

Capacity-Achieving Sequences for the Erasure Channel

Peter Oswald *

Amin Shokrollahi †

September 26, 2000

Abstract

This paper starts a systematic study of capacity-achieving sequences of low-density parity-check codes for the erasure channel. We introduce a class \mathcal{A} of analytic functions and develop a procedure to obtain degree distributions for the codes. We show various properties of this class which will help us to construct new distributions from old ones. We then study certain types of capacity-achieving sequences and introduce new measures for their optimality. For instance, it turns out that the right-regular sequence is capacity-achieving in a much stronger sense than, e.g., the Tornado sequence. This also explains why numerical optimization techniques tend to favor graphs with only one degree of check nodes. Using our methods, we attack the problem of reducing the fraction of degree 2 variable nodes, which has important practical implications. It turns out that one can produce capacity achieving sequences for which this fraction remains below any constant, albeit at the price of slower convergence to capacity.

1 Introduction

Low-density parity-check codes have attracted a lot of attention lately. Very simple and efficient decoding algorithms and the near capacity performance of the codes with respect to these algorithms have made them one of the most powerful classes of codes known to date. Despite recent advances in the asymptotic analysis of these codes [1, 2, 3], for all nontrivial channels except for the erasure channel it is still unknown whether there exist sequences of these codes that meet the Shannon capacity. The case of the erasure channel is the simplest to analyze, and a thorough understanding of this case seems to be a prerequisite for understanding the more general situation. Moreover, [3] showed that many concepts that were first developed for the erasure channel carry over to the case of other more complicated channels. For this reason, we will start in this paper a systematic study of capacity achieving sequences of low-density codes over the erasure channel.

In [4] the authors introduced a simple algorithm for correcting erasures in a low-density parity-check code. To describe the result, we need some piece of notation. We visualize low-density

*Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA, email: poswald@research.bell-labs.com

†Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA, email: amin@shokrollahi.com

parity-check codes as bipartite graphs between a set of left nodes called variable nodes and a set of right nodes called check nodes. An edge in this graph is said to have left-degree i if it is connected to a variable node of degree i . Similarly, it has right-degree i if it is connected to a check node of degree i . Let λ_i and ρ_i denote the fraction of edges of left-degree and right-degree i , respectively. A degree distribution for the graph is then the pair (λ, ρ) , where $\lambda = \lambda(x) = \sum_i \lambda_i x^{i-1}$ and $\rho = \rho(x) = \sum_i \rho_i x^{i-1}$. The main result of [4] states that their simple recovery algorithm is successful on a random graph with degree distribution (λ, ρ) and initial erasure probability δ if

$$\delta \lambda(1 - \rho(1 - x)) < x \tag{1}$$

on the interval $(0, \delta)$. The capacity of the erasure channel with probability δ is $1 - \delta$.

On the other hand, (λ, ρ) determine the rate of the code as $R = 1 - \int_0^1 \rho(x) dx / \int_0^1 \lambda(x) dx$. We will call R the rate of the pair (λ, ρ) .

The first design goal is thus to produce pairs (λ, ρ) such that the supremum $\delta(\lambda, \rho)$ of all δ that satisfy (1) is very close to $1 - R$. We call a sequence (λ^n, ρ^n) of degree distributions of rate R capacity-achieving (c.a. for short) if the corresponding sequence $\delta^n = \delta(\lambda^n, \rho^n)$ converges to its upper bound $1 - R$ as n tends to infinity.

How can we produce c.a. distributions (λ^n, ρ^n) ? A closer study of two examples of c.a. distributions in the literature is very helpful. For the first sequence, called the *Tornado sequence*, $\lambda^n(x)$ is the initial segment of the series $-\ln(1 - x)$ properly normalized to give $\lambda^n(1) = 1$, and $\rho^n(x)$ is the initial segment of an exponential $\exp(\mu(1 - x))$, where μ is computed from the rate-constraint [4]. For the second sequence, called the *right-regular sequence*, we have $\rho^n(x) = x^n$ and $\lambda^n(x)$ is related to the power series $(1 - x)^{1/m}$ for some m [5].

Roughly speaking, in both cases we start with a $f(x)$ represented by a Taylor series with non-negative coefficients on $[0, 1]$ and satisfying the normalizations $f(0) = 0$, $f(1) = 1$, for which

$$\mathcal{T}f(x) := 1 - f^{-1}(1 - x) \tag{2}$$

has again a converging Taylor series expansion with non-negative coefficients. The existence of the inverse function f^{-1} needed in (2) is automatic from the conditions. We therefore define the set

$$\mathcal{P} := \{f(x) = \sum_1^\infty f_k x^k, \quad x \in [0, 1] \mid f_k \geq 0, f(1) = 1\}, \tag{3}$$

and the set \mathcal{A} as the maximal subset of \mathcal{P} invariant under the action of \mathcal{T} :

$$\mathcal{A} := \{f \in \mathcal{P} \mid \mathcal{T}f \in \mathcal{P}\}. \tag{4}$$

To further motivate these definitions, observe that (1) is essentially equivalent to

$$\lambda(x) \leq \frac{1}{\delta} \mathcal{T} \rho(x), \quad x \in [0, 1]. \quad (5)$$

Indeed, replacing x by $1 - \rho(1 - x)$ in (5) shows its equivalence to $\delta \lambda(1 - \rho(1 - x)) \leq x$, $x \in [0, 1]$. Thus, (1) implies (5) while the validity of (5) for some δ_0 implies (1) at least for all $\delta < \delta_0$. Evidently, defining $\delta(\lambda, \rho)$ as the maximum value of all δ that satisfy (5) leads to the same value as before. For our convenience, we will from now on base our considerations on (5). In particular, we say that (λ, ρ) *affords* δ if (5) holds.

Thus, knowing some $f \in \mathcal{A}$ we can find pairs (λ, ρ) and δ afforded by them by justifying two separate inequalities, $\lambda(x) \leq \delta^{-1} \mathcal{T} f(x)$ and $f(x) \leq \rho(x)$. This can be considered an easy task because the right-hand sides of these inequalities are known and belong to \mathcal{P} , i.e., have already Taylor series with non-negative coefficients. Later in Section 2 we will introduce a general method for obtaining degree distributions from elements of \mathcal{A} along these lines. Note that the nonlinear operator \mathcal{T} has two interesting properties: first, its square is the identity, and second, it commutes with composition \circ in the sense that $\mathcal{T}(f \circ g) = \mathcal{T}g \circ \mathcal{T}f$. Using these identities, we will be able to construct infinitely many c.a. sequences starting from known ones.

From a practical point of view achieving capacity is not sufficient. Suppose that (λ^n, ρ^n) is a sequence of degree distributions of rate R . What we would like to know is how fast, if at all, the maximal $\delta = \delta^n$ afforded by (λ^n, ρ^n) converges to $1 - R$ as $n \rightarrow \infty$. This problem was studied in [5], where it was shown that if a_r is the average degree of the check nodes ($1/a_r = \int_0^1 \rho(x) dx$) and $\epsilon = 1 - \delta/(1 - R)$, where δ is afforded by (λ, ρ) , then $a_r \geq \log(\epsilon)/\log(R)$, or, equivalently,

$$\epsilon \geq R^{a_r}.$$

In that paper a sequence of degree distributions is called *asymptotically quasi-optimal* if a_r stays bounded by $\mu \log(1/\epsilon)$ where μ is a constant depending on R . It was shown that the Tornado sequence and the right-regular sequence are both asymptotically quasi-optimal.

A much more refined notion of optimality would be to directly compare R^{a_r} and ϵ . We call a sequence of degree distributions *asymptotically optimal* if ϵ/R^{a_r} stays bounded by some constant depending on the rate. Obviously, asymptotical optimality implies asymptotical quasi-optimality. Using this notion, we will show later in Section 3 that the right-regular sequence is asymptotically optimal while the Tornado sequence turns out to be only asymptotically quasi-optimal. Using the

composition operation, we will also construct infinitely many sequences of asymptotically quasi-optimal degree distributions.

It is proved in [6] that for a c.a. sequence (λ^n, ρ^n) the sequence $\delta^n \lambda^n (1 - \rho^n (1 - x)) - x$ and all its derivatives up to any fixed order k converge uniformly to zero in $[0, 1)$ as n goes to infinity. Examining the first derivative, this shows that the product $\delta^n \lambda_2^n (\rho^n)'(1)$ converges to 1 as n goes to infinity. As a result, no c.a. sequence has the property that for all sufficiently large n we have $\lambda_2^n = 0$.

On the other hand, from a practical point of view it is advantageous to have as little degree 2 variable nodes as possible, since these are the nodes that get corrected last. In particular, if the fraction of degree 2 nodes is at most $1 - R$, then one can construct the graph in such a way that all the redundant symbols fall within the set of degree 2 variable nodes, and one does not need to correct these nodes at all. This greatly accelerates the decoding procedure. Using our techniques, we will construct in Section 4 asymptotically quasi-optimal sequences of degree distributions whose fraction of degree 2 variable nodes is strictly less than $1 - R$.

This paper is organized as follows.

In the next section, we will derive the basic properties of the set \mathcal{A} and describe how to obtain degree distributions from it. In Section 3 we will address the problem of convergence speed of c.a. sequences of degree distributions, based on suitable elements of \mathcal{A} . Section 4 discusses the problem of reducing the fraction of degree 2 variable nodes. In Section 5 we will provide numerical evidence supporting our theoretical findings. In conclusion, we summarize our results and pose some open problems. Issues regarding the feasibility of the proposed algorithm, further properties of \mathcal{T} , as well as some algebraic criteria for $f \in \mathcal{A}$ are summarized in the appendices.

2 Degree distributions from \mathcal{A}

We start with more details on the properties of \mathcal{A} defined by (4). Note that by definition $f \in \mathcal{A}$ implies absolute convergence of the Taylor series of both f and $\mathcal{T}f$. From this and properties of power series and inverse functions, the following lemma is straightforward.

Lemma 1. (a) $f \in \mathcal{A}$ if and only if $\mathcal{T}f \in \mathcal{A}$.

(b) \mathcal{T} is an involution, i.e., $\mathcal{T}^2 f = f$.

(c) If $f, g \in \mathcal{A}$ then the composition $(f \circ g)(x) = f(g(x))$ also belongs to \mathcal{A} , and

$$\mathcal{T}(f \circ g) = \mathcal{T}g \circ \mathcal{T}f.$$

(d) If $f \in \mathcal{A}$ then

$$S_{a,b}f(x) = \frac{f(ax+b) - f(b)}{f(a+b) - f(b)} \in \mathcal{A}$$

for any $0 < a \leq a+b \leq 1$.

(e) For any $f \in \mathcal{P}$, we have

$$\int_0^1 \mathcal{T}f(x) dx = \int_0^1 f(x) dx.$$

It is easy to check that each of the families of functions

$$f(x) = x^n, \quad n = 1, 2, \dots, \quad (6)$$

$$f(x) = \frac{e^{ax} - 1}{e^a - 1}, \quad a > 0, \quad (7)$$

and

$$f(x) = \frac{(1-b)x}{1-bx}, \quad 0 < b < 1, \quad (8)$$

belong to \mathcal{A} . The first two families correspond to the right-regular and Tornado sequences, while (8) has the property $\mathcal{T}f = f$.¹ Many more examples can be constructed from these families by using properties (c) and (d) of Lemma 1. For example, applying $S_{a,b}$ to $f(x) = x^2$ generates the family of quadratic polynomials in \mathcal{A} :

$$f(x) = cx^2 + (1-c)x, \quad 0 < c < 1. \quad (9)$$

Below, we will systematically explore sequences of the form $f(x) = \phi(x^n)$, $n = 1, 2, \dots$, generated by composition of some $\phi \in \mathcal{A}$ with the sequence (6). More examples can be found in later sections.

