A Master Theorem for Discrete Divide and Conquer Recurrences*

Wojciech Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907

January 20, 2011

NSF CSoI

SODA, 2011

*Research supported by NSF Science & Technology Center, and Humboldt Foundation.

Outline

1. Divide and Conquer
2. Example: Boncelet's Algorithm
3. Continuous Relaxation of the Recurrence
4. Master Theorem
5. Examples
6. Sketch of Proof.

Divide and Conquer

Divide and Conquer:
A divide and conquer algorithm splits the input into several smaller subproblems, solving each subproblem separately, and then knitting together to solve the original problem.

Complexity:
A problem of size n is divided into $m \geq 2$ subproblems of size $\lfloor p_j n + \delta_j \rfloor$ and $\lceil p_j n + \delta_j' \rceil$ and each subproblem contributes b_j, b_j' fraction to the final solution; there is a cost a_n associated with combining subproblems.

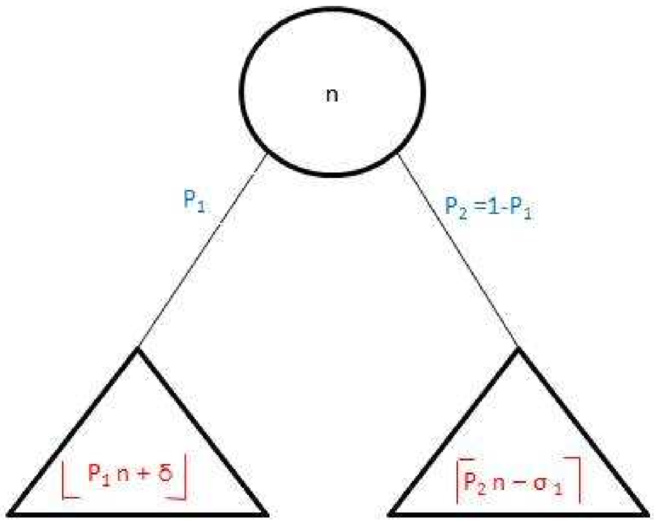
Total Cost:
The total cost $T(n)$ satisfies the **discrete divide and conquer** recurrence:

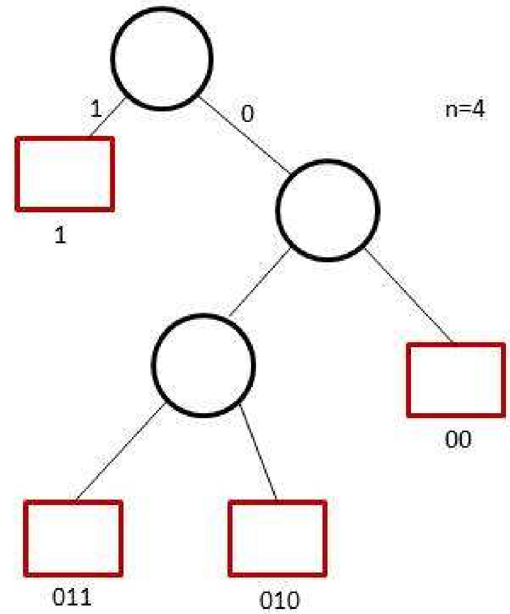
$$T(n) = a_n + \sum_{j=1}^{m} b_j T\left(\lfloor p_j n + \delta_j \rfloor\right) + \sum_{j=1}^{m} b_j' T\left(\left\lceil p_j n + \delta_j' \right\rceil\right) \qquad (n \geq 2)$$

where $0 \leq p_j < 1$ (e.g., $\sum_{i=1}^{m} p_i = 1$).

Example: Boncelet's Algorithm

Boncelet's algorithm is a variable-to-fixed data compression scheme:

