mathcal{N}_\tau(s)$ with $s = (1, -1, 0, -1, 1, 1)$ and $\tau = 1$.

We are interested in finding the reconstruction signal in the neighborhood $\mathcal{N}_\tau(s)$ that can be coded most efficiently using differential coding, i.e., we are faced with the following optimization problem:

$$\text{find } t^* \in \mathcal{N}_\tau(s) \text{ s. t. } H[D(t^*)] = \min_{t \in \mathcal{N}_\tau(s)} H[D(t)].$$

The set $\mathcal{N}_\tau(s)$ has $(2\tau + 1)^n$ elements, therefore, it is computationally infeasible to determine the optimal element by enumeration when $\tau > 0$ and n is large. To the best of our knowledge there is no algorithm that solves this optimization problem in an amount of time that depends polynomially on the signal size n. We conjecture that this problem is computationally intractable. The problem was presented in [1] where an iterative optimization method based on dynamic programming is proposed that yields a local minimum. Results of this and several other near-lossless coding methods can be found in [2].

If the probability distribution $P[D(t^*)]$ could be determined it could either be used as part of the codec scheme (in this case s would be a training sequence), or it had to be sent as side information for the encoding of s. In this letter we derive bounds for the coding gain $H[D(s)] - H[D(t^*)]$. The optimal general bound is given in Section II; bounds for some special cases are given in Section III.

II. Lower Bound on the Entropy

We begin by examining the entropy of sequences $t \in \mathcal{N}_\tau(s)$, in particular those that come from quantizing a given signal $s \in \mathbb{Z}^n$. Let Q be a quantization function with tolerance τ, i.e., $Q: \mathbb{Z} \to \mathbb{Z}$, and $|Q(z) - z| \leq \tau$ for all $z \in \mathbb{Z}$. Thus $Q(s) \in \mathcal{N}_\tau(s)$, where $Q(s)$ stands for the signal $[Q(s_i)]_{1 \leq i \leq n}$. From [3] we have the following result concerning the maximal entropy gain from quantization:

$$H[Q(s)] \geq H(s) - \log_2(2\tau + 1) \tag{1}$$

for any quantization function Q with tolerance τ and any signal $s \in \mathbb{Z}^n$.

However, this result does not apply to all $t \in \mathcal{N}_\tau(s)$, since for arbitrary $t \in \mathcal{N}_\tau(s)$ it is possible for $i \neq j$ and $s_i = s_j$ that $t_i \neq t_j$, i.e., t may not necessarily be derived from s by quantization. Yet, we show that for all $t \in \mathcal{N}_\tau(s)$ there exists a quantization function Q_t with tolerance τ such that

$$H(t) \geq H[Q_t(s)]. \tag{2}$$

First we show the following:

Lemma 1: Given $n \in \mathbb{N}$, $i, j \in \{1, \cdots, n\}$, and $p_1, \cdots, p_n, q_1, q_2 \geq 0$ such that $q_1 + q_2 + \sum_{k=1}^{n} p_k = 1$. Set $P_0 = (p_1, \cdots, p_i + q_1, \cdots, p_j + q_2, \cdots, p_n)$, $P_1 = (p_1, \cdots, p_i + q_1 + q_2, \cdots, p_j, \cdots, p_n)$, $P_2 = (p_1, \cdots, p_i, \cdots, p_j + q_1 + q_2, \cdots, p_n)$. Then

$$H(P_1) \leq H(P_0) \quad \text{or} \quad H(P_2) \leq H(P_0).$$

Proof: Define $h(x) := -x \log_2 x$. In the above inequalities we can first subtract $\sum_{k \neq i, j} h(p_k)$ on each side. Let $p_i \leq p_j$. Since the derivative of $h(x)$ on $(0, 1)$ is strictly decreasing, it follows that $h(p_i + q_1) - h(p_i) \geq h(p_j + q_1 + q_2) - h(p_j + q_2)$. This gives the inequality $H(P_2) \leq H(P_0)$. In the other case, $p_i > p_j$, we obtain $H(P_1) \leq H(P_0)$ in the same manner. ■

Lemma 1 is now used to prove inequality (2). Take a fixed symbol $z \in \mathbb{Z}$ and define $M_z := \{i | s_i = z\}$. Then for $i \in M_z$ it follows that $t_i \in \{z - \tau, \cdots, z + \tau\}$. $P(t)$ is given by $p_j(t) = 1/n|\{i|t_i = j\}|$ and for $|k| \leq \tau$ the probability $p_{z+k}(t)$, $|k| \leq \tau$, is split up into $p_{z+k}(t) = \tilde{p}_{z+k}(t) + q_k$ with $q_k = 1/n|\{i \in M_z | t_i = z + k\}|$. Then, by iteratively applying Lemma 1, we concentrate all q_k's on one symbol without increasing the entropy. This single value defines $Q(z)$. When this is done for each symbol $z \in \mathbb{Z}$ that occurs in s we obtain inequality (2).

So far we were only concerned with the sequence s itself, now we turn our attention to the sequence of differences $D(s)$.

