

## Probabilistic recurrence relations revisited

Shiva Chaudhuri<sup>a,\*</sup>, Devdatt Dubhashi<sup>b,1</sup>

<sup>a</sup> *Max-Planck-Institute für Informatik, Im Stadtwald, 66123 Saarbrücken, Germany*

<sup>b</sup> *BRICS<sup>2</sup>, Department of Computer Science, University of Aarhus, Denmark*

---

### Abstract

The performance attributes of a broad class of randomised algorithms can be described by a recurrence relation of the form

$$T(x) = a(x) + T(H(x)),$$

where  $a$  is a function and  $H(x)$  is a random variable. For instance,  $T(x)$  may describe the running time of such an algorithm on a problem of size  $x$ . Then  $T(x)$  is a random variable, whose distribution depends on the distribution of  $H(x)$ . To give high probability guarantees on the performance of such randomised algorithms, it suffices to obtain bounds on the tail of the distribution of  $T(x)$ . Karp derived tight bounds on this tail distribution, when the distribution of  $H(x)$  satisfies certain restrictions. In this paper, we give a simple proof of bounds similar to that of Karp using standard tools from elementary probability theory, such as Markov's inequality, stochastic dominance and a variant of Chernoff bounds applicable to unbounded geometrically distributed variables. Further, we extend the results, showing that similar bounds hold under weaker restrictions on  $H(x)$ . As an application, we derive performance bounds for an interesting class of algorithms that was outside the scope of the previous results.

---

### 1. Introduction and motivation

Consider a randomised algorithm that works as follows: on an input of size  $x$ , it performs  $a(x)$  work to generate a subproblem of size  $H(x)$  (where  $H(x)$  is a random variable taking values in  $[0, x]$ , whose distribution depends on the algorithm) and then solves the subproblem recursively. Then, the running time of the algorithm may be described by the (probabilistic) recurrence relation

$$T(x) = a(x) + T(H(x)). \quad (1)$$

---

\* Corresponding author. E-mail: shiva@mpi-sb.mpg.de. Supported by the ESPRIT Basic Research Actions Program of the EC under contract No. 7141 (project ALCOM II).

<sup>1</sup> Work done while at the Max-Planck-Institute für Informatik supported by the ESPRIT Basic Research Actions Program of the EC under contract No. 7141 (project ALCOM II).

<sup>2</sup> Basic Research in Computer Science, Centre of the Danish National Research Foundation.

Hence,  $T(x)$  is a random variable whose distribution depends on the distribution of  $H(x)$ . The performance of the randomised algorithm can be described in terms of certain statements on the distribution of this random variable. For instance, one may compute the expected running time, or we may give more precise information on the tail of the distribution of this random variable.

Such a recursion also describes succinctly, the size or structure of certain randomly generated combinatorial structures, for instance, the structure of random permutations of objects or the sizes of cliques generated by a random greedy process.

In the literature, the analysis of many randomised algorithms fit this framework (see Section 2 below for some typical examples, or numerous ones exhibited in [3]). However, their analyses are frequently carried out by disparate ad hoc techniques. Karp [3] recognised that all these algorithms can be analysed uniformly in the above framework and gave general theorems which could be applied in the fashion of a “cook-book” substitution to give the desired performance guarantees on the algorithms. To state the hypothesis and results of Karp, we introduce some notations and definitions.

In the following,  $T(x)$  satisfies Eq. (1), where  $a$  is a fixed function,  $H(x)$  is a random variable taking values in  $[0, x]$ , and  $E[H(x)] \leq m(x)$ , for a fixed function,  $m$ , satisfying  $0 \leq m(x) \leq x$ . Also,  $a$  and  $m$  are non-decreasing functions. The equation

$$\tau(x) = a(x) + \tau(m(x)) \quad (2)$$

can be regarded as the deterministic counterpart of the probabilistic recurrence (1). Intuitively, it is an equation governing the expected values. Whenever this equation has a solution, it has a unique least non-negative solution  $u(x)$ , given by  $u(x) = \sum_{i \geq 0} a(m^{(i)}(x))$ , where we define  $m^{(0)}(x) := x$  and  $m^{(i+1)}(x) := m(m^{(i)}(x))$  for  $i \geq 0$ . Karp proved [3, Theorems 1.1, 1.2].

