Growth properties of power-free languages

Arseny M. Shur

Ural State University, Ekaterinburg, Russia

November 10, 2009

Counting Maps for Formal Languages

The study of growth properties of formal languages can be traced back to Morse and Hedlund, 1938-40.

Since then: lots of computer science papers about counting maps (words, factors, subsequences, patterns, palindromes, etc in languages or infinite words are counted).

All these counting maps are called complexities. We count only words (combinatorial complexity): for $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$.

At the same time, algebraists studied growth maps (=combinatorial complexity) and growth functions (formal power series with the coefficients given by growth maps) for the languages arised from different algebras.

Counting Maps for Formal Languages

The study of growth properties of formal languages can be traced back to Morse and Hedlund, 1938-40.

Since then: lots of computer science papers about counting maps (words, factors, subsequences, patterns, palindromes, etc in languages or infinite words are counted).

All these counting maps are called complexities. We count only words (combinatorial complexity): for $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$.

At the same time, algebraists studied growth maps (=combinatorial complexity) and growth functions (formal power series with the coefficients given by growth maps) for the languages arised from different algebras.

Counting Maps for Formal Languages

The study of growth properties of formal languages can be traced back to Morse and Hedlund, 1938-40.

Since then: lots of computer science papers about counting maps (words, factors, subsequences, patterns, palindromes, etc in languages or infinite words are counted).

All these counting maps are called complexities. We count only words (combinatorial complexity): for $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$.

At the same time, algebraists studied growth maps (=combinatorial complexity) and growth functions (formal power series with the coefficients given by growth maps) for the languages arised from different algebras.

Counting Maps for Formal Languages

The study of growth properties of formal languages can be traced back to Morse and Hedlund, 1938-40.

Since then: lots of computer science papers about counting maps (words, factors, subsequences, patterns, palindromes, etc in languages or infinite words are counted).

All these counting maps are called complexities. We count only words (combinatorial complexity): for $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$.

At the same time, algebraists studied growth maps (=combinatorial complexity) and growth functions (formal power series with the coefficients given by growth maps) for the languages arised from different algebras.

Combinatorial Complexity and Growth Rate

Combinatorial complexity: $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$

Growth rate: $\alpha(L) = \limsup_{n\to\infty} (C_L(n))^{1/n}$.

If L is factorial (closed under taking factors of the words), then $\alpha(L) = \lim_{n\to\infty} (C_L(n))^{1/n} = \inf (C_L(n))^{1/n}$.

- $\alpha(L) = 0$: L is finite
- $\alpha(L) > 1$: L is exponential
- $\alpha(L) = 1$: L is subexponential

Combinatorial Complexity and Growth Rate

Combinatorial complexity: $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$

Growth rate: $\alpha(L) = \limsup_{n\to\infty} (C_L(n))^{1/n}$.

If L is factorial (closed under taking factors of the words), then $\alpha(L) = \lim_{n\to\infty} (C_L(n))^{1/n} = \inf (C_L(n))^{1/n}$.

- $\alpha(L) = 0$: L is finite
- $\alpha(L) > 1$: L is exponential
- $\alpha(L) = 1$: L is subexponential

Combinatorial Complexity and Growth Rate

Combinatorial complexity: $L \in \Sigma^*$, $C_L(n) = |L \cap \Sigma^n|$

Growth rate: $\alpha(L) = \limsup_{n\to\infty} (C_L(n))^{1/n}$.

If L is factorial (closed under taking factors of the words), then $\alpha(L) = \lim_{n\to\infty} (C_L(n))^{1/n} = \inf\,(C_L(n))^{1/n}$.

- $\alpha(L) = 0$: L is finite
- $\alpha(L) > 1$: L is exponential
- $\alpha(L) = 1$: L is subexponential

Power-Free Languages

Integral powers: $mama = (ma)^2$.
Fractional powers: $template = (templa)^{4/3}$.
Exponents: $\exp(mama) = 2$, $\exp(template) = 4/3$
Roots: $\mathrm{root}(mama) = ma$, $\mathrm{root}(template) = templa$

Power-freeness: a word is β-free (β^+-free) if the exponents of all its factors are strictly less than β (resp., at most β). We always write "β-free", assuming β to be an extended rational. A β-free language consists of all β-free words over a fixed alphabet. We denote it $L(k, \beta)$.

