2007 Conference on Analysis of Algorithms, AofA 07 DMTCS proc. **AH**, 2007, 53–72

Randomized Optimization: a Probabilistic Analysis

Jean Cardinal and Stefan Langerman[†] and Guy Louchard

Université Libre de Bruxelles
Computer Science Department, CP 212
B-1050 Brussels, Belgium
{ jcardin, slanger, louchard } @ulb.ac.be

In 1999, Chan proposed an algorithm to solve a given optimization problem: express the solution as the minimum of the solutions of several subproblems and apply the classical randomized algorithm for finding the minimum of r numbers. If the decision versions of the subproblems are easier to solve than the subproblems themselves, then a faster algorithm for the optimization problem may be obtained with randomization. In this paper we present a precise probabilistic analysis of Chan's technique.

Keywords: –

1 Introduction

In [Cha99], Chan considers the following classical randomized algorithm for finding the minimum of r numbers $\min\{A[1], \ldots, A[r]\}$:

Randomization in the above algorithm ensures that the assignation step $t := A[i_k]$; is executed only a few number of times on average compared to the test $A[i_k] < t$. In fact, we know that the expected number of assignations is given by the rth harmonic number H_r, and hence that the algorithm runs in time $O(r\mathcal{D} + \ln rT)$ expected time, where $\mathcal{D}$ and T are the complexities of respectively the decision and assignation steps.

This can be useful in the context of minimization or maximization problems. It is often the case that we can find more efficient algorithms for the decision version of the problem, i.e. answering for instance whether the minimum value is greater than a given value t, than for the optimization version. Transforming the decision algorithm into an optimization algorithm usually increases the complexity, typically multiplying it by a factor (poly)logarithmic in the problem size.

Chan showed that if we have a method for dividing an optimization problem A into subproblems $A[1], A[2], \ldots, A[r]$ of roughly equal sizes, and that the optimal value for A is defined as the minimum value of the subproblems, then the above algorithm can be used as a method for solving an optimization

[†]Chercheur Qualifié FNRS.

Algorithm 1

Input: r
Output: $\min\{A[1], \ldots, A[r]\}$
randomly pick a permutation $\{i_1, \ldots, i_r\}$ of $\{1, \ldots, r\}$;
$t := \infty$;
for $k = 1$ to r **do**
  if $A[i_k] < t$ **then**
    $t := A[i_k]$;
  end if;
end for;
return t;

problem in randomized expected time $\mathcal{O}(\mathcal{D})$, where $\mathcal{D}$ is the time needed for solving the decision version of problem A. So up to a constant factor, we solve the optimization problem as efficiently as the decision problem. This is achieved simply by making recursive calls for the assignation steps in the algorithm outlined above. Recursion stops as soon as the problem size drops below a given constant, at which point we solve the problem directly.

More precisely, Chan proves the following Lemma. Reusing Chan's notations, we define Π as the *problem space*, and for a problem $P \in \Pi$, we denote by $w(P) \in \mathbb{R}$ its *solution*, and by $|P|$ its *size* (a positive integer). It is further assumed that the solution of a problem of constant size can be computed in constant time. The Lemma is given as follows. To compare with Chan [Cha99], we must set his parameter $\alpha = 1/b$.

Lemma 1.1 *Let $b > 1, \lambda > 0$, and r be constants, and let $\mathcal{D}(.)$ be a function such that $\mathcal{D}(n)/n^{\lambda}$ is monotone increasing in n. Given any problem $P \in \Pi$, suppose that within $\mathcal{D}(|P|)$ time,*
(i) we can decide whether $w(P) < t$ for any given $t \in \mathbb{R}$, and
(ii) we can construct r subproblems $P_1, \ldots, P_r \in \Pi$, each of size at most $\lceil |P|/b \rceil$, so that

$$w(P) = \min\{w(P_1), \ldots, w(P_r)\}.$$

Then for any problem $P \in \Pi$, we can compute the solution $w(P)$ in $\mathcal{O}(\mathcal{D}(|P|))$ time.

In [Cha99], Chan gives a suprisingly long series of geometric optimization problems where this technique can be efficiently applied.

In this paper, we want to present a precise probabilistic analysis of Chan's technique. First, we give an exact expression for the dominant terms. Next, we provide successive contributions to the asymptotic algorithm cost. The particularity of the associated recurrence is that it deals with the $\lceil \rceil$ operator. This entails interesting analysis and convergence difficulties. The paper is organized as follows: section 2 summarizes the notations we need in the sequel and presents the main results. Section 3 is devoted to an illustrative application: two-dimensional linear programming. Section 4 presents the probabilistic analysis of the algorithm. A preliminary analysis of the fundamental recurrence is given in section 5 and a more precise analysis is given in section 6, including the periodicities, the convergence conditions and some numerical experiments. Section 7 concludes the paper.

2 Notations and main results

We split the initial problem of size n into r subproblems of size roughly n/b: we assume that the size of a subproblem is given by $\lceil n/b \rceil$, with a constant factor b. Usually, $r = b$. In this case, there is of course a small overlapping of subproblems, if n is not an exact power of b, but we will neglect this effect in the analysis. $\mathcal{D}(P)$ and $T(P)$ are now random variables. We also use the following notations:

- P_i: ith subproblem.
- $w(P_i)$: value of P_i.
- I_i : $[\![w(P_i)$ is a record$]\!]$, assumed to be iid random variables.
- $\mathcal{D}(P)$: cost of decision: $[\![w(P) \le t]\!]$, for some real t, assumed to be *independent* of $w(P)$ and to depend *only* on the size n of P.
- $T(P)$: cost of computing $w(P)$, assumed to be *independent* of $w(P)$ and to depend *only* on the size n of P.
- $D(n) = \mathbb{E}(\mathcal{D}(P))$, assumed to depend *only* on the size n of P.
- $M(n) := \mathbb{E}(T(P))$, assumed to depend *only* on the size n of P.
- $\lceil n/b \rceil$: size of a subproblem of a problem of size n.
- r: decomposition factor, usually large in most applications (but this is not crucial).
- $D(n) = \kappa_0 n^\lambda (\ln n)^\delta, \lambda \ge 1$. We assume δ is an nonnegative integer. We have chosen this form as it is the most common one in many applications.
- $\mathbb{V}$ denotes the variance.
- $G(n) := \mathbb{V}(\mathcal{D}(P))$, assumed to depend *only* on the size n of P.
- $C_0 := \log_b(a) = \frac{\ln a}{\ln b}$.
- $C_1 := \frac{a}{b^\lambda}$;
- $\chi_k := \frac{2\pi \mathbf{i} k}{\ln b}, \xi_k := \frac{-2\pi \mathbf{i} k + \ln a}{\ln b} = C_0 - \chi_k$.

