On Fixed Point Equations over Commutative Semirings

Javier Esparza, Stefan Kiefer, and Michael Luttenberger

Universität Stuttgart
Institute for Formal Methods in Computer Science
Stuttgart, Germany
{esparza,kiefersn,luttenml}@informatik.uni-stuttgart.de

Abstract. Fixed point equations $\mathbf{x} = \mathbf{f}(\mathbf{x})$ over ω-continuous semirings can be seen as the mathematical foundation of interprocedural program analysis. The sequence $\mathbf{0}, \mathbf{f}(\mathbf{0}), \mathbf{f}^2(\mathbf{0}), \ldots$ converges to the least fixed point $\mu\mathbf{f}$. The convergence can be accelerated if the underlying semiring is commutative. We show that accelerations in the literature, namely Newton's method for the arithmetic semiring [3] and an acceleration for commutative Kleene algebras due to Hopkins and Kozen [4], are instances of a general algorithm for arbitrary commutative ω-continuous semirings. In a second contribution, we improve the $\mathcal{O}(3^n)$ bound of [4] and show that their acceleration reaches $\mu\mathbf{f}$ after n iterations, where n is the number of equations. Finally, we apply the Hopkins-Kozen acceleration to itself and study the resulting hierarchy of increasingly fast accelerations.

1 Introduction

Interprocedural program analysis is the art of extracting information about the executions of a procedural program without executing it, and fixed point equations over ω-continuous semirings can be seen as its mathematical foundation. A program can be mapped (in a syntax-driven way) to a system of fixed point equations over an abstract semiring containing one equation for each program point. Depending on the information on the program one wants to compute, the carrier and the abstract semiring operations can be instantiated so that the desired information is the least solution of the system. To illustrate this, consider a (very abstractly defined) program consisting of one single procedure X. This procedure can either do an action a and terminate, or do an action b and call itself twice. Schematically:

$$X \xrightarrow{a} \varepsilon \qquad X \xrightarrow{b} XX$$

The abstract equation corresponding to this program is

$$x = r_a + r_b \cdot x \cdot x \tag{1}$$

where $+$ and $\cdot$ are the abstract semiring operations. In order to compute the language $L(X)$ of terminating executions of the program, we instantiate the

semiring as follows: The carrier is $2^{\{a,b\}^*}$ (the set of languages over the alphabet $\{a,b\}$), $r_a = \{a\}$, $r_b = \{b\}$, $+$ is set union, and $\cdot$ is language concatenation. It is easy to prove that $L(X)$ is the least solution of (1) under this interpretation. But we can also be interested in other questions. We may wish to compute the *Parikh image* of $L(X)$, i.e., the set of vectors $(n_a, n_b) \in \mathbb{N}^2$ such that some terminating execution of the program does exactly n_a a's and n_b b's, respectively. For this, we take $2^{\mathbb{N}^2}$ as carrier, $r_a = \{(1,0)\}$, $r_b = \{(0,1)\}$, define $+$ as set union and $\cdot$ by $X \cdot Y = \{(x_a + y_a, x_b + y_b) \mid (x_a, x_b) \in X, (y_a, y_b) \in Y\}$. We may also be interested in quantitative questions. For instance, assume that the program X executes a with probability p and b with probability $(1-p)$. The probability that X eventually terminates is the least solution of (1) interpreted over $\mathbb{R}^+ \cup \{0, \infty\}$ with $r_a = p$, $r_b = (1-p)$, and the standard interpretation of $+$ and $\cdot$ (see for instance [2, 3]). If instead of the probability of termination we are interested in the probability of the most likely execution, we just have to reinterpret $+$ as the max operator.

The semirings corresponding to all these interpretations share a property called *ω-continuity* [7]. This property allows to apply the Kleene fixed point theorem and to prove that the least solution of a system of equations $\mathbf{x} = \mathbf{f}(\mathbf{x})$ is the supremum of the sequence $\mathbf{0}, \mathbf{f}(\mathbf{0}), \mathbf{f}^2(\mathbf{0}), \ldots$, where $\mathbf{0}$ is the vector whose components are all equal to the neutral element of $+$. If the carrier of the semiring is finite, this yields a procedure to compute the solution. However, if the carrier is infinite, the procedure rarely terminates, and its convergence can be very slow. For instance, the approximations to $L(X)$ are all finite sets of words, while $L(X)$ is infinite. Another example is the probability case with $p = 1/2$; the least fixed point (the least solution of $x = 1/2x^2 + 1/2$) is 1, but $\mathbf{f}^k(0) \leq 1 - \frac{1}{k+1}$ for every $k \geq 0$, which means that the Kleene scheme needs 2^i iterations to approximate the solution within i bits of precision[1].

Due to the slow convergence of $(\mathbf{f}^k(\mathbf{0}))_{k \geq 0}$, it is natural to look for "accelerations". Loosely speaking, an acceleration is a procedure of low complexity that on input $\mathbf{f}$ yields a function $\mathbf{g}$ having the same least fixed point $\mu\mathbf{f}$ as $\mathbf{f}$, but such that $(\mathbf{g}^k(\mathbf{0}))_{k \geq 0}$ converges faster to $\mu\mathbf{f}$ than $(\mathbf{f}^k(\mathbf{0}))_{k \geq 0}$. In [4], Hopkins and Kozen present a very elegant acceleration—although they do not use this term—that works for *every* commutative and idempotent ω-continuous semiring[2], i.e., for every ω-continuous semiring in which $\cdot$ is commutative and $+$ is idempotent (this is the case for both the Parikh image and the probability of the most likley computation). They prove that, remarkably, the acceleration is guaranteed to terminate. More precisely, they show that the fixed point is always reached after at most $\mathcal{O}(3^n)$ iterations, where n is the number of equations.

In this paper we further investigate the Hopkins-Kozen acceleration. In the first part of the paper we show that, in a certain formal sense, this acceleration was already discovered by Newton more than 300 years ago. In the arithmetic semiring, where the carrier is $\mathbb{R}^+ \cup \{0, \infty\}$ and $+$ and $\cdot$ have their usual meanings,

[1] This example is adapted from [3].

[2] Actually, in [4] the result is proved for commutative Kleene algebras, an algebraic structure more general than our semirings (cf. Section 4.1).

one can compute the least solution of $\mathbf{x} = \mathbf{f}(\mathbf{x})$ as a zero of $\mathbf{f}(\mathbf{x}) - \mathbf{x}$. Due to this connection, Newton's numerical method for approximating the zeros of a differentiable function (see e.g. [8]) can also be seen as an acceleration for the arithmetic case, which has been been studied by Etessami and Yannakakis in [3] in a different context. Here we show that the Hopkins-Kozen acceleration and Newton's are two particular instances of an acceleration for equations over arbitrary commutative ω-continuous semirings [7] and, in this sense, "the same thing".

In a second contribution, we improve the $\mathcal{O}(3^n)$ bound of [4] and show that the acceleration is actually much faster: the fixed point is already reached after n iterations. Finally, in a third contribution we investigate the possibility of "accelerating the acceleration". We study a hierarchy $\{\mathcal{H}_i\}_{i\geq 1}$ of increasingly faster accelerations, with $\mathcal{H}_1$ as the Hopkins-Kozen acceleration, and show that k iterations of the i-th acceleration can already be matched by ki iterations of the basic acceleration.

In Section 2 we introduce commutative ω-continuous semirings following [7]. In Section 3 we introduce the Hopkins-Kozen acceleration and Newton's method. In Section 4 we present our generalisation and derive both the Hopkins-Kozen acceleration and Newton's method as particular cases. In Section 5 we prove that the Hopkins-Kozen acceleration terminates after n-steps. The hierarchy of accelerations is studied in Section 6. For space reasons all proofs have been moved to an appendix.

2 ω-Continuous Semirings

A *semiring* is a quintuple $\langle A, +, \cdot, 0, 1\rangle$ s.t.

(i) $\langle A, +, 0\rangle$ is a commutative monoid,
(ii) $\langle A, \cdot, 1\rangle$ is a monoid,
(iii) $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$ for all $a, b, c \in A$,
(iv) $0 \cdot a = a \cdot 0$ for all $a \in A$.

A semiring is

- *commutative* if $a \cdot b = b \cdot a$ for all $a, b \in A$;
- *idempotent* if $a + a = a$ for all $a \in A$;
- *naturally ordered* if the relation $\leq$ given by $a \leq b \Leftrightarrow \exists c \in A : a + c = b$ is a partial order (this relation is always reflexive and transitive, but not necessarily antisymmetric);
- *complete* if it is possible to define "infinite sums" as an extension of finite sums, that are associative, commutative and distributive with respect to $\cdot$ as are finite sums. The formal axioms are given in [7]. In complete semirings, the unary *-operator is defined by $a^* = \sum_{j\geq 0} a^j$. Notice that $a^* = 1 + aa^*$;
- *ω-continuous* if it is naturally ordered, complete, and for all sequences $(a_i)_{i\in\mathbb{N}}$ with $a_i \in A$

$$\sup\left\{\sum_{i=0}^{n} a_i \mid n \in \mathbb{N}\right\} = \sum_{i\in\mathbb{N}} a_i.$$

Notation 1. *We abbreviate commutative ω-continuous semiring to* cc-semiring.

Remark 1. For our proofs the existence and ω-continuity of *countable* sums is sufficient. While in the idempotent case there is the term of commutative *closed semirings* for such structures (see [5]), it seems that there is no such term in the non-idempotent case.

Examples of semirings include $\langle \mathbb{N} \cup \{0, \infty\}, +, \cdot, 0, 1\rangle$, $\langle \mathbb{R}^+ \cup \{0, \infty\}, +, \cdot, 0, 1\rangle$, $\langle \mathbb{N} \cup \{0, \infty\}, \min, +, \infty, 0\rangle$ and $\langle 2^{\Sigma^*}, \cup, \cdot, \emptyset, \varepsilon\rangle$. They are all ω-continuous. The last two have an idempotent $+$-operation (min resp. $\cup$), and all but the last one are commutative.

2.1 Systems of Power Series

Let A be an ω-continuous semiring and let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a set of variables. We write $\mathbf{x}$ for the vector $(x_1, \ldots, x_n)^\top$. For every $i \in \{1, \ldots, n\}$, let $f_i(\mathbf{x})$ be a *(semiring) power series with coefficients in A*, i.e., a countable sum of products of elements of $A \cup \mathcal{X}$, and let $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x}))^\top$. We call $\mathbf{x} = \mathbf{f}(\mathbf{x})$ a *system of power series over A*. A vector $\bar{\mathbf{x}} \in A^n$ with $\mathbf{f}(\bar{\mathbf{x}}) = \bar{\mathbf{x}}$ is called a *solution* or a *fixed point* of $\mathbf{f}$.