Combinatorial complexity of these languages is intensively studied since 1980's. Particular languages were considered, no universal instrument was known.

Power-Free Languages

Integral powers: $mama = (ma)^2$.
Fractional powers: $template = (templa)^{4/3}$.
Exponents: $\exp(mama) = 2$, $\exp(template) = 4/3$
Roots: $\mathrm{root}(mama) = ma$, $\mathrm{root}(template) = templa$

Power-freeness: a word is β-free (β^+-free) if the exponents of all its factors are strictly less than β (resp., at most β). We always write "β-free", assuming β to be an extended rational. A β-free language consists of all β-free words over a fixed alphabet. We denote it $L(k, \beta)$.

Combinatorial complexity of these languages is intensively studied since 1980's. Particular languages were considered, no universal instrument was known.

Power-Free Languages

Integral powers: $mama = (ma)^2$.
Fractional powers: $template = (templa)^{4/3}$.
Exponents: $\exp(mama) = 2$, $\exp(template) = 4/3$
Roots: $\mathrm{root}(mama) = ma$, $\mathrm{root}(template) = templa$

Power-freeness: a word is β-free (β^+-free) if the exponents of all its factors are strictly less than β (resp., at most β). We always write "β-free", assuming β to be an extended rational. A β-free language consists of all β-free words over a fixed alphabet. We denote it $L(k, \beta)$.

Combinatorial complexity of these languages is intensively studied since 1980's. Particular languages were considered, no universal instrument was known.

Finite and Infinite: Dejean's Conjecture

To study combinatorial complexity of power-free languages, it is useful to know which of them are infinite. If a β-free language over a k-letter alphabet is infinite, and $\gamma > \beta$, then the γ-free language over this alphabet is also infinite. But what is the minimal such β? The answer was conjectured as follows:

Conjecture (Dejean, 1972)

The minimal infinite β-free languages are: $L(3, \frac{7}{4}^+)$, $L(4, \frac{7}{5}^+)$, and $L(k, \frac{k}{k-1}^+)$ for $k = 2$ and any $k \geq 5$.

Finally proved in 2009. Different particular cases are due to Thue, Dejean, Pansiot, Moulin-Ollagnier, Mohammad-Noori, Currie, Carpi, Rampersad, and Rao.

Finite and Infinite: Dejean's Conjecture

To study combinatorial complexity of power-free languages, it is useful to know which of them are infinite. If a β-free language over a k-letter alphabet is infinite, and $\gamma > \beta$, then the γ-free language over this alphabet is also infinite. But what is the minimal such β? The answer was conjectured as follows:

Conjecture (Dejean, 1972)

The minimal infinite β-free languages are: $L(3, \frac{7}{4}^+)$, $L(4, \frac{7}{5}^+)$, and $L(k, \frac{k}{k-1}^+)$ for $k = 2$ and any $k \geq 5$.

Finally proved in 2009. Different particular cases are due to Thue, Dejean, Pansiot, Moulin-Ollagnier, Mohammad-Noori, Currie, Carpi, Rampersad, and Rao.

Finite and Infinite: Dejean's Conjecture

To study combinatorial complexity of power-free languages, it is useful to know which of them are infinite. If a β-free language over a k-letter alphabet is infinite, and $\gamma > \beta$, then the γ-free language over this alphabet is also infinite. But what is the minimal such β? The answer was conjectured as follows:

Conjecture (Dejean, 1972)

The minimal infinite β-free languages are: $L(3, \frac{7}{4}^+)$, $L(4, \frac{7}{5}^+)$, and $L(k, \frac{k}{k-1}^+)$ for $k = 2$ and any $k \geq 5$.

Finally proved in 2009. Different particular cases are due to Thue, Dejean, Pansiot, Moulin-Ollagnier, Mohammad-Noori, Currie, Carpi, Rampersad, and Rao.