In section 5 we prove the following preliminary result

Theorem 2.1 *If*

$$\lambda > C_0,$$

then

$$M(n) \sim \kappa_1 n^\lambda (\ln n)^\delta \left[\frac{1}{1 - C_1} - \frac{\delta \ln b}{\ln n} \frac{1}{(1 - C_1)^2} + \mathcal{O}(1/\ln n^2) \right],$$

with

$$\kappa_1 = \frac{r \kappa_0}{b^\lambda}.$$

A more precise analysis is given in section 6, with the Mellin-Perron Lemma. We prove the following results (Δ_1 and $Z(s)$ are also defined in that section):

Theorem 2.2 *The contribution to the mean $M(n)$ due to C_0 is given by*

$$\psi_0(n) - \psi_0(n-1)$$

with

$$\psi_0(n) = \frac{n^{C_0+1}[\Delta_1 + rZ(C_0)/a]}{C_0(C_0+1)\ln b},$$

Theorem 2.3 *The total contribution of $\xi_k, k \neq 0$ to the mean $M(n)$ is given by*

$$\psi_P(n) - \psi_P(n-1)$$

with the Fourier series

$$\psi_P(n) = n^{C_0+1} \sum_{k \neq 0} \alpha_k e^{-\chi_k \ln n},$$

and

$$\alpha_k := \frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k(\xi_k+1)\ln b}.$$

The conditions for absolute convergence of the series, for the shifting of integration line in some integrals and the error terms are also detailed in section 6.

3 Illustration

Chan's method has a lot of applications in the field of geometric optimization, a number of which are given in his paper [Cha99]. We now describe a simple, illustrative example.

One of the most well-known geometric optimization problem is linear programming. Lots of contributions have been made in the field of algorithms for linear programming (or more generally LP-type) problems in the case where the number of constraints is n, but the number of variables is a constant. Many of them are randomized incremental and consider each constraint in turn in random order. This is the case of Seidel's algorithm, for instance [Sei91]. Those algorithms have been improved, then generalized to so-called *LP-type* problems, in which the fundamental structure of linear programs is preserved [SW92].

For the purpose of illustration, we will consider the application of Chan's method to 2-dimensional linear programming. The variables are x and y, and the set of constraints $\mathcal{C}$ is a set of lines. The goal is to find a point (x, y) with minimum y above the lines in $\mathcal{C}$. In order to apply Chan's technique, we need two ingredients: a decision algorithm, and a way of decomposing the problem into subproblems.

The decision problem asks whether the optimum point is above some given horizontal line of equation $y = t$. Intersecting this line with the halfplanes above each line $c \in \mathcal{C}$ yields a set of horizontal halflines. Depending on the slope of the line c, these halflines are either bounded to the left or to the right. The decision algorithm proceeds by computing the maximum of the origins of halflines going to the right, and the minimum of the origins of the lines going to the left. The answer to the decision problem is yes if and only if the maximum is no greater than the minimum. This takes linear deterministic time, and the variance of $\mathcal{D}(P)$ is zero.

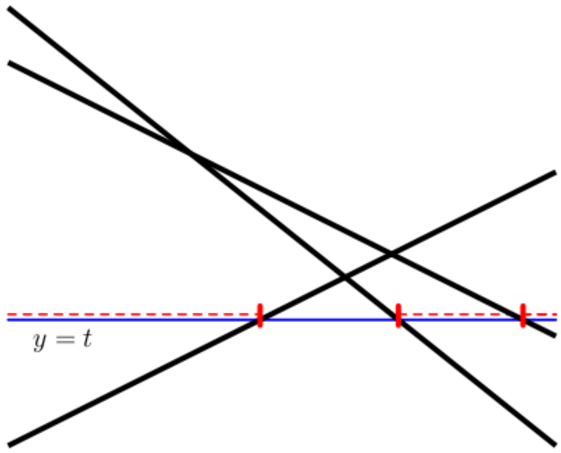


Fig. 1: The decision algorithm: intersections of the constraints with the line of equation $y = t$ yield a set of halflines (dashed). We have $A[i] < t$ iff the halflines have a common point.

Now we can decompose the problem as follows. Partition the set of constraints $\mathcal{C}$ in k equal-sized groups $\{\mathcal{C}_i\}_{i=1}^{k}$. We consider the $r = \binom{k}{2}$ subproblems P_i defined by the unions $\mathcal{C}_i \cup \mathcal{C}_j, i \neq j$. Hence each subproblem has size $\lceil 2n/k \rceil$ and we set $b = k/2$. Since the optimal solution is defined by two constraints, and each pair of constraints is in at least one subproblem, the optimum value is the maximum of the optimal values of the r subproblems. Applying the above analysis with the following parameters: $\kappa_0 = 1$, $\lambda = 1$, $\delta = 0$, $r = \binom{k}{2}$, $b = k/2$, $a = H_r$, $\kappa_1 = r\kappa_0/b^\lambda = k-1$, $\kappa_1 \frac{1}{1-C_1} \simeq \frac{k-1}{1-2\frac{H_r}{k}}$ yields

$$M(n) \sim \frac{k-1}{1-2H_{\binom{k}{2}}/k} n.$$

Note that k must be chosen so that $C_0 < 1$. The first such value is $k = 8$, yielding $C_0 \simeq 0.98$. A better choice for k is one minimizing the constant in front of the linear term. This is $k = 18$, yielding $M(n) \sim 45.145n$.

