A Master Theorem for Discrete Divide and Conquer Recurrences

MICHAEL DRMOTA and WOJCIECH SZPANKOWSKI
TU Wien and Purdue University

Divide-and-conquer recurrences are one of the most studied equations in computer science. Yet, *discrete* versions of these recurrences, namely

$$T(n) = a_n + \sum_{j=1}^{m} b_j T\left(\lfloor p_j n + \delta_j \rfloor\right) + \sum_{j=1}^{m} \bar{b}_j T\left(\left\lceil p_j n + \overline{\delta}_j \right\rceil\right)$$

for some known sequence a_n and given $b_j, \overline{b}_j$, p_j and $\delta_j, \overline{\delta}_j$, present some challenges. The discrete nature of this recurrence (represented by the floor and ceiling functions) introduces certain oscillations not captured by the traditional Master Theorem, for example due to Akra and Bazzi who primary studied the continuous version of the recurrence. We apply powerful techniques such as Dirichlet series, Mellin-Perron formula, and (extended) Tauberian theorems of Wiener-Ikehara to provide a complete and precise solution to this basic computer science recurrence. We illustrate applicability of our results on several examples including a popular and fast arithmetic coding algorithm due to Boncelet for which we estimate its average redundancy and prove the Central Limit Theorem for the phrase length. To the best of our knowledge, *discrete* divide and conquer recurrences were not studied in this generality and such detail; in particular, this allows us to compare the redundancy of Boncelet's algorithm to the (asymptotically) optimal Tunstall scheme.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Computations on discrete structures*; G.2.1 [**Discrete Mathematics**]: Combinatorics—*Generating functions*; *counting problems*

General Terms: Algorithms

Additional Key Words and Phrases: Divide-and-conquer recurrence, mergesort, Karatsuba algorithm, Strassen algorithm, Boncelet's data compression algorithm, Dirichlet series, Mellin-Perron formula, Tauberian theorem

1. INTRODUCTION

Divide and conquer is a very popular strategy to design algorithms. It splits the input into several smaller subproblems, solving each subproblem separately, and then

The work of these authors was supported in part by the Austrian Science Foundation FWF Grant No. S9604, by NSF Science and Technology Center Grant CCF-0939370, NSF Grants DMS-0800568 and CCF-0830140, AFOSR Grant FA8655-11-1-3076, and NSA Grant H98230-11-1-0141.

Author's address: M. Drmota, Inst. Discrete Mathematics and Geometry, TU Wien, A-1040 Wien, Austria, email: michael.drmota@tuwien.ac.at; W. Szpankowski, Department of Computer Science, Purdue University, West Lafayette, IN 47907, e-mail: spa@cs.purdue.edu.

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 20YY ACM 0000-0000/20YY/0000-0001 $5.00

knitting together to solve the original problem. Typical examples include heapsort, mergesort, discrete Fourier transform, queues, sorting networks, compression algorithms, and so forth [Flajolet and Sedgewick 2008; Cormen et al. 1990; Knuth 1998; Roura 2001; Szpankowski 2001]. While it is relatively easy to determine the general growth order for the algorithm complexity, a precise asymptotic analysis is often appreciably more subtle. Our goal is to present such an analysis for *discrete* divide and conquer recurrences.

The complexity of a divide and conquer algorithm is well described by its divide and conquer recurrence. We assume that the problem is split into m subproblems. It is natural to assume that there is a cost associated with combining subproblems together to find the solution. We denote such a cost by a_n, where n is the size of the original problem. In addition, each subproblem may contribute in a different way to the final solution; we represent this by coefficients b_j and $\overline{b}_j$ for $1 \le j \le m$. Finally, we postulate that the original input n is divided into subproblems of size $\lfloor h_j(n) \rfloor$ and $\lceil \overline{h}_j(n) \rceil$, $1 \le j \le m$, where $h_j(x)$ and $\overline{h}_j(x)$ are functions that satisfy $h_j(x) \sim \overline{h}_j(x) \sim p_j x$ for $x \to \infty$ and for some $0 < p_j < 1$. We aim at presenting precise asymptotic solutions of *discrete* divide and conquer recurrences of the following form [Cormen et al. 1990]

$$T(n) = a_n + \sum_{j=1}^{m} b_j T\left(\lfloor h_j(n) \rfloor\right) + \sum_{j=1}^{m} \overline{b}_j T\left(\lceil \overline{h}_j(n) \rceil\right) \qquad (n \ge 2). \tag{1}$$

A popular approach to solve this recurrence is to relax it to a *continuous* version of the following form (hereafter we assume $\overline{b}_j = 0$ for simplicity)

$$T(x) = a(x) + \sum_{j=1}^{m} b_j T(h_j(x)), \qquad x > 1, \tag{2}$$

where $h_j(x) \sim p_j x$ with $0 < p_j < 1$, and solve it using a Master Theorem as for example in [Cormen et al. 1990; Roura 2001]. This is usually quite powerful and provides the order of growth for $T(x)$. The most general solution of (2) is due to Akra and Bazzi [Akra and Bazzi 1998] who proved (under certain regularity assumptions, namely that $a'(x)$ is of polynomial growth and that $h_j(x) - p_j x = O(x/(\log x)^2)$)

$$T(x) = \Theta\left(x^{s_0}\left(1 + \int_1^x \frac{a(u)}{u^{s_0+1}} du\right)\right),$$

where s_0 is a unique real root of

$$\sum_j b_j p_j^s = 1. \tag{3}$$

Actually this also leads directly to

$$T(n) = \Theta\left(n^{s_0}\left(1 + \sum_{j=1}^{n} \frac{a_j}{j^{s_0+1}}\right)\right)$$

in the discrete version provided that $a_{n+1} - a_n$ is at most of polynomial growth.

For more precise results of the continuous version one can apply Mellin transform techniques [Flajolet et al. 1995; Flajolet and Sedgewick 2008; Szpankowski 2001]. Indeed, let

$$t(s) = \int_0^\infty T(x)x^{s-1}dx$$

be the Mellin transform of $T(x)$. Then using standard properties of the Mellin transform applied to the (slightly simplified) divide and conquer recurrence $T(x) = a(x) + \sum_{j=1}^m b_j T(p_j x)$ we arrive at

$$t(s) = \frac{a(s) + g(s)}{1 - \sum_{j=1}^m b_j p_i^{-s}},$$

where $a(s)$ is the Mellin transform of $a(x)$, and $g(s)$ is an additional function due to the initial conditions. Suppose that $a(s)$ and $g(s)$ are analytic for $\Re(s) \in (c,d)$ such that $s_0 \in (c,d)$, where s_0 is the root of $1 = \sum_j b_j p_i^{-s}$. Then we recover the asymptotics of $T(x)$ showing that

$$T(x) \sim Cx^{s_0} \quad \text{or} \quad T(x) \sim \Psi(\log x)x^{s_0}$$

where C is a constant, and $\Psi(t)$ is a (in general) discontinuous periodic function when the logarithms $\log p_j$ are rationally related (i.e., all ratios $(\log p_i)/(\log p_j)$ are rational, see Definition 1)

Discrete versions of the divide and conquer recurrence, given by (1) are more subtle and require a different approach. We will use Dirichlet series (closely related to the Mellin transform) that better captures the discrete nature of the recurrence, and then apply Tauberian theorems (and also the Mellin-Perron formula) to obtain asymptotics for $T(n)$. Precise results are presented in Theorem 2 in Section 3. In the particular case of sequences a_n of the form $a_n = Cn^a(\log n)^b$ (with $C > 0$ and $a, b \geq 0$) Theorem 2 has a more applicable form that we state in Theorem 1 of Section 2.

As in the continuous case the solution depends crucially on the relation between $\log p_1, \ldots, \log p_m$; when $\log p_1, \ldots, \log p_m$ are rationally related the final solution will exhibit some oscillation that disappears when $\log p_1, \ldots, \log p_m$ are irrationally related. This phenomenon was already observed for other discrete recurrences [Delange 1975; Fayolle et al. 1986; Flajolet and Sedgewick 2008; Grabner and Hwang 2005; Szpankowski 2001].

As a featured application of our results and techniques developed for solving the general discrete divide and conquer recurrence, we shall present a comprehensive analysis of a data compression algorithm due to Boncelet [Boncelet 1993], where we need even more precise results than stated in Theorem 2. Boncelet's algorithm is a variable-to-fixed data compression scheme. One of the best variable-to-fixed scheme belongs to Tunstall [Tunstall 1967]; another variation is due to Khodak [Khodak 1969]. Boncelet's algorithm is based on the divide and conquer strategy, and therefore is very fast and easy to implement. The question arises how it compares to the (asymptotically) optimal Tunstall algorithm. In Theorem 3 and Corollary 1 we provide an answer by first computing the redundancy of the Boncelet scheme (i.e., the excess of code length over the optimal code length) and compare it to the redundancy of the Tunstall code. In this case precise asymptotics of the Boncelet recurrence are crucial. We also prove in Theorem 7 of Section 6 that the phrase length of the Boncelet's algorithm obeys the central limit law, as for the Tunstall algorithm [Drmota et al. 2010]. This result actually generalizes divide and conquer recurrences to divide and conquer *functional recurrences* and allows us – under some additional assumptions – to prove a general central limit theorem.

The literature on continuous divide and conquer recurrence is very extensive. We mention here [Akra and Bazzi 1998; Choi and Golin 2001; Cormen et al. 1990]. The discrete version of the recurrence has received much less attention, especially with respect to precise asymptotics. Flajolet and Golin [Flajolet and Golin 1954] and Cheung et al. [Cheung et al. 2008] use similar techniques to ours, however, their recurrence is a simpler one with $p_1 = p_2 = 1/2$. Erdős et al. [Erdös et al. 1987] apply renewal theory and Hwang [Hwang 2000] (cf. also [Grabner and Hwang 2005; Hwang and Janson 2011]) analytic techniques when dealing with similar recurrences. The approach presented in this paper is generalized and somewhat simplified by using a combination of methods such as Tauberian theorems and Mellin-Perron techniques. To the best of our knowledge, there is no comprehensive analysis of the discrete divide and conquer recurrences and therefore there is no precise redundancy analysis for the Boncelet algorithm.

The paper is organized as follows. In the next section we present a short (but more applicable) version of our main result with several applications. The main result is then formulated in Section 3. The results for the Boncelet coding algorithm are presented in Sections 4 and 6. All proofs are delayed till Sections 5. Furthermore, in Appendix A we discuss analytic continuations properties of certain Dirichlet series, and in the Appendix B we present the Wiener-Ikehara Tauberian theorem and several extensions.

2. A DISCRETE MASTER THEOREM AND APPLICATIONS

In this section, we first present a version of the discrete master theorem for special toll functions $a_n = Cn^a(\log n)^b$ ($C > 0$, $a, b \geq 0$). Then we discuss a number of examples illustrating our master theorem.

2.1 A Simplified Discrete Master Theorem

We now consider a special toll function $a_n = Cn^a(\log n)^b$ and formulate our master theorem in this case. Its proof will follow from our general Discrete Master Theorem 2 presented in Section 3 and proved in Section 5.

It turns out that asymptotic behavior of $T(n)$ may depend on a relation between $\log p_1, \ldots \log p_m$. Therefore, we need the following definition.

Definition 1. *We say that* $\log p_1, \ldots, \log p_m$ *are rationally related if there exists a positive real number* L *such that* $\log p_1, \ldots, \log p_m$ *are integer multiples of* L*, that is,* $\log p_j = -n_j L$, $n_j \in \mathbb{Z}_{>0}$, *(*$1 \leq j \leq m$*). Equivalently this means that all ratios* $(\log p_i)/(\log p_j)$ *are rational. Without loss of generality we can assume that* L *is as large as possible which implies that* $\gcd(n_1, \ldots, n_m) = 1$*. Similarly, we say that* $\log p_1, \ldots, \log p_m$ *are irrationally related if they are not rationally related.*

Example. If $m = 1$, then we are always in the rationally related case. In the binary case $m = 2$, the numbers $\log p_1$, $\log p_2$ are rationally related if and only if the ratio $(\log p_1)/(\log p_2)$ is rational.

Now we are in the position to formulate our first (simplified) discrete master theorem.

Theorem 1 (Discrete Master Theorem — Special Case). *Let $T(n)$ be the divide and conquer recurrence defined in (1) with $a_n = Cn^a(\log n)^b$ ($C > 0$, $a, b \geq 0$) such that:*

(A1) b_j and $\overline{b}_j$ are non-negative with $b_j + \overline{b}_j > 0$,

(A2) $h_j(x)$ and $\overline{h}_j(x)$ are increasing and non-negative functions such that $h_j(x) = p_j x + O(x^{1-\delta})$ and $\overline{h}_j(x) = p_j x + O(x^{1-\delta})$ for positive $p_j < 1$ and $\delta > 0$, with $h_j(n) < n$ and $\overline{h}_j(n) \leq n - 1$ for all $n \geq 2$.

Furthermore, let s_0 be the unique real solution of the equation

$$\sum_{j=1}^{m} (b_j + \overline{b}_j) p_j^{s_0} = 1.$$

Then the sequence $T(n)$ has the following asymptotic behavior:

(i) If $a > s_0$, then

$$T(n) = \begin{cases} C'n^a(\log n)^b + O\left(n^a(\log n)^{b-1}\right) & \text{if } b > 0, \\ C'n^a + O(n^{a-\delta'}) & \text{if } b = 0, \end{cases}$$

where $\delta' = \min\{a - s_0, \delta\}$ and

$$C' = \frac{C}{1 - \sum_{j=1}^{m} (b_j + \overline{b}_j) p_j^a}.$$

(ii) If $a = s_0$, then

$$T(n) = C''n^a(\log n)^{b+1} + O\left(n^a(\log n)^b\right)$$

with

$$C'' = \frac{C}{(b+1)\sum_{j=1}^{m} (b_j + \overline{b}_j) p_j^a \log(1/p_j)}.$$

(iii) If $a < s_0$ (or if we just assume that $a_n = O(n^a)$ for some $a < s_0$ as long as a_n is a non-negative and non-decreasing sequence), then for $\log p_1, \ldots, \log p_m$ irrationally related

$$T(n) \sim C'''n^{s_0},$$

where C''' is a positive constant. If $\log p_1, \ldots, \log p_m$ are rationally related and if we also assume that

$$T(n+1) - T(n) = O\left(n^{s_0 - \eta}\right) \tag{4}$$

for some $\eta > 1 - \delta$, then

$$T(n) = \Psi(\log n)\, n^{s_0} + O\left(n^{s_0 - \eta'}\right)$$

where $\Psi(t)$ is a positive and periodic continuous function with period L and $\eta' > 0$.

Remark 1. It should be remarked that the order of magnitude of $T(n)$ can be checked easily by the Akra-Bazzi theorem [Akra and Bazzi 1998]. In particular, if we just know an upper bound for a_n which is of the form $a_n = O(n^a(\log n)^b)$ – even if a_n is not necessarily increasing – the Akra-Bazzi theorem provides an upper bound for $T(n)$ which is of form stated in Theorem 1. Hence the theorem can be easily adapted to cover a_n of the form $a_n = Cn^a(\log n)^b + O((n^{a_1}(\log n)^{b_1})$ with $a_1 < a$ or with $a_1 = a$ but $b_1 < b$. We split up the solution $T(n)$ into $T(n) = T_1(n) + T_2(n)$, where $T_1(n)$ corresponds to $a_n^{(1)} = Cn^a(\log n)^b$, for which we can apply Theorem 1, and $T_2(n)$ corresponds to the error term $a_n^{(2)} = O((n^{a_1}(\log n)^{b_1})$, for which we apply the Akra-Bazzi theorem.

The same idea can be used for a bootstrapping procedure. Theorem 1 provides the asymptotic leading term for $T(n)$ that is (for example, in case *(i)*) of the form $C'n^a(\log n)^b$. Hence, by setting $T(n) = C'n^a(\log n)^b + S(n)$ we obtain a recurrence for $S(n)$ that is precisely of the form (1) with a new sequence a_n that is of smaller order than the previous one. At this step we can either apply Theorem 1 a second time or the Akra-Bazzi theorem.

Remark 2. Theorem 1 can be extended to the case $a_n = Cn^a(\log n)^b$, where $a > 0$ and b is an arbitrary real number. The same result holds with the only exception $a = s_0$ and $b = -1$. In this case we obtain

$$T(n) = C''n^a \log\log n + O\left(n^a(\log n)^{-1}\right)$$

with

$$C'' = \frac{C}{\sum_{j=1}^m b_j p_j^a \log(1/p_j)}.$$

Remark 3. The third case *(iii)*: $a < s_0$, is of particular interest. Let us consider first the irrationally related case. Even in this case it is not immediate to describe the constant C''' explicitly. It depends heavily on a_n and also on $T(n)$ and can be written as

$$C''' = \frac{\tilde{A}(s_0) + \sum_{j=1}^m b_j(G_j(s_0) - E_j(s_0)) + \sum_{j=1}^m \overline{b}_j(\overline{G}_j(s_0) - \overline{E}_j(s_0))}{s_0 \sum_{j=1}^m (b_j + \overline{b}_j)p_j^{s_0}\log(1/p_j)} \tag{5}$$

with

$$\tilde{A}(s) = \sum_{n=1}^{\infty} \frac{a_{n+2} - a_{n+1}}{n^s},$$

and

$$G_j(s) = \sum_{n<n_j(1)} \frac{T\left(\lfloor h_j(n+2)\rfloor\right) - T\left(\lfloor h_j(n+1)\rfloor\right)}{n^s} + \frac{T(2) - T\left(\lfloor h_j(n_j(1)+1)\rfloor\right)}{n_j(1)} \tag{6}$$

$$E_j(s) = \sum_{k=1}^{\infty}(T(k+2) - T(k+1))\left(\frac{1}{(k/p_j)^s} - \frac{1}{n_j(k)^s}\right), \tag{7}$$

where $n_j(k) = \max\{n \ge 1 : h_j(n+1) < k+2\}$, and

$$\overline{G}_j(s) = \sum_{n<\overline{n}_j(1)} \frac{T\left(\lceil \overline{h}_j(n+2)\rceil\right) - T\left(\lceil \overline{h}_j(n+1)\rceil\right)}{n^s} + \frac{T(2) - T\left(\lceil \overline{h}_j(n_j(1)+1)\rceil\right)}{n_j(1)},$$

$$\overline{E}_j(s) = \sum_{k=1}^{\infty}(T(k+2)-T(k+1))\left(\frac{1}{(k/p_j)^s} - \frac{1}{\overline{n}_j(k)^s}\right),$$

where $\overline{n}_j(k) = \min\{n \ge 1 : \overline{h}_j(n+2) > k+1\}$. We will show in the proof that the series $\tilde{A}(s_0)$, $E_j(s_0)$ and $\overline{E}_j(s_0)$ actually converge. It should be also mentioned that there is no general error term in the asymptotic relation $T(n) \sim C''' n^{s_0}$.

In the rationally related case the periodic function $\Psi(t)$ has a convergent Fourier series $\Psi(t) = \sum_k c_k e^{2k\pi i x/L}$, where the Fourier coefficients are given by

$$c_k = \frac{\tilde{A}(s_k) + \sum_{j=1}^m b_j(G_j(s_k) - E_j(s_k)) + \sum_{j=1}^m \overline{b}_j(\overline{G}_j(s_k) - \overline{E}_j(s_k))}{s_k \sum_{j=1}^m (b_j + \overline{b}_j) p_j^{s_0} \log(1/p_j)}, \tag{8}$$

where $s_k = s_0 + 2k\pi i/L$. In particular the constant coefficient c_0 equals C'''. Note that it cannot be deduced from this representation that the Fourier series is convergent. This makes the problem really subtle.

Remark 4. It turns out that the assumption (4): $T(n+1) - T(n) = O(n^{s_0-\eta})$, which we call *small growth condition*, is essential for the result $T(n) \sim \Psi(\log n)\, n^{s_0}$ actually holds, see the *counter examples* in Example 3. On the other hand the small growth condition (4) is actually easy to check in practice. (Note also that in most applications we have $\delta = 1$ so that any $\eta > 0$ is sufficient.)