Given two vectors $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in A^n$, we write $\bar{\mathbf{x}} \leq \bar{\mathbf{y}}$ if $\bar{x}_i \leq \bar{y}_i$ (w.r.t. the natural order of A) in every component. The *least fixed point of* $\mathbf{f}$, denoted by $\mu\mathbf{f}$, is the fixed point $\bar{\mathbf{x}}$ with $\bar{\mathbf{x}} \leq \bar{\mathbf{y}}$ for every fixed point $\bar{\mathbf{y}}$. It exists and can be computed by the following theorem.

Theorem 1 (Kleene fixed point theorem, cf. [7]). *Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be a system of power series over an ω-continuous semiring. Then $\mu\mathbf{f}$ exists and $\mu\mathbf{f} = \sup_{k \in \mathbb{N}} \mathbf{f}^k(\mathbf{0})$.*

3 Two Acceleration Schemes

Loosely speaking, an acceleration is a procedure that on input $\mathbf{f}$ yields a function $\mathbf{g}$ having the same least fixed point $\mu\mathbf{f}$ as $\mathbf{f}$, but converging "faster" to it, meaning that $\mathbf{f}^k(\mathbf{0}) \leq \mathbf{g}^k(\mathbf{0})$ for every $k \geq 0$. In order to exclude trivial accelerations like $\mathbf{g}(\mathbf{x}) = \mu\mathbf{f}$, a formal definition should require the procedure to have low complexity with respect to some reasonable complexity measure. Since such a definition would take too much space and would not be relevant for our results, we only use the term "acceleration" informally.

We describe two accelerations for different classes of cc-semirings. Both of them are based on the notion of derivatives. Given a polynomial or a power series $f(\mathbf{x})$, its derivative $\frac{\partial f}{\partial x_i}$ with respect to the variable x_i is defined as follows, where $a \in A$ and g, g_j, h are polynomials or power series (see also [4]):

$$\frac{\partial a}{\partial x_i} = 0 \qquad \frac{\partial}{\partial x_i}(g + h) = \frac{\partial g}{\partial x_i} + \frac{\partial h}{\partial x_i} \qquad \frac{\partial}{\partial x_i}(g \cdot h) = \frac{\partial g}{\partial x_i} \cdot h + g \cdot \frac{\partial h}{\partial x_i}$$

$$\frac{\partial x_j}{\partial x_i} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \qquad \frac{\partial}{\partial x_i} \sum_{j \in \mathbb{N}} g_j = \sum_{j \in \mathbb{N}} \frac{\partial g_j}{\partial x_i}$$

As usual, the *Jacobian* of a vector $\mathbf{f}(\mathbf{x})$ is the $n \times n$-matrix $\mathbf{f}'(\mathbf{x})$ given by $\mathbf{f}'(\mathbf{x})_{ij} = \dfrac{\partial f_i}{\partial x_j}$.

3.1 The Hopkins-Kozen Acceleration

In [4] Hopkins and Kozen introduce an acceleration of the Kleene procedure for *idempotent* cc-semirings and prove that it reaches the fixed point after finitely many steps. Given a system of power series $\mathbf{x} = \mathbf{f}(\mathbf{x})$, the *Hopkins-Kozen sequence* is defined by

$$\begin{aligned} \boldsymbol{\kappa}^{(0)} &= \mathbf{f}(\mathbf{0}) \\ \boldsymbol{\kappa}^{(k+1)} &= \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^* \cdot \boldsymbol{\kappa}^{(k)} \end{aligned}$$

Theorem 2 (Hopkins and Kozen [4]). *Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be a system of power series over an idempotent cc-semiring. There is a function $N : \mathbb{N} \to \mathbb{N}$ with $N(n) \in \mathcal{O}(3^n)$ s.t. $\boldsymbol{\kappa}^{(N(n))} = \mu\mathbf{f}$, where n is the number of variables of the system.*

Actually, [4] prove the theorem for commutative Kleene algebras, whose axioms are weaker than those of idempotent cc-semirings. There is no notion of infinite sums in the Kleene algebra axioms, especially the Kleene star operator $*$ and its derivative are defined axiomatically.

Example 1. Let $\langle 2^{\{a\}^*}, +, \cdot, 0, 1\rangle$ denote the cc-semiring $\langle 2^{\{a\}^*}, \cup, \cdot, \emptyset, \{\varepsilon\}\rangle$. For simplicity, we write a^i instead of $\{a^i\}$. Consider the equation system

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2^2 + a \\ x_1^2 \end{pmatrix} = \mathbf{f}(\mathbf{x}) \text{ with } \mathbf{f}'(\mathbf{x})^* = (x_1 x_2)^* \begin{pmatrix} 1 & x_2 \\ x_1 & 1 \end{pmatrix}.$$

The Hopkins-Kozen acceleration reaches the least fixed point $\mu\mathbf{f}$ after two steps:

$$\boldsymbol{\kappa}^{(0)} = (a, 0)^\top, \quad \boldsymbol{\kappa}^{(1)} = (a, a^2)^\top, \quad \boldsymbol{\kappa}^{(2)} = (a^3)^*(a, a^2)^\top.$$

It is easy to check that $\boldsymbol{\kappa}^{(2)}$ is a fixed point of $\mathbf{f}$. By Theorem 2 we have $\boldsymbol{\kappa}^{(2)} = \mu\mathbf{f}$.

3.2 Newton's Acceleration

Newton's method for approximating the zeros of a differentiable real function $\mathbf{g}(\mathbf{x})$ is one of the best known methods of numerical analysis. It computes the sequence

$$\begin{aligned} \mathbf{x}^{(0)} &= \mathbf{s} \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - \mathbf{g}'(\mathbf{x}^{(k)})^{-1} \cdot \mathbf{g}(\mathbf{x}^{(k)}) \end{aligned}$$

starting at the *seed* $\mathbf{s}$. Under certain conditions on $\mathbf{g}(\mathbf{x})$ and on the seed $\mathbf{s}$ (typically the seed must be "close enough" to the solution) the sequence converges to a solution of the equation $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

In order to approximate a solution of an equation system $\mathbf{x} = \mathbf{f}(\mathbf{x})$ over the reals, we can apply Newton's method to the function $\mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{x}$, which gives the sequence

$$\begin{aligned}\mathbf{x}^{(0)} &= \mathbf{0}\\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + (\mathbf{1} - \mathbf{f}'(\mathbf{x}^{(k)}))^{-1}(\mathbf{f}(\mathbf{x}^{(k)}) - \mathbf{x}^{(k)}).\end{aligned}$$

4 An Acceleration for Arbitrary cc-Semirings

We show that the Hopkins-Kozen and Newton's accelerations are two instances of a general acceleration for arbitrary cc-semirings, which we call the cc-scheme. The proof relies on lemmata from [4] and [3], which we reformulate and generalise so that they hold for arbitrary cc-semirings.

The *cc-scheme* is given by:

$$\begin{aligned}\boldsymbol{\nu}^{(0)} &= \mathbf{0}\\ \boldsymbol{\nu}^{(k+1)} &= \boldsymbol{\nu}^{(k)} + \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}),\\ &\qquad \text{where } \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) \text{ is any vector s.t. } \boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)})\end{aligned}$$

The scheme leaves the choice of $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)})$ free, but there is always at least one $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)})$ satisfying the condition (see Lemma 2 below).

The following theorem states that the cc-scheme accelerates the Kleene scheme $(\mathbf{f}^k(\mathbf{0}))_{k\in\mathbb{N}}$.

Theorem 3. *Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be a system of power series over a cc-semiring. Then the iterates $\boldsymbol{\nu}^{(k)}$ of the cc-scheme exist and satisfy $\mathbf{f}^k(\mathbf{0}) \leq \boldsymbol{\nu}^{(k)} \leq \mu\mathbf{f}$ for all $k \geq 0$.*

The proof uses the following fundamental property of derivatives in cc-semirings:

Lemma 1 (Taylor's Theorem, cf. [4]). *Let $\mathbf{f}(\mathbf{x})$ and $\mathbf{d}$ be vectors of power series over a cc-semiring. Then*

$$\mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x}) \cdot \mathbf{d} \leq \mathbf{f}(\mathbf{x} + \mathbf{d}) \leq \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x} + \mathbf{d}) \cdot \mathbf{d}.$$

The following lemma assures the existence of a suitable $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)})$ for each k.

Lemma 2. *Let $\boldsymbol{\nu}^{(k)}$ be the k-th iterate of the cc-scheme. For all $k \geq 0$: $\mathbf{f}(\boldsymbol{\nu}^{(k)}) \geq \boldsymbol{\nu}^{(k)}$. So, there is a $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)})$ such that $\boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)})$.*

What remains to show for Theorem 3 is $\mathbf{f}^k(\mathbf{0}) \leq \boldsymbol{\nu}^{(k)} \leq \mu\mathbf{f}$. See Appendix A for detailed proofs.

In the rest of the section we show that the Hopkins-Kozen acceleration and Newton's acceleration are special cases of the cc-scheme.

4.1 Idempotent cc-Semirings

If addition is idempotent, we have $x \leq y$ iff $x + y = y$, as $x \leq y$ implies that there is a d with $x + d = y$ so that $x + y = x + (x + d) = x + d = y$. By Lemma 2 we have $\boldsymbol{\nu}^{(k)} \leq \mathbf{f}(\boldsymbol{\nu}^{(k)})$. In the cc-scheme

$$\begin{aligned}\boldsymbol{\nu}^{(0)} &= \mathbf{0}\\ \boldsymbol{\nu}^{(k+1)} &= \boldsymbol{\nu}^{(k)} + \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}),\\ &\qquad \text{where } \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) \text{ is any vector s.t. } \boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)})\end{aligned}$$

we therefore may choose $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)})$. Moreover, since $\mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \geq \mathbf{1}$ by the definition of the Kleene star and since $\boldsymbol{\nu}^{(k)} \leq \mathbf{f}(\boldsymbol{\nu}^{(k)})$ by Lemma 2 we get

$$\boldsymbol{\nu}^{(k)} \leq \mathbf{f}(\boldsymbol{\nu}^{(k)}) \leq \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \mathbf{f}(\boldsymbol{\nu}^{(k)})$$

and by idempotence

$$\boldsymbol{\nu}^{(k)} + \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \mathbf{f}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \mathbf{f}(\boldsymbol{\nu}^{(k)})\,.$$

So the cc-scheme collapses in the idempotent case to

$$\begin{aligned}\boldsymbol{\nu}^{(0)} &= \mathbf{0}\\ \boldsymbol{\nu}^{(k+1)} &= \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \mathbf{f}(\boldsymbol{\nu}^{(k)}).\end{aligned}$$

In other words, $\boldsymbol{\nu}^{(k+1)}$ results from $\boldsymbol{\nu}^{(k)}$ by applying the operator $\mathcal{N}_{\mathbf{f}}(\mathbf{x}) := \mathbf{f}'(\mathbf{x})^* \cdot \mathbf{f}(\mathbf{x})$. Recall that the Hopkins-Kozen sequence is given by

$$\begin{aligned}\boldsymbol{\kappa}^{(0)} &= \mathbf{f}(\mathbf{0})\\ \boldsymbol{\kappa}^{(k+1)} &= \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^* \cdot \boldsymbol{\kappa}^{(k)}.\end{aligned}$$

So it is obtained by repeatedly applying the Hopkins-Kozen operator $\mathcal{H}_{\mathbf{f}}(\mathbf{x}) := \mathbf{f}'(\mathbf{x})^* \cdot \mathbf{x}$, starting from $\mathbf{f}(\mathbf{0})$. While the two sequences are not identical, the following theorem shows that they are essentially the same.