4 Probabilistic analysis

4.1 Some useful results

- $\mathbb{E}(I_i) = \mathbb{P}(I_i = 1) = \frac{1}{i}$, and $I_i, I_j, i \neq j$ are independent (Records master theorem)
- Let us consider the fundamental recurrence,

$$M(n) = aM(\lceil n/b \rceil) + f(\lceil n/b \rceil), \text{ with } f(n) = rD(n) = r\kappa_0 n^\lambda (\ln n)^\delta, a \geq 1, b > 1, M(1) = 1. \tag{1}$$

If

$$\lambda > C_0,$$

then, by Theorem 2.1

$$M(n) \sim \kappa_2 n^\lambda (\ln n)^\delta, \text{ with } \kappa_2 = \kappa_1 \frac{1}{1-C_1}, \kappa_1 = \frac{r\kappa_0}{b^\lambda}. \quad (2)$$

4.2 The mean

We have

$$\begin{aligned} T(P) &= \sum_1^r I_i T(P_i) + \sum_1^r \mathcal{D}(P_i), \\ \mathbb{E}(T(P)) &= \sum_1^r \frac{1}{i}\mathbb{E}(T(P_i)) + rD(\lceil n/b \rceil), b > 1, \\ M(n) &= aM(\lceil n/b \rceil) + r\kappa_0 \lceil n/b \rceil^\lambda (\ln(\lceil n/b \rceil))^\delta, \quad (3) \\ &\text{with } a = H_r = \sum_1^r \frac{1}{i}. \text{ If } r \text{ is large, then } a \sim \ln r + \gamma. \end{aligned}$$

Hence, if

$$\lambda > C_0, \quad (4)$$

then, by (2)

$$M(n) \sim \kappa_2 n^\lambda (\ln n)^\delta.$$

We usually have $r \sim b$, so $a < b$ and $0 < C_0 < 1$, hence (4) will be satified in this case. But (4) is sufficient in the general case $r \neq b$.

4.3 The variance

We have

$$T(P) - M(n) = \sum_1^r \left[I_i T(P_i) - \frac{1}{i} M(\lceil n/b \rceil) \right] + \sum_1^r [\mathcal{D}(P_i) - D(\lceil n/b \rceil].$$

Set

$$V(n) := \mathbb{E}[T(P) - M(n)]^2.$$

By our independence assumptions, this simplifies as

$$V(n) = \sum_1^r \mathbb{E}\left[I_i T(P_i) - \frac{1}{i} M(\lceil n/b \rceil) \right]^2 + \sum_1^r \mathbb{E}[\mathcal{D}(P_i) - D(\lceil n/b \rceil)]^2.$$

Set

$$G(n) := \mathbb{V}(\mathcal{D}(P)) = \mathbb{E}[\mathcal{D}(P) - D(n)]^2.$$

We have

$$\begin{aligned} \mathbb{E}\left[I_i T(P_i) - \frac{1}{i} M(\lceil n/b \rceil) \right]^2 &= \mathbb{E}\left[\frac{1}{i} T^2(P_i) - \frac{1}{i^2} M^2(\lceil n/b \rceil) \right] \\ = \mathbb{E}\left[\frac{1}{i} T^2(P_i) - \frac{1}{i} M^2(\lceil n/b \rceil) \right] &+ \left(\frac{1}{i} - \frac{1}{i^2} \right) M^2(\lceil n/b \rceil), \end{aligned}$$

and

$$\mathbb{V}(T(P_i)) = \mathbb{E}\left[T^2(P_i) - M^2(\lceil n/b \rceil)\right] = V(\lceil n/b \rceil).$$

Hence,

$$V(n) \sim aV(\lceil n/b \rceil) + (a - H_r^{(2)})M^2(\lceil n/b \rceil) + rG(\lceil n/b \rceil) \sim aV(\lceil n/b \rceil) + a_1 n^{2\lambda}(\ln n)^{2\delta} + rG(\lceil n/b \rceil),$$

with

$$a_1 := (H_r - H_r^{(2)})\kappa_2^2/b^{2\lambda}.$$

If r is large, as $\sum_1^\infty 1/i^2 = \pi^2/6$, we have

$$a_1 \sim (\ln r + \gamma - \pi^2/6)\kappa_2^2/b^{2\lambda}.$$

The behaviour of $V(n)$ depends on the relative growth of $n^{2\lambda}$ and $G(n)$.

The other moments can be computed, with some extra effort.

If r is large, the number of records is known to be Gaussian, and we conjecture that $T(P)$ should be Gaussian.

4.4 Deeper tree

To enforce condition (4), we can use a tree of depth l,(i.e. if we compress l levels of recurrence into one), we derive with large r,

$$M(n) = [\ln r^l + \gamma]M(\lceil n/b^l \rceil) + r^l D(\lceil n/b^l \rceil).$$

$$\text{So } \tilde{a}(l) = \ln r^l + \gamma,$$

$$\text{we must have } \lambda > \frac{\ln[l \ln r + \gamma]}{l \ln b}. \tag{5}$$

But the right-hand side of (5) is decreasing with l, so it is always possible to satisfy (5).

We obtain

$$\begin{aligned}
\tilde{\kappa}_1(l) &= r^l \kappa_0 / b^{l\lambda}, \\
\tilde{\kappa}_2(l) &= r^l \kappa_0 / b^{l\lambda} \frac{1}{1 - \tilde{a}(l)/b^{l\lambda}}, \\
M(l, n) &\sim \tilde{\kappa}_2(l) n^\lambda (\ln n)^\delta.
\end{aligned} \tag{6}$$

For some parameters values, (6) possesses a minimum for some l^*. For instance, for $r = 10, \lambda = 2, b = 2$, we have $l^* = 2$.

5 Preliminary Analysis of (1), proof of theorem 2.1

Proceeding as in Bentley et al. [BHS80], Cormen et al. [CLR90], Flajolet and Sedgewick [FS96], with more precision and details, we first assume that n is an exact power of b. We iterate (1) and obtain

$$M(n) = n^{\log_b a} + \sum_{0}^{\log_b n - 1} a^j f(n/b^{j+1}). \tag{7}$$

Set $k := \log_b n$. We have

$$\sum_0^{k-1} a^j f(n/b^{j+1}) = r\kappa_0 \sum_0^{k-1} a^j \frac{n^\lambda}{b^{(j+1)\lambda}} (\ln n - (j+1)\ln b)^\delta.$$

Set

$$C_1 := a/b^\lambda < 1,$$

by (4). This leads to

$$\begin{aligned} M(n) &= n^{C_0} + \kappa_1 n^\lambda \left[\sum_{j=0}^{\infty} C_1^j (\ln n - (j+1)\ln b)^\delta - C_1^{k-1} \sum_{j=1}^{\infty} C_1^j (\ln n - k\ln b - j\ln b)^\delta \right] \\ &= n^{C_0} + \kappa_1 n^\lambda \left[\sum_{j=0}^{\infty} C_1^j \sum_{u=0}^{\delta} \binom{\delta}{u} (\ln n)^{\delta-u} (-(j+1)\ln b)^u - C_1^{k-1} \sum_{j=1}^{\infty} C_1^j (-j\ln b)^\delta \right]. \end{aligned}$$

