Asymptotic Limits of a New Type of Maximization Recurrence with an Application to Bioinformatics

Kun-Mao Chao[1], An-Chiang Chu[1], Jesper Jansson[2], Richard S. Lemence[2,3], and Alban Mancheron[4]

[1] Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan 106
kmchao@csie.ntu.edu.tw, anchiang@gmail.com
[2] Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan
Jesper.Jansson@ocha.ac.jp, rslemence@gmail.com
[3] Institute of Mathematics, College of Science, University of the Philippines, Diliman, Quezon City, 1101 Philippines
[4] Université Montpellier 2, LIRMM/CNRS, 161 rue Ada, 34095 Montpellier Cedex 5, France
alban.mancheron@lirmm.fr

Abstract. We study the asymptotic behavior of a new type of maximization recurrence, defined as follows. Let k be a positive integer and $p_k(x)$ a polynomial of degree k satisfying $p_k(0) = 0$. Define $A_0 = 0$ and for $n \geq 1$, let $A_n = \max_{0 \leq i < n}\{A_i + n^k\, p_k(\frac{i}{n})\}$. We prove that $\lim_{n\to\infty} \frac{A_n}{n^k} = \sup\{\frac{p_k(x)}{1-x^k} : 0 \leq x < 1\}$. We also consider two closely related maximization recurrences S_n and S'_n, defined as $S_0 = S'_0 = 0$, and for $n \geq 1$, $S_n = \max_{0\leq i<n}\{S_i + \frac{i(n-i)(n-i-1)}{2}\}$ and $S'_n = \max_{0\leq i<n}\{S'_i + \binom{n-i}{3} + 2i\binom{n-i}{2} + (n-i)\binom{i}{2}\}$. We prove that $\lim_{n\to\infty} \frac{S_n}{n^3} = \frac{2\sqrt{3}-3}{6} \approx 0.077350...$ and $\lim_{n\to\infty} \frac{S'_n}{3\binom{n}{3}} = \frac{2(\sqrt{3}-1)}{3} \approx 0.488033...$, resolving an open problem from Bioinformatics about rooted triplets consistency in phylogenetic networks.

1 Introduction

A *recurrence relation* (or *recurrence*, for short) is an equation of the form $T_n = f(T_{n-1}, T_{n-2}, \ldots, T_0, n)$, where f is a specified function and n is an unspecified positive integer, along with the values $T_0, T_1, \ldots, T_m$ for some finite, non-negative integer m. Intuitively, a recurrence describes how the value of T_n for any n depends on n and the values of the elements in the sequence $T_0, T_1, \ldots, T_{n-1}$.

Recurrences are central to the analysis of algorithms [3]. In particular, when recursion is involved, the worst-case running time T_n of an algorithm for an input of size n can often be expressed in terms of T_{n_1}, T_{n_2}, $\ldots$, T_{n_k}, where $n_1, n_2, \ldots, n_k < n$, which naturally yields a recurrence. It can be argued that recurrences are as important to Theoretical Computer Science as differential equations are to Physics. Over the years, elegant techniques for solving various types

M. Agrawal, S.B. Cooper, and A. Li (Eds.): TAMC 2012, LNCS 7287, pp. 177–188, 2012.

of *linear* recurrences (i.e., recurrences for which the function f mentioned above is a linear function) have been developed, and are now part of most standard undergraduate and graduate algorithm theory courses [3]. However, much less is known about how to solve *nonlinear* recurrences, and no general technique that works for all types of nonlinear recurrences exists. Instead, people have focused on asymptotically bounding the values of T_n as $n \to \infty$ for various special cases such as *minimization* recurrences of the form $T_n = \min_{1 \le i < n}\{T_i + T_{n-i}\} + g(n)$, where g is some auxiliary function, and *maximization* recurrences that use the max-function [5,9,12,13,15]. Interestingly, such recurrences have shown up in many different problems concerning random trees, Huffman coding, binomial group testing, dynamic programming, dichotomous search problems, the design of electrical circuits, binary search trees, quicksort, parallel divide-and-conquer algorithms, computational geometry, and tree-drawing.

In this paper, we contribute to the existing repertoire of tools for analyzing nonlinear recurrences. To be precise, we develop a technique for bounding the asymptotic behavior of a new type of maximization recurrence, defined as follows. Let k be a positive integer and $p_k(x)$ a polynomial of degree k satisfying $p_k(0) = 0$. Define $A_0 = 0$ and for $n \ge 1$, let

$$\boxed{A_n = \max_{0 \le i < n} \left\{ A_i + n^k \, p_k(\tfrac{i}{n}) \right\}}$$

We also consider two closely related maximization recurrences S_n and S'_n, defined as $S_0 = S'_0 = 0$, and for $n \ge 1$,

$$\boxed{S_n = \max_{0 \le i < n} \left\{ S_i + \tfrac{i(n-i)(n-i-1)}{2} \right\}}$$

and

$$\boxed{S'_n = \max_{0 \le i < n} \left\{ S'_i + \tbinom{n-i}{3} + 2i\tbinom{n-i}{2} + (n-i)\tbinom{i}{2} \right\}}$$

where $\binom{x}{y} = 0$ if $x < y$. (At this point, the reader may like to verify that some consecutive values of S'_n are: $S'_0 = 0$, $S'_1 = 0$, $S'_2 = 0$, $S'_3 = 2$, $S'_4 = 7$, $S'_5 = 16$, $S'_6 = 32$, $S'_7 = 55$, $S'_8 = 87$, $S'_9 = 130$, $S'_{10} = 184$, $\ldots$, and this sequence does not appear to follow any regular pattern.)

Below, we derive non-trivial, constant values of the expressions $\lim_{n\to\infty} A_n/n^k$, $\lim_{n\to\infty} S_n/n^3$, and $\lim_{n\to\infty} S'_n/3\binom{n}{3}$.

