Optimal Regular Expressions for Permutations

Antonio Molina Lovett and Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
ajmolina@uwaterloo.ca
shallit@uwaterloo.ca

December 18, 2018

Abstract

The permutation language P_n consists of all words that are permutations of a fixed alphabet of size n. Using divide-and-conquer, we construct a regular expression R_n that specifies P_n. We then give explicit bounds for the length of R_n, which we find to be $4^n n^{-(\lg n)/4+\Theta(1)}$, and use these bounds to show that R_n has minimum size over all regular expressions specifying P_n.

1 Introduction

Given a regular language L defined in some way, it is a challenging problem to find good upper and lower bounds on the size of the smallest regular expression specifying L. (In this paper, by a regular expression, we always mean one using the operations of union, concatenation, and Kleene closure only.) Indeed, as a computational problem, it is known that determining the shortest regular expression corresponding to an NFA is PSPACE-hard [10]. Jiang and Ravikumar proved the analogous result for DFAs [9]. For more recent results on inapproximability, see [7].

For nontrivial families of languages, only a handful of results are already known. For example, Ellul et al. [5] showed that the shortest regular expression for the language $\{w \in \{0,1\}^n \ : \ |w|_1 \text{ is even}\}$ is of length $\Omega(n^2)$. Here $|w|_1$ denotes the number of occurrences of the symbol 1 in the word w. (A simple divide-and-conquer strategy provides a matching upper bound.) Chistikov et al. [3] showed that the regular language

$$\{ij \ : \ 1 \leq i < j \leq n\}$$

can be specified by a regular expression of size exactly $n(\lfloor \log_2 n \rfloor + 2) - 2^{\lfloor \log_2 n \rfloor + 1}$, and furthermore this bound is optimal. Mousavi [11] developed a general program for computing lower bounds on regular expression size for the binomial languages

$$B(n, k) = \{w \in \{0, 1\}^n : |w|_1 = k\}.$$

Let n be a positive integer, and define $\Sigma_n = \{1, 2, \ldots, n\}$. In this paper we study the finite language P_n consisting of all permutations of Σ_n. Thus, for example,

$$P_3 = \{123, 132, 213, 231, 312, 321\}.$$

We are interested in regular expressions that specify P_n. In counting the length of regular expressions, we adopt the conventional measure of *alphabetic length* (see, for example, [4]): the length of a regular expression is the number of occurrences of symbols of the alphabet Σ_n. Thus, other symbols, such as parentheses and `+`, are ignored.

A brute-force solution, which consists of listing all the members of P_n and separating them by the union symbol `+`, evidently gives a regular expression for P_n of alphabetic length $n \cdot n!$. This can be improved to $n! \sum_{0 \le i < n} 1/i! \sim e \cdot n!$ by tail recursion, where $E(S)$ represents a regular expression for all permutations of the symbols of S:

$$E(S) = \sum_{i \in S} i(E(S - \{i\})); \quad E(i) = i.$$

For example, for P_4 this gives

```
1(2(34+43)+3(24+42)+4(23+32))+2(1(34+43)+3(14+41)+4(13+31))+
  3(1(24+42)+2(14+41)+4(12+21))+4(1(23+32)+2(13+31)+3(12+21)).
```

Can we do better?

Ellul et al. [5] proved the following weak lower bound: every regular expression for P_n has alphabetic length at least 2^{n-1}. In this note we derive an upper bound through divide-and-conquer. We then show that the regular expression this strategy produces is, in fact, actually optimal. This improves the result from [5]. The language P_n is of particular interest because its complement has short regular expressions, as shown in [5]. For other results concerning context-free grammars for P_n, see [5, 1, 2, 6].

2 Divide-and-conquer

Consider the following divide-and-conquer strategy. Let S be an alphabet of cardinality n. We consider all subsets $T \subseteq S$ of cardinality $\lfloor n/2 \rfloor$. For each subset we recursively determine a regular expression for the permutations of T, a regular expression for the permutations of $S - T$, and concatenate them together. This gives

$$E(S) = \sum_{\substack{T \subseteq S \\ |T| = \lfloor n/2 \rfloor}} (E(T))(E(S - T)); \quad E(i) = i. \tag{1}$$

Finally, we define $R_n = E(\Sigma_n)$.