**Theorem 1** (Karp [3]). *Consider the probabilistic recurrence (1). Let  $m(x)$  and  $a(x)$  be continuous functions satisfying (1)  $m(x)/x$  is nondecreasing and (2)  $a(x)$  is strictly increasing on  $\{x \mid a(x) > 0\}$ .*

- *Let  $b$  be the terminating point of the recurrence with  $a(x) = 0$ ,  $x < b$  and  $a(x) = 1$ ,  $x \geq b$ . Let  $c_t := \min\{x \mid u(x) \geq t\}$ . Then, for every real  $x$  and every  $l \geq 1$ ,*

$$\Pr[T(x) \geq u(x) + l] \leq \left(\frac{m(x)}{x}\right)^{l-1} \frac{m(x)}{c_{u(x)}}.$$

- *Then for every positive real  $x$  and every positive integer  $w$ ,*

$$\Pr[T(x) \geq u(x) + wa(x)] \leq (m(x)/x)^w.$$

This theorem gives very precise bounds on the performance attributes of algorithms. It also admits a fine-tuned tradeoff between the relaxation permitted in the running time and the high probability guarantee. However, the method used to prove the result, while ingenious, offers no intuition about why the result holds, and the proofs are somewhat difficult to follow. Further, the conditions (1) and (2) and continuity in the

theorem are technical artifices introduced by the methods of proof. In particular, for weaker conditions on  $m(x)/x$ , very similar bounds hold, as shown in Theorem 2, below. Specifically, condition (1) prevents the application of Theorem 1 whenever  $m(x)$  grows more slowly than  $x$ . For instance, it prevents a direct application of Karp's results to an interesting class of randomized algorithms based on a probabilistic strategy called the *Rödl Nibble* [2].

We give an alternative analysis that yields comparable, although somewhat weaker, bounds. We essentially reduce the problem to the analysis of waiting times between successes in a sequence of Bernoulli trials. The reduction is obtained using three major components: Markov's inequality, stochastic dominance and a variant of the Chernoff bound applicable to *unbounded* but geometrically distributed random variables. The structure of the proof is thus strongly intuitive, reflecting the behaviour of the randomised process. It is also quite general, in that when  $m(x)/x$  is non-decreasing, it yields bounds comparable to Theorem 1, and when  $m(x)/x$  satisfies a weaker condition, the same proof yields exponentially decreasing bounds. In particular, it covers the case of the Rödl Nibble algorithms mentioned above. Our results, by comparison with Theorem 1 above are:

**Theorem 2.** *Let  $\Delta = \Delta(x) := \max_{b \leq y \leq x} (m(y)/y)$ , where  $b$  is the terminating point of the recurrence (1). Then,*

$$\Pr[T(x) \geq u(x) + la(x)] \leq C(\Delta(x))^{(l-1)/2},$$

where  $C := C(a, m, x)$  is independent of  $l$ . The exact form of  $C(a, m, x)$  is given in Corollary 5.

**Remark.** Notice that there are no assumptions on  $a$  and  $m$  other than that they are non-decreasing. Further the case  $a(x) = [x \geq b]$  is covered by the same statement.

In the case that  $m(x)/x$  is non-decreasing,  $\Delta(x) = m(x)/x$  and we get the following bounds to compare with those of Theorem 1:

$$\Pr[T(x) \geq lu(x)] \leq C(a, m, x)(m(x)/x)^{(l-1)/2}.$$

Our bounds are not quite as precise as Karp's; ours are weaker by constant factors in the exponent. However, in many applications, these constants are not crucial. We do not have results for the case considered by Karp when more than one recursive call is made, e.g. Quicksort.

## 2. An example application

Our theorems can be applied in a "cook-book" fashion to yield high-probability statements about the running time of randomised algorithms or the size and structure of randomly generated combinatorial structures, as in the examples in Karp's paper

[3, Section 2]. Our bounds are somewhat weaker than those of Karp as is evident in the statement of the two bounds in the previous section.

However, we now give an example of an application to a class of problems where Karp's bound does not apply owing to the fact that the assumption that  $m(x)/x$  is non-decreasing in Karp's Theorem does not apply.