1.1 Motivation

Our motivation for studying the maximization recurrences in this paper originates from a combinatorial problem in Bioinformatics related to *phylogenetic networks* and *rooted triplets consistency*. This subsection describes the background; for further technical details, see [2] and [11].

One of the many objectives of Bioinformatics is to develop new concepts and tools that can help researchers visualize the evolutionary history of a set of species. Traditionally, *phylogenetic trees* (rooted, unordered, distinctly leaf-labeled trees in which every internal node has at least two children) have been used for this purpose [4]. As might be expected, it is computationally prohibitive in general to infer a reliable phylogenetic tree for a large set of species directly. A promising alternative is the *supertree approach* [1,8] which first infers highly accurate phylogenetic trees for many small, overlapping subsets of the species and then applies a combinatorial algorithm to merge them into a single phylogenetic tree. One variant of the supertree approach takes as input a set $\mathcal{R}$ of *rooted triplets* (binary phylogenetic trees with exactly three leaves each) whose leaf label sets overlap, and tries to construct a phylogenetic tree that is consistent with the maximum possible number of rooted triplets from $\mathcal{R}$, where a rooted triplet t is said to be *consistent* with a phylogenetic tree T if t is an embedded subtree of T. Gąsieniec *et al.* [6] presented a polynomial-time algorithm that outputs a phylogenetic tree which is consistent with at least $1/3$ of the rooted triplets in any input set $\mathcal{R}$, and also showed that for a particular set $\mathcal{R}$ of rooted triplets, no phylogenetic tree can be consistent with more than $1/3$ of its elements (to see this, just take the set $\mathcal{R}_n$ of all $3\binom{n}{3}$ rooted triplets over a fixed leaf label set of cardinality n, for any $n \geq 3$). In this sense, the algorithm of Gąsieniec *et al.* [6] is worst-case optimal for phylogenetic trees.

Due to certain evolutionary events such as hybridization that sometimes occur in nature, not all evolution is treelike. Therefore, the phylogenetic tree model was recently extended to *phylogenetic networks* that permit nodes to have more than one parent (see, e.g., the surveys in [10,14]). One important special type of phylogenetic network, introduced by Wang *et al.* [16] and later termed "galled-tree" by Gusfield *et al.* [7], requires all cycles in the underlying undirected graph to be node-disjoint. (Galled-trees are also known in the literature as "level-1 networks" [10,11,14].) Obviously, galled-trees can express more complicated evolutionary relationships than phylogenetic trees. To measure how much more powerful galled-trees really are, we can compare the optimal $1/3$ bound stated above for phylogenetic trees to the corresponding bound for galled-trees, and this leads to the recurrence S'_n studied in the present paper. More precisely, Jansson *et al.* [11] proved that for any $n \geq 3$, no galled-tree can be consistent with more than a fraction of $S'_n/3\binom{n}{3}$ of the elements in the set $\mathcal{R}_n$ of all rooted triplets over a fixed leaf label set of cardinality n. Later, Byrka *et al.* [2] gave a polynomial-time algorithm that constructs a galled-tree consistent with at least $S'_n/3\binom{n}{3}$ of the rooted triplets in any input set $\mathcal{R}$.

Jansson *et al.* [11] showed that for large enough values of n, it holds that $S'_n/3\binom{n}{3} < 0.4883$. On the other hand, Byrka *et al.* [2] proved that $S'_n/3\binom{n}{3} > 0.4800$ for all n. However, both groups of authors were unable to derive tight asymptotic bounds on $S'_n/3\binom{n}{3}$, and this has been one of the remaining open problems for galled-trees. Computations have suggested that $S'_n/3\binom{n}{3}$ is closer to the upper bound 0.4883 than the lower bound 0.4800, and indeed, we settle the issue in Section 3 by proving that $\lim_{n\to\infty} \frac{S'_n}{3\binom{n}{3}} = \frac{2(\sqrt{3}-1)}{3} \approx 0.488033...$ Observe that this improves the 5/12-ratio mentioned on p. 311 of [10] and the 48%-ratio mentioned on p. 135 of [14].

The other two recurrences introduced in this paper, S_n and A_n, were studied because of their connections to S'_n. As shown in Lemma 2 in Section 2 below, the bound for $S'_n/3\binom{n}{3}$ follows immediately from the bound for S_n/n^3, which is slightly easier to compute. A_n is a special case of a generalization of S_n.

1.2 Related Work

The appearance of nonlinear recurrence relations eluding exact solutions in diverse fields of study has motivated many previous papers, including [5,9,12,13,15], to investigate their asymptotic properties on a case-by-case basis. For example, Fredman and Knuth [5] considered minimization recurrences of the form $T_n = \min_{1\leq i<n}\{a \cdot T_i + b \cdot T_{n-i}\} + g(n)$, and Kapoor and Reingold [12] extended their results and also studied analogous maximization recurrences. In [13], Li and Reingold considered exact solutions and upper bounds for a special type of recurrence of the form $T_n = \max_{1\leq i<n}\{T_i + T_{n-i} + \min\{g(i), g(n-i)\}\}$ involving minimization and maximization simultaneously, and in [9], Hwang and Tsai derived asymptotic approximations of this recurrence for more general auxiliary functions g. Saha and Wagh [15] studied a recurrence of the form $T_n = \min_{1\leq i<n}\{\max\{T_i + a \cdot i, T_{n-i}\} + b\}$. Nevertheless, due to the irregular and often unpredictable behavior of nonlinear recurrences, general techniques for analyzing them still seem far from reach.