Thus, for example, we get

$$R_4 = \texttt{(12+21)(34+43)+(13+31)(24+42)+(23+32)(14+41)+} \\ \texttt{(14+41)(23+32)+(24+42)(13+31)+(34+43)(12+21)}$$

for P_4.

The alphabetic length of the resulting regular expression R_n for all permutations of Σ_n is then $f(n)$, where

$$f(n) = \begin{cases} 1, & \text{if } n = 1; \\ \binom{n}{\lfloor n/2 \rfloor}\big(f(\lfloor n/2 \rfloor) + f(\lceil n/2 \rceil)\big), & \text{if } n > 1. \end{cases}$$

The first few values of $f(n)$ are given in the table below.

n	$f(n)$
1	1
2	4
3	15
4	48
5	190
6	600
7	2205
8	6720
9	29988
10	95760

It is sequence A320460 in the *On-Line Encyclopedia of Integer Sequences* [13].

It seems hard to determine a simple closed-form expression for $f(n)$. Nevertheless, we can roughly estimate it as follows, at least when $n = 2^m$ is a power of 2:

$$\begin{aligned} f(2^m) &= 2\binom{2^m}{2^{m-1}} f(2^{m-1}) \\ &= 2^m \binom{2^m}{2^{m-1}}\binom{2^{m-1}}{2^{m-2}} \cdots \binom{2}{1} \\ &= 2^m \frac{(2^m)!}{(2^{m-1})!\,(2^{m-2})!\,\cdots\,2!\,1!}. \end{aligned}$$

Substituting the Stirling approximation $n! \sim \sqrt{2\pi n}(n/e)^n$ and simplifying, we get that $f(2^m)$ is roughly equal to

$$4^{2^m} e^{-1} \pi^{(1-m)/2} 2^{-(m^2-5m+6)/4}.$$

To make this precise, and make it work when n is not a power of 2, however, takes more work.

The rest of the paper is organized as follows: in Section 3, we prove that our regular expression is in fact optimal, assuming one result that is proven at the end of the paper. In Section 4, we establish some inequalities related to Stirling's formula. In Section 5, we connect these inequalities to $f(n)$ and obtain the estimate mentioned in the abstract. Finally, in Section 6 we use our obtained bounds on $f(n)$ to provide the missing piece in our optimality proof.

3 Optimality

In order to show that our regular expression has minimum possible length, we use the following property of $f(n)$ that we prove in Section 6:

Lemma 1. *If $n \geq 1$, then every integer $0 < k < n$ satisfies $\binom{n}{k}(f(k) + f(n-k)) \geq f(n)$. Equality occurs if and only if $k = \lfloor n/2 \rfloor$ or $k = \lceil n/2 \rceil$.*

For $n \geq 1$ and $1 \leq k \leq n!$, define $\ell(n,k)$ to be the minimum alphabetic length of a regular expression specifying a subset of P_n, where the subset has cardinality at least k.

Lemma 2. *If $n \geq 1$ and $1 \leq k \leq n!$, then $\ell(n,k)/k \geq \ell(n,n!)/n! \geq f(n)/n!$.*

Proof. We prove this by induction over the lexicographical ordering of pairs (n,k). This is easy for our base case $n = k = 1$, as the best regular expression is a single character. We thus suppose $n \geq 2$.

Consider a regular expression for a subset of P_n of cardinality at least $k \geq 1$ that has minimum alphabetic length. Clearly no such expression will involve ϵ or $\emptyset$. We now consider the possibilities for the last (outermost) operation in the regular expression. Clearly the only relevant possibilities are union and concatenation.

If the last operation is a union, then it is the union of two subsets of P_n of cardinalities $k_1, k_2 \geq 1$ where $k_1 + k_2 \geq k$, and $k_1, k_2 < k$ by minimality. Then we get

$$\begin{aligned}\frac{\ell(n,k)}{k} &\geq \frac{\ell(n,k_1) + \ell(n,k_2)}{k_1 + k_2} \\ &\geq \min\left\{\frac{\ell(n,k_1)}{k_1}, \frac{\ell(n,k_2)}{k_2}\right\} \\ &\geq \frac{\ell(n,n!)}{n!} \geq \frac{f(n)}{n!}.\end{aligned}$$