To our knowledge, efficient criteria stated, e.g., in terms of the Taylor coefficients of a function $f \in \mathcal{P}$ to decide whether or not it belongs to the class \mathcal{A} are not known in general. Some comments on this question are made in Appendix D.

What we outline next is how to produce for given $f \in \mathcal{A}$ and $R \in (0, 1)$ a suitable pair of left- and right-degree distributions (λ, ρ) of rate R . Roughly speaking, ρ and λ are defined by scaled

¹Actually, the Tornado sequence introduced in [4] is slightly different, but the asymptotic properties of that sequence and ours are identical.

sections of the Taylor expansion of f and $g =: \mathcal{T}f$, respectively. A simplifying assumption for our procedure to work is that f has an analytic extension to a neighborhood of $x = 1$. This assumption is not severe in practice, as it is satisfied for all examples considered in this paper and in particular for polynomial f . In particular, it implies that the Taylor coefficient g_1 is always strictly positive.

Throughout the paper, we use the notation $[t]$ for the integer part of a real number t , and set $\{t\} = t - [t] \in [0, 1)$. If $f(x) = \sum_{k \geq 1} f_k x^k \in \mathcal{P}$, then we set

$$T_t f(x) := \sum_{k \leq [t]} f_k x^k + \{t\} f_{[t]+1} x^{[t]}, \quad \hat{T}_t f(x) = \frac{T_t f(x)}{T_t f(1)} \quad (10)$$

for the Taylor polynomial $T_t f$ of degree t of $f \in \mathcal{A}$ and its normalization defined by $\hat{T}_t f(1) = 1$, respectively. Note that this extends the standard definition of Taylor polynomials for integer degree to non-integer t in a continuous way. We also set

$$\hat{I}f(t) = \int_0^1 \hat{T}_t f(x) dx. \quad (11)$$

Observe that $\hat{T}_t f(x)$ and $\hat{I}f(t)$ are well-defined only for $t > k_0 - 1$, where k_0 is defined by the index of the first non-zero Taylor coefficient f_k of $f \in \mathcal{P}$.

Let $f = \sum_{k \geq 1} f_k x^k$ and $g = \sum_{k \geq 1} g_k x^k$ are defined by absolutely converging Taylor series on $[0, 1]$. Then we say that g supercedes f and denote it by $f \prec g$, if for all $m \geq 1$ we have $\sum_{k \leq m} f_k \leq \sum_{k \leq m} g_k$. We have the following

Lemma 2. (1) *Suppose that $f \prec g$. Then $f(x) \leq g(x)$ on $[0, 1]$.*

(2) *If $f \in \mathcal{P}$ and $k_0 - 1 < t \leq s < \infty$, then $f(x) \leq \hat{T}_s f(x) \leq \hat{T}_t f(x)$ pointwise for $x \in [0, 1]$, and $\int_0^1 f(x) dx \leq \hat{I}f(s) \leq \hat{I}f(t)$.*

Proof. (1) Set $s_m := \sum_{k \leq m} f_k$ and $s'_m := \sum_{k \leq m} g_k$. The inequality $x^k - x^{k+1} \geq 0$, $x \in [0, 1]$, and an Abel transformation prove (1):

$$f(x) = \sum_{k=1}^{\infty} (s_k - s_{k-1}) x^k = \sum_{k=1}^{\infty} s_k (x^k - x^{k+1}) \leq \sum_{k=1}^{\infty} s'_k (x^k - x^{k+1}) = \sum_{k=1}^{\infty} (s'_k - s'_{k-1}) x^k = g(x).$$

(2) Since $0 < T_t f(1) \leq T_s f(1) \leq 1$, an easy calculation reveals that $f \prec \hat{T}_s f \prec \hat{T}_t f$, and (2) follows from (1). \square

Our algorithm for computing the pair (λ, ρ) is as follows.

Algorithm 1. Given $f \in \mathcal{A}$, $R \in (0, 1)$, and an integer N , this algorithm computes a pair (λ, ρ) of rate R and a real number δ such that $\rho(x)$ is of degree N , the maximal value of $\rho(x) - f(x)$ on $(0, 1)$ is bounded above by $\sum_{k>N} f_k / \sum_{k \leq N} f_k$, and such that (λ, ρ) affords δ .

- (1) If $f(x)$ is a polyomial, then set $\rho(x) := f(x)$, otherwise set $\rho(x) := \hat{T}_N f(x)$. Let $a_r := 1 / \int_0^1 \rho(x) dx$.
- (2) Let $g(x) = \mathcal{T}f(x) = \sum_{k \geq 1} g_k x^k$. Using the coefficients g_k , compute t such that $\hat{I}(t) = \frac{1}{(1-R)a_r}$ where $\hat{I}(t) := \hat{I}g(t)$ is defined by (11).
- (3) Set $\delta := T_t g(1)$ and $\lambda(x) := \hat{T}_t g(x) = T_t g(x) / \delta$.

Proposition 1. Suppose that $f(x)$ is analytic at 1, and that a_r computed in Step (1) of the above algorithm satisfies $(1-R)a_r \geq 2$. Then the algorithm correctly computes its output, i.e., the output satisfies the specifications of the algorithm.

Proof. (1) First we prove the upper bound for the non-negative function $\rho(x) - f(x)$ on $(0, 1)$. Indeed, from Lemma 2(2) we have for $f \in \mathcal{P}$

$$f(x) \leq \rho(x) = \hat{T}_N f(x) \leq \frac{f(x)}{T_N f(1)}, \quad x \in [0, 1], \quad (12)$$

from which the assertion follows immediately.

(2) Next we prove that the computation of Step (2) is always successful. To see this, recall that $g_1 > 0$. Thus, the function $\hat{I}(t)$ is well-defined for $t > 0$, continuous and monotonically decreasing by Lemma 2. Note that

$$\hat{I}(k) = \frac{\sum_{l=1}^k g_l / (l+1)}{\sum_{l=1}^k g_l}$$

for integer $t = k$, and we have $\hat{I}(1) = 1/2$ and $\hat{I}(t) \geq \lim_{t \rightarrow \infty} \hat{I}(t) = \int_0^1 g(x) dx = \int_0^1 f(x) dx \geq 1/a_r$. Hence, if $\frac{1}{(1-R)a_r} \leq 1/2$, there is a value of $t \geq 1$ for which $\hat{I}(t)$ equals $\frac{1}{(1-R)a_r}$. More precisely, such a value $t = k + s$ can be found by first searching for the smallest integer k such that $\hat{I}(k) \geq \frac{1}{(1-R)a_r} \geq \hat{I}(k+1)$, and then determining $s \in [0, 1]$ such that

$$\sum_{l=1}^k \frac{g_l}{l+1} + s \frac{g_{k+1}}{k+2} = \frac{1}{(1-R)a_r} \left(\sum_{l=1}^k g_l + s g_{k+1} \right).$$

Thus, all what is needed for Step (2) to succeed, is to compute the Taylor coefficients g_k as we go along. Since $f(x)$ is analytic at 1, the coefficients of $g(x)$ can be computed via the algorithm outlined in Appendix A.

(3) Now we show that (λ, ρ) affords δ . Indeed, we have

$$\lambda(x) = \delta^{-1} T_t g(x) \leq \delta^{-1} \mathcal{T} f(x) = \delta^{-1} (1 - f^{-1}(1 - x)) \leq \delta^{-1} (1 - \rho^{-1}(1 - x)),$$

which is equivalent to (5). The last estimation step follows from the definition of \mathcal{T} and the lower inequality in (12). This completes the proof. \square

In summary, our procedure outlined in Step (1)-(3) above is applicable to any rate $0 < R < 1$ and $f \in \mathcal{A}$ satisfying the additional analyticity condition at $x = 1$, i.e., the validity of (28) for all small x , and the compatibility condition

$$\int_0^1 f(x) dx < (1 - R)/2. \tag{13}$$

Indeed, (13) makes sure that, by choosing N large enough, we can always satisfy the remaining condition $(1 - R)a_r \geq 2$ in Proposition 1. It leads to a pair (λ, ρ) of left- and right-degree distributions of rate R , and to a δ afforded by it. Applying the scheme to suitable sequences f^n from \mathcal{A} , we can construct c.a. sequences (λ^n, ρ^n) for any rate R , in a systematic way, and obtain quantitative estimates for the convergence speed in $\delta^n \rightarrow 1 - R$. This will be demonstrated in the following sections. A straightforward numerical implementation of Algorithm 1 was used to compute the examples in Section 5. Its running time depends heavily on how long it takes to compute all the coefficients of g_k necessary (see Appendix A), as we need to satisfy $\hat{I}(t) = \frac{1}{a_r(1-R)}$, i.e.,

$$\sum_{k \leq [t]} \frac{g_k}{k+1} + \{t\} \frac{g_{[t]+1}}{[t]+2} = \frac{1}{(1-R)a_r} \left(\sum_{k \leq [t]} g_k + \{t\} g_{[t]+1} \right),$$

in Step (2) above. This depends ultimately on the function $f(x)$ and, in particular, on its behavior near $x = 1$. For most of our applications, however, we have analytic expressions for the g_k which makes it rather trivial to estimate the right value of t .

3 Convergence speed of c.a. sequences

This section is devoted to estimating how close the performance of the decoding algorithm for the code given by the pair (λ, ρ) is to the capacity of the erasure channel. In other words, we want to study how close the maximal δ afforded by this pair is to $1 - R$ where R is the rate of the pair. We introduce

$$\epsilon(\lambda, \rho) := 1 - \frac{\delta(\lambda, \rho)}{1 - R}.$$

As stated in the introduction, the following lower bound holds [5]

$$\epsilon(\lambda, \rho) \geq R^{a_r}, \quad a_r = \left(\int_0^1 \rho(x) dx \right)^{-1}. \quad (14)$$

This shows, that in order to have $\epsilon \rightarrow 0$ or, equivalently, $1 - \delta \rightarrow R$, we must have $a_r \rightarrow \infty$ for fixed R . Also, (14) suggests the use of either

$$\mu(\lambda, \rho) = \frac{a_r \log(R)}{\log(\epsilon(\lambda, \rho))} \quad (15)$$

or

$$\Delta(\lambda, \rho) = \frac{\epsilon(\lambda, \rho)}{R^{a_r}} \quad (16)$$

to quantitatively measure the closeness of capacity and rate of any particular pair (λ, ρ) , hence also of c.a. sequences. In refining the definition introduced in [5], we will call a sequence (λ^n, ρ^n) *asymptotically quasi-optimal with constant $\mu \geq 1$* if

$$\limsup_{n \rightarrow \infty} \mu(\lambda^n, \rho^n) = \mu.$$

Clearly, the closer μ is to 1 the faster δ^n converges to $1 - R$. In addition, we call a sequence (λ^n, ρ^n) *asymptotically optimal with constant $\Delta \geq 1$* if

$$\limsup_{n \rightarrow \infty} \Delta(\lambda^n, \rho^n) = \Delta.$$

Trivially, asymptotically optimal sequences with any Δ are asymptotically quasi-optimal with constant 1. As it turns out these definitions allow to better classify c.a. sequences, and see distinctive differences between, e.g., the Tornado and right-regular sequences.