1.3 Main Results and Organization of the Paper

We establish the relationships among the three recurrences A_n, S_n, and S'_n in Section 2. Then, in Section 3, we prove that $\lim\limits_{n\to\infty} \frac{S_n}{n^3} = \frac{2\sqrt{3}-3}{6} \approx 0.077350...$ and that $\lim\limits_{n\to\infty} \frac{S'_n}{3\binom{n}{3}} = \frac{2(\sqrt{3}-1)}{3} \approx 0.488033...$ Next, in Section 4, we consider the ratio A_n/n^k. We show that $\lim\limits_{n\to\infty} \frac{A_n}{n^k} = \sup\{\frac{p_k(x)}{1-x^k} : 0 \leq x < 1\}$. Finally, Section 5 discusses generalizations of our techniques and an open problem.

2 Preliminaries

The two recurrences S_n and S'_n are related as follows.

Lemma 1. *For all $n \geq 0$, it holds that $S_n = S'_n - \binom{n}{3}$.*

Proof. By induction on n. For $n = 0$, we have $S_0 = S'_0 = 0$.

Next, suppose that $S_k = S'_k - \binom{k}{3}$ for all $k < n$. Then, since $\binom{n}{3} = \binom{n-i}{3} + i\binom{n-i}{2} + (n-i)\binom{i}{2} + \binom{i}{3}$ for every $0 \leq i < n$, we can rewrite S'_n as $S'_n = \binom{n}{3} + \max\limits_{0\leq i<n}\{i\binom{n-i}{2} + S'_i - \binom{i}{3}\}$. By the induction hypothesis: $S'_n - \binom{n}{3} = \max\limits_{0\leq i<n}\{i\binom{n-i}{2} + S'_i - \binom{i}{3}\} = \max\limits_{0\leq i<n}\{i\binom{n-i}{2} + S_i\} = S_n$. $\square$

Lemma 2. $\lim\limits_{n\to\infty} \frac{S'_n}{3\binom{n}{3}} = \lim\limits_{n\to\infty} \frac{2S_n}{n^3} + \frac{1}{3}$.

Proof. From Lemma 1, we have $\lim_{n\to\infty} \frac{S'_n}{3\binom{n}{3}} = \lim_{n\to\infty} \frac{S_n}{3\binom{n}{3}} + \frac{1}{3} = \lim_{n\to\infty} \frac{2S_n}{n^3} + \frac{1}{3}$. □

Next, we consider the relationship between the recurrences S_n and A_n. Another (equivalent) way to write S_n is:

$$S_n = \max_{0\le i<n} \{S_i + n^3 \cdot p_3(\frac{i}{n}) + n^2 \cdot p_2(\frac{i}{n})\},$$

where $p_3(x) = \frac{x(1-x)^2}{2}$ and $p_2(x) = \frac{-x(1-x)}{2}$. Looking at S_n defined in this way, we are tempted to extend it to a more general type of recurrence as follows. Let k be a positive integer and let $p_0(x), p_1(x), \ldots, p_k(x)$ be polynomials such that $p_d(x)$ is a polynomial of degree d for every $d \in \{0, 1, \ldots, k\}$. Set $G_0 = p_0(0)$, and for $n \geq 1$, define:

$$G_n = \max_{0\le i<n} \{G_i + \sum_{d=0}^{k} n^d p_d(\frac{i}{n})\}.$$

Now, if we restrict the recurrence G_n to the special case where $p_d(x) = 0$ for all $d \in \{0, 1, \ldots, k-1\}$ and $p_k(0) = 0$, we obtain precisely the recurrence A_n.

3 The Asymptotic Behavior of S_n and S'_n

In order to analyze the asymptotic behavior of S_n/n^3, we define $s_n = S_n/n^3$ and rewrite S_n in terms of s_n. This gives $s_0 = 0$, and for $n \geq 1$:

$$s_n = \max_{0\le i<n} \{s_{n,i}\}, \text{ where } s_{n,i} = p_3(\frac{i}{n}) + \frac{1}{n} \cdot p_2(\frac{i}{n}) + s_i \cdot (\frac{i}{n})^3.$$

Here, p_3 and p_2 are the polynomials $p_3(x) = \frac{x(1-x)^2}{2}$ and $p_2(x) = \frac{-x(1-x)}{2}$, introduced in Section 2. Consider the function $\frac{p_3(x)}{1-x^3}$. It has a unique maximum value on the interval $[0, 1)$. Call this value α and let β be the point where α is obtained, i.e., $\frac{p_3(\beta)}{1-\beta^3} = \alpha$. By straightforward calculations, we have $\alpha = \frac{2\sqrt{3}-3}{6}$, $\beta = \frac{\sqrt{3}-1}{2}$. In this section, we shall prove that $\lim_{n\to\infty} s_n = \alpha$.

First, we introduce two sequences l_n, u_n $(n \geq 0)$ and show that they provide a lower bound and an upper bound, respectively, on each term in the sequence s_n. Let $l_0 = u_0 = 0$ and, for $n \geq 1$, define:

$$\begin{cases} l_n = \max_{0\le i<n} \{l_{n,i}\}, & \text{where } l_{n,i} = p_3(\frac{i}{n}) + \frac{1}{n} p_2(\frac{i}{n}) + \alpha(\frac{i}{n} - \frac{1}{n})^3, \\ u_n = \max_{0\le i<n} \{u_{n,i}\}, & \text{where } u_{n,i} = p_3(\frac{i}{n}) + \frac{1}{n} p_2(\frac{i}{n}) + \alpha(\frac{i}{n})^3. \end{cases}$$