If the last operation is a concatenation, then it is the concatenation of two regular expressions for subsets of P_{n_1} and P_{n_2} of cardinalities k_1 and k_2 respectively (possibly after changing alphabets) where $n_1 + n_2 = n$ and $k_1 k_2 \geq k$. By minimality, we have n_1, n_2, k_1, k_2

all positive, so $n_1, n_2 < n$. We now obtain

$$\begin{aligned}
\frac{\ell(n,k)}{k} &\geq \frac{\ell(n_1,k_1)+\ell(n_2,k_2)}{k_1k_2} \\
&\geq \frac{\ell(n_1,n_1!)+\ell(n_2,k_2)}{n_1!\ k_2} \\
&\geq \frac{\ell(n_1,n_1!)+\ell(n_2,n_2!)}{n_1!\ n_2!} \\
&\geq \frac{f(n_1)+f(n_2)}{n_1!\ n_2!} \\
&= \frac{1}{n!}\binom{n}{n_1}(f(n_1)+f(n-n_1)) \\
&\geq \frac{1}{n!}\binom{n}{\lfloor n/2\rfloor}(f(\lfloor n/2\rfloor)+f(\lceil n/2\rceil)) && \text{(by Lemma 1)} \\
&= \frac{f(n)}{n!}.
\end{aligned}$$

In both cases, we get the desired inequalities for these choices of n and k, completing our induction. □

Theorem 3. *Let $n \geq 1$. Over all regular expressions for the permutation language P_n, the regular expression R_n given by our divide-and-conquer strategy achieves the minimum alphabetic length.*

Proof. By our construction from Section 2, the regular expression R_n specifies the entirety of P_n and has alphabetic length $f(n)$. We thus get the upper bound $\ell(n,n!) \leq f(n)$. By Lemma 2, we have the matching lower bound $\ell(n,n!) \geq f(n)$. Thus, R_n has minimum possible alphabetic length for a regular expression specifying P_n. □

4 Analysis

In what follows we use ln to denote the natural logarithm, and lg to denote logarithms to the base 2.

Define $S : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ to be the usual Stirling approximation [12]:

$$S(x) = \sqrt{2\pi x}(x/e)^x.$$

Lemma 4. *For every $x \geq 1$, we have the bounds*

$$S\big(x+\frac{1}{2}\big)^2 \leq S(x)S(x+1) \leq e^{1/(2x)}S\big(x+\frac{1}{2}\big)^2.$$

Proof. The first two derivatives of $\ln S(x)$ are

$$\frac{d}{dx}\ln S(x) = \frac{1}{2x} + \ln x$$
$$\frac{d^2}{dx^2}\ln S(x) = -\frac{1}{2x^2} + \frac{1}{x},$$

and so we see that $\ln S(x)$ is convex (that is, its derivative is increasing) for all $x > 1/2$. Thus, by Jensen's inequality [8] and exponentiating, we obtain the lower bound

$$S(x)S(x+1) \geq S\big(x + \frac{1}{2}\big)^2$$

for all $x \geq 1$.

Now, using the mean value theorem twice, we get that

$$(\ln S(x)) + \frac{1}{2}\mu \leq \ln S(x + \frac{1}{2}) \tag{2}$$
$$\ln S(x+1) \leq (\ln S(x + \frac{1}{2})) + \frac{1}{2}M, \tag{3}$$

where

$$\mu = \inf_{z \in [x, x+\frac{1}{2}]} (\ln S(z))' = \frac{1}{2x} + \ln x$$
$$M = \sup_{z \in [x+\frac{1}{2}, x+1]} (\ln S(z))' = \frac{1}{2(x+1)} + \ln(x+1).$$

Adding the inequalities (2) and (3), we get

$$\begin{aligned}(\ln S(x)) + (\ln S(x+1)) &\leq (2\ln S(x + 1/2)) - \mu/2 + M/2 \\ &\leq (2\ln S(x+1/2)) + \frac{1}{2}\ln(\frac{x+1}{x}) \\ &\leq (2\ln S(x+1/2)) + \frac{1}{2x}.\end{aligned}$$

This gives us the inequality

$$S(x)S(x+1) \leq e^{1/(2x)} S(x+1/2)^2$$

for all $x \geq 1$. ∎

Next, for $\alpha \in \mathbb{R}$, define the function $g_\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ by

$$g_\alpha(x) = \frac{4^x}{x^{(\lg x)/4}} x^\alpha.$$

Our goal is to show that f can be approximated by g_α for some choice of α.