### 2.1. Edge colouring of graphs

In [2], a randomised distributed edge-colouring algorithm is described, based on a probabilistic strategy called the *Rödl Nibble*. The algorithm proceeds in stages. At each stage, each vertex has available to it, a *palette* of colours. Each vertex then chooses a small subset ("nibble") of incident edges to colour, and tentatively assigns them a colour chosen uniformly and independently at random from its current palettes. The colour becomes final if it is admissible at the other endpoint and there are no other edges whose tentative colours conflict with it. The edges which are successfully coloured are then deleted and the palettes are correspondingly updated. It can be shown in that the palette sizes (and hence the vertex degrees) obey the following *decay law*: If  $\Delta_k$  denotes the (expected) palette size at stage  $k$  and  $\Delta$  is the maximum degree of the input graph, then

$$\Delta_{k+1} \leq \exp\left(-\frac{\alpha}{\Delta} \Delta_k\right) \Delta_k.$$

Hence, for the number of rounds of the distributed protocol, we have a recurrence of the form

$$T(n) = 1 + T(H(n))$$

with  $E[H(n)] \leq \exp(-(\alpha/\Delta)n)n$ . In this example, the function  $m(x)/x = \exp(-(\alpha/\Delta)n)$  is a *decreasing* function and hence Karp's Theorem 1 is inapplicable. Applying our theorem, and stopping the recurrence when  $\Delta_k = \lambda\Delta$ , as is needed in the algorithm in [2], we get the tail probability bounds:

$$\Pr\left[T(\Delta) > \frac{1}{\alpha\lambda} \ln\left(\frac{1}{\lambda\Delta}\right)\right] \leq C \exp(-\alpha\lambda(l-1)/2).$$

### 3. Some probabilistic lemmas

A set of variables  $X_1, \dots, X_n$  is *stochastically dominated* by a set of variables  $Y_1, \dots, Y_n$  if

$$E[f(X_1, \dots, X_n)] \leq E[f(Y_1, \dots, Y_n)]$$

for all non-decreasing functions  $f$  (that is, non-decreasing in each argument) [6]. We will use the following criterion for stochastic dominance [6, Section 17C]:

**Proposition 3.** Let  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_n$  be random variables such that for all real  $t$ ,

$$\Pr[X_1 > t] \leq \Pr[Y_1 > t],$$

and for each  $i \geq 1$  and for all  $x_1 \leq y_1, \dots, x_i \leq y_i$ ,

$$\Pr[X_{i+1} > t \mid X_1 = x_1, \dots, X_i = x_i] \leq \Pr[Y_{i+1} > t \mid Y_1 = y_1, \dots, Y_i = y_i].$$

Then the variables  $X_1, \dots, X_n$  are stochastically dominated by the variables  $Y_1, \dots, Y_n$ .

The following lemma, gives Chernoff-like bounds for the sum of random variables with a geometric distribution. We note that a similar lemma with  $p = \frac{1}{2}$  was proved in [1], using generating functions. Before we extract the bounds that are actually useful to us in this paper, we give a direct simple proof of *exact* bounds on a somewhat more general version that may be useful in other applications.

**Lemma 4.** Let  $\mathbf{Z} := (Z_1, \dots, Z_n)$  be a collection of independent random variables which are geometrically distributed in the following way: for each  $i$ ,  $1 \leq i \leq n$ , there exist non-negative reals  $z_i$  such that for any positive integer  $l$ ,

$$\Pr[Z_i = lz_i] = (1 - p)p^{l-1}$$

for a real  $p$ ,  $0 < p < 1$ . Then, letting  $Z := Z_1 + \dots + Z_n$ , and  $z := z_1 + \dots + z_n$ ,

1. If  $z_i = z^*$  for each  $i$ , then for any  $t \geq 0$ ,

$$\Pr[Z \geq t] = \left(\frac{1-p}{p}\right)^n \sum_{k \geq \frac{t}{z^*}} p^k \binom{k+n-1}{n}.$$

2. If  $z_1 > z_i$  for  $i > 1$ , then for any  $t \geq 0$ ,

$$\Pr[Z \geq t] = F(p; n; z_1, \dots, z_n) p^{t/z_1 - 1},$$

where

$$F(p; n; z_1, \dots, z_n) := \prod_{1 < i \leq n} \frac{1-p}{1-p^{1-z_i/z_1}} p^{-z_i/z_1}.$$