Growth of Power-free Languages: Known Results

- $L(2, 2^+)$ has polynomial complexity (Restivo, Salemi, 1984) which is $\Omega(n^{1,22})$ and $O(n^{1,37})$ (Lepisto, 1996), but cannot be expressed as $\Theta(n^\alpha)$ for a single number α (Cassaigne, 1993);
- $L(2, \beta)$ has polynomial complexity whenever $2^+ \leq \beta \leq \frac{7}{3}$ and exponential complexity if $\beta \geq \frac{7}{3}^+$ (Karhumaki, Shallit, 2003), while all infinite languages $L(k, \beta)$ with $k \geq 3$ have exponential complexity (Conjecture by Shallit);
- upper bounds for the growth rates of some particular power-free languages are known, like 1,4576 for $L(2, 3)$ (Edlin, 1999), 1,2299 for $L(2, \frac{7}{3}^+)$ (Karhumaki, Shallit, 2003), 1,301788 for $L(3, 2)$ (Ochem, 2006), as well as some "existential" lower bounds for such growth rates.

Growth of Power-free Languages: Known Results

- $L(2, 2^+)$ has polynomial complexity (Restivo, Salemi, 1984) which is $\Omega(n^{1{,}22})$ and $O(n^{1{,}37})$ (Lepisto, 1996), but cannot be expressed as $\Theta(n^\alpha)$ for a single number α (Cassaigne, 1993);
- $L(2, \beta)$ has polynomial complexity whenever $2^+ \le \beta \le \frac{7}{3}$ and exponential complexity if $\beta \ge \frac{7}{3}^+$ (Karhumaki, Shallit, 2003), while all infinite languages $L(k, \beta)$ with $k \ge 3$ have exponential complexity (Conjecture by Shallit);
- upper bounds for the growth rates of some particular power-free languages are known, like $1{,}4576$ for $L(2, 3)$ (Edlin, 1999), $1{,}2299$ for $L(2, \frac{7}{3}^+)$ (Karhumaki, Shallit, 2003), $1{,}301788$ for $L(3, 2)$ (Ochem, 2006), as well as some “existential” lower bounds for such growth rates.

Growth of Power-free Languages: Known Results

- $L(2,2^+)$ has polynomial complexity (Restivo, Salemi, 1984) which is $\Omega(n^{1{,}22})$ and $O(n^{1{,}37})$ (Lepisto, 1996), but cannot be expressed as $\Theta(n^\alpha)$ for a single number α (Cassaigne, 1993);
- $L(2,\beta)$ has polynomial complexity whenever $2^+ \le \beta \le \frac{7}{3}$ and exponential complexity if $\beta \ge \frac{7}{3}^+$ (Karhumaki, Shallit, 2003), while all infinite languages $L(k,\beta)$ with $k \ge 3$ have exponential complexity (Conjecture by Shallit);
- upper bounds for the growth rates of some particular power-free languages are known, like 1,4576 for $L(2,3)$ (Edlin, 1999), 1,2299 for $L(2,\frac{7}{3}^+)$ (Karhumaki, Shallit, 2003), 1,301788 for $L(3,2)$ (Ochem, 2006), as well as some "existential" lower bounds for such growth rates.

Growth of Power-free Languages: Known Results

- $L(2,2^+)$ has polynomial complexity (Restivo, Salemi, 1984) which is $\Omega(n^{1,22)}$ and $O(n^{1,37})$ (Lepisto, 1996), but cannot be expressed as $\Theta(n^\alpha)$ for a single number α (Cassaigne, 1993);
- $L(2,\beta)$ has polynomial complexity whenever $2^+ \le \beta \le \frac{7}{3}$ and exponential complexity if $\beta \ge \frac{7}{3}^+$ (Karhumaki, Shallit, 2003), while all infinite languages $L(k,\beta)$ with $k \ge 3$ have exponential complexity (Conjecture by Shallit);
- upper bounds for the growth rates of some particular power-free languages are known, like $1{,}4576$ for $L(2,3)$ (Edlin, 1999), $1{,}2299$ for $L(2,\frac{7}{3}^+)$ (Karhumaki, Shallit, 2003), $1{,}301788$ for $L(3,2)$ (Ochem, 2006), as well as some “existential” lower bounds for such growth rates.

Upper Bounds: Reduction

Power-free languages are factorial (closed under taking factors of words). Factorial languages can be defined in terms of antidictionaries (sets of minimal forbidden words).