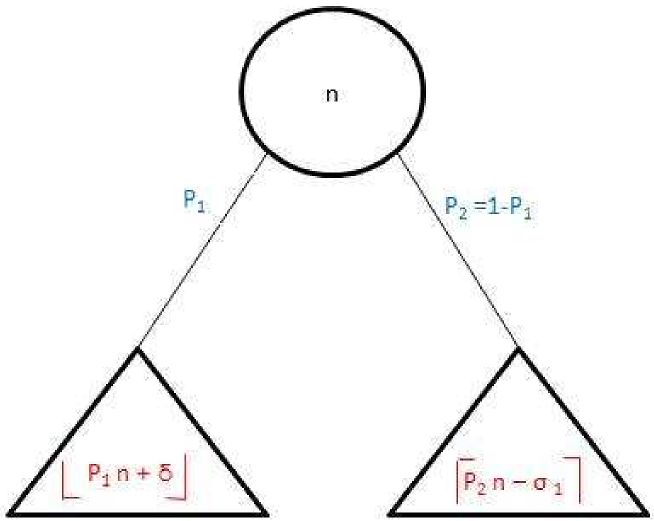


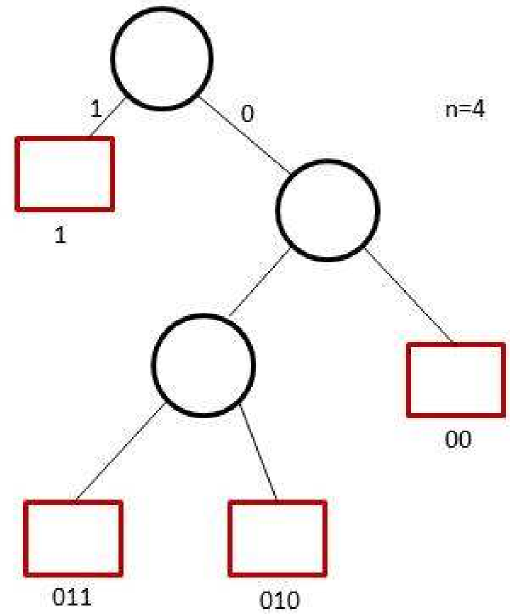


1. A variable-to-fixed length encoder partitions a source string over an m-ary alphabet into variable-length phrases.
2. Each phrase belongs to a given dictionary.
3. A dictionary is represented by a complete parsing tree.
4. The dictionary entries correspond to the leaves of the parsing tree.

Example: Boncelet's Algorithm

Boncelet's algorithm is a variable-to-fixed data compression scheme:





1. A variable-to-fixed length encoder partitions a source string over an m-ary alphabet into variable-length phrases.
2. Each phrase belongs to a given dictionary.
3. A dictionary is represented by a complete parsing tree.
4. The dictionary entries correspond to the leaves of the parsing tree.

Example: A binary string ($m = 2$) with symbol probabilities p_1 and p_2. The expected phrase length $d(n)$ satisfies:

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

Continuous Relaxation

We relax the discrete nature of the recurrence and consider a continuous version:

$$T(x) = a(x) + \sum_{j=1}^{m} b_j T(p_j x)), \quad x > 1, \quad b_j' = 0.$$

Akra and Bazzi (1998) proved that

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

where s_0 is a unique real root of $\sum_j b_j {p_j}^{s_0} = 1$.

Continuous Relaxation

We relax the discrete nature of the recurrence and consider a continuous version:

$$T(x) = a(x) + \sum_{j=1}^{m} b_j T(p_j x)), \quad x > 1, \quad b_j' = 0.$$

Akra and Bazzi (1998) proved that

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

where s_0 is a unique real root of $\sum_j b_j p_j^{s_0} = 1$.

Indeed, by taking Mellin transform of the relaxed recurrence:

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

we find (for some $a(s)$ and $g(s)$)

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

An application of the Wiener-Ikehara theorem leads to

$$T(x) \sim C x^{s_0} \quad \text{with} \quad C = \frac{a(-s_0) + g(-s_0)}{\sum_j b_j p_j^{s_0} \log(1/p_j)}.$$

Discrete Divide & Conquer Recurrence by Dirichlet Series

For a sequence $c(n)$ define the Dirichlet series as

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

provided it exists for $\Re(s) > \sigma_c$ for some $\sigma_c \geq -\infty$.

Theorem 1 (Perron-Mellin Formula)**.** *For all* $\sigma > \sigma_c$ *and all* $x > 0$

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

where $[\![P]\!]$ *is* 1 *if* P *is a true proposition and* 0 *otherwise.*

Discrete Divide & Conquer Recurrence by Dirichlet Series

For a sequence $c(n)$ define the Dirichlet series as

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

provided it exists for $\Re(s) > \sigma_c$ for some $\sigma_c \geq -\infty$.
Theorem 1 (Perron-Mellin Formula). *For all $\sigma > \sigma_c$ and all $x > 0$*

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

where $[\![P]\!]$ is 1 if P is a true proposition and 0 otherwise.