Let $\phi \in \mathcal{A}$ be analytic at $x = 1$, and $0 < R < 1$ be given. We consider the sequence (λ^n, ρ^n) of fixed rate R generated from the sequence

$$f^n(x) := \phi(x^n), \quad n \geq n_0, \quad (17)$$

by Algorithm 1. For short, we say that this sequence is generated by ϕ . Let us comment on the feasibility of this definition. Obviously, all f^n satisfy the additional analyticity condition at $x = 1$. We assume that n_0 is chosen such that the compatibility condition (13) holds. This is always possible since

$$n \int_0^1 \rho^n(x) dx = n \int_0^1 \phi(x^n) dx = \int_0^1 \frac{\phi(t)}{t} t^{1/n} dt \rightarrow \int_0^1 \frac{\phi(t)}{t} dt, \quad n \rightarrow \infty, \quad (18)$$

which implies $\int_0^1 f^n(x) dx \rightarrow 0$ as $n \rightarrow \infty$. Finally, if ϕ is a polynomial of degree K then we choose $N = Kn$ and set $\rho^n = f^n$; otherwise $N = N(n)$ is such that a sufficiently fast convergence in $T_N f^n(1) \rightarrow 1$ can be guaranteed. More precisely,

$$T_N f^n(1) \geq \frac{1}{1 + \phi(R^n)}, \quad n \geq n_0, \quad (19)$$

is sufficient.

Theorem 1. *If a sequence (λ^n, ρ^n) of rate $R \in (0, 1)$ is generated by $\phi \in \mathcal{A}$ as described above, then it is asymptotically quasi-optimal with constant*

$$\mu \leq \mu^\phi := \frac{1}{\int_0^1 \frac{\phi(x)}{x} dx}.$$

Before elaborating on the proof of this theorem, we introduce an auxiliary construction which will be used in it. Suppose that $f \in \mathcal{A}$ and let $g := \mathcal{T}f$. For $a \in (0, 1]$ we define $g_a(x) := g(ax)/g(a)$. By Lemma 1, we have $g_a \in \mathcal{A}$. Let $I(a) := \int_0^1 g_a(x) dx$. The functions $g_a(x)$ and $I(a)$ possess properties similar to those of their counterparts $\hat{T}_t g(x)$ and $\hat{I}(t)$ appearing in Algorithm 1.

Lemma 3. *For $a \in (0, 1]$ we have:*

- (1) $g(x) \leq g_a(x) \leq \frac{g(x)}{g(a)}$.
- (2) $I(a)$ is continuous and monotonically decreasing.
- (3) Suppose that t and a are such that $I(a) = \hat{I}(t)$. Then $g(a) \leq T_t g(1)$.

Proof. (1) Obvious by Lemma 2(1).

(2) By Lemma 2 it suffices to show that for $0 < b \leq a \leq 1$ we have $g_a \prec g_b$. Let $s_m(x) := \sum_{k \leq m} g_k x^k$ and $\hat{s}_m(x) := \sum_{k > m} g_k x^k$. Then, we need to show that for all $m \geq 1$:

$$\frac{s_m(a)}{s_m(a) + \hat{s}_m(a)} \leq \frac{s_m(b)}{s_m(b) + \hat{s}_m(b)},$$

or, equivalently, $s_m(a)\hat{s}_m(b) \leq s_m(b)\hat{s}_m(a)$. Now it is easily seen that for any $\ell > m$ we have $b^\ell s_m(a) \leq a^\ell s_m(b)$ (since $b^{\ell-k} \leq a^{\ell-k}$ for any $k < \ell$). This proves the result.

(3) In light of the monotonicity of the functions $\hat{I}(t)$ and $I(a)$, the statement is equivalent to proving that $g(a) = T_t g(1)$ implies $I(a) \leq \hat{I}(t)$ or, what is the same by definition of these functions, that

$$g(a) = T_t g(1) \quad \implies \quad \int_0^1 g(ax) dx \leq \int_0^1 T_t g(x) dx.$$

We use again Lemma 2(1) which reduces this statement to showing $g(ax) \prec T_t g(x)$ provided that $g(a) = T_t g(1)$. Set $k = \lfloor t \rfloor$ and let us look at the partial sums s_m and s'_m of the sequences of Taylor coefficients of $g(ax) = \sum_{l=1}^{\infty} (g_l a^l) x^l$ and $T_t g(x)$, respectively. For $m \leq k$, we have

$$s_m = \sum_{l=1}^m g_l a^l \leq \sum_{l=1}^m g_l = s'_m$$

since $0 < a \leq 1$, while for $m \geq k + 1$ we obviously have $s_m \leq g(a) = T_t g(1) = s'_m$ by definition of k and (10). This gives the assertion of (3). \square

Proof. (of Theorem 1). Let us first assume that ϕ is a polynomial. Let $f^n = \phi(x^n)$ as above, and $N = Kn$, where K is the degree of ϕ . Further, let the rate R be fixed. We apply Algorithm 1 on the triple f^n, N, R and obtain $\rho^n = f^n$, λ^n , and δ^n . Note that according to (18) the quantity $a_r^n := (\int_0^1 \rho^n(x) dx)^{-1}$ satisfies

$$\frac{a_r^n}{n} \rightarrow \mu^\phi, \quad n \rightarrow \infty. \quad (20)$$

For the most part of the proof we will suppress dependencies on n to ease the notation. Thus, the above data will be abbreviated by f, ρ, λ, δ , and a_r .

The aim of the proof is to show that $\lim_{n \rightarrow \infty} a_r \log(R) / \log(1 - \delta / (1 - R))$ is at most μ^ϕ . Since it is somewhat difficult to get a handle on δ directly, we will use the auxiliary construction above. That is, we will construct a number $\alpha \leq \delta$ and prove that $a_r \log(R) / \log(1 - \alpha / (1 - R))$ has a limit which is at most μ^ϕ . This will be sufficient to prove our claim.

Let $0 < b < 1$ be such that $I(b) = \frac{1}{(1-R)a_r}$. Such a b exists for $n \geq n_0$ since according to Lemma 3(2) the function $I(b)$ is continuous and monotonous, with a range given by

$$\int_0^1 g(x) dx = I(1) \leq I(b) \leq I(+0) = 1/2,$$

which contains the right-hand side of the equation strictly in its interior (the upper limit for the range is $1/2$ since $g_1 > 0$ by the analyticity assumption).

Since $\delta = T_t g(1)$ for a $t \geq 1$ such that $\hat{I}(t) = \frac{1}{(1-R)a_r}$ (see Algorithm 1), Lemma 3(3) implies that $\alpha := g(b) \leq \delta$. Moreover,

$$\begin{aligned} \int_0^b g(x) dx &= b - \int_{1-b}^1 f^{-1}(x) dx = 1 - \int_{f(1-\alpha)}^1 f(x) dx - f(1-\alpha) \\ &= \int_{1-\alpha}^1 f(x) dx + (1-\alpha)f(1-\alpha) - f(1-\alpha) = \frac{1}{a_r} - \int_0^{1-\alpha} f(x) dx - \alpha f(1-\alpha), \end{aligned}$$

where we have applied the identity $1 - \int_{f(t)}^1 f^{-1}(x)dx = tf(t) + \int_t^1 f(x)dx$ valid for all $f \in \mathcal{P}$ with $t = 1 - \alpha$, and used the fact that $1 - b = f(1 - \alpha)$.

Now we can substitute into our equation for b

$$\frac{1}{(1-R)a_r} = \int_0^1 g_b(x)dx = \frac{1}{g(b)b} \int_0^b g(x)dx,$$

find an expression for $\alpha = g(b)$ from it, and use the result to estimate $\epsilon := 1 - \delta/(1-R)$:

$$\epsilon \leq 1 - \frac{\alpha}{1-R} = 1 - \frac{a_r}{b} \int_0^b g(x)dx = \frac{f(1-\alpha)}{1-f(1-\alpha)} \left(a_r \left(\alpha + \frac{\int_0^{1-\alpha} f(x)dx}{f(1-\alpha)} \right) - 1 \right). \quad (21)$$

But for any $\phi(x) = \sum_{k=1}^{\infty} \phi_k x^k \in \mathcal{A}$, $t \in (0, 1)$, and $n \geq 1$ we have

$$0 < f(t) = \sum_{k=1}^{\infty} \phi_k t^{kn} \leq t^n \sum_{k=1}^{\infty} \phi_k = t^n,$$

and

$$0 < \int_0^t f(x) dx = \sum_{k=1}^{\infty} \phi_k \frac{t^{kn+1}}{kn+1} \leq (n+1)^{-1} t f(t).$$

We use these inequalities with $t = 1 - \alpha$ and (20) in (21) to obtain

$$\epsilon \leq \frac{(1-\alpha)^n a_r}{1-(1-\alpha)^n} \left(\alpha + \frac{1-\alpha}{n+1} \right) \leq C \mu^\phi (1-\alpha)^n (n\alpha+1), \quad n \rightarrow \infty, \quad (22)$$

where $C \geq 1$ is an absolute constant.

This inequality is sufficient for our purposes. Indeed, if we temporarily assume that

$$n\alpha \rightarrow \infty, \quad n \rightarrow \infty, \quad (23)$$

is already established, we conclude from (22) that ϵ converges to 0, and hence the sequence (λ^n, ρ^n) obtained from $f(x^n)$ by Algorithm 1 is c.a. This follows from

$$(n\alpha+1)(1-\alpha)^n = (n\alpha+1) \left(1 - \frac{n\alpha}{n}\right)^n \leq (n\alpha+1) e^{-cn\alpha} \rightarrow 0, \quad n \rightarrow \infty,$$

which holds for some $c < 1$. In turn, $\epsilon \rightarrow 0$ implies $\alpha \rightarrow 1 - R > 0$ which shows that (23) holds in a much stronger sense. Finally, from this we see that the dominating term in the previous upper estimate for ϵ is $(1-\alpha)^n$, so we obtain the assertion of Theorem 1:

$$\mu(\lambda^n, \rho^n) = \frac{a_r \log(R)}{\log(\epsilon)} \leq (1 + o(1)) \frac{a_r}{n} \frac{\log(R)}{\log(1-\alpha)} \rightarrow \mu^\phi, \quad n \rightarrow \infty. \quad (24)$$

We still need to establish (23). On the contrary, assume that $\alpha = O(1/n)$ for $n \rightarrow \infty$. From our above transformations we have

$$\frac{n+1}{a_r(1-R)} = \frac{n+1}{bg(b)} \int_0^b g(x)dx = \frac{n+1}{\alpha(1-f(1-\alpha))} \left(\int_{1-\alpha}^1 f(x)dx - \alpha f(1-\alpha) \right).$$

Now recall that $f(x) = \phi(x^n)$, where $\phi \in \mathcal{A}$ is analytic at $x = 1$. Thus, $1 - c_0x \leq \phi(1-x) \leq 1 - \hat{c}_0x$ for $0 \leq x \leq \alpha$, where $c_0 = \phi'(1)$ is independent of n , and $c_0 - c\alpha \leq \hat{c}_0 \leq c_0$ holds for all sufficiently small α and some absolute constant $c > 0$. Without loss of generality, we can assume that $c_0 > 1$ since $c_0 = 1$ implies $\phi(x) = x$ (this case is covered by the stronger Theorem 3 below for which an independent proof is given in Appendix B). This gives

$$f(1-\alpha) = (1-c_1\alpha)^n, \quad \int_{1-\alpha}^1 f(x)dx = \frac{1 - (1-c_2\alpha)^{n+1}}{n+1},$$

where $c_1, c_2 = c_0 + O(\alpha)$ are new constants depending on α , and tending to $c_0 > 1$ as n tends to infinity. Substitution yields

$$\begin{aligned} \frac{n+1}{a_r(1-R)} &= \frac{1 - (1-c_2\alpha)^{n+1} - (n+1)\alpha(1-c_1\alpha)^n}{\alpha(1 - (1-c_1\alpha)^n)} \\ &= \frac{\sum_{k=0}^n (c_2(1-c_2\alpha)^k - (1-c_1\alpha)^n)}{c_1\alpha \sum_{k=0}^{n-1} (1-c_1\alpha)^k}, \end{aligned}$$

where we have used the identity $1 - y^k = (1-y) \sum_{l=0}^{k-1} y^l$ twice. Due to the assumptions on the asymptotic behavior of α , c_1 , and c_2 (e.g., we have $(c_1 - c_2)\alpha = O(\alpha^2) = O(n^{-2})$), we see that $1 \geq (1 - c_1\alpha)^n = (1 + o(1))(1 - c_2\alpha)^n \geq c > 0$ for some constant c independent of n . Thus, the numerator in the right-hand side of last displayed equation is bounded from below by $cn(c_2 - (1 + o(1))) \geq c'n$, where we have used that $c_2 \rightarrow c_0 > 1$. On the other hand, the denominator is majorized by $c_2\alpha n \leq C'$. Since the constants c', C' are independent of n , we see that the whole expression in the right-hand side tends to infinity which contradicts the fact that the left-hand side tends to the finite limit $\mu^\phi/(1-R)$, see (20).

