Exercises in Termination

Dieter Hofbauer and Johannes Waldmann

manuscript, Intl. School on Rewriting 2008

1 Derivational Complexity

1.1 Three rewriting systems that look alike

Find lower bounds for the derivational complexity of:

- $R_1 = \{ba \to acb, bc \to abb\}$

  Answer: looping, this is SRS/HofWald/1. To find the loop: note that for $w \in \{a, b, c\}^* : bw \to^* \phi(w)b$ where $\phi : a \mapsto ac, b \mapsto b, c \mapsto ab$, and if $w \notin b^*$, then we have indeed at least one step. By iteration, $\forall k \geq 0 : b^k w \to^* \phi^k(w)b^k$. Then verify that $\phi^5(a)$ contains $b^5 a$ as a scattered subword. Then conclude that $b^5 a$ starts a loop. This is an example of a "D0L loop" (Waldmann, 2007).

- $R_2 = \{ba \to acb, bc \to cbb\}$

  Answer: at least tower of exponentials, this is SRS/Zantema/z018, see Hofbauer/Waldmann RTA06

- $R_3 = \{ba \to aab, bc \to cbb\}$

  Answer: doubly exponential, proof: matrix interpretation, two times

Hint: one is doubly exponential, one is multiply exponential, one is non-terminating.

A lower bound is proved by presenting a family of derivations that achieves the desired length.

1.2 How To Count

(H. Zantema) Give an example of a length-preserving string rewriting system that has exponential derivational complexity, by simulating a binary counter. Use letters 0,1 and a "carry" symbol.

(Hint: there is a solution with total size (sum of lenghts of all lhs and rhs) 12. Can you do better?)

Answer: $R = \{0 \to 1, 1 \to c, 0c \to 10, 1c \to c0\}$.

- show exactly a family of derivations of exponential length

  Answer: For each k, there is a derivation $0^k \to^{\geq 2^k - 1} 1^k$ Induction step: $0^{k+1} = 00^k \to^* 01^k \to^* 01^{k-1}c \to^* 0c0^{k-1} \to 10^k \to^* 11^k$.

- give a termination proof

  Answer: There is a linear interpretation (for the reversed system) that removes all rules at the same time. Hint: all interpretation functions are of shape $x \mapsto 3x+?$.

1.3 Such A Loong Looop

Find a looping derivation for $R_k = \{10^k \to 0^k 1^k 0\}$.

Answer: starts from 10^{k^2}, see Geser: Loops of Superexponential Length, RTA-02; Existence (but not minimality) of the loop can be verified quickly using the D0L approach (Waldmann).

1.4 All (Natural) Degrees

(D. Hofbauer) We choose a number $d \in \mathbb{N}$ with $d \geq 2$ and consider the system $R_d = \{xy \to zx \mid z < x\}$ over $\Sigma_d = \{1, \dots, d\}$.

Show that the derivational complexity of R_d is $\Theta(n^d)$.

Of course, this requires two proofs:

- there is a derivation of the given length

  Answer: There is derivation of length $\Theta(n^d)$ from $d?^n$. first step: $d?^{n+1} \to (d-1)d?^n$, first subderivation: $\to^* (d-1)?^{n+1}$, second subderivation: $\to ?^{n+1}$,

- there is no longer derivation

  Answer: assign weight n^d (roughly) to letter d at position n, counting from the right end.

Hint (for both parts): the x symbol (the largest in each rule) moves from left to right.

Extra question: find the smallest subset of R_d that still has the same properties.

1.5 In Between Days

Find a string or term rewriting system whose derivational complexity is polynomially bounded, but not big-Theta of any polynomial. E.g. construct a system of complexity $n \mapsto n \cdot \log n$.

> Answer: (J. Endrullis) Idea: a Turing machine moves over a string of sa^*f. Going right, it replaces every second a by b. Going left, it does nothing.
> Detail: the intended tape contents is $s\{A,B\}^*\{e,o,l\}\{a,b\}^*f$ and rules are $ea \to Bo, eb \to Be, oa \to Ae, ob \to Bo, of \to lf, Al \to la, Bl \to lb, sl \to se$ where e is even and o is odd (counting a when going right).
> Complexity is indeed at most quadratic: There is a two-dim upper triangular matrix interpretation.
> The required $n \log n$ derivation really exists: starting from $sea^{2^k}f$, we reach $seb^{2^k}f$ in $\Theta(k \times 2^k)$ steps.
> But is the complexity really less than quadratic? We have to watch out for tape contents that don't fit the intended pattern.
> By brute force enumeration of overlap closures, I don't find any closure $uv \to^+ vu$. Such a closure would imply quadratic derivations.