Lemma 5. *Let $\alpha > 0$. Then for every $x \geq 4^\alpha$ we have*

$$e^{-1/(2\sqrt{x})}\frac{5}{2}g_\alpha(x+\frac{1}{2}) \leq g_\alpha(x) + g_\alpha(x+1) \leq e^{1/(2\sqrt{x})}\frac{5}{2}g_\alpha(x+\frac{1}{2}).$$

Proof. We again compute the logarithmic derivative:

$$\frac{d}{dx}\ln g_\alpha(x) = \frac{\alpha - (\lg x)/2}{x} + \ln 4.$$

For $x \geq 4^\alpha$, this derivative is at most $\ln 4$, so by the mean value theorem,

$$\ln g_\alpha(x+1) - \ln 2 \leq \ln g_\alpha(x+\frac{1}{2}) \leq \ln g_\alpha(x) + \ln 2.$$

Exponentiating, we get

$$\frac{1}{2}g_\alpha(x+1) \leq g_\alpha(x+\frac{1}{2}) \leq 2g_\alpha(x). \tag{4}$$

Next, we note that the derivative of $\ln x$ exceeds that of $\sqrt{x}$ for $0 < x < 4$ and is less for $x > 4$. So, since $\ln 4 < \sqrt{4}$, we have $\ln x < \sqrt{x}$ for all $x > 0$. Hence

$$\begin{aligned}\frac{d}{dx}\ln g_\alpha(x) &= \frac{\alpha}{x} - \frac{\ln x}{2x\ln 2} + \ln 4 \\ &\geq \ln 4 - \frac{\sqrt{x}}{(2\ln 2)x} \\ &\geq \ln 4 - 1/\sqrt{x}\end{aligned}$$

for all $x > 0$. Thus, by the mean value theorem and exponentiating again, we get

$$\frac{1}{2}e^{1/(2\sqrt{x})}g_\alpha(x+1) \geq g_\alpha(x+\frac{1}{2}) \geq 2e^{-1/(2\sqrt{x})}g_\alpha(x). \tag{5}$$

We can now combine the inequalities (4) and (5) for $x \geq 4^\alpha$ to get

$$\begin{aligned}e^{-1/(2\sqrt{x})}\frac{5}{2}g_\alpha(x+\frac{1}{2}) &\leq \frac{1}{2}g_\alpha(x+\frac{1}{2}) + 2e^{-1/(2\sqrt{x})}g_\alpha(x+\frac{1}{2}) \\ &\leq g_\alpha(x) + g_\alpha(x+1) \\ &\leq 2e^{1/(2\sqrt{x})}g_\alpha(x+\frac{1}{2}) + \frac{1}{2}g_\alpha(x+\frac{1}{2}) \\ &\leq e^{1/(2\sqrt{x})}\frac{5}{2}g_\alpha(x+\frac{1}{2}),\end{aligned}$$

which gives us both desired bounds. □

We now show an identity relating g_α and S.

Lemma 6. *Suppose $x > 0$ and $\beta > 0$. If $\alpha = \lg\beta + 1/4 - (\lg\pi)/2$, then*

$$\beta\frac{S(2x)}{S(x)^2}g_\alpha(x) = g_\alpha(2x).$$

Proof. We have

$$\begin{aligned}
\beta \frac{S(2x)}{S(x)^2} g_\alpha(x) &= \beta \frac{\sqrt{4\pi x}(2x/e)^{2x}}{(\sqrt{2\pi x}(x/e)^x)^2} \frac{4^x}{x^{\lg x/4}} x^\alpha \\
&= \beta \frac{4^x}{\sqrt{\pi x}} \frac{4^x}{x^{\lg x/4}} x^\alpha \\
&= \frac{2^{\lg \beta + 1/4}}{\pi^{1/2} x^{1/4} x^{1/4} 2^{1/4}} \frac{4^{2x}}{x^{\lg x/4}} x^\alpha \\
&= 2^{\lg \beta + 1/4 - \lg \pi/2} \frac{4^{2x}}{2^{(1+\lg x)/4} x^{(1+\lg x)/4}} x^\alpha \\
&= \frac{4^{2x}}{(2x)^{(\lg 2x)/4}} (2x)^\alpha \\
&= g_\alpha(2x).
\end{aligned}$$

□

5 Bounds on $f(n)$

In this section we obtain an estimate for $f(n)$, the size of the optimal regular expression for P_n.

