Expressivity of Quadratic and Subquadratic Word Equations

Matthew Konefal

Loughborough University, UK

19th June 2026

Word Equations

- Made of *variables* $\Xi = \{X, Y, \ldots\}$ and *terminals* $\Sigma = \{\mathsf{a}, \mathsf{b}, \ldots\}$.

$$X\mathsf{ba}Z\mathsf{b}Y \doteq YZ\mathsf{aa}W.$$

- *Solutions*: assignments to the variables making the sides identical.

$$\mathsf{aababbaabab} \doteq \mathsf{aababbaabab}.$$

  - This is an example of a “conjugacy equation”.

Word Equations: Language Expressers

- Each variable defines a language: the set of values it may be assigned.

  For example,

  - $X \doteq YY\mathsf{aba}$

    ...the set of words obtained by appending "aba" to a square
  - $X\mathsf{ab}Y \doteq Y\mathsf{ba}X$

    ...the set $\mathcal{S}$ of Standard Sturmian words, (which are all *balanced*).

Word Equations: Language Expressers

- The resulting class of languages $\mathcal{L}(\mathrm{WE})$ is poorly understood. Even its interesection with REG is not known.
- Tools for showing $L \notin \mathcal{L}(\mathrm{WE})$ based on the length-set of L, its factor complexity, codes, “anchored and unanchored factors”,...
- Unpleasant properties. For membership, one (naïvely) has to determine the solubility of a WE. And...

Theorem (Day et al., 2023)

The problem of determining whether $L \in \mathcal{L}(WE)$ is regular is undecidable.

Closure Properties of $\mathcal{L}(\mathbf{WE})$

- Languages expressed by (finite) Boolean combinations of word equations are expressible by a single word equation.

$$\begin{array}{c} X \doteq aY \\ Y \doteq Zb \end{array} \quad \Longrightarrow \quad X \doteq aZb$$

- We get closure under concatenation, reversal, *etc.*, but not under complementation, (injective) morphic image, Kleene star or shuffle.

Restricting the Form

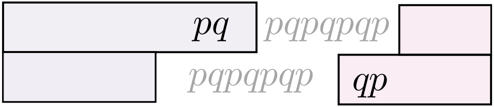
- Practical examples of word equations often have restricted form.
- How does this impact the expressivity and closure properties?

Strictly Regular-Ordered Equations

- Here, both sides of the equation have variables $Y_1, ..., Y_n$ in that order.

$$\mathsf{a}Y_1\mathsf{bba}XY_2\mathsf{ca} \doteq Y_1\mathsf{abb}X\mathsf{caa}Y_2$$

- Solution: first determine a "left-overlap" and a "right-overlap" for X. X carries one overlap to the other, so its form is determined...



We can describe $\mathcal{L}(\mathrm{SROWE})$ completely...

A Description of $\mathcal{L}(\mathbf{SROWE})$

Theorem

$\mathcal{L}(SROWE)$ contains exactly the unions $\bigcup_{i \in \{1,\ldots,n\}} (p_i q_i)^* p_i$ such that

1. *the $p_i q_i$ are all primitive,*
2. *the Parikh vectors of the $p_i q_i$ are all collinear,*
3. *"conjugate-fullness" is satisfied: Where $p_i q_i = tu$, $q_j p_j = ut$, the union must include $(p_i q_i)^* t$.*

e.g., (1) rules out $(\mathsf{aa})^*$, (2) rules out $\mathsf{a}^*|\mathsf{b}^*$, and (3) rules out $(\mathsf{ab})^*|(\mathsf{ba})^*$.

A Description of $\mathcal{L}(\mathrm{SROW}$

Since the left-overlaps sets are Parikh-consistent: each is generated from the previous by $S \mapsto u^{-1}\,\mathrm{CyclicClosure}(S)v$.

Theorem

$\mathcal{L}(SROWE)$ contains exactly the u ... such that

1. *the $p_i q_i$ are all primitive,*
2. *the Parikh vectors of the $p_i q_i$ are all collinear,*
3. *"conjugate-fullness" is satisfied: Where $p_i q_i = tu$, $q_j p_j = ut$, the union must inclu*

Since the left- and right-overlaps are independent of each other.

e.g., (1) rules out $(\mathrm{aa})^*$, (2 ... rules out $(\mathrm{ab})^*|(\mathrm{ba})^*$.