1.6 Some Unnatural Degrees

(H. Zantema) Find a rewriting system with derivational complexity $\Theta(n^d)$ where d is not an integer.

Hint: for instance, $d = \log_2(3)$.

Can you do with one rule only?

Answer: A Turing machine with intended tape contents $sb^*\{l,r\}a^*f$ and derivations $sla^{2n}f \to^* sla^{3n}f$.
This system should achieve $\Theta(n^{\log_q p})$ for co-prime p, q. It is similar to SRS/Zantema/z079. $R = \{sla \to sra, ra \to br, rf \to lf, b^q l \to la^p\}$
Zantema adds: the one-rule system $a^q b \to ba^p$ should have similar behaviour.

1.7 Quadratic used to be Easy

Give an example of a string rewriting system with quadratic derivational complexity where all rules are length-increasing.

(So the obvious $ab \to ba$ is ruled out.)

2 Match Bounds

2.1 Right. Forward!

Verify that for $k \geq 2$, the system $R_k = \{baba^k \to a^{k+1}babab\}$ is RFC-matchbounded by 2. Construct the automaton, it is not too hard.

Note: this problem is treated (without match-bounds) in Section 6.6 of Geser's Habil thesis.

2.2 How High Can You Get

For any $n \in \mathbb{N}$, construct a SRS R_n that is exactly match-bounded by n.

- (easy) number of rules and size of alphabet may depend on n

  Answer: alphabet is $\{a_0, \dots, a_n\}$ and rules are $\{a_0 \to a_1, \dots, a_{n-1} \to a_n$.

- alphabet is fixed (but number of rules may depend on n)

  Answer: same idea as before, but encode the letters over some fixed (e.g. binary) alphabet. Be careful to avoid overlaps. One encoding that works is $a_k \mapsto 0^k 1^{n-k}$.

- (hard) both alphabet and number of rules is fixed (so, length of rules depends on n)

Answer: J. Endrullis (diploma thesis) proved that for $w = ba^k b$, the system $aw \to wwa$ is RFC-matchbounded by k, by explicitely giving the automaton. It is in fact compatible with $aw \to w^+ a$.

2.3 Match Left and Match Right

(research problem)

We define a method to annotate positions in strings with heights: in a rule application $l \to r$ with $|l| > 1$ and $|r| > 1$, choose any nontrivial representation $l = l_1 l_2, r = r_1 r_2$, and then annotate with match heights for $l_1\ tor_1$ and $l_2 \to r_2$ separately (the r_1 part gets $1 + \min l_1$, the r_2 part gets $1 + \min l_2$). Any derivation annotated in that manner is called a split-match derivation. Prove that a system is split-match-bounded iff it is match-bounded.

2.4 Near Ground Level (Plain)

Determine the class of one-rule string rewriting systems that are match-bounded by one. (Discuss what overlaps are allowed between lhs and rhs.)

2.5 Near Ground Level (RFC)

Determine the class of one-rule string rewriting systems that are RFC-matchbounded by zero (i.e. they have no R-redex in the RFC closure automaton, assuming that the "extension rules" produce match height zero).

Here is an example: $R = \{abaa \to a^3bbab\}$. We compute $\text{RFC}(R) = (a^3b^2)^*ab$ and this contains no R-redex.

Answer: In the notation of McNaughton, see Geser (Habil, Theorem 6.6), such a system R is called *left-barren.*

2.6 The size without the bound

Is the following questions decidable: does a finite string rewriting system R admit a match-bound certificate

- with given size (number of states) s and given bound b?

Answer: trivially decidable. Just enumerate all certificate candidates

- with given bound (but any size)?

Answer: nontrivially dedicable. Construct a certificate by computing the exact closure under bounded (by b) rewriting of Σ^*, using Endrullis/Hofbauer/Waldmann papers, then check whether there is any uncovered redex (that would produce a reduct of height $b+1$).

- with given size (but any bound)?

Answer: Waldmann, unpublished: given the size s, the largest bound that must be checked is obtained if all edges have different labels. Note the label set can always be taken as a contiguous prefix of $0, 1, 2, \ldots$, since we only apply min and max and this ignores any gaps. So this idea does not work for arctics or naturals.