Example: Define $c(n) = T(n+2) - T(n+1)$. Then

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

for some $c > \sigma_{\widetilde{T}}$ with

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

where $\Re(s) > \sigma_{\widetilde{T}}$.

Assumptions

Let a_n be a nondecreasing sequence. Define

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

which is postulated to exists for $\Re(s) > \sigma_a$.

Example. Define $a_n = n^\sigma (\log n)^\alpha$. Then

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

where $\tilde{F}(s)$ is analytic for $\Re(s) > \sigma - 1$ and $\Gamma(s)$ is the gamma function.

Define s_0 to be the **unique real** root of

$$\sum_{j=1}^{m} (b_j + b_j') \, p_j^{\,s} = 1.$$

Other zeros depend on the relation among $\log(1/p_1), \ldots, \log(1/p_m)$.

Rationally and Irrationally Related Numbers

Definition 1. (i) $\log(1/p_1), \ldots, \log(1/p_m)$ *are* *rationally related* *if* $\log(1/p_1), \ldots, \log(1/p_m)$ *are integer multiples of* L, *that is,* $\log(1/p_j) = n_j L$, $n_j \in \mathbb{Z}$, $(1 \leq j \leq m)$.
(ii) *Otherwise* $\log(1/p_1), \ldots, \log(1/p_m)$ *are* *irrationally related*.

Example. If $m = 1$, then we are always in the rationally related case.
For $m = 2$, if $\log(1/p_1)/\log(1/p_2) = m/n$, $(m, n$ integers), then rationally related.

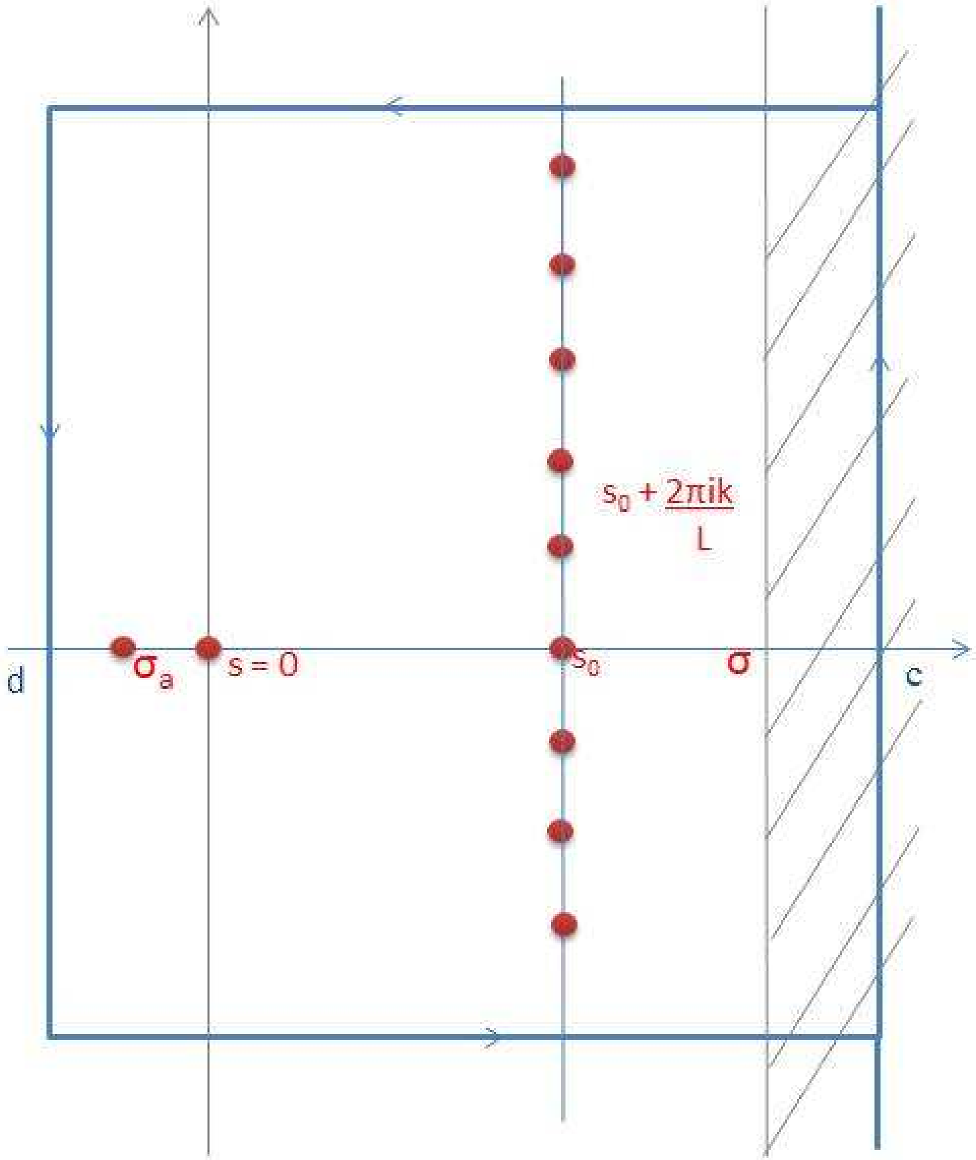
Lemma 1. (i) *If* $\log(1/p_1), \ldots, \log(1/p_m)$ *are* *irrationally related*, *then* s_0 *is the* *only solution* *on* $\Re(s) = s_0$.