This completes the proof of Theorem 1 for polynomial ϕ . The general case goes through in the same way, by choosing the maximal degree $N = N(n)$ for defining ρ^n in Step (1) of our algorithm large enough such that the asymptotic estimates for the resulting α and ϵ are preserved. From a theoretical point of view, it is important to make sure that N can be chosen *a priori*, i.e., solely based on knowledge about the input f^n and R (in practice, this is not a very pressing issue, as we could take by default N such that $T_N f^n(1) = 1$ within machine precision). Again, we neglect the dependence on n in the notation. Note first that the choice of N enters Step (2)-(3) of

Algorithm 1 only through the value of a_r . More precisely, instead of solving $\hat{I}(t) = \frac{1}{(1-R)a_r}$ for t , where $a_r = (\int_0^1 f(x) dx)^{-1}$, and then setting $\delta = T_t g(1)$, we use

$$\hat{I}(\tilde{t}) = \xi \frac{1}{(1-R)a_r}, \quad \tilde{\delta} = T_{\tilde{t}} g(1),$$

where $1 \leq \xi \leq 1/T_N f(1)$ by (12). Following our above strategy, we now define \tilde{b} and $\tilde{\alpha} := g(\tilde{b})$ by $I(\tilde{b}) = \xi \frac{1}{(1-R)a_r}$. Then we have $\tilde{\alpha} \geq \tilde{\delta}$, and repeating the estimation leading to (21) we get

$$\tilde{\epsilon} := 1 - \frac{\tilde{\delta}}{1-R} \leq 1 - \frac{\tilde{\alpha}}{1-R} = \frac{f(1-\tilde{\alpha})}{\xi(1-f(1-\tilde{\alpha}))} \left(\frac{\xi-1}{f(1-\tilde{\alpha})} + a_r \left(\tilde{\alpha} + \frac{\int_0^{1-\tilde{\alpha}} f(x) dx}{f(1-\tilde{\alpha})} \right) - \xi \right). \quad (25)$$

Now, in order to preserve the above asymptotic estimates leading to the statement of Theorem 1, it is sufficient to ensure that $\xi \rightarrow 1$ is such that $\xi-1 \leq f(1-\tilde{\alpha})$. Indeed, the expression $(\xi-1)/f(1-\tilde{\alpha})-\xi$ is then negative, and can be neglected in (25). Thus, the analog of (22) holds (with an additional factor $1/\xi = 1 + o(1)$ in the right-hand side). Since $n\tilde{\alpha} \rightarrow \infty$ can be proved as before, we get the desired result. Since (19) implies

$$\xi - 1 \leq \frac{1}{T_N f(1)} - 1 \leq \phi(R^n) \leq \phi((1-\tilde{\alpha})^n) = f(1-\tilde{\alpha}),$$

we have completed the proof. □

We remark that Theorem 1 provides only an upper bound for the constant μ of quasi-optimality. However, the numerical evidence presented in Section 5 strongly suggests that the above estimation techniques are rather sharp. They can also be applied to other families from \mathcal{A} . The following result will be stated without proof. It addresses the sequence (7), and gives a more quantitative statement on the asymptotic quasi-optimality of the Tornado sequence previously established in [4].

Theorem 2. *Let (λ^a, ρ^a) be the sequence of degree distributions of rate $R \in (0, 1)$ obtained from the family $f^a(x) = (e^{ax} - 1)/(e^a - 1)$, $a \rightarrow \infty$, by Algorithm 1. This sequence is asymptotically quasi-optimal with constant*

$$\mu \leq \mu_{\text{Torn}}(R) := \frac{\ln(1/R)}{1-R}.$$

In contrast to the bounds in Theorem 1, the bound of Theorem 2 is rate-dependent. Note that $\mu_{\text{Torn}}(R) \rightarrow \infty$ as $R \rightarrow 0$ and $\mu_{\text{Torn}}(R) \rightarrow 1$ as $R \rightarrow 1$. Numerical results for this family are given in Section 5. As a final comment, let us mention that the family (8) does not produce a quasi-optimal c.a. sequence; this is left as an exercise to the reader.

The only sequence (17) that is guaranteed by Theorem 1 to lead to an asymptotically quasi-optimal c.a. sequence (λ^n, ρ^n) with constant 1 is the right-regular sequence generated by $f^n(x) = x^n$. Indeed, for any $\phi \in \mathcal{A}$ we have $\phi(x) \prec x$ and therefore by Lemma 2 $\mu^\phi \geq 1$, with equality only for $\phi(x) = x$. For the right-regular sequence, a stronger result holds:

Theorem 3. *Let (λ^n, ρ^n) be the right-regular sequence of degree distributions of rate $R \in (0, 1)$ obtained from the family (6) by Algorithm 1. This sequence is asymptotically optimal with constant*

$$\Delta = e^\gamma = 1.78107241\dots,$$

where $\gamma = 0.57721566\dots$ is the Euler constant.

The proof of this theorem can be found in Appendix B.

Compared with the statements before it and in light of the numerical evidence, Theorem 3 exhibits the excellent asymptotic behavior of the right-regular sequence. We do not know of any asymptotically optimal c.a. sequence with a constant $\Delta < e^\gamma$.

Appendix C gives a more general view of composing c.a. sequences with elements in \mathcal{A} .

4 Reduction of degree 2 nodes

In this short section, we concentrate on the asymptotic behavior of the fraction of degree 2 variable nodes of the c.a. sequences generated by the family (17). As discussed in the introduction, this fraction should be as small as possible to allow fast decoding. For an arbitrary left-degree distribution λ , it is defined by the coefficient Λ_2 of the Taylor series

$$\frac{\int_0^t \lambda(x) dx}{\int_0^1 \lambda(x) dx} = \Lambda_2 t^2 + \Lambda_3 t^3 + \dots$$

The stability condition [5, 3] states in the case of the erasure channel that if (λ, ρ) affords δ , then $\delta \lambda_2 \rho'(1) \leq 1$, where $\lambda(x) = \lambda_2 x + O(x^2)$. Hence, we have

$$\Lambda_2 = \frac{(1-R)a_r \lambda_2}{2} \leq \frac{(1-R)a_r}{2\delta(\lambda, \rho)\rho'(1)}.$$

If we apply this inequality to a c.a. sequence (λ^n, ρ^n) , then using $\delta^n = \delta(\lambda^n, \rho^n) \rightarrow 1-R$ we obtain

$$\lim_{n \rightarrow \infty} \Lambda_2^n \leq \lim_{n \rightarrow \infty} \frac{a_r^n}{2(\rho^n)'(1)}, \tag{26}$$

if the limits exist (otherwise, the relation holds at least with \lim replaced by \limsup or \liminf). Moreover, if the c.a. sequence (λ^n, ρ^n) is generated from a sequence $f^n \in \mathcal{A}$ by Algorithm 1, then we can replace ρ by f and prove the existence of the limits and equality in (26).

Theorem 4. *Let the sequence (λ^n, ρ^n) of rate R be generated by $\phi \in \mathcal{A}$, where ϕ is analytic at $x = 1$. Then*

$$\lim_{n \rightarrow \infty} \Lambda_2^n = \Lambda^\phi := \frac{\mu^\phi}{2\phi'(1)},$$

where $\mu^\phi = (\int_0^1 x^{-1}\phi(x) dx)^{-1}$ is the same as in Theorem 1.

Proof. Obviously, $(\rho^n)'(1) = n\phi(1)$ and from (20) we have $a_r^n/n \rightarrow \mu^\phi$. Substituting into (26) which holds in this case with equality, we get the statement. \square

Theorem 1 and Theorem 4 point at a certain relationship between the constant of asymptotic quasi-optimality and the fraction of degree 2 variable nodes, at least for c.a. sequences generated by (17). For the right-regular sequence, we have

$$\lim_{n \rightarrow \infty} \Lambda_2^n = \frac{1}{2},$$

which should lead to fast decoding algorithms if $R < 1/2$. For the Tornado sequence we get the same limit value by directly using (26). In order to get to smaller limit values, suitable also for $R \geq 1/2$, one needs to explore other $\phi \in \mathcal{A}$. By using the family (8), we get the following

Theorem 5. *For any $0 < R < 1$, there exists a c.a. sequence (λ^n, ρ^n) of rate R which is asymptotically quasi-optimal and has a limit*

$$\lim_{n \rightarrow \infty} \Lambda_2^n < 1 - R.$$

Proof. Set $\phi(x) = (1 - b)x/(1 - bx)$ for some $0 < b < 1$. Then $\phi'(1) = 1/(1 - b)$ and

$$\int_0^1 x^{-1}\phi(x) dx = \frac{1 - b}{b} |\log(1 - b)|.$$

Thus, according to Theorem 4, for the sequence (λ^n, ρ^n) generated by this ϕ we have

$$\Lambda^\phi = \frac{b}{2|\log(1 - b)|} \rightarrow 0, \quad b \rightarrow 1.$$

Consequently, by choosing a b close to 1, we can make this limit smaller than $1 - R$. Theorem 1 guarantees the asymptotical quasi-optimality for any b , however, with a constant that quickly grows as $b \rightarrow 1$:

$$\mu^\phi = \frac{b}{(1 - b)|\log(1 - b)|} \rightarrow \infty, \quad b \rightarrow 1.$$

This proves Theorem 5. □

Other choices might be tried. For example, if we pick ϕ from the family (9) we compute

$$\Lambda^\phi = \frac{1}{(1+c)(2-c)}, \quad \mu^\phi = \frac{2}{2-c}, \quad 0 < c < 1.$$

Taking $\phi(x) = (x + x^2)/2$ leads to the smallest Λ_2^ϕ for the family (9). We have

$$\Lambda^\phi = \frac{4}{9} = 0.4444\dots, \quad \mu^\phi = \frac{4}{3} = 1.3333\dots,$$

which is, however, slightly worse than the values obtained from the family (8) where we obtain for $b = 0.2136702$

$$\Lambda^\phi = 0.4444\dots, \quad \mu^\phi = 1.1304\dots.$$

Using composition of the previous ϕ with itself gives $\phi(x) = (2x + 3x^2 + 2x^3 + x^4)/8$, for which

$$\Lambda^\phi = \frac{64}{159} = 0.4025\dots, \quad \mu^\phi = \frac{96}{53} = 1.8113\dots,$$

and so on.