For example, if there exists n_0 such that for all $n \ge n_0$ there exist j such that

$$\lfloor h_j(n) \rfloor = \lfloor h_j(n+1) \rfloor \quad (\text{and } b_j > 0) \quad \text{or} \quad \lceil \overline{h}_j(n) \rceil = \lceil \overline{h}_j(n+1) \rceil \quad (\text{and } \overline{b}_j > 0), \tag{9}$$

then the small growth condition (4) is satisfied for all $\eta < s_0 - s'$, where s' is defined as the maximum of a and the real solutions of the equations

$$\sum_{\ell=1}^{m} (b_\ell + \overline{b}_\ell) p_\ell^s - b_j p_j^s = 1$$

for which $b_j > 0$, $1 \le j \le m$, and

$$\sum_{\ell=1}^{m} (b_\ell + \overline{b}_\ell) p_\ell^s - \overline{b}_j p_j^s = 1$$

for which $\overline{b}_j > 0$, $1 \le j \le m$. (The difference sequence $S(n) = T(n+1) - T(n)$ satisfies a divide-and-conquer-like recurrence with a trivial upper bound of the form $Cn^{s_0-\eta}$.)

In particular, if $p_1 = p$ and $p_2 = 1 - p$ and if

$$h_1(n) = p_1 n + \delta \quad \text{and} \quad \overline{h}_2(n) = p_2 n - \delta$$

(where $b_1 > 0$ and $\overline{b}_2 > 0$) then it is easy to check that (9) holds and consequently the small growth condition (4) is satisfied. This means that we do not have to care about (4) for recurrences of the form

$$T(n) = a_n + b\,T(\lfloor pn + \delta \rfloor) + \overline{b}\,T(\lceil (1-p)n - \delta \rceil).$$

The proof runs as follows. Assume that $\lfloor p_1 n + \delta \rfloor < \lfloor p_1(n+1) + \delta \rfloor$. This means that (with $m = \lfloor p_1 n + \delta \rfloor$) $m \le p_1 n + \delta < m + 1 \le p_1(n+1) + \delta$. If we set $x =$

$\{p_1n+\delta\} = p_1n+\delta-m$, then we have $x+p_1 \ge 1$ or $p_2 \le x$. Since $h_1(n)+\overline{h}_2(n) = n$ it follows that $\overline{h}_2(n) = \lceil \overline{h}_2(n) \rceil - x$. Hence $\overline{h}_2(n+1) = \overline{h}_2(n) + p_2 \le \lceil \overline{h}_2(n) \rceil$ and consequently $\lceil \overline{h}_j(n) \rceil = \lceil \overline{h}_j(n+1) \rceil$.

Note that if p_1 is irrational, then we can also work with $h_1(n) = p_1n + \delta$ and $h_2(n) = p_2n - \delta$ (where $b_1 > 0$ and $b_2 > 0$).

Remark 5. In several applications of case *(ii)* the second order term is of interest, however, in contrast to the leading term the behavior of the second order term depends again on the arithmetic properties of the $\log p_i$. For example, if $a = b = s_0 = 0$, that is, the recurrence is of the form

$$T(n) = \sum_{j=1}^{m} b_j T(\lfloor p_j n \rfloor) + C$$

with $b_1 + \cdots + b_m = 1$ and if we are in the irrationally related case, then

$$T(n) = C'' \log n + C_2'' + o(1),$$

where

$$C'' = \frac{C}{\sum_{j=1}^{m} b_j \log(1/p_j)}$$

and C_2'' is a constant that can be computed similarly as C'''; see Example 10. Furthermore, if we can assume a corresponding *small growth condition* of the form $T(n+1) - T(n) = O(n^{-\eta})$ for some $\eta > 0$, then

$$T(n) = C'' \log n + \Psi(\log n) + o(1),$$

where $\Psi(t)$ is a continuous periodic function.

A similar statement holds in the case $a = s_0 = 1$, $b = 0$. Here the corresponding small growth condition is $T(n+1) - T(n) = O(n^{1-\eta})$ (for some $\eta > 0$), see Examples 6 and 10.

2.2 Applications

We first illustrate our theorem on a few simple divide and conquer recurrences. Some of these examples are also discussed in [Leighton 1996], where the growth order of $T(n)$ is determined. Note that we only consider examples for the case *(iii)* and *(ii)*, since they are more interesting.

Example 1. Consider the recurrence

$$T(n) = 2\,T(\lfloor n/2 \rfloor) + 3\,T(\lfloor n/6 \rfloor) + n \log n.$$

Here we have $a = b = 1$. Furthermore the equation

$$2 \cdot 2^{-s} + 3 \cdot 6^{-s} = 1$$

has the (real) solution $s_0 = 1.402\ldots > 1$ and it is easy to check that $\log(1/2)$ and $\log(1/6)$ are irrationally related. Namely, if $\log(1/2)/\log(1/6)$ were rational, say c/d then it would follow that $2^d = 6^c$. However, this equation has no non-zero integer solution. Hence by case *(iii)* we obtain

$$T(n) \sim C''' n^{s_0} \qquad (n \to \infty)$$

for some constant $C''' > 0$. By using (5) we find numerically that $C''' = 5.61\ldots$. Note that $n_1(k) = 2k+2$ and $n_2(k) = 6k+10$. Figure 1 shows the precise behavior of $T(n)$.

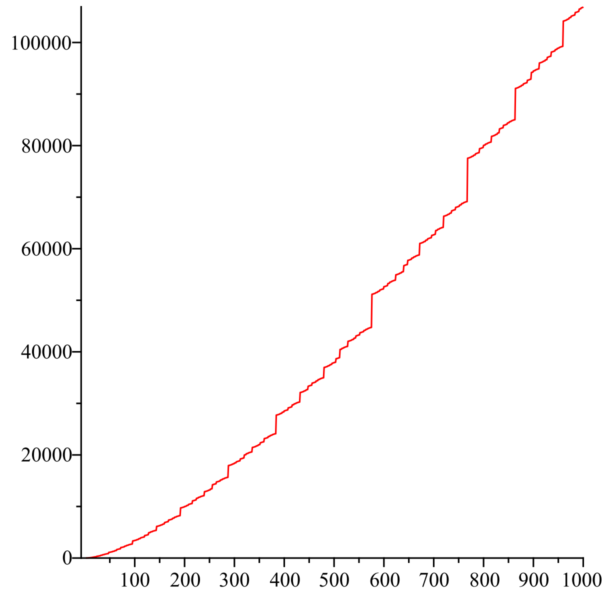

Fig. 1. $T(n)$ versus n from Example 1.

Example 2. The recurrence

$$T(n) = 2T(\lfloor n/2 \rfloor) + 1 \quad (\text{with } T(1) = 1)$$

is formally of the kind covered by Theorem 1: we have $a = b = 0$ and $s_0 = 1 > 0$. Since $m = 1$ we are in the rationally related case. However, it is easy to check that

$$T(n) = 2^{\lfloor \log_2 n \rfloor + 1} - 1.$$

In particular we have $T(2^k) = 2^{k+1} - 1$ and $T(2^k - 1) = 2^k - 1$. Consequently, the small growth condition (4) is not satisfied. Actually we can write

$$T(n) = n\Psi(\log_2 n) - 1$$

with $\Psi(t) = 2^{1-\{t\}}$, where $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of x, that is, the assertion of Theorem 1 holds formally, however, the the periodic function is discontinuous at $t = 0$.

Next consider the same kind of recurrence with a different sequence a_n, namely

$$T(n) = 2\,T(\lfloor n/2 \rfloor) + \log_2 n$$

with $T(1) = 0$. Here $a = 0$, $b = 1$, and $s_0 = 1 > 0$. Again the small growth property (4) is not satisfied. By induction it follows that $T(n)$ has the following explicit representation:

$$T(n) = \sum_{0 \le k \le \log_2 n} 2^k \log_2\left(\lfloor n/2^k \rfloor\right) = n \sum_{0 \le \ell \le \lfloor \log_2 n \rfloor} 2^{-\{\log_2 n\} - \ell} \log_2\left(\lfloor 2^{\{\log_2 n\}+\ell} \rfloor\right).$$

Hence we have

$$T(n) = n\Psi(\log_2 n) + O(\log n),$$

where

$$\Psi(t) = 2^{-\{t\}} \sum_{\ell \ge 0} 2^{-\ell} \log_2\left(\lfloor 2^{\{t\}+\ell} \rfloor\right)$$

is a periodic function that is discontinuous for the countable (and dense) set $\{\{\log_2 m\} : m \in \mathbb{Z}_{\ge 1}\}$.

Example 3. The two recurrences of Example 2 give rise to the conjecture that Theorem 1 might be generalized to the case, where the small growth property (4) is not satisfied. However, this is certainly not true as the two following **counter examples** show. They indicate that there is probably no easy characterization when we have $T(n) \sim n^{s_0}\Psi(\log n)$ for a periodic function $\Psi(t)$.

For example consider the recurrence

$$T(n) = 2\,T\left(\left\lfloor \frac{2}{3} n \right\rfloor\right)$$

with $T(1) = 1$. Here we have $T(n) = 2^k$ for $n_k \le n < n_{k+1}$, where $n_0 = 1$ and $n_{k+1} = \left\lceil \frac{3}{2} n_k \right\rceil$. It is clear that $c_1(3/2)^k \le n_k \le c_2(3/2)^k$ for some positive constants c_1, c_2, however, the precise behaviour of n_k is *erratic* so that we cannot expect a precise behaviour of the kind $n_k = c(3/2)^k + O(1)$ and consequently not a representation of the form $T(n) \sim n^{s_0}\Psi(\log n)$, since the *jumps* from $T(n_k - 1)$ to $T(n_k)$ cannot be covered with the help of a single function $\Psi(t)$.

Next consider the recurrence

$$T(n) = 2\,T\left(\left\lfloor \frac{n + 2\sqrt{n} + 1}{2} \right\rfloor\right) \qquad (n \ge 6)$$

with $T(n) = 1$ for $1 \le n \le 5$. Here we have (again) $T(n) = 2^k$ for $n_k \le n < n_{k+1}$, where $n_1 = 6$ and $n_{k+1} = \lceil 2n_k + 1 - 2\sqrt{2n_k} \rceil$. It follows that n_k is asymptotically of the form $n_k = c_1 2^k - c_2 2^{k/2} + O(1)$ (for certain positive constants c_1, c_2). Consequently it is again not possible to represent $T(n)$ asymptotically as $T(n) \sim n\Psi(\log n)$.

Example 4. The recurrence

$$T(n) = T(\lfloor n/2 \rfloor) + 2\,T(\lceil n/2 \rceil) + n$$

is related to the Karatsuba algorithm [Karatsuba and Ofman 1963; Knuth 1998]. Here we have $s_0 = \log(1/3)/\log(1/2) = 1.5849\ldots$ and $s_0 > a = 1$. Furthermore, since $m = 1$, we are in the rationally related case. Here the small growth condition (4) is satisfied so that we can apply Theorem 1 to obtain

$$T(n) = \Psi(\log n)\, n^{\frac{\log 3}{\log 2}} \cdot (1 + o(1)) \qquad (n \to \infty)$$

with some continuous periodic function $\Psi(t)$.

In a similar manner, the Strassen algorithm [Cormen et al. 1990; Strassen 1969] for matrix multiplications results in the following recurrence

$$T(n) = T(\lfloor n/2 \rfloor) + 6\,T(\lceil n/2 \rceil) + n^2.$$

Here we have $m = 1$, $s_0 = \log 7/\log 2 \approx 2.81$ and $a = 2$, and again we get an representation of the form

$$T(n) = \Psi(\log n)\, n^{\frac{\log 7}{\log 2}} \cdot (1 + o(1)) \qquad (n \to \infty)$$

with some periodic function $\Psi(t)$.

Example 5. The next two examples show that a small change in the recurrence might change the asymptotic behaviour significantly. First let

$$T(n) = T(\lfloor n/2 \rfloor) + T(\lceil n/4 \rceil)$$

with $T(1) = 1$. Here we have $s_0 = \log((1+\sqrt{5})/2)\log 2 \approx 0.6942$ and we are in the rationally related case. Furthermore it follows easily that $T(n+1) - T(n) \le 1$. Hence the small growth condition (4) is satisfied and we obtain

$$T(n) \sim n^{s_0}\Psi(\log_2 n)$$

for a continuous periodic function $\Psi(t)$.

However, if we just replace the appearing ceiling function by the floor function, that is,

$$\tilde{T}(n) = \tilde{T}(\lfloor n/2 \rfloor) + \tilde{T}(\lfloor n/4 \rfloor) \qquad \text{for } n \ge 4$$

and $\tilde{T}(1) = \tilde{T}(2) = \tilde{T}(3) = 1$, then the small growth condition (4) is not satisfied . We get $\tilde{T}(n) = F_k$ for $2^{k-1} \le n < 2^k$, where F_k denotes the k-th Fibonacci number. This leads to

$$\tilde{T}(n) \sim n^{s_0}\tilde{\Psi}(\log_2 n),$$

where $\tilde{\Psi}(t) = ((1+\sqrt{5})/2)^{1-\{t\}}/\sqrt{5}$ is discontinuous for $t = 0$; see also Figure 2.

Example 6. The recurrences

$$\begin{aligned}
T(n) &= T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n - 1,\\
Y(n) &= Y(\lfloor n/2 \rfloor) + Y(\lceil n/2 \rceil) + \lfloor n/2 \rfloor,\\
U(n) &= U(\lfloor n/2 \rfloor) + U(\lceil n/2 \rceil) + n - \frac{\lfloor n/2 \rfloor}{\lceil n/2 \rceil + 1} + \frac{\lceil n/2 \rceil}{\lfloor n/2 \rfloor + 1}\lfloor n/2 \rfloor
\end{aligned}$$

are related to Mergesort (see [Flajolet and Golin 1954]). For all three recurrences we have $a = s_0 = 1$ and we are (again) in the rationally related case. Here it is immediate to derive a-priori bounds of the form $T(n+1) - T(n) = O(\log n)$ (and corresponding ones for $Y(n)$ and $U(n)$).

Hence, we obtain asymptotic expansions of the form

$$C\,n\log n + n\Psi(\log n) + o(n) \qquad (n \to \infty),$$

where $C = 1/\log 2$ for $T(n)$ and $U(n)$ and $C = 1/(2\log 2)$ for $Y(n)$, and $\Psi(t)$ is a continuous periodic function.

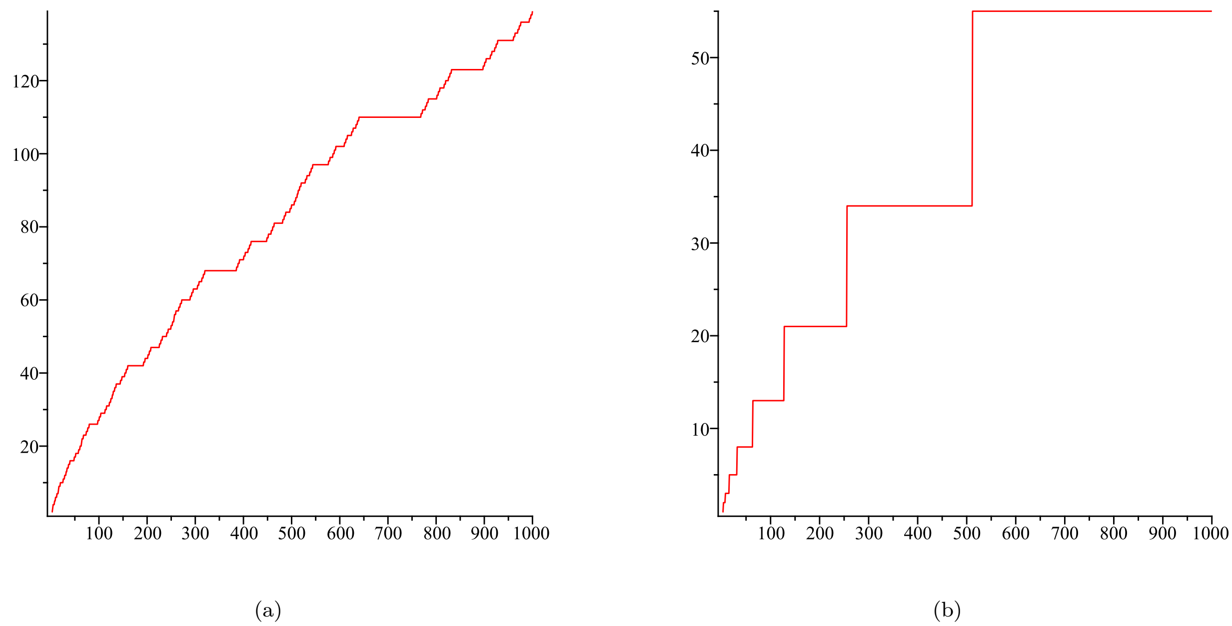

Fig. 2. Illustration to Example 5: (a) recurrence $T(n) = T(\lfloor n/2 \rfloor) + T(\lceil n/4 \rceil)$, (b) recurrence $T(n) = T(\lfloor n/2 \rfloor) + T(\lfloor n/2 \rfloor)$.

Example 7. Consider now

$$T(n) = T(\lfloor n/2 \rfloor) + \log n.$$

Here we have $a = s_0 = 0$ and consequently, according to case *(ii)* we have

$$T(n) = \frac{1}{2\log 2}(\log n)^2 + O(\log n).$$

Example 8. Next consider the recurrence

$$T(n) = 2\,T(\lfloor n/2 \rfloor) + \frac{8}{9}T(\lfloor 3n/4 \rfloor) + \frac{n^2}{\log n}.$$

Here $a = s_0 = 2$. Hence, by the extended case *(ii)* (described in Remark 2) we have

$$T(n) = \frac{2}{\log 2 + \log(4/3)} n^2 \log\log n + O(n^2/\log n).$$

Example 9. The solution of the recurrence

$$T(n) = \frac{1}{3}T\left(\left\lfloor \frac{n}{3} + \frac{1}{2} \right\rfloor\right) + \frac{2}{3}T\left(\left\lceil \frac{2n}{3} - \frac{1}{2} \right\rceil\right) + 1$$

with initial value $T(1) = 0$ is asymptotically given by

$$T(n) = \frac{1}{H}\log n + C + o(1),$$

with $H = \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2} \approx 0.6365$ and some constant C. By Remark 5, we are in the irrationally related case. With the help of Theorem 3 we can compute $C \approx -0.0813$. Its precise form is $C = -\alpha/H$, where

$$\begin{aligned}\alpha &= \sum_{m\ge 1}\frac{T(m+2)-T(m+1)}{3}\left(\log\left\lceil 3m+\frac{5}{2}\right\rceil - \log(3m)\right)\\ &+ 2\sum_{m\ge 1}\frac{T(m+2)-T(m+1)}{3}\left(\log\left\lfloor \frac{3}{2}m+\frac{5}{4}\right\rfloor - \log(\frac{3m}{2})\right)\\ &+ \frac{\log 3}{3} - H - \frac{\frac{1}{3}\log^2 3+\frac{2}{3}\log^2\frac{3}{2}}{2H} \approx 0.0518.\end{aligned}$$

We will use this example for computing the redundancy of the binary Boncelet code with $p = 1/3$ in Section 4.

Example 10. Let $s_2(k)$ denote the binary sum-of-digits function of a non-negative integer k and let $T(n) = \sum_{k<n} s_2(k)$ the partial sums. Then $T(n)$ satisfies the recurrence

$$T(n) = T(\lfloor n/2\rfloor) + T(\lceil n/2\rceil) + \lfloor n/2\rfloor$$

and $T(1) = 0$. It is a well known fact (originally due to [Delange 1975]) that $T(n)$ is given by

$$T(n) = \frac{1}{2}n\log_2 n + n\Psi(\log_2 n),$$

where $\Psi(t)$ a periodic function that is even continuous. There are several different representations for $\Psi(t)$. For example we have

$$\Psi(t) = 2^{-\{t\}}\sum_{\ell\ge 0}2^{-\ell}g\left(2^{\ell+\{t\}}\right) + \frac{1-\{x\}}{2},$$

where

$$g(t) = \int_0^t\left(\lfloor 2\{s/2\}\rfloor - \frac{1}{2}\right)ds.$$

Example 11. The recurrence

$$T(n) = \frac{1}{4}T(\lfloor n/2\rfloor) + \frac{1}{4}T(\lceil n/2\rceil) + \frac{1}{n}$$

is not covered by Theorem 1 since a_n is decreasing. Hence, $T(n)$ is not increasing, either. However, we can adapt the proof methods of Theorem 1. Formally we have $a = s_0 = -1 < 0$ and, since $m = 1$, we are in the rationally related case. It follows that

$$T_n = \frac{1}{\log 2}\frac{\log n}{n} + \frac{\Psi(\log n)}{n} + o\left(\frac{1}{n}\right)$$

with a periodic function $\Psi(t)$.

3. DIRICHLET SERIES AND DISCRETE DIVIDE AND CONQUER RECURRENCES

In this section we present the full version of our theorem concerning the precise asymptotic behavior of discrete divide and conquer recurrence.