(ii) *If* $\log(1/p_1), \ldots, \log(1/p_m)$ *are* *rationally related*, *then there are infinitely many solutions*

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

where $\log(1/p_j)$ *are all integer multiples of* L.

Evaluation of $T(n)$: A Bird View



$$\widetilde{T}(s) = \frac{\widetilde{A}(s) + B(s)}{1 - \sum_{j=1}^{m} (b_j + b'_j)\, p_j^s},$$

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

Main Master Theorem

Theorem 2 (DISCRETE MASTER THEOREM). *Let* $a_n = Cn^{\sigma_a}(\log n)^{\alpha}$ *with* $\min\{\sigma, \alpha\} \geq 0$.

(i) If $\log(1/p_1), \ldots, \log(1/p_m)$ *are irrationally related, then*

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

(ii) *If* $\log(1/p_1), \ldots, \log(1/p_m)$ *are rationally related, then* $T(n)$ *behaves as in the irrationally related case with the following two exceptions:*

$$T(n) = \begin{cases} C_2 \log n + \Psi_2(\log n) + o(1) & \text{if } \sigma_a < s_0 = 0, \\ \Psi_4(\log n)\, n^{s_0} \cdot (1 + o(1)) & \text{if } \sigma_a < s_0 \text{ and } s_0 > 0, \end{cases}$$

where C_2 *is positive and* $\Psi_2(t)$, $\Psi_4(t)$ *are periodic functions with period* L *(with usually countably many discontinuities).*

Extensions and Remarks

1. We can handle any a_n sequence with Dirichlet series $\widetilde{A}(s)$:

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

$\tilde{F}(s)$ is analytic, $g_0(\sigma_a) \neq 0$, β_j non-negative integers, and α_0 real.
Then the solution $T(n)$ of the divide and conquer recurrence has the form:

$$T(n) \sim C\, n^{\sigma'} (\log n)^{\alpha'} (\log\log n)^{\beta'} \quad or \quad T(n) \sim \Psi(\log n)\, n^{s_0}$$

$\sigma' = \max\{\sigma, s_0\}$), depending whether $\log p_1, \ldots \log p_m$ are irrationally or rationally related.

2. The periodic function $\Psi(t)$ has the following building blocks

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

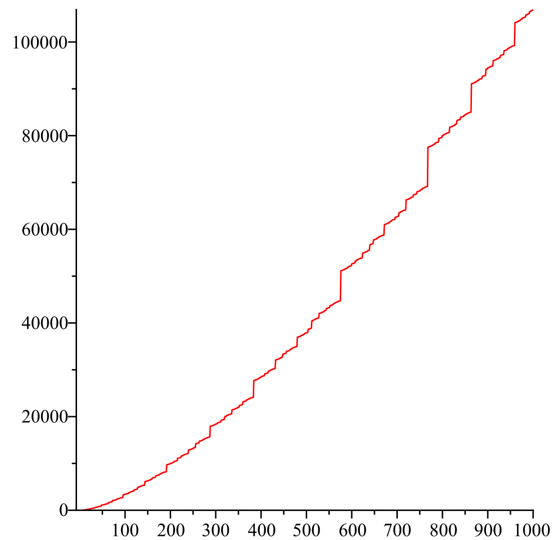
where $\lambda > 1$ and B_n is such that $\sum_{n\geq 1} B_n \lambda^{-(\log n)/L}$ converges absolutely.
This function is discontinuous at