Theorem 4.

1. *For all* $k > 0 : \boldsymbol{\kappa}^{(k-1)} \leq \boldsymbol{\nu}^{(k)} \leq \boldsymbol{\kappa}^{(k)}$.
2. *For all* $k \geq 0 : \boldsymbol{\kappa}^{(k)} = \mathcal{H}^k_{\mathbf{f}}(\mathbf{f}(\mathbf{0})) = \mathcal{N}^k_{\mathbf{f}}(\mathbf{f}(\mathbf{0}))$.

4.2 The Semiring over the Nonnegative Reals

We now consider the cc-semiring $\langle \mathbb{R}^+ \cup \{0, \infty\}, +, \cdot, 0, 1\rangle$. In order to instantiate the cc-scheme, we have to choose $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)})$ so that $\boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)})$ holds. By Lemma 2 we have $\boldsymbol{\nu}^{(k)} \leq \mathbf{f}(\boldsymbol{\nu}^{(k)})$, and so we can take $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}$. The cc-acceleration becomes

$$\begin{aligned}\boldsymbol{\nu}^{(0)} &= \mathbf{0}\\ \boldsymbol{\nu}^{(k+1)} &= \boldsymbol{\nu}^{(k)} + \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot (\mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)})\,.\end{aligned}$$

It is easy to see that for any nonnegative real-valued square matrix $\mathbf{A}$, if $\sum_{k\in\mathbb{N}} \mathbf{A}^k = \mathbf{A}^*$ has only finite entries, then $(\mathbf{1} - \mathbf{A})^{-1}$ exists and equals $\mathbf{A}^*$. If this is the case for $\mathbf{A} = \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*$, then Newton's method coincides with the cc-acceleration for the reals and thus converges to $\mu\mathbf{f}$. In [3] Etessami and Yannakakis give sufficient conditions for $\mathbf{f}'(\boldsymbol{\nu}^{(k)})^* = (\mathbf{1} - \mathbf{f}'(\boldsymbol{\nu}^{(k)}))^{-1}$ when $\mathbf{f}$ is derived from a recursive Markov chain.

5 Convergence Speed in Idempotent Semirings

In the first subsection we want to analyse how many steps the Newton iteration or, equivalently, the Hopkins-Kozen iteration needs to reach $\mu\mathbf{f}$ when we consider an idempotent cc-semiring $\langle A, +, \cdot, 0, 1\rangle$, i.e. we have the additional equation $1 + 1 = 1$. In the subsequent subsection we then generalise the obtained results to the setting of commutative Kleene algebras.

5.1 Idempotent cc-Semirings

In this subsection $\mathbf{f}$ again denotes a system of n power series in the variables $\mathcal{X} = \{x_1, \ldots, x_n\}$, i.e. we have

$$f_i(\mathbf{x}) = \sum_{\boldsymbol{\iota} \in \mathbb{N}^n} c_{\boldsymbol{\iota}}^{(i)} \mathbf{x}^{\boldsymbol{\iota}},$$

where $\mathbf{x}^{\boldsymbol{\iota}}$ denotes the product $x_1^{\iota_1} \cdot \ldots \cdot x_n^{\iota_n}$ and $c_{\boldsymbol{\iota}}^{(i)} \in A$ for all $\boldsymbol{\iota} \in \mathbb{N}^n$ and $1 \leq i \leq n$. We define the concept of *derivation trees* of our system $\mathbf{f}$ as in formal language theory.

Notation 2. *In the following, if u is a node of a tree t, we identify u with the subtree of t rooted at u. In particular, t is also used to denote t's root. A* chain *in a tree t is any path from t's root to a leaf of t. The* height *$h(t)$ of t is then the length of its longest chain, e.g. a tree consisting only of a single node has height 0.*

Definition 1. *A* partial derivation tree *t of x_i is a labelled tree satisfying these conditions:*

- *every node of t is labelled by either an element of A or an element of $\mathcal{X}$,*
- *its root is labelled by x_i, and*
- *for each node u of t labeled by some variable x_k the following holds: Let $p_u(\mathbf{x})$ be the product of the labels of u's children. Then p_u is a summand of f_k, i.e. there exists a $\boldsymbol{\iota} \in \mathbb{N}^n$ with $c_{\boldsymbol{\iota}}^{(k)} \neq 0$ and $c_{\boldsymbol{\iota}}^{(k)} \mathbf{x}^{\boldsymbol{\iota}} = p_u(\mathbf{x})$.*

We call a partial derivation tree t a derivation tree *if no leaf of t is labelled by a variable. The* yield *$Y(t)$ of a derivation tree t is the product of the labels of its leaves.*

As in the case of formal languages we have the following

Theorem 5.

1. *The sum of yields of all derivation trees of x_i with height at most h equals $(\mathbf{f}^h(\mathbf{0}))_i$.*
2. *The sum of yields of all derivation trees of x_i equals $(\mu\mathbf{f})_i$.*

In the following we show that because of commutativity and idempotence already a special class of derivation trees is sufficient to reach $\mu\mathbf{f}$.

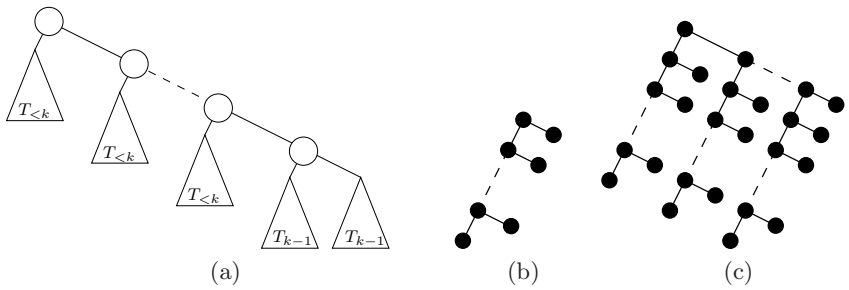

Fig. 1. (a) shows the general structure of a tree of dimension k, where $T_{<k}$ (T_{k-1}) represents any tree of dimension $< k$ ($= k-1$). (b) and (c) give some idea of the topology of one-, resp. two-dimensional trees in general.

Definition 2. *We define the* dimension $\dim(t)$ *of a tree* t *as follows.*

1. *A tree of height* 0 *or* 1 *has dimension* 0.
2. *Let* t *be a tree of height* $h(t) > 1$ *with children* $c_1, c_2, \ldots, c_s$ *where* $\dim(c_1) \geq \dim(c_2) \geq \ldots \dim(c_s)$. *Let* $d_1 = \dim(c_1)$. *If* $s > 1$, *let* $d_2 = \dim(c_2)$, *otherwise let* $d_2 = 0$. *Then we define*

$$\dim(t) := \begin{cases} d_1 + 1 & \text{if } d_1 = d_2 \\ d_1 & \text{if } d_1 > d_2. \end{cases}$$

(See also Fig. 5.1).

Remark 2. For every derivation tree t we have $h(t) > \dim(t)$.

Definition 3. *Let* t *be a derivation tree. We denote with* $V(t)$ *the number of distinct variables appearing as labels in* t. *We call* t compact *if* $\dim(t) \leq V(t)$.

In the following, we state two central lemmata that lead to the main result of this section. Lemma 3 tells us that it is sufficient to consider only compact derivation trees. Lemma 4 shows the connection between the dimension of a derivation tree and the Hopkins-Kozen sequence.

Lemma 3. *For each derivation tree* t *of* x_i *there is a compact derivation tree* t' *of* x_i *with equal yield.*

Lemma 4. *Let* t *be a derivation tree of* x_i *s.t.* $\dim(t) \leq k$. *Then* $Y(t) \leq (\boldsymbol{\kappa}^{(k)})_i$.[3]

The proof of Lemma 3 bears some similarity to the proof of the pumping lemma for context free languages. Let us call a partial derivation tree a *pumping* tree (p-tree) if it has exactly one leaf which bears the same label as its root

[3] In fact one can similarly show that $(\boldsymbol{\kappa}^{(k)})_i$ equals exactly the sum of yields of all derivation trees of x_i of dimension less than or equal to k.

and all other leaves are labelled by elements of A. Because of commutativity, reallocating such a p-tree from one subtree of t to another one does not change t's yield. We use a reallocation procedure to inductively reduce the dimension of t's subtrees, which eventually results in decreasing the dimension of t itself.

Theorem 6. *Let $\mathbf{f} : A^n \to A^n$ be a system of power series over an idempotent cc-semiring $\langle A, +, \cdot, 0, 1\rangle$. Then $\mu\mathbf{f} = \boldsymbol{\kappa}^{(n)}$.*

Proof. First recall that by Theorem 3 ($\boldsymbol{\nu}^{(k)} \leq \mu\mathbf{f}$) and Theorem 4 ($\boldsymbol{\kappa}^{(k-1)} \leq \boldsymbol{\nu}^{(k)} \leq \boldsymbol{\kappa}^{(k)}$) we have $\boldsymbol{\kappa}^{(n)} \leq \mu\mathbf{f}$. Obviously, $V(t) \leq n$ for every derivation tree t of x_i. Lemma 3 allows to assume that t is compact, i.e. $\dim(t) \leq V(t) \leq n$. Lemma 4 thus implies $Y(t) \leq (\boldsymbol{\kappa}^{(n)})_i$. Therefore the sum of yields of derivation trees of x_i is less than or equal to $(\boldsymbol{\kappa}^{(n)})_i$. But Theorem 5 tells us that this sum is already $(\mu\mathbf{f})_i$. Hence $\mu\mathbf{f} \leq \boldsymbol{\kappa}^{(n)} \leq \mu\mathbf{f}$. □

Remark 3. The bound of this theorem is tight as can be shown by a generalisation of Example 1: If $\mathbf{f}(\mathbf{x}) = (x_2^2 + a, x_3^2, \ldots, x_n^2, x_1^2)^\top$, then $(\boldsymbol{\kappa}^{(k)})_1 = a$ for $k < n$, but $a^{2^n} \leq (\boldsymbol{\kappa}^{(n)})_1 = (\mu\mathbf{f})_1$.