We first obtain a dominant contribution

$$\begin{aligned} &\kappa_1 n^\lambda (\ln n)^\delta \left[\frac{1}{1-C_1} - \frac{\delta \ln b}{\ln n} \left(\frac{C_1}{(1-C_1)^2} + \frac{1}{1-C_1} \right) + \mathcal{O}(1/\ln n^2) \right] \\ &= \kappa_1 n^\lambda (\ln n)^\delta \left[\frac{1}{1-C_1} - \frac{\delta \ln b}{\ln n} \frac{1}{(1-C_1)^2} + \mathcal{O}(1/\ln n^2) \right]. \end{aligned} \tag{8}$$

The next contribution, as $C_1^k = n^{C_0}/n^\lambda$, is given by

$$n^{C_0} - n^{C_0}\kappa_1/C_1 (\ln b)^\delta (-1)^\delta \sum_{u=0}^{\delta} \left\{ \begin{matrix} \delta \\ u \end{matrix} \right\} \frac{u! C_1^u}{(1-C_1)^{u+1}} := n^{C_0} C_2, \tag{9}$$

where $\left\{ \begin{matrix} \cdot \\ \cdot \end{matrix} \right\}$ denotes the Stirling number of the second kind. For instance, $\delta = 3$ leads to

$$n^{C_0} \left[1 + \kappa_1 (\ln b)^3 \frac{C_1^2 + 4C_1 + 1}{(1-C_1)^4} \right].$$

If n is not a power of b, we must replace n/b^j by n_j such that $n_j = \lceil n_{j-1}/b \rceil$, but, as shown in [CLR90],

$$n_j < n/b^j + \frac{b}{b-1},$$

so

$$n_j^\lambda < (n/b^j)^\lambda + (n/b^j)^{\lambda-1} C_3 \lambda + \dots,$$

for some constant C_3, and the correction to (8) is less than

$$\kappa_1 n^\lambda (\ln n)^\delta \frac{C_3 \lambda}{1 - bC_1} \frac{1}{n},$$

which is negligible with respect to the dominant contribution. Equ. (8) proves theorem 2.1.

6 A more precise analysis of (1), proof of theorems 2.2, 2.3.

6.1 The analysis

We proceed now in the spirit of Flajolet and Golin [FG94] and Grabner and Hwang [GH05]. See also many other examples in Flajolet et al. [FGKP94] and Panny and Prodinger [PP95]. Let the Dirichlet generating function of Δ_n defined as

$$W(s) := \sum_1^\infty \frac{\Delta_n}{n^s}.$$

We have the Mellin-Perron Lemma: Assume that $W(s)$ converges absolutely for $\Re(s) > \mu$. Then,

$$\psi(n) := \frac{n}{2\pi\mathbf{i}} \int_{\mu+1-\mathbf{i}\infty}^{\mu+1+\mathbf{i}\infty} W(s) n^s \frac{ds}{s(s+1)} = \sum_{k=1}^{n-1} (n-k)\Delta_k = \sum_{k=1}^{n-1} \sum_{l=1}^{k} \Delta_l. \tag{10}$$

Here, we have

$$\begin{aligned}
M(bm) &= aM(m) + rD(m), \\
M(bm+1) &= aM(m+1) + rD(m+1), \\
&\vdots \\
M(bm+b) &= aM(m+1) + rD(m+1), \\
M(bm+b+1) &= aM(m+2) + rD(m+2) \\
&\vdots
\end{aligned}$$

Of course, practically, if $2 \leq m \leq b$, we should divide the set of tasks into m tasks of size 1, but, to simplify, we use expression (1) for all cases.

Hence, with

$$\begin{aligned}
\Delta_m &:= M(m+1) - M(m), \\
U_m &:= D(m+1) - D(m), \\
M(1) &= 1, \\
\Delta_1 &= M(2) - 1 = a - 1 + rD(1), \Delta_2 = 0, \dots, \Delta_{b-1} = 0
\end{aligned}$$

we derive

$$\begin{aligned}
\Delta_{bm} &= a\Delta_m + rU_m, \\
\Delta_{bm+1} &= 0, \\
&\vdots \\
\Delta_{bm+b-1} &= 0, \\
\Delta_{bm+b} &= a\Delta_{m+1} + rU_{m+1}, \\
&\vdots
\end{aligned}$$

So

$$W(s) - \Delta_1 = a\frac{W(s)}{b^s} + r\frac{Z(s)}{b^s},$$

with

$$Z(s) := \sum_1^\infty \frac{U_m}{m^s}.$$

As we will see, $Z(s)$ converges absolutely for $\Re(s) > \lambda$. We have

$$W(s) = \frac{\Delta_1 + rZ(s)/b^s}{1 - a/b^s}.$$

Note that in [FG94] and [GH05], the authors use $\nabla\Delta$ instead of Δ in order to obtain directly $M(n)$ at the end. But, here, it does not seem possible to apply this technique efficiently.

Now (10) leads to

$$\psi(n) = \frac{n}{2\pi\mathbf{i}}\int_{\mu+1-\mathbf{i}\infty}^{\mu+1+\mathbf{i}\infty} \frac{\Delta_1 + rZ(s)/b^s}{1 - a/b^s} n^s \frac{ds}{s(s+1)} = \sum_{k=2}^n [M(k) - 1]. \tag{11}$$

and

$$M(n) = \psi(n) - \psi(n-1).$$

Now

$$\begin{aligned} U_m &= \kappa_0\left[m^\lambda + \lambda m^{\lambda-1} + \binom{\lambda}{2}m^{\lambda-2} + \dots\right]\left[\ln m + 1/m - 1/m^2 + \dots\right]^\delta - \kappa_0 m^\lambda(\ln m)^\delta \\ &= \kappa_0\left[\lambda m^{\lambda-1}(\ln m)^\delta + \delta m^{\lambda-1}(\ln m)^{\delta-1} + \binom{\lambda}{2}m^{\lambda-2}(\ln m)^\delta + \mathcal{O}\left(m^{\lambda-2}(\ln m)^{\delta-1}\right)\right]. \end{aligned}$$

This leads to

$$Z(s) = \kappa_0\left[\lambda(-1)^\delta\zeta^{\delta/}(s-\lambda+1) + \delta(-1)^{\delta-1}\zeta^{(\delta-1)/}(s-\lambda+1) + \binom{\lambda}{2}(-1)^\delta\zeta^{\delta/}(s-\lambda+2) + \dots\right] \tag{12}$$

Now

$$\zeta(s) = \frac{1}{s-1} + \gamma_0 - \gamma_1(s-1) + \dots$$

where γ_i are Stieltjes constants. Hence the dominant singularity of (11) is at $s = \lambda$. So, by standard arguments (see details in Flajolet and Golin [FG94] or Flajolet and Sedgewick [FS06]) , this leads to the following contribution (conditions for shifting the line of integration are investigated below).