**Proof.** We have that

$$\begin{aligned} \Pr[Z \geq t] &= \sum_{t_1 + \dots + t_n \geq t} \Pr[Z_1 = t_1] \cdots \Pr[Z_n = t_n] \\ &= \sum_{l_1 z_1 + \dots + l_n z_n \geq t} \Pr[Z_1 = l_1 z_1] \cdots \Pr[Z_n = l_n z_n] \\ &= \sum_{l_1 z_1 + \dots + l_n z_n \geq t} \prod_{1 \leq i \leq n} (1-p)p^{l_i - 1} \\ &= \left(\frac{1-p}{p}\right)^n \sum_{l_1 z_1 + \dots + l_n z_n \geq t} p^{l_1 + \dots + l_n}. \end{aligned} \tag{3}$$

1. If  $z_i = z^*$  for each  $1 \leq i \leq n$ , then

$$\begin{aligned} \sum_{l_1 z_1 + \dots + l_n z_n \geq t} p^{l_1 + \dots + l_n} &= \sum_{(l_1 + \dots + l_n) z^* \geq t} p^{l_1 + \dots + l_n} \\ &= \sum_{k \geq t/z^*} \sum_{l_1 + \dots + l_n = k} p^{l_1 + \dots + l_n} \\ &= \sum_{k \geq t/z^*} p^k \binom{k+n-1}{n}. \end{aligned}$$

Substituting this into Eq. (3) gives the first part.

2. If  $z_1 > z_i$  for  $i > 1$ , then we have

$$\begin{aligned} \sum_{l_1 z_1 + \dots + l_n z_n \geq t} p^{l_1 + \dots + l_n} &= \sum_{l_n \geq 1} \dots \sum_{l_2 \geq 1} \sum_{l_1 \geq (t - (l_2 z_2 + \dots + l_n z_n) / z_1)} p^{l_1 + \dots + l_n} \\ &= \sum_{l_n \geq 1} \dots \sum_{l_2 \geq 1} \sum_{m \geq 0} p^{m + (t - (l_2 z_2 + \dots + l_n z_n) / z_1) + l_2 + \dots + l_n} \\ &= \sum_{l_n \geq 1} p^{(1 - z_n/z_1) l_n} \dots \sum_{l_2 \geq 1} p^{(1 - z_2/z_1) l_2} \sum_{m \geq 0} p^m p^{t/z_1} \\ &= \frac{p^{1 - z_n/z_1}}{1 - p^{1 - z_n/z_1}} \dots \frac{p^{1 - z_2/z_1}}{1 - p^{1 - z_2/z_1}} \frac{1}{1 - p} p^{t/z_1}. \end{aligned}$$

Substituting into Eq. (3) and simplifying gives the second part.  $\square$

The form actually useful to us here is obtained by substituting  $t := z + lz_1$  into the second part.

**Corollary 5.** For the variables  $Z, Z_i$  with  $z_1 > z_2 > \dots$ , we have for  $l \geq 1$ ,

$$\Pr[Z \geq z + lz_1] \leq Cp^{l-1},$$

where  $C := C(z_1, \dots, z_n, p) = \prod_{i>1} (1 - p) / (1 - p^{1 - z_i/z_1})$ .

### 4. Proof of the main theorem

#### 4.1. Intuition

The probabilistic recurrence (1) defines a sequence of random variables  $X_0, X_1, \dots$  with

$$\begin{aligned} X_0 &:= x, \\ X_{i+1} &:= H(X_i), \quad i \geq 0. \end{aligned}$$

One can think of  $X_i$  is the current value of the problem size at stage  $i$  of the recurrence.

Let  $x =: y_0 > y_1 > \dots > 1$  be a sequence of reals to be specified shortly. Divide the process into *phases*, where phase  $i$  consists of those stages  $j$ , at which the random variable  $X_j$  lies between  $y_i$  and  $y_{i+1}$ ,  $i \geq 0$ . Define a “success” in phase  $i$  to be

a passage from the interval  $[y_i, y_{i+1})$  to the next one, and denote by  $p_i$ , the corresponding probability of success. Thus phase  $i$  consists of the stages between one “success” and the next one. If phase  $i$  lasts  $k$  stages, we have that the work done in phase  $i$ ,  $S_i := \sum_{y_{i+1} < X_j \leq y_i} a(X_j) \leq ka(y_i)$ . If the experiments were independent, we would have that  $\Pr[k \geq l] \leq (1 - p_i)^{l-1}$ , hence also  $\Pr[S_i \geq la(y_i)] \leq (1 - p_i)^{l-1}$ . Now, once again, assuming the experiments in different phases are independent, we can bound the probability that the sum of the  $S_i$ s exceeds any given value by employing Corollary 5 giving probability bounds for the sum of geometrically distributed random variables, to get the required tail probability bounds.