3.1 Divide and Conquer Recurrence

Let us recall the general form of divide and conquer recurrences that we will analyze. For $m \ge 1$, let $b_1, \ldots, b_m$ and $\overline{b}_1, \ldots, \overline{b}_m$ be positive real numbers and $h_j(x)$ and $\overline{h}_j(x)$ non-decreasing positive functions with $h_j(x) = p_j x + O(x^{1-\delta})$ and $\overline{h}_j(x) = p_j x + O(x^{1-\delta})$ for some positive numbers $p_j < 1$ and some $\delta > 0$ (for $1 \le j \le m$). We consider a (general) divide and conquer recurrence: given $T(0) \le T(1)$ for $n \ge 2$ we set

$$\begin{aligned} T(n) &= a_n + \sum_{j=1}^{m} b_j T\left(\lfloor h_j(x) \rfloor\right) + \sum_{j=1}^{m} \overline{b}_j T\left(\lceil \overline{h}_j(x) \rceil\right) \qquad (n \ge 2), \qquad (10) \\ &= a_n + \sum_{j=1}^{m} b_j T\left(\lfloor p_j x + O(x^{1-\delta}) \rfloor\right) + \sum_{j=1}^{m} \overline{b}_j T\left(\lceil p_j x + O(x^{1-\delta}) \rceil\right) \end{aligned}$$

where $(a_n)_{n\ge 2}$ is a known *non-negative* and *non-decreasing* sequence. We also assume that $h_j(n) < n$ and $\overline{h}_j(n) \le n-1$ (for $n \ge 2$ and $1 \le j \le m$) so that the recurrence is well defined. It follows by induction that $T(n)$ is nondecreasing, too. In order to solve recurrence (10), we use Dirichlet series [Apostol 1976; Szpankowski 2001]. In fact, in the proof presented in Section 5 we make use of the following Dirichlet series

$$\widetilde{T}(s) = \sum_{n=1}^{\infty} \frac{T(n+2) - T(n+1)}{n^s} \qquad (11)$$

from which we can calculate $\sum_{i=1}^{n-2} (T(i+2) - T(i+1)) = T(n) - T(2)$.

For an asymptotic solution of recurrence (10), we will make some assumptions regarding the Dirichlet series of the known sequence a_n. We postulate that the abscissa of absolute convergence σ_a of the Dirichlet series

$$\widetilde{A}(s) = \sum_{n=1}^{\infty} \frac{a_{n+2} - a_{n+1}}{n^s} \qquad (12)$$

is finite (or $-\infty$), that is, $\widetilde{A}(s)$ represents an analytic function for $\Re(s) > \sigma_a$. For example, if we know that a_n is non-decreasing and

$$a_n = O(n^a (\log n)^b)$$

for some real number a and b, then $\widetilde{A}(s)$ converges (absolutely) for $\Re(s) > a$. In particular, we have $\sigma_a \le a$.

Analytically, these observations follow from the fact, proved in Section 5, that the Dirichlet series $\widetilde{T}(s)$ can be expressed as

$$\widetilde{T}(s) = \frac{\widetilde{A}(s) + B(s)}{1 - \sum_{j=1}^{m} (b_j + \overline{b}_j)\, p_j^s} \qquad (13)$$

for some analytic function $B(s)$ and $\widetilde{A}(s)$ defined in (12). For the asymptotic analysis, we appeal to the Tauberian theorem by Wiener-Ikehara and an analysis based on the Mellin-Perron formula (see Appendix B and Section 5.3). Both approaches rely on the singular behavior of $\widetilde{T}(s)$. By the Mellin-Perron formula, we shall

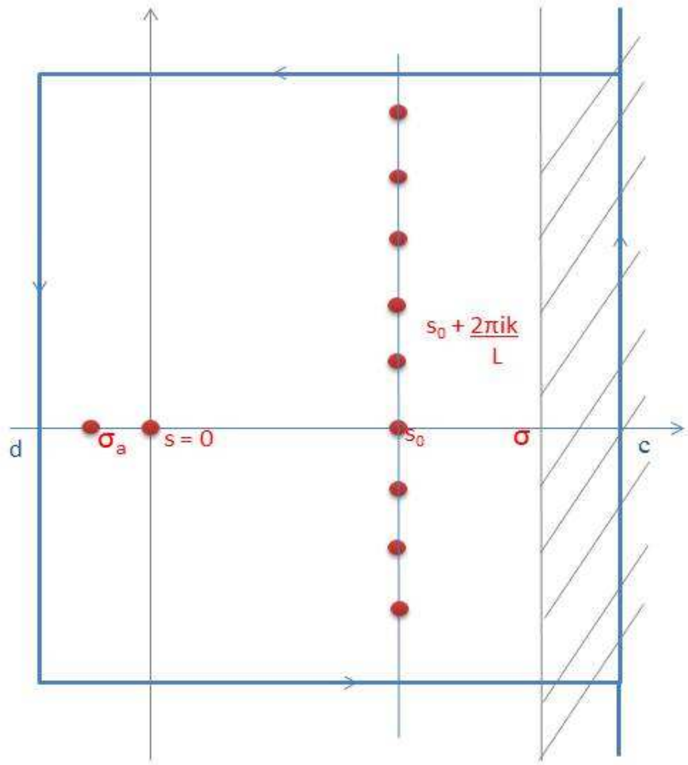

Fig. 3. Illustration to the asymptotic analysis of the divide and conquer recurrence

observe that

$$T(n) = T(2) + \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \widetilde{T}(s) \frac{(n - \frac{3}{2})^s}{s} \, ds. \tag{14}$$

Hence, the asymptotic behavior of $T(n)$ depends on the singular behavior of $\widetilde{A}(s)$, on the singularity at $s = 0$, and on the roots of the denominator in (13), that is, roots of the *characteristic equation*

$$\sum_{j=1}^{m} (b_j + \overline{b}_j) \, p_j^s = 1. \tag{15}$$

We denote by s_0 the unique real solution of this equation.

A master theorem has usually three (major) parts. In the first case the (asymptotic) behavior of a_n dominates the asymptotics of $T(n)$, in the second case, there is an *interaction* between the internal structure of the recurrence and the sequence a_n (resonance), and in the third case the behavior of the solution is driven by the recurrence and does not depend on a_n; see the three cases of Theorem 1. This also corresponds to an interplay between the poles $s = 0$, $s = \sigma_a$ and s_0 that determines the asymptotic behavior as illustrated in Figure 3. In fact, the pole of the largest value dictates asymptotics and determines the leading term.

We will handle these cases separately. If $s_0 < \sigma_a$ or if $s_0 = \sigma_a$ (that is, we are in the first two cases) we have to assume some regularity properties about the sequence a_n in order to cope with the asymptotics of $T(n)$. We assume that $\widetilde{A}(s)$

has a certain extension to a region that contains the line $\Re(s) = \sigma_a$ with a pole-like singularity at $s = \sigma_a$. To be more precise, we will assume that there exist functions $\tilde{F}(s)$, $g_0(s), \ldots, g_J(s)$ that are analytic in a region that contains the half plane $\Re(s) \ge \sigma_a$ such that

$$\widetilde{A}(s) = g_0(s) \frac{\left(\log \frac{1}{s-\sigma_a}\right)^{\beta_0}}{(s-\sigma_a)^{\alpha_0}} + \sum_{j=1}^{J} g_j(s) \frac{\left(\log \frac{1}{s-\sigma_a}\right)^{\beta_j}}{(s-\sigma_a)^{\alpha_j}} + \tilde{F}(s), \tag{16}$$

where $g_0(\sigma_a) \neq 0$, β_j are non-negative integers, α_0 is real, and $\alpha_1, \ldots, \alpha_J$ are complex numbers with $\Re(\alpha_j) < \alpha_0$ $(1 \le j \le J)$, and β_0 is non-negative if α_0 is contained in the set $\{0, -1, -2, \ldots\}$.

As demonstrated in Appendix A, this is certainly the case if a_n is a linear combination of sequences of the form

$$n^a (\log n)^b$$

(or related to such sequences with floor and ceiling functions). For example, if b is not a negative integer, then the corresponding Dirichlet series $\widetilde{A}(s)$ of the sequence $a_n = n^a (\log n)^b$ can be expressed as

$$\widetilde{A}(s) = a \frac{\Gamma(b+1)}{(s-a)^{b+1}} + \frac{\Gamma(b+1)}{(s-a)^b} + \tilde{F}(s),$$

where $\tilde{F}(s)$ is analytic for $\Re(s) > a - 1$, see Theorem 9 of Appendix A. Therefore, if $a \neq 0$, then

$$\sigma_a = a \quad \text{and} \quad \alpha_0 = b + 1$$

and if $a = 0$ and $b \neq 0$ (and not a negative integer), then

$$\sigma_a = a = 0 \quad \text{and} \quad \alpha_0 = b.$$

Of course, if $a = b = 0$ then $\tilde{A}(s) = 0$ and $\sigma_a = -\infty$.

If $s_0 = \sigma_a$ or if $s_0 > \sigma_a$, then the zeros of the characteristic equation (15) determines the asymptotic behavior. It turns out – as already seen – we need to consider two different scenarios depending whether $\log p_1, \ldots, \log p_m$ are rationally related or not (cf. Definition 1). This governs the location of the roots of our characteristic equation (15). The following property of the roots of (15) is due to [Schachinger 2001] (cf. also [Drmota et al. 2010; Flajolet et al. 2010]).

Lemma 1. *Let s_0 be the unique real solution of equation (15). Then all other solutions s' of (15) satisfy $\Re(s') \le s_0$.*
(i) *If $\log p_1, \ldots, \log p_m$ are irrationally related, then s_0 is the only solution of (15) on $\Re(s) = s_0$.*
(ii) *If $\log p_1, \ldots, \log p_m$ are rationally related, then there are infinitely many solutions s_k, $k \in \mathbb{Z}$, with $\Re(s_k) = s_0$ which are given by*

$$s_k = s_0 + k \frac{2\pi i}{L} \qquad (k \in \mathbb{Z}),$$

where $L > 0$ is the largest real number such that $\log p_j$ are all integer multiples of L. Furthermore, there exists $\delta > 0$ such that all remaining solutions of (15) satisfy $\Re(s) \le s_0 - \delta$.

3.2 General Discrete Master Theorem

We are now ready to formulate our main results regarding the asymptotic solutions of discrete divide and conquer recurrences. Note that the irrational case is easier to handle whereas the rational case needs additional assumptions on the Dirichlet series. Nevertheless these assumptions are usually easy to establish in practice.

As discussed, our Discrete Master Theorem shows that for sequences a_n of practical importance such as

$$a_n = n^a (\log n)^b$$

the solution $T(n)$ of the divide and conquer recurrence grows as

$$T(n) \sim C\, n^{a'} (\log n)^{b'} (\log\log n)^{c'} \tag{17}$$

with $a' = \max\{a, s_0\}$) when $\log p_1, \ldots, \log p_m$ are irrationally related. For rationally related $\log p_1, \ldots, \log p_m$, it is either of the form (17) or (if $s_0 > a$) there appears an oscillation in the form of

$$T(n) \sim \Psi(\log n)\, n^{s_0} \tag{18}$$

with a continuous periodic function $\Psi(t)$. The proof of the following general result will be given in Section 5.

Theorem 2 (Discrete Master Theorem – Full Version)**.** *Let $T(n)$ be the divide and conquer recurrence defined in (10) such that:*

(A1) b_j and $\overline{b}_j$ are non-negative with $b_j + \overline{b}_j > 0$,

(A2) $h_j(x)$ and $\overline{h}_j(x)$ are increasing and non-negative functions such that $h_j(x) = p_j x + O(x^{1-\delta})$ and $\overline{h}_j(x) = p_j x + O(x^{1-\delta})$ for positive $p_j < 1$ and $\delta > 0$, with $h_j(2) < 2$, $\overline{h}_j(2) \le 1$,

(A3) the sequence $(a_n)_{n\ge 2}$ is non-negative and non-decreasing.

(A4) Let σ_a denote the abscissa of absolute convergence of the Dirichlet series $\widetilde{A}(s)$ and s_0 be the real root of (15). If $\sigma_a \ge s_0 \ge 0$ assume further that $\widetilde{A}(s)$ has a representation of the form (16), where $\tilde{F}(s)$, $g_0(s), \ldots, g_J(s)$ are analytic in a region that contains the half plane $\Re(s) \ge \sigma_a$ with $g_0(\sigma_a) \ne 0$, α_0 is real and $\Re(\alpha_j) < \alpha_0$ $(1 \le j \le J)$.

(i) If $\log p_1, \ldots, \log p_m$ are irrationally related, then as $n \to \infty$

$$T(n) = \begin{cases} C_1 + o(1) & \text{if } \sigma_a < 0 \text{ and } s_0 < 0,\\ C_2 \log n + C_2' + o(1) & \text{if } \sigma_a < s_0 \text{ and } s_0 = 0,\\ C_3 (\log n)^{\alpha_0+1} (\log\log n)^{\beta_0 - I} \cdot (1+o(1)) & \text{if } \sigma_a = s_0 = 0,\\ C_4\, n^{s_0} \cdot (1+o(1)) & \text{if } \sigma_a < s_0 \text{ and } s_0 > 0,\\ C_5 n^{s_0} (\log n)^{\alpha_0} (\log\log n)^{\beta_0 - I} \cdot (1+o(1)) & \text{if } \sigma_a = s_0 > 0,\\ C_6\, (\log n)^{\alpha_0} (\log\log n)^{\beta_0 - I} (1+o(1)) & \text{if } \sigma_a = 0 \text{ and } s_0 < 0,\\ C_7\, n^{\sigma_a} (\log n)^{\alpha_0 - 1} (\log\log n)^{\beta_0 - I} \cdot (1+o(1)) & \text{if } \sigma_a > s_0 \text{ and } \sigma_a > 0, \end{cases} \tag{19}$$

with positive real constants $C_1, C_2, C_3, C_4, C_5, C_6, C_7$, where in particular if $\alpha_0 \notin \{0, -1, -2, \ldots\}$ we have $I = 0$. We only have $I = 1$ if $\alpha_0 \in \{0, -1, -2, \ldots\}$, $\beta_0 > 0$

and if in the corresponding cases of (19) we have

$$\begin{array}{l} \sigma_a = s_0 = 0 \text{ and } \alpha_0 \le -2, \\ \sigma_a = s_0 > 0 \text{ and } \alpha_0 \le -1, \\ \text{if } \sigma_a = 0,\ s_0 < 0, \text{ and } \alpha_0 \le -1, \text{ or} \\ \sigma_a > s_0 \text{ and } \sigma_a > 0. \end{array}$$

(ii) *If* $\log p_1, \ldots, \log p_m$ *are rationally related and if in the case* $s_0 = \sigma_a$ *the Fourier series, with* L *defined in Lemma 1(ii),*

$$\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\widetilde{A}(s_0 + 2\pi ik/L)}{s_0 + 2\pi ik/L} e^{2\pi ikx/L} \tag{20}$$

is convergent for $x \in \mathbb{R}$ *and represents an integrable function, then* $T(n)$ *behaves as in the irrationally related case with the following two exceptions:*

$$T(n) = \begin{cases} C_2 \log n + \Psi_2(\log n) + O(n^{-\eta'}) & \text{if } \sigma_a < s_0 \text{ and } s_0 = 0, \\ \Psi_4(\log n)\, n^{s_0} + O(n^{s_0-\eta'}) & \text{if } \sigma_a < s_0 \text{ and } s_0 > 0, \end{cases} \tag{21}$$

where C_2 *and* $\Psi_4(t)$ *are positive and* $\Psi_2(t), \Psi_4(t)$ *are continuous periodic functions with period* L *and* $\eta' > 0$*, provided that the small growth condition*

$$T(n+1) - T(n) = O(n^{s_0-\eta}) \tag{22}$$

holds for some $\eta > 1 - \delta$*.*

Remark 6. We should point out that the periodic functions $\Psi_2(t)$ and $\Psi_4(t)$ that appear in the second part of Theorem 2 have building blocks of the form

$$\lambda^{-t/L} \sum_{n\ge1} B_n \frac{\lambda^{\left\lfloor \frac{t-\log n}{L} \right\rfloor + 1}}{\lambda - 1}$$

for some $\lambda > 1$ and a sequence B_n such that the series $\sum_{n\ge1} B_n \lambda^{-(\log n)/L}$ is absolutely convergent. This representation suggests that the periodic functions should have countably many discontinuities and, thus, should not have absolutely convergent Fourier series. Nevertheless the *small growth condition* (22) ensures that the final periodic function is actually continuous (see Lemma 4), however, this property is not immediate. Actually we can expect Hölder continuous functions, that is $|\Psi(s) - \Psi(t)| \le C|s-t|^\eta$ for some $\eta > 0$ and absolutely convergent Fourier series, see [Grabner and Hwang 2005].

Remark 7. The condition (20) for $\widetilde{A}(s)$ looks artificial. However, it is really needed in the proof in order to control the polar singularities of $\widetilde{T}(s)$ at s_k, $k \in \mathbb{Z} \setminus \{0\}$. Nevertheless it is no real restriction in practice. As shown in Appendix A the condition (20) is satisfied for sequences of the form $a_n = n^a (\log n)^b$.

We now briefly show how Theorem 1 can be deduced from Theorem 2 which we prove in Section 5.

Case $a > s_0$:

We recall that if $a_n = Cn^a(\log n)^b$, then we have $\sigma_a = a$ if a or b are different from zero and $\sigma_a = -\infty$ if $a = b = 0$. Suppose that we are in the first case. By Theorems 9 and 10 of Appendix A the Dirichlet series $\tilde{A}(s)$ satisfies the assumptions of Theorem 2. Recall that $\alpha_0 = b+1$ and that $\beta_0 = I = 0$. Consequently the last case of (19) applies and we obtain the asymptotic leading term for $T(n) \sim C'n^a(\log n)^b$. The constant C' can be either determined from the proof of Theorem 2 or more directly by inserting $C'n^a(\log n)^b$ into the recurrence and *comparing coefficients* – we leave the easy details to the reader.

In order to obtain the remainder term we follow Remark 1. We set $T(n) = C'n^a(\log n)^b + S(n)$ and using the relation

$$\left(p_j n + O(n^{1-\delta})\right)^a \left(\log(p_j n + O(n^{1-\delta})\right)^b = n^a(\log n)^b p_j^a + O\left(n^a(\log n)^{b-1}\right)$$

we obtain (if $b > 0$) a recurrence for $S(n)$ of the form

$$S(n) = \sum_{j=1}^{m} b_j S\left(\lfloor h_j(n) \rfloor\right) + \sum_{j=1}^{m} \overline{b}_j S\left(\lceil \overline{h}_j(n) \rceil\right) + O\left(n^a(\log n)^{b-1}\right).$$

Hence by the Akra-Bazzi theorem it follows that $S(n) = O\left(n^a(\log n)^{b-1}\right)$, too. If $b = 0$, then the error term is replaced by $O(n^{a-\delta})$ and we get even a better estimate for $S(n)$.

Finally if $a = b = 0$ and $s_0 < 0$, then we are in the first case of Theorem 2 and we observe that $T(n) = C' + o(1)$, where $C' = C/\left(1 - \sum_{j=1}^{m}(b_j + \overline{b}_j)\right)$. By setting $T(n) = C' + S(n)$ we get a homogeneous recurrence for $S(n)$ and the Akra-Bazzi theorem proves $S(n) = O(n^{s_0})$.

Case $a = s_0$:
If $a > 0$ or $b > 0$, then $\sigma_a = a$. If $a > 0$, then we apply the fifth case of Theorem 2 (with $\alpha_0 = b+1$ and $\beta_0 = I = 0$) and obtain $T(n) \sim C''n^a(\log n)^{b+1}$. As in case *(i)* we obtain C'' explicitly and also the error term with a bootstrapping procedure.

If $a = 0$ and $b > 0$, then we have $\sigma_a = s_0 = 0$ and $\alpha_0 = b$. Consequently the third case of Theorem 2 applies. Again the constant as well the error term are derived as above.

Case $a < s_0$:
Observe that $a < s_0$ implies $\sigma_a < s_0$. Consequently we can apply the fourth case of Theorem 2 and obtain $T(n) \sim C'''n^{s_0}$ in the irrationally related case and $T(n) \sim \Psi(\log n)n^{s_0}$ in the rationally related case.

Finally we mention that the extensions of Theorem 1 discussed in Remarks 2–4 are also covered by Theorem 2. We leave the details to the reader.

