

TRAJECTORY-BASED OPERATIONS

by

MICHAEL DOMARATZKI

A thesis submitted to the
School of Computing
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada

August 2004

Copyright © Michael Domaratzki, 2004

Abstract

Shuffle on trajectories was introduced by Mateescu *et al.* [147] as a method of generalizing several studied operations on words, such as the shuffle, concatenation and insertion operations. This natural construction has received significant and varied attention in the literature. In this thesis, we consider several unexamined areas related to shuffle on trajectories.

We first examine the state complexity of the shuffle on trajectories. We find that the density of the set of trajectories is an appropriate measure of the complexity of the associated operation, since low density sets of trajectories yield less complex operations.

We introduce the operation of deletion along trajectories, which serves as an inverse to shuffle on trajectories. The operation is also of independent interest, and we examine its closure properties. The study of deletion along trajectories also leads to the study of language equations and systems of language equations with shuffle on trajectories.

The notion of shuffle on trajectories also has applications to the theory of codes. Each shuffle on trajectories operation defines a class of languages. Several of these language classes are important in the theory of codes, including the prefix-, suffix-, biprefix-codes and the hypercodes. We investigate these classes of languages, decidability questions, and related binary relations.

We conclude with results relating to iteration of shuffle and deletion on trajectories. We characterize the smallest language closed under shuffle on trajectories or deletion along trajectories, as well as generalize the notion of primitive words and primitive roots. Further examination of language equations are also possible with the iterated counterparts of shuffle and deletion along trajectories.

Acknowledgments

I am grateful for everything my supervisor Dr. Salomaa has done for me during the course of our time together. His support has been outstanding and he has dedicated many hours to helping me improve this work, and to our collaborations.

I am grateful to the anonymous referees who have made suggestions for the journal and conference versions of the results presented here.

I would also like to thank the members of my examining committee, Dr. Ilie, Dr. Kashyap, Dr. Skillicorn and Dr. Tennent, for their comments and suggestions which have improved this thesis.

I am in debt to Alexander Okhotin for discussions, collaboration, his answering of my questions, making elegant figures, and not killing me, though I'm sure he would have preferred it.

The following people have also helped me through discussions on the topics in this thesis: Mark Daley, Masami Ito, Alexandru Mateescu, Jeff Shallit, and Petr Sosík.

kristy, amelia and jasper, for everything.

A proof is a proof. What kind of a proof? It's a proof.

A proof is proof, and when you have a good proof,

it's because it's proven.

–Jean Chretien (Sept. 5, 2002).

Co-Authorship

The work in Chapter 4 is joint work with my supervisor, Dr. K. Salomaa [43, 46]. Portions of Chapter 7, most notably Sections 7.3, 7.5 and 7.7, are also joint work with Dr. Salomaa [44].

Statement of Originality

I, Michael Domaratzki, certify that all results in this thesis are original, unless otherwise noted. Specifically, those results due to other authors which have appeared in the literature have been cited as necessary.

The work in this thesis has either appeared [43, 36, 37, 39, 40] or is to appear [44, 46] in the literature.

Contents

Abstract	i
Acknowledgments	ii
Co-Authorship	iii
Statement of Originality	iv
Contents	v
List of Figures	xii
1 Introduction	1
1.1 Formal Languages and Operations: Introduction and Motivation	1
1.2 Descriptive Complexity of Languages	3
1.3 Codes and Trajectories	4
1.4 Language Equations	5
1.5 Iteration	6
1.6 Organization	7
2 Preliminary Definitions	8
2.1 Formal Language Theory	8

2.2	Regular Languages	10
2.3	Grammars	13
2.4	Complexity Theory	14
2.5	Decidability	16
2.6	Families of Languages	17
2.7	Shuffle on Trajectories	18
2.7.1	Examples	19
2.7.2	Algebraic Properties	20
3	Related Work	21
3.1	Introduction	21
3.2	Shuffle	21
3.2.1	Iteration	22
3.2.2	Decomposition	23
3.2.3	Grammar Formalisms	24
3.3	Insertion and Deletion Operations	25
3.3.1	Insertion Operations	25
3.3.2	Deletion Operations	27
3.3.3	Interaction	28
3.3.4	Iteration	29
3.3.5	Decomposition and Related Language Equations	30
3.4	Shuffle on Trajectories	30
3.4.1	Infinite Words	30
3.4.2	Fairness	31
3.4.3	Related Concepts	31
3.4.4	Splicing on Routes	33
3.4.5	Concurrent Work	34