In the next four lemmas, we show that the following chain of inequalities holds for every integer $n \geq 1$:

$$\alpha(1 - \frac{1}{n})^3 \leq l_n \leq s_n \leq u_n \leq \alpha.$$

Lemma 3. *For all $n \geq 0$, $u_n \leq \alpha$.*

Proof. By the definition of α, we have $\frac{p_3(x)}{1-x^3} \leq \alpha$, for $0 \leq x < 1$. This yields $p_3(x) + \alpha x^3 \leq \alpha$, for $0 \leq x < 1$. Since u_n is defined as $\max\limits_{0 \leq i < n} \{p_3(\frac{i}{n}) + \frac{1}{n} p_2(\frac{i}{n}) + \alpha(\frac{i}{n})^3\}$ and $p_2(x) \leq 0$ for all $0 \leq x < 1$, we have $u_n \leq \alpha$. □

Lemma 4. *For all $n \geq 0$, $s_n \leq u_n$.*

Proof. By induction on n. For $n = 0$, $u_0 = s_0 = 0$. Next, suppose $s_m \leq u_m$ for all $m < n$. For each integer $0 \leq i < n$, by Lemma 3, we have $s_{n,i} - u_{n,i} = s_i(\frac{i}{n})^3 - \alpha(\frac{i}{n})^3 \leq s_i(\frac{i}{n})^3 - u_i(\frac{i}{n})^3 = (s_i - u_i)(\frac{i}{n})^3 \leq 0$. Therefore,$s_n = \max\limits_{1 \leq i < n} \{s_{n,i}\} = \max\limits_{1 \leq i < n} \{u_{n,i} + (s_{n,i} - u_{n,i})\} \leq \max\limits_{1 \leq i < n} \{u_{n,i}\} = u_n$. □

Lemma 5. *For all $n \geq 1$, $l_n \geq \alpha(1 - \frac{1}{n})^3$.*

Proof. For $n \leq 15$, the inequality can be verified by computation. For $n \geq 16$, we show that $l_n \geq \alpha(1 - \frac{1}{n})^3$. First note that:

(*1) Since $\beta - \frac{1}{n} \leq \frac{\lfloor \beta n \rfloor}{n} \leq \beta = \frac{\sqrt{3}-1}{2} < \frac{1}{2}$ and $p_2(x)$ is decreasing on $[0, \frac{1}{2}]$, we have $p_2(\frac{\lfloor \beta n \rfloor}{n}) \geq p_2(\beta)$.

(*2) We have $p_3(x) \geq p_3(\beta)$ for $x \in [0.302, \beta]$. For $n \geq 16$, $\frac{\lfloor \beta n \rfloor}{n} > \beta - \frac{1}{n} \geq \beta - \frac{1}{16} > 0.302$, therefore we have $p_3(\frac{\lfloor \beta n \rfloor}{n}) > p_3(\beta)$.

Then, it follows that:

$$
\begin{aligned}
& l_n - \alpha(1 - \frac{1}{n})^3 \geq l_{n,\lfloor \beta n \rfloor} - \alpha(1 - \frac{1}{n})^3 \\
=& p_3(\frac{\lfloor \beta n \rfloor}{n}) \underbrace{-p_3(\beta) + \alpha(1 - \beta^3)}_{=0} + \frac{1}{n} p_2(\frac{\lfloor \beta n \rfloor}{n}) + \alpha((\frac{\lfloor \beta n \rfloor}{n} - \frac{1}{n})^3 - (1 - \frac{1}{n})^3) \\
=& \underbrace{p_3(\frac{\lfloor \beta n \rfloor}{n}) - p_3(\beta)}_{\geq 0, \text{ by } (*2)} + \alpha((\frac{\lfloor \beta n \rfloor}{n} - \frac{1}{n})^3 - \beta^3 + 1 - (1 - \frac{1}{n})^3) + \frac{1}{n} p_2(\frac{\lfloor \beta n \rfloor}{n}) \\
\geq& \alpha((\frac{\lfloor \beta n \rfloor}{n} - \frac{1}{n})^3 - (\beta - \frac{2}{n})^3 + \frac{3 - 6\beta^2}{n} + \frac{12\beta - 3}{n^2} - \frac{7}{n^3}) + \frac{1}{n} \cdot \underbrace{p_2(\frac{\lfloor \beta n \rfloor}{n})}_{\geq p_2(\beta), \text{ by } (*1)} \\
\geq& \underbrace{\alpha((\frac{\lfloor \beta n \rfloor}{n} - \frac{1}{n})^3 - (\beta - \frac{2}{n})^3)}_{\geq 0} + \underbrace{\frac{\alpha}{n}(3 - 6\beta^2 + \frac{p_2(\beta)}{\alpha} + \frac{12\beta - 3}{n} + \frac{-7}{n^2})}_{\geq 0, \text{ for } n \geq 3} \geq 0.
\end{aligned}
$$

□

Lemma 6. *For all $n \geq 1$, $s_n \geq l_n$.*

Proof. By induction on n. For $n = 0$, $s_0 = l_0 = 0$. Next, suppose $s_m \geq l_m$, for all $m < n$. For each integer $0 \leq i < n$, by Lemma 5, we have $s_{n,i} - l_{n,i} = s_i(\frac{i}{n})^3 - \alpha(\frac{i}{n} - \frac{1}{n})^3 = s_i(\frac{i}{n})^3 - \alpha(1 - \frac{1}{i})^3(\frac{i}{n})^3 \geq (s_i - l_i)(\frac{i}{n})^3 \geq 0$. Therefore, $\max\limits_{0 \leq i < n} \{s_{n,i}\} \geq \max\limits_{0 \leq i < n} \{l_{n,i}\}$, which gives $s_n \geq l_n$. □

We now obtain the main result of this section:

Theorem 1. $\lim_{n\to\infty} \frac{S_n}{n^3} = \lim_{n\to\infty} s_n = \alpha = \frac{2\sqrt{3}-3}{6} \approx 0.077350....$