The dominant contribution to the mean $M(n)$ is given by

$$\psi_D(n) - \psi_D(n-1),$$

where

$$\begin{aligned} \psi_D(n) := r\kappa_0 n &\left[\frac{\lambda n^\lambda(\ln n)^\delta + \delta n^\lambda(\ln n)^{\delta-1}}{b^\lambda\lambda(\lambda+1)(1-C_1)}\right. \\ &\left.+\frac{\lambda\delta n^\lambda(\ln n)^{\delta-1}}{b^\lambda\lambda(\lambda+1)}\left(-\frac{C_1\ln b}{(1-C_1)^2} - \frac{1}{\lambda(1-C_1)} - \frac{1}{(\lambda+1)(1-C_1)} - \frac{\ln b}{1-C_1}\right)\right]. \end{aligned} \tag{13}$$

$$\psi(n) = \psi_D(n) + \mathcal{O}(n^{\lambda+1}(\ln n)^{\delta-2})$$

After some algebra (we omit the details), this exactly leads to the dominant term (8). Of course, all $(\ln n)^j$ terms can be automatically computed.

The next singularity of (12) is at $s = \lambda - 1$. If $\lambda - 1 > C_0$, the corresponding contribution to $M(n)$, of order $n^{\lambda-1}$ can be computed as previously.

If λ is integer, the last useful terms of (12) correspond to $\zeta^{i/}(s)$ for some integer i. The singularity at $s = 1$ should allow the shift of the line of integration to the left from $1+\varepsilon$ to $1-\varepsilon, \varepsilon > 0$. But the growth property of $Z(s)$ is related to the one of $\zeta(s)$. Let $s = \sigma + \mathbf{i}t$. Then (see Whittaker and Watson [WW27, p. 275])

$$\begin{aligned} \text{if } \sigma > 1, &|\zeta(s)| = \mathcal{O}(1), \\ \text{if } 1-\delta \le \sigma \le 1+\delta, \delta > 0, &|\zeta(s)| = \mathcal{O}(|t|^{1-\sigma} \log |t|), \\ \text{if } 0 < \sigma < 1, &|\zeta(s)| = \mathcal{O}(|t|^{1-\sigma}), \\ \text{if } \sigma < 0, &|\zeta(s)| = \mathcal{O}(|t|^{1/2-\sigma}\zeta(1-s)) = \mathcal{O}(|t|^{1/2-\sigma}). \end{aligned}$$

So, with $\sigma = 1$,

$$\left| \frac{\zeta(s)}{s(s+1)} \right| = \mathcal{O}(|t|^{-2} \log |t|),$$

and the shift is allowed.

Let us now turn to the singularity of (11) with simple poles:

$$\xi_k := \frac{-2\pi \mathbf{i} k + \ln a}{\ln b} = C_0 - \chi_k,$$

with

$$\chi_k := \frac{2\pi \mathbf{i} k}{\ln b},$$

To simplify the analysis, we will only consider the case $\delta = 0$. The general more complicated case can be analyzed similarly. We are now interested in the case $\sigma = C_0 - \varepsilon, \varepsilon > 0$.

- If $\lambda = 1$,

  $$|\zeta(s - \lambda + 1)| = \mathcal{O}(|t|^{1-C_0+\varepsilon}),$$

  the integrand of (11) is of order $\mathcal{O}(|t|^{1-C_0+\varepsilon-2}) = \mathcal{O}(|t|^{-1-C_0+\varepsilon})$ and we can safely shift the line of integration from $C_0 + \varepsilon$ to $C_0 - \varepsilon$.

- If $\lambda = 2$,

  - if $0 < C_0 - \varepsilon < 1$

    $$|\zeta(s - \lambda + 1)| = \mathcal{O}(|t|^{1/2-C_0+\varepsilon+1}),$$

    the integrand is of order $\mathcal{O}(|t|^{-1/2-C_0+\varepsilon})$.

    If $1/2 < C_0 - \varepsilon < 1$ then we can safely shift the line of integration to $C_0 - \varepsilon$.

    If $0 < C_0 - \varepsilon < 1/2$, we *do not* have convergence of (11) for $s \in [C_0 - \varepsilon + \mathbf{i}R \ldots C_0 - \varepsilon + \mathbf{i}\infty] \cup [C_0 - \varepsilon - \mathbf{i}\infty \ldots C_0 - \varepsilon - \mathbf{i}R]$, R large and we *cannot* shift the line of integration.

– if $1 < C_0 - \varepsilon < 2$, the integrand is of order $\mathcal{O}(|t|^{-C_0+\varepsilon})$ and we can safely shift the line of integration to $C_0 - \varepsilon$.

Let us now consider the contribution of the poles ξ_k. For $k = 0$, we get the following result, which proves theorem 2.2:

The contribution to the mean $M(n)$ due to C_0 is given by

$$\psi_0(n) - \psi_0(n-1)$$

with

$$\psi_0(n) = \frac{n^{C_0+1}[\Delta_1 + rZ(C_0)/a]}{C_0(C_0+1)\ln b}, \tag{14}$$

This gives

$$\psi_0(n) - \psi_0(n-1) \sim C_4 n^{C_0} + \mathcal{O}(n^{C_0-1}), \text{ with} \tag{15}$$

$$C_4 = \frac{n^{C_0}[\Delta_1 + rZ(C_0)/a]}{C_0 \ln b}$$

For $k \neq 0$, we obtain the following result, which proves theorem 2.3:

The total contribution of $\xi_k, k \neq 0$ to the mean $M(n)$ is given by

$$\psi_P(n) - \psi_P(n-1)$$

with the Fourier series

$$\psi_P(n) = n^{C_0+1} \sum_{k \neq 0} \alpha_k e^{-\chi_k \ln n},$$

and

$$\alpha_k := \frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k(\xi_k+1)\ln b}.$$

The growth order of the coefficients depends on λ.

- If $\lambda = 1$,

$$|\alpha_k| = \mathcal{O}(|k|^{-1-C_0}),$$

  and the series is absolutely convergent

- If $\lambda = 2$,

  – if $0 < C_0 < 1$, then

$$|\alpha_k| = \mathcal{O}(|k|^{-1/2-C_0}).$$

    If $C_0 > 1/2$, the series is absolutely convergent.
    If $C_0 < 1/2$, the series is *not* absolutely convergent.

  – if $1 < C_0 < 2$, then

$$|\alpha_k| = \mathcal{O}(|k|^{-C_0}),$$

    the series is absolutely convergent.