That composition seems to decrease Λ_2^ϕ is not by accident as is shown by the following

Proposition 2. *Suppose that ϕ_1 and ϕ_2 are in \mathcal{A} . Then*

$$\Lambda_2^{\phi_1 \circ \phi_2} \leq \Lambda_2^{\phi_1}. \tag{27}$$

Proof. To see this, use $(\phi_1 \circ \phi_2)'(1) = \phi_1'(1)\phi_2'(1)$ and

$$\int_0^1 \frac{\phi_1(\phi_2(x))}{x} dx = \int_0^1 \frac{\phi_1(t)}{t} \psi(t) dt, \quad t = \phi_2(x), \quad \psi(t) := \frac{\phi_2(x)}{x\phi_2'(x)}.$$

The inequality (27) follows now from the observation that $\psi(t)$ is monotonously decreasing on $(0, 1]$ for any $\phi_2 \in \mathcal{A}$, which gives $\psi(t) \geq \psi(1) = \phi_2'(1)^{-1}$ for $t \in (0, 1]$. Note that the order in the composition matters since, in general, $\Lambda_2^{\phi_1 \circ \phi_2} \leq \Lambda_2^{\phi_2}$ does not hold. □

5 Numerical examples

The first three tables illustrate the quality of the asymptotic estimates in Theorems 1-3. They all have been obtained following the algorithm of Section 2, except that the Taylor coefficients of $g = Tf$ in Step (2) of Algorithm 1 have been computed by explicit formulas. These are available

$R = 1/3$							
a_r	M	a	δ	ϵ	μ	Δ	Λ_2
4	2	3.594	0.41106	0.38341	3.438	7.82	0.8776
5	5	4.801	0.58236	0.12646	2.125	9.19	0.5912
6	13	5.903	0.62936	0.05596	1.905	13.03	0.5369
7	28	6.953	0.64823	0.02765	1.837	19.82	0.5172
8	57	7.978	0.65717	0.01425	1.809	30.96	0.5085
9	114	8.990	0.66170	0.00746	1.794	48.81	0.5042
10	226	9.995	0.66405	0.00392	1.784	77.10	0.5022
11	443	10.998	0.66529	0.00206	1.776	121.55	0.5012
12	864	11.999	0.66595	0.00108	1.769	191.08	0.5006
13	1682	13.000	0.66629	0.00056	1.762	299.34	0.5003
$R = 1/2$							
a_r	M	a	δ	ϵ	μ	Δ	Λ_2
4	1	3.594	0.27063	0.45875	3.558	5.25	0.9999
5	4	4.801	0.39509	0.20981	2.219	5.58	0.6536
6	8	5.903	0.44546	0.10908	1.877	6.30	0.5689
7	15	6.953	0.46951	0.06098	1.735	7.37	0.5356
8	27	7.978	0.48236	0.03529	1.658	8.75	0.5195
9	46	8.990	0.48959	0.02081	1.611	10.47	0.5111
10	78	9.995	0.49380	0.01240	1.579	12.57	0.5065
11	132	10.998	0.49628	0.00744	1.556	15.15	0.5038
12	221	11.999	0.49776	0.00448	1.538	18.29	0.5023
13	367	13.000	0.49865	0.00270	1.524	22.11	0.5014
14	609	14.000	0.49918	0.00163	1.512	26.75	0.5008
15	1008	15.000	0.49951	0.00099	1.503	32.40	0.5005
16	1666	16.000	0.49970	0.00060	1.495	39.23	0.5003
17	2751	17.000	0.49982	0.00036	1.488	47.56	0.5002
18	4541	18.000	0.49989	0.00022	1.481	57.65	0.5001
19	7492	19.000	0.49993	0.00013	1.476	69.92	0.5001
20	13234	20.000	0.49996	0.00008	1.471	84.79	0.5000
$R = 2/3$							
a_r	M	a	δ	ϵ	μ	Δ	Λ_2
6	2	5.903	0.16894	0.49317	3.442	3.88	1.0000
7	3	6.953	0.23464	0.29608	2.332	3.90	0.7144
8	5	7.978	0.27228	0.18315	1.911	3.90	0.6136
9	8	8.990	0.29372	0.11883	1.713	4.00	0.5708
10	12	9.995	0.30690	0.07930	1.600	4.16	0.5433
11	18	10.998	0.31537	0.05389	1.527	4.36	0.5286
12	26	11.999	0.32097	0.03709	1.477	4.58	0.5193
13	38	13.000	0.32474	0.02578	1.441	4.85	0.5132
14	55	14.000	0.32732	0.01805	1.414	5.14	0.5092
15	78	15.000	0.32910	0.01270	1.393	5.47	0.5064
16	111	16.000	0.33034	0.00898	1.377	5.83	0.5045
17	157	17.000	0.33121	0.00637	1.363	6.22	0.5032
18	220	18.000	0.33182	0.00453	1.352	6.65	0.5023
19	310	19.000	0.33226	0.00322	1.343	7.11	0.5016
20	434	20.000	0.33257	0.00230	1.335	7.62	0.5011

Table 1: Tests for the Tornado sequence, generated by (7).

$R = 1/3$						
a_r	M	δ	ϵ	μ	Δ	Λ_2
3	2	0.500000	0.250000	2.377	5.10	1.0000
4	9	0.641100	0.038350	1.348	2.79	0.6933
5	33	0.659958	0.010063	1.194	2.34	0.6314
6	110	0.664649	0.003026	1.136	2.17	0.6018
7	348	0.666029	0.000957	1.106	2.08	0.5839
8	1077	0.666460	0.000309	1.088	2.03	0.5716
9	3296	0.666599	0.000101	1.075	1.99	0.5626
10	10026	0.666645	0.000033	1.065	1.96	0.5556
$R = 1/2$						
a_r	M	δ	ϵ	μ	Δ	Λ_2
4	2	0.333333	0.333333	2.524	3.81	1.0000
5	5	0.449009	0.101981	1.518	2.71	0.6960
6	12	0.479776	0.040448	1.297	2.34	0.6253
7	28	0.491035	0.017929	1.207	2.17	0.5940
8	60	0.495820	0.008360	1.159	2.07	0.5762
9	125	0.497997	0.004006	1.130	2.01	0.5648
10	256	0.499025	0.001949	1.111	1.98	0.5566
11	522	0.499521	0.000957	1.097	1.95	0.5505
12	1058	0.499764	0.000473	1.086	1.93	0.5457
13	2135	0.499883	0.000234	1.078	1.92	0.5418
14	4300	0.499942	0.000116	1.071	1.90	0.5395
15	8649	0.499971	0.000058	1.065	1.89	0.5357
$R = 2/3$						
a_r	M	δ	ϵ	$\tilde{\mu}$	Δ	Λ_2
6	2	0.200000	0.400000	2.655	3.15	1.0000
7	3	0.265741	0.202778	1.779	2.67	0.7317
8	6	0.296314	0.111057	1.476	2.36	0.6441
9	10	0.311497	0.065510	1.339	2.21	0.6019
10	16	0.319948	0.040156	1.261	2.11	0.5788
11	26	0.324917	0.025250	1.212	2.04	0.5642
12	41	0.327955	0.016136	1.179	1.99	0.5544
13	64	0.329857	0.010429	1.155	1.96	0.5474
14	98	0.331068	0.006795	1.137	1.94	0.5421
15	150	0.331849	0.004453	1.123	1.92	0.5381
16	227	0.332356	0.002931	1.112	1.90	0.5349
17	344	0.332688	0.001935	1.103	1.89	0.5323
18	521	0.332907	0.001280	1.096	1.88	0.5301
19	786	0.333051	0.000848	1.089	1.87	0.5282
20	1184	0.333146	0.000563	1.084	1.87	0.5266

Table 2: Tests for the right-regular sequence, generated by (6)

a_r	c	n	M	δ	ϵ	μ	Δ	Λ_2
6	0.375	4	9	0.466270	0.067459	1.542	3.89	0.5849
7	0.314	5	19	0.484497	0.031007	1.397	3.75	0.5497
8	0.271	6	39	0.492661	0.014677	1.314	3.64	0.5324
9	0.481	6	56	0.494241	0.011518	1.398	5.79	0.5121
10	0.429	7	107	0.497063	0.005873	1.349	5.96	0.5030
11	0.386	8	205	0.498509	0.002981	1.311	6.07	0.4974
12	0.352	9	397	0.499245	0.001511	1.281	6.20	0.4939
13	0.487	9	503	0.499354	0.001292	1.355	10.57	0.4863
14	0.450	10	952	0.499666	0.000668	1.327	10.94	0.4831
15	0.418	11	1821	0.499828	0.000343	1.303	11.25	0.4809
16	0.391	12	3506	0.499912	0.000176	1.282	11.50	0.4795
17	0.490	12	4192	0.499922	0.000156	1.344	20.43	0.4754
18	0.462	13	7990	0.499960	0.000080	1.323	21.08	0.4737

Table 3: Tests for the sequence generated from $\{cx^2n + (1 - c)x^n\}$ and rate $R = 1/2$.

since we have simple analytic expressions for the inverse functions f^{-1} and therefore for $g = \mathcal{T}f$ in all three cases considered. Table 1 and 2 concern the Tornado sequence from Theorem 2 and the right-regular sequence from Theorem 3. To make the tables easier to compare, we have chosen the values a in the family (7) such that the a_r -values are integer, as is the case for the family (6). Note that the columns for $\mu = \mu(\lambda, \rho)$ and $\Delta = \Delta(\lambda, \rho)$ demonstrate the tightness of the bounds established in Section 3. Recall that according to Theorem 2, the limit value for μ should not exceed $\mu_{\text{Torn}}(1/3) = 1.648\dots$, $\mu_{\text{Torn}}(1/2) = 1.386\dots$, $\mu_{\text{Torn}}(2/3) = 1.216\dots$ for $R=1/3, 1/2, 2/3$, respectively. Slight differences with the corresponding tables from [4], [5] stem from the fact that our algorithm for actually producing polynomial pairs (λ, ρ) is slightly different. Also note that the faster convergence of the right-regular sequence comes at a price: the polynomials λ for the right-regular sequence have significantly higher maximal degree M than their counterparts from Table 2.

To obtain the values in Table 3, we have experimented with the family (9) to choose suitable ϕ for use in conjunction with (17) and small n . Consequently, we had two parameters, c and n , which were used to achieve integer a_r , for better comparison with the previous tables. We also tried to keep c close to $1/2$ to demonstrate the behavior of the fraction of degree 2 variable nodes Λ_2 which for $c = 1/2$ would be expected to approach the value $4/9 = 0.4444\dots$ for $n \rightarrow \infty$.

The next two tables illustrate the result of Theorem 5. According to its proof, by taking the sequence (17) with ϕ from the family (8) we should expect smaller Λ^ϕ -values if $b \in (0, 1)$ is increased. We

show calculations for $R = 1/2, 2/3$ and $b = 0.2, 0.5, 0.8$, restricting our attention to small values of a_r and n . Note that for $b = 0$ we obtain the sequence (6) for which values are reported in Table 2. The theoretical findings of Theorem 5 are backed to full extent: the Λ^ϕ - as well as the μ^ϕ -values remain R -independent. As to the dependence on b , we see that with increasing b , the Λ_2 -values become smaller, at the expense of worse capacities, compare the large Δ -values. Altogether, Tables 3-5 indicate that, despite the promising theoretical results, very small Λ_2 -values can only be obtained at the expense of comparably bad ϵ -values resp. higher maximal polynomial degree N and M .