$$t = \{\log n/L\},$$

where $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of a real number x.

Examples

Example 1. **Irrationally Related; Case 4**:



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

Here $\sigma_a = 1$ since $a_n = n \log n$.
The equation

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

has the (real) solution $s_0 = 1.402\ldots > 1$, and finally $\log(1/2)/\log(1/6)$ are irrationally related. Thus by our **Master Theorem** Case 4

$$T(n) \sim C n^{s_0}$$

for some constant $C > 0$

Examples

Example 2. **Irrationally Related; Case 6**:
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 $\sigma_a = s_0 = 2$, and we deal with irrationally related case. Furthermore,

$$\widetilde{A}(s) = s \log \frac{1}{s-2} + G(s)$$

for $G(s)$ analytic for $\Re(s) > 1$. By Master Theorem Case 6

$$T(n) \sim C n^2 \log\log n.$$

Example 3. **Rationally Related ($m = 1$); Case 3**:
Next consider

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

Here $\sigma_a = s_0 = 0$, and we have rational case $(m = 1)$. Since

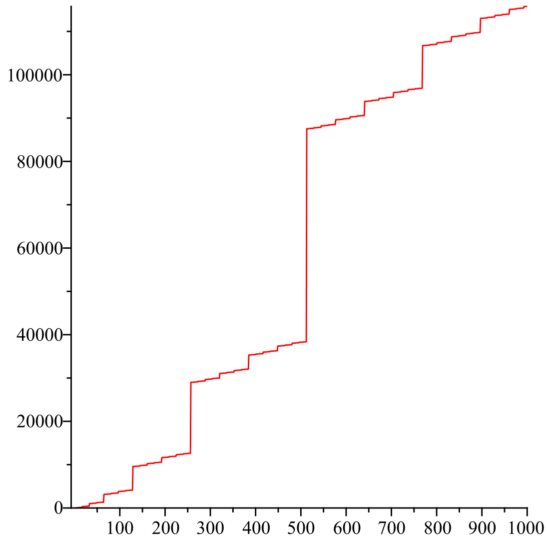
$$\widetilde{A}(s) = \frac{1}{s} + G(s)$$

we conclude

$$T(n) \sim C(\log n)^2.$$

Examples

Example 4: Karatsuba algorithm: Rationally Related $(m = 1)$:



$$T(n) = 3T(\lceil n/2 \rceil) + n$$

Here, $s_0 = (\log 3)/(\log 2) = 1.5849\ldots$ and $s_0 > \sigma_a = 1$. Thus

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

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

Examples

Example 5. **Rationally Related** $(m = 1)$. The recurrence

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

is not covered by our Master Theorem but our methodology still works. Here $\sigma_a = s_0 = -1 < 0$. It follows that

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

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

Example 6: Mergesort. Rationally Related.
The mergesort recurrences are

$$T(n) = T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n - 1,$$

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

Here $\sigma_a = s_0 = 1$ and we deal with the rationally related case. By our Master Theorem (cf. Flajolet & Golin, 1994)

$$\begin{aligned} T(n) &= \frac{1}{\log 2} n \log n + n\Psi(\log n) + o(n), \\ Y(n) &= \frac{1}{2\log 2} n \log n + n\Psi(\log n) + o(n). \end{aligned}$$

Boncelet's Algorithm Revisited

Let a sequence X be generated by a memoryless source over alphabet $\mathcal{A}$ of size m with symbol probabilities p_i, $i \in \mathcal{A}$.

Using the Boncelet's parsing tree, we parse X into phrases $\{v_1, \dots v_n\}$ of length $\ell(v_1), \dots, \ell(v_n)$ with phrase probabilities $P(v_1), \dots, P(v_n)$.