Theorem 7. *For all $n \geq 1$ we have*

$$0.195 \frac{4^n}{n^{(\lg n)/4}} n^{5/4 - (\lg \pi)/2} \leq f(n) \leq \frac{1}{4} \frac{4^n}{n^{(\lg n)/4}} n^{(\lg 5) - 3/4 - (\lg \pi)/2}.$$

Further, when n is a power of two, we get the following upper bound, matching the general lower bound.

$$f(n) \leq \frac{1}{4} \frac{4^n}{n^{\lg n/4}} n^{5/4 - (\lg \pi)/2}.$$

Proof. Recall the Stirling approximation

$$e^{1/(12n+1)} S(n) \leq n! \leq e^{1/12n} S(n); \tag{6}$$

see [12]. Now suppose that $f(n) \leq r_n g_\alpha(n)$ and $f(n+1) \leq r_{n+1} g_\alpha(n+1)$, where $n \geq \max\{1, 4^\alpha\}$, for some non-decreasing function $r : \mathbb{N} \to \mathbb{R}_{>0}$. Then by combining Lemma 4, Lemma 5, and equation (6), we get

$$\begin{aligned}
f(2n+1) &= \binom{2n+1}{n} (f(n) + f(n+1)) \\
&\leq e^{\frac{1}{12(2n+1)} - \frac{1}{12n+1} - \frac{1}{12(n+1)+1}} \frac{S(2n+1)}{S(n)S(n+1)} (r_n g_\alpha(n) + r_{n+1} g_\alpha(n+1)) \\
&\leq \frac{5}{2} r_{n+1} e^{1/(2\sqrt{n})} \frac{S(2n+1)}{S(n+1/2)^2} g_\alpha(n+1/2)
\end{aligned}$$

and

$$
\begin{aligned}
f(2n) &= \binom{2n}{n}(f(n)+f(n)) \\
&\le 2r_n e^{\frac{1}{12(2n)}-2\frac{1}{12n+1}} \frac{S(2n)}{S(n)^2} g_\alpha(n) \\
&\le 2r_n \frac{S(2n)}{S(n)^2} g_\alpha(n).
\end{aligned}
$$

For the case where n is a power of two only, we use $\beta = 2$ and $r_n = C$, we set

$$
\alpha = \lg \beta + 1/4 - (\lg \pi)/2 = 5/4 - (\lg \pi)/2,
$$

so $\alpha > 0$ and $4^\alpha < 2$. Now Lemma 6 gives us the identity $2\frac{S(2x)}{S(x)^2} g_\alpha(x) = g_\alpha(2x)$. Then by induction we have $f(n) \le C g_\alpha(n)$ for all $n \ge 1$, where C is any constant that satisfies this bound for $n < 4$. In particular, $C = \frac{1}{4}$ works, so we have $f(n) \le \frac{1}{4} g_{5/4-\lg \pi/2}(n)$ for all $n \ge 1$ that are powers of two.

Next, for general n, we use $\beta = 5/2$ and $r_n = Ce^{-\frac{\sqrt{5}}{(4-\sqrt{10})\sqrt{n}}}$, we set

$$
\alpha = \lg \beta + 1/4 - \lg \pi/2 = \lg 5 - 3/4 - \lg \pi/2,
$$

so $\alpha > 0$ and $4^\alpha < 4$. Now Lemma 6 gives us the identity $\frac{5}{2}\frac{S(2x)}{S(x)^2} g_\alpha(x) = g_\alpha(2x)$. For $n \ge 4$, we get

$$
\begin{aligned}
r_{n+1} e^{1/(2\sqrt{n})} &= C e^{\frac{1}{2\sqrt{n}} - \frac{\sqrt{5}}{(4-\sqrt{10})\sqrt{n+1}}} \\
&\le C\big(e^{\frac{1}{\sqrt{n}}}\big)^{\frac{1}{2} - \frac{\sqrt{5}}{4-\sqrt{10}}\frac{\sqrt{4}}{\sqrt{5}}} \\
&= C\big(e^{\frac{1}{\sqrt{n}}}\big)^{-\frac{\sqrt{10}}{2(4-\sqrt{10})}} \\
&= C e^{-\frac{\sqrt{5}}{(4-\sqrt{10})\sqrt{2n}}} \\
&\le r_{2n+1}.
\end{aligned}
$$