For example, $L = \mathtt{Fact}((ab)^*)$ is defined by $M = \{aa, bb\}$.

Languages with regular (e.g., finite) antidictionary are regular.

Method of approximation:

Let M be the antidictionary for L, $M_j = M \cap \Sigma^{\leq j}$, and let L_j be the factorial language over Σ with the finite antidictionary M_j. Then

$$M_1 \subseteq \ldots \subseteq M_j \subseteq \ldots \subseteq M, \ \bigcup_{j=1}^{\infty} M_j = M, \ \ L \subseteq \ldots \subseteq L_j \subseteq \ldots \subseteq L_1, \ \bigcap_{j=1}^{\infty} L_j = L,$$

and for any n $C_L(n) = \ldots = C_{L_n}(n) \leq \ldots \leq C_{L_1}(n)$.

Then get $C_{L_j}(n) \to C_L(n)$ and $\alpha(L_j) \to \alpha(L)$ as $j \to \infty$.

Upper Bounds: Reduction

Power-free languages are factorial (closed under taking factors of words). Factorial languages can be defined in terms of antidictionaries (sets of minimal forbidden words).
For example, $L = \mathtt{Fact}((ab)^*)$ is defined by $M = \{aa, bb\}$.

Languages with regular (e.g., finite) antidictionary are regular.

Method of approximation:
Let M be the antidictionary for L, $M_j = M \cap \Sigma^{\le j}$, and let L_j be the factorial language over Σ with the finite antidictionary M_j. Then

$$M_1 \subseteq \ldots \subseteq M_j \subseteq \ldots \subseteq M, \ \bigcup_{j=1}^{\infty} M_j = M, \ \ L \subseteq \ldots \subseteq L_j \subseteq \ldots \subseteq L_1, \ \bigcap_{j=1}^{\infty} L_j = L,$$

and for any n $C_L(n) = \ldots = C_{L_n}(n) \le \ldots \le C_{L_1}(n)$.
Then get $C_{L_j}(n) \to C_L(n)$ and $\alpha(L_j) \to \alpha(L)$ as $j \to \infty$.

Upper Bounds: Reduction

Power-free languages are factorial (closed under taking factors of words). Factorial languages can be defined in terms of antidictionaries (sets of minimal forbidden words).

For example, $L = \mathtt{Fact}((ab)^*)$ is defined by $M = \{aa, bb\}$.

Languages with regular (e.g., finite) antidictionary are regular.

Method of approximation:

Let M be the antidictionary for L, $M_j = M \cap \Sigma^{\leq i}$, and let L_j be the factorial language over Σ with the finite antidictionary M_j. Then

$$M_1 \subseteq \ldots \subseteq M_j \subseteq \ldots \subseteq M, \ \bigcup_{j=1}^{\infty} M_j = M, \quad L \subseteq \ldots \subseteq L_j \subseteq \ldots \subseteq L_1, \ \bigcap_{j=1}^{\infty} L_j = L,$$

and for any n $C_L(n) = \ldots = C_{L_n}(n) \leq \ldots \leq C_{L_1}(n)$.

Then get $C_{L_j}(n) \to C_L(n)$ and $\alpha(L_j) \to \alpha(L)$ as $j \to \infty$.

Upper Bounds for Growth Rates: Algorithm

We have: power-free language $L = L(k, \beta)$

We need efficient algorithms (1) to build a dfa for the approximating language L_j and (2) to calculate the growth rate of a regular language from a dfa.

Theorem 1 (S., 2008)

A dfa for calculating $\alpha(L_j)$ can be built in $O(N \log N)$ time and $O(N)$ space, where $N = jk(\alpha(L))^{j/\beta}$.

Theorem 2 (S., 2008)

The growth rate of a regular language L recognized by a consistent dfa $\mathcal{A}$ can be calculated with the absolute error at most δ, $0 < \delta < 1$ in time $O(-\log \delta \cdot |\mathcal{A}|)$ using $O(|\mathcal{A}|)$ additional space.

Upper Bounds for Growth Rates: Algorithm

We have: power-free language $L = L(k, \beta)$

We need efficient algorithms (1) to build a dfa for the approximating language L_j and (2) to calculate the growth rate of a regular language from a dfa.