4. BONCELET'S ARITHMETIC CODING ALGORITHM

We present a novel application of our analytic approach to discrete divide and conquer recurrences by computing the redundancy of a practical variable-to-fixed compression algorithm due to Boncelet [Boncelet 1993]. To recall, a variable-to-fixed length encoder partitions the source string, say over an m-ary alphabet $\mathcal{A}$, into a concatenation of variable-length phrases. Each phrase belongs to a given dictionary of source strings. A uniquely parsable dictionary is represented by a *complete parsing tree*, i.e., a tree in which every internal node has all m children nodes. The dictionary entries correspond to the *leaves* of the associated parsing tree. The encoder represents each parsed string by the fixed length binary code word corresponding to its dictionary entry. There are several well known variable-to-fixed algorithms; e.g., Tunstall and Khodak schemes (cf. [Drmota et al. 2010; Khodak 1969; Tunstall 1967]). Boncelet's algorithm, described next, is a practical and computationally fast algorithm that is coming into use. Therefore, we compare its redundancy to the (asymptotically) optimal Tunstall's algorithm.

Boncelet describes his algorithm in terms of a parsing tree. For fixed n (representing the number of leaves in the parsing tree and hence also the number of distinct phrases), the algorithm in each step creates two subtrees of predetermined number of leaves (phrases). Thus at the root, n is split into two subtrees with the number of leaves, respectively, equal to $n_1 = \lfloor p_1 n + \frac{1}{2} \rfloor$ and $n_2 = \lfloor p_2 n + \frac{1}{2} \rfloor$. This continues recursively until only 1 or 2 leaves are left. Note that this splitting procedure does not assure that $n_1 + n_2 = n$. For example if $p_1 = \frac{3}{8}$ and $p_2 = \frac{5}{8}$, then $n = 4$ would be split into $n_1 = 2$ and $n_2 = 3$. Therefore, we propose to modify the splitting as follows $n_1 = \lfloor p_1 n + \delta \rfloor$ and $n_2 = \lceil p_2 n - \delta \rceil$ for some $\delta \in (0, 1)$ that satisfies $2p_1 + \delta < 2$.

Let $\{v_1, \ldots v_n\}$ denote phrases of the Boncelet code that correspond to the paths from the root to leaves of the parsing tree, and let $\ell(v_1), \ldots, \ell(v_n)$ be the corresponding phrase lengths. Observe that while the parsing tree in the Boncelet's algorithm is fixed, a randomly generated sequence is partitioned into random length phrases. Therefore, one can talk about the probabilities of phrases denoted as $P(v_1), \ldots, P(v_n)$. Here we restrict the analysis to a binary alphabet and denote the probabilities by $p := p_1$ and $q := p_2 = 1 - p$.

For sequences generated by a binary memoryless source, we aim at understanding the probabilistic behavior of the phrase length that we denote as D_n. Its probability generating function is defined as

$$C(n, y) = \mathbb{E}[y^{D_n}]$$

which can also be represented as

$$C(n, y) = \sum_{j=1}^{n} P(v_j) y^{\ell(v_j)}.$$

The Boncelet's splitting procedure leads to the following recurrence on $C(n, y)$ for $n \geq 2$

$$C(n, y) = p\, y\, C\left(\lfloor pn + \delta \rfloor, y\right) + q\, y\, C\left(\lceil qn - \delta \rceil, y\right) \tag{23}$$

with initial conditions $C(0, y) = 0$ and $C(1, y) = 1$.

Next let $d(n)$ denote the average phrase length

$$\mathbb{E}[D_n] := d(n) = \sum_{j=1}^{n} P(v_j)\, \ell(v_j)$$

which is also given by $d(n) = C'(n, 1)$ (where the derivative is taken with respect to y) and satisfies the recurrence

$$d(n) = 1 + p_1 d\left(\lfloor p_1 n + \delta \rfloor\right) + p_2 d\left(\lceil p_2 n - \delta \rceil\right) \tag{24}$$

with $d(0) = d(1) = 0$. This recurrence falls exactly under our general divide and conquer recurrence, hence Theorem 2 applies.

Theorem 3. *Consider a binary memoryless source with positive probabilities $p_1 = p$ and $p_2 = q$ and the entropy rate $H = p\log(1/p) + q\log(1/q)$. Let $d(n) = \mathbb{E}[D_n]$ denote the expected phrase length of the binary Boncelet code.*
(i) *If the ratio $(\log p)/(\log q)$ is irrational, then*

$$d(n) = \frac{1}{H}\log n - \frac{\alpha}{H} + o(1), \tag{25}$$

where

$$\alpha = \tilde{E}'(0) - \tilde{G}'(0) - H - \frac{H_2}{2H}, \tag{26}$$

$H_2 = p\log^2 p + q\log^2 q$, and $\tilde{E}'(0)$ and $\tilde{G}'(0)$ are the derivatives at $s = 0$ of the Dirichlet series defined in (32) of Section 5 and Remark 3 of Section 2.
(ii) *If $(\log p)/(\log q)$ is rational, then*

$$d(n) = \frac{1}{H}\log n - \frac{\alpha + \Psi(\log n)}{H} + O(n^{-\eta}), \tag{27}$$

where $\Psi(t)$ is a periodic function and $\eta > 0$.

For practical data compression algorithms, it is important to achieve low redundancy defined as the excess of the code length over the optimal code length nH. For variable-to-fixed codes, the average redundancy is expressed as [Drmota et al. 2010; Savari and Gallager 1997]

$$R_n = \frac{\log n}{\mathbb{E}[D_n]} - H = \frac{\log n}{d(n)} - H$$

since every phrase of average length $d(n)$ requires $\log n$ bits to point to a dictionary entry. Our previous results imply immediately the following corollary.

Corollary 1. *Let R_n denote the redundancy of the binary Boncelet code with positive probabilities $p_1 = p$ and $p_2 = q$.*
(i) *If the ratio $(\log p)/(\log q)$ is irrational, then*

$$R_n = \frac{H\alpha}{\log n} + o\left(\frac{1}{\log n}\right) \tag{28}$$

with α defined in (26).
(ii) *If $(\log p)/(\log q)$ is rational, then*

$$R_n = \frac{H(\alpha + \Psi(\log n))}{\log n} + o\left(\frac{1}{\log n}\right) \tag{29}$$

where $\Psi(t)$ is a periodic function.

We should compare the redundancy of Boncelet's algorithm to the asymptotically optimal Tunstall algorithm. From [Drmota et al. 2010; Savari and Gallager 1997] we know that the redundancy of the Tunstall code is

$$R_n^T = \frac{H}{\log n}\left(-\log H - \frac{H_2}{2H}\right) + o\left(\frac{1}{\log n}\right)$$

(provided that $(\log p)/(\log q)$ is irrational; in the rational case there is also a periodic term in the leading asymptotics). This should be compared to the redundancy of the Boncelet algorithm.

Example. Consider $p = 1/3$ and $q = 2/3$. Then the recurrence for $d(n)$ is precisely the same as that of Example 9. Consequently $\alpha \approx 0.0518$ while for the Tunstall code the corresponding constant is equal to $-\log H - \frac{H_2}{2H} \approx 0.0496$.

5. ANALYSIS AND ASYMPTOTICS

We prove here a general asymptotic solution of the divide and conquer recurrence (cf. Theorem 2). We first derive the appropriate Dirichlet series and apply Tauberian theorems for the irrationally related case, then discuss the Perron-Mellin formula, and finally finish the proof of Theorem 2 for the rationally related case.

5.1 Dirichlet Series

As discussed in the previous section, the proof makes use of the Dirichlet series

$$\widetilde{T}(s) = \sum_{n=1}^{\infty} \frac{T(n+2) - T(n+1)}{n^s},$$

where we apply Tauberian theorems and the Mellin-Perron formula to obtain asymptotics for $T(n)$ from singularities of $\widetilde{T}(s)$.

By partial summation and using a-priori upper bounds for the sequence $T(n)$, it follows that $\widetilde{T}(s)$ converges (absolutely) for $s \in \mathbb{C}$ with $\Re(s) > \max\{s_0, \sigma_a, 0\}$, where s_0 is the real solution of the equation (15), and σ_a is the abscissa of absolute convergence of $\tilde{A}(s)$.

Next we apply the recurrence relation (10) to $\widetilde{T}(s)$. To simplify our presentation, we assume that $\overline{b}_j = 0$, that is, we consider only the floor function on the right hand side of the recurrence (10); those parts that contain the ceiling function can be handled in the same way. We thus obtain

$$\widetilde{T}(s) = \widetilde{A}(s) + \sum_{j=1}^{m} b_j \sum_{n=1}^{\infty} \frac{T\left(\lfloor h_j(n+2) \rfloor\right) - T\left(\lfloor h_j(n+1) \rfloor\right)}{n^s}.$$

For $k \geq 1$ set

$$n_j(k) := \max\{n \geq 1 : h_j(n+1) < k+2\}.$$

By definition it is clear that $n_j(k+1) \geq n_j(k)$ and

$$n_j(k) = \frac{n}{p_j} + O\left(k^{1-\delta}\right). \tag{30}$$

Furthermore, by setting $G_j(s)$ as in (6) of Remark 3 we obtain

$$\sum_{n=1}^{\infty} \frac{T\left(\lfloor h_j(n+2) \rfloor\right) - T\left(\lfloor h_j(n+1) \rfloor\right)}{n^s} = G_j(s) + \sum_{k=1}^{\infty} \frac{T(k+2) - T(k+1)}{n_j(k)^s}.$$

We now compare the last sum to $p_j^s \widetilde{T}(s)$:

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{T(k+2)-T(k+1)}{n_j(k)^s} &= \sum_{k=1}^{\infty} \frac{T(k+2)-T(k+1)}{(k/p_j)^s} \\ &\quad - \sum_{k=1}^{\infty} (T(k+2)-T(k+1)) \left(\frac{1}{(k/p_j)^s} - \frac{1}{n_j(k)^s} \right) \\ &= p_j^s \widetilde{T}(s) - E_j(s),\end{aligned}$$

where $E_j(s)$ is defined in (7) of Remark 3. Defining

$$E(s) = \sum_{j=1}^{m} b_j E_j(s) \quad \text{and} \quad G(s) = \sum_{j=1}^{m} b_j G_j(s)$$

we finally obtain the relation

$$\widetilde{T}(s) = \frac{\widetilde{A}(s) + G(s) - E(s)}{1 - \sum_{j=1}^m b_j\, p_j^s}. \tag{31}$$

As mentioned above, (almost) the same procedure applies if some of the $\overline{b}_j$ are positive, that is, the ceiling function also appear in the recurrence equation. The only difference to (31) is that we arrive at a representation of the form

$$\widetilde{T}(s) = \frac{\widetilde{A}(s) + \tilde{G}(s) - \tilde{E}(s)}{1 - \sum_{j=1}^m (b_j + \overline{b}_j)\, p_j^s}, \tag{32}$$

with a properly modified functions $\tilde{G}(s)$ and $\tilde{E}(s)$, however, they have the same analyticity properties as in (31), compare also with Remark 3.

By our previous assumptions, we know the analytic behaviors of

$$\widetilde{A}(s) \qquad \text{and} \qquad \left(1 - \sum_{j=1}^{m} (b_j + \overline{b}_j)\, p_j^s \right)^{-1} :$$

$\widetilde{A}(s)$ has a pole-like singularity at $s = \sigma_a$ (if $\sigma_a \ge s_0$) and a proper continuation to a complex domain that contains the (punctuated) line $\Re(s) = \sigma_a$, $s \ne \sigma_a$. On the other hand,

$$\frac{1}{1 - \sum_{j=1}^m (b_j + \overline{b}_j)\, p_j^s}$$

has a polar singularity at $s = s_0$ (and infinitely many other poles on the line $\Re(s) = s_0$ if the numbers $\log p_j$ are rationally related), and also a meromorphic continuation to a complex domain that contains the line $\Re(s) = s_0$. Furthermore, $G(s)$ is an entire function. It suffices to discuss $E_j(s)$. First observe that (30) implies

$$\frac{1}{(k/p_j)^s} - \frac{1}{n_j(k)^s} = O\left(\frac{1}{(k/p_j)^{\Re(s)+\delta}} \right).$$

By partial summation (and by using again the a-priori estimates), it follows immediately that the series

$$\sum_{k=1}^{\infty} (T(k+2) - T(k+1)) \frac{1}{(k/p_j)^{\Re(s)+\delta}}$$

converges for $\Re(s) > \max\{s_0, \sigma_a, 0\} - \delta$. Since $T(n)$ is an increasing sequence, this implies (absolute) convergence of the series $E_j(s)$, just representing an analytic function in this region, too.

In order to recover (asymptotically) $T(n)$ from $\widetilde{T}(s)$ we need to apply several different techniques discussed in the next subsection. The main analytic tools are Tauberian theorems (of Wiener-Ikehara which is discussed in detail in Appendix B) and the Mellin-Perron formula (Theorem 4).

5.2 Tauberian Theorems

We are now ready to prove several parts of Theorem 2 with the help of Tauberian theorems of Wiener-Ikehara type (see Appendix B). We recall that such theorems apply in general to the so-called Mellin-Stieltjes transform

$$\int_{1-}^{\infty} v^{-s}\, d\overline{c}(v) = s \int_{1}^{\infty} \overline{c}(v) v^{-s-1}\, dv$$

of a non-negative and non-decreasing function $\overline{c}(v)$. If $c(n)$ is a sequence of non-negative numbers, then the Dirichlet series $C(s) = \sum_{n\ge 1} c(n)n^{-s}$ is just the Mellin-Stieltjes transform of the function $\overline{c}(v) = \sum_{n\le v} c(n)$:

$$C(s) = \sum_{n\ge 1} c(n)n^{-s} = \int_{1-}^{\infty} v^{-s}\, d\overline{c}(v) = s \int_{1}^{\infty} \overline{c}(v) v^{-s-1}\, dv.$$

Informally, a Tauberian theorem is a correspondence between the singular behavior of $\frac{1}{s}C(s)$ and the asymptotic behavior of $\overline{c}(v)$. In the context of Tauberian theorems of Wiener-Ikehara type one assumes that $C(s)$ continues analytically to a proper region, has only one (real) singularity s_0 on the *critical line* $\Re(s) = s_0$, and the singularity is of special type (for example a polar or algebraic singularity, see Appendix B).

We recall that $\widetilde{T}(s)$ is the Dirichlet series of the sequence $c(n) = T(n+2) - T(n+1)$. Hence

$$T(n) = \overline{c}(n-2) + T(2).$$

Consequently, if we know the asymptotic behavior of $\overline{c}(v)$ we also find that of $T(n)$. Notice that $\widetilde{T}(s)$ is given by (32). Hence the dominant singularity of $\frac{1}{s}\widetilde{T}(s)$ is either zero, or induced by the singular behavior of $\widetilde{A}(s)$, or induced by the zeros of the denominator

$$1 - \sum_{j=1}^{m} (b_j + \overline{b}_j) p_j^s.$$

Here it is essential to assume that the $\log p_j$ are *irrationally related*. Precisely in this case the denominator has only the real zero s_0 on the line $\Re(s) = s_0$. Hence

Tauberian theorems can be applied in the irrationally related case if $s_0 \ge \sigma_a$. (For the rational case we will apply a different approach to cover the case $s_0 \ge \sigma_a$.)

Our conclusions for the proof of the first part of Theorem 2 are summarized as follows:

(1) $\sigma_a < 0$ and $s_0 < 0$:
This is indeed a trivial case, since the dominant singularity is at $s = 0$ and the series $\widetilde{T}(s)$ converges for $s = 0$:

$$\widetilde{T}(0) = \sum_{n \ge 1} (T(n+2) - T(n+1)),$$

hence

$$T(n) = C_1 + o(1),$$

where $C_1 = T(2) + \widetilde{T}(0)$.

(2) $\sigma_a < s_0$ and $s_0 = 0$:
We can apply directly a proper version of the Wiener-Ikehara theorem (Theorem 12 of Appendix B) that proves

$$T(n) = C_2 \log n \cdot (1 + o(1)).$$

Observe, that $s = 0$ is a double pole of $\frac{1}{s}\widetilde{T}(s)$ that induces the $\log n$-term in the asymptotic expansion. Note that this does not prove the full version that is stated in Theorem 2. By applying Theorem 5 of the next subsection (that is based on a more refined analysis) we also arrive at an asymptotic expansion of the form

$$T(n) = C_2 \log n + C_2' + o(1).$$

(3) $\sigma_a = s_0 = 0$:
In this case, we obtain at $s = 0$ the dominant singular term of $\frac{1}{s}\widetilde{T}(s)$ that is given by

$$C \, \frac{(\log(1/s))^{\beta_0}}{s^{\alpha_0 + 2}} \quad \text{with} \quad C = \frac{-g_0(0)}{\sum_{j=1}^m (b_j + \overline{b}_j) \log p_j}$$

Hence, an application of Theorem 13 of Appendix B provides the asymptotic leading term for $T(n)$. Recall that we have to handle separately the case when α_0 is contained in the set $\{-2, -3, \ldots\}$ (and $\beta_0 > 0$). In this case, only logarithmic singularities remain.

(4) $\sigma_a < s_0$ and $s_0 > 0$:
Here the classical version of the Wiener-Ikehara theorem (Theorem 11 of Appendix B) applies. Note again that it is crucial that the denominator has only one pole on the line $\Re(s) = s_0$.

(5) $\sigma_a = s_0 > 0$:
Here the function $\frac{1}{s}\widetilde{T}(s)$ has the dominant singular term

$$C \, \frac{(\log(1/(s - \sigma_a)))^{\beta_0}}{(s - \sigma_a)^{\alpha_0 + 1}}$$

for some constant $C > 0$ (and there are no other singularities on the line $\Re(s) = s_0$). Thus, an application of Theorem 13 of Appendix B provides the asymptotic leading term for $T(n)$. Observe that we have to handle separately the case when α_0 is contained in the set $\{-1, -2, \ldots\}$ (and $\beta_0 > 0$).

(6) $\sigma_a = 0$ and $s_0 < 0$:
The analysis of this case is very close to the previous one. The dominant singular term of $\frac{1}{s}\widetilde{T}(s)$ is of the form

$$C\,\frac{(\log(1/s))^{\beta_0}}{s^{\alpha_0+1}}.$$

(7) $\sigma_a > s_0$ and $\sigma_a > 0$:
In this case the singular behavior of $\widetilde{A}(s)$ dominates the asymptotic behavior of $\frac{1}{s}\widetilde{T}(s)$. An application of Theorem 13 of Appendix B provides the asymptotic leading term of $T(n)$.

5.3 Mellin-Perron Formula

One disadvantage of the use of Tauberian theorems is that they provide (usually) only the asymptotic leading term and no error terms. In order to provide error terms or second order terms one has to use more refined methods. In the framework of Dirichlet series we can apply the Mellin-Perron formula that we recall next.

Below we shall use Iverson's notation $[\![P]\!]$ which is 1 if P is a true proposition and 0 else.

Theorem 4 (see [Apostol 1976])**.** *For a sequence $c(n)$ define the Dirichlet series*

$$C(s) = \sum_{n=1}^{\infty} \frac{c(n)}{n^s}$$

and assume that the abscissa of absolute convergence σ_a is finite or $-\infty$. Then for all $\sigma > \sigma_a$ and all $x > 0$

$$\sum_{n<x} c(n) + \frac{c(\lfloor x \rfloor)}{2}[\![x \in \mathbb{Z}]\!] = \lim_{T\to\infty} \frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT} C(s)\frac{x^s}{s}\,ds.$$

Note that – similarly to the Tauberian theorems – the Mellin-Perron formula enables us to obtain precise information about the function $\overline{c}(v) = \sum_{n\le v} c(n)$ if we know the behavior of $\frac{1}{s}C(s)$. In our context we have $c(n) = T(n+2) - T(n)$, that is,

$$T(n) = T(2) + \lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT} \widetilde{T}(s)\frac{(n-\frac{3}{2})^s}{s}\,ds \tag{33}$$

with

$$\widetilde{T}(s) = \sum_{n=1}^{\infty} \frac{T(n+2) - T(n+1)}{n^s}.$$

As a first application we apply the Mellin-Perron formula of Theorem 4 for Dirichlet series of the form

$$C(s) = \sum_{n\ge 1} c(n) n^{-s} = \frac{B(s)}{1 - \sum_{j=1}^m b_j p_j^s}, \tag{34}$$

where we assume that the $\log p_j$ are *not* rationally related and where $B(s)$ is analytic in a region that contains the real zero s_0 of the denominator. This theorem can be also applied to the proof of some parts of Theorem 2; in particular for the (irrationally related) cases

$$\begin{array}{l} \text{if } \sigma_a < 0 \text{ and } s_0 < 0, \\ \text{if } \sigma_a < s_0 \text{ and } s_0 = 0, \text{ and} \\ \text{if } \sigma_a < s_0 \text{ and } s_0 > 0. \end{array}$$

Note that Theorem 5 provides a second order term in the case $\sigma_a < s_0 = 0$, see also Remark 9.

