Solving fixed-point equations over semirings

Javier Esparza

Technische Universität München

Joint work with

Michael Luttenberger and Maximilian Schlund

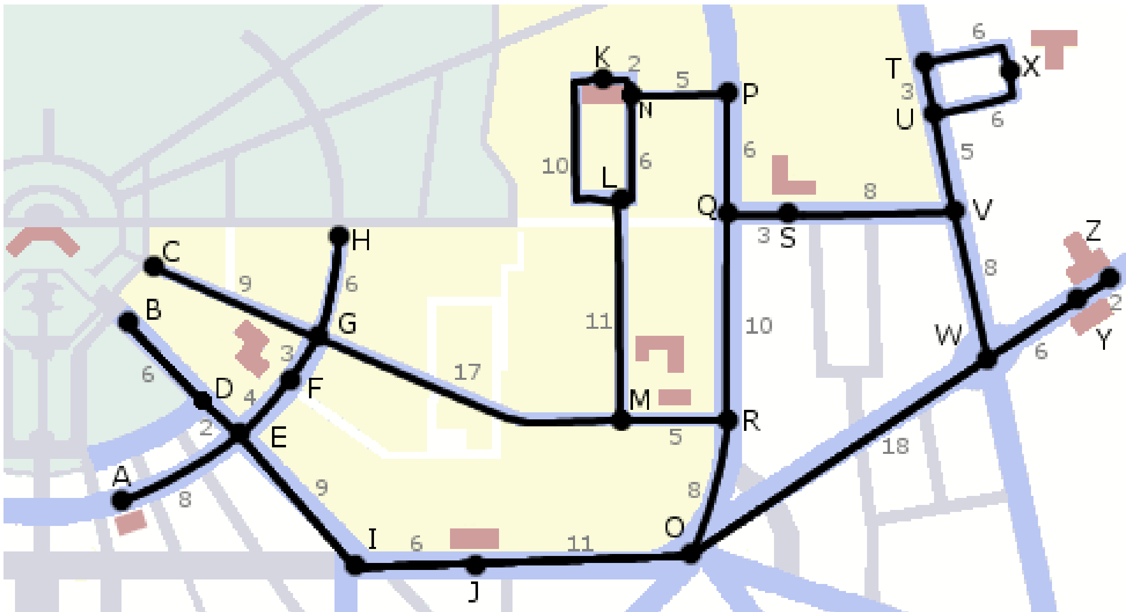
Fixed-point equations

We study systems of equations of the form

$$\begin{aligned} X_1 &= f_1(X_1, \ldots, X_n) \\ X_2 &= f_2(X_1, \ldots, X_n) \\ &\cdots \\ X_n &= f_n(X_1, \ldots, X_n) \end{aligned}$$

where the f_i's are "polynomial expressions".

Shortest paths



Lengths d_i of shortest paths from vertex 0 to vertex i in graph $G = (V, E)$ are the largest solution of

$$d_i = \min_{(i,j)\in E}(d_i, d_j + w_{ji})$$

where w_{ij} is the distance from i to j.

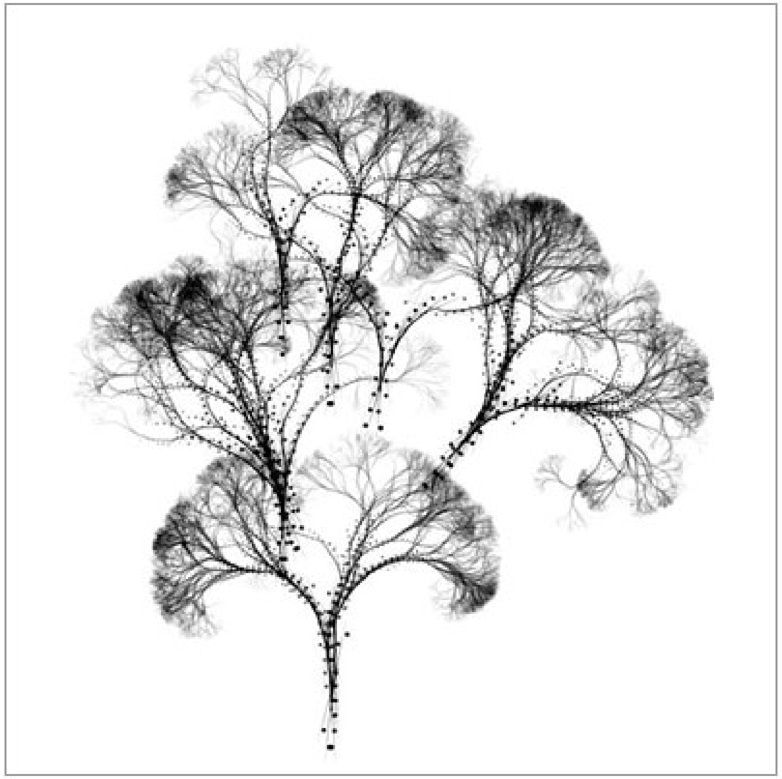
Context-free languages

Context-free grammar