A Hierarchy by Number of Auxiliary Variables

- *Idea:* The more variables that can appear to the left of X, the greater the diameter the set of left-overlaps can be under a certain metric δ. *(The taxicab distance of double Parikh-vectors)*
- Indeed, the map $S \mapsto u^{-1}\,\mathrm{CyclicClosure}(S)v$ increases the δ-diameter by at most 8. Hence

$$\prod_{i=1}^{n}(Y_i\mathsf{a})\cdot X\prod_{i=1}^{n-1}(\mathsf{g}Z_i)\cdot\mathsf{f}Z_n(\mathsf{deabca})^n \doteq (\mathsf{abcade})^n\cdot Y_1\mathsf{f}\prod_{i=2}^{n}(Y_i\mathsf{g})\cdot X\cdot\prod_{i=1}^{n}(\mathsf{a}Z_i),$$

which has $2n$ auxiliary variables, requires these to express the language of its X.

Regular-Ordered Equations

- Now we allow *single-occurrence* variables too:

$$X\mathsf{aa}YW \doteq XZ\mathsf{b}Y$$

Proposition

$\mathcal{L}(ROWE)$ contains only finite unions of languages:

- $(pq)^*p$ *and*
- $w_1\Sigma^*w_2\Sigma^*w_3\cdots\Sigma^*w_n$.

... but this is an over-approximation.

Finite Languages

- Where ρ is long and unbordered, $\{w_1, ..., w_n\}$ is expressed by X in the conjugacy ROWE

$$w_1 \rho w_2 \rho \cdots w_n \rho Y \doteq Y X \rho Z_1 \rho Z_2 \cdots \rho Z_{n-1}.$$

Open Problem (Nowotka & Saarela, 2018)

Which finite languages can be expressed using one-variable word equations?

A Partial Description of $\mathcal{L}(\mathrm{ROWE})$

Proposition

Let L **include no pattern language***. Then $L \in \mathcal{L}(ROWE)$ iff L is finite or a finite union $\bigcup_i (p_i q_i)^* p_i$ satisfying conjugate-fullness.*

⇒ Parikh-collinearity no longer necessary. Languages such as a*|b* are doable using a "two-sided" version of the previous argument.

- Some unions of pattern languages can be made (e.g. Σ^+). Others, *e.g.*, aΣ^*|bΣ^* seem tricky.

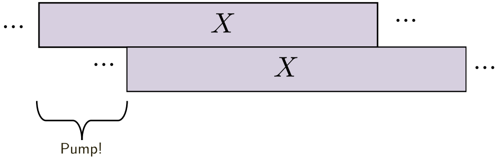
Regular Word Equations

- Now we allow *permutations* of the above variable-arrangements:

$$XZY \doteq YW\mathrm{b}X$$

Lemma

In any infinite $L \in \mathcal{L}(RWE)$, the maximal number of runs in words of L is either 1 or ∞.



Non-Membership in $\mathcal{L}(\mathrm{RWE})$

Lemma

In any infinite $L \in \mathcal{L}(RWE)$, the maximal number of "runs" in words of L is either 1 or ∞.

- This shows, *e.g.*, $\mathsf{a}^*\mathsf{b}^* \notin \mathcal{L}(\mathrm{RWE})$, and allows some non-closure properties of $\mathcal{L}(\mathrm{RWE})$ to be established.

- Seems like conjugate-fullness still holds, which is odd. For instance, we think $(\mathsf{ab})^*|(\mathsf{ba})^* \notin \mathcal{L}(\mathrm{RWE})$.

Quadratic Equations

- Now we also allow variables that occur *exactly twice, both on same side*:

$$Y\mathsf{b}X\mathsf{a}X \doteq Z\mathsf{bc}Z\mathsf{a}Y\mathsf{bb}$$

- A well-studied class with many applications.

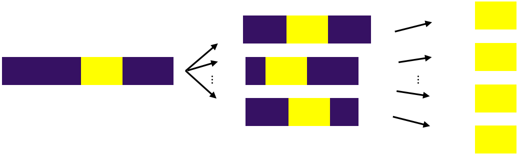
Owen's Lemma

A powerful tool for showing non-membership in $\mathcal{L}(\mathrm{QWE})$:

Lemma

Let $L \in \mathcal{L}(QWE)$. Then there are finitely many QWEs $\alpha \doteq \beta$ such that L is the union of the corresponding "solution-word languages" $\{h(\alpha) : h \text{ is a solution of } \alpha \doteq \beta\}$

- *Proof Idea*: We can chop out the relevant part of the QWE using Nielsen transformations:



Owen's Lemma

Proposition

Let $n > 2$. *Then* $(\mathsf{a}^n)^* \notin \mathcal{L}(QWE)$.

- *Proof*: Otherwise, by Owen's Lemma there'd exist a QWE E having as its set of solution-words infinitely many a^{mn} $(m \in \mathbb{N})$.

aaaaaaaaaaaaaaaaaaaaa
aaaaaaaaaaaaaaaaaaaaa

We're free to pump "a"s into any variable, (as long as we keep the sides the same length). No variable appears more than twice, so we can break the divisibility by n.

