Singularity of optimal prefix codes for pairs of geometrically-distributed random variables.

Frédérique Bassino, Julien Clément, Gadiel Seroussi, Alfredo Viola

Singularity of optimal prefix codes for pairs of geometrically-distributed random variables

Frédérique Bassino*, Julien Clément†, Gadiel Seroussi‡§¶ and Alfredo Viola§*‖
*LIPN UMR 7030, Université Paris 13 - CNRS, France; *Email: bassino@lipn.univ-paris13.fr
†GREYC UMR 6072, CNRS, Université de Caen, ENSICAEN, France; †Email: julien.clement@info.unicaen.fr
‡HP Labs, Palo Alto, California, USA; ¶Email: gseroussi@ieee.org
§Universidad de la República, Montevideo, Uruguay; ‖Email: viola@fing.edu.uy

***Abstract*—Optimal prefix codes are studied for pairs of independent, integer-valued symbols emitted by a source with a geometric probability distribution of parameter $q \in (0, 1)$. By encoding pairs of symbols, it is possible to reduce the redundancy penalty of symbol-by-symbol encoding, while preserving the simplicity of the encoding and decoding procedures typical of Golomb codes and their variants. It is shown that optimal codes for these so-called two-dimensional geometric distributions are *singular*, in the sense that a prefix code that is optimal for one value of the parameter q cannot be optimal for any other value of q. This is in sharp contrast to the one-dimensional case, where a countable family of optimal codes covers all values of the parameter q, and each code is optimal over a positive-length interval of the parameter. Thus, in the two-dimensional case, it is infeasible to give a compact characterization of optimal codes for all values of q, and practical solutions are likely to be based on discrete sequences of codes with good coverage of the parameter interval, such as those previously presented in [1]. Motivated by these findings, we extend the results of [1], giving a more explicit characterization of the sequence of optimal codes, and studying the asymptotic behavior of their redundancy as $q \to 1$.**

I. Introduction

The family of *Golomb codes* [2] was shown, in [3], to comprise the optimal codes for all *geometric distributions*, i.e., distributions of the nonnegative integers with probabilities of the form

$$p(i) = (1 - q)q^i\,, \quad i \geq 0,$$

where q is a parameter in the real interval $(0, 1)$. These distributions are useful in many practical contexts, e.g., when encoding *run lengths* (the original motivation in [2]), and in image compression when encoding prediction residuals, which are well-modeled by *two-sided geometric distributions*. Optimal codes for the latter were characterized in [4], based on some combinations and variants of Golomb codes. Codes based on the Golomb construction have the practical advantage of allowing the encoding of a symbol i using a simple explicit computation on the integer value of i, without recourse to nontrivial data structures or tables. This has led to their adoption in many practical applications (cf. [5],[6]).

Symbol-by-symbol encoding, however, can incur significant redundancy relative to the entropy of the distribution. One way to mitigate this problem, while keeping the simplicity and low latency of the encoding and decoding operations, is to consider short blocks of $d{>}1$ symbols, and use a prefix code for the blocks. In this paper, we are interested in optimal prefix codes for *pairs* (blocks of length $d{=}2$) of independent, identically distributed geometric random variables, namely, distributions on pairs of nonnegative integers (i, j) with probabilities of the form

$$P(i, j) = (1 - q)^2 q^{i+j} \quad i, j \geq 0. \tag{1}$$

We refer to this distribution as a *two-dimensional geometric distribution (TDGD)*, defined on the alphabet of integer pairs $\mathcal{A} = \{\,(i, j) \mid i, j \geq 0\,\}$. For succinctness, we denote a TDGD of parameter q by TDGD(q).

Aside from its practical aspects, the problem is also of intrinsic combinatorial interest, given the scarcity of constructive results on optimal codes for families of distributions over countable alphabets [7][8]. It was proved in [9] (see also [10]) that, if the entropy[1] $-\sum_{i\geq 0} P(i)\log P(i)$ of a distribution over the nonnegative integers is finite, optimal codes exist and can be obtained, in the limit, from Huffman codes for truncated versions of the alphabet. However, the proof does not give a general way for effectively constructing optimal codes. An algorithmic approach to building optimal codes is presented in [8], covering geometric distributions and various generalizations. The approach, though, is not applicable to TDGDs, as explicitly noted in [8]. The first general construction of optimal codes for TDGDs was presented in [1], where optimal codes were constructed for countable sequences of parameters of the form $q = 2^{-k},\ k \geq 1$ (covering the range $q \leq \frac{1}{2}$), and $q = 2^{-\frac{1}{k}},\ k \geq 1$ (covering the range $q \geq \frac{1}{2}$). The codes in [1] admit relatively simple encoding/decoding procedures, and provide good coverage of the interval $0 < q < 1$, yielding the expected reduction in redundancy relative to the one-dimensional Golomb codes.