Proof. By Lemmas 3–6, we have $\alpha(1-\frac{1}{n})^3 \le s_n \le \alpha$. Therefore, $\alpha = \lim_{n\to\infty} \alpha(1-\frac{1}{n})^3 \le \lim_{n\to\infty} s_n \le \alpha$, i.e., $\lim_{n\to\infty} s_n = \alpha$. □

Finally, using Theorem 1 together with Lemma 2 gives:

Corollary 1. $\lim_{n\to\infty} \frac{S'_n}{3\binom{n}{3}} = \frac{2(\sqrt{3}-1)}{3} \approx 0.488033....$

Remark. Corollary 1 gives a strengthening of the inapproximability bound in Theorem 8 in [11]; just change the "0.4883" to any real number strictly larger than $\frac{2(\sqrt{3}-1)}{3}$, for example "0.488034". Moreover, we can strengthen Lemma 5 in [2] (which says that $S'_n/3\binom{n}{3} > 0.4800$) and the resulting approximation ratio in Theorem 2 in [2] by observing that $S'_n/3\binom{n}{3} = 2\frac{S_n}{n^3}\cdot\frac{n^2}{(n-1)(n-2)}+\frac{1}{3} \ge 2\alpha(\frac{n-1}{n})^3\cdot\frac{n^2}{(n-1)(n-2)} + \frac{1}{3}$ by Lemmas 5 and 6, and then rewriting it as $2\alpha\cdot\frac{(n-1)^2}{(n-2)n} + \frac{1}{3} > 2\alpha + \frac{1}{3} = \frac{2(\sqrt{3}-1)}{3}$. In other words, $S'_n/3\binom{n}{3} > \frac{2(\sqrt{3}-1)}{3} \approx 0.488033....$

4 The Asymptotic Behavior of A_n

The asymptotic behavior of A_n depends on the properties of $p_k(x)/(1-x^k)$. We define $\alpha_p = \sup\{p_k(x)/(1-x^k) : 0 \le x < 1\}$, when $\sup\{p_k(x)/(1-x^k) : 0 \le x < 1\} < \infty$.[1] There are four possible cases:

(C1) $\sup\{p_k(x)/(1-x^k) : 0 \le x < 1\} = \infty$.
(C2) $\sup\{p_k(x)/(1-x^k) : 0 \le x < 1\} = \alpha_p < \infty$, and $\lim_{x\to 1^-} \frac{p_k(x)}{1-x^k} = \alpha_p$.
(C3) $\sup\{p_k(x)/(1-x^k) : 0 \le x < 1\} = \alpha_p = 0$, and $\frac{p_k(0)}{1-0^k} = \alpha_p = 0$.
(C4) $\sup\{p_k(x)/(1-x^k) : 0 \le x < 1\} = \alpha_p < \infty$, and there exists a β_p, where $0 < \beta_p < 1$, such that $\frac{p_k(\beta_p)}{1-\beta_p^k} = \alpha_p$.

The definition of A_n is $\max_{0\le i<n}\{n^k p_k(\frac{i}{n}) + A_i\}$, for $n > 0$. If we substitute A_n $(m-1)$ times recursively, we get

$$A_n = \max_{0\le i_2<i_1<n}\{n^k p_k(\frac{i_1}{n}) + i_1^k p_k(\frac{i_2}{i_1}) + A_{i_2}\} = \cdots$$
$$= \max_{0\le i_m<\cdots<i_1<i_0}\{\sum_{t=0}^{m-1} i_t^k p_k(\frac{i_{t+1}}{i_t}) + A_{i_m}\}.$$

[1] Note that we use "sup" instead of "max" for the following reason. For some $p_k(x)$, e.g., $k = 3, p_3(x) = -x^3 + x$, there is no maximum value for $\frac{p_k(x)}{1-x^k}$, $0 \le x < 1$. However, there exists an upper bound for $\frac{p_k(x)}{1-x^k}$, $0 \le x < 1$.

By choosing $i_t = n - t$, we define L_n with $L_0 = 0$, and for $n \geq 1$,

$$L_n = n^k p_k(\frac{n-1}{n}) + L_{n-1}.$$

We substitute L_n $(m-1)$ times, which gives: $L_n = \sum_{t=0}^{n-1}(n-t)^k p_k(\frac{n-t-1}{n-t})$.

Since A_n is taking the maximum value among all parameters $\{i_t\}$, we have $A_n \geq L_n$. For case (C1), we show that $\lim_{n\to\infty} \frac{L_n}{n^k} = \infty$ in Lemma 7. It follows that $\lim_{n\to\infty} \frac{A_n}{n^k} = \infty$. For case (C2), we show that $\lim_{n\to\infty} \frac{L_n}{n^k} = \alpha_p$ and A_n also has an upper bound α_p. Therefore, $\lim_{n\to\infty} \frac{A_n}{n^k} = \alpha_p$.

Lemma 7. *If* $\sup\{\frac{p_k(x)}{1-x^k} : 0 \leq x < 1\} = \infty$*, then* $\lim_{n\to\infty} \frac{A_n}{n^k} = \infty$.

Proof. Assume that $p_k(x) = \sum_{i=1}^{k} c_i x^i$. We observe that $(n-t)^k p_k(\frac{n-t-1}{n-t})$ is a polynomial of $(n-t)$ with degree at most k. Furthermore, the coefficient of $(n-t)^k$ in $(n-t)^k p_k(\frac{n-t-1}{n-t}) = \sum_{i=1}^{k} c_i(n-t-1)^i(n-t)^{k-i}$ equals $\sum_{i=1}^{k} c_i = p_k(1)$. For the reason that $\lim_{x\to 1^-} \frac{p_k(x)}{1-x^k} = \infty$, we have $p_k(1) > 0$.