Owen's Lemma

Proposition

Let $n > 2$*. Then* $(\mathsf{a}^n)^* \notin \mathcal{L}(QWE)$*.*

- The missing case $n = 2$ is open! Owen's Lemma doesn't apply to it.

(All this holds with a replaced by any $w \in \Sigma^+$).

Regular Languages in $\mathcal{L}(\mathbf{QWE})$

- *Initial hypothesis:* Finite unions of languages $w_1\mathcal{R}_1^* w_2 \cdots \mathcal{R}_n^*$ where R_i is Σ or a primitive word.
- I can make a cofinite subset of such L using a "quadratic pair of equations". One selects the element of the union; the other determines the structure of words in L.
- *Tricky example:* aΣ^*|bΣ^*.

Regular Languages in $\mathcal{L}(\mathbf{QWE})$

Lemma

Finite unions of languages $w_1 u_1^* w_2 \cdots u_n^*$*, where the* u_i *are primitive, are in* $\mathcal{L}(QWE)$

- *Proof:* Assert that $(X\rho Y\rho \cdots Z\rho)^2$ is conjugate to $(\ell_1\rho \cdots \ell_n\rho) \cdot (r_1\rho \cdots r_n\rho)$, where

$$\ell_i = w_{i,1} u_{i,1} T_{i,1} u_{i,2} T_{i,2} \cdots u_{i,n} T_{i,n}$$

$$r_i = w_{i,1} T_{i,1} u_{i,1} T_{i,2} u_{i,2} \cdots T_{i,n} u_{i,n}.$$

- Once we have pattern languages in the union, we can't find a suitable ρ.

Closure Properties of $\mathcal{L}(\mathrm{QWE})$

Proposition

$\mathcal{L}(QWE)$ is not closed under intersection.

- *Counterexample:* the intersection $\mathcal{S} \cap \mathsf{a}^*\mathsf{b}\mathsf{a}^*\mathsf{b}\mathsf{a}^*$ is $L := \{\mathsf{a}^n\mathsf{b}\mathsf{a}^m\mathsf{b}\mathsf{a}^n : n \leq m \leq n+1\}$.
- *Proof:* Use Owen's Lemma. Suppose some QWE E had, as its solution-word language, an infinite subset of L. Most placements of variables in E would let us pump the blocks to break the synchronisation. Brute force through the remaining cases!
- *Open Problem:* Is $\mathcal{L}(\mathrm{QWE})$ closed under union?

A Hierarchy by Number of Auxiliary Variables

- Karhumäki et al. showed that

$$\mathcal{L}(\mathrm{WE})|_{k \text{ aux. variables}} \supset \mathcal{L}(\mathrm{WE})|_{k-1 \text{ aux. variables}}$$

- Their witness also separates the equivalent hierarchy for QWEs. It is the language is the language of Z in the QWE

$$ZZ \doteq \mathsf{a}_0 Y_1 \mathsf{a}_1 Y_2 \mathsf{a}_2 \cdots Y_k \mathsf{a}_k \mathsf{a}_0 \mathsf{a}_1 Y_1 \mathsf{a}_2 Y_2 \cdots \mathsf{a}_k Y_k,$$

which, it turns out, requires k auxiliary variables.

Further Work

- Look at expressivity of *relations* by (sub)quadratic equations,
- Some missing and difficult-looking closure properties,
- Tricky examples to resolve,
- More hierarchies to separate,
- Complexity matters,

⋮

Thank You!

Joel D. Day, Vijay Ganesh, Nathan Grewal, Matthew Konefal, and Florin Manea.
A closer look at the expressive power of logics based on word equations.
Theory Comput. Syst., 68(3):322–379, 2024.

Joel D. Day and Matthew Konefal.
Word equations with length constraints via weak arithmetics and matrix reachability problems.
In Pierre Ganty and Alessio Mansutti, editors, Reachability Problems, pages 51–67, Cham, 2026. Springer Nature Switzerland.

Juhani Karhumäki, Filippo Mignosi, and Wojciech Plandowski.
The expressibility of languages and relations by word equations.
J. ACM, 47(3):483–505, 2000.

Juhani Karhumäki, Filippo Mignosi, and Wojciech Plandowski.
On the expressibility of languages by word equations with a bounded number of variables.
Bulletin of the Belgian Mathematical Society-Simon Stevin, 8(2):293–305, 2001.

Dirk Nowotka and Aleksi Saarela.
An optimal bound on the solution sets of one-variable word equations and its consequences.
In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic, volume 107 of LIPIcs, pages 136:1–136:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.