Theorem 5. *Suppose that $0 < p_j < 1$, $1 \le j \le m$, are given such that $\log p_j$, $1 \le j \le m$, are not rationally related and let s_0 denote the real solution of the equation*

$$\sum_{j=1}^{m} b_j p_j^s = 1,$$

where $b_j > 0$, $1 \le j \le m$. Let $C(s) = \sum_{n\ge 1} c(n) n^{-s}$ be a Dirichlet series with non-negative coefficients $c(n)$ that has a representation of the form (34), that is,

$$C(s) = \sum_{n\ge 1} c(n) n^{-s} = \frac{B(s)}{1 - \sum_{j=1}^{m} b_j p_j^s}$$

where $B(s)$ is an analytic function for $\Re(s) \ge s_0 - \eta$ for some $\eta > 0$ and satisfies $B(s) = O(|s|^\alpha)$ in this region for some $\alpha < 1$. Then

$$\sum_{n \le v} c(n) = \begin{cases} \dfrac{B(0)}{1 - \sum_{j=1}^{m} b_j} + o(1) & \text{if } s_0 < 0, \\ \dfrac{B(0)}{H(0)} \log v + \dfrac{B'(0) + B(0) H_2/H}{H} + o(1) & \text{if } s_0 = 0, \\ \dfrac{B(s_0)}{s_0 H(s_0)} v^{s_0} \left(1 + o(1)\right) & \text{if } s_0 > 0 \end{cases},$$

where $H(s) = -\sum_{j=1}^{m} b_j p_j^s \log p_j$ with $H = H(0)$, and $H_2(s) = \sum_{j=1}^{m} b_j p_j^s (\log p_j)^2$ with $H_2 = H_2(0)$.

We quickly check that Theorem 5 is applicable for $\widetilde{T}(s)$ in the above mentioned cases. Since $\tilde{A}(s)$ and $\tilde{G}(s)$ are convergent Dirichlet series with non-negative coefficients that stay bounded for $\Re(s) \ge s_0 - \eta$, however, for $\tilde{E}(s)$ we only get an estimate of the form $\tilde{E}(s) = O(|s|^\alpha)$ for some $\alpha < 1$ (we leave the details to the reader). Thus, everything fits together.

Remark 8. Note that the case $s_0 < 0$ is immediate since the series is convergent for $s = 0$. Furthermore, the case $s_0 > 0$ is covered by the Wiener-Ikehara theorem. The case $s_0 = 0$ is the most interesting case. Here the Wiener-Ikehara theorem provides only the asymptotic leading term. However, the assumptions of Theorem 5 are much stronger than those needed for the Wiener-Ikehara theorem. Actually we can cover also multiple poles and obtain also corresponding (non-leading) terms in the asymptotic expansion. In order to present these kind of techniques we will consider the cases $s_0 > 0$ and $s_0 = 0$ in the following proof.

Proof. We start with the case $s_0 > 0$. We will use the Mellin-Perron formula of Theorem 4, however, we cannot use it directly, since there are convergence problems. Namely, if we shift the line of integration $\Re(s) > s_0$ to the left (to $\Re(s) = \sigma < s_0$) and collect residues we obtain (with $\mathcal{Z} = \{s \in \mathbb{C} : \sum_{j=1}^m b_j p_j^s = 1\}$)

$$\begin{aligned}
\sum_{n\le v} c(n) &= \lim_{T\to\infty} \sum_{s'\in\mathcal{Z},\ \Re(s')>\sigma,\, |\Im(s')|<T} \mathrm{Res}(C(s)\,\frac{v^s}{s},\ s=s') \\
&\quad + \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\sigma-iT}^{\sigma+iT} C(s)\frac{v^s}{s}\,ds \\
&= \lim_{T\to\infty} \sum_{s'\in\mathcal{Z},\ \Re(s')>\sigma,\, |\Im(s')|<T} \frac{B(s')v^{s'}}{s'H(s')} \\
&\quad + \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\sigma-iT}^{\sigma+iT} \frac{B(s)}{1-\sum_{j=1}^m b_j p_j^s}\frac{v^s}{s}\,ds
\end{aligned}$$

provided that the series of residues converges and the limit $T \to \infty$ of the last integral exists. The problem is that neither the series nor the integral above are necessarily absolutely convergent since the integrand is only of order $1/s$. We have to introduce the auxiliary function

$$\overline{c}_1(v) = \int_0^v \left(\sum_{n\le w} c(n)\right) dw$$

which is also given by

$$\overline{c}_1(v) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} C(s)\frac{v^{s+1}}{s(s+1)}\,ds = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{B(s)}{1-\sum_{j=1}^m b_j p_j^s}\cdot\frac{v^{s+1}}{s(s+1)}\,ds,$$

for $c > s_0$. Note that there is no need to consider the limit $T \to \infty$ in this case since the series and the integral are now absolutely convergent. Hence, the above procedure works without any convergence problem. In order to make the presentation of our analysis slightly easier we additionally assume that the region of analyticity of $B(s)$ is large enough such that it covers all zeros in $\mathcal{Z}$ and also the point -1. We now shift the line of integration to $\Re(s) = \sigma < \min\{-1, s_0\}$. Then we have to consider the (absolutely convergent) sum of residues

$$\sum_{s'\in\mathcal{Z},} \mathrm{Res}\left(C(s)\,\frac{v^{s+1}}{s(s+1)},\ s=s'\right) = \sum_{s'\in\mathcal{Z}} \frac{B(s')}{s'(s'+1)H(s')}v^{s'+1},$$

the residues at $s = 0$ and $s = -1$:

$$\frac{B(0)}{1-\sum_{j=1}^m b_j}v, \quad -\frac{B(-1)}{1-\sum_{j=1}^m b_j p_j^{-1}},$$

and the remaining (absolutely convergent) integral

$$\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} C(s)\frac{v^{s+1}}{s(s+1)}\,ds = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} \frac{B(s)}{1-\sum_{j=1}^m b_j p_j^s}\frac{v^{s+1}}{s(s+1)}\,ds = O(v^{1+\sigma}).$$

Thus, we obtain

$$\overline{c}_1(v) = \frac{B(s_0)}{s_0(s_0+1)H(s_0)}(1+Q(\log v))v^{1+s_0} + O(v^{1+s_0-\eta})$$

for some $\eta > 0$, where

$$Q(x) = \sum_{s' \in \mathcal{Z}\setminus\{s_0\}} \frac{s_0(s_0+1)H(s_0)B(s')}{s'(s'+1)H(s')B(s_0)} e^{x(s'-s_0)}.$$

It is easy to show that $Q(x) \to 0$ as $x \to \infty$ (cf. also [Schachinger 2001, Lemma 4] and [Szpankowski 2001]). Indeed, suppose that $\varepsilon > 0$ is given. Then there exists $S_0 = S_0(\varepsilon) > 0$ such that

$$\sum_{s' \in \mathcal{Z},\, |s'| > S_0} \left| \frac{s_0(s_0+1)H(s_0)B(s')}{s'(s'+1)H(s')B(s_0)} \right| < \frac{\varepsilon}{2}.$$

Further, since $\Re(s') < s_0$ for all $s' \in \mathcal{Z}\setminus\{s_0\}$, and by the assumption of irrationality the zeros are not on the critical line $\Re(s) = s_0$ (except the real one), it follows that there exists $x_0 = x_0(\varepsilon) > 0$ with

$$\left| \sum_{s' \in \mathcal{Z}\setminus\{s_0\},\, |s'| \le S_0} \frac{s_0(s_0+1)H(s_0)B(s')}{s'(s'+1)H(s')B(s_0)} e^{x(s'-s_0)} \right| < \frac{\varepsilon}{2}$$

for $x \ge x_0$. Hence $|Q(x)| < \varepsilon$ for $x \ge x_0(\varepsilon)$.

Note that we cannot obtain the rate of convergence for $Q(x)$. This means that we just get

$$\overline{c}_1(v) = \frac{B(s_0)}{s_0(s_0+1)H(s_0)} \cdot v^{1+s_0} + o(v^{1+s_0})$$

as $v \to \infty$, where s' with $\Re(s') < s_0$ contribute to the error term. However, since, $\sum_{n \le v} c(n)$ is monotonely increasing in v (by assumption) it also follows that

$$\sum_{n \le v} c(n) \sim \frac{B(s_0)}{s_0 H(s_0)}\, v^{s_0},$$

compare with the case $s_0 = 0$ that we discuss next.

Now suppose that $s_0 = 0$ which means that $C(s)$ has a double pole as $s = 0$. We can almost use the same analysis as above and obtain the asymptotic expansion

$$\overline{c}_1(v) = \frac{B(0)}{H} v \log v + \frac{B'(0) - B(0) + B(0)H_2/H}{H}\, v + o(v).$$

It is now an easy exercise to derive from this expansion the final result

$$\sum_{n \le v} c(n) = \frac{B(0)}{H} \log v + \frac{B'(0) + B(0)H_2/H}{H} + o(1) \qquad (35)$$

in the following way. For simplicity we write $\overline{c}_1(v) = C_1 v \log v + C_2 v + o(v)$. By the assumption

$$|\overline{c}_1(v) - C_1 v \log v + C_2 v| \le \varepsilon v$$

for $v \ge v_0$. Set $v' = \varepsilon^{1/2} v$, then by monotonicity we obtain (for $v \ge v_0$)

$$\begin{aligned}\sum_{n\le v} c(n) \ &\le \ \frac{\overline{c}_1(v+v') - \overline{c}_1(v)}{v'} \le \frac{1}{v'}\left(C_1(v+v')\log(v+v') + C_2(v+v') - \right.\\ &\qquad \left. C_1 v\log v - C_2 v\right) + \varepsilon\frac{2v+v'}{v'}\\ &= \ C_1\log(v+v') + C_2 + C_1\frac{v}{v'}\log\left(1+\frac{v'}{v}\right) + \varepsilon\frac{2v+v'}{v'}\\ &= \ C_1\log v + C_2 + C_1 + O\left(\varepsilon^{1/2}\right),\end{aligned}$$

where the O-constant is an absolute one. In a similar manner, we obtain the corresponding lower bound (for $v \ge v_0 + v_0^{1/2}$). Hence, it follows that

$$\left|\sum_{n\le v} c(n) - C_1\log v - C_1 - C_2\right| \le C'\varepsilon^{1/2}$$

for $v \ge v_0 + v_0^{1/2}$. This proves $\sum_{n\le v} c(n) = C_1\log v + C_1 + C_2 + o(1)$ and consequently (35).

Remark 9. The advantage of the preceding proof is its flexibility. For example, we can apply the procedure for multiple poles and are able to derive asymptotic expansions of the form

$$\sum_{n\le v} c(n) = \sum_{j=0}^{K} A_j\,\frac{(\log v)^j}{j!}\,v^{s_0} + o(v^{s_0}).$$

Furthermore we can derive asymptotic expansions that are uniform in an additional parameter when we have some control on the singularities in terms of the additional parameter. We will use this generalization in the proof of the central limit theorem for the phrase lengths of the Boncelet code (Theorem 7).

In principle it is also possible to obtain bounds for the error terms. However, they depend heavily on Diophantine approximation properties of the vector $(\log p_1, \ldots \log p_m)$, see [Flajolet et al. 2010].

5.4 The Rationally Related Case

Unfortunately, the previous method generally is not applicable when there are several poles (or infinitely many poles) on the line $\Re(s) = s_0$. This means that we cannot use the above procedure when the $\log p_j$ are rationally related. The reason is that it does not follow *automatically* that an asymptotic expansion of the form

$$\overline{c}_1(v) = \int_0^v \overline{c}(w)\,dw \sim \Psi_1(\log v)\cdot v^{s_0+1}$$

implies

$$\overline{c}(v) \sim \Psi(\log v)\cdot v^{s_0}$$

for certain periodic functions Ψ and Ψ_1, even if $\overline{c}(v)$ is non-negative and non-decreasing.

Therefore we will apply an alternative approach which is – in some sense – more direct and applies *only* in this case, but it proves a convergence result for $c(v)$ of the form

$$\overline{c}(v) = \sum_{n\le v} c(n) \sim \Psi(\log v)\, v^{s_0}.$$

Suppose that $\log p_j = -n_j L$ for coprime integers n_j and a real number $L > 0$. Then the equation $1 = \sum_{j=1}^m b_j p_j^s$ with the only real solution s_0 becomes an algebraic equation

$$1 - \sum_{j=1}^{m} b_j z^{n_j} = 0 \qquad \text{with } z = e^{-Ls}.$$

with a single (dominating) real root $z_0 = e^{-Ls_0}$. We can factor this polynomial as

$$1 - \sum_{j=1}^{m} b_j z^{n_j} = (1 - e^{Ls_0} z)P(z), \qquad P(e^{-Ls}) \ne 0,$$

and obtain also a partial fraction decomposition of the form

$$\frac{1}{1-\sum_{j=1}^m b_j z^{n_j}} = \frac{1/P(e^{-Ls_0})}{1-e^{Ls_0}z} + \frac{Q(z)}{P(z)}.$$

Therefore, to study (33) with $C(s)$ as in (33) it is natural in this context to consider Mellin-Perron integrals of the form

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT} \frac{B(s)}{1-e^{-Ls}\lambda}\,\frac{x^s}{s}\,ds$$

for some complex number $\lambda \ne 0$ and a Dirichlet series $B(s)$. The corresponding result is stated below in Theorem 6.

To derive asymptotics of the above integral (cf. Theorem 6 below) we need the following two lemmas. The first lemma (Lemma 2) is also the basis of the proof of the Mellin-Perron formula (cf. [Apostol 1976; Szpankowski 2001]). For the reader's convenience we provide a short proof of Lemma 2.

Lemma 2. *Suppose that a and c are positive real numbers. Then*

$$\left|\frac{1}{2\pi i}\int_{c-iT}^{c+iT} a^s\frac{ds}{s} - 1\right| \le \frac{a^c}{\pi T\log a} \qquad (a>1),$$

$$\left|\frac{1}{2\pi i}\int_{c-iT}^{c+iT} a^s\frac{ds}{s}\right| \le \frac{a^c}{\pi T\log(1/a)} \qquad (0<a<1),$$

$$\left|\frac{1}{2\pi i}\int_{c-iT}^{c+iT} a^s\frac{ds}{s} - \frac{1}{2}\right| \le \frac{C}{T} \qquad (a=1).$$

Proof. Suppose first that $a > 1$. By considering the contour integral of the function $F(s) = a^s/s$ around the rectangle with vertices $-A-iT, c-iT, c+iT, -A+iT$ and letting $A \to \infty$ one directly obtains the representation

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT} a^s\frac{ds}{s} = \mathrm{Res}(a^s/s; s=0) + \frac{1}{2\pi i}\int_{-\infty}^{c}\frac{a^{x+iT}}{x+iT}\,dx + \frac{1}{2\pi i}\int_{-\infty}^{c}\frac{a^{x-iT}}{x-iT}\,dx.$$

Since

$$\left|\frac{1}{2\pi i}\int_{-\infty}^{c}\frac{a^{x\pm iT}}{x\pm iT}\,dx\right| \le \frac{a^c}{\pi T\log a}$$

we directly obtain the bound in the case $a > 1$.

The case $0 < a < 1$ can be handled in the same way. And finally, in the case $a = 1$ the integral can be explicitly calculated (and estimated).

Lemma 3. *Suppose that L is a positive real number, λ a complex number different from 0 and 1, and c a real number with $c > \frac{1}{L}\log|\lambda|$. Then we have for all real numbers $x > 1$*

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{1}{1-e^{-Ls}\lambda}\cdot\frac{x^s}{s}\,ds = \frac{\lambda^{\left\lfloor\frac{\log x}{L}\right\rfloor+1}-1}{\lambda-1} - \frac{1}{2}\lambda^{\left\lfloor\frac{\log x}{L}\right\rfloor}[\![\log x/L\in\mathbb{Z}]\!]. \tag{36}$$

Proof. By assumption we have $|\lambda e^{-Ls}| < 1$. Thus, by using a geometric series expansion we get for all $x > 1$ such that $\log x/L$ is not an integer

$$\begin{aligned}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{1}{1-e^{-Ls}\lambda}\cdot\frac{x^s}{s}\,ds &= \sum_{k\ge 0}\lambda^k\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\left(\frac{x}{e^{Lk}}\right)^s\frac{ds}{s}\\ &= \sum_{k\le\frac{\log x}{L}}\lambda^k + O\left(\frac{1}{T}\sum_{k\ge 0}\frac{|\lambda|^k\left(\frac{x}{e^{Lk}}\right)^c}{\left|\log\left(\frac{x}{e^{Lk}}\right)\right|}\right)\\ &= \frac{\lambda^{\left\lfloor\frac{\log x}{L}\right\rfloor+1}-1}{\lambda-1} + O\left(\frac{1}{T}\frac{x^c}{1-\frac{1}{e^{Lc}}|\lambda|}\right).\end{aligned}$$

In the second line above we use the first part of Lemma 2 replacing the integral by 1 plus the error term. Similarly we can proceed if $\log x/L$ is an integer which implies (36).

Theorem 6. *Let L be a positive real number, λ be a non-zero complex number, and suppose that*

$$B(s) = \sum_{n\ge 1}B_n n^{-s}$$

is a Dirichlet series that is absolutely convergent for $\Re(s) > \frac{1}{L}\log|\lambda| - \eta$ for some $\eta > 0$. Then

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{B(s)}{1-e^{-Ls}\lambda}\frac{x^s}{s}\,ds = \sum_{n\ge 1}B_n\frac{\lambda^{\left\lfloor\frac{\log(x/n)}{L}\right\rfloor+1}}{\lambda-1} - \frac{1}{2}\sum_{n\ge 1}B_n\lambda^{\left\lfloor\frac{\log(x/n)}{L}\right\rfloor}[\![\log(x/n)/L\in\mathbb{Z}]\!] \tag{37}$$

$$+\,O\left(x^{\frac{1}{L}\log|\lambda|-\eta}\right).$$

if $|\lambda| > 1$, *and*

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{B(s)}{1-e^{-Ls}}\frac{x^s}{s}\,ds = \sum_{n\ge 1} B_n\left(\left\lfloor\frac{\log(x/n)}{L}\right\rfloor+1\right) \qquad (38)$$
$$-\frac{1}{2}\sum_{n\ge 1} B_n [\![\log(x/n)/L\in\mathbb{Z}]\!] + O\left(x^{-\eta}\right).$$

if $\lambda = 1$.

Proof. We split the integral into an infinite sum of integrals according to the series $B(s)=\sum_{n\ge 1} B_n n^{-s}$ and apply (36) for each term by replacing x by x/n.

First assume that $\log(x/n)/L$ is not an integer for $n\ge 1$. Hence, if $x > ne^{Lk}$, then we have

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{n^{-s}}{1-e^{-Ls}\lambda}\frac{x^s}{s}\,ds = \frac{\lambda^{\left\lfloor\frac{\log(x/n)}{L}\right\rfloor+1}-1}{\lambda-1}+$$
$$O\left(\frac{1}{T}\sum_{k\ge 0}\frac{|\lambda|^k\left(\frac{x}{e^{Lk}n}\right)^c}{\left|\log\left(\frac{x}{e^{Lk}n}\right)\right|}\right),$$

and if $x < ne^{Lk}$, then we just have

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{n^{-s}}{1-e^{-Ls}\lambda}\frac{x^s}{s}\,ds = O\left(\frac{1}{T}\sum_{k\ge 0}\frac{|\lambda|^k\left(\frac{x}{e^{Lk}n}\right)^c}{\left|\log\left(\frac{x}{e^{Lk}n}\right)\right|}\right).$$

Further, for given x there are only finitely many pairs (k,n) with

$$\left|\frac{x}{e^{Lk}n}-1\right| < \frac{1}{2}.$$

Hence, the series

$$\sum_{n\ge 1}\sum_{k\ge 0} B_n\frac{|\lambda|^k\left(\frac{x}{e^{Lk}n}\right)^c}{\left|\log\left(\frac{x}{e^{Lk}n}\right)\right|}$$

is convergent. Consequently, we find

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{\sum_{n\ge 1} B_n n^{-s}}{1-e^{-Ls}\lambda}\frac{x^s}{s}\,ds = \frac{1}{\lambda-1}\sum_{n<x} B_n\left(\lambda^{\left\lfloor\frac{\log(x/n)}{L}\right\rfloor+1}-1\right)+O(1)$$

(and a similar expression if there are integers $n\ge 1$ for which $\log(x/n)/L$ is an integer). Finally, since

$$\sum_{n<x} B_n = O\left(n^{\frac{1}{L}\log|\lambda|-\eta}\right)$$

and

$$\sum_{n>x} B_n n^{-\frac{1}{L}\log|\lambda|} = O\left(x^{-\eta}\right)$$

it follows that

$$\frac{1}{\lambda-1}\sum_{n<x} B_n\left(\lambda^{\left\lfloor\frac{\log(x/n)}{L}\right\rfloor+1}-1\right)=\frac{1}{\lambda-1}\sum_{n\ge 1} B_n\left(\lambda^{\left\lfloor\frac{\log(x/n)}{L}\right\rfloor+1}\right)+O\left(n^{\frac{1}{L}\log|\lambda|-\eta}\right)$$

(and similarly if there are integers $n \ge 1$ for which $\log(x/n)/L$ is an integer). This proves (40).