In fact, the experiments are not necessarily independent. However, below, we show that the assumption of independence is unnecessary, and that one can obtain the same conclusions by considering a set of independent variables that stochastically dominate the sequence  $X_i, i \geq 0$ .

For each phase  $i$ , think of starting an independent copy of the stochastic process described by (1) with the starting value  $y_i$ . Keeping in mind that a “success” is a passage from the interval  $[y_i, y_{i+1})$  to the next one, define a sequence of independent stochastic processes, one corresponding to each phase as follows: in phase  $i$ , start with the value  $y_i$  and repeat the “experiment” corresponding to the original stochastic process. If we have “success”, terminate the process for this phase, otherwise reset to the starting value,  $y_i$ . It is intuitively clear that the variables obtained by patching together these independent copies of the stochastic process stochastically dominate the original variables  $X_i, i \geq 0$ . We shall give a simple proof of this assertion using Proposition 3.

As to the independent process corresponding to phase  $i$ , this corresponds to a waiting time in a simple experiment consisting of repeated Bernoulli trials. Hence, we have a variable which is geometrically distributed that stochastically dominates the work in phase  $i$ . Finally, we put everything together using Lemma 4 for stochastic bounds on the sum of geometrically distributed variables.

#### 4.2. A generic bound

Define a sequence of variables  $Z_i, i \geq 0$  as follows:

$$Z_0 := y_0,$$

$$[Z_{j+1} | Z_j = y_i] := \begin{cases} y_i & \text{with probability } 1 - p, \\ y_{i+1} & \text{with probability } p. \end{cases}$$

**Lemma 6** (Stochastic Dominance). *The sequence  $Z_i, i \geq 0$  stochastically dominates the sequence  $X_i, i \geq 0$ .*

**Proof.** Apply Proposition 3. Since  $X_0 = x = Z_0$ , we merely verify that for any  $i \geq 0$  and for any  $x_0 \leq z_0, \dots, x_i \leq z_i$ ,

$$\Pr[X_{i+1} > t | X_0 = x_0, \dots, X_i = x_i] \leq \Pr[Z_{i+1} > t | Z_0 = z_0, \dots, Z_i = z_i],$$

which, by the memory-less nature of the processes, amounts to verifying that

$$\Pr[X_{i+1} > t \mid X_i = x_i] \leq \Pr[Z_{i+1} > t \mid Z_i = z_i].$$

This is easily seen to be the case by the definitions of the processes.  $\square$

Thus, for purposes of stochastic bounds, we can concentrate on the variables  $Z_i$ ,  $i \geq 0$ . For each phase  $i$ , let  $U_i := \sum_{Z_j=y_i} a(y_i)$ . Then for  $l \geq 1$ ,

$$\Pr[U_i = ly_i] = (1 - p)^{l-1} p.$$

Moreover, the variables  $U_i$  are independent of each other. Hence, we can apply Lemma 4 or Corollary 5 to  $U := \sum_i U_i$  to get

$$\Pr\left[U > \sum_i a(y_i) + la(y_0)\right] \leq C(1 - p)^{l-1}, \quad l \geq 1. \quad (4)$$

### 4.3. The parameters

Now we shall specify the sequence  $y_i$ ,  $i \geq 0$ . Given a function  $m: \mathbb{R} \rightarrow \mathbb{R}$  such that  $m(x) \leq x$ , define an auxiliary function  $\hat{m}$  as follows: First, define  $\Delta$  as in Theorem 2:

$$\Delta(x) := \max_{b \leq y \leq x} (m(y)/y)$$

and then set

$$\hat{m}(x) := m(x)/\sqrt{\Delta(x)}.$$

The function  $\hat{m}$  interpolates between the values  $x$  and  $m(x)$  in such a way that both of the following properties hold: (1) Applied twice,  $\hat{m}$  drops below  $m$  and (2) there is a finite probability for the event “ $H(X_{i+1}) < \hat{m}(X_i)$ ” via Markov’s inequality. The following proposition establishes (1).