However, as was observed in [1], the families of codes described did not contain *all* the optimal codes for TDGDs, and the work left open the question of whether a compact characterization of all the optimal codes (as available for the one-dimensional case [3] and its two-sided variant [4]) was possible. In this paper, we answer this question in the negative. Specifically, we show that optimal codes for TDGDs are *singular*, in the sense that if a code $\mathcal{T}_q$ is optimal for TDGD(q), then $\mathcal{T}_q$ *is not* optimal for TDGD(q_1) for any parameter value $q_1 \neq q$. Consequently, any set containing optimal codes for

[1] $\log x$ denotes the base-2 logarithm of x.

all values of q would be uncountable, and, thus, it would be infeasible to give a compact characterization of such a set.[2]

The main result of the paper is presented in Section III-A, after covering some definitions and preliminary material in Section II. To prove our main result, we derive, in Section III-B, some structural properties of optimal trees for TDGDs. In particular, we study how leaves corresponding to a given probability value can be distributed in the tree depth levels, and bound the number of consecutive levels that can comprise exclusively internal nodes of the tree (i.e., levels without leaves, or *gaps*). With these tools on hand, we prove our main result by showing that if a certain tree $\mathcal{T}_q$ is optimal for TDGD(q), and $q_1 \neq q$, then a modification $\mathcal{T}'_q$ of $\mathcal{T}_q$ gives a shorter expected code length than $\mathcal{T}_q$ under TDGD(q_1). Thus, $\mathcal{T}_q$ is not optimal for TDGD(q_1). It follows from these findings that, in practice, one is likely to use families of optimal codes for discrete sequences of the parameter q, such as those considered in [1]. With this motivation, in Section IV we extend the results of [1], and present a more explicit characterization of the optimal codes for $q = 2^{-1/k}$. This characterization allows us to study precisely the (oscillatory) behavior of the redundancy of these codes as $q \to 1$, and when the codes are used for all values of q (picking, for each value of q, the code in the family that minimizes the redundancy). Finally, in Section V we present some conclusions and directions of further research.

II. PRELIMINARIES

We are interested in encoding the alphabet $\mathcal{A}$ of integer pairs (i, j), $i, j \geq 0$, using a binary prefix code C. As usual, we associate C with a rooted (infinite) binary tree, whose leaves correspond, bijectively, to symbols in $\mathcal{A}$, and where each branch is labeled with a binary digit. The binary codeword assigned to a symbol is "read off" the labels on the path from the root to the corresponding leaf. The *depth* of a node x in a tree T, denoted $\text{depth}_T(x)$, is the number of branches on the path from the root to x. By extension, the depth (or *height*) of a finite tree is defined as the maximal depth of any of its nodes. A *level* of T is the set of all nodes at a given depth ℓ (we refer to this set as *level* ℓ). Let n_ℓ^T denote the number of leaves in level ℓ of T (we may omit the superscript T when clear from the context). We refer to the sequence $\{n_\ell^T\}_{\ell \geq 0}$ as the *profile* of T. Two trees will be considered *equivalent* if their profiles are identical. Thus, for a code C, we are only interested in its tree profile, or, equivalently, the *length distribution* of its codewords. Given the profile of a tree, and an ordering of $\mathcal{A}$ in decreasing probability order, it is always possible to define a canonical tree (say, by assigning leaves in alphabetical order) that uniquely defines a code for $\mathcal{A}$. Therefore, with a slight abuse of terminology, we will not distinguish between a code and its corresponding tree profile, and will refer to the same object sometimes as a tree and sometimes as a code. All trees considered in this paper are binary. A tree is *full* if every node in the tree is either a leaf or the parent of two children.[3] A tree is *balanced* (or *uniform*) if it has depth k, and 2^k leaves, for some $k \geq 0$. We will restrict the use of the term *subtree* to whole subtrees of T, i.e., subtrees consisting of a node and all its descendants in T.