5.2 Generalisation to Commutative Kleene Algebras

Notation 3. *Let M be any set. Then RExp_M denotes the set of* regular expressions *generated by the elements of M. We write $R_M : \mathsf{RExp}_M \to 2^{M^*}$ for their* canonical interpretation *as languages.*

For this subsection, let $\mathbf{f}$ denote a system of n *regular expressions* $f_i \in \mathsf{RExp}_{K\cup\mathcal{X}}$. We are again interested in the least solution $\mu\mathbf{f}$ of $\mathbf{x} = \mathbf{f}(\mathbf{x})$, but this time over the *commutative Kleene algebra* $\langle K, +, \cdot, ^*, 0, 1\rangle$. A *commutative Kleene algebra* is an idempotent commutative semiring $\langle K, +, \cdot, 0, 1\rangle$ where the *-operator is only required to satisfy these two equations for all $a, b, c \in K$:

$$1 + aa^* \leq a^* \quad \text{and} \quad a + bc \leq c \to b^*a \leq c.$$

In [4] it is proved that $\mu\mathbf{f}$ can be computed by applying the Hopkins-Kozen operator $\mathcal{H}_\mathbf{f}$ to $\mathbf{f}(\mathbf{0})$ for a finite number of times. In addition, $\mathcal{H}_\mathbf{f}^i(\mathbf{f}(\mathbf{0})) \leq \mu\mathbf{f}$ for all $i \in \mathbb{N}$.

As in the setting of cc-semirings the Hopkins-Kozen operator is defined by $\mathcal{H}_\mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})^*\mathbf{x}$. For $\mathcal{H}$ to be well defined over Kleene algebras, one has to define the partial derivatives $\frac{\partial}{\partial x_j}$ over $\mathsf{Reg}_{K\cup\mathcal{X}}$. This is done in [4] just as in the case of cc-semirings (see the beginning of Section 3), however the equation for the $\sum$-operator is replaced by the axiom $\frac{\partial \alpha^*}{\partial x_j} = \alpha^* \frac{\partial \alpha}{\partial x_j}$ for $\alpha \in \mathsf{RExp}_{K\cup\mathcal{X}}$.

We lift the result of the previous subsection to commutative Kleene algebras, improving the $\mathcal{O}(3^n)$ bound in [4]. More precisely we show that

$$\mathbf{f}(\mathcal{H}_\mathbf{f}^n(\mathbf{f}(\mathbf{0}))) = \mathcal{H}_\mathbf{f}^n(\mathbf{f}(\mathbf{0})). \tag{2}$$

In order to prove (2) we appeal to Redko's theorem [1] that essentially states that an equation of terms over any commutative Kleene algebra holds if it holds

under the *canonical commutative interpretation*. See Appendix B for a technical justification of this fact. Let Σ be the finite set of elements of K appearing in $\mathbf{f}$. The canonical commutative interpretation $\mathsf{c}_\Sigma : \mathsf{RExp}_\Sigma \to 2^{\mathbb{N}^\Sigma}$ is then defined by

$$\mathsf{c}_\Sigma(\alpha) = \{\#w \mid w \in R_\Sigma(\alpha)\},$$

where $\#w$ is the Parikh-vector of $w \in \Sigma^*$, i.e. $a \in \Sigma$ appears exactly $(\#w)_a$-times in w. We omit the subscript of c_Σ in the following. The idempotent cc-semiring of sets of Parikh-vectors $\mathcal{C}_\Sigma$ is defined by $\mathcal{C}_\Sigma = \langle 2^{\mathbb{N}^\Sigma}, \cup, +, \emptyset, \{\mathbf{0}\}\rangle$ with $A+B = \{a+b \mid a \in A, b \in B\}$ for all $A, B \subseteq \mathbb{N}^\Sigma$ and $\sum S = \bigcup S$ for all $S \subseteq 2^{\mathbb{N}^\Sigma}$. By Redko's theorem, we can prove (2) by showing $\mathsf{c}(\mathbf{f}(\mathcal{H}^n_{\mathbf{f}}(\mathbf{f}(\mathbf{0})))) = \mathsf{c}(\mathcal{H}^n_{\mathbf{f}}(\mathbf{f}(\mathbf{0})))$ over $\mathcal{C}_\Sigma$.

For any function $g : \mathsf{RExp}_\Sigma \to \mathsf{RExp}_\Sigma$, let g^{c} denote the commutative interpretation of g as a map over $\mathcal{C}_\Sigma$, i.e. $\mathsf{c}(g(\alpha)) = g^{\mathsf{c}}(\mathsf{c}(\alpha))$ for all $\alpha \in \mathsf{RExp}_\Sigma$. In particular $\mathsf{c}(\alpha^*) = \bigcup_{i\in\mathbb{N}} \mathsf{c}(\alpha^i)$. Notice that this definition is consistent with the axiomatic definition of derivatives of *-expressions, since

$$\mathsf{c}(\frac{\partial}{\partial x_i}(\alpha^*)) = \mathsf{c}(\alpha^* \frac{\partial}{\partial x_i}(\alpha)) = \bigcup_{j\in\mathbb{N}} \mathsf{c}(\alpha^j) \frac{\partial}{\partial x_i}(\mathsf{c}(\alpha)) = \frac{\partial}{\partial x_i}\left(\bigcup_{j\in\mathbb{N}} \mathsf{c}(\alpha^j)\right) = \frac{\partial}{\partial x_i}(\mathsf{c}(\alpha^*)).$$

We then have $(\mathcal{H}_{\mathbf{f}})^{\mathsf{c}} = \mathcal{H}_{\mathbf{f}^{\mathsf{c}}}$. Furthermore, by Theorem 6, $\mathcal{H}^n_{\mathbf{f}^{\mathsf{c}}}(\mathbf{f}^c(\emptyset))$ solves the equation system $\mathbf{x} = \mathbf{f}^{\mathsf{c}}(\mathbf{x})$ over $\mathcal{C}_\Sigma$. Combined, we have

$$\mathsf{c}(\mathbf{f}(\mathcal{H}^n_{\mathbf{f}}(\mathbf{f}(\mathbf{0})))) = \mathbf{f}^{\mathsf{c}}((\mathcal{H}^n_{\mathbf{f}})^{\mathsf{c}}(\mathbf{f}^{\mathsf{c}}(\emptyset))) = \mathbf{f}^{\mathsf{c}}(\mathcal{H}^n_{\mathbf{f}^{\mathsf{c}}}(\mathbf{f}^c(\emptyset))) = \mathcal{H}^n_{\mathbf{f}^{\mathsf{c}}}(\mathbf{f}^c(\emptyset)) = \mathsf{c}(\mathcal{H}^n_{\mathbf{f}}(\mathbf{f}(\mathbf{0}))).$$

This proves the following theorem.

Theorem 7. *Let $\mathbf{f} \in \mathsf{RExp}^n_{K\cup\mathcal{X}}$ define a system $\mathbf{x} = \mathbf{f}(\mathbf{x})$ over a commutative Kleene algebra $\langle K, +, \cdot, ^*, 0, 1\rangle$. Then $\mu\mathbf{f} = \boldsymbol{\kappa}^{(n)}$.*

6 A Hierarchy of Accelerations

In this section we apply the Hopkins-Kozen acceleration to itself. Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be an equation system of degree-2-polynomials over a commutative Kleene algebra. Any polynomial equation system (even with *-expressions) can be reduced to this "Chomsky normal" form by introducing auxiliary variables.

Recall the Hopkins-Kozen operator $\mathcal{H}_{\mathbf{f}}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})^*\mathbf{x}$. As shown in [4] and in the previous section, the sequence $\mathcal{H}^i_{\mathbf{f}}(\mathbf{f}(\mathbf{0}))$ is "faster" than $\mathbf{f}^i(\mathbf{f}(\mathbf{0}))$ to the extent that the fixed point iteration of $\mathcal{H}_{\mathbf{f}}$ reaches $\mu\mathbf{f}$ in a finite number of steps, whereas the fixed point iteration of $\mathbf{f}$ may not reach $\mu\mathbf{f}$. We study in this section how fast accelerations $\mathcal{H}_{\mathcal{H}_{\mathbf{f}}}, \mathcal{H}_{\mathcal{H}_{\mathcal{H}_{\mathbf{f}}}}, \ldots$ are compared to $\mathcal{H}_{\mathbf{f}}$. We write $\mathcal{H}_1$ for $\mathcal{H}_{\mathbf{f}}$ and $\mathcal{H}_{i+1}$ for $\mathcal{H}_{\mathcal{H}_i} = (\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}_i(\mathbf{x})))^*\mathbf{x}$. In the following we mean $\mathcal{H}_{\mathbf{f}}$ when the subscript of $\mathcal{H}$ is omitted. Our hierarchy theorem states that using $\mathcal{H}_i$ once amounts to using $\mathcal{H}$ i-times:

Theorem 8. *For all $i \geq 1 : \mathcal{H}_i(\mathbf{x}) = \mathcal{H}^i(\mathbf{x})$.*

Combined with Theorem 6 we conclude that the least fixed point $\mu\mathbf{f}$ can be computed by (a) iteratively applying $\mathcal{H}$ to $\mathbf{f}(\mathbf{0})$ (n times) or (b) computing the operator $\mathcal{H}^n$ and applying it to $\mathbf{f}(\mathbf{0})$ once or (c) computing the operator $\mathcal{H}_n$ and applying it to $\mathbf{f}(\mathbf{0})$ once. A discussion which method is most appropriate depends on the the particular applications and is beyond the scope of this paper.

Example 2. We continue Example 1 where we have shown that $\mathcal{H}^2(\mathbf{f}(\mathbf{0})) = \mu\mathbf{f}$. Now we illustrate Theorem 8 by showing that $\mathcal{H}_2(\mathbf{f}(\mathbf{0})) = \mu\mathbf{f}$. We have

$$\mathcal{H}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})^*\mathbf{x} = (x_1x_2)^* \begin{pmatrix} x_1 + x_2^2 \\ x_1^2 + x_2 \end{pmatrix} ,$$

$$\mathcal{H}'(\mathbf{x}) = (x_1x_2)^* \begin{pmatrix} 1 + x_2^3 & x_1^2 + x_2 \\ x_2^2 + x_1 & 1 + x_1^3 \end{pmatrix} ,$$

$$\mathcal{H}'(\mathbf{f}(\mathbf{0})) = \begin{pmatrix} 1 & a^2 \\ a & 1 + a^3 \end{pmatrix} \quad \text{and} \quad \mathcal{H}'(\mathbf{f}(\mathbf{0}))^* = (a^3)^* \begin{pmatrix} 1 & a^2 \\ a & 1 \end{pmatrix} .$$

So $\mathcal{H}_2(\mathbf{f}(\mathbf{0})) = \mathcal{H}'(\mathbf{f}(\mathbf{0}))^*\mathbf{f}(\mathbf{0}) = (a^3)^* \begin{pmatrix} a \\ a^2 \end{pmatrix} = \mu\mathbf{f}$.