Since $L_n = \sum_{t=0}^{n-1}(n-t)^k p_k(\frac{n-t-1}{n-t})$, L_n is a polynomial of n with degree $k+1$. Therefore, $\lim_{n\to\infty} \frac{A_n}{n^k} \geq \lim_{n\to\infty} \frac{L_n}{n^k} = \infty$. □

Lemma 8. *If* $\sup\{\frac{p_k(x)}{1-x^k} : 0 \leq x < 1\} = \alpha_p < \infty$ *and* $\lim_{x\to 1^-} \frac{p_k(x)}{1-x^k} = \alpha_p$*, then* $\lim_{n\to\infty} \frac{A_n}{n^k} = \alpha_p$.

Proof. The proof of the upper bound of A_n is at most α_p is similar to that of Lemma 4.

Assume that $p_k(x) = \sum_{i=1}^{k} c_i x^i$. The coefficient of $(n-t)$ in $(n-t)^k p_k(\frac{n-t-1}{n-t})$ equals $p_k(1)$. However, for the reason that $\lim_{x\to 1^-} \frac{p_k(x)}{1-x^k} = \alpha_p$, we have $p_k(1) = 0$. Hence, L_n is a polynomial with degree at most k.

Furthermore, the coefficient of $(n-t)^{k-1}$ in $(n-t)^k p_k(\frac{n-t-1}{n-t}) = \sum_{i=1}^{k} c_i(n-t-1)^i(n-t)^{k-i}$ is $\sum_{i=1}^{k} -ic_i = -p'_k(1)$. We have the coefficient of n^k in L_n equals that in $\sum_{t=0}^{n-1} -p'_k(1) \cdot (n-t)^{k-1}$. Then the coefficient of n^k in L_n equals $\frac{-p'_k(1)}{k}$.

Since $(x-1)$ is a factor of $p_k(x)$, let $q_k(x) = \frac{p_k(x)}{x-1}$. Then $\frac{d}{dx}p_k(x) = \frac{d}{dx}(q_k(x)(x-1)) = q_k(x) + (x-1)\frac{d}{dx}(q_k(x))$. Hence, $p'_k(1) = q_k(1)$. Moreover,

$$\alpha_p = \lim_{x\to 1^-} \frac{p_k(x)}{1-x^k} = \lim_{x\to 1^-} \frac{(x-1)q_k(x)}{1-x^k} = \lim_{x\to 1^-} \frac{-q_k(x)}{1+x+\cdots+x^{k-1}} = \frac{-q_k(1)}{k}.$$

Finally, we have $\lim_{n\to\infty} \frac{L(n)}{n^k} = \frac{-p_k'(1)}{k} = \frac{-q_k(1)}{k} = \alpha_p$. Then, $\lim_{n\to\infty} \frac{A_n}{n^k} = \alpha_p$. □

Lemma 9. *If* $\sup\{\frac{p_k(x)}{1-x^k} : 0 \le x < 1\} = 0$*, then* $A_n = 0$.

Proof. By induction on n. For $n = 0$, it holds that $A_0 = 0$. Next, suppose that $A_m = 0$ for all $m < n$. Then, since $\alpha_p = 0$, we have $p_k(x) \le 0$, for $0 \le x \le 1$, therefore

$$A_n = \max_{0\le i<n} \{n^k p_k(\frac{i}{n}) + A_i\} \le \max_{0\le i<n} \{A_i\} = 0. \qquad \square$$

To study the asymptotic value of A_n/n^k in case (C4), we define $a_n = A_n/n^k$, and rewrite the recurrence for A_n in terms of a_n as follows. Let $a_0 = 0$ and, for $n \ge 1$,

$$a_n = \max_{0\le i<n} \{a_{n,i}\}, \text{ where } a_{n,i} = p_k(\frac{i}{n}) + a_i(\frac{i}{n})^k.$$

To find a lower bound of a_n, we rewrite a_n by recursively substituting it $(m-1)$ times, for some value of m to be specified later.

$$\begin{aligned} a_n &= \max_{0\le i_1<n} \{p_k(\frac{i_1}{n}) + (\frac{i_1}{n})^k a_{i_1}\} = \max_{0\le i_2<i_1<n} \{p_k(\frac{i_1}{n}) + (\frac{i_1}{n})^k (p_k(\frac{i_2}{i_1}) + (\frac{i_2}{i_1})^k a_{i_2})\} \\ &= \cdots = \max_{0\le i_m<\cdots<i_1<i_0=n} \{(\sum_{t=0}^{m-1} (\frac{i_t}{n})^k p_k(\frac{i_{t+1}}{i_t})) + (\frac{i_m}{n})^k a_{i_m}\}. \end{aligned}$$

By choosing $i_t = \lfloor \beta_p^t n \rfloor$ for a_n, we define $l_{n,m} = (\sum_{t=0}^{m-1} (\frac{\lfloor \beta_p^t n\rfloor}{n})^k p_k(\frac{\lfloor \beta_p^{t+1} n\rfloor}{\lfloor \beta_p^t n\rfloor})) + (\frac{\lfloor \beta_p^m n\rfloor}{n})^k a_{i_m}$. For the condition that $\lfloor \beta_p^{t-1} n\rfloor > \lfloor \beta_p^t n\rfloor$ with $t < m$ to hold, we need m to satisfy $\beta_p^{m-1} n \ge \beta_p^m n + 1$, i.e., $n > \frac{1}{\beta_p^{m-1}(1-\beta_p)}$. Since a_n is taking the maximum value among all parameters $\{i_t\}$, we have $a_n \ge l_{n,m}$.