We call $s(i, j) = i + j$ the *signature* of $(i, j) \in \mathcal{A}$. For a given signature $s = s(i, j)$, there are $s+1$ pairs with signature s, all with the same probability, $P(s) = (1 - q)^2 q^s$, under the distribution (1) on $\mathcal{A}$. Given a code C, symbols of the same signature can be freely permuted without affecting the properties of interest to us (e.g., average code length). Thus, for simplicity, we can also regard the correspondence between leaves and symbols as one between leaves and elements of the *multiset*

$$\hat{\mathcal{A}} = \{0, 1, 1, 2, 2, 2, \ldots, \underbrace{s, \ldots, s}_{s+1 \text{ times}}, \ldots\}. \tag{2}$$

In studying a tree, we do not distinguish between different occurrences of a signature s; for actual encoding, the $s+1$ leaves labeled with s would be mapped to the symbols $(0, s), (1, s-1), \ldots, (s, 0)$ in some fixed order. In the sequel, we will often ignore normalization factors for the signature probabilities $P(s)$ (in cases where normalization is inconsequential), and will use instead *weights* $w(s) = q^s$.

Consider a tree T for $\mathcal{A}$. Let U be a subtree of T, and let $s(x)$ denote the signature associated with a leaf x of U. Let $F(U)$ denote the set of leaves of U. We define the *weight*, $w_q(U)$, and *cost*, $\mathcal{L}_q(U)$, of U, respectively, as

$$w_q(U) = \sum_{x \in F(U)} q^{s(x)}, \text{ and } \mathcal{L}_q(U) = \sum_{x \in F(U)} \text{depth}_U(x) q^{s(x)}$$

(the subscript q may be omitted when clear from the context). When $U = T$, we have $w_q(T) = (1-q)^{-2}$, and $(1-q)^2 \mathcal{L}_q(T)$ is the average code length of T. A tree T is *optimal* for TDGD(q) if $\mathcal{L}_q(T) \leq \mathcal{L}_q(T')$ for any tree T'.

The *concatenation* of two trees T and U, denoted $T \cdot U$, is obtained by attaching a copy of U to each leaf of T. Regarded as a code, $T \cdot U$ consists of all the possible concatenations $t \cdot u$ of a word $t \in T$ with one $u \in U$. A *quasi-uniform* tree with k leaves, denoted Q_k, is a full tree with $2^{\lceil \log k \rceil} - k$ leaves at level $\lfloor \log k \rfloor$ and $2k - 2^{\lceil \log k \rceil}$ leaves at level $\lfloor \log k \rfloor + 1$. Such a tree is optimal for a uniform distribution on k symbols.

The Golomb code of order $k \geq 1$ [2], denoted G_k, encodes an integer i by concatenating $Q_k(i \bmod k)$ with a *unary* encoding of $\lfloor i/k \rfloor$ (e.g., $\lfloor i/k \rfloor$ zeros followed by a one). The first-order Golomb code G_1 is just the unary code. Thus, we have $G_k = Q_k \cdot G_1$.

In the case of one-dimensional geometric distributions, the unit interval $(0, 1)$ is partitioned into an infinite sequence of semi-open intervals $(q_{m-1}, q_m]$, $m \geq 1$, such that the Golomb code G_m is optimal for all values of the distribution

[2]Loosely, by a compact characterization we mean one in which each code is characterized by a finite number of finite parameters, which drive the corresponding encoding/decoding procedures.

[3]We use the usual "family" terminology for trees: nodes have children, parents, ancestors and descendants. We also use the common convention of visualizing trees with the root at the top and leaves at the bottom. Thus, ancestors are "up," and descendants are "down." Full trees are sometimes referred to in the literature as *complete*.

parameter q in the interval $q_{m-1} < q \le q_m$. Specifically, for $m \ge 0$, q_m is the (unique) nonnegative root of the equation $q^m + q^{m+1} - 1 = 0$ [3]. Thus, we have $q_0 = 0, q_1 = (\sqrt{5}-1)/2 \approx 0.618, q_2 \approx 0.755$, etc. A similar property holds in the case of two-sided geometric distributions [4], where the two-dimensional parameter space is partitioned into a countable sequence of patches of positive area such that all the distributions with parameter values in a given patch admit the same optimal code. Our main result, presented in the next section, shows that the situation is completely different in the case of TDGDs, where a given code can be optimal for *at most one* value of the parameter q.

