О верхних аппроксимациях в моноидах регулярных языков

С. А. Григорьев*, С. А. Афонин‡

Доказывается, что множество верхних аппроксимаций регулярного языка элементами заданного конечно порожденного моноида регулярных языков является рациональным и вычислимым подмножеством этого моноида.

Ключевые слова: Регулярные языки, аппроксимация языков, рациональная подстановка, задача представимости регулярного языка, моноиды регулярных языков.

Одной из классических задач теории формальных языков является задача представимости. Она состоит в том, чтобы по заданным множеству B регулярных языков и регулярному языку L ответить, может ли L быть выражен из B применением некоторого набора операций. В работе [1] строится множество нижних аппроксимаций L, то есть множество всех конечных конкатенаций языков из B таких, что эта конкатенация содержится в L. Настоящая работа посвящена двойственной задаче построения множества верхних аппроксимаций L, то есть конкатенаций языков из B, которые *содержат* L.

Постановка задачи Зафиксируем два конечных алфавита Σ и Δ. Обозначим за φ некоторую *рациональную подстановку* из Δ в Σ, то есть отображение $\Delta \to \mathrm{Reg}(\Sigma) \setminus \{\varnothing\}$, продолженное до гомоморфизма $\varphi : \langle \Delta^*, \cdot \rangle \to \langle \mathrm{Reg}(\Sigma), \cdot \rangle$, где $\cdot$ обозначает операцию конкатенации.

Будем обозначать пустое слово как ε. Определим:

$$\Delta_0 := \{\delta \in \Delta \mid \varepsilon \in \varphi(\delta)\}, \quad \Delta_1 := \Delta \setminus \Delta_0.$$

Длину слова v будем обозначать $|v|$. Через $\min(L)$ будем обозначать длину самого короткого слова в языке $L \neq \varnothing$. Когда слово v является (разреженным) подсловом w, будем писать $v \sqsubseteq w$.

*Григорьев Степан Александрович — Аспирант МГУ им. Ломоносова, e-mail: stpngrgr@gmail.com, ORCID: 0009-0002-1607-9166.

Grigorev Stepan Aleksandrovich — PhD student, Moscow State University.

‡*Афонин Сергей Александрович* — Ведущий научный сотрудник лаборатории автоматизации экспериментальных исследований НИИ механики МГУ им. Ломоносова, e-mail: serg@msu.ru, ORCID: 0000-0003-3058-9269.

Afonin Sergey Aleksandrovich — leading researcher, Moscow State University, Institute of Mechanics, Experimental research automation Laboratory.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

Образ $\varphi(\Delta^*)$ является конечно порожденным подмоноидом моноида регулярных языков $\langle \mathrm{Reg}(\Sigma), \cdot \rangle$. Элемент E этого подмоноида будем называть *верхней аппроксимацией* $L \in \mathrm{Reg}(\Sigma)$, если $L \subseteq E$. Обозначим через $\overline{M}_\varphi(L)$ множество прообразов всех верхних аппроксимаций L в $\varphi(\Delta^*)$. Формально:

$$\overline{M}_\varphi(L) := \{u \in \Delta^* \mid \varphi(u) \supseteq L\}.$$

Отметим, что $\overline{M}_\varphi(L)$ отличается от объекта, строящегося в [1], направлением знака $\subseteq$.