Easily, we also get $2r_n \le \frac{5}{2} r_{2n}$. Thus, by induction we have $f(n) \le r_n g_\alpha(n)$ for all $n \ge 12$, where C is chosen to make this work for $12 \le n < 24$. In particular, $C = \frac{1}{4}$ works again. Further, since we have $r_n < C$ for all $n \ge 1$, we also have $f(n) \le \frac{1}{4} g_{\lg 5 - 3/4 - \lg \pi/2}(n)$ for all $n \ge 12$. Finally, we check manually that this last inequality holds for $1 \le n < 12$ too, and thus for all $n \ge 1$.

All that remains is the lower bound. We get similar recurrences, supposing $f(n) \ge r_n g_\alpha(n)$ and $f(n+1) \ge r_{n+1} g_\alpha(n+1)$, where $n \ge \max\{1, 4^\alpha\}$ for some non-increasing

$r : \mathbb{N} \to \mathbb{R}_{>0}$. Then by a similar argument as for the upper bounds, we have

$$\begin{aligned} f(2n+1) &= \binom{2n+1}{n}(f(n)+f(n+1)) \\ &\geq \frac{5}{2} r_n e^{\frac{1}{12(2n+1)+1} - \frac{1}{12n} - \frac{1}{12(n+1)}} e^{-1/(2\sqrt{n})} e^{-1/(2n)} \frac{S(2n+1)}{S(n+1/2)^2} g_\alpha(n+1/2) \\ &\geq \frac{5}{2} r_n e^{-1/(2\sqrt{n})-2/(3n)} \frac{S(2n+1)}{S(n+1/2)^2} g_\alpha(n+1/2) \end{aligned}$$

and

$$\begin{aligned} f(2n) &= \binom{2n}{n}(f(n)+f(n)) \\ &\geq 2 r_n e^{\frac{1}{24n+1} - \frac{2}{12n}} \frac{S(2n)}{S(n)^2} g_\alpha(n) \\ &\geq 2 r_n e^{-1/(6n)} \frac{S(2n)}{S(n)^2} g_\alpha(n). \end{aligned}$$

This time, we set $\beta = 2$ with $r_n = Ce^{1/3n}$ (indeed non-increasing), and $\alpha = 5/4 - \lg \pi/2$, noting $4^\alpha < 4$. Now for $n = 9$, we have $\ln \frac{5}{4} \geq \frac{2}{9} = \frac{1}{2\sqrt{n}} + \frac{1}{2n}$. Since the right-hand side of this inequality is non-increasing in n, the bound holds for all $n \geq 9$. This implies

$$\begin{aligned} \frac{5}{2} e^{-1/(2\sqrt{n})-2/(3n)} r_n &\geq \frac{5}{2} e^{-1/(6n)-\ln(5/4)} r_n \\ &= 2r_{2n} \\ &\geq 2r_{2n+1}. \end{aligned}$$

Further, $2r_n e^{-1/6n} = 2r_{2n}$, so by induction we have $f(n) \geq r_n g_\alpha(n)$ for all $n \geq 17$, where C is chosen to satisfy this bound for $17 \leq n < 34$. In particular, $C = 0.195$ works. Since $r_n > C$ for all $n \geq 17$, we also have $f(n) \geq 0.195 g_{5/4-\lg \pi/2}(n)$ for all $n \geq 17$. Finally, we check manually that this works for all $1 \leq n < 17$ too, and thus for all $n \geq 1$. □

6 Optimality revisited

We now give a simple lower bound on the growth of f.

Lemma 8. *We have $f(n+1) \geq 3f(n)$ for all $n \geq 1$.*

Proof. We prove this by induction on n. It is easy to verify the base case $f(2) = 4 \geq 3 = 3f(1)$. Otherwise $n > 1$. Suppose the desired inequality holds for all smaller values of n. If

$n \geq 2$ is odd, then let $1 \leq m < n$ satisfy $2m+1 = n$. Then

$$\begin{aligned} f(n+1) &= f(2m+2) \\ &= 2\binom{2m+2}{m+1} f(m+1) \\ &= 2\frac{2m+2}{m+1}\binom{2m+1}{m} f(m+1) \\ &= \binom{2m+1}{m}(f(m+1) + 3f(m+1)) \\ &\geq \binom{2m+1}{m}(3f(m) + 3f(m+1)) \\ &= 3f(2m+1) \\ &= 3f(n). \end{aligned}$$