In the case of absolute convergence, we have uniform convergence of the Fourier series and $\psi_P(n)$ is continuous, non differentiable, and periodic, as in the average case analysis of Flajolet and Golin [FG94].

We must now carefully analyze the error term on $\psi_P(n) - \psi_P(n-1)$ due to the shift of the integration line to $C_0 - \varepsilon$. Indeed, we want to obtain an error at most $\mathcal{O}(n^{C_0-\gamma}), \gamma > 0$, as (15) gives a contribution of order n^{C_0}. We start from (11), this gives

$$\psi(n) - \psi(n-1) = \frac{1}{2\pi\mathbf{i}} \int_{-\infty}^{+\infty} n^{\tau+\mathbf{i}t} \phi(t) \frac{\Delta_1 + rZ(v)/b^v}{(1-a/b^v)v(v+1)} dt,$$

with

$$\begin{aligned} \tau &= C_0 + 1 - \varepsilon, v = C_0 - \varepsilon + \mathbf{i}t, \\ \phi(t) &:= 1 - (1 - 1/n)^{\tau+\mathbf{i}t}. \end{aligned}$$

We must have

$$\int |\phi(t)| \left| \frac{\Delta_1 + rZ(v)/b^v}{(1-a/b^v)v(v+1)} \right| dt = \mathcal{O}(1/n^{1-\beta}),$$

with $0 < \beta < 1$, so that $\gamma = \varepsilon - \beta > 0$. Assume

$$\left| \frac{\Delta_1 + rZ(v)/b^v}{(1-a/b^v)v(v+1)} \right| = \mathcal{O}\left(\frac{1}{|t|^{1+\nu}} \right).$$

Possible values of ν can be derived from above. After some algebra, we get

$$|\phi(t)| = (2(1 - \cos[\ln(1-1/n)t])(1 - \tau/n) + \tau^2/n^2)^{1/2}.$$

We split the interval $t \in [0 \ldots \infty]$ into $[0 \ldots n^{1-\eta}]$,$[n^{1-\eta} \ldots \infty]$,$0 < \eta < 1$ (similarly for $t \in [-\infty \ldots 0]$). This gives a first bound

$$C_5 \int_0^{n^{1-\eta}} \frac{t}{n} \frac{dt}{t^{1+\nu}} = \frac{C_5}{n^{\eta+(1-\eta)\nu}} \text{ for some constant } C_5,$$

hence a first parameter

$$\beta_1 = 1 - \eta - (1-\eta)\nu.$$

The other bound is given by

$$C_6 \int_{n^{1-\eta}}^{\infty} \frac{dt}{t^{1+\nu}} = \frac{C_6}{n^{(1-\eta)\nu}} \text{ for some constant } C_6,$$

hence a second parameter

$$\beta_2 = 1 - (1-\eta)\nu.$$

Note that $\beta_1 < \beta_2$, so we must have $\beta_2 < \varepsilon$, or $(1-\eta)\nu > 1 - \varepsilon$. We can choose η as small as we want, so we have the constraint $\nu > 1 - \varepsilon$. So

- If $\lambda = 1$, $\nu = C_0 - \varepsilon$, but this would impose $C_0 > 1$, which is not allowed,

- If $\lambda = 2$,

  If $1/2 < C_0 - \varepsilon < 1$, $\nu = -1/2 + C_0 - \varepsilon$, so we must have $C_0 > 3/2$.

  If $0 < C_0 - \varepsilon < 1/2$, recall that we do not have convergence of ψ_P.

  If $1 < C_0 - \varepsilon < 2$, $\nu = C_0 - \varepsilon - 1$, but this would impose $C_0 > 2$, which is not allowed.

We must now analyze

$$\psi_P(n) - \psi_P(n-1).$$

This would give

$$\begin{aligned}
\tilde{\psi}_P(n) = \psi_P(n) - \psi_P(n-1) &= \sum_{k\neq 0} \left[n^{\xi_k+1} - (n-1)^{\xi_k+1}\right] \frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k(\xi_k+1)\ln b} \\
&= \sum_{k\neq 0} \left[n^{C_0+1}e^{-\chi_k \ln n} - (n-1)^{C_0+1}e^{-\chi_k \ln(n-1)}\right] \frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k(\xi_k+1)\ln b} \\
&\sim \sum_{k\neq 0} \left[n^{C_0+1}e^{-\chi_k \ln n} - n^{C_0+1}\left(1 - \frac{C_0+1}{n}\right)e^{-\chi_k \ln n}\left(1 + \frac{\chi_k}{n}\right)\right] \frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k(\xi_k+1)\ln b} \\
&\sim n^{C_0} \sum_{k\neq 0} e^{-\chi_k \ln n}[C_0 + 1 - \chi_k] \frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k(\xi_k+1)\ln b} + \mathcal{O}(n^{C_0-1}) \\
&= n^{C_0}\psi_P^{(0)}(n) + \mathcal{O}(n^{C_0-1}),
\end{aligned}$$

with

$$\psi_P^{(0)}(n) := \sum_{k\neq 0} e^{-\chi_k \ln n}\left[\frac{\Delta_1 + rZ(\xi_k)/a}{\xi_k \ln b}\right]. \tag{16}$$

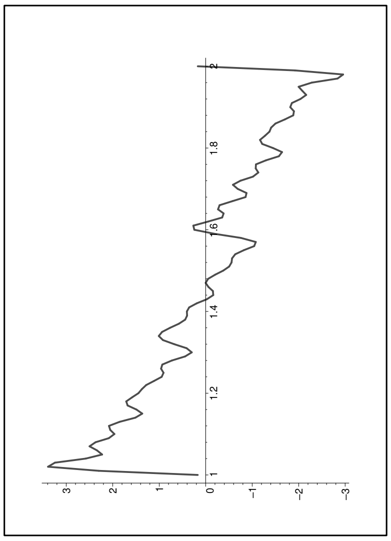
One one side, $|\Delta_1/\xi_k| = \mathcal{O}(1/k)$, on the other side, the contribution of $Z(\xi_k)$ is such that

- If $\lambda = 1$, the coefficients of this series are of order k^{-C_0} and we have *no* absolute convergence if $C_0 < 1$.
- If $\lambda = 2$ and $0 < C_0 < 1$ the coefficients of this series are of order $k^{1/2-C_0}$ and we do not have absolute convergence.
- If $\lambda = 2$ and $1 < C_0 < 2$ the coefficients of this series are of order k^{1-C_0} and we do not have absolute convergence.