III. Singularity of Optimal Codes

A. Main result

Theorem 1: Let q and q_1 be real numbers in the interval $(0,1)$, with $q \ne q_1$, and let $\mathcal{T}_q$ be an optimal tree for TDGD(q). Then, $\mathcal{T}_q$ is *not* optimal for TDGD(q_1).

We will prove Theorem 1 through a series of lemmas, which explore structural properties of optimal trees for TDGDs.

B. Levels and gaps

Lemma 1: Leaves with a given signature s are found in at most two consecutive levels of $\mathcal{T}_q$.

Proof: Let d_0 and d_1 denote, respectively, the minimum and maximum depths of a leaf with signature s in $\mathcal{T}_q$. Assume, contrary to the claim of the lemma, that $d_1 > d_0 + 1$. We transform $\mathcal{T}_q$ into a tree $\mathcal{T}'_q$ as follows. Pick a leaf with signature s at level d_0, and one at level d_1. Place both signatures s as children of the leaf at level d_0, which becomes an internal node. Pick any signature s' from a level strictly deeper than d_1, and move it to the vacant leaf at level d_1. Tracking changes in the code lengths corresponding to the affected signatures, and their effect on the cost, we have

$$\mathcal{L}_q(\mathcal{T}'_q) = \mathcal{L}_q(\mathcal{T}_q) + q^s(d_0 - d_1 + 2) - q^{s'}\delta, \qquad (3)$$

where δ is a positive integer. By our assumption, the quantity multiplying q^s in (3) is non-positive, and we have $\mathcal{L}_q(\mathcal{T}'_q) < \mathcal{L}_q(\mathcal{T}_q)$, contradicting the optimality of $\mathcal{T}_q$. Therefore, we must have $d_1 \le d_0 + 1$. ∎

A *gap* in a tree T is a non-empty set of consecutive levels containing only internal nodes of T, and such that both the level immediately above the set and the level immediately below it contain at least one leaf each. The corresponding *gap size* is defined as the number of levels in the gap. It follows immediately from Lemma 1 that in an optimal tree, if the largest signature above a gap is s, then the smallest signature below the gap is $s+1$.

Lemma 2: Let $\mathcal{T}_q$ be an optimal tree for TDGD(q), and let $k = 1 + \lfloor \log q^{-1} \rfloor$. Then, for all sufficiently large s, the size g of any gap between leaves of signature s and leaves of signature $s+1$ in $\mathcal{T}_q$ satisfies $g \le k-1$.

Proof: Assume that an optimal tree $\mathcal{T}_q$ is given.
Case $q > \frac{1}{2}$. In this case, $k = 1$, and the claim of the lemma means that there can be *no gaps* in the tree from a certain level on. Assume that there is a gap between level d with signatures s, and level d' with signatures $s+1$, $d' - d \ge 2$. By Lemma 1, all signatures $s+1$ are either in level d' or in level $d'+1$. By rearranging nodes within levels, we can assume that there is a subtree of $\mathcal{T}_q$ of height at most two, rooted at a node v of depth $d'-1 \ge d+1$, and containing at least two leaves of signature $s+1$. Hence, the weight of the subtree satisfies $w(v) \ge 2q^{s+1} > q^s$, and switching a leaf s on level d with node v on level $d'-1$ decreases the cost of $\mathcal{T}_q$, in contradiction with its optimality (when switching nodes, we carry also any subtrees rooted at them). Therefore, there can be no gap between the level containing signatures s and $s+1$, as claimed. Notice that this holds for all values of s, regardless of level.
Case $q = \frac{1}{2}$. In this case, the TDGD is dyadic, the optimal profile is uniquely determined, and it and has no gaps (the optimal profile is that of $G_1 \cdot G_1$).
Case $q < \frac{1}{2}$. Assume that $s \ge 2^k - 2$, and that there is a gap of size g between signatures s at level d, and signatures $s+1$ at level $d+g+1$. Signatures $s+1$ may also be found at level $d+g+2$. By a rearrangement of nodes that preserves optimality, and by our assumption on s, we can assume that there is a subtree of $\mathcal{T}_q$ rooted at a node v at level $d+g+1-k$, and containing at least 2^k leaves with signature $s+1$, including some at level $d+g+1$. Thus, we have $w(v) \ge 2^k q^{s+1} > q^s = w(s)$, the second inequality following from the definition of k. Therefore, we must have $d+g+1-k \le d$, or equivalently, $g \le k-1$, for otherwise exchanging v and s would decrease the cost, contradicting the optimality of $\mathcal{T}_q$. ∎