Otherwise, if $n \geq 2$ is even, then let $1 \leq m < n$ satisfy $2m = n$. We note that $4m+2 \geq 3m+3$, so $2\frac{2m+1}{m+1} \geq 3$. We then have

$$\begin{aligned} f(n+1) &= f(2m+1) \\ &= \binom{2m+1}{m}(f(m) + f(m+1)) \\ &= \frac{2m+1}{m+1}\binom{2m}{m}(f(m) + f(m+1)) \\ &\geq \frac{2m+1}{m+1}\binom{2m}{m}(f(m) + 3f(m)) \\ &= 2\frac{2m+1}{m+1} f(2m) \\ &\geq 3f(2m) \\ &= 3f(n). \end{aligned}$$

□

Armed with this inequality and the bounds given by Theorem 7 of Section 5, we are ready to complete our proof of the optimality of R_n. We recall Lemma 1, which is what we have left to show:

Lemma 1. *If $n \geq 1$, then every integer $0 < k < n$ satisfies $\binom{n}{k}(f(k) + f(n-k)) \geq f(n)$. Equality occurs if and only if $k = \lfloor n/2 \rfloor$ or $k = \lceil n/2 \rceil$.*

Proof. We easily check the cases $n < 12$ by hand. Let $n \geq 12$ be arbitrary. It suffices to consider the cases where $k \leq \lfloor n/2 \rfloor$, as those where $k \geq \lceil n/2 \rceil$ are symmetric. Equality for the case $k = \lfloor n/2 \rfloor$ is given by the definition of $f(n)$.

Suppose that $n/6 \leq k < \lfloor n/2 \rfloor$. Then $n > 9$, so $3n - 3 > 2n + 6$. Hence we get

$$\frac{n/2 - 1/2}{n/2 + 3/2} > 2/3$$

and so

$$\frac{\lfloor n/2\rfloor}{\lceil n/2\rceil+1} > 2/3. \tag{7}$$

Also, from $k \geq n/6$ we get $3k+3 \geq \lceil n/2\rceil + 2$, and so

$$\frac{k+1}{\lceil n/2\rceil+2} \geq 1/3. \tag{8}$$

Then

$$\begin{aligned}
&\binom{n}{k}(f(k)+f(n-k)) \\
&\geq \binom{n}{k} f(n-k) \\
&\geq \binom{n}{k} 3^{\lfloor n/2\rfloor - k} f(\lceil n/2\rceil) && \text{(by Lemma 8)} \\
&\geq 3^{\lfloor n/2\rfloor - k} \binom{n}{k} \frac{f(\lfloor n/2\rfloor)+f(\lceil n/2\rceil)}{2} \\
&= \frac{3^{\lfloor n/2\rfloor - k}}{2} \prod_{k\leq j<\lfloor n/2\rfloor} \frac{j+1}{n-j} \binom{n}{\lfloor n/2\rfloor} (f(\lfloor n/2\rfloor)+f(\lceil n/2\rceil)) \\
&= \frac{3^{\lfloor n/2\rfloor - k}}{2} \frac{\prod_{k<j\leq\lfloor n/2\rfloor} j}{\prod_{\lceil n/2\rceil<j\leq n-k} j} f(n) \\
&= \frac{3^{\lfloor n/2\rfloor - k}}{2} \frac{\lfloor n/2\rfloor}{\lceil n/2\rceil+1} \frac{\prod_{k+1\leq j\leq\lfloor n/2\rfloor-1} j}{\prod_{\lceil n/2\rceil+2\leq j\leq n-k} j} f(n) \\
&> \frac{3^{\lfloor n/2\rfloor - k}}{2} (2/3) \prod_{1\leq j\leq\lfloor n/2\rfloor-k-1} \frac{k+j}{\lceil n/2\rceil+1+j} f(n) && \text{(by (7))} \\
&\geq 3^{\lfloor n/2\rfloor - k-1} \left(\frac{k+1}{\lceil n/2\rceil+2}\right)^{\lfloor n/2\rfloor-k-1} f(n) \\
&\geq 3^{\lfloor n/2\rfloor - k-1} (1/3)^{\lfloor n/2\rfloor-k-1} f(n) && \text{(by (8))} \\
&= f(n).
\end{aligned}$$