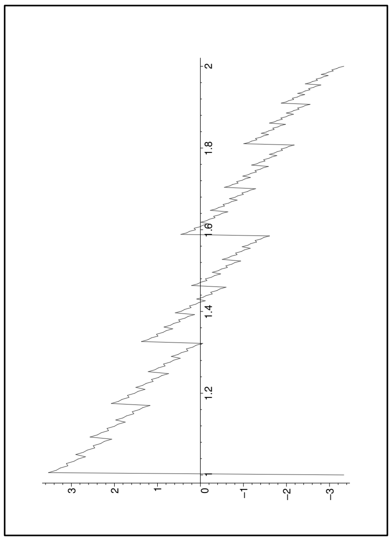
Nevertheless, in some cases, the numerical experiments give a rather good fit: see the next subsection.

6.2 Numerical experiments

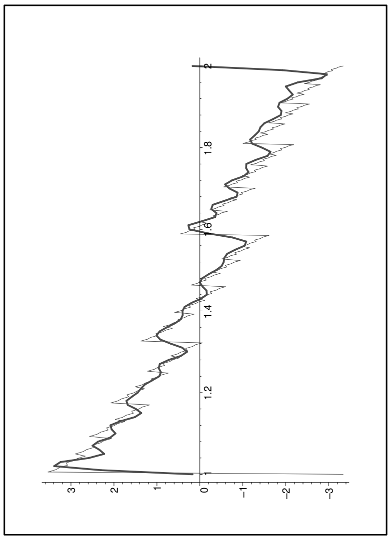
We have made several numerical experiments. They are not related to particular examples, but only chosen on order to give a flavour of different possible behaviours.



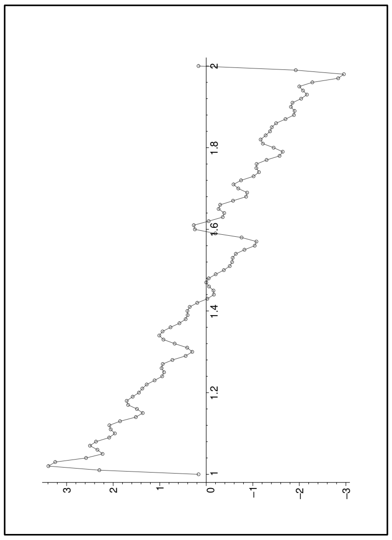
(a) $\psi_P^{(0)}(i)$ in function of $\log_b i$



(b) $(M(i) - (8))/(0.9 i^{C_0})$ in function of $\log_b i$

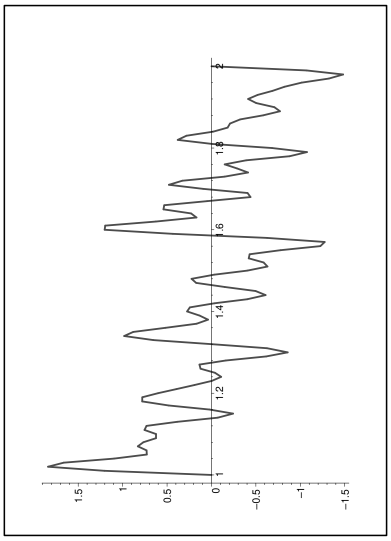


(c) $\psi_P^{(0)}(i)$ and $(M(i) - (8))/(0.9 i^{C_0})$ in function of $\log_b i$

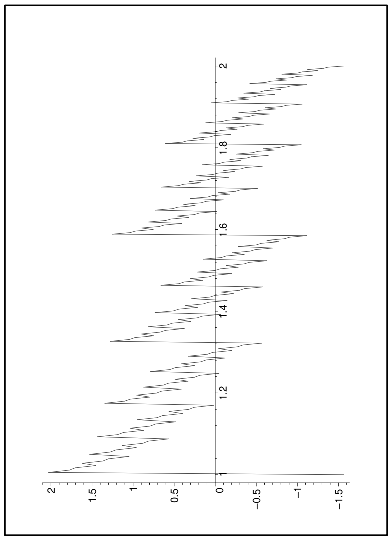


(d) $(\psi_P(i) - \psi_P(i-1))/i^{C_0}$ (circle) and $\psi_P^{(0)}(i)$ (line) in function of $\log_b i$

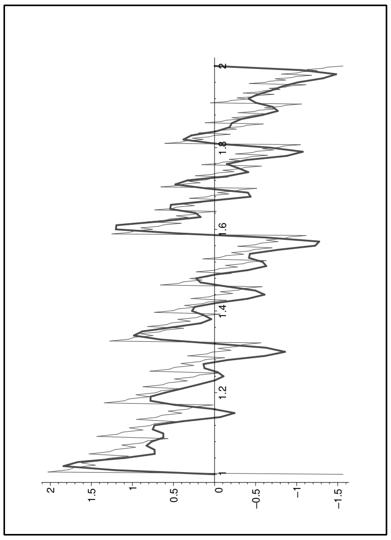
Fig. 2: Case 1



(a) $\psi_P^{(0)}(i)$ in function of $\log_b i$

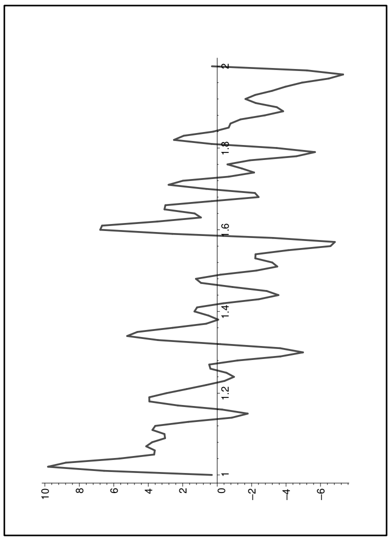


(b) $(M(i)-(8)-16i^{C_0})/(5/2i^{C_0})$ in function of $\log_b i$

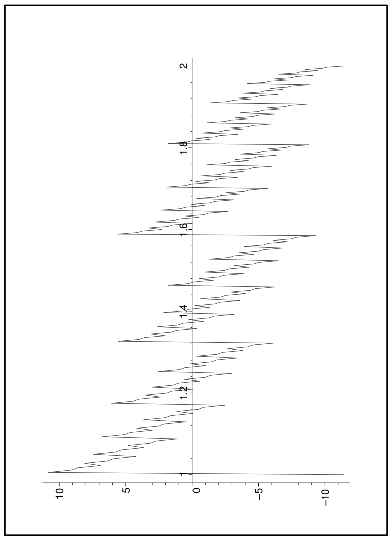


(c) $\psi_P^{(0)}(i)$ and $(M(i)-(8)-16i^{C_0})/(5/2i^{C_0})$ in function of $\log_b i$