6 Conclusion

In this paper we have started a systematic study of capacity achieving sequences for the erasure channel. We have introduced a class \mathcal{A} of analytic functions which give rise to valid degree distributions in an algorithmic way. We introduced an operator \mathcal{T} on this class and showed various properties of both \mathcal{A} and \mathcal{T} . A thorough and systematic study of these objects is likely to shed new light on a partial classification of capacity achieving sequences.

We further introduced a method for constructing infinitely many capacity achieving sequences. To date, the only known sequences were the Tornado and the right-regular sequence. We introduced two measures of optimality for capacity achieving sequences and showed that the right regular sequence is capacity achieving in a much stronger sense than the Tornado sequence. This result may explain why numerical optimization techniques [7] tend to produce sequences that are close to being right regular. Given that this also holds for other types of channels [3], we conjecture that right regular sequences are fundamentally superior on symmetric channels.

We also introduced a technique to reduce the fraction of degree 2 nodes in a c.a. sequence. It is well-known that decoding algorithms based on message passing tend to correct low-degree variable nodes last. In particular, variable nodes of degree 2 are corrected at the very end, and are responsible for slow convergence of the iteration. By having a small fraction of degree 2 nodes, one can push them into redundant nodes and hence accelerate the decoding in practical settings.

$b = 0.2$							
n	a_r	M	δ	ϵ	μ	Δ	Λ_2
3	4.4122	2	0.3705	0.2590	2.2638	5.514	0.7939
4	5.5344	7	0.4574	0.0853	1.5583	3.953	0.6050
5	6.6559	16	0.4824	0.0351	1.3774	3.540	0.5518
6	7.7771	35	0.4921	0.0158	1.2988	3.456	0.5268
7	8.8980	75	0.4963	0.0074	1.2569	3.528	0.5122
8	10.0188	157	0.4982	0.0036	1.2315	3.689	0.5027
$b = 0.5$							
n	a_r	M	δ	ϵ	μ	Δ	Λ_2
2	4.0576	2	0.2611	0.4778	3.8076	7.956	0.9712
3	5.5116	5	0.4198	0.1604	2.0877	7.319	0.5471
4	6.9605	12	0.4690	0.0620	1.7351	7.722	0.4638
5	8.4072	27	0.4866	0.0267	1.6085	9.066	0.4319
6	9.8527	58	0.4939	0.0122	1.5511	11.319	0.4156
$b = 0.8$							
n	a_r	M	δ	ϵ	μ	Δ	Λ_2
2	6.5143	6	0.3928	0.2143	2.9315	19.590	0.4146
3	9.0323	14	0.4564	0.0873	2.5673	45.700	0.3299
4	11.5360	32	0.4812	0.0376	2.4376	111.700	0.2997

Table 4: Illustrations for Theorem 5: $R = 1/2$.

$b = 0.2$							
n	a_r	M	δ	ϵ	μ	Δ	Λ_2
5	6.6559	2	0.2272	0.3185	2.3590	4.734	0.7814
6	7.7771	4	0.2785	0.1646	1.7479	3.855	0.6206
7	8.8980	8	0.3019	0.0944	1.5288	3.483	0.5615
8	10.0188	13	0.3144	0.0567	1.4157	3.296	0.5311
9	11.1395	21	0.3216	0.0351	1.3490	3.217	0.5131
10	12.2602	33	0.3259	0.0222	1.3061	3.206	0.5016
11	13.3808	52	0.3286	0.0143	1.2767	3.241	0.4936
12	14.5013	80	0.3302	0.0093	1.2555	3.309	0.4879
$b = 0.5$							
n	a_r	M	δ	ϵ	μ	Δ	Λ_2
4	6.9605	3	0.2122	0.3634	2.7877	6.110	0.6833
5	8.4072	5	0.2657	0.2028	2.1362	6.130	0.5273
6	9.8527	8	0.2943	0.1170	1.8620	6.356	0.4649
7	11.2974	14	0.3100	0.0699	1.7216	6.821	0.4338
8	12.7417	22	0.3190	0.0431	1.6432	7.555	0.4161
9	14.1856	35	0.3243	0.0271	1.5947	8.541	0.4050
10	15.6293	55	0.3276	0.0173	1.5631	9.805	0.3976
$b = 0.8$							
n	a_r	M	δ	ϵ	μ	Δ	Λ_2
2	6.5143	2	0.1310	0.6069	5.2890	8.516	0.8286
3	9.0323	5	0.2267	0.3200	3.2145	12.466	0.4428
4	11.5360	9	0.2708	0.1875	2.7943	20.156	0.3550
5	14.0330	14	0.2956	0.1133	2.6126	33.517	0.3165
6	16.5265	23	0.3101	0.0696	2.5150	56.628	0.2961

Table 5: Illustrations for Theorem 5: $R = 2/3$.

7 Open problems

Below is a list of open problems for the interested reader. They are divided into two classes. The problems in the first class are probably easy to solve.

1. Prove equality in Theorems 1 and 2.
2. Prove (or disprove) that composition with $\phi \in \mathcal{A}$ preserves the convergence type of a c.a. sequence, i.e., c.a. sequences are transformed into c.a. sequences, and asymptotical quasi-optimality is preserved (with a different constant). From Theorem 1 and Theorem 3 it follows that asymptotical optimality is generally not preserved.
3. Show that $\Lambda^\phi \rightarrow 0$ implies $\mu^\phi \rightarrow \infty$. More generally, make the statement "Small Λ_2 implies large ϵ " more quantitative.
4. For which $\phi \in \mathcal{A}$ does the sequence $f^n = \phi \circ \dots \circ \phi$ (n times) lead to a c.a. or asymptotically quasi-optimal sequence? For $\phi(x) = x^2$ the answer is yes, for ϕ from (8) the answer is no (for asymptotic quasi-optimality).
5. Write a code that fully explores the design principles of our paper, i.e., implements Algorithm 1 in a stable and run-time efficient way, explores composition principles to full power, adds postprocessing to reduce M, N , includes multi-objective optimization for simultaneously reducing ϵ and Λ_2 , etc.

The problems in the following second class are probably much harder to solve.

1. Give a constructive characterization of \mathcal{A} . Here we feel that there are possibly related results in classical complex function theory which need to be recovered.
2. Give an algorithm to check whether a polynomial ϕ belongs to \mathcal{A} .
3. Prove that the right-regular sequence is in some sense optimal. For example, prove that there is no c.a. sequence (λ^n, ρ^n) of rate R for which $\lim_{n \rightarrow \infty} \Delta(\lambda^n, \rho^n) < e^\gamma$.
4. Prove that the approach via \mathcal{A} is universal, e.g., prove that if (λ, ρ) has rate R and affords δ then there is another pair generated by a $f \in \mathcal{A}$ with the same rate, and approximately the same (or better) δ , and approximately the same maximal and average degrees.

References

- [1] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, “Analysis of low density codes and improved designs using irregular graphs,” in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pp. 249–258, 1998.
- [2] T. Richardson and R. Urbanke, “The capacity of low-density parity check codes under message-passing decoding,” *IEEE Trans. Inform. Theory*, 2000. To appear.
- [3] T. Richardson, A. Shokrollahi, and R. Urbanke, “Design of capacity-approaching irregular low-density parity-check codes,” *IEEE Trans. Inform. Theory*, 2000. To appear.
- [4] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, “Practical loss-resilient codes,” in *Proceedings of the 29th annual ACM Symposium on Theory of Computing*, pp. 150–159, 1997.
- [5] A. Shokrollahi, “New sequences of linear time erasure codes approaching the channel capacity,” in *Proceedings of the 13th International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes* (M. Fossorier, H. Imai, S. Lin, and A. Poli, eds.), Lecture Notes in Computer Science, no. 1719, pp. 65–76, 1999.
- [6] A. Shokrollahi, Capacity-achieving sequences, in *IMA Volumes in Mathematics and its Applications*, vol. 123, pp. 153–166. 2000.
- [7] A. Shokrollahi and R. Storn, “Design of efficient erasure codes with differential evolution,” in *Proceedings of ISIT’00*, p. 5, 2000.
- [8] W. Walter, *Analysis*, vol. 1 und 2. Springer Verlag, Berlin, 1992.

A Computing $\mathcal{T}f$

In this section, we will briefly explain how to compute the Taylor expansion of $\mathcal{T}f$ for a given $f \in \mathcal{A}$ which is analytic at 1.

Since we assume that $f \in \mathcal{A}$, we know that

$$g(x) = \mathcal{T}f(x) = \sum_{k=1}^{\infty} g_k x^k$$

has an absolutely converging Taylor series, where $g_k \geq 0$ and $\sum_{k=1}^{\infty} g_k = 1$. By (2), the g_k can be computed from the power series expansion of f^{-1} at $x = 1$ which, in turn, can be computed from the coefficients \tilde{f}_k in

$$f(1+x) = 1 + \sum_{k \geq 1} \tilde{f}_k x^k, \quad (28)$$

For example, $g_1 = \tilde{f}_1^{-1}$, $g_2 = 2\tilde{f}_1^{-3}\tilde{f}_2$, and so on. The validity of (28) in a neighborhood of $x = 0$ needs to be assumed, as the example $f(x) = 1 - (1-x)^{1/2} \in \mathcal{A}$ shows, it cannot be concluded from $f \in \mathcal{A}$.

Alternatively, we can design an iterative scheme based on a standard transformation of (2) into a fixpoint equation. Indeed, since $g = \mathcal{T}f$ is equivalent to $f(1-g(x)) = 1-x$, we can write

$$g(x) = \frac{1}{\tilde{f}_1} \left(x + \sum_{k=2}^{\infty} (-1)^k \tilde{f}_k g(x)^k \right),$$

where we again have assumed (28). The finite-dimensional analogue of this fixpoint equation is

$$T_M g = \frac{1}{\tilde{f}_1} T_M \tilde{g}, \quad \tilde{g}(x) = x + \sum_{k=2}^{\infty} (-1)^k \tilde{f}_k T_M g(x)^k, \quad (29)$$

which can be shown to give the first M coefficients of g from any initial guess in at most M steps, provided that no rounding occurs.

B Proof of Theorem 3

Using the same notation as before, but dropping the superscript n we have $\rho(x) := \rho^n(x) = f^n(x) = x^n$, $a_r := a_r^n = n+1$, and

$$g(x) := g^n(x) = 1 - (1-x)^{1/n} = \sum_{l=1}^{\infty} g_l x^l,$$

where

$$g_l = \frac{1}{n^l} \beta_l := \frac{1}{n^l} \prod_{r=1}^{l-1} \left(1 - \frac{1}{nr} \right), \quad l \geq 1.$$

Let us set

$$\hat{I}_k = \frac{\sigma_k}{s_k}, \quad s_k = \sum_{l=1}^k g_l, \quad \sigma_k = \sum_{l=1}^k \frac{g_l}{l+1}, \quad k \geq 1.$$

and $\hat{I}_0 = 1/2$, $s_0 = \sigma_0 = 0$. Obviously, \hat{I}_k is a decreasing sequence with limit a_r^{-1} (this follows from the corresponding properties of the function $\hat{I}(t)$ defined in Step (2) of Algorithm 1) while $\{s_k\}$ and $\{\sigma_k\}$ are increasing sequences, with limits $g(1) = 1$, and $\int_0^1 g(x) dx = \int_0^1 x^n dx = a_r^{-1}$, respectively.