- with no further information? Remark: if you can choose the starting language, then this is undecidable (Middeldorp), but here the question is for Σ^*.

3 Matrix Method(s)

(we restrict to the "standard" matrix method (shape E) where top left and bottom right entry of matrices are positiv, and top right entry of difference is positive.)

3.1 Small Matrices

Give the smallest matrix proof dimension for these systems:

- $R_1 = \{aa \to bbb\}$

Answer: dimesion one is enough: $a = 3, b = 2$, then $a^2 = 9 > b^3 = 8$. Note that after taking logarithms, this is just an additive weight assignment.

- $R_2 = \{ab \to ba\}$

Answer: dimesion one cannot work since the system has no compatible additive weight function. dimension two is OK with $a : x \mapsto 2x, b : x \mapsto x + 1$. note that each dimension two interpretation has entries ≥ 1 on the main diagonal, thus for any $x \in \Sigma, w \in \Sigma^* : [xw] \geq [w]$. Therefore, interpretations of dimension two really prove simple termination.

- $R_3 = \{aa \to aba\}$.

Answer: since this is not simply terminating, dimension two is not enough. it works with 3. $a : (x, y) \mapsto (x + y, 1), b : (x, y) \mapsto (x, 0)$. This interpreation actually counts the number of (overlapping) blocks a^2, which is exactly the number of redexes, and which decreases by exactly one in each step. Verify this numerically.

Hint: one, two, three.

The easy part is (probably) to find the interpretation, the harder part is to prove that no smaller interpretation exists. You need to prove the theorem that matrix interpretations of dimensions ≤ 2 in fact show *simple* termination.

3.2 All Degrees

(see exercise in complexity)

Show that assigning weight $\binom{p}{k}$ to letter k standing at position p (counting from the right) gives a compatible interpretation (the sum of weights is decreasing).

Show that this weight can be computed by a matrix interpretation. Hint: the matrices are upper triangular.

3.3 A Path Is Enough

Characterize those one-rule SRS that have a matrix interpretation that just consists of the path corresponding to the left-hand side, all edges labelled by one. (And of course the Σ-loops at start and end.)

Answer: We have to check that for each reduct from start s to any state p, there is also a redex from s to p. Note that such a path will cycle in s and then go straight towards p. Then the condition is: each suffix of a reduct that is a prefix of the redex, must also be a suffix of the redex. A symmetrical condition holds for paths to the final state. In all, we get $\mathrm{OVL}(l, r) \cup \mathrm{OVL}(r, l) \subseteq \mathrm{OVL}(l, l)$, which is Kurth's Kriterion E = Theorem 2.8 in Geser's Habil thesis. The proof is still identical: we get termination because the number of l factors decreases. This is exactly what the matrix interpretation counts.

3.4 Just One Bit Of Information

(research problem) Characterize (some of) those systems that admit a matrix interpretation where all entries in all matrices (interpretations of symbols and rules) are in $\{0, 1\}$. (This is what would happen if you restrict the bit width in the SAT translation to one.)

3.5 Some Calculus

(Using some computer algebra system,) take any matrix interpretation (e.g. the 5-dimensional for z001), pick one position and replace the entry there by a variable. Then discuss (plot) the difference(s) between lhs and rhs interpretations as functions of that variable.

4 Automated Termination

4.1 That's a Classic

SRS/Zantema/z001 ... z128 is a classical problem set in termination of string rewriting. E.g. z001 is Zantema's Problem $\{a^2b^2 \to b^3a^3\}$, Without looking into your computer or the internet, answer the following:

- how many problems are these? (Hint: some are missing from the enumeration.

  Answer: 16 and 99 are missing

  Bonus points: when (what year) were they removed and why?)

- what number is Zantema's Other Problem $\{a^2 \to bc, b^2 \to ac, c^2 \to ab\}$? When was it first solved? By what program, using what method?

  Answer: z086, 2006, jambox, matrix interpretation

  (Bonus points: why is this proof not visible in the results table?)

  Answer: proof was found in first round (60 seconds) but overwritten by MAYBE from second round (120 seconds).

- what problems are looping?

  Answer: 42,44,46,96,127,128

- what problem is non-looping non-terminating?

  Answer: 73

- what problem is Ackermannian? was it solved?

  Answer: 90, yes: TTT via root labelling

4.2 Common Knowledge

What is the common property of SRS/Gebhardt/* (besides being hard)?

Answer: these are two-rule length-preserving SRS over a binary alphabet, each lhs and rhs has length four.