Theorem 1 (S., 2008)

A dfa for calculating $\alpha(L_j)$ can be built in $O(N \log N)$ time and $O(N)$ space, where $N = jk(\alpha(L))^{j/\beta}$.

Theorem 2 (S., 2008)

The growth rate of a regular language L recognized by a consistent dfa $\mathcal{A}$ can be calculated with the absolute error at most δ, $0 < \delta < 1$ in time $O(-\log \delta \cdot |\mathcal{A}|)$ using $O(|\mathcal{A}|)$ additional space.

Upper Bounds for Growth Rates: Algorithm

We have: power-free language $L = L(k, \beta)$

We need efficient algorithms (1) to build a dfa for the approximating language L_j and (2) to calculate the growth rate of a regular language from a dfa.

Theorem 1 (S., 2008)

A dfa for calculating $\alpha(L_j)$ can be built in $O(N \log N)$ time and $O(N)$ space, where $N = jk(\alpha(L))^{j/\beta}$.

Theorem 2 (S., 2008)

The growth rate of a regular language L recognized by a consistent dfa $\mathcal{A}$ can be calculated with the absolute error at most δ, $0 < \delta < 1$ in time $O(-\log \delta \cdot |\mathcal{A}|)$ using $O(|\mathcal{A}|)$ additional space.

Lower bounds

The same idea fails for lower bounds: all regular subsets of a power-free language are finite.

Alternative way: estimate the number of words in $(L_j - L)$, using the properties of the dfa built for L_j in Theorem 1.

Theorem 3 (S., 2009)

Suppose that $\beta \geq 2$ and a number γ satisfies $\gamma + \frac{1}{\gamma^{\lfloor j/\beta \rfloor - 1}(\gamma - 1)} \leq \alpha(L_j)$. Then $\gamma \leq \alpha(L)$.

Alas! The condition $\beta \geq 2$ is essential.

Practical efficiency: 5 or more sure digits for $\alpha(L)$ in several minutes on a PC with 2GB memory.

Lower bounds

The same idea fails for lower bounds: all regular subsets of a power-free language are finite.

Alternative way: estimate the number of words in $(L_j - L)$, using the properties of the dfa built for L_j in Theorem 1.

Theorem 3 (S., 2009)

Suppose that $\beta \geq 2$ and a number γ satisfies $\gamma + \frac{1}{\gamma^{\lfloor j/\beta \rfloor - 1}(\gamma - 1)} \leq \alpha(L_j)$. Then $\gamma \leq \alpha(L)$.

Alas! The condition $\beta \geq 2$ is essential.

Practical efficiency: 5 or more sure digits for $\alpha(L)$ in several minutes on a PC with 2GB memory.

Lower bounds

The same idea fails for lower bounds: all regular subsets of a power-free language are finite.

Alternative way: estimate the number of words in $(L_j - L)$, using the properties of the dfa built for L_j in Theorem 1.

Theorem 3 (S., 2009)

Suppose that $\beta \geq 2$ and a number γ satisfies $\gamma + \frac{1}{\gamma^{\lfloor j/\beta \rfloor - 1}(\gamma - 1)} \leq \alpha(L_j)$. Then $\gamma \leq \alpha(L)$.

Alas! The condition $\beta \geq 2$ is essential.

Practical efficiency: 5 or more sure digits for $\alpha(L)$ in several minutes on a PC with 2GB memory.

Lower bounds

The same idea fails for lower bounds: all regular subsets of a power-free language are finite.

Alternative way: estimate the number of words in $(L_j - L)$, using the properties of the dfa built for L_j in Theorem 1.

Theorem 3 (S., 2009)

Suppose that $\beta \geq 2$ and a number γ satisfies $\gamma + \frac{1}{\gamma^{\lfloor j/\beta \rfloor - 1}(\gamma - 1)} \leq \alpha(L_j)$. Then $\gamma \leq \alpha(L)$.

Alas! The condition $\beta \geq 2$ is essential.

Practical efficiency: 5 or more sure digits for $\alpha(L)$ in several minutes on a PC with 2GB memory.

Lower bounds

The same idea fails for lower bounds: all regular subsets of a power-free language are finite.

Alternative way: estimate the number of words in $(L_j - L)$, using the properties of the dfa built for L_j in Theorem 1.