**Proposition 7.** For all  $x \geq 0$ ,

$$\hat{m}(\hat{m}(x)) \leq m(x).$$

**Proof.** We compute

$$\begin{aligned} \hat{m}(\hat{m}(x)) &= m(\hat{m}(x))/\sqrt{\Delta(\hat{m}(x))} \\ &\leq \hat{m}(x)\sqrt{\Delta(\hat{m}(x))} \quad \text{as } m(y) \leq \Delta(y)y \\ &= m(x)\sqrt{\Delta(\hat{m}(x))/\Delta(x)} \quad \text{by definition of } \hat{m}(\cdot) \\ &\leq m(x) \quad \text{as } \Delta(z) \text{ is non-decreasing.} \quad \square \end{aligned}$$



Note that via Markov’s inequality on the third line below,

$$\begin{aligned}
 p_i &:= \Pr[H(y_i) \leq y_{i+1}] \\
 &= 1 - \Pr[H(y_i) > y_{i+1}] \\
 &\geq 1 - \frac{m(y_i)}{\hat{m}(y_i)} \\
 &= 1 - \sqrt{\Delta(y_i)} \\
 &\geq 1 - \sqrt{\Delta(x)} =: p.
 \end{aligned} \tag{5}$$

#### 4.4. The final step

Let  $\hat{u}(x) := \sum_i a(\hat{m}_{(i)}(x))$ . Suppose  $a(x)$  is strictly increasing (an assumption that will be removed shortly). Then from (4), and (5), we have

$$\Pr[U > \hat{u}(x) + la(x)] \leq C\Delta(x)^{(l-1)/2},$$

where  $C := C(a, m, x)$  is independent of  $l$ . Note that

$$\begin{aligned}
 \hat{u}(x) &:= \sum_{i \geq 0} a(\hat{m}^{(i)}(x)) \\
 &= \sum_{i \geq 0} a(\hat{m}^{(2i)}(x)) + a(\hat{m}^{(2i+1)}(x)) \\
 &\leq \sum_{i \geq 0} 2a(m^{(i)}(x)) \\
 &= 2u(x).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \Pr[U > u(x) + la(x)] &\leq \Pr \left[ U > \hat{u}(x) + \left( l - \frac{u(x)}{a(x)} \right) a(x) \right] \\
 &\leq C(\Delta(x))^{\frac{(l - \frac{u(x)}{a(x)}) - 1}{2}} \\
 &= C(\Delta(x))^{l-1/2} (\Delta(x))^{-u(x)/a(x)}.
 \end{aligned}$$

Finally, by Lemma 6, we can transfer this bound to our original variables  $X_i$ ,  $i \geq 0$  which are stochastically dominated by the variables  $Z_i$ ,  $i \geq 0$ , to get our final result:

**Theorem 2.** Let  $\Delta = \Delta(x) := \max_{b \leq y \leq x} (m(y)/y)$ , where  $b$  is the terminating point of the recurrence (1). Then,

$$\Pr[T(x) \geq u(x) + la(x)] \leq C(\Delta(x))^{(l-1)/2},$$

where  $C := C(a, m, x)$  is independent of  $l$ .

**Proof.** We need an additional technical comment for the case when  $a(x)$  is not strictly increasing. Since the generic bound (4) does not apply directly in that case. We employ

the following limiting argument: let  $\varepsilon > 0$  be arbitrary and consider  $a'(x) := a(x) + \varepsilon x$ . This is strictly increasing and so one can apply the bound (4) to  $a'$ . Now pass to the limit  $\varepsilon \rightarrow 0$ .  $\square$

**Remark 1.** The theorem (and also Karp’s Theorem) are essentially tight upto constants in the exponent under the weak hypothesis on only the expectation as the following example demonstrates. Let  $a(x) := x$ , and suppose the r.v. has the “two-point” distribution  $\Pr[H(X) = 0] = \frac{1}{2} = \Pr[H(X) = X]$ . So,  $E[H(X) | X] = X/2$ . One can easily compute that  $\Pr[T(x) = lx] = 2^{-l}$  for any positive integer  $l$ , hence  $\Pr[T(x) \geq lx] = 2^{-(l-1)}$ . Our theorem gives  $\Pr[T(x) \geq lx] = \Pr[T(x) \geq (l/2)2x] \leq C2^{-(l-1)/2}$ . Of course, with more information on the distribution, one can improve the probability bound, as the trivial example  $\Pr[H(x) = m(x)] = 1$  indicates.