Let t and δ be the parameters determined in Step (2) and Step (3) of Algorithm 1, respectively. Since $\hat{I}_k = \hat{I}(k)$, it follows from Step (2) of Algorithm 1 that the integer $k = \lfloor t \rfloor$ must satisfy

$$\hat{I}_k \geq \frac{1}{(n+1)(1-R)} \geq \hat{I}_{k+1}. \quad (30)$$

After we have estimated k from this relation, we can further estimate δ and $\epsilon(\lambda, \rho)$ since Step (3) of Algorithm 1 implies

$$s_k \leq \delta \leq s_{k+1}. \quad (31)$$

To implement this strategy, we need to access the Taylor coefficients g_l of g , and obtain sharp estimates for σ_k and s_k . The latter task is usually performed using

$$s_k = 1 - \hat{s}_k, \quad \sigma_k = a_r^{-1} - \hat{\sigma}_k, \quad \hat{s}_k = \sum_{l=k+1}^{\infty} g_l, \quad \hat{\sigma}_k = \sum_{l=k+1}^{\infty} \frac{g_l}{l+1}.$$

For notational convenience, let θ_l denote a variable which may vary from formula to formula but everywhere satisfies the bound $|\theta_l| \leq c_0 l^{-1}$ with some absolute constant $0 < c_0 < \infty$. Analogously, we will use the notation θ_k and θ_n . Note that these quantities also may have different signs in different places. By definition of the Euler constant γ , we have

$$\sum_{r=1}^{l-1} \frac{1}{r} = \ln l + \gamma + O\left(\frac{1}{l}\right),$$

see [8, vol. 1, p. 349], which leads to

$$-\ln \beta_l = \frac{\ln l + \gamma + \theta_l + \theta_n}{n},$$

where the θ_n -term can be chosen independently of l and comes from the summation of the higher order correction terms in $\ln(1 - 1/(nr)) = 1/(nr) + O((nr)^{-2})$. Thus,

$$g_l = \frac{(1 + \theta_l/n)e^{-\gamma_n/n}}{nl^{1+1/n}}, \quad l \geq 1.$$

For convenience, we have set $\gamma_n = \gamma + \theta_n$. From this, sufficiently precise asymptotic formula for \hat{s}_k and $\hat{\sigma}_k$ can be obtained. For example,

$$\hat{s}_k = (1 + \theta_k/n)e^{-\gamma_n/n} \left(\frac{1}{n} \sum_{l=k+1}^{\infty} l^{-(1+1/n)} \right) = (1 + \theta_k/n)e^{-\gamma_n/n} k^{-1/n}.$$

Analogously,

$$\hat{\sigma}_k = (1 + \theta_k/n)e^{-\gamma_n/n} \left(\frac{1}{n} \sum_{l=k+1}^{\infty} \frac{l^{-(1+1/n)}}{(l+1)} \right) = \frac{(1 + \theta_k)e^{-\gamma_n/n}}{n+1} k^{-(1+1/n)}.$$

Consequently,

$$\hat{s}_k - (n+1)\hat{\sigma}_k = (1+\theta_k)e^{-\gamma n/n}k^{-1/n} = (1+\theta_k)\hat{s}_k,$$

and

$$(\hat{s}_k)^{n+1} = (1+\theta_k/n)^{n+1}e^{-(1+1/n)\gamma n}k^{-(1+1/n)} = (1+\theta_k)e^{-\gamma n}(n+1)\hat{\sigma}_k. \quad (32)$$

Let us first mention that for $n \rightarrow \infty$, the index k corresponding to the solution of (30) grows exponentially with n . Indeed, from the lower bound in (31), (30), and the above estimates we obtain

$$\hat{s}_k \geq 1 - \delta > R = 1 - (1 - R) \geq 1 - \frac{s_{k+1}}{(n+1)\sigma_{k+1}} = \frac{\hat{s}_{k+1} - (n+1)\hat{\sigma}_{k+1}}{1 - (n+1)\hat{\sigma}_{k+1}} = (1+\theta_k)\hat{s}_k$$

and, consequently, $\hat{s}_k = (1+\theta_k)R$ and

$$k > e^{-\gamma n} \left(\frac{1+\theta_k}{R} \right)^n.$$

Since k cannot stay bounded as $n \rightarrow \infty$, we definitely have $(1+\theta_k)/R > q > 1$ which implies exponential growth $k \geq cq^n$. In particular, $k \geq n^2$ for large enough n , and the expressions θ_k are majorized by θ_n .

Now we are able to conclude the argument for Theorem 3. From (30) and the above mentioned properties of $\hat{s}_k, \hat{\sigma}_k$, we have

$$1 - \frac{s_k}{1-R} = 1 - \frac{s_{k+1}}{1-R} + \frac{g_{k+1}}{1-R} \leq 1 - (n+1)\sigma_{k+1} + \frac{g_{k+1}}{1-R} = (n+1)\hat{\sigma}_{k+1} + \frac{g_{k+1}}{1-R}.$$

According to the asymptotic behavior of g_k and $\hat{\sigma}_k$, we have $g_{k+1}/(1-R) = \theta_n(n+1)\hat{\sigma}_{k+1}$, and we can continue by using (32))

$$1 - \frac{s_k}{1-R} \leq (1+\theta_n)(n+1)\hat{\sigma}_{k+1} = (1+\theta_n)e^{\gamma n}(\hat{s}_{k+1})^{n+1} = (1+\theta_n)(1+\theta_k)^n e^{\gamma n} R^{n+1}.$$

By definition of the quantity $\Delta(\lambda, \rho)$ and (31) we thus obtain

$$\Delta(\lambda, \rho) = \frac{1 - \delta/(1-R)}{R^{n+1}} \leq \frac{1 - s_k/(1-R)}{R^{n+1}} \leq (1+\theta_n)(1+\theta_k)^n e^{\gamma n} \rightarrow e^\gamma,$$

where we have used that $(1+\theta_k)^n \rightarrow 1$ since $k \geq n^2$ for large enough n . This gives the upper bound for Δ . The lower bound follows in exactly the same way, by starting from

$$\Delta(\lambda, \rho) \geq \frac{1 - s_{k+1}/(1-R)}{R^{n+1}},$$

and estimating $1 - s_{k+1}/(1-R) = 1 - s_k/(1-R) - g_{k+1}/(1-R) \geq (n+1)\hat{\sigma}_n - g_{k+1}/(1-R)$ from below. This finally proves the statement of Theorem 3. \square

C Composition principles

Let (λ, ρ) be a pair of degree distributions of rate R , and $0 < \delta < 1 - R$ afforded by it, i.e., (5) is satisfied. Take any $\phi \in \mathcal{A}$ and denote $\psi = \mathcal{T}\phi$. What we claim is that the new pair $(\hat{\lambda}, \hat{\rho})$ defined by

$$\hat{\rho} = \phi \circ \rho, \quad \hat{\lambda} = \lambda \circ \psi,$$

satisfies again (5), with the same δ . Indeed, by Lemma 1 we have

$$\delta \hat{\lambda}(x) = \delta \lambda(\psi(x)) \leq T\rho \circ \mathcal{T}\phi(x) = T(\phi \circ \rho)(x) = T\hat{\rho}(x).$$

This gives the claim. Unfortunately, without further assumptions we cannot control the rate, i.e., we are not sure whether the rate goes up or down. Recall that rate R implies

$$(1 - R) \int_0^1 \lambda(x) dx = \int_0^1 \rho(x) dx = \int_0^1 \mathcal{T}\rho(x) dx,$$

while the rate of the new pair is defined by the equation

$$(1 - \hat{R}) = \frac{\int_0^1 \mathcal{T}\hat{\rho}(x) dx}{\int_0^1 \hat{\lambda}(x) dx} = \frac{\int_0^1 \mathcal{T}\rho(\psi(x)) dx}{\int_0^1 \lambda(\psi(x)) dx} = \frac{\int_0^1 \phi'(1 - y) \mathcal{T}\rho(y) dy}{\int_0^1 \phi'(1 - y) \lambda(y) dy},$$

where we have used the transformation $y = \psi(x)$ and $\psi'(x)^{-1} = (\psi^{-1})'(y) = \phi'(1 - y)$.

It would be desirable to fix this problem, e.g., under the assumption that (λ, ρ) was produced from a (polynomial) $f \in \mathcal{A}$ by our algorithm:

$$f \in \mathcal{A} \rightarrow (f, \mathcal{T}f) \rightarrow \begin{cases} (\lambda_f, \rho_f, \delta_f) \\ (\tilde{\lambda}_f, \tilde{\rho}_f, \tilde{\delta}_f) \end{cases}.$$

The tilde-notation refers to the theoretical construction used in the proof of Theorem 1 which is easier accessible for further estimation. One could try to compare the pair $(\hat{\lambda}, \hat{\rho}) = (\lambda_f \circ \psi, \phi \circ \rho_f)$ to either $(\lambda_{\phi \circ f}, \rho_{\phi \circ f})$ or $(\tilde{\lambda}_{\phi \circ f}, \tilde{\rho}_{\phi \circ f})$.

The final statement which we anticipate to be able to prove is that the mapping

$$(\lambda, \rho) \rightarrow (\lambda \circ \psi, \phi \circ \rho), \quad \psi = \mathcal{T}\phi, \quad \phi \in \mathcal{A},$$

transforms c.a sequences into c.a sequences and also preserves asymptotic quasi-optimality. In light of our theorems, there is no hope to preserve the constant of asymptotic quasi-optimality and, in particular, the asymptotic optimality of a c.a. sequence.

D Criteria for $f \in \mathcal{A}$

To decide on whether a given $f \in \mathcal{P}$ belongs to \mathcal{A} heavily depends on verifying

$$g_k \geq 0, \quad k \geq 3, \quad (33)$$

where g_k are the Taylor series coefficients of $g = \mathcal{T}f$ (obviously, g_1 and g_2 are non-negative for any $f \in \mathcal{P}$). Since g is the inverse function of $\psi(x) = 1 - f(1 - x)$, the g_k can be expressed by finite formula containing the derivatives of f at $x = 1$ (below we will always assume that f is analytic in a neighborhood of this point). We will derive a recursion which could be helpful to check (33) in particular cases. Let $y = \psi(x)$ and, thus, $x = g(y)$, $x, y \in [0, 1]$. Then

$$g'(y) = (\psi^{-1})'(y) = \frac{1}{\psi'(x)} = \frac{1}{f'(1-x)} (= \frac{dx}{dy}).$$

Thus, $g'(y)$ has an explicit expression as a function of x , by induction this holds for higher derivatives, too:

$$g^{(n)}(y) = \frac{dg^{(n-1)}(y)}{dx} \frac{dx}{dy} = \frac{dg^{(n-1)}(y)}{dx} \frac{1}{\psi'(x)}$$

Since $g_n = (n!)^{-1}g^{(n)}(0)$, and $y = 0$ corresponds to $x = 0$, we are in the game. For future use, define

$$s(x) = \frac{-\psi''(x)}{\psi'(x)^2} =, \quad s_n(x) = \frac{\psi'(x)^{n-2}}{\psi''(x)^{n-1}} \psi^{(n)}(x), \quad n \geq 3, \quad (34)$$

and

$$s = s(1) = \frac{f''(1)}{f'(1)^2}, \quad s_n = s_n(1) = \frac{f'(1)^{n-2}}{f''(1)^{n-1}} f^{(n)}(1), \quad n \geq 3. \quad (35)$$