4	Descriptive Complexity	36
4.1	Introduction	36
4.2	General State Complexity Bounds	37
4.3	Slenderness and Trajectories	40
4.3.1	Perfect Shuffle	41
4.3.2	Bounds on Slender Trajectories	44
4.4	Future Directions	53
4.4.1	Polynomial Density Trajectories	53
4.4.2	Exponential Density Trajectories	53
4.4.3	Other Open Problems	54
4.5	Conclusions	54
5	Deletion along Trajectories	55
5.1	Introduction	55
5.2	Definitions	56
5.3	Closure and Characterization Results	57
5.3.1	Recognizing Deletion Along Trajectories	61
5.3.2	Equivalence of Trajectories	62
5.4	Regularity-Preserving Sets of Trajectories	63
5.5	i -Regularity	72
5.6	Filtering and Deletion along Trajectories	75
5.7	Splicing on Routes	76
5.8	Inverse Word Operations	80
5.8.1	Left Inverse	81
5.8.2	Right Inverse	83
5.9	Conclusions	84

6	Trajectory-Based Codes	85
6.1	Introduction	85
6.2	Definitions	87
6.3	General Properties of T -codes	88
6.4	The Binary Relation defined by Trajectories	92
6.4.1	Anti-symmetry	93
6.4.2	Reflexivity	94
6.4.3	Positivity	94
6.4.4	ST-Strictness	94
6.4.5	Cancellativity	95
6.4.6	Compatibility	96
6.4.7	Transitivity	98
6.4.8	Monotonicity	103
6.4.9	Well-Foundedness	105
6.5	Transitivity and Bases	106
6.6	Convexity and Transitivity	110
6.7	Closure Properties	113
6.7.1	Closure under Boolean Operations	113
6.7.2	Closure under Catenation	114
6.7.3	Closure under Inverse Morphism	115
6.7.4	Closure under Reversal	118
6.8	Maximal T -codes	118
6.8.1	Decidability and Maximal T -Codes	119
6.8.2	Transitivity and Embedding T -codes	123
6.9	Finiteness of all T -codes	127
6.9.1	Finiteness of Regular T -codes	128
6.9.2	Finiteness of Context-free T -codes	130

6.9.3	Finiteness of T -codes	132
6.9.4	Decidability and Finiteness Conditions	137
6.9.5	Up and Down Sets	137
6.9.6	T -Convexity Revisited	139
6.10	Conclusions	140
7	Language Equations	141
7.1	Introduction	141
7.2	Solving One-Variable Equations	142
7.2.1	Solving $X \sqcup_T L = R$ and $X \rightsquigarrow_T L = R$	142
7.2.2	Solving $L \sqcup_T X = R$ and $L \rightsquigarrow_T X = R$	143
7.2.3	Solving $\{x\} \sqcup_T L = R$	144
7.2.4	Solving $\{x\} \rightsquigarrow_T L = R$	144
7.3	Decidability of Shuffle Decompositions	146
7.3.1	1-thin sets of trajectories	151
7.4	Solving Quadratic Equations	152
7.5	Existence of Trajectories	153
7.6	Undecidability of One-Variable Equations	155
7.7	Undecidability of Shuffle Decompositions	159
7.8	Undecidability of Existence of Trajectories	162
7.9	Systems of Language Equations	165
7.10	Conclusions	169
8	Iteration of Trajectory Operations	171
8.1	Introduction	171
8.2	Definitions	172
8.3	Iterated Shuffle on Trajectories	173
8.3.1	Left-Associativity and a Simplified Definition	173

8.3.2	Some examples	175
8.3.3	Iteration and Density	176
8.4	Iterated Deletion	177
8.4.1	Iterated Scattered Deletion	179
8.4.2	Density and Iterated Deletion	184
8.5	Additional Closure Properties	184
8.6	T -Closure of a Language	186
8.6.1	Shuffle- T Quotient	186
8.6.2	T -closure	188
8.6.3	Codes and Shuffle-Closed Languages	190
8.7	Deletion Closure of a Language	191
8.7.1	Del- T Quotient	191
8.7.2	T -del-closure	193
8.8	T -Shuffle Base	196
8.9	Shuffle-Free Languages	198
8.10	T -Primitive Words	200
8.10.1	T -Primitivity and T -roots	200
8.10.2	Freeness and Uniqueness of T -Primitive Roots	202
8.11	Language Equations Revisited	208
8.11.1	Arden-like Equations	209
8.11.2	A Language Equation for $(\sqcup_T)^+(L)$	211
8.12	Conclusions	213
9	Conclusions	214
9.1	Results and Focus	214
9.2	Open Problems	216
9.3	Further Research Directions	217

9.3.1	Confluence of ω_T	217
9.3.2	Codes Defined by Multiple Sets of Trajectories	218
9.3.3	Semantic Trajectory-Based Operations	218
	Bibliography	221
	Index	241

List of Figures

2.1	A DFA, illustrated.	12
2.2	Some examples of shuffle on trajectories and their algebraic properties.	20
4.1	A two-state NFA accepting the set $T = 0^*1^*$ of trajectories.	40
5.1	Construction of the words in M_1 and M_2 from the action of M	70
6.1	Two factorizations of t	134
8.1	Summary of minimum-density regular languages and regular sets of trajectories demonstrating non-closure properties for iterated shuffle on trajectories.	178
8.2	Summary of minimum-density regular languages and regular sets of trajectories demonstrating non-closure properties for iterated deletion along trajectories.	184