If $\lambda = 1$ we first observe that (36) rewrites to

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{1}{1-e^{-Ls}}\frac{x^s}{s}\,ds=\left\lfloor\frac{\log x}{L}\right\rfloor+1-\frac{1}{2}[\![\log x/L\in\mathbb{Z}]\!].$$

Now the proof of (38) is very similar to that of (40).

Remark 10. The representations (37) and (38) have nice interpretations. When $|\lambda| > 1$ set

$$\Psi(t)=\lambda^{-t/L}\sum_{n\ge 1} B_n\frac{\lambda^{\left\lfloor\frac{t-\log n}{L}\right\rfloor+1}}{\lambda-1}-\frac{\lambda^{-t/L}}{2}\sum_{n\ge 1} B_n\lambda^{\left\lfloor\frac{t-\log n}{L}\right\rfloor}[\![(t-\log n)/L\in\mathbb{Z}]\!]. \quad (39)$$

Then $\Psi(t)$ is a periodic function of bounded variation with period L, that has (usually) countably many discontinuities for $t = L\{\log n/L\}$, $n \ge 1$ (where we recall $\{x\} = x - \lfloor x\rfloor$ is the fractional part of x). We arrive at

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{B(s)}{1-e^{-Ls}\lambda}\frac{x^s}{s}\,ds=x^{\frac{1}{L}\log\lambda}\,\Psi(\log x)+O\left(x^{\frac{1}{L}\log|\lambda|-\eta}\right).$$

Formally, this representation also follows by adding the residues of

$$B(s)/(1-e^{-Ls}\lambda)$$

at $s = s_0 + 2k\pi i/L$ ($k \in \mathbb{Z}$) which are the zeros of $1 - e^{Ls}\lambda = 0$. This means the leading asymptotic follows (in this case) from a *formal residue calculus*.

Furthermore, if we go back to the original problem, where we have to discuss a function of the form

$$\frac{B(s)}{1-\sum_{j=1}^m b_j p_j^s},$$

for $\log p_j$ rationally related, then we have

$$\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{B(s)}{1-\sum_{j=1}^m b_j p_j^s}\frac{x^s}{s}\,ds=x^{s_0}\,\Psi(\log x)+O\left(x^{s_0-\eta}\right).$$

As mentioned above we split up the integral with the help of a partial fraction decomposition of the rational function

$$\frac{1}{1-\sum_{j=1}^m b_j z^{n_j}}.$$

The *leading term* can be handled directly with the help of Theorem 3. The remaining terms can use again (36) and obtains (finally) a second error term of order $O\left(x^{s_0-\eta}\right)$.

Remark 11. If $\lambda = 1$ then the situation is even simpler. Set

$$C = \frac{1}{L} \sum_{n \ge 1} B_n$$

and

$$\tilde{\Psi}(t) = \sum_{n \ge 1} B_n \left(-\left\{ \frac{t - \log n}{L} \right\} + 1 \right) - \frac{1}{2} \sum_{n \ge 1} B_n [\![(t - \log n)/L \in \mathbb{Z}]\!] - \frac{1}{L} \sum_{n \ge 1} B_n \log n.$$

Then $\tilde{\Psi}(t)$ is a periodic function with period L, that has (usually) countably many discontinuities for $t = L\{\log n/L\}$, $n \ge 1$, and we have

$$\frac{1}{2\pi i} \lim_{T \to \infty} \int_{c-iT}^{c+iT} \frac{B(s)}{1 - e^{-Ls}} \frac{x^s}{s}\, ds = C \log x + \tilde{\Psi}\left(\log x\right) + O\left(x^{-\eta}\right).$$

Hence, by applying the same partial fraction decomposition as above we also obtain (if $s_0 = 0$ and if the $\log p_j$ are rationally related)

$$\frac{1}{2\pi i} \lim_{T \to \infty} \int_{c-iT}^{c+iT} \frac{B(s)}{1 - \sum_{j=1}^{m} b_j p_j^s} \frac{x^s}{s}\, ds = C \log x + \tilde{\Psi}\left(\log x\right) + O\left(x^{-\eta}\right).$$

We now can go back to our analysis, and recall that $\widetilde{T}(s)$ is given by (32). Hence Theorem 6 covers (together with the subsequent Remarks 10 and 11) the part $B(s) = \tilde{A}(s) + \tilde{G}(s)$, since they represent absolute convergent Dirichlet series. The remaining function $\tilde{E}(s)$ is actually much more difficult to handle. To simplify our presentation, we will only discuss the case $s_0 > 0$ that corresponds to $\lambda > 1$, the case $s_0 = 0$ can be handled in a similar way, see again Remark 11.

First of all, for function of the form

$$\overline{B}(s) = \sum_{n \ge 1} \overline{B}_n \left(\frac{1}{(n/p)^s} - \frac{1}{h_n^s} \right),$$

where $\overline{B}_n \ge 0$ and h_n are integers satisfying $h_n = n/p + O(n^{1-\delta})$, we find the representation

$$\begin{aligned} \frac{1}{2\pi i} \lim_{T \to \infty} \int_{c-iT}^{c+iT} \frac{\overline{B}(s)}{1 - e^{-Ls}\lambda} \frac{x^s}{s}\, ds \;=\; & \frac{1}{1-\lambda^{-1}} \sum_{n < px} \overline{B}_n \lambda^{\left\lfloor \frac{\log(px/n)}{L} \right\rfloor} - \frac{1}{1-\lambda^{-1}} \sum_{h_n < x} \overline{B}_n \lambda^{\left\lfloor \frac{\log(x/h_n)}{L} \right\rfloor} \\ & + \frac{1}{2} \sum_{n \le px} \overline{B}_n \lambda^{\left\lfloor \frac{\log(px/n)}{L} \right\rfloor + 1} [\![\log(px/n)/L \in \mathbb{Z}]\!] \\ & - \frac{1}{2} \sum_{h_n \le x} \overline{B}_n \lambda^{\left\lfloor \frac{\log(x/h_n)}{L} \right\rfloor + 1} [\![\log(x/h_n)/L \in \mathbb{Z}]\!]. \end{aligned} \tag{40}$$

Set

$$\begin{aligned} D_n(x) := \frac{1}{1-\lambda^{-1}} \overline{B}_n \left(\lambda^{\left\lfloor \frac{\log(px/n)}{L} \right\rfloor} - \lambda^{\left\lfloor \frac{\log(x/h_n)}{L} \right\rfloor} \right) & + \frac{1}{2} \overline{B}_n \lambda^{\left\lfloor \frac{\log(px/n)}{L} \right\rfloor + 1} [\![\log(px/n)/L \in \mathbb{Z}]\!] \\ & - \frac{1}{2} \overline{B}_n \lambda^{\left\lfloor \frac{\log(x/h_n)}{L} \right\rfloor + 1} [\![\log(x/h_n)/L \in \mathbb{Z}]\!]. \end{aligned}$$

We next show that (in our context) the right hand side of (40) can be approximated by the infinite sum together with an error term:

$$\sum_{n\ge 1} D_n(x) + O\left(x^{s_0-\eta'}\right)$$

for some $\eta' > 0$.

Recall that we have $\overline{B}_n = T(n+2) - T(n+1)$ and $\lambda = e^{Ls_0}$ and by the small growth assumption $\overline{B}_n = O(n^{s_0-\eta})$ for some $\eta > 1-\delta$. Hence we have

$$\sum_{n:\ px<n<px+O(x^{1-\delta})} \overline{B}_n \lambda^{\left\lfloor \frac{\log(x/h_n)}{L} \right\rfloor} = O\left(x^{s_0-\eta} x^{1-\delta}\right) = O\left(x^{s_0-(\eta+\delta-1)}\right).$$

Next observe that $D_n(x)$ can only be non-zero if there is an integer m such that $\log(px/n) \le Lm \le \log(x/h_n)$ (or the other way round). Since we are only interested in those n with $n \ge px$ this means that $m \le 0$. Now fix an integer $m \le 0$ and let I_m be the set of integers n with the property that $D_n(x) \ne 0$. It is clear that all integers $n \in I_m$ have to satisfy $n \sim pxe^{-mL}$, and since $h_n = n/p + O(n^{1-\delta})$ the cardinality of I_m is bounded by $O\left((pxe^{-mL})^{1-\delta}\right)$. Consequently it follows that

$$\begin{aligned}
\sum_{n\ge px} D_n(x) &= O\left(\sum_{m\le 0} \left(pxe^{-mL}\right)^{s_0-\eta} \left(pxe^{-mL}\right)^{1-\delta} \lambda^m\right) \\
&= O\left(x^{s_0-(\eta+\delta-1)} \sum_{m\le 0} \lambda^{m\frac{\eta+\delta-1}{s_0}} O\left(x^{s_0-(\eta+\delta-1)}\right)\right).
\end{aligned}$$

Now if we define a periodic function $\overline{\Psi}(t)$ by

$$\begin{aligned}
\overline{\Psi}(t) = {}& \frac{\lambda^{-t/L}}{1-\lambda^{-1}} \sum_{n\ge 1} \overline{B}_n \left(\lambda^{\left\lfloor \frac{t-\log(n/p)}{L}\right\rfloor} - \lambda^{\left\lfloor \frac{t-\log h_n}{L}\right\rfloor}\right) - \frac{\lambda^{-t/L}}{2} \sum_{n\ge 1} \overline{B}_n \lambda^{\left\lfloor \frac{t-\log(n/p)}{L}\right\rfloor} [\![(t-\log(n/p))/L \in \mathbb{Z}]\!] \\
&+ \frac{\lambda^{-t/L}}{2} \sum_{n\ge 1} \overline{B}_n \lambda^{\left\lfloor \frac{t-\log h_n}{L}\right\rfloor} [\![(t-\log h_n)/L \in \mathbb{Z}]\!], \qquad (41)
\end{aligned}$$

then

$$\frac{1}{2\pi i} \lim_{T\to\infty} \int_{c-iT}^{c+iT} \frac{\overline{B}(s)}{1-e^{-Ls}\lambda} \frac{x^s}{s}\, ds = x^{\frac{1}{L}\log\lambda}\, \overline{\Psi}\left(\log x\right) + O\left(x^{\frac{1}{L}\log|\lambda| - \eta'}\right).$$

Summing up, we can handle all parts of $\widetilde{T}(s)$ (given by (32)) with the help of these techniques. If $s_0 > \sigma_a$ and $s_0 > 0$, we obtain

$$\frac{1}{2\pi i} \lim_{T\to\infty} \int_{c-iT}^{c+iT} \widetilde{T}(s) \frac{x^s}{s}\, ds = x^{s_0} \Psi(\log x) + O\left(x^{s_0-\eta'}\right), \qquad (42)$$

where $\eta' > 0$ and $\Psi(t)$ is a positive and periodic function (with period L) that has building blocks of the forms (39) and (41).

If we now set $x = n$ then we obtain

$$\frac{T(n+1)+T(n+2)}{2} = n^{s_0} \Psi(\log n) + O\left(n^{s_0-\eta'}\right).$$

However, by applying the small growth property (22) this also implies that

$$T(n) = n^{s_0}\Psi(\log n) + O\left(n^{s_0-\eta'}\right). \tag{43}$$

Remark 12. We note that the small growth property (22) was essential in the proof. If (22) is not satisfied, then it is not clear that the sum $\sum_n D_n(x)$ converges. Actually, the *counter examples* from Example 3 show that it might diverge (although the partial sums are bounded).

The final goal is to present some properties of the periodic function $\Psi(t)$.

Lemma 4. *Suppose that $s_0 > \sigma_a$, $s_0 > 0$, that we are in the rationally related case, and that the small growth property (22) is satisfied. Then the periodic function $\Psi(t)$ of (42) is continuous and of bounded variation. It is has a convergent Fourier series $\Psi(t) = \sum_k c_k e^{2k\pi i x/L}$, where*

$$c_k = \frac{\tilde{A}(s_k) + \sum_{j=1}^m b_j(G_j(s_k) - E_j(s_k)) + \sum_{j=1}^m \overline{b}_j(\overline{G}_j(s_k) - \overline{E}_j(s_k))}{s_k \sum_{j=1}^m (b_j + \overline{b}_j) p_j^{s_0}\log(1/p_j)},$$

with $s_k = s_0 + 2k\pi i/L$, as presented in (8).

Proof. By definition the function $\Psi(t)$ is a finite sum of absolutely (and locally uniformly) convergent series of functions $\Psi_k(t)$ of bounded variation, for example

$$\Psi_k(t) = C\lambda^{-t/L}(a_{k+1}-a_k)\left(\lambda^{\lfloor (t-\log(n/p))/L\rfloor} - \frac{1}{2}\lambda^{\lfloor (t-\log(n/p))/L\rfloor}[\![(t-\log k)/L \in \mathbb{Z}]\!]\right)$$

see (39) and (41). Consequently $\Psi(t)$ is of bounded variation, too. Furthermore since functions $\Psi_k(t)$ of these series have the property that the limits $\Psi_k(t+0) = \lim_{u\to 0+}\Psi_k(t+u)$ and $\Psi_k(t-0) = \lim_{u\to 0+}\Psi_k(t-u)$ exist and satisfy $\Psi_k(t) = \frac{1}{2}\left(\Psi_k(t+0) + \Psi_k(t-0)\right)$ for all t, the same property holds for $\Psi(t)$. Consequently, in order to prove continuity we only have to check that $\Psi(t+0) = \Psi(t-0)$ for all potential discontinuities of $\Psi(t)$.

Again by definition the only potential discontinuities of $\Psi(t)$ are those t for which there exists an integer m with $t \equiv \log m \bmod L$. (Recall that $p_j = e^{-Ln_j}$ so that $\log(k/p_j)/L = \log(k)/L + n_i$. Furthermore $n_j(k)$ is always an integer.) For all other t all functions $\Psi_k(t)$ are continuous for all k so that $\Psi(t)$ is continuous, too.

We now compare $T(\lfloor me^{Lr} - 1\rfloor)$ and $T(\lfloor me^{Lr}+2\rfloor)$ for a fixed integer m and for integers $r\to\infty$. By (43) if follows that

$$\begin{aligned} T(\lfloor me^{Lr}-1\rfloor) &= \lfloor me^{Lr}-1\rfloor^{s_0}\Psi\left(\log\lfloor me^{Lr}-1\rfloor\right) + O\left((me^{Lr})^{s_0-\eta'}\right) \\ &= m^{s_0}\lambda^r\Psi\left(\log m - c_r' e^{-Lr}/m\right) + O\left(m^{s_0}\lambda^{r(1-\eta'/s_0)}\right), \end{aligned}$$

where c_r' is a sequence of positive numbers that are bounded by $1 \le c_r' \le 2$ (at least for sufficiently large r). Similarly we find

$$T(\lfloor me^{Lr}+2\rfloor) = m^{s_0}\lambda^r\Psi\left(\log m + c_r'' e^{-Lr}/m\right) + O\left(m^{s_0}\lambda^{r(1-\eta'/s_0)}\right),$$

for a corresponding sequence of positive numbers c_r'' that are bounded by $1 \le c_r'' \le 3$ (for sufficiently large r). Finally be the small growth property (22) we have

$$T(\lfloor me^{Lr}+2\rfloor) - T(\lfloor me^{Lr}-1\rfloor) = O\left(m^{s_0}\lambda^{r(1-\eta/s_0)}\right).$$

Putting these things together it follow that

$$\lim_{r\to\infty} \left(\Psi\left(\log m + c_r'' e^{-Lr}/m\right) - \Psi\left(\log m - c_r' e^{-Lr}/m\right)\right) = 0$$

which implies $\Psi(t+0) = \Psi(t-0)$ for t that satisfy $t \equiv \log m \bmod L$. As mentioned above this implies continuity of $\Psi(t)$.

In order to obtain the Fourier series expansion of $\Psi(t)$ we just have to check the corresponding ones for the summands $\Psi_k(t)$ (that are immediate) and sum over k. It is also easy to check that the Fourier coefficients coincide with (8). However, since all functions $\Psi_k(t)$ are discontinuous the corresponding Fourier series are not absolutely convergent. Therefore it cannot be deduced (in this way) that the Fourier series of $\Psi(t)$ is absolutely convergent.

We recall that the case $s_0 = 0$ (that corresponds to $\lambda = 1$) can be handled in a similar way.

5.5 Finishing the Proof

It remains to complete the proof of Theorem 2 in the *rationally related* case. Actually we only have to (re)consider the cases, where $s_0 \ge \sigma_a$. Namely, if $\sigma_a > s_0$, then the zeros of the equation (15) do not contribute to the leading analytic behavior of $\widetilde{T}(s)$ and we can apply proper Tauberian theorems. In what follows we comment on the differences in the cases of interest.

2. $\sigma_a < s_0$ and $s_0 = 0$:
   This case is basically handled in Theorem 6, in particular see Remark 11. As mentioned above we also have to adapt the considerations following Remark 11 to the case $s_0 = 0$ (which is immediate). There is also a proper variant of Lemma 4 that ensures continuity of $\Psi_2(t)$.
3. $\sigma_a = s_0 = 0$:
   In this case we apply proper generalizations of Tauberian theorems. Recall that in this case the dominant singular term of $\frac{1}{s}\widetilde{T}(s)$ is given by

$$C\,\frac{(\log(1/s))^{\beta_0}}{s^{\alpha_0+2}}$$

   and there are infinitely many simple poles at $s = 2\pi i k/L$ ($k \in \mathbb{Z} \setminus \{0\}$). Of course we have $\alpha_0 \ge 0$, otherwise the sequence a_n would not be non-decreasing. Here we need a slightly modified version of Theorem 12 or Theorem 13, resp., that can be found in Appendix B. Here the proof requires that the Fourier series (20) converges and represents an integrable function, see also Remark 13. However, this property is only required in the proof, the asymptotic leading term does not depend on the Fourier series (20).
4. $\sigma_a < s_0$ and $s_0 > 0$:
   Here we use Theorem 6, Remark 10 together with the discussion preceding Lemma 4, in particular we apply (43). Finally Lemma 4 implies that $\Psi_4(t)$ is continuous.
5. $\sigma_a = s_0$ and $s_0 > 0$:
   This case is very similar to Case 3.

6. EXTENSION: CENTRAL LIMIT LAW FOR BONCELET'S ALGORITHMS

So far we mostly dealt with divide and conquer recurrences representing a parameter of interest, e.g., the average code length. However, in Section 4 we encounter a simple divide and conquer functional recurrence (23) for the underlying generating function $C(n, y)$. This recurrence can be solved for y in a compact set and — under some additional assumptions regarding the growth of the mean and the variance – we can establish a general central limit law. In this section we do exactly this for the Boncelet algorithm, as an illustration of the power of our technique.

We prove the following result.

Theorem 7. *Consider a biased memoryless source (i.e., $p \neq q$) generating a sequence of length n parsed by the Boncelet algorithm. The phrase length D_n satisfies the central limit law, that is,*

$$\frac{D_n - \frac{1}{H}\log n}{\sqrt{\left(\frac{H_2}{H^3} - \frac{1}{H}\right)\log n}} \to N(0,1),$$

where $N(0,1)$ denotes the standard normal distribution, and

$$\mathbb{E}[D_n] = \frac{\log n}{H} + O(1), \qquad \mathbb{V}\mathrm{ar}[D_n] \sim \left(\frac{H_2}{H^3} - \frac{1}{H}\right)\log n$$

for $n \to \infty$.

Proof. We recall that $C(n, y)$ satisfies the recurrence (23) with initial conditions $C(0, y) = 0$ and $C(1, y) = 1$. It is clear that for every fixed positive real number y we can apply Theorem 2. However, we have to be careful since we need an asymptotic representation for $C(n, y)$ uniformly for y in an interval that contains 1 in its interior. Note that $C(n, 1) = 1$.