Theorem 1: For all $t \in \mathcal{N}_\tau(s)$ the following inequality holds:

$$H[D(t)] \geq H[D(s)] - \log_2(4\tau + 1). \tag{3}$$

Proof: There are $4\tau + 1$ possible values for the ith component of $D(t)$, namely $(s_{i+1} - s_i + k)_{-2\tau \leq k \leq 2\tau}$. Thus, $D(t) \in \mathcal{N}_{2\tau}[D(s)]$ for all $t \in \mathcal{N}_\tau(s)$, and it follows that

$$\min_{u \in \mathcal{N}_{2\tau}[D(s)]} H(u) \leq \min_{t \in \mathcal{N}_\tau(s)} H[D(t)].$$

In order to find a lower bound for $H(u)$, where $u \in \mathcal{N}_{2\tau}[D(s)]$, we make use of inequality (2) with a suitable quantization function Q_u for $D(s)$ and tolerance 2τ, obtaining

$$H\{Q_u[D(s)]\} \leq H(u).$$

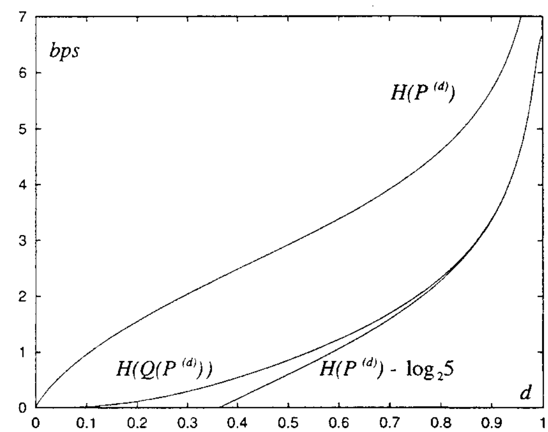

Fig. 2. Comparison of the entropy values for $P^{(d)}$, the lower bound given in Theorem 1, and the tighter bound derived in Section III ($\tau = 1$). The x axis gives the values for d, the y axis gives the entropy value in bits per sample *(b/s)*.

Now, inequality (1) yields

$$H[D(s)] - \log_2(4\tau + 1) \leq H\{Q_u[D(s)]\}.$$

Finally, by joining the above three inequalities we obtain the result of the theorem. ■

The bound of Theorem 1 can be attained, e.g., with the signal in Fig. 1 and $\tau = 1$, and, therefore, it is the best general lower bound.

III. Special Cases

The maximal coding gain of $\log_2(4\tau + 1)$ can only be attained when the distribution $P[D(s)]$ is uniform. This follows from the proof of inequality (1) in [3]. Typically, the distribution of $D(s)$ will be more peaked. In this section, we analyze the maximal drop in entropy when the distribution of $D(s)$ has the following form. Define $P^{(d)}$ for $d \in (0, 1)$ via

$$p_j^{(d)} = \begin{cases} c \cdot d^{|j|}, & |j| \leq m \\ 0, & \text{otherwise} \end{cases}$$

where c is the appropriate normalization factor. This class is a discrete version of the Laplace distribution. For $d = 0$ one obtains the Dirac distribution, when d tends to one the distribution becomes uniform. As in the proof of Theorem 1 we determine a bound for the entropy by using an optimal quantization function with tolerance 2τ. For this class of distributions the optimal quantization function has to be a uniform quantizer. This can be proven with a similar argument as in Lemma 1. We have computed the entropies $H[P^{(d)}]$ and $H\{Q[P^{(d)}]\}$ with $m = 255$, $\tau = 1$, and Q being the optimal quantization function with tolerance 2τ. The distribution $Q[P^{(d)}] = [\tilde{p}_j^{(d)}]_{j \in \mathbb{Z}}$ is defined via $\tilde{p}_j^{(d)} = \sum_{k \in \{i | Q(i) = j\}} p_k^{(d)}$. The results are given in Fig. 2. For a wide range of d-values the improved entropy bound is almost as low as predicted by inequality (3).

For an arbitrary distribution a lower bound that is tighter than the one given in (3) can be obtained by using the method of [3] to determine the optimal quantization function. This can be done in linear time with a dynamic programming approach.

A numerical simulation performed on a signal s of length 100 000 with underlying distribution $P^{(d)}$, $d = 0.9$, and $m = 255$, gives the following results. The empirical entropy of $D(s)$ is 5.7 bits per sample *(b/s)* and the lower bound for $\tau = 1$ is found to be 3.4 b/s. The coding rate obtained with the method of [1] is 4.1 b/s without taking into account the coding costs for the side information. Since there is no feasible method known to determine the optimal rate, it is not possible to indicate by what amount the lower bound underestimates the optimal rate. It depends, of course, on the length of the signal s whether it is worth to transmit the probability tables.

IV. Conclusion and Further Work

We have addressed the question of what compression can be gained when a signal is near-losslessly encoded using differential coding. Our main contribution is the derivation of the optimal lower bound on the entropy of the sequence of differences of consecutive samples in the reconstruction signal. This bound is obtained by analyzing the distribution of the differences in the original signal. However, while this result represents the best possible bound, in most cases it cannot be attained since the actual order in which the signal values are given is important in practice but not used in the theorem. Further work is directed toward a generalization of the theorem for differential coding schemes using arbitrary predictors, and toward a theoretical study of the effect of uniform quantization of the original signal to near-lossless differential encoding.

References

[1] L. Ke and M. W. Marcellin, "Near-lossless image compression: Minimum-entropy, constrained-error DPCM," in *Proc. IEEE Int. Conf. on Image Processing*, Washington, DC, Oct. 1995, pp. 298–301.

[2] H. Hartenstein, R. Herz, and D. Saupe, "A comparative study of L_∞-distortion limited image compression algorithms," in *Proc. Picture Coding Symp.*, Berlin, Germany, Sept. 1997, pp. 51–55.

[3] M. J. Gormish and J. T. Gill, "Discrete minimum entropy quantization," *IEEE Trans. Image Processing*, vol. 4, pp. 1314–1317, Sept. 1995.