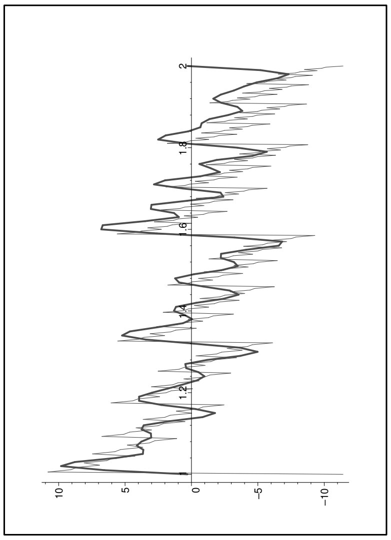
Fig. 3: Case 2



(a) $\psi_P^{(0)}(i)$ in function of $\log_b i$



(b) $(M(i) - (8) - 48i^{C_0})/(2i^{C_0})$ in function of $\log_b i$



(c) $\psi_P^{(0)}(i)$ and $(M(i) - (8) - 48i^{C_0})/(2i^{C_0})$ in function of $\log_b i$

Fig. 4: Case 3

Case 1 : $\kappa_0 = 1, r = 2, \lambda = 1, b = 2, \delta = 0$. In that case, ψ_P is convergent, the shift is allowed, but the error term should be too large. This leads to $a = 3/2, \kappa_1 = 1, C_0 = 0.58\ldots, C_1 = 0.75\ldots, C_4 = 0.04\ldots, C_2 = -3$. We let $n = 1\ldots 2^{17}$. $\psi_P^{(0)}(i)$ in function of $\log_b i$ is given in Figure 2(a). $(M(i) - (8))/(0.9i^{C_0})$ in function of $\log_b i$, $i = 2^{16}\ldots 2^{17}$, is given in Figure 2(b), the constants $0, 0.9$ are chosen to give the best fit, but we have no justification up to now. Both functions are given in Figure 2(c). The fit is rather good. A comparison between $(\psi_P(n) - \psi_P(n-1))/n^{C_0}$ and $\psi_P^{(0)}(n)$ is given in Figure 2(d), they are undistinguishable.

Case 2 : $\kappa_0 = 1, r = 2, \lambda = 1, b = 2, \delta = 1$ leads to $a = 3/2, \kappa_1 = 1, C_0 = 0.58\ldots, C_1 = 0.75\ldots, C_4 = 20.1\ldots, C_2 = 12.1\ldots$. We let $n = 1\ldots 2^{17}$. $\psi_P^{(0)}(i)$ in function of $\log_b i$ is given in Figure 3(a).
$(M(i) - (8) - 16i^{C_0})/(5/2i^{C_0})$ in function of $\log_b i$, $i = 2^{16}\ldots 2^{17}$, is given in Figure 3(b), the constants $16, 5/2$ are chosen to give the best fit, but again we have no justification up to now. Both functions are given in Figure 3(c).

Case 3 : $\kappa_0 = 1, r = 10, \lambda = 2, b = 2, \delta = 1$ leads to $a = 2.92\ldots, \kappa_1 = 5/2, C_0 = 1.55\ldots, C_1 = 0.73\ldots, C_4 = 32.7\ldots, C_2 = 25.1\ldots$. We let $n = 1\ldots 2^{17}$.$\psi_P^{(0)}(i)$ in function of $\log_b i$ is given in Figure 4(a). $(M(i) - (8) - 48i^{C_0})/(2i^{C_0})$ in function of $\log_b i$, $i = 2^{16}\ldots 2^{17}$, is given in Figure 4(b), the constants $48, 2$ are chosen to give the best fit, but again we have no justification up to now. Both functions are given in Figure 4(c).

7 Conclusion

Two problems remain open:

- Even when we have no proved absolute convergence for $\psi_P^{(0)}(n)$, the series appear to converge in some cases, with a rather good behaviour either with respect to $\psi_P(n) - \psi_P(n-1)$ or with respect to the observed $M(i)$. What are the specific conditions for that?

- How to derive the constants leading to the best fits?

H.K. Hwang suggested to use the subdivision

$$\lfloor (n+j)/b \rfloor, j = 0\ldots b-1,$$

which indeed covers exactly $[1\ldots n]$. But now the records probabilities are different from $1/i$, they depend on the size of each subproblem (two different sizes), which themselves depend on n and j. We do not have a constant a anymore. And we rarely know the distribution of $T(P)$.

Acknowledgments

The pertinent remarks of the referees led to corrections and improvement in the presentation. H.K. Hwang provided many useful suggestions.

References

[BHS80] J. L. Bentley, D. Haken, and J. B. Saxe. A general method for solving divide-and conquer recurrences. *SIGACT News*, 15:36–44, Fall 1980.

[Cha99] T. M. Chan. Geometric applications of a randomized optimization technique. *Discrete and Computational Geometry*, 22:547–567, 1999.

[CLR90] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1990.

[FG94] P. Flajolet and M. Golin. Mellin transforms and asymptotics: the mergesort recurrence. *Acta Informatica*, 31:673–696, 1994.

[FGKP94] P. Flajolet, P. Grabner, P. Kirschenhofer, and H. Prodinger. Mellin transforms and asymptotics: Digital sums. *Theoretical Computer Science*, 123(2):291–314, 1994.

[FS96] P. Flajolet and R. Sedgewick. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, 1996.

[FS06] P. Flajolet and B. Sedgewick. *Analytic combinatorics*. To appear, 2006. available at http://algo.inria.fr/flajolet/Publications/.

[GH05] P. J. Grabner and H-K. Hwang. Digital sums and divide-and-conquer recurrences: Fourier expansions and absolute convergence. *Constructive Approximation*, 21(2):149–179, 2005.

[PP95] W. Panny and H. Prodinger. Bottom-up mergesort. A detailed analysis. *Algorithmica*, 14:340–354, 1995.

[Sei91] R. Seidel. Small-dimensional linear programming and convex hulls made easy. *Discrete & Computational Geometry*, 6(5):423–434, 1991.

[SW92] M. Sharir and E. Welzl. A combinatorial bound for linear programming and related problems. In *Proceedings of the 9th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 577 of *Lecture Notes in Computer Science*, pages 569–579. Springer, 1992.

[WW27] E.T. Whittaker and G.N. Watson. *A Course of Modern Analysis*. Cambridge University press (reprinted 1999), 1927.