To deal with $C(n, y)$, one has to consider the Dirichlet series

$$C(s, y) = \sum_{n=1}^{\infty} \frac{C(n+2, y) - C(n+1, y)}{n^s}.$$

For simplicity we just consider here the case $y > 1$. (The case $y \le 1$ can be handled in a similar way.) Then $C(s, y)$ converges for $\Re(s) > s_0(y)$, where $s_0(y)$ denotes the real zero of the equation $y(p^{s+1} + q^{s+1}) = 1$. We find

$$C(s, y) = \frac{(y-1) - \tilde{E}(s, y)}{1 - y(p^{s+1} + q^{s+1})},$$

where

$$\tilde{E}(s, y) = py \sum_{k=1}^{\infty} (C(k+2, y)) - C(k+1, y)) \left(\frac{1}{(k/p)^s} - \frac{1}{\left(\left\lceil \frac{k+2-\delta}{p} \right\rceil - 2\right)^s} \right)$$
$$+ qy \sum_{k=1}^{\infty} (C(k+2, y)) - C(k+1, y)) \left(\frac{1}{(k/q)^s} - \frac{1}{\left(\left\lfloor \frac{k+1+\delta}{q} \right\rfloor - 1\right)^s} \right)$$

converges for $\Re(s) > s_0(y) - 1$ and satisfies $\tilde{E}(0, y) = 0$ and $\tilde{E}(s, 1) = 0$.

Suppose first that we are in the irrational case. Then by the Wiener-Ikehara theorem only the residue at $s_0(y)$ contributes to the main asymptotic leading term. (Recall that we consider the case $y > 1$). We thus have

$$C(n,y) \sim \mathrm{Res}\left(\frac{((y-1) - \tilde{E}(s,y))(n-3/2)^s}{s(1-y(p^{s+1}+q^{s+1}))}; s = s_0(y) \right)$$
$$= \frac{((y-1) - \tilde{E}(s_0(y),y))(n-3/2)^{s_0(y)}}{-s_0(y)(\log(p)p^{s_0(y)+1} + \log(q)q^{s_0(y)+1}))}(1+o(1)).$$

The essential but non-trivial observation is that this asymptotic relation holds uniform for y in an interval around 1. In order to make this precise we can use the Mellin-Perron formula from Theorem 4

$$C(n,y) = C(2,y) + \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} C(s,y)\frac{\left(n-\frac{3}{2}\right)^s}{s}\,ds$$

and apply the methods presented in the proof of Theorem 5, compare also with [Drmota et al. 2010]. We observe that the sum of residues (that is denoted by $Q(x)$ in the proof of Theorem 5) converges to 0 uniformly in y. This follows from the fact that the zeros of the equation $y(p^{s+1}+q^{s+1}) = 1$ vary continuously in y. Hence, if y in contained in some (compact) interval Y and $s_{\mathrm{nr}}(y)$ denotes one of the non-real zeros, then

$$\min_{y\in Y} \Re(s_{\mathrm{nr}}(y) - s_0(y)) > 0.$$

Hence we find

$$C(n,y) = (1+O(y-1))n^{s_0(y)}(1+o(1))$$

uniformly for real y that are contained in an interval around 1. Finally by using the local expansion

$$s_0(y) = \frac{y-1}{H} + \left(\frac{H_2}{2H^3} - \frac{1}{H}\right)(y-1)^2 + O((y-1)^3), \tag{44}$$

and by setting $y = e^{t/(\log n)^{1/2}}$ we obtain

$$n^{s_0(y)} = \exp\left(\log n\left(\frac{y-1}{H} - \left(\frac{1}{H} - \frac{H_2}{2H^3}\right)(y-1)^2 + O(|z-1|^3)\right)\right)$$
$$= \exp\left(\frac{1}{H}t\sqrt{\log n} + \frac{1}{H}\frac{t^2}{2} - \left(\frac{1}{H} - \frac{H_2}{2H^3}\right)t^2 + O(t^3/\sqrt{\log n})\right)$$
$$= \exp\left(\frac{1}{H}t\sqrt{\log n} + \left(\frac{H_2}{H^3} - \frac{1}{H}\right)\frac{t^2}{2} + O(t^3/\sqrt{\log n})\right),$$

and consequently

$$\mathbb{E}\left[e^{D_n t/\sqrt{\log n}}\right] = C\left(n, e^{t/\sqrt{\log n}}\right) = \exp\left(\frac{1}{H}t\sqrt{\log n} + \left(\frac{H_2}{H^3} - \frac{1}{H}\right)\frac{t^2}{2}\right)(1+o(1)).$$

Hence, we arrive at

$$\mathbb{E}\left[e^{t(D_n - \frac{1}{H}\log n)/\sqrt{\log n}}\right] = e^{-(t/H)\sqrt{\log v}}\mathbb{E}\left[e^{D_n t/\sqrt{\log n}}\right] \tag{45}$$
$$= e^{\frac{t^2}{2}\left(\frac{H_2}{H^3} - \frac{1}{H}\right)} + o(1).$$

By the convergence theorem for the Laplace transform (see [Szpankowski 2001]) this proves the normal limiting distribution as $n \to \infty$ and also convergence of (centralized) moments.

In the rational case we can use a similar procedure. However, we have to use a proper variation of the proof of Theorem 3, from which we obtain estimates that are uniform in (real) y. Formally, we just have to add the residues coming from the zeros $s_k(y) = s_0(y) + k2\pi i/L$ for $k \neq 0$ (where $L > 0$ is the largest real number such that $\log p$ and $\log q$ are integer multiples of L). These terms lead to an additional contribution of the form

$$\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{((y-1) - \tilde{E}(s_k(y), y))(n-3/2)^{s_k(y)}}{-s_k(y)(\log(p)p^{s_k(y)+1} + \log(q)q^{s_k(y)+1}))} = O(|y-1|n^{s_0(y)}).$$

Since $(y-1) - \tilde{E}(s_k(y), y) = O(|y-1|)$, this suggest that these additional terms are bounded by $O(|y-1|n^{s_0(y)})$. Actually this can be checked rigorously by adapting the methods from Section 5. Hence, if we set $y = e^{t/(\log n)^{1/2}}$ this term is asymptotically negligible and the central limit theorem follows also in the rational case.

In passing we observe that the phrase length D_n follows the same central limit law as the Tunstall algorithm [Drmota et al. 2010].

Appendix

A. ANALYTIC CONTINUATION OF DIRICHLET SERIES

Dirichlet series of special sequences are frequently used in the present paper. In particular we are interested in the Dirichlet series of sequences of the form

$$c(n) = n^\sigma (\log n)^\alpha.$$

It is clear that the Dirichlet series $C(s) = \sum_{n\ge1} c(n)n^{-s}$ converges (absolutely) for complex s with $\Re(s) > \sigma + 1$. We also know that the abscissa of absolute convergence is given by $\sigma_a = \sigma + 1$. However, it is not immediate that $C(s)$ has a certain analytic continuation to a larger region (that does not contain the singularity $s = \sigma_a$. Nevertheless, such continuation properties do hold (see [Grabner and Thuswaldner 1996]).

Theorem 8. *Suppose that σ and α are real numbers and let $C(s)$ be the Dirichlet series*

$$C(s) = \sum_{n\ge2} n^\sigma (\log n)^\alpha n^{-s}.$$

(i) *If α is not a negative integer, then $C(s)$ can be represented as*

$$C(s) = \frac{\Gamma(\alpha+1)}{(s-\sigma-1)^{\alpha+1}} + G(s),$$

where $G(s)$ is an entire function.

(ii) *If $\alpha = -k$ is a negative integer, then we have*

$$C(s) = \frac{(-1)^k}{(k-1)!}(s-\sigma-1)^{k-1}\log(s-\sigma+1) + G(s),$$

where $G(s)$ is an entire function.

Proof. We do not provide a full proof but sketch the arguments from [Grabner and Thuswaldner 1996] where even a slightly more general situation was considered. Furthermore it is sufficient to consider the case $\sigma = 0$.

First it follows form the Euler Maclaurin summation that $C(s)$ can be represented (for $\Re(s) > 1$) as

$$C(s) = \int_2^\infty \frac{(\log v)^\alpha}{v^s}\,dv + \frac{(\log 2)^\alpha}{2^{s+1}} + \int_2^\infty \left(\{v\} - \frac{1}{2}\right)\left(\alpha(\log v)^{\alpha-1} - s(\log v)^\alpha\right) v^{-s-1}\,dv,$$

where the second integral on the right hand side represents a function that is analytic for $\Re(s) > 0$. Furthermore, by using the substitution $z = (s-1)\log v$ the first integral can be rewritten as

$$\int_2^\infty \frac{(\log v)^\alpha}{v^s}\,dv = (s-1)^{-\alpha-1}\int_{(s-1)\log 2}^\infty z^\alpha e^{-z}\,dz.$$

The latter integral is precisely the incomplete Γ-function.

If α is not a negative integer, then [Abramowitz and Stegun 1964]

$$\int_w^\infty z^\alpha e^{-z}\,dz = \Gamma(\alpha+1) - w^{\alpha+1}\sum_{m=0}^\infty \frac{(-1)^m}{m!}\frac{w^m}{(m+\alpha+1)}$$

and if $\alpha = -k$ is a negative integer, then [Abramowitz and Stegun 1964]

$$\int_w^\infty z^{-k} e^{-z}\,dz = \Gamma_{k-1}(-k+1) + \frac{(-1)^k}{(k-1)!}\log(w) - w^{\alpha+1}\sum_{m=0,\, m\ne k-1}^\infty \frac{(-1)^m}{m!}\frac{w^m}{(m+\alpha+1)},$$

where $\Gamma_k(z) = \Gamma(z) - (-1)^k/(k!(k+z))$. Hence the conclusion follows.

Note that the above method is quite flexible. For example, if

$$c(n) = n^\sigma(\log n)^\alpha + O(n^{\sigma-\delta})$$

for some $\delta > 0$, then we obtain a similar representation except that $G(s)$ is not any more an entire function but a function that is analytic for $\Re(s) > \sigma + 1 - \delta$.

It is now easy to apply Theorem 8 to sequences of the form

$$c(n) = a_{n+2} - a_{n+1},$$

where

$$a_n = n^a(\log n)^b.$$

Theorem 9. *Suppose that $a_n = n^a(\log n)^b$, where a and b are real numbers, and let $\widetilde{A}(s)$ be the Dirichlet series*

$$\widetilde{A}(s) = \sum_{n\ge 1} \frac{a_{n+2} - a_{n+1}}{n^s}.$$

(i) *If b is not a negative integer, then $\widetilde{A}(s)$ can be represented as*

$$\widetilde{A}(s) = a\frac{\Gamma(b+1)}{(s-a)^{b+1}} + \frac{\Gamma(b+1)}{(s-a)^b} + G(s),$$

where $G(s)$ is analytic for $\Re(s) > a-1$.
(ii) *If $b=-k$ is a negative integer, then we have*

$$\begin{aligned}\widetilde{A}(s) &= a\frac{(-1)^k}{(k-1)!}(s-a)^{k-1}\log(s-a)\\ &\quad + \frac{k(-1)^k}{(k-1)!}(s-a)^k\log(s-a)+G(s),\end{aligned}$$

where $G(s)$ is analytic for $\Re(s) > a-1$.

Proof. This follows from the simple fact that

$$\begin{aligned}a_{n+2}-a_{n+1} &= an^{a-1}(\log n)^b\left(1+O(n^{-1})\right)\\ &\quad + bn^{a-1}(\log n)^{b-1}\left(1+O(n^{-1})\right).\end{aligned}$$

Note that Theorem 9 is even more flexible than Theorem 8. For example, we can also consider sequences of the form $a_n = (\lfloor \rho n+\tau\rfloor)^a$ for some ρ with $0<\rho<1$ (or similarly defined sequences). In this case one could argue, as in Section 5.1, that

$$\widetilde{A}(s) = \rho^s B(s) + R(s),$$

where $B(s)$ is the Dirichlet series of the differences $(n+2)^a-(n+1)^a$ and $R(s)$ is analytic for $\Re(s)>a-1$.

Finally we show that condition (20) of Theorem 2 is satisfied for sequences $a_n = n^a(\log n)^b$.

Theorem 10. *Suppose that $a_n = n^a(\log n)^b$ and let $\widetilde{A}(s)$ denote the corresponding Dirichlet series. Then the Fourier series*

$$\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\widetilde{A}(a+2\pi ik/L)}{a+2\pi ik/L}e^{2\pi ikx/L} \tag{A.1}$$

is convergent for $x\in\mathbb{R}$ and represents an integrable function.

Proof. We restrict ourselves to the case $a=1$, which means that the sequence $a_{n+2}-a_{n+1}$ consists (mainly) of the two terms $(\log n)^b$ and $(\log n)^{b-1}$. To simplify the presentation, we only discuss the function

$$A(s) = \sum_{n\ge 2}(\log n)^b n^{-s}$$

instead of $\widetilde{A}(s)$ (and neglect the error terms, since they be handled easily).

Following the proof of Theorem 8 we have to discuss the three integrals

$$\begin{aligned}A_1(s) &= \int_2^\infty \frac{(\log v)^b}{v^s}\,dv,\\ A_2(s) &= \int_2^\infty \left(\{v\}-\frac12\right) b(\log v)^{b-1}v^{-s-1}\,dv,\\ A_3(s) &= s\int_2^\infty \left(\{v\}-\frac12\right)(\log v)^b v^{-s-1}\,dv.\end{aligned}$$

Let us start with $A_3(s)$ which we represent as

$$A_3(s) = s\int_0^\infty v^{-s}h(v)\,dv,$$

where $h(v) = 0$ for $0 \le v < 2$ and $h(v)/v$ is of bounded variation on $[2,\infty)$. (Note that in our case, $h(v)$ is not continuous if v is an integer.) Set

$$F(x) = L\sum_{m\in\mathbb{Z}} h(e^{x+mL}).$$

Then $F(x)$ is periodic (with period L) and also of bounded variation. Hence it has a convergent Fourier series with Fourier coefficients (see [Korner 1989])

$$\begin{aligned}
f_k &= \frac{1}{L}\int_0^L F(x)e^{-2\pi ikx/L}\,dx = \int_0^L \sum_{m\in\mathbb{Z}} h(e^{x+mL})e^{-(x+mL)2\pi ik/L}\,dx \\
&= \int_{-\infty}^\infty h(e^x)\,e^{-x2\pi ik/L}\,dx = \int_0^\infty h(v)\,v^{-(1+2\pi ik/L)}\,dv \\
&= \frac{A_3(1+2\pi ik/L)}{1+2\pi ik/L}.
\end{aligned}$$

Consequently the Fourier series with Fourier coefficients $A_3(1+2\pi ik/L)/(1+2\pi ik/L)$ is convergent. Furthermore, it is integrable, since the set of discontinuities of $F(x)$ is countable and $F(x)$ equals its Fourier series at all points of continuity (here we use the fact that $f_k = O(1/k)$).

Similarly we can handle $A_2(s)$. We represent it as

$$A_2(s) = \int_2^\infty \overline{h}(v)v^{-s}\,dv,$$

where $\overline{h}(v)/v$ is of bounded variation on $[2,\infty)$. Here the corresponding periodic function is given by

$$\overline{F}(x) = L\int_2^\infty \overline{h}(v)\frac{e^{-L\{(x-\log v)/L\}}}{v(1-e^{-L})}\,dv.$$

Finally, we have to consider $A_1(s)$. By Theorem 8 we know that $A_1(s)$ has an analytic continuation to the slit region $\mathbb{C}\setminus(-\infty,1]$. In particular it follows that the limit

$$\lim_{\varepsilon\to 0+} A_1(1+\varepsilon+2\pi ik/L)$$

exists and equals to (the analytically continued value) $A_1(1+2\pi ik/L)$. By partial integration it follows that (for real t)

$$\int_2^\infty \frac{(\log v)^b}{v^{1+\varepsilon+it}}\,dv = \frac{(\log 2)^b}{\varepsilon+it}2^{-\varepsilon-it} + \frac{b}{\varepsilon+it}\int_2^\infty \frac{(\log v)^{b-1}}{v^{1+\varepsilon+it}}\,dv.$$

This implies that

$$A_1(1+it) = O\left(\frac{1}{t}\right).$$

Consequently the Fourier series with Fourier coefficients $A_1(1+2\pi ik/L)/(1+2\pi ik/L)$, $k\ne 0$, converges absolutely. This completes the proof of the Theorem.

B. TAUBERIAN THEOREMS

The main analytic problem in the present paper is to obtain asymptotic information on the partial sums

$$\overline{c}(v) = \sum_{n \le v} c(n)$$

from analytic properties of the Dirichlet series

$$C(s) = \sum_{n \ge 1} c(n) n^{-s}.$$

The classical Tauberian theorem of Wiener-Ikehara, as presented in Theorem 11, is a very strong tool in this context. Actually it applies to the Mellin-Stieltjes transforms (see [Korevaar 2002]) that is closely related to Dirichlet series:

$$C(s) = \sum_{n \ge 1} c(n) n^{-s} = \int_{1-}^{\infty} v^{-s} \, d\overline{c}(v).$$

Theorem 11 (Wiener-Ikehara; cf. [Korevaar 2002])**.** *Let $\overline{c}(v)$ be non-negative and non-decreasing on $[1,\infty)$ such that the Mellin-Stieltjes transform*

$$C(s) = \int_{1-}^{\infty} v^{-s} \, d\overline{c}(v) = s \int_{1}^{\infty} \overline{c}(v) v^{-s-1} \, dv$$

exists for $\Re(s) > 1$. Suppose that for some constant $A_0 > 0$, the analytic function

$$F(s) = \frac{1}{s} C(s) - \frac{A_0}{s-1} \qquad (\Re(s) > 1)$$

has a continuous extension to the closed half-plane $\Re(s) \ge 1$. Then

$$\overline{c}(v) \sim A_0 v$$

as $v \to \infty$.

Theorem 11 is quite flexible. For example, it is sufficient to assume that $\overline{c}(v)(\log v)^b$ is non-decreasing for some real b (and $v \ge 2$). Furthermore it is clear that it generalizes directly to the case when $C(s)$ converges for $\Re(s) > s_0$ and has a continuous extension to the closed half-plane $\Re(s) \ge s_0$ (for $s_0 \ge 0$). It also applies if $C(s)$ behaves like a pole of higher order for $s \to s_0$, however, the asymptotic result has to be adjusted accordingly.

Theorem 12. *Let $\overline{c}(v)$ be non-negative and non-decreasing on $[1,\infty)$ such that the Mellin-Stieltjes transform $C(s)$ exists for $\Re(s) > s_0$ for some $s_0 \ge 0$ and suppose that there exist real constants $A_0, \ldots, A_K$ (with $A_K > 0$) such that*

$$\tilde{F}(s) = \frac{1}{s} C(s) - \sum_{j=0}^{K} \frac{A_j}{(s-s_0)^{j+1}} \tag{B.1}$$

has a continuous extension to the closed half-plane $\Re(s) \ge s_0$. Then we have

$$\overline{c}(v) \sim \frac{A_K}{K!} (\log v)^K v^{s_0} \qquad (v \to \infty). \tag{B.2}$$

We indicate how Theorem 12 can be deduced from (a slight variation of) Theorem 11 when $K = 2$ and $s_0 = 1$. Let

$$\frac{1}{s}C(s) = \int_1^\infty \overline{c}(v)v^{-s-1}\,ds = \frac{A_1}{(s-1)^2} + \frac{A_0}{s-1} + \tilde{F}(s)$$

with some $A_1 > 0$ and some function $\tilde{F}(s)$ that is analytic for $\Re(s) > 1$ and has a continuous extension to the half plane $\Re(s) \ge 1$. By subtracting $A_0/(s-1)$ and by splitting up the integral into two parts we obtain

$$\int_2^\infty (\overline{c}(v) - A_0 v)\, v^{-s-1}\,dv = \frac{A_1}{(s-1)^2} + \tilde{F}(s) \\ - \int_1^2 (\overline{c}(v) - A_0 v)\, v^{-s-1}\,dv$$

Hence, by integrating with respect to s (from 2 to s) we have

$$\int_2^\infty \left(\frac{\overline{c}(v) - A_0 v}{\log v}\right) v^{-s-1}\,dv = \frac{A_1}{s-1} - A_1 - \int_2^s \tilde{F}(t)\,dt + \int_2^\infty \left(\frac{\overline{c}(v) - A_0 v}{\log v}\right) v^{-3}\,dv \\ + \int_2^s \int_1^2 (\overline{c}(v) - A_0 v) v^{-t-1}\,dv\,dt.$$

We can apply a slight generalization of Theorem 11 to $(\overline{c}(v) - A_0 v)/\log v$. Note that the right hand side is of the form $A_1/(s-1) + \overline{F}(s)$, where $\overline{F}(s)$ has a continuous continuation to the half plane $\Re(s) \ge 1$. The point is that the function $(\overline{c}(v) - A_0 v)/\log v$ is not necessarily non-negative and non-decreasing. However, there is certainly a constant $C > 0$ such that $(\overline{c}(v) - A_0 v)/\log v + Cv \ge 0$, and A_1 on the right hand side can be replaced by $A_1 + C$. Furthermore, the proof of Theorem 11 has some flexibility. As mentioned above the proof of Theorem 11 can be easily modified so that it also applies to a function of the form $(\overline{c}(v) - A_0 v)/\log v$, where it is only assumed that $\overline{c}(v)$ is non-decreasing [Korevaar 2002].