Obviously,

$$g''(y) = -\frac{\psi''(x)}{\psi'(x)^3} = \frac{s(x)}{\psi'(x)}, \quad g'''(y) = -\frac{\psi'''(x)}{\psi'(x)^4} + 3\frac{\psi''(x)^2}{\psi'(x)^5} = 3\frac{s(x)^2}{\psi'(x)} \left(1 - \frac{1}{3}s_3(x)\right),$$

which lets us guess the following formula:

$$g^{(n)}(y) = (2n-3)!! \frac{s(x)^{n-1}}{\psi'(x)} p_n(s_3(x), \dots, s_n(x)), \quad n \geq 3. \quad (36)$$

where p_n is a polynomial satisfying a recursion to be derived next. Indeed, (36) holds for $n = 3$ with $p_3(t) = 1 - t/3$. Let us compute $g^{(n+1)}(x)$, assuming that (36) holds for some $n \geq 3$. We have (for simplicity, we drop the arguments of all functions involved)

$$g^{(n+1)} = \frac{(2n-3)!!}{\psi'} \left[p_n(s^n + (n-1)\frac{s^{n-2}s'}{\psi'} + \frac{s^{n-1}}{\psi'} \sum_{k=3}^n \frac{\partial p_n}{\partial s_k} s'_k \right].$$

But

$$s' = 2(\psi'')^2 - \psi' \psi''' (\psi')^{-3} = \psi' s^2 (2 - s_3) ,$$

and

$$\begin{aligned} s'_k &= \frac{(\psi')^{k-2}}{(\psi'')^{k-1}} \psi^{(k+1)} + ((k-2) \frac{(\psi')^{k-3}}{(\psi'')^{k-2}} - (k-1) \frac{(\psi')^{k-2} \psi'''}{(\psi'')^k}) \psi^{(k)} \\ &= \psi' s (((k-1)s_3 - (k-2))s_k - s_{k+1}) . \end{aligned}$$

Substitution gives

$$g^{(n+1)} = (2n-3)!! \frac{s^n}{\psi'} \left[(1 + (n-1)(2-s_3))p_n + \sum_{k=3}^n (((k-1)s_3 - (k-2))s_k - s_{k+1}) \frac{\partial p_n}{\partial s_k} \right]$$

and validates (36) for n replaced by $n+1$, where

$$\begin{aligned} p_{n+1}(s_3, \dots, s_{n+1}) &= (1 - \frac{n-1}{2n-1} s_3) p_n(s_3, \dots, s_n) + \\ &+ \sum_{k=3}^n \frac{((k-1)s_3 - (k-2))s_k - s_{k+1}}{2n-1} \frac{\partial p_n}{\partial s_k}(s_3, \dots, s_n) . \end{aligned} \tag{37}$$

Thus, (33) holds if and only if

$$p_n(s_3, \dots, s_n) \geq 0 \quad \forall n \geq 3 , \tag{38}$$

is satisfied for the sequence $\{s_k, k \geq 3\}$ defined by (35). This does not sound particularly helpful, as we replaced one sequence of inequalities by another one. However, one may try to further analyze (38) and reduce it to more explicit criteria. E.g., if $f \in \mathcal{P}$ is a cubic polynomial then only s_3 is different from 0, and (38) degenerates into a sequence of univariate polynomial inequalities. Setting $s_k = 0$ for all $k \geq 4$ in (37), we arrive at a simplified recursion for $p_n = p_n(s_3)$:

$$p_{n+1}(s_3) = (1 - \frac{n-1}{2n-1} s_3) p_n(s_3) + \frac{(2s_3-1)s_3}{2n-1} p'_n(s_3)$$

This formula implies

$$p_n(s_3) > 0 , \quad s_3 \in [0, 1/2] . \tag{39}$$

Let us sketch the argument for (39). For convenience, let

$$s_3 = \frac{1}{2} - t , \quad q_n(t) = (2n-3)!! p_n(s_3) .$$

This gives

$$\begin{aligned}
q_{n+1}(t) &= ((3n-1)/2 + (n-1)t)q_n(t) + t(1-2t)q'_n(t), \\
q'_{n+1}(t) &= (n-1)q_n(t) + ((3n+1)/2 + (n-5)t)q'_n(t) + t(1-2t)q''_n(t), \\
&\dots \\
q_{n+1}^{(r)}(t) &= r(n-2r+1)q_n^{(r-1)}(t) \\
&\quad + ((3n+2r-1)/2 + (n-1-4r)t)q_n^{(r)}(t) + t(1-2t)q_n^{(r+1)}(t), \\
&\dots
\end{aligned}$$

for all $n \geq 3$ and $r \geq 0$, where we have set $q_n^{(-1)}(t) \equiv 0$. By induction, one checks that the degree of $q_{n+1}(t)$ equals $[n/2]$ (we leave this as an exercise), thus, only the range $0 \leq r \leq [n/2]$ needs to be considered. Set $t = 0$ in the above formulas:

$$q_{n+1}^{(r)}(0) = r(n-2r+1)q_n^{(r-1)}(0) + \frac{3n+2r-1}{2}q_n^{(r)}(0), \quad r = 0, \dots, [n/2],$$

and $q_{n+1}^{(r)}(0) = 0$ for $r > [n/2]$. Both terms in the right-hand side of this recursion have positive coefficients for all r of interest. Since $q_3(t) = 3 - (1/2 - t) = 5/2 + t$ satisfies $q_3^{(r)}(0) \geq 0$ for all $r \geq 0$, by induction we obtain

$$q_n^{(r)}(0) \geq 0, \quad r \geq 0 \implies q_n(t) \geq 0, \quad t \geq 0,$$

for all $n \geq 3$ (strict positivity follows from the observation that $q_n(0) > 0$ for all $n \geq 3$). This proves (39).

Thus, what we have proved is the following sufficient condition: If a cubic polynomial f belongs to \mathcal{P} and

$$2f'(1)f'''(1) \leq f''(1)^2, \tag{40}$$

then $f \in \mathcal{A}$. Numerical evidence indicates that the condition (40) is also necessary. It seems that

$$E_N = \{s_3 \geq 0 : p_n(s_3) \geq 0, 3 \leq n \leq N\}$$

coincides with the interval $[0, 1/2 + t_N]$, where $t_N < 0$ is the smallest positive zero of the polynomial $q_N(-t)$. Numerically, we found $t_3 = 5/2$, $t_4 = 1$, $t_5 = 0.6193$, $t_{10} = 0.2243$, $t_{20} = 0.1088$, $t_{40} = 0.0582$ and so on. A formal proof of the necessity is open.

At present, we have no general idea on how to obtain efficient characterizations of the condition (38), even in the case of polynomials of degree $m = 4$. Nevertheless, (37) could be used to reject

(or get confident about) a particular f by computing the sequence $\{s_k\}$ and substituting it in the recursion. E.g., Figure 1 depicts the two-dimensional region

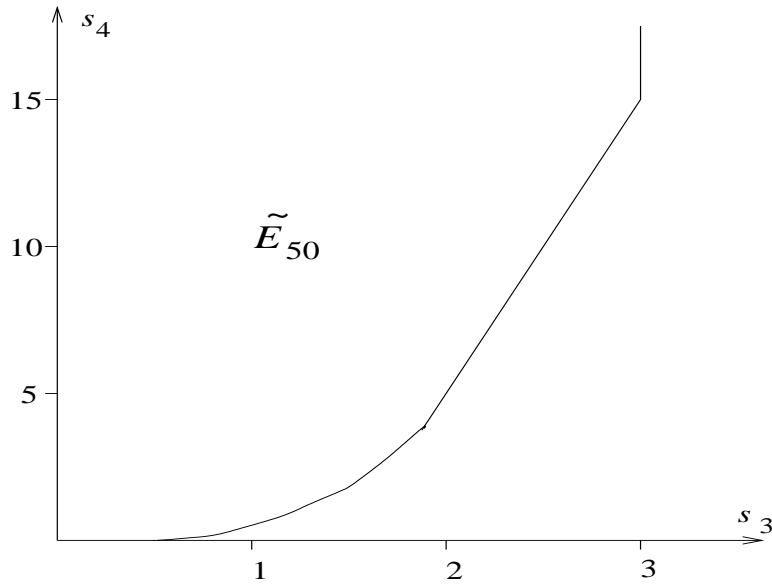


Figure 1: Good approximation to the region of feasible (s_3, s_4) for $m = 4$

$$\tilde{E}_N = \{(s_3, s_4) : p_n(s_3, s_4) \geq 0, 3 \leq n \leq N, s_3, s_4 \geq 0\}$$

for $N = 50$. Based on our experiments, we believe that it is a very accurate approximation to the set of all (s_3, s_4) satisfying (38) for polynomials of degree 4, except for a small neighborhood of $(s_3, s_4) = (1/2, 0)$. Note that $\tilde{E}_3 = [0, 3] \times [0, \infty)$ (thus, $s_3 \leq 3$ is necessary). Since

$$p_4(s_3, s_4) = 1 - \frac{2}{3}s_3 + \frac{1}{15}s_4,$$

we have

$$\tilde{E}_4 = \{(s_3, s_4) : 0 \leq s_3 \leq 3, s_4 \geq \max(0, 10s_3 - 15)\}.$$

Inequalities become more involved for larger N but the tendency is that $\tilde{E}_N \setminus \tilde{E}_{N+1}$ represents an increasingly smaller region near the point $(1/2, 0)$ when $N \rightarrow \infty$.

E Fixed points of \mathcal{T}

In this section we will briefly discuss the fixed points of the operator \mathcal{T} , i.e., functions f such that $\mathcal{T}f = f$. We already have one family of fixed points, see (8). Note that by definition of \mathcal{T} , the graph of a fixed point, restricted to the square $[0, 1]^2$ in the (x, y) -plane, is symmetric about the

line $x + y = 1$. Thus, in the new coordinate system $t = y + x - 1$, $s = y - x$ fixed points $y = f(x)$ of \mathcal{T} are described by the simple formula $s = -h(t)$, $t \in [-1, 1]$, where $h(t)$ satisfies

$$0 \leq h(t) < 1, \quad h(t) = h(-t), \quad t \in [0, 1], \quad h(-1) = h(1) = 0. \quad (41)$$

This leads to the following necessary condition:

Proposition 3. *If $f = \mathcal{T}f$ for some $f \in \mathcal{P}$ then $f(x) = t(x) + 1 - x$, where $t(x)$ satisfies the identity*

$$t(x) = 2x - 1 - h(t(x)), \quad x \in [0, 1],$$

with a function $h(t)$ for which (41) holds.

Choosing appropriate functions h may lead to new family of fixed points. For example, if we set $h(t) = \alpha(1 - t^2)$ in Proposition 3 then we obtain

$$f(x) = (1 + 1/(2\alpha)) - x - \sqrt{(1 + 1/(2\alpha))^2 - 2x/\alpha} \in \mathcal{A}, \quad 0 < \alpha \leq 1/2.$$

In particular, for $\alpha = 1/2$ we obtain another remarkable $f \in \mathcal{A}$:

$$f(x) = 2 - x - 2\sqrt{1-x} = (1 - \sqrt{1-x})^2 = 2 \sum_{k=2}^{\infty} \frac{(2k-3)!!}{(2k)!!} x^k.$$

Obviously, this f is not analytic at $x = 1$. Since $\mathcal{T}f = f$ and the Taylor coefficients of f are explicitly known, we could still apply the strategy of Algorithm 1. Note that this would result in $\Lambda_2 = \lambda_2 = \rho_2 = 0$.