To show that $l_{n,m}$ converges to α_p, we replace α_p by $p_k(\beta_p) + \alpha_p \beta_p^k$ $(m - 1)$ times and find an expression for α_p which looks similar to the formula for $l_{n,m}$.

$$\begin{aligned} \alpha_p &= p_k(\beta_p) + \alpha_p \beta^k = p_k(\beta_p) + \beta_p^k(p_k(\beta_p) + \alpha_p \beta_p^k) = \cdots \\ &= (\sum_{t=0}^{m-1} \beta_p^{tk} p_k(\beta_p)) + \beta_p^{mk} \alpha_p. \end{aligned}$$

In the next lemma, we show that $l_{n,m}$ is close to α_p based on two observations: (1) $\beta_p^{mk}\alpha_p$ is very small for sufficiently large m; and (2) when $\lfloor \beta_p^t n\rfloor$ is large, $\frac{\lfloor \beta_p^{t+1} n\rfloor}{\lfloor \beta_p^t n\rfloor}$ is close to β_p and then we have $\beta_p^{tk} p_k(\beta_p)$ is close to $(\frac{\lfloor \beta_p^t n\rfloor}{n})^k p_k(\frac{\lfloor \beta_p^{t+1} n\rfloor}{\lfloor \beta_p^t n\rfloor})$.

Lemma 10. *If* $\sup\{\frac{p_k(x)}{1-x^k} : 0 \le x < 1\} = \alpha_p < \infty$ *and there exists a* β_p*, where* $0 < \beta_p < 1$*, such that* $\frac{p_k(\beta_p)}{1-\beta_p^k} = \alpha_p$*, then* $\lim_{n\to\infty} \frac{A_n}{n^k} = \alpha_p$.

Proof. The proof of α_p being the upper bound of a_n is similar to that of Lemma 4. To pave the way for the lower bound of a_n, we introduce two notations M_1 and M_2.

Consider the Taylor series expansion for $p_k(x)$ in β_p: $p_k(x) = p_k(\beta_p) + \sum_{i=1}^{k} \frac{f^{(i)}(\beta_p)}{i!}(x-\beta_p)^i$. For $0 \le x < 1$, we have

$$\begin{aligned}|p_k(x) - p_k(\beta_p)| \le & (x-\beta_p)\sum_{i=1}^{k} |\frac{p_k^{(i)}(\beta_p)}{i!}(x-\beta_p)^{i-1}| \\ \le & (x-\beta_p)\sum_{i=1}^{k} |\frac{p_k^{(i)}(\beta_p)}{i!}| \quad \text{(because } 0 < x, \beta_p < 1) \\ \le & (x-\beta_p)M_1, \quad \text{where } M_1 = \sum_{i=1}^{k} |\frac{p_k^{(i)}(\beta_p)}{i!}|. \end{aligned} \tag{1}$$

Since $p_k(x)$ is a polynomial, there exists a maximum value of $p_k(x)$ on the interval [0,1]. Let

$$M_2 = \max_{0 \le x \le 1} \{p_k(x)\}. \tag{2}$$

Furthermore, for the reason that $0 < \beta < 1$, we have:

$$\beta_p^{tk}(\beta_p - \frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor}) \le \frac{\beta_p^{tk}}{\lfloor \beta_p^t n \rfloor} \le \frac{2\beta_p^{tk}}{\beta_p^t n} \le \frac{2\beta_p^{t(k-1)}}{n} \le \frac{2}{n}, \text{ and} \tag{3}$$

$$\beta_p^{tk} - (\frac{\lfloor \beta_p^t n \rfloor}{n})^k = (\beta_p^t - \frac{\lfloor \beta_p^t n \rfloor}{n}) \sum_{i=0}^{k-1} ((\beta_p^t)^i (\frac{\lfloor \beta_p^t n \rfloor}{n})^{k-1-i}) \le \frac{k}{n}. \tag{4}$$

Since M_1, M_2, α_p and β_p are fixed values, for all $\epsilon > 0$, there exists a positive integer m such that:

$$\beta_p^{mk}(2mM_1 + kmM_2 + \alpha_p) < \epsilon. \tag{5}$$

For $n \ge \max\{\lceil \frac{1}{\beta_p^{mk}} \rceil, \lceil \frac{1}{\beta_p^{m-1}(1-\beta_p)} \rceil\}$, we have

$$\begin{aligned} & |\alpha_p - a_n| \\ \le & |\alpha_p - l_{n,m}| \quad \text{(because } l_{n,m} \le a_n \le \alpha) \\ \le & |\sum_{t=0}^{m-1} (\beta_p^{tk} f(\beta_p) - (\frac{\lfloor \beta_p^t n \rfloor}{n})^k p_k(\frac{\lfloor \beta_p^{t+1} n \rfloor}{\lceil \beta_p^t n \rceil}))| + |\beta_p^{mk}\alpha_p - (\frac{\lfloor \beta_p^m n \rfloor}{n})^k a_m| \\ \le & |\sum_{t=0}^{m-1} (\beta_p^{tk} p_k(\beta_p) \underbrace{-\beta_p^{tk} p_k(\frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor}) + \beta_p^{tk} p_k(\frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor})}_{=0} - (\frac{\lceil \beta_p^t n \rceil}{n})^k p_k(\frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor}))| \\ & + \beta_p^{mk}\alpha_p \end{aligned}$$

$$= \sum_{t=0}^{m-1} |\beta_p^{tk}(p_k(\beta_p) - p_k(\frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor})) + (\beta_p^{tk} - (\frac{\lfloor \beta_p^t n \rfloor}{n})^k) p_k(\frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor})| + \beta_p^{mk}\alpha_p$$

$$\leq \sum_{t=0}^{m-1} (|\beta_p^{tk}(\beta_p - \frac{\lfloor \beta_p^{t+1} n \rfloor}{\lfloor \beta_p^t n \rfloor}) M_1| + |(\beta_p^{tk} - (\frac{\lfloor \beta_p^t n \rfloor}{n})^k) M_2|) + \beta_p^{mk}\alpha_p \quad \text{(by (1), (2))}$$

$$\leq \frac{1}{n} \sum_{t=0}^{m-1} |2M_1 + kM_2| + \beta_p^{mk}\alpha_p \qquad \text{(by (3), (4))}$$

$$\leq \beta_p^{mk}(m(2M_1 + kM_2) + \alpha_p) \leq \epsilon \qquad \text{(by } n \geq \lceil \frac{1}{\beta_p^{mk}} \rceil \text{ and (5))} \qquad \square$$

Combining Lemmas 7 – 10, we obtain the following result.