Note that the cases $s_0 > 0$ and $s_0 = 0$ of Theorem 2 have to be handled separately.[1] Furthermore, the case $s_0 < 0$ is not applicable in this setting. Namely if $c(v) > 0$ and non-decreasing, then $C(s)$ cannot converge for s with $\Re(s) < 0$. Note also that we cannot expect a more precise asymptotic expansion in this generality. For example if $\overline{c}(v) = (A_1 \log v + A_0 + \sin(\log^2 v))v$ with $A_1 > 2$. Then $\overline{c}(v)$ is positive and non-decreasing, so (B.1) is satisfied but we do not have $\overline{c}(v) = (A_1 \log v + A_0 + o(1))v$.

Remark 13. The above mentioned proof method of Theorem 12 also applies to situations, where $\frac{1}{s}C(s)$ has a representation of the form

$$\frac{1}{s}C(s) = \int_1^\infty \overline{c}(v)v^{-s-1}\,ds = \frac{A_1}{(s-1)^2} + \sum_{m\in\mathbb{Z}} \frac{A_{0,m}}{s + im\tau - 1} + \tilde{F}(s)$$

with some $A_1 > 0$ and some function $\tilde{F}(s)$ that is analytic for $\Re(s) > 1$ and has a continuous extension to the half plane $\Re(s) \ge 1$. Furthermore we have to assume

[1] The approach we present works for $s_0 > 0$. For $s_0 = 0$ we have to adjust parts of the proof of Theorem 11.

that the Fourier series

$$\sum_{m\in\mathbb{Z}} A_{0,m} e^{im\tau x}$$

is convergent and represents an integrable function. Note that this condition corresponds to the condition (20) in Theorem 2. Under these assumptions the previous proof works, too, and it follows that $\overline{c}(v) \sim A_1 v \log v$.

This kind of reasoning is precisely what is needed in Section 5.5, where we completed the proof of Theorem 2 in the rationally related case.

There are even more general versions by [Delange 1954] that cover singularities of algebraic-logarithmic type that we state next. Note that this theorem requires an analytic continuation property and not only a continuity property.

Theorem 13 ([Delange 1954]). *Let $\overline{c}(v)$ be non-negative and non-decreasing on $[1,\infty)$ such that the Mellin-Stieltjes transform $C(s)$ exists for $\Re(s) > s_0$ for some $s_0 > 0$ and suppose that there exist functions $\tilde{F}(s)$, $g_0(s),\ldots,g_J(s)$ that are analytic in a region that contains half plane $\Re(s) \ge s_0$ such that*

$$\frac{1}{s}C(s) = g_0(s)\frac{\left(\log\frac{1}{s-s_0}\right)^{\beta_0}}{(s-s_0)^{\alpha_0}} + \sum_{j=1}^{J} g_j(s)\frac{\left(\log\frac{1}{s-s_0}\right)^{\beta_j}}{(s-s_0)^{\alpha_j}} + \tilde{F}(s),$$

where $g_0(s_0) \ne 0$, β_j are non-negative integers, α_0 is real but not a negative integer when it is non-zero, and $\alpha_1,\ldots,\alpha_J$ are complex numbers with $\Re(\alpha_j) < \alpha_0$. Furthermore $\beta_0 > 0$ if α_0 is contained in the set $\{0,-1,-2,\ldots\}$. Then, as $v\to\infty$,

$$\overline{c}(v) \sim \frac{g_0(s_0)}{\Gamma(\alpha_0)}(\log v)^{\alpha_0-1}(\log\log v)^{\beta_0} v^{s_0} \tag{B.3}$$

if α_0 is not contained in the set $\{0,-1,-2,\ldots\}$ and

$$\overline{c}(v) \sim (-1)^{\alpha_0}(-\alpha_0)!\beta_0 g_0(s_0)(\log v)^{\alpha_0-1}(\log\log v)^{\beta_0-1} v^{s_0} \tag{B.4}$$

if α_0 is contained in the set $\{0,-1,-2,\ldots\}$ and $\beta_0 > 0$.

Interestingly, Theorem 11 generalizes – partly – to the case, where there are infinitely many *poles* on the line $\Re(s) = s_0$, where one obtains a fluctuating factor in the asymptotic expansion.

The drawback of this generalization is that it only applies if the appearing periodic function has an absolutely convergent Fourier series. Unfortunately we cannot check this in general although we can expect this property provided that the small growth assumption (22) is satisfied, see also [Grabner and Hwang 2005].

Anyway, we could not find such a theorem in the literature, so we present it here.

Theorem 14. *Let $\overline{c}(v)$ be non-negative and non-decreasing on $[1,\infty)$ such that the Mellin-Stieltjes transform $C(s)$ exists for $\Re(s) > s_0$, where $s_0 > 0$. Assume that the function*

$$\tilde{F}(s) = \frac{1}{s}C(s) - \sum_{m\in\mathbb{Z}} \frac{A_m}{s - s_0 - im\tau}, \tag{B.5}$$

with some real $\tau > 0$ and real coefficients A_m, where $A_0 > 0$, has a continuous extension to the closed half-plane $\Re(s) \ge s_0$. Furthermore assume that the Fourier series

$$\Psi(x) = \sum_{m\in\mathbb{Z}} A_m\, e^{im\tau x}$$

is absolutely convergent and has bounded derivative. Then

$$\overline{c}(v) \sim \Psi(\log v)\, v^{s_0} \qquad (v \to \infty). \tag{B.6}$$

The proof is an extension of the approach from [Korevaar 2002]. For the reader's convenience we give it here. Let

$$K_\lambda(t) = \frac{1-\cos(\lambda t)}{\pi \lambda t^2} = \frac{\lambda}{2\pi}\left(\frac{\sin(\lambda t/2)}{\lambda t/2}\right)^2$$

denote the Fejer kernel.

Lemma 5. *Let $\kappa > 0$ and*

$$h(t) = \sum_{m\in\mathbb{Z}} A_m e^{im\tau t}$$

be an absolutely convergent Fourier series with bounded derivative. Then

$$\int_0^\infty K_\lambda(u-t)h(t)\,dt = h(u) + o(1) \qquad (\lambda \to \infty) \tag{B.7}$$

uniformly for $u \ge 1$.

Proof. Let $m_0 = m_0(\varepsilon)$ be defined by

$$\sum_{|m|>m_0(\varepsilon)} |A_m| < \varepsilon$$

and suppose that for $\lambda_0 = \lambda_0(\varepsilon) > \kappa m_0(\varepsilon)$ we have

$$\sum_{|m|\le m_0(\varepsilon)} |mA_m| < \varepsilon\, \lambda_0(\varepsilon).$$

Furthermore we note that for $u \ge 1$

$$\int_u^\infty K_\lambda(t)\,dt = O\left(\frac{1}{\lambda}\right) \qquad (\lambda \to \infty).$$

Consequently it follows for $\lambda \ge \lambda_0(\varepsilon)$

$$\begin{aligned}
\int_0^\infty K_\lambda(u-t)h(t)\,dt &= \sum_{m\in\mathbb{Z}} A_m \int_0^\infty K_\lambda(u-t)e^{im\tau t}\,dt \\
&= \sum_{m\in\mathbb{Z}} A_m \int_{-\infty}^\infty K_\lambda(t)e^{im\tau(u-t)}\,dt + O\left(\int_u^\infty K_\lambda(t)\,dt\right) \\
&= \sum_{m\in\mathbb{Z}} A_m e^{im\kappa u}\hat{K}_\lambda(\kappa m) + O\left(\frac{1}{\lambda}\right) \\
&= \sum_{m\in\mathbb{Z}} A_m e^{im\kappa u}\left(1-\frac{|\kappa m|}{\lambda}\right) + O\left(\frac{1}{\lambda}\right) \\
&= \sum_{|m|\le\lambda/\kappa} A_m e^{im\kappa u} + O\left(\frac{1}{\lambda}\sum_{|m|\le\lambda/\kappa}|mA_m|\right) + O\left(\frac{1}{\lambda}\right).
\end{aligned}$$

Since $\lambda_0 > \kappa m_0$ we have

$$\left| h(u) - \sum_{|m|\le\lambda/\kappa} A_m e^{im\kappa u}\right| \le \sum_{|m|>m_0(\varepsilon)} |A_m| < \varepsilon.$$

Furthermore

$$\begin{aligned}
\frac{1}{\lambda}\sum_{|m|\le\lambda/\kappa}|mA_m| &\le \frac{1}{\lambda}\sum_{|m|\le m_0}|mA_m| + \frac{1}{\lambda}\sum_{m_0<|m|\le\lambda/\kappa}|mA_m| \\
&< \varepsilon + \sum_{|m|>m_0}|A_m| \\
&< 2\varepsilon.
\end{aligned}$$

Of course this proves (B.7).

Lemma 6. *Let $\ell(t)$ be non-negative for $t \ge 0$ such that the Laplace transform*

$$L(z) = \int_0^\infty \ell(t)e^{-zt}\,dt$$

exists for $\Re(z) > 0$. Suppose further that there exists a bounded and integrable function $h(t)$ (for $t \ge 0$) with the property that

$$G(z) = L(z) - H(z)$$

has a continuous extension to the closed half-plane $\Re(z) \ge 0$, where $H(z)$ denotes the Laplace transform of $h(t)$. Then

$$\lim_{u\to\infty}\left(\int_0^\infty K_\lambda(u-t)\ell(t)\,dt - \int_0^\infty K_\lambda(u-t)h(t)\,dt\right) = 0$$

Proof. Let $\hat{K}_\lambda(y) = \max\{1-|y|/\lambda, 0\}$ denote the Fourier transform of $K_\lambda(t)$

which is non-negative and has support $[-\lambda, \lambda]$. Then we have for $x > 0$

$$\int_0^\infty K_\lambda(u-t)\ell(t)e^{ixt}\,dt = \int_0^\infty K_\lambda(u-t)h(t)e^{ixt}\,dt + \frac{1}{2\pi}\int_{-\lambda}^{\lambda} \hat{K}_\lambda(y)G(x+iy)e^{iuy}\,dy.$$

By assumption the right hand side has a finite limit as $x \to 0+$. Hence, by the monotone convergence theorem it follows that $K_\lambda(u-t)\ell(t)$ is integrable over $(0,\infty)$ and it follows that

$$\int_0^\infty K_\lambda(u-t)\ell(t)\,dt = \int_0^\infty K_\lambda(u-t)h(t)\,dt + \frac{1}{2\pi}\int_{-\lambda}^{\lambda} \hat{K}_\lambda(y)G(iy)e^{iuy}\,dy.$$

Finally, the Riemann-Lebesgue lemma implies

$$\lim_{u\to\infty} \frac{1}{2\pi}\int_{-\lambda}^{\lambda} \hat{K}_\lambda(y)G(iy)e^{iuy}\,dy = 0.$$

This proves the lemma.

With the help of these preliminaries we prove Theorem 12.

Proof Proof of Theorem 14. We set $\ell(t) = e^{-s_0 t}a(e^t)$. Then for $\Re(z) > 0$

$$L(z) = \int_0^\infty \ell(t)e^{-zt}\,dt = \int_1^\infty a(v)v^{-(s_0+z)-1}\,dv = \frac{A(s_0+z)}{s_0+z}.$$

Furthermore observe that the Laplace transform of $\Psi(t)$ (for $\Re(z) > 0$) is given by

$$H(z) = \int_0^\infty \Psi(t)e^{-tz}\,dt = \sum_{m\in\mathbb{Z}} \frac{A_m}{z - im\tau}$$

Hence, by assumption the function

$$G(z) = L(z) - H(z) = \frac{A(s_0+z)}{s_0+z} - \sum_{m\in\mathbb{Z}} \frac{A_m}{z - im\tau}$$

has a continuous extension to the half-plane $\Re(z) \ge 0$. Consequently by Lemma 6

$$\lim_{u\to\infty} \left(\int_0^\infty K_\lambda(u-t)e^{-s_0 t}a(e^t)\,dt - \int_0^\infty K_\lambda(u-t)\Psi(t)\,dt \right) = 0$$

Since $\Psi(t)$ is bounded it also follows that the second integral is uniformly bounded in λ and u. Hence

$$\limsup_{u\to\infty} \int_0^\infty K_\lambda(u-t)\ell(t)\,dt \le C$$

for some constant that is uniform in λ. Since $a(v)$ is positive and non-decreasing it follows that

$$\int_0^\infty K_\lambda(u-t)\ell(t)\,dt \ge \ell(u-1/\sqrt{\lambda})e^{-2s_0/\sqrt{\lambda}} \int_{-1/\sqrt{\lambda}}^{1/\sqrt{\lambda}} K_\lambda(t)\,dt = \ell(u-1/\sqrt{\lambda})\left(1 + O(1/\sqrt{\lambda})\right)$$

and consequently

$$\limsup_{u\to\infty} \ell(u-1/\sqrt{\lambda}) \le C\left(1+O(1/\sqrt{\lambda})\right).$$

This shows that $\ell(t)$ is a bounded function.

Now, for given $\varepsilon > 0$ choose $\lambda_0 = \lambda_0(\varepsilon) > 1/\varepsilon^2$ such that

$$\left|\int_0^\infty K_{\lambda_0}(u-t)\Psi(t)\,dt - \Psi(u)\right| < \varepsilon.$$

Since $\Psi(t)$ has bounded derivative we also have $|\Psi(u)-\Psi(u-1/\sqrt{\lambda_0})| \le C/\sqrt{\lambda_0} \le C\varepsilon$. Putting these estimates together it follows that

$$\limsup_{u\to\infty}\left(\ell(u-1/\sqrt{\lambda_0})\left(1+O(1/\sqrt{\lambda_0})\right) - \Psi(u-1/\sqrt{\lambda_0})\right) \le (1+C)\varepsilon$$

and consequently

$$\limsup_{u\to\infty}\left(\ell(u)-\Psi(u)\right) \le 0.$$

Similarly we obtain estimates from below. We just have to observe that

$$\begin{aligned}\int_0^\infty K_\lambda(u-t)\ell(t)\,dt &\le \ell(u+1/\sqrt{\lambda})e^{2s_0/\sqrt{\lambda}}\int_{-1/\sqrt{\lambda}}^{1/\sqrt{\lambda}} K_\lambda(t)\,dt + O\left(\int_{1/\sqrt{\lambda}} K_\lambda(t)\,dt\right)\\ &= \ell(u+1/\sqrt{\lambda})\left(1+O(1/\sqrt{\lambda})\right) + O\left(1/\sqrt{\lambda}\right)\end{aligned}$$

and obtain in the same way

$$\liminf_{u\to\infty}\left(\ell(u)-\Psi(u)\right) \ge 0.$$

Hence, $\ell(u) = \Psi(u) + o(1)$ and consequently $a(v) = \left(\Psi(\log v) + o(1)\right) v^{s_0}$. Finally, since $a(v)$ is non-decreasing we have $\min \Psi(u) > 0$ and consequently $a(v) \sim \Psi(\log v) v^{s_0}$.

Acknowledgment

We thank referees for many constructive comments. We also thanks Dr. Y. Reznik for pointing to us Boncelet's algorithm.

REFERENCES

ABRAMOWITZ, M. AND STEGUN, I. 1964. *Handbook of Mathematical Functions.* Dover, New York.

AKRA, M. AND BAZZI, L. 1998. On the solution of linear recurrence equations. *Computational Optimization and Applications 10*, 195–201.

APOSTOL, T. 1976. *Introduction to analytic number theory.* Springer, New York.

BONCELET, C. G. 1993. Block arithmetic coding for source compression. *IEEE Transactions on Information Theory 39*, 1546–1554.

CHEUNG, Y., FLAJOLET, P., GOLIN, M., AND LEE, C. 2008. Multidimensional divide-and-conquer and weighted digital sums. In *ANALCO*.

CHOI, V. AND GOLIN, M. J. 2001. Lopsided trees. i. analyses. *Algorithmica 31*, 240–290.

CORMEN, T., LEIERSON, C., AND RIVEST, R. 1990. *Introduction to Algorithms.* MIT Press, Cambridge.

Delange, H. 1954. Généralisation du théorème de ikehara. *Ann. Sci. Éc. Norm. Supér. 71*, 213–242.

Delange, H. 1975. Sur la fonction sommatoire de la fonction "somme des chiffres. *Enseignement Math. 21*, 31–47.

Drmota, M., Reznik, Y., and Szpankowski, W. 2010. Tunstall code, khodak variations, and random walks. *IEEE Transactions on Information Theory 56*, 2928–2937.

Erdös, P., Hildebrand, A., Odlyzko, A., Pudaite, P., and Reznick, B. 1987. The asymptotic behavior of a family of sequences. *Pacific Journal of Mathematics 126*, 227–241.

Fayolle, G., Flajolet, P., and Hofri, M. 1986. On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel. *Adv. Appl. Prob. 18*, 441–472.

Flajolet, P. and Golin, M. 1954. Mellin transforms and asymptotics: The mergesort recurrence. *Acta Informatica 31*, 673–696.

Flajolet, P., Gourdon, X., and Dumas, P. 1995. Mellin transforms and asymptotics: harmonic sums. *Theoretical Computer Science 144*, 3–58.

Flajolet, P., Roux, M., and Vallee, B. 2010. Digital trees and memoryless sources: from arithmetics to analysis. In *Discrete Math. Theor. Comput. Sci. Proc.* Vienna, Austria, 233–260. 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms.

Flajolet, P. and Sedgewick, R. 2008. *Analytic Combinatorics*. Cambridge University Press.

Grabner, P. and Hwang, H. K. 2005. Digital sums and divide-and-conquer recurrences: Fourier expansions and absolute convergence. *Constr. Approx. 21*, 149–179.

Grabner, P. and Thuswaldner, J. M. 1996. Analytic continuation of a class of dirichlet series. *Abh. Math. Sem. Univ. Hamburg 66*, 281–287.

Hwang, H. K. 2000. Distribution of the number of factors in random ordered factorizations of integers. *Journal of Number Theory 81*, 61–92.

Hwang, H. K. and Janson, S. 2011. A central limit theorem for random ordered factorizations of integers. *Electronic J. Probab. 16*, 347–361.

Karatsuba, A. and Ofman, Y. 1963. Multiplication of many-digital numbers by automatic computers. *Doklady Akad. Nauk SSSR 7*, 595–596.

Khodak, G. L. 1969. Connection between redundancy and average delay of fixed-length coding. In *All-Union Conference on Problems of Theoretical Cybernetics*.

Knuth, D. E. 1998. *The Art of Computer Programming. Seminumerical Algorithms*, Third ed. Vol. 2. Addison-Wesley.

Korevaar, J. 2002. A century of complex tauberian theory. *Bull. Amer. Math. Soc. 39*, 475–531.

Korner, T. 1989. *Fourier Analysis*. Cambridge University Press.

Leighton, T. 1996. Notes on better master theorems for divide-and-conquer recurrences. In *MIT*.

Roura, S. 2001. Improved master theorems for divide-and-conquer recurrences. *Journal of the ACM 48*, 170–205.

Savari, S. and Gallager, R. 1997. Generalized tunstall codes for sources with memory. *IEEE Trans. Info. Theory 43*, 658–668.

Schachinger, W. 2001. Limiting distributions for the costs of partial match retrievals in multidimensional tries. *Random Structures and Algorithms 17*, 428–459.

Strassen, V. 1969. Gaussian elimination is not optimal. *Numer. Math. 13*, 354–356.

Szpankowski, W. 2001. *Average Case Analysis of Algorithms on Sequences*. John Wiley.

Tunstall, B. P. 1967. Synthesis of noiseless compression codes. In *Georgia Inst. Tech.*