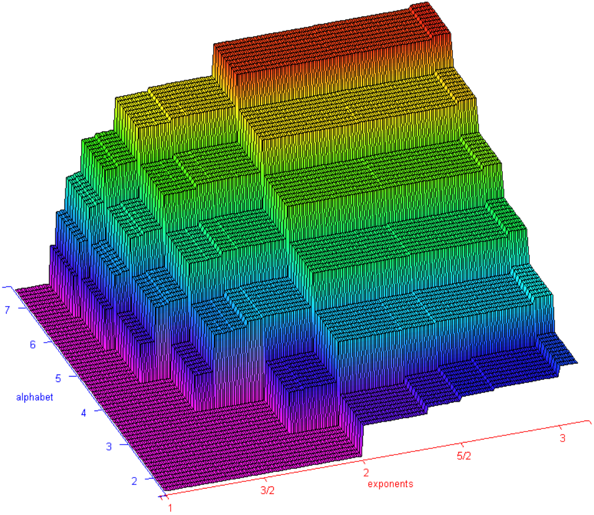
Theorem 3 (S., 2009)

Suppose that $\beta \geq 2$ and a number γ satisfies $\gamma + \frac{1}{\gamma^{\lfloor j/\beta \rfloor - 1}(\gamma - 1)} \leq \alpha(L_j)$. Then $\gamma \leq \alpha(L)$.

Alas! The condition $\beta \geq 2$ is essential.

Practical efficiency: 5 or more sure digits for $\alpha(L)$ in several minutes on a PC with 2GB memory.

Growth Rate Graph for Power-Free Languages



Asymptotic Formulas for Growth Rates

Let $\alpha(k,\beta)$ be the growth rate of $L(k,\beta)$. Combining the techniques for the upper and lower bounds we can prove the following result:

Theorem 4 (S., 2009)

For any integer $n \geq 2$ the following equalities hold:
$\alpha(k,n^+) = k - \frac{1}{k^{n-1}} + \frac{1}{k^n} - \frac{1}{k^{2n-2}} + O(\frac{1}{k^{2n-1}})$,
$\alpha(k,n{+}1) = k - \frac{1}{k^{n-1}} + \frac{1}{k^n} + O(\frac{1}{k^{2n-1}})$.
Besides this, $\alpha(k,2) = k - 1 - \frac{1}{k} - \frac{1}{k^2} + O(\frac{1}{k^3})$.

We also give a conjecture about small exponents:

Conjecture

For any integer $n \geq 0$ there exist the limits
$\alpha_n = \lim\limits_{k\to\infty} \alpha(k, \frac{k-n}{k-n-1}^+) = \lim\limits_{k\to\infty} \alpha(k, \frac{k-n-1}{k-n-2})$.
Moreover, $\alpha_0 \approx 1{,}242$, $\alpha_1 \approx 2{,}326$, $\alpha_2 \approx 3{,}376$.

Asymptotic Formulas for Growth Rates

Let $\alpha(k,\beta)$ be the growth rate of $L(k,\beta)$. Combining the techniques for the upper and lower bounds we can prove the following result:

Theorem 4 (S., 2009)

For any integer $n \geq 2$ the following equalities hold:
$\alpha(k,n^+) = k - \frac{1}{k^{n-1}} + \frac{1}{k^n} - \frac{1}{k^{2n-2}} + O(\frac{1}{k^{2n-1}})$,
$\alpha(k,n{+}1) = k - \frac{1}{k^{n-1}} + \frac{1}{k^n} + O(\frac{1}{k^{2n-1}})$.
Besides this, $\alpha(k,2) = k - 1 - \frac{1}{k} - \frac{1}{k^2} + O(\frac{1}{k^3})$.

We also give a conjecture about small exponents:

Conjecture

For any integer $n \geq 0$ there exist the limits
$\alpha_n = \lim\limits_{k\to\infty} \alpha(k, \frac{k-n}{k-n-1}^+) = \lim\limits_{k\to\infty} \alpha(k, \frac{k-n-1}{k-n-2})$.
Moreover, $\alpha_0 \approx 1{,}242, \alpha_1 \approx 2{,}326, \alpha_2 \approx 3{,}376$.

THANK YOU!