7 Conclusions

We have studied the Hopkins-Kozen acceleration scheme for solving fixed point equations $\mathbf{x} = \mathbf{f}(\mathbf{x})$ over commutative Kleene algebras [4]. We have shown that, maybe surprisingly, the scheme is tightly related to Newton's method for approximating a zero of a differentiable real function. Loosely speaking, the scheme is the result of generalising Newton's method to commutative ω-continuous semirings in a very straightforward way, and then instantiating this generalisation to the case in which addition is idempotent. In the proof we very much profit from a result by Etessami and Yannakakis on using Newton's method to solve fixed point equations derived from recursive Markov chains [3]. At the same time, our result extends Etessami and Yannakakis' result to arbitrary commutative ω-continuous semirings, a much more general algebraic setting.

We have also proved that the Hopkins-Kozen scheme terminates after n iterations for a system of n equations, improving on the $O(3^n)$ bound of [4]. As in [4], our bound holds for arbitrary commutative Kleene algebras.

Finally, we have studied the result of applying the scheme to itself, leading to a sequence of faster and faster accelerations. The Hopkins-Kozen scheme can be "arbitrarily faster" than the basis scheme derived from Kleene's theorem (the scheme computing $(\mathbf{f}^k(\mathbf{0}))_{k\geq 0}$) because it is guaranteed to terminate, while Kleene's scheme is not. We have shown that, on the contrary, the reduction in the number of iterations achieved by subsequent accelerations is very moderate: one iteration of the scheme obtained by applying k times the acceleration to itself is already matched by k iterations of the Hopkins-Kozen scheme.

Our work can be extended in several directions. Our proof of the new bound relies on formal languages concepts, and is therefore very non-algebraic. We

intend to search for an algebraic proof. We also plan to investigate accelerations for the non-commutative case.

Acknowledgements

We thank Volker Diekert for helpful discussions and the anonymous referees for useful comments.

References

1. J.H. Conway. *Regular Algebra and Finite Machines*. Chapman and Hall, 1971.
2. J. Esparza, A. Kučera, and R. Mayr. Model checking probabilistic pushdown automata. In *LICS 2004*. IEEE Computer Society, 2004.
3. K. Etessami and M. Yannakakis. Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations. In *STACS*, pages 340–352, 2005.
4. M. W. Hopkins and D. Kozen. Parikh's theorem in commutative Kleene algebra. In *Logic in Computer Science*, pages 394–401, 1999.
5. D. Kozen. On Kleene algebras and closed semirings. In B. Rovan, editor, *Proc. Math. Found. Comput. Sci.*, volume 452 of *Lecture Notes in Computer Science*, pages 26–47, Banska-Bystrica, Slovakia, 1990. Springer-Verlag.
6. D. Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. In *Logic in Computer Science*, pages 214–225, 1991.
7. W. Kuich. *Handbook of Formal Languages*, volume 1, chapter 9: Semirings and Formal Power Series: Their Relevance to Formal Languages and Automata, pages 609 – 677. Springer, 1997.
8. J.M. Ortega. *Numerical Analysis: A Second Course*. Academic Press, New York, 1972.

A Proofs

A.1 Proof of Theorem 1

Theorem 1 (Kleene fixed point theorem, cf. [7]). *Let* $\mathbf{x} = \mathbf{f}(\mathbf{x})$ *be a system of power series over an* ω*-continuous semiring. Then* $\mu\mathbf{f}$ *exists and* $\mu\mathbf{f} = \sup_{k\in\mathbb{N}} \mathbf{f}^k(\mathbf{0})$.

Proof. This theorem is similarly stated in [7] for semiring polynomials. In order to generalise it to power series, we only need to argue that infinite sums are ω-*continuous functions*, as are addition and multiplication. A function $f : A \to A$ is said to be ω-continuous if for any sequence $(s_i)_{i\in\mathbb{N}}$ of elements of A s.t. $s_i \leq s_{i+1}$,

$$f(\sup_{i\in\mathbb{N}} s_i) = \sup_{i\in\mathbb{N}} f(s_i).$$

We show that any power series is ω-continuous, i.e., let $f_j = c_j x^j$ for $j \in \mathbb{N}$ and $c_j \in A$, and let $f = \sum_{j\in\mathbb{N}} f_j$. Any power series can be written in this sum form.

Because we have $\sum_{j\in\mathbb{N}} f_j(\sup_i s_i) \geq \sum_{j\in\mathbb{N}} f_j(s_i)$ for all i, also

$$\sum_{j\in\mathbb{N}} f_j(\sup_i s_i) \geq \sup_i \sum_{j\in\mathbb{N}} f_j(s_i)$$

holds. For the other direction, we have for all $n \in \mathbb{N}$:

$$\begin{aligned} \textstyle\sum_{j=0}^{n} f_j(\sup_i s_i) &= \textstyle\sum_{j=0}^{n} \sup_i f_j(s_i) && (\omega\text{-continuity of } f_j) \\ &= \textstyle\sup_i \sum_{j=0}^{n} f_j(s_i) && (\omega\text{-continuity of finite sums}) \\ &\leq \textstyle\sup_i \sum_{j=0}^{\infty} f_j(s_i). \end{aligned}$$

Because the structure A is ω-continuous, it follows that

$$\sum_{j\in\mathbb{N}} f_j(\sup_i s_i) \leq \sup_i \sum_{j\in\mathbb{N}} f_j(s_i).$$

□

A.2 Proof of Lemma 1

Lemma 1 (Taylor's Theorem, cf. [4] (stronger version)). *Let* $\mathbf{f}(\mathbf{x})$ *and* $\mathbf{d}$ *be vectors of power series over a cc-semiring. Then*

$$\mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x}) \cdot \mathbf{d} \leq \mathbf{f}(\mathbf{x} + \mathbf{d}) \leq \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x} + \mathbf{d}) \cdot \mathbf{d}.$$

Moreover, in the idempotent case we have $\mathbf{f}(\mathbf{x} + \mathbf{d}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x} + \mathbf{d}) \cdot \mathbf{d}$.

Proof. The proof handles only the first inequation and the case $n = 1$ explicitly. The general result is obtained by iteration as in [4].

We first consider the case of f being a (finite) polynomial. We argue by induction over the polynomial structure.

Case 1: $f(x) = a$, where a is a constant or a variable different from x. Then $f'(x) = 0$ and $f(x + d) = f(x) = a$.
Case 2: $f(x) = x$. Then $f'(x) = 1$ and $f(x + d) = x + d = f(x) + 1 \cdot d$.
Case 3: $f(x) = g(x) + h(x)$. Then $f'(x) = g'(x) + h'(x)$ and (by induction) $f(x+d) = g(x+d)+h(x+d) \geq g(x)+g'(x)\cdot d+h(x)+h'(x)\cdot d = f(x)+f'(x)\cdot d$.
Case 4: $f(x) = g(x)h(x)$. Then $f'(x) = g'(x)h(x) + g(x)h'(x)$ and (by induction) $f(x + d) = g(x + d)h(x + d) \geq (g(x) + g'(x)d)(h(x) + h'(x)d) \geq g(x)h(x) + (g'(x)h(x) + g(x)h'(x))d = f(x) + f'(x)d$.

Now, to generalise the result to power series, we exploit the fact that the power series over an ω-continuous semiring form again an ω-continuous semiring [7]. Let $f(x)$ be any power series, i.e., $f(x) = \sum_{j\in\mathbb{N}} f_j(x)$ with $f_j(x) = c_j x^j$ where the c_j are constants. By the proof above we have $f_j(x)+f_j'(x)\cdot d \leq f_j(x+d)$ and therefore $\sum_{j=0}^{n} f_j(x)+f_j'(x)\cdot d \leq \sum_{j=0}^{n} f_j(x+d) \leq f(x+d)$. By ω-continuity, $f(x) + f'(x) \cdot d \leq f(x + d)$ follows.

The result in the idempotent case is proved similarly to [4]. □

A.3 Proof of Lemma 2

Lemma 2 (stronger version). *Let $\boldsymbol{\nu}^{(k)}$ be the k-th iterate of the cc-scheme. For all $k \geq 0 : \mathbf{f}(\boldsymbol{\nu}^{(k)}) \geq \boldsymbol{\nu}^{(k)} \geq \mathbf{f}^k(\mathbf{0})$.*
In particular, since $\mathbf{f}(\boldsymbol{\nu}^{(k)}) \geq \boldsymbol{\nu}^{(k)}$, there is a $\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)})$ s.t. $\boldsymbol{\nu}^{(k)}+\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) = \mathbf{f}(\boldsymbol{\nu}^{(k)})$.

Proof. By induction on k. The base case $k = 0$ is obvious. For the step:

$$\begin{aligned}
\mathbf{f}(\boldsymbol{\nu}^{(k+1)}) &\geq \mathbf{f}(\boldsymbol{\nu}^{(k)}) + \mathbf{f}'(\boldsymbol{\nu}^{(k)})\mathbf{f}'(\boldsymbol{\nu}^{(k)})^*\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) && \text{(Taylor)}\\
&= \boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) + \mathbf{f}'(\boldsymbol{\nu}^{(k)})\mathbf{f}'(\boldsymbol{\nu}^{(k)})^*\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) && \text{(ind. hyp.)}\\
&= \boldsymbol{\nu}^{(k)} + \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*\boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) && (a^* = 1 + aa^*)\\
&= \boldsymbol{\nu}^{(k+1)}\\
&\geq \boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) && (\mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \geq \mathbf{1})\\
&= \mathbf{f}(\boldsymbol{\nu}^{(k)})\\
&\geq \mathbf{f}(\mathbf{f}^k(\mathbf{0})) && \text{(ind. hyp., monot. of } \mathbf{f})\\
&= \mathbf{f}^{k+1}(\mathbf{0})
\end{aligned}$$

□

A.4 Proof of Theorem 3

Lemma 5. *Let $\mathbf{f}(\mathbf{x}) \geq \mathbf{x}$. For all $d \geq 0$ there exists a vector $\boldsymbol{\delta}^{(d)}(\mathbf{x})$ s.t.*

$$\begin{aligned}
&\mathbf{f}^d(\mathbf{x}) + \boldsymbol{\delta}^{(d)}(\mathbf{x}) = \mathbf{f}^{d+1}(\mathbf{x}) \quad \textit{and}\\
&\qquad \boldsymbol{\delta}^{(d)}(\mathbf{x}) \geq \mathbf{f}'(\mathbf{f}^{d-1}(\mathbf{x})) \cdot \ldots \cdot \mathbf{f}'(\mathbf{x}) \cdot \boldsymbol{\delta}^{(0)}(\mathbf{x}) \geq \mathbf{f}'(\mathbf{x})^d \cdot \boldsymbol{\delta}^{(0)}(\mathbf{x}).
\end{aligned}$$

Proof. By induction on d. For $d = 0$ there is an appropriate $\boldsymbol{\delta}^{(0)}(\mathbf{x})$ by assumption. Let $d \geq 0$.

$$\begin{aligned}
\mathbf{f}^{d+2}(\mathbf{x}) &= \mathbf{f}(\mathbf{f}^d(\mathbf{x}) + \boldsymbol{\delta}^{(d)}(\mathbf{x})) && \text{(ind. hyp.)}\\
&\geq \mathbf{f}^{d+1}(\mathbf{x}) + \mathbf{f}'(\mathbf{f}^d(\mathbf{x})) \cdot \boldsymbol{\delta}^{(d)}(\mathbf{x}) && \text{(Taylor)}\\
&\geq \mathbf{f}^{d+1}(\mathbf{x}) + \mathbf{f}'(\mathbf{f}^d(\mathbf{x})) \cdot \ldots \cdot \mathbf{f}'(\mathbf{x}) \cdot \boldsymbol{\delta}^{(0)}(\mathbf{x}) && \text{(ind. hyp.)}
\end{aligned}$$

Therefore, there exists a $\boldsymbol{\delta}^{(d+1)}(\mathbf{x}) \geq \mathbf{f}'(\mathbf{f}^d(\mathbf{x})) \cdot \ldots \cdot \mathbf{f}'(\mathbf{x}) \cdot \boldsymbol{\delta}^{(0)}(\mathbf{x})$. Since $\mathbf{f}'$ is monotone and $\mathbf{x} \leq \mathbf{f}(\mathbf{x}) \leq \mathbf{f}^2(\mathbf{x}) \leq \ldots$, the second inequality also holds. □

Notice that by virtue of Lemma 2, Lemma 5 applies to the iterates of the cc-scheme.