Предметом данной работы является изучение свойств $\overline{M}_\varphi(L)$. Основным ее результатом является следующая теорема:

Теорема 1. *Для любого регулярного языка L и рациональной подстановки φ язык $\overline{M}_\varphi(L)$ регулярен и существует алгоритм его построения.*

Регулярность Пусть $w := w_1 \dots w_n$, $w_i \in \Delta$. Тогда определим Δ_0–*расширение* w как

$$\mathrm{Ext}(w, \Delta_0) := \Delta_0^* w_1 \Delta_0^* w_2 \Delta_0^* w_3 \dots w_n \Delta_0^*.$$

Поскольку при «накачке» слова $w \in \Delta^*$ буквами из Δ_0, образ относительно φ может только расшириться, можно заключить, что всякое w входит в $\overline{M}_\varphi(L)$ вместе со своим расширением. Отсюда

$$\overline{M}_\varphi(L) = \bigcup_{w \in \overline{M}_\varphi(L)} \mathrm{Ext}(w, \Delta_0).$$

Определим $\sqsubseteq_0$ так: $v \sqsubseteq_0 w \Leftrightarrow w \in \mathrm{Ext}(v, \Delta_0)$. Очевидно $v \sqsubseteq_0 w \Rightarrow \mathrm{Ext}(v, \Delta_0) \supseteq \mathrm{Ext}(w, \Delta_0)$. Пусть $v \in \Delta_1^*$, а $C(v, L)$ – множество минимальных относительно $\sqsubseteq_0$ элементов $V(v) := \mathrm{Ext}(v, \Delta_0) \cap \overline{M}_\varphi(L)$. Поскольку на $V(v)$ отношения $\sqsubseteq$ и $\sqsubseteq_0$ совпадают, из леммы Хигмана [3] следует конечность $C(v, L)$. Заметим, что множество

$$S(L) := \{v \in \Delta_1^* \mid \exists w \in \overline{M}_\varphi(L) : v \sqsubseteq_0 w\}$$

также будет конечным, поскольку для любого $w \in \overline{M}_\varphi(L)$ должно быть верно: $\min(\varphi(w)) \leqslant \min(L)$, а при $v \in \Delta_1^*$ верно, что $\min(\varphi(v)) \geqslant |v|$. Обозначим за $C(L) := \bigcup_{v \in S(L)} C(v, L)$. Тогда из определений $C(v, L)$ и $S(L)$

$$\bigcup_{w \in C(L)} \mathrm{Ext}(w, \Delta_0) \subseteq \overline{M}_\varphi(L) \subseteq \bigcup_{w \in C(L)} \mathrm{Ext}(w, \Delta_0),$$

а сам $\bigcup_{w \in C(L)} \mathrm{Ext}(w, \Delta_0)$ регулярен как конечное объединение регулярных языков.

Алгоритм построения Выше показано, что $\forall v \in S(L)\ |v| \leqslant \min(L)$. Обозначим $\overline{S}(L) := \{v \in \Delta_1^* \mid |v| \leqslant \min(L)\}$. Чтобы построить $\overline{M}_\varphi(L)$ необходимо уметь находить все минимальные относительно $\sqsubseteq_0$ элементы $\bigcup\limits_{v \in \overline{S}(L)} \mathrm{Ext}(v, \Delta_0)$, образ которых содержит L.

Автоматом с расстояниями называется набор $\mathfrak{A} := \langle Q, \Sigma, I, F, \tau, d\rangle$, где $\langle Q, \Sigma, I, F, \tau\rangle$ задают ε-недетерминированный автомат, а функцию $d : Q \times \Sigma \times Q \to \{0, 1, \infty\}$ называют функцией расстояния. Эта функция присваивает вес каждому переходу автомата. Вес пути π в автомате $\mathfrak{A}$ определяется как сумма весов всех переходов, составляющих π. Вес слова $d(w), w \in \Sigma^*$ определяется как минимум из весов допускающих путей для слова w. Автомат $\mathfrak{A}$ *ограничен* если $\exists N \in \mathbb{N} : \forall w\ d(w) \leqslant N$. В [2] доказано, что задача проверки ограниченности $\mathfrak{A}$ алгоритмически разрешима.