Next, suppose $1 \leq k < n/6$. Then $4k < n - n/3$ and so $4 < \frac{n-1}{k}$. Then if $k > 1$,

$$\begin{aligned}
\binom{n}{k} &= n \prod_{2\leq j\leq k} \frac{n-1-k+j}{j} \\
&\geq n\left(\frac{n-1}{k}\right)^{k-1} \\
&\geq n4^{k-1}.
\end{aligned}$$

We note also that $\binom{n}{1} = n4^0$, so we have $\binom{n}{k} \geq n4^{k-1}$ for all $1 \leq k < n/6$. We note from our proof of Lemma 5 that the derivative of $\ln g_\alpha(x)$ is at most $\ln 4$ for $x \geq 4^\alpha$. In particular, for $\alpha = 5/4 - \lg \pi/2$, this derivative is at most $\ln 4$ for all $x \geq 2$, and so $4^k g_\alpha(n-k) \geq g_\alpha(n)$ here (as $n - k > 5n/6 \geq 10$). Thus

$$\begin{aligned}
&\binom{n}{k}(f(k) + f(n-k)) \\
&\geq \binom{n}{k} f(n-k) \\
&\geq n4^{k-1}\left(0.195 g_{5/4-\lg \pi/2}(n-k)\right) && \text{(by Theorem 7)} \\
&\geq n4^{k-1}\left(0.195 \frac{g_{5/4-\lg \pi/2}(n)}{4^k}\right) \\
&= 0.195 n \frac{1}{4} g_{5/4-\lg \pi/2}(n) \\
&= 0.195 n \frac{1}{4} n^{2-\lg 5} g_{\lg 5 - 3/4 - \lg \pi/2}(n) \\
&\geq 0.195 n^{3-\lg 5} f(n) && \text{(by Theorem 7)} \\
&> f(n) && \text{(since } n \geq 12 > 0.195^{-\frac{1}{3-\lg 5}}\text{).}
\end{aligned}$$

□

References

[1] P. R. J. Asveld, Generating all permutations by context-free grammars in Chomsky normal form, *Theoret. Comput. Sci.* **354** (2006), 118–130.

[2] P. R. J. Asveld, Generating all permutations by context-free grammars in Greibach normal form, *Theoret. Comput. Sci.* **409** (2008), 565–577.

[3] D. Chistikov, S. Ivan, A. Lubiw, and J. Shallit, Fractional coverings, greedy coverings, and rectifier networks. In H. Vollmer and B. Vallée, editors, *34th Symposium on Theoretical Aspects of Computer Science (STACS 2017)*, Vol. 66 of *LIPIcs*, pp. 23:1–23:14. Schloss Dagstuhl — Leibniz-Zentrum für Informatik, 2017.

[4] A. Ehrenfeucht and P. Zeiger, Complexity measures for regular expressions, *J. Comput. System Sci.* **12** (1976), 134–146.

[5] K. Ellul, B. Krawetz, J. Shallit, and M.-w. Wang, Regular expressions: new results and open problems, *J. Autom. Lang. Combin.* **10** (2005), 407–437.

[6] Y. Filmus, Lower bounds for context-free grammars, *Info. Proc. Letters* **111** (2011), 895–898.

[7] G. Gramlich and G. Schnitger, Minimizing nfas and regular expressions, *J. Comput. System Sci.* **73** (2007), 908–923.

[8] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Math.* **30** (1906), 175–193.

[9] T. Jiang and B. Ravikumar, Minimal NFA problems are hard, *SIAM J. Comput.* **22** (1993), 1117–1141.

[10] A. R. Meyer and L. J. Stockmeyer, The equivalence problem for regular expressions with squaring requires exponential space, In *Proc. 13th Ann. IEEE Symp. on Switching and Automata Theory*, pp. 125–129. IEEE, 1972.

[11] H. Mousavi, Lower bounds on regular expression size. Preprint available at https://arxiv.org/abs/1712.00811, 2017.

[12] H. Robbins, A remark on Stirling's formula, *Amer. Math. Monthly* **62** (1955), 26–29.

[13] N. J. A. Sloane et al., The on-line encyclopedia of integer sequences. Available at https://oeis.org, 2018.