Lemma 6. *For all $k \geq 0 : \boldsymbol{\nu}^{(k)} \leq \mu\mathbf{f}$.*

Proof. By induction on k.

$$\begin{array}{lll}
\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + \mathbf{f}'(\boldsymbol{\nu}^{(k)})^* \cdot \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) & & \\
= \boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty} \mathbf{f}'(\boldsymbol{\nu}^{(k)})^d \cdot \boldsymbol{\delta}(\boldsymbol{\nu}^{(k)}) & \text{(definition of } ^*) & \\
\leq \boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty} \boldsymbol{\delta}^{(d)}(\boldsymbol{\nu}^{(k)}) & \text{(Lemma 5)} & \\
= \sup_{d \geq 0} \mathbf{f}^d(\boldsymbol{\nu}^{(k)}) & (\omega\text{-continuity}) & \\
\leq \mu\mathbf{f} & \text{(ind. hyp., monotonicity:} & \\
 & \boldsymbol{\nu}^{(k)} \leq \mathbf{f}(\boldsymbol{\nu}^{(k)}) \leq \mathbf{f}^2(\boldsymbol{\nu}^{(k)}) \leq \cdots \leq \mu\mathbf{f}) &
\end{array}$$

□

Theorem 3. *Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be a system of power series over a cc-semiring. Then the iterates $\boldsymbol{\nu}^{(k)}$ of the cc-scheme exist and satisfy $\mathbf{f}^k(\mathbf{0}) \leq \boldsymbol{\nu}^{(k)} \leq \mu\mathbf{f}$ for all $k \geq 0$.*

Proof. The proof follows from Lemma 2 and Lemma 6. □

A.5 Proof of Theorem 4

We first show the following lemma.

Lemma 7. *For all $i \geq 0 : \mathcal{H}_{\mathbf{f}}^{i+1}(\mathbf{x}) = \mathbf{f}'(\mathcal{H}_{\mathbf{f}}^{i}(\mathbf{x}))^*\mathbf{x}$.*

Proof. By induction on k. The base case is trivial. Let $k \geq 1$. Then

$$\begin{array}{lll}
\mathcal{H}_{\mathbf{f}}^{k+1}(\mathbf{x}) = \mathcal{H}_{\mathbf{f}}(\mathcal{H}_{\mathbf{f}}^{k}(\mathbf{x})) = \mathbf{f}'(\mathcal{H}_{\mathbf{f}}^{k}(\mathbf{x}))^*\mathcal{H}_{\mathbf{f}}^{k}(\mathbf{x}) & \text{(by definition)} & \\
= \mathbf{f}'(\mathcal{H}_{\mathbf{f}}^{k}(\mathbf{x}))^*\mathbf{f}'(\mathcal{H}_{\mathbf{f}}^{k-1}(\mathbf{x}))^*\mathbf{x} & \text{(by induction hyp.)} & \\
= \mathbf{f}'(\mathcal{H}_{\mathbf{f}}^{k})^*\mathbf{x} & (\mathcal{H}_{\mathbf{f}}^{k-1}(\mathbf{x}) \leq \mathcal{H}_{\mathbf{f}}^{k}(\mathbf{x})). &
\end{array}$$

□

Set $\mathcal{B}_{\mathbf{f}}(\mathbf{x}) := \mathbf{f}'(\mathbf{x})^* \cdot \mathbf{f}(\mathbf{0})$. With the help of Lemma 7 we may prove the following stronger version of Theorem 4.

Theorem 4 (stronger version).

1. *For all $k > 0 : \boldsymbol{\kappa}^{(k-1)} \leq \boldsymbol{\nu}^{(k)} \leq \boldsymbol{\kappa}^{(k)}$.*
2. *For all $k \geq 0 : \boldsymbol{\kappa}^{(k)} = \mathcal{H}_{\mathbf{f}}^{k}(\mathbf{f}(\mathbf{0}) = \mathcal{N}_{\mathbf{f}}^{k}(\mathbf{f}(\mathbf{0})) = \mathcal{B}_{\mathbf{f}}^{k}(\mathbf{f}(\mathbf{0}))$.*

Proof.

(a) By induction on k. We show the two inequations separately starting with the first one. We have $\boldsymbol{\kappa}^{(0)} = \mathbf{f}(\mathbf{0}) \leq \mathbf{f}'(\mathbf{0})^*\mathbf{f}(\mathbf{0}) = \boldsymbol{\nu}^{(1)}$ and, inductively, $\boldsymbol{\kappa}^{(k)} = \mathbf{f}'(\boldsymbol{\kappa}^{(k-1)})^*\boldsymbol{\kappa}^{(k-1)} \leq \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*\boldsymbol{\nu}^{(k)} \leq \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*\mathbf{f}(\boldsymbol{\nu}^{(k)}) = \boldsymbol{\nu}^{(k+1)}$.
For the second inequation we have $\boldsymbol{\nu}^{(0)} = \mathbf{0} \leq \mathbf{f}(\mathbf{0}) = \boldsymbol{\kappa}^{(0)}$ and, inductively, $\boldsymbol{\nu}^{(k+1)} = \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*\mathbf{f}(\boldsymbol{\nu}^{(k)}) \leq \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\mathbf{f}(\boldsymbol{\kappa}^{(k)}) \leq \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\boldsymbol{\kappa}^{(k)} = \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\boldsymbol{\kappa}^{(k)} = \boldsymbol{\kappa}^{(k+1)}$, where the last inequation follows from Taylor's Theorem: $\mathbf{f}(\boldsymbol{\kappa}^{(k)}) \leq \mathbf{f}(\mathbf{0}) + \mathbf{f}'(\boldsymbol{\kappa}^{(k)})\boldsymbol{\kappa}^{(k)} \leq \boldsymbol{\kappa}^{(k)} + \mathbf{f}'(\boldsymbol{\kappa}^{(k)})\boldsymbol{\kappa}^{(k)} \leq \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\boldsymbol{\kappa}^{(k)}$.

(b) By induction on k. We have obviously

$$\boldsymbol{\kappa}^{(0)} = \mathcal{H}^0_{\mathbf{f}}(\mathbf{f}(\mathbf{0})) = \mathcal{N}^0_{\mathbf{f}}(\mathbf{f}(\mathbf{0})) = \mathcal{B}^0_{\mathbf{f}}(\mathbf{f}(\mathbf{0})) = \mathbf{f}(\mathbf{0}).$$

Let $k \geq 0$. Using Lemma 7 and the induction hypothesis we have

$$\boldsymbol{\kappa}^{(k+1)} = \mathcal{H}^{k+1}_{\mathbf{f}}(\mathbf{f}(\mathbf{0})) = \mathbf{f}'(\mathcal{H}^k_{\mathbf{f}}(\mathbf{x}))^*\mathbf{x} \mid_{\mathbf{x}=\mathbf{f}(\mathbf{0})} = \mathcal{B}_{\mathbf{f}}(\mathcal{H}^k_{\mathbf{f}}(\mathbf{f}(\mathbf{0}))) = \mathcal{B}^{k+1}_{\mathbf{f}}(\mathbf{f}(\mathbf{0}))$$

and

$$\begin{aligned}
\mathcal{N}^{k+1}_{\mathbf{f}}(\mathbf{f}(\mathbf{0})) &= \mathcal{N}_{\mathbf{f}}(\mathcal{N}^k_{\mathbf{f}}(\mathbf{f}(\mathbf{0}))) && \text{(by definition)}\\
&= \mathcal{N}_{\mathbf{f}}(\boldsymbol{\kappa}^{(k)}) && \text{(by induction hyp.)}\\
&= \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\mathbf{f}(\boldsymbol{\kappa}^{(k)}) && \text{(by definition)}\\
&= \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*(\mathbf{f}(\mathbf{0}) + \mathbf{f}'(\boldsymbol{\kappa}^{(k)})\boldsymbol{\kappa}^{(k)}) && \text{(Taylor)}\\
&= \boldsymbol{\kappa}^{(k+1)} + \mathbf{f}'(\boldsymbol{\kappa}^{(k)})^*\mathbf{f}'(\boldsymbol{\kappa}^{(k)})\boldsymbol{\kappa}^{(k)} && \text{(see above)}\\
&\quad \begin{cases} \leq \boldsymbol{\kappa}^{(k+1)} + \mathcal{H}_{\mathbf{f}}(\boldsymbol{\kappa}^{(k)}) \\ \geq \boldsymbol{\kappa}^{(k+1)} \end{cases}\\
&= \boldsymbol{\kappa}^{(k+1)}.
\end{aligned}$$

□

A.6 Proof of Theorem 5

Theorem 5.