Next, we bound the rate of change of signature magnitudes as a function of depth in an optimal tree. Together with the bound on gap sizes in Lemma 2, this will lead to the proof of Theorem 1. It follows from Lemma 1 that for every signature $s \ge 0$ there is a level of $\mathcal{T}_q$ containing at least one half of the $s+1$ leaves with signature s. We denote the depth of this level by $L(s)$ (with some fixed policy for ties).

Lemma 3: Let s be a signature, and $\ell \ge 2$ a positive integer such that $s \ge 2^{\ell+2}-1$, and such that $L(s') = L(s)+\ell$ for some signature $s' > s$. Then, in an optimal tree $\mathcal{T}_q$ for TDGD(q), we have

$$\frac{\ell-2}{\log q^{-1}} \le \; s' - s \; \le \frac{\ell+1}{\log q^{-1}} \; . \qquad (4)$$

Proof: Since $s' > s \ge 2^{\ell+2}-1 > 2^{\ell-1}-1$, by the definition of $L(s')$, there are more than $2^{\ell-2}$ leaves with signature s' at level $L(s')$. We perform the following transformation

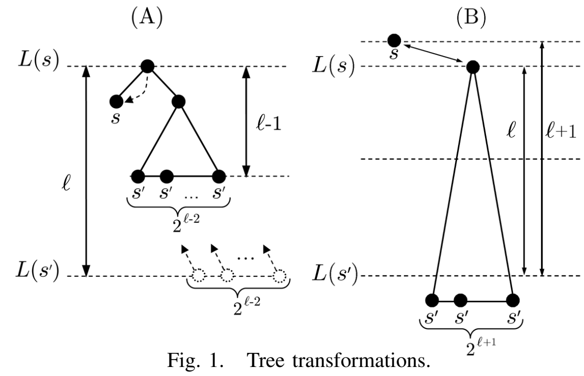

Fig. 1. Tree transformations.

(depicted in Figure 1(A)) on the tree $\mathcal{T}_q$, yielding a modified tree $\mathcal{T}'_q$: choose a leaf with signature s at level $L(s)$, and graft to it a tree with a left subtree consisting of a leaf with signature s ("moved" from the root of the subtree), and a right subtree that is a balanced tree of height $\ell-2$ with $2^{\ell-2}$ leaves of signature s'. These signatures come from $2^{\ell-2}$ leaves at level $L(s')$ of $\mathcal{T}_q$, which are removed. It is easy to verify that the modified tree $\mathcal{T}'_q$ defines a valid, albeit incomplete, code for the alphabet of a TDGD. Next, we estimate the change, Δ, in cost due to this transformation. We have

$$\Delta = \mathcal{L}_q(\mathcal{T}'_q) - \mathcal{L}_q(\mathcal{T}_q) \le q^s - 2^{\ell-2}q^{s'} . \quad (5)$$

The term q^s is due to the increase, by one, in the code length for the signature s, which causes an increase in cost, while the term $-2^{\ell-2}q^{s'}$ is due to the decrease in code length for $2^{\ell-2}$ signatures s', which produces a decrease in cost. The lower bound in (4) follows from (5) by imposing the condition $\Delta \ge 0$, which must hold if $\mathcal{T}_q$ is optimal.[4]

To prove the upper bound, we apply a different modification to $\mathcal{T}_q$. Here, we locate $2^{\ell+1}$ signatures s' at level $L(s')$, and rearrange the level so that these signatures are the leaves of a balanced tree of height $\ell+1$, rooted at depth $L(s)-1$. The availability of the required number of leaves at level $L(s')$ is guaranteed by the conditions of the lemma. We then exchange the root of this subtree with a leaf of signature s at level $L(s)$. The situation, after the transformation, is depicted in Figure 1(B). The resulting cost change is given by

$$\Delta = \mathcal{L}_q(\mathcal{T}'_q) - \mathcal{L}_q(\mathcal{T}_q) \le -q^s + 2^{\ell+1}q^{s'} ,$$