Theorem 2. $\lim_{n\to\infty} \frac{A_n}{n^k} = \sup\{\frac{p_k(x)}{1-x^k} : 0 \leq x < 1\}$.

Remark. When we take $k = 3$ and $p_3(x) = \frac{x(1-x)(1-x)}{2}$ in A_n, we have $\lim_{n\to\infty} A_n/n^3 = (2\sqrt{3}-3)/6$, which is equal to $\lim_{n\to\infty} S_n/n^3$. We can see that the term $p_2(x)$ in S_n has no effect on the asymptotic behavior of S_n.

5 Concluding Remarks

We note that to analyze *minimization* recurrences analogous to A_n, we can apply our technique from Section 4 as follows. Suppose that $B_n = \min_{0\leq i<n} \{n^k p_k(\frac{i}{n}) + B_i\}$. Let $A_n = -B_n$. Then $A_n = \max_{0\leq i<n} \{n^k \cdot (-p_k(\frac{i}{n})) + A_i\}$, and Theorem 2 gives:

Corollary 2. $\lim\limits_{n\to\infty} \frac{B_n}{n^k} = \inf\{\frac{p_k(x)}{1-x^k} : 0 \leq x < 1\}$.

We conclude this paper by mentioning two open problems. First, to derive a closed-form expression for the exact value of S_n or to determine that such a formula does not exist is an open problem. Second, for the general case of G_n (see Section 2), we can set $g_n = G_n/n^k$ and rewrite the recurrence relation as:

$$g_n = \max_{0\leq i<n} \{ (\sum_{d=0}^{k} \frac{1}{n^{k-d}} p_d(\frac{i}{n})) + g_i(\frac{i}{n})^k \}.$$

For $d < k$, the term $p_d(\frac{i}{n})$ is multiplied by $\frac{1}{n^{k-d}}$. For sufficiently large n, the part $p_d(\frac{i}{n})$ has a small effect on g_n, for $d < k$. Hence, we conjecture that the asymptotic behavior of g_n is the same as that of a_n.

Acknowledgements. An-Chiang Chu was supported in part by the 2011 Summer Visiting Program from the Interchange Association, Japan. An-Chiang Chu and Kun-Mao Chao were supported in part by NSC grants 98-2221-E-002-081-MY3 and 100-2221-E-002-131-MY3 from the National Science Council, Taiwan. Jesper Jansson and Richard S. Lemence were funded by the Special Coordination Funds for Promoting Science and Technology, Japan.

References

1. Bininda-Emonds, O.R.P.: The evolution of supertrees. Trends in Ecology and Evolution 19(6), 315–322 (2004)
2. Byrka, J., Gawrychowski, P., Huber, K.T., Kelk, S.: Worst-case optimal approximation algorithms for maximizing triplet consistency within phylogenetic networks. Journal of Discrete Algorithms 8(1), 65–75 (2010)
3. Cormen, T., Leiserson, C., Rivest, R., Stein, C.: Introduction to Algorithms, 3rd edn. The MIT Press, Massachusetts (2009)
4. Felsenstein, J.: Inferring Phylogenies. Sinauer Associates, Inc., Sunderland (2004)
5. Fredman, M.L., Knuth, D.E.: Recurrence relations based on minimization. Journal of Mathematical Analysis and Applications 48(2), 534–559 (1974)
6. Gąsieniec, L., Jansson, J., Lingas, A., Östlin, A.: On the complexity of constructing evolutionary trees. Journal of Combinatorial Optimization 3(2-3), 183–197 (1999)
7. Gusfield, D., Eddhu, S., Langley, C.: Efficient reconstruction of phylogenetic networks with constrained recombination. In: Proceedings of the Computational Systems Bioinformatics Conference (CSB2 2003), pp. 363–374 (2003)
8. Henzinger, M.R., King, V., Warnow, T.: Constructing a tree from homeomorphic subtrees, with applications to computational evolutionary biology. Algorithmica 24(1), 1–13 (1999)
9. Hwang, H.-K., Tsai, T.-H.: An asymptotic theory for recurrence relations based on minimization and maximization. Theoretical Computer Science 290(3), 1475–1501 (2003)
10. Huson, D.H., Rupp, R., Scornavacca, C.: Phylogenetic Networks: Concepts, Algorithms and Applications. Cambridge University Press (2010)
11. Jansson, J., Nguyen, N., Sung, W.: Algorithms for combining rooted triplets into a galled phylogenetic network. SIAM Journal on Computing 35(5), 1098–1121 (2006)
12. Kapoor, S., Reingold, E.M.: Recurrence relations based on minimization and maximization. Journal of Mathematical Analysis and Applications 109(2), 591–604 (1985)
13. Li, Z., Reingold, E.M.: Solution of a divide-and-conquer maximin recurrence. SIAM Journal on Computing 18(6), 1188–1200 (1989)
14. Morrison, D.: Introduction to Phylogenetic Networks. RJR Productions (2011)
15. Saha, A., Wagh, M.D.: Minmax recurrences in analysis of algorithms. In: Proceedings of Southeastcon 1993. IEEE (1993)
16. Wang, L., Ma, B., Li, M.: Fixed topology alignment with recombination. Discrete Applied Mathematics 104(1-3), 281–300 (2000)