1. *The sum of yields of all derivation trees of x_i with height at most h equals $(\mathbf{f}^h(\mathbf{0}))_i$.*
2. *The sum of yields of all derivation trees of x_i equals $(\mu\mathbf{f})_i$.*

Proof. For $h = 0$ this holds by definition as there are no derivation trees of height 0 and $\mathbf{f}^0(\mathbf{0}) = \mathbf{0}$. Let $\mathcal{T}^h_i$ be the set of all derivations of x_i with height at most h. $\mathcal{T}_i$ then denotes the set of all derivation trees of x_i. We then have

$$\begin{aligned}
(\mathbf{f}^{h+1}(\mathbf{0}))_i &= f_i((\sum_{t\in\mathcal{T}^h_j} Y(t))_{1\leq j\leq n})\\
&= \sum_{\boldsymbol{\iota}\in\mathbb{N}^n} (c^{(i)}_{\boldsymbol{\iota}}\mathbf{x}^{\boldsymbol{\iota}} \mid_{\mathbf{x}=(\sum_{t\in\mathcal{T}^h_j} Y(t))_{1\leq j\leq n}}) + c^{(i)}_{\mathbf{0}}\\
&= (\sum_{t\in\mathcal{T}^{h+1}_j} Y(t))_{1\leq j\leq n}.
\end{aligned}$$

One simply has to realise that $c_{\boldsymbol{\iota}}^{(i)}\mathbf{x}^{\boldsymbol{\iota}}$ evaluated at $(\sum_{t\in\mathcal{T}_j^h} Y(t))_{1\leq j\leq n}$ equals the sum of yields of all derivation trees t' of x_i with $p_{t'}(\mathbf{x}) = c_{\boldsymbol{\iota}}^{(i)}\mathbf{x}^{\boldsymbol{\iota}}$ and $h(t') \leq h+1$. A similar argument shows that the vector $\mathbf{x}_0 = (\sum_{t\in\mathcal{T}_j} Y(t))_{1\leq j\leq n}$ is a fixed point of $\mathbf{f}$. As shown in the first part we have $\mathbf{x}_0 = \sup_{h\in\mathbb{N}} \mathbf{f}^h(\mathbf{0})$. With $\mathbf{f}^h(\mathbf{0}) \leq \mu\mathbf{f}$ we then may conclude $\mu\mathbf{f} = \mathbf{x}_0$. □

A.7 Proof of Lemma 3

Lemma 3. *For each derivation tree t of x_i there is a compact derivation tree t' of x_i with equal yield.*

Proof. First, let us introduce the following notation: we write $t = t_1 \cdot t_2$ to denote that the root of t_2 has the same label as some leaf l of t_1, and that t can be obtained by identifying l with t_2.

We proceed by induction on the dimension followed by an induction on the number of nodes. If $\dim(t) = 0$, t is compact and we are done. Assume $\dim(t) > V(t)$. Let $s_1, s_2, \ldots, s_r$ be the children of t with $\dim(s_1) \geq \dim(s_2) \geq \ldots \geq \dim(s_r)$. By induction on the number of nodes, we may assume that every child is compact, i.e. $\dim(s_i) \leq V(s_i)$. We then have by definition of dimension

$$V(t) + 1 \leq \dim(t) \leq \dim(s_1) + 1 \leq V(s_1) + 1 \leq V(t) + 1.$$

Hence, we have $\dim(t) = \dim(s_1) + 1$ which by definition implies $\dim(s_1) = \dim(s_2) = V(t) = V(s_1) = V(s_2)$. As $h(s_2) > \dim(s_2)$, we find a chain in s_2 which passes at least two inner nodes both labelled by, say, x_j. This means, we may factor s_2 into $t_1^b \cdot (t_2^b \cdot t_3^b)$. We also have $V(s_1) = V(s_2)$, thus we find an inner node of s_1 labelled by x_j which allows us to write $s_1 = t_1^a \cdot t_2^a$. We define t' by reallocating t_2^b to s_1, i.e., by induction on the number of nodes, we find a compact derivation tree s_1', resp. s_2', equivalent to $t_1^a \cdot (t_2^b \cdot t_3^a)$, resp. $t_1^b \cdot t_3^b$, and substitute this one for s_1, resp. s_2. We then have $V(s_2') \leq V(s_2) = V(s_1) = V(s_1')$ and, by compactness of s_1' and s_2', $\dim(s_1') \leq \dim(s_1) = \dim(s_2) \geq \dim(s_2')$.

We iterate this procedure until the dimension of t drops by one, i.e. t becomes compact. This has to happen as the number of nodes of (the current) s_2 strictly decreases in every iteration, and the number of nodes is, obviously, an upper bound for the height, and, therefore, for the dimension of a tree. □

A.8 Proof of Lemma 4

Lemma 4. *Let t be a derivation tree of x_i with $\dim(t) \leq k$. Then $Y(t) \leq (\boldsymbol{\kappa}^{(k)})_i$.*

Proof. We proceed by induction on the height. If $h(t) = 1$, we have obviously $Y(t) = f_i(\mathbf{0}) \leq (\boldsymbol{\kappa}^{(k)})_i$.

Hence, assume $h(t) > 1$, so $\dim(t) > 0$. Let $s_1, \ldots, s_r$ be the children of t being derivation trees themselves with yields $Y_l = Y(s_l)$. Again we assume $\dim(s_1) \geq \dim(s_2) \geq \ldots \geq \dim(s_s)$. We note that we either have $\dim(t) >$

$\dim(s_1) \geq \dim(s_2)$ or $\dim(t) = \dim(s_1) > \dim(s_2)$, i.e. $\dim(s_2) \leq k-1$. We have $p_t(\mathbf{x}) = c_{\boldsymbol{\iota}}^{(i)} \mathbf{x}^{\boldsymbol{\iota}}$ for some $\boldsymbol{\iota} \in \mathbb{N}^n$ by definition of derivation tree. Then $Y(t) = c_{\boldsymbol{\iota}}^{(i)} Y_1 \cdot \ldots \cdot Y_s$. Let x_j be the label of s_1. By induction hypothesis, $Y_1 \leq (\boldsymbol{\kappa}^{(k)})_j$, and $Y_2, \ldots, Y_s$ are subsumed in their respective components of $\boldsymbol{\kappa}^{(k-1)}$. As $\frac{\partial p_t}{\partial x_j}$ results from $p_t(\mathbf{x})$ by removing exactly one occurrence of x_j (the "occurrence" of c_1), we have $c_{\boldsymbol{\iota}}^{(i)} Y_2 \cdot \ldots \cdot Y_r \leq \frac{\partial p_t}{\partial x_j}(\boldsymbol{\kappa}^{(k-1)})$. Put together: $Y(t) = c_{\boldsymbol{\iota}}^{(i)} Y_1 \cdot \ldots \cdot Y_r \leq \frac{\partial p_t}{\partial x_j}(\boldsymbol{\kappa}^{(k-1)}) \cdot Y_1 \leq \frac{\partial f_i}{\partial x_j}(\boldsymbol{\kappa}^{(k-1)}) \cdot (\boldsymbol{\kappa}^{(k)})_j \leq (\boldsymbol{\kappa}^{(k)})_i$. □

A.9 Proof of Theorem 8

Theorem 8. *For all* $i \geq 1 : \mathcal{H}_i(\mathbf{x}) = \mathcal{H}^i(\mathbf{x})$.

For the proof of Theorem 8 we use tensor algebra to calculate with vectors of polynomials and their derivatives. All the tensors we are dealing with have exactly one *contravariant* index. In particular, vectors of polynomials (such as $\mathbf{f}$) are rank-1 tensors $\mathbf{T}^\alpha$. For ease of notation we often omit indices that do not change. In addition, we use the Einstein summation convention, i.e., indices appearing twice in a single term, once contravariant and once covariant, are implicitly summed over. This allows to identify a tensor with its entries. Additional covariant indices are introduced when taking a derivative of a tensor $\mathbf{T}$: $\mathbf{T}' = \mathbf{T}'_\alpha = \frac{\partial}{\partial x_\alpha}\mathbf{T}$.

A general tensor-tensor multiplication is written and defined as $\mathbf{S}_\alpha \mathbf{T}^\alpha$ where the resulting tensor has the indices of $\mathbf{S}$ and $\mathbf{T}$ except for the repeated index α. Using tensor algebra and the definition of derivatives (Section 3) we can state and prove the following derivation rules:

Lemma 8 (derivation rules).

(a) product rule: *Let* $\mathbf{S}, \mathbf{T}$ *be tensors. Then* $(\mathbf{S}_\alpha \mathbf{T}^\alpha)' = \mathbf{S}'_{\alpha\beta}\mathbf{T}^\alpha + \mathbf{S}_\alpha \mathbf{T}'^\alpha_\beta$.
(b) chain rule: *Let* $\mathbf{V} = \mathbf{V}^\alpha$ *be a vector. Then* $(\mathbf{T}(\mathbf{V}))' = (\mathbf{T}'(\mathbf{V}))_\alpha \mathbf{V}'^\alpha_\beta$.
(c) matrix star rule: *Let* $\mathbf{M}$ *be a matrix. Then* $(\mathbf{M}^{*\alpha}_\beta)' = \mathbf{M}^{*\alpha}_\gamma \mathbf{M}'^\gamma_{\delta\eta} \mathbf{M}^{*\delta}_\beta$, *where* $\mathbf{M}'^\gamma_{\delta\eta} = \frac{\partial}{\partial x_\eta}(\mathbf{M}^\gamma_\delta)$.

Proof. The proofs of parts (a) and (b) are standard. We prove only part (c), since it requires more care. The entries of $\mathbf{M}$ are regular expressions over $\Sigma \cup \mathcal{X}$. So are the entries of $\mathbf{M}^*$ (following the construction in [6]). In order to show $\frac{\partial}{\partial x_\eta}(\mathbf{M}^{*\alpha}_\beta) = \mathbf{M}^{*\alpha}_\gamma \mathbf{M}'^\gamma_{\delta\eta} \mathbf{M}^{*\delta}_\beta$, we exploit again Redko's theorem, i.e., it suffices to show $\mathsf{c}\left(\frac{\partial}{\partial x_\eta}(\mathbf{M}^{*\alpha}_\beta)\right) = \mathsf{c}\left(\mathbf{M}^{*\alpha}_\gamma \mathbf{M}'^\gamma_{\delta\eta} \mathbf{M}^{*\delta}_\beta\right)$, where c denotes the canonical commutative interpretation, i.e., the Parikh images.

For convenience of notation we identify in the following a Parikh vector with any word having this Parikh vector. Let $w' \in \mathsf{c}\left(\frac{\partial}{\partial x_\eta}(\mathbf{M}^{*\alpha}_\beta)\right)$. By definition of derivatives, w' can be obtained from some $w \in \mathsf{c}\left(\mathbf{M}^{*\alpha}_\beta\right)$ by deleting one occurrence of x_η. Say, $w = w^\alpha_{\theta_1} w^{\theta_1}_{\theta_2} \cdots w^{\theta_{N-2}}_{\theta_{N-1}} w^{\theta_{N-1}}_\beta$ s.t. $w^x_y \in \mathsf{c}\left(\mathbf{M}^x_y\right)$, and say, $w' = w^\alpha_{\theta_1} w^{\theta_1}_{\theta_2} \cdots w^{\theta_{K-2}}_{\theta_{K-1}} u'^{\theta_{K-1}}_{\theta_K} w^{\theta_K}_{\theta_{K+1}} \cdots w^{\theta_{N-2}}_{\theta_{N-1}} w^{\theta_{N-1}}_\beta$, where $u'^{\theta_{K-1}}_{\theta_K}$ be obtained

from $w^{\theta_{K-1}}_{\theta_K}$ (for a $K \in \{1,\dots,N\}$) by deleting one occurrence of x_η. Then we have $u'^{\theta_{K-1}}_{\theta_K} \in \mathsf{c}\left(\mathbf{M}'^{\gamma}_{\delta\eta}\right)$ and thus $w' \in \mathsf{c}\left(\mathbf{M}^{*\alpha}_{\ \gamma}\mathbf{M}'^{\gamma}_{\delta\eta}\mathbf{M}^{*\delta}_{\ \beta}\right)$.