and the upper bound follows by requiring $\Delta \ge 0$. ■

C. Proof of Theorem 1

Proof: We assume, without loss of generality, that $q_1 > q$, and we write $q_1 = q(1+\varepsilon)$, $0 < \varepsilon < q^{-1}-1$. In $\mathcal{T}_q$, choose a sufficiently large signature s (the meaning of "sufficiently large" will be specified in the sequel), and a node of signature s at level $L(s)$. Let $s' > s$ be a signature such that $\ell \triangleq L(s') - L(s) \ge 2$. We apply the transformation of Figure 1(A) to $\mathcal{T}_q$, yielding a modified tree $\mathcal{T}'_q$. We claim that when weights are taken with respect to TDGD(q_1), and with an appropriate choice of the parameter ℓ, $\mathcal{T}'_q$ will have strictly lower cost than $\mathcal{T}_q$. Therefore, $\mathcal{T}_q$ is not optimal for TDGD(q_1). To prove the claim, we compare the costs of $\mathcal{T}_q$ and $\mathcal{T}'_q$ with respect to TDGD(q_1). Reasoning as in the proof of the lower bound in Lemma 3, we write

$$\begin{aligned}\Delta &= \mathcal{L}_{q_1}(\mathcal{T}'_q) - \mathcal{L}_{q_1}(\mathcal{T}_q) \le q_1^s - 2^{\ell-2}q_1^{s'} \\ &= q_1^s\left(1-2^{\ell-2}q_1^{s'-s}\right) \le q_1^s\left(1-2^{\ell-2}q_1^{\frac{\ell+1}{\log q^{-1}}}\right), \quad (6)\end{aligned}$$

where the last inequality follows from the upper bound in Lemma 3. It follows from (6) that we can make Δ negative if

$$\ell - 2 + \frac{\ell+1}{\log q^{-1}}\log q_1 > 0.$$

Writing q_1 in terms of q and ε, and after some algebraic manipulations, the above condition is equivalent to

$$\ell > 3\frac{\log q^{-1}}{\log(1+\varepsilon)} - 1 . \quad (7)$$

Hence, choosing a large enough value of ℓ, we get $\Delta < 0$, and we conclude that the tree $\mathcal{T}_q$ is not optimal for TDGD(q_1), subject to an appropriate choice of s, which we discuss next.

The argument above relies strongly on Lemma 3. We recall that in order for this lemma to hold, ℓ and the signature s must satisfy the condition $s \ge 2^{\ell+2}-1$. Now, it could happen that, after choosing ℓ according to (7) and then s according to the condition of Lemma 3, the level $L(s)+\ell$ does not contain $2^{\ell-2}$ signatures s' as required (e.g., when the level is part of a gap). This would force us to increase ℓ, which could then make s violate the condition of the lemma. We would then need to increase s, and re-check ℓ, in a potentially vicious circle. The bound on gap sizes of Lemma 2 allows us to avoid this trap. The bound in the lemma depends only on q and thus, for a given TDGD, it is a constant, say g_q. Thus, first, we choose a value ℓ_0 satisfying the constraint on ℓ in (7). Then, we choose $s \ge 2^{\ell_0+g_q+4}$. Now, we try $\ell = \ell_0, \ell_0+1, \ell_0+2, \ldots,$ in succession, and check whether level $L(s)+\ell$ contains enough of the required signatures. By Lemmas 1 and 2, an appropriate level $L(s')$ will be found for some $\ell \le \ell_0+g_q+2$. For such a value of ℓ, we have $2^{\ell+2}-1 \le 2^{\ell_0+g_q+4}-1 < s$, satisfying the condition of Lemma 3. This condition, in turn, guarantees also that there are at least $2^{\ell-2}$ signatures s' at $L(s')$, as required. ■

Notice that in the above proof, although we show that the tree $\mathcal{T}'_q$ is better than $\mathcal{T}_q$ for TDGD(q_1), thus establishing the sub-optimality of $\mathcal{T}_q$, no claim is made on the optimality of $\mathcal{T}'_q$ (in fact, since the transformation of Figure 1(A) produces an incomplete tree, we know that $\mathcal{T}'_q$ can be improved upon). The proof does not provide a construction of the optimal tree for TDGD(q_1).