Мы можем показать, что для заданного регулярного Δ–языка $X \in \mathrm{Reg}(\Delta)$ можно построить автомат с расстояниями $\mathfrak{A}(X)$ над Σ, который принимает язык $\bigcup\limits_{w \in X} \varphi(w)$, а функция расстояния задана таким образом, что вес слова u равен длине самого короткого слова из X, образ которого содержит u:

$$d(u) = \min\{|v| \mid (v \in X) \wedge (u \in \varphi(v))\}.$$

Для этого в автомате, распознающем X, заменим каждый переход по букве δ на автомат, распознающий $\varphi(\delta)$, а функцию расстояния положим равной 1 только на переходах «входящих в новую копию $\varphi(\delta)$».

Тогда при $X = \mathrm{Ext}(w, \Delta_0)$ для некоторого $w \in \Delta^*$, можем рассмотреть пересечение автомата $\mathfrak{A}(X)$ с автоматом, распознающим L. Если язык, распознаваемый полученным автоматом, не совпадает с L, или этот автомат не ограничен, то для любого слова u из $\mathrm{Ext}(w, \Delta_0)$ верно $L \not\subseteq \varphi(u)$. Обратно: если полученный автомат распознает L и ограничен, то $\exists u \in \mathrm{Ext}(w, \Delta_0) : L \subseteq \varphi(u)$.

Пусть $v \in \Delta_1^*$. Построим множество $C(v, L)$ минимальных относительно $\sqsubseteq_0$ слов из $\mathrm{Ext}(v, \Delta_0) \cap \overline{M}_\varphi(L)$. Будем последовательно перебирать $w \in \mathrm{Ext}(v, \Delta_0) \setminus E$, где E изначально равно $\varnothing$, в порядке роста длин и проверять описанное выше условие. Если $\forall u \in \mathrm{Ext}(w, \Delta_0) \setminus E\ \ L \not\subseteq \varphi(u)$, то весь $\mathrm{Ext}(w, \Delta_0)$ можно удалить из дальнейшего рассмотрения, то есть добавить его к E. А если $L \subseteq \varphi(w)$, то $w \in C(v, L)$. Тогда такой w можно добавить в результат, а множество $\mathrm{Ext}(w, \Delta_0)$ также удалить из рассмотрения.

Поскольку таких минимальных слов конечное число, алгоритм остановится. Тогда применив этот алгоритм ко всем словам v из Δ_1^* длины не больше $\min(L)$ и объединив полученные результаты, мы построим

искомое множество S:

$$\overline{M}_\varphi(L) = \bigcup_{w \in S} \mathrm{Ext}(w, \Delta_0).$$

Список литературы

[1] Calvanese, Diego et al., "Rewriting of regular expressions and regular path queries", *Proc. 18th PODS symposium*, 1999. DOI: 10.1145/303976.303996, 194–204.

[2] Hashiguchi, Kosaburo, "Limitedness theorem on finite automata with distance functions", *Journal of computer and system sciences*, **24**:2 (1982), 233–244.

[3] Higman, Graham, "Ordering by divisibility in abstract algebras", *Proceedings of the London Mathematical Society*, **3**, 3rd ser.:1 (1952), 326–336.

Статья поступила 9 октября 2026 г.

On upper approximations in monoids of regular languages

S. A. Grigorev, S. A. Afonin

We prove that the set of upper approximations of a regular language by elements of a given finitely generated monoid of regular languages is a rational effectively constructible set.

Keywords: Regular languages, approximation of languages, rational substitution, representability problem for regular languages, monoids of regular languages.

References

[1] Calvanese, Diego et al., "Rewriting of regular expressions and regular path queries", *Proc. 18th PODS symposium*, 1999. DOI: 10.1145/303976.303996, 194–204.

[2] Hashiguchi, Kosaburo, "Limitedness theorem on finite automata with distance functions", *Journal of computer and system sciences*, **24**:2 (1982), 233–244.

[3] Higman, Graham, "Ordering by divisibility in abstract algebras", *Proceedings of the London Mathematical Society*, **3**, 3rd ser.:1 (1952), 326–336.

Received on October 9, 2026