The other direction ($\mathbf{M}^{*\alpha}_{\ \gamma}\mathbf{M}'^{\gamma}_{\delta\eta}\mathbf{M}^{*\delta}_{\ \beta} \leq \frac{\partial}{\partial x_\eta}(\mathbf{M}^{*\alpha}_{\ \beta})$) is shown similarly. □

The following lemma has already been implicitly used in Section 5 for vectors and holds for arbitrary tensors.

Lemma 9 (Taylor's theorem in commutative Kleene algebra, cf. [4]). *Let $\mathbf{V} = \mathbf{V}^\alpha$ be a vector. Then $\mathbf{T}(\mathbf{V}) = \mathbf{T}(\mathbf{0}) + (\mathbf{T}'(\mathbf{V}))_\alpha \mathbf{V}^\alpha$.*

Proof. This is proved in [4] for vectors $\mathbf{T}$. It is easy to see that it holds for general tensors as well. □

In the following, we mean $\mathcal{H}(\mathbf{x})$ when the argument is omitted, and similarly for other operators. Theorem 8 then states $\mathcal{H}_i = \mathcal{H}^i$. To prove it, we rely on Lemmata 8 and 9 and on the fact that the set of matrices of a fixed dimension over a Kleene algebra forms again a (non-commutative) Kleene algebra [6]. We prove the inequations $\mathcal{H}_i \geq \mathcal{H}^i$ and $\mathcal{H}_i \leq \mathcal{H}^i$ of Theorem 8 in turn. For both of them we use the following equation.

Lemma 10. *For $i \geq 1$: $\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i) = \mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}''\{\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\} + \mathbf{f}'(\mathcal{H}^{i-1})^*$, where $\mathbf{f}''\{\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\}$ denotes the matrix $\mathbf{M}^\alpha_\beta = \mathbf{f}''^\alpha_{\delta\eta}(\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}))^\delta_\beta(\mathcal{H}^i)^\eta$.*

Proof.

$$\begin{aligned}
\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i) &= \tfrac{\partial}{\partial \mathbf{x}}(\mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{x}) && \text{(Lemma 7)}\\
&= \tfrac{\partial}{\partial \mathbf{x}}(\mathbf{f}'(\mathcal{H}^{i-1})^*)\mathbf{x} + \mathbf{f}'(\mathcal{H}^{i-1})^* && \text{(Lemma 8 a)}\\
&= \mathbf{f}'(\mathcal{H}^{i-1})^*\tfrac{\partial}{\partial \mathbf{x}}(\mathbf{f}'(\mathcal{H}^{i-1}))\mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{x} + \mathbf{f}'(\mathcal{H}^{i-1})^* && \text{(Lemma 8 c)}\\
&= \mathbf{f}'(\mathcal{H}^{i-1})^*\tfrac{\partial}{\partial \mathbf{x}}(\mathbf{f}'(\mathcal{H}^{i-1}))\mathcal{H}^i + \mathbf{f}'(\mathcal{H}^{i-1})^* && \text{(Lemma 7)}\\
&= \mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}''\{\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\} + \mathbf{f}'(\mathcal{H}^{i-1})^* && \text{(Lemma 8 b) } \square
\end{aligned}$$

Lemma 11. *For $i \geq 1 : \mathcal{H}_i \geq \mathcal{H}^i$.*

Proof. By induction on i. The base case $i = 1$ is trivial. Let $i \geq 1$.

$$\begin{aligned}
\mathcal{H}_{i+1} &= \left[\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}_i)\right]^*\mathbf{x} && \text{(by definition)}\\
&\geq \left[\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i)\right]^*\mathbf{x} && \text{(ind. hyp.)}\\
&= \left[\mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}''\{\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\} + \mathbf{f}'(\mathcal{H}^{i-1})^*\right]^*\mathbf{x} && \text{(Lemma 10)}\\
&= \left[\mathbf{f}'(\mathcal{H}^{i-1}) + \mathbf{f}''\{\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\}\right]^*\mathbf{x} && \text{(Kleene algebra)}\\
&\geq \left[\mathbf{f}'(\mathbf{0}) + \mathbf{f}''\{\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\}\right]^*\mathbf{x} && \text{(Kleene algebra)}\\
&\geq \left[\mathbf{f}'(\mathbf{0}) + \mathbf{f}''\mathcal{H}^i\right]^*\mathbf{x} && (\mathcal{H}^{i-1} \geq \mathbf{x})\\
&= \left[\mathbf{f}'(\mathcal{H}^i)\right]^*\mathbf{x} && \text{(Lemma 9)}\\
&= \mathcal{H}^{i+1} && \text{(Lemma 7)} \quad \square
\end{aligned}$$

For the other direction (see Lemma 13 below) the following lemma is used.

Lemma 12. *For $i < k : (\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i))^*\mathcal{H}^k \leq \mathcal{H}^k$.*

Proof. By the axioms of Kleene algebra [6] it suffices to show $\frac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i)\mathcal{H}^k \leq \mathcal{H}^k$. This is proved by induction on i. The base case $i = 0$ is trivial, because $\mathcal{H}^0 = \mathbf{x}$. Let $1 \leq i < k$.

$$\begin{aligned}
\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i)\mathcal{H}^k &= (\mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}''\{\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1}), \mathcal{H}^i\} + \mathbf{f}'(\mathcal{H}^{i-1})^*)\mathcal{H}^k && \text{(Lemma 10)}\\
&= \mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}''\{\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^{i-1})\mathcal{H}^k, \mathcal{H}^i\} + \mathbf{f}'(\mathcal{H}^{i-1})^*\mathcal{H}^k && \text{(distributivity)}\\
&\leq \mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}''\{\mathcal{H}^i, \mathcal{H}^k\} + \mathbf{f}'(\mathcal{H}^{i-1})^*\mathcal{H}^k && \text{(ind. hyp.)}\\
&\leq \mathbf{f}'(\mathcal{H}^{i-1})^*\mathbf{f}'(\mathcal{H}^i)\mathcal{H}^k + \mathbf{f}'(\mathcal{H}^{i-1})^*\mathcal{H}^k && \text{(Lemma 9)}\\
&\leq \mathbf{f}'(\mathcal{H}^{k-1})^*\mathcal{H}^k && (\mathcal{H}^{i-1} \leq \mathcal{H}^i)\\
&= \mathcal{H}^k && \text{(Lemma 7) } \square
\end{aligned}$$

Lemma 13. *For $i \geq 1 : \mathcal{H}_i \leq \mathcal{H}^i$.*

Proof. By induction on i. The base case $i = 1$ is trivial. Let $i \geq 1$.

$$\begin{aligned}
\mathcal{H}_{i+1} &= \left(\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}_i)\right)^*\mathbf{x} && \text{(by definition)}\\
&\leq \left(\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i)\right)^*\mathbf{x} && \text{(ind. hyp.)}\\
&\leq \left(\tfrac{\partial}{\partial \mathbf{x}}(\mathcal{H}^i)\right)^*\mathcal{H}^{i+1} && (\mathbf{x} \leq \mathcal{H}^{i+1})\\
&\leq \mathcal{H}^{i+1} && \text{(Lemma 12) } \square
\end{aligned}$$

This finishes the proof of Theorem 8.

B Redko's Theorem and Commutative Kleene Algebras

There is a number of inequivalent definitions of Kleene algebras. This includes *C-algebras* and *Kleene algebras in the sense of Kozen* the latter of which we simply refer to as *Kleene algebras*.

Both definitions require an algebraic structure $(K, +, \cdot, {}^*, 0, 1)$ that is an idempotent semiring under $+, \cdot, 0, 1$. In addition, different sets of axioms are required.

A C-algebra [1] must satisfy the following axioms:

$$\begin{array}{ll}
C11 & (a+b)^* = (a^*b)^*a^*\\
C12 & (ab)^* = 1 + a(ba)^*b\\
C13 & (a^*)^* = a^*\\
C14.n & a^* = a^{n*}a^{<n} \quad (n > 0).
\end{array}$$

A Kleene algebra [6] on the other hand must satisfy the following axioms:

$$\begin{array}{lll}
K1 & 1 + aa^* \leq a^* & \\
K2 & 1 + a^*a \leq a^* & \\
K3 & a + bc \leq c \quad \rightarrow & b^*a \leq c\\
K4 & a + cb \leq c \quad \rightarrow & ab^* \leq c,
\end{array}$$

where $\leq$ refers to the natural partial order on K.

It was shown by Kozen [6] that the axioms of Kleene algebra are *complete* for the algebra of regular languages. That means, if an equation $\alpha = \beta$ between

regular expressions holds under the canonical interpretation over the regular languages, then it holds in any Kleene algebra. It is easy to see that equations $C11 - C14$ hold under the canonical interpretation. Therefore any Kleene algebra is a C-algebra.

The axioms of C-algebra are not complete, i.e., they are too weak to derive some equation valid under the canonical interpretation [1]. However, if two more axioms (C^+1 and C^+2, see below) describing commutativity are added, the resulting system of axioms (defining *commutative C-algebras*) becomes complete for the algebra of commutative regular languages. In other words, if the Parikh images of languages $L(\alpha)$ and $L(\beta)$ are equal, then $\alpha = \beta$ can be proved using only the axioms of commutative C-algebras. The additional axioms are:

$$\begin{array}{ll} C^+1 & ab = ba \\ C^+2 & a^*b^* = (ab)^*(a^* + b^*). \end{array}$$

The completeness of commutative C-algebras is called *Redko's theorem*. Conway's monograph [1] contains a proof of this theorem.

We want to show that the system of axioms of Kleene algebra plus the commutativity axiom $ab = ba$ (defining *commutative Kleene algebras*) is complete for commutative regular languages as well. Appealing to Redko's theorem, we only have to show that equation C^+2 is a theorem of commutative Kleene algebra.

We use the identity $a^*b^* = (a+b)^*$ which is a theorem of commutative Kleene algebra [4]. Since $(a + b)^* \geq (ab)^*(a^* + b^*)$ holds in any Kleene algebra, we only need to show $(a + b)^* \leq (ab)^*(a^* + b^*)$. With $K3$ it suffices to show

$$1 + (a + b)(ab)^*(a^* + b^*) \leq (ab)^*(a^* + b^*).$$

We show this inequality for each term of the sum at the left hand side. For 1 it obviously holds. We also have $a(ab)^*a^* = (ab)^*aa^* \leq (ab)^*a^*$ using commutativity and $K1$. Similarly, $a(ab)^*b^* = (ab)^*ab^* = (ab)^*a+(ab)^*abb^* \leq (ab)^*a+(ab)^*b^* \leq (ab)^*(a^* + b^*)$. Here we used that $b^* = 1 + bb^*$ is a theorem of Kleene algebra. The other inequalities follow symmetrically.