IV. Asymptotic redundancy of codes for $q=2^{-1/k}$

Optimal codes for TDGD(q) with $q = 2^{-1/k}$, $k \ge 1$, were described in [1]. In the following theorem, we present a more explicit description of these codes, which will allow us to study the asymptotic behavior of their redundancy in more detail.

Theorem 2: An optimal prefix code C_k for TDGD(q), with $q = 2^{-1/k}$, $k \ge 1$, is given by

$$C_k(i,j) = T_k(i \bmod k, j \bmod k)\cdot G_1(\lfloor \tfrac{i}{k}\rfloor)\cdot G_1(\lfloor \tfrac{j}{k}\rfloor),$$

where G_1 is the unary code, and T_k is an optimal code for the finite alphabet of pairs (i',j') with associated weights $w(i',j') = q^{i'+j'}$, $0 \le i',j' \le k-1$. The code T_k is characterized in terms of its length profile as follows: Define $Q = k^2 - \lceil k(k-1)/4\rceil$, $M = \lceil \log Q\rceil$, and the function

$$\Delta(x) = 2k^2 - 2^{M+1} + x(x+1) - \frac{(k-x-2)(k-x-1)}{2} .$$

[4] Clearly, the condition $s \ge 2^{\ell-1}-1$ would have sufficed to prove the lower bound; the stricter condition of the lemma is required for the upper bound, and was adopted here for uniformity.

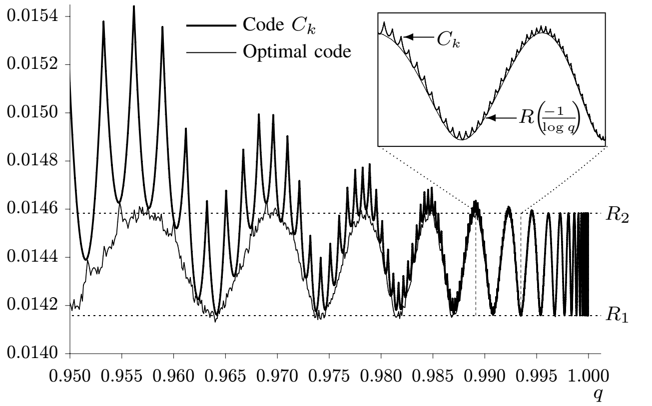

Fig. 2. Redundancy as $q\to 1$ ($k\to\infty$). Dashed lines show the asymptotic limits R_1 and R_2. The inset closes up further on a narrow segment, showing the redundancy of the codes C_k vs. the asymptotic estimate (8).

Let x_0 denote the largest real root of $\Delta(x)$, and $\xi = \lfloor x_0 \rfloor$. Set

$$(j,r) = \begin{cases} (\xi, \left\lfloor \frac{-\Delta(j)+1}{2} \right\rfloor) & \text{if } -\Delta(\xi) \le 2\xi, \\ (\xi+1, 0) & \text{otherwise}, \end{cases}$$

and

$$c_k = k^2 - 2^M + \frac{j(j+1)}{2} + r\,.$$

Then, T_k has $2^M - k^2 + c_k$ codewords at level $M-1$, $2k^2 - 2^M - 3c_k$ codewords at level M, and $2c_k$ codewords at level $M+1$.

A proof of Theorem 2 can be found in [11]. The explicit description of the codes C_k in the theorem allows us to derive expressions for the average code length, and the redundancy of the codes as $k \to \infty$ ($q \to 1$). The asymptotic behavior of the redundancy in this regime is oscillatory, as is also the case for Golomb codes [3]. The limiting behavior of the redundancy is precisely characterized in the following corollary.

Corollary 1: Let $\lambda_k = 2^{M(k)}/k^2$, where $M(k)$ is the value M defined in Theorem 2. As $k \to \infty$, the redundancy of the code C_k at $q = 2^{-1/k}$ is

$$\begin{aligned} R(k) \;&=\; \tfrac{1}{2}\left(1+\log\lambda_k\right) + 2^{1-2\sqrt{\lambda_k - \frac{1}{2}}}\left(1 + \tfrac{2}{\log e}\sqrt{\lambda_k - \tfrac{1}{2}}\right) \\ &\quad -\log(e\log e) + o(1)\,. \end{aligned} \tag{8}$$

Remark. We have $\frac{3}{4} \lessapprox \lambda_k \lessapprox \frac{3}{2}$, where $\lessapprox$ denotes inequality up to asymptotically negligible terms. For large k, as k increases, λ_k sweeps its range decreasing from $\frac{3}{2}$ to $\frac{3}{4}$, at which point $M(k)$ increases by one, and λ_k resets to $\frac{3}{2}$, starting a new cycle.