**Remark 2.** One might consider fine-tuning the parameters. Thus, we could define for an arbitrary  $k \geq 1$ ,

$$\hat{m}_k(x) := \frac{m(x)}{(\Delta(x))^{1-1/k}}$$

Then one can show analogously that  $\hat{u}_k(x) \leq (1/k)u(x)$  and that  $p \geq 1 - (\Delta(x))^{1-1/k}$ . Then the probability bound would be

$$\Pr[U > u(x) + la(x)] \leq (\Delta(x))^{(1-1/k)(l - (1-1/k)\hat{u}_k(x)/a(x) - 1)}$$

It turns out that  $k := 2$  is a fairly good choice and one cannot significantly improve the analysis in this way.

**Remark 3.** Normally, one would like to see a large deviation result of the form  $\Pr[T(x) > E[T(x)] + \dots] < \dots$ . So the natural question is: how is the solution to the deterministic Eq. (2) related to  $E[T(x)]$ . We can give the following partial answer:

**Proposition 8.** *Let  $a$  and  $m$  both be concave functions. Then  $E[T(x)] \leq u(x)$ .*

**Proof.** The stochastic process described by the probabilistic recurrence (1), determines a sequence of non-increasing random variables

$$x =: X_0, X_1, \dots, X_i, \dots$$

such that

$$E[X_{i+1} | X_i] \leq m(X_i) \tag{6}$$

for each  $i \geq 0$ . Hence, we have

$$\begin{aligned} E[X_{i+1}] &= E[E[X_{i+1} | X_i]] \\ &\leq E[m(X_i)] \quad \text{using (6)} \\ &\leq m(E[X_i]) \quad \text{since } m \text{ is concave, using Jensen's inequality.} \end{aligned}$$

By induction then

$$E[X_i] \leq m^{(i)}(x) \quad (7)$$

for each  $i \geq 0$ .

Finally then, since

$$T(x) = \sum_{i \geq 0} a(X_i),$$

we have

$$\begin{aligned} E[T(x)] &= \sum_{i \geq 0} E[a(X_i)] \\ &\leq \sum_{i \geq 0} a(E[X_i]) \quad \text{since } a \text{ is concave} \\ &\leq \sum_{i \geq 0} a(m^{[i]}(x)) \quad \text{using (7)} \\ &= u(x). \quad \square \end{aligned}$$

Hence, in this situation ( $a, m$  concave), Theorem 1 yields the large deviation bounds in the usual form. However it would be nice to replace these conditions on  $a, m$  by more natural ones or perhaps to remove them altogether. We note that Proposition 8 was independently observed by Prabhakar Ragde (private communication).

## 5. Conclusion

We have shown that by applying standard tools from Probability Theory, namely Markov's Inequality, Stochastic Dominance and a Chernoff Bound for unbounded variables, we can obtain tail probability bounds on the performance of randomised algorithms comparable to those derived by Karp.

It is likely that the same techniques can be applied to probabilistic recurrence relations describing algorithms that generate more than one subproblem, and to versions of the recurrences describing the performance of *parallel* algorithms.

## Acknowledgements

We thank Torben Hagerup for pointing us to the work in [2]. We thank two conscientious referees for their detailed comments, suggestions and constructive criticism that greatly improved our presentation. We also thank the guest editor of the special issue, Ricardo Baeza-Yates for his patience.

## References

- [1] M. Dietzfelbinger and F. Meyer auf der Heide, A new universal class of hash functions and dynamic hashing in real time, *17th ICALP*, Lecture Notes in Computer Science, Vol. 443 (Springer, Berlin, 1990) 6–19.
- [2] D. Dubhashi, D. Grable and A. Panconesi, Near optimal distributed edge colouring via the Nibble method, *Theoret. Comput. Sci.* (special issue), to appear.
- [3] R.M. Karp, Probabilistic recurrence relations, *J. ACM* **41** (1994) 1136–1150.
- [4] R.M. Karp, E. Upfal and A. Wigderson, The complexity of parallel search, *J. Comput. System Sci.* **36** (1988) 225–253.
- [5] M. Luby, A simple parallel algorithm for the maximal independent set problem, *SIAM J. Comput.* **15** (1986) 1036–1053.
- [6] I.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* (Academic Press, New York, 1979).
- [7] A. Panconesi and A. Srinivasan, Fast randomized algorithms for distributed edge coloring, in: *Proc. ACM Symp. on Principles of Distributed Computing* (1992) 251–262.
- [8] R. Raman, The power of collision: randomized parallel algorithms for chaining and integer sorting, in: *Proc. 10th Ann. FST&TCS Conf.* Lecture Notes in Computer Science, Vol. 472 (Springer, Berlin, 1990) 161–175.