Phrase Length and its Probability Generating Function:
Let D_n denote the phrase length and define the probability generating function as

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

It satisfies the following discrete divide and conquer recurrence:

$$C(n, y) = y \sum_{i=1}^{m} p_i C([p_i n + \delta_i], y)$$

The expected phrase length $d(n) = \mathbf{E}[D_n] = C'(n, 1)$ satisfies the following discrete divide and conquer recurrence:

$$d(n) = 1 + \sum_{i=1}^{m} p_i d([p_i n + \delta_i])$$

with $d(0) = \cdots = d(m-1) = 0$.

Main Results for Boncelet's Algorithm

Theorem 3. *Consider an m-ary memoryless source with probabilities $p_i > 0$ and the entropy rate $H = \sum_{i=1}^{m} p_i \log(1/p_i)$.*

(i) *If $\log(1/p_1), \dots \log(1/p_m)$ are irrationally related, then*

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

where

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

$H_2 = \sum_{i=1}^{m} p_i \log^2 p_i$, and $E'(0)$ is the derivative at $s = 0$ of a Dirichlet series $E(s)$ arises from the **discrete** *nature of the recurrence.*

(ii) *If $\log(1/p_1), \dots \log(1/p_m)$ are rationally related, then*

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

for some $\eta > 0$, where $\Psi(t)$ is a periodic function of bounded variation that has usually an infinite number of discontinuities.

Redundancy of the Boncelet's Algorithm

Corollary 1. *Let R_n denote the redundancy of the Boncelet code:*

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

(i) *If* $\log(1/p_1), \ldots \log(1/p_m)$ *are irrationally related, then*

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

(ii) *If* $\log(1/p_1), \ldots \log(1/p_m)$ *are rationally related, then*

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

Tunstall Code Redundancy:

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

for irrational case; in the rational case there is aperiodic function.

Example. Consider $p = 1/3$ and $q = 2/3$. Then one computes $\alpha = E'(0) - H - \frac{H_2}{2H} \approx 0.322$ while for the Tunstall code $-\log H - \frac{H_2}{2H} \approx 0.0496$.

Limiting Distribution for the Phrase length

Theorem 4. *Consider a memoryless source generating a sequence of length n parsed by the Boncelet algorithm. If $(p_1, \ldots, p_m)$ is not the uniform distribution, then 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

$$\mathbf{E}[D_n] = \frac{\log n}{H} + O(1),$$

$$\mathbb{V}\mathrm{ar}\, D_n \sim \left(\frac{H_2}{H^3} - \frac{1}{H}\right)\log n$$

for $n \to \infty$.

Sketch of Proof

1. From the recurrence we have

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

But defining

$$n = \left\lfloor \frac{k+2-\delta_j}{p_j} \right\rfloor - 2$$

for some integer k, we arrive at

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

for an explicit (and simple) analytic function $G_j(s)$.

Sketch of Proof – Continuation

2. We now compare the last sum to $p_j^s\widetilde{T}(s)$ and obtain

$$\sum_{k=1}^{\infty}\frac{T(k+2)-T(k+1)}{\left(\left\lfloor\frac{k+2-\delta_j}{p_j}\right\rfloor-2\right)^s}=\sum_{k=1}^{\infty}\frac{T(k+2)-T(k+1)}{(k/p_j)^s}=p_j^s\widetilde{T}(s)-E_j(s),$$

where

$$E_j(s)=\sum_{k=1}^{\infty}(T(k+2)-T(k+1))\left(\frac{1}{(k/p_j)^s}-\frac{1}{\left(\left\lfloor\frac{k+2-\delta_j}{p_j}\right\rfloor-2\right)^s}\right).$$

3. Defining

$$E(s)=\sum_{j=1}^{m}b_jE_j(s)\quad\text{and}\quad G(s)=\sum_{j=1}^{m}b_jG_j(s)$$

we finally obtain our final formula

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

Sketch of Proof – Binary Boncelet's Algorithm

1. Let $s_0(y)$ be the real zero of

$$y(p^{s+1} + q^{s+1}) = 1, \quad q = 1 - p.$$

2. Define

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

which from the basic recurrence becomes

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

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

3. By Mellin-Perron formula and residue theorem we can prove that

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

where

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

That's It

THANK YOU