The limits of oscillation of the function $R(k)$ can be obtained by numerical computation, yielding $R_1 \triangleq \liminf_{k\to\infty} R(k) = 0.014159\ldots$ and $R_2 \triangleq \limsup_{k\to\infty} R(k) = 0.014583\ldots$. The corresponding limits for the redundancy of the Golomb codes are, respectively, $R_1' = 0.025101\ldots$ and $R_2' = 0.032734\ldots$ [3].

Corollary 1 applies to the discrete sequence of redundancy values at the points $q = 2^{-1/k}$. It is not difficult to prove that the same behavior, and in particular the limits R_1 and R_2, apply also to the continuous redundancy curve obtained when using the best code C_k at each arbitrary value of q. This follows from the readily verifiable fact that as q varies in the interval $2^{-1/k} \le q \le 2^{-1/(k+1)}$, the maximal variation in both the code length under C_k and the distribution entropy is bounded by $O(k^{-1})$. The redundancy of the codes C_k for all values of q, as $q \to 1$, is plotted in Figure 2, which also shows the limits R_1 and R_2, and the redundancy of the optimal prefix code for each value of q (which was estimated numerically, to a precision exceeding the resolution of the plot). The figure suggests that the oscillatory behavior observed for the redundancy of C_k might apply also to the redundancy of the optimal code for each value of q. This is clearly true for the limit superior R_2. The question remains open, however, for the limit inferior R_1, which is an upper bound for the limit inferior of the optimal redundancy.

V. Conclusion

We have shown that optimal prefix codes for two-dimensional geometric distributions are singular, in the sense that a given code can be optimal for at most one value of the parameter q. Consequently, it is infeasible to give a compact characterization of optimal codes for all values of q, and practical solutions are likely to be based on countable sequences of codes, such as those presented in [1]. Thus, it is of interest to find other parameter sets for which optimal codes can be characterized, and, in particular, to determine whether optimal codes can be characterized for a dense set, such as the rationals. In a somewhat dual direction, it would be interesting to determine other "natural" families of distributions which exhibit similar singularity properties (the results here are not difficult to extend to d-dimensional geometric distributions, with $d \ge 2$).

References

[1] F. Bassino, J. Clement, G. Seroussi, and A. Viola, "Optimal prefix codes for pairs of geometrically-distributed random variables," in *Proc. ISIT 2006*, (Seattle, Washington), pp. 2667–2671, 2006.

[2] S. W. Golomb, "Run length encodings," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 399–401, 1966.

[3] R. G. Gallager and D. C. Van Voorhis, "Optimal source codes for geometrically distributed integer alphabets," *IEEE Trans. Inf. Theory*, vol. IT-21, pp. 228–230, 1975.

[4] N. Merhav, G. Seroussi, and M. J. Weinberger, "Optimal prefix codes for sources with two-sided geometric distributions," *IEEE Trans. Inf. Theory*, vol. 46, pp. 229–236, 2000.

[5] R. F. Rice, "Some practical universal noiseless coding techniques," Tech. Rep. JPL-79-22, JPL, Pasadena, CA, 1979.

[6] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Proc.*, vol. 9, pp. 1309–1324, 2000.

[7] J. Abrahams, "Code and parse trees for lossless source encoding," *Commun. Inf. Syst.*, vol. 1, pp. 113–146, 2001.

[8] M. J. Golin and K. K. Ma, "Algorithms for constructing infinite Huffman codes," Technical Report HKUST-TCSC-2004-07, HKUST, Hong Kong, China, July 2004.

[9] T. Linder, V. Tarokh, and K. Zeger, "Existence of optimal prefix codes for infinite source alphabets," *IEEE Trans. Inf. Theory*, vol. 43, pp. 2026–2028, 1997.

[10] A. Kato, T. S. Han, and H. Nagaoka, "Huffman coding with an infinite alphabet," *IEEE Trans. Inf. Theory*, vol. 42, pp. 977–984, 1996.

[11] F. Bassino, J. Clement, G. Seroussi, and A. Viola, "Optimal prefix codes for pairs of geometrically-distributed random variables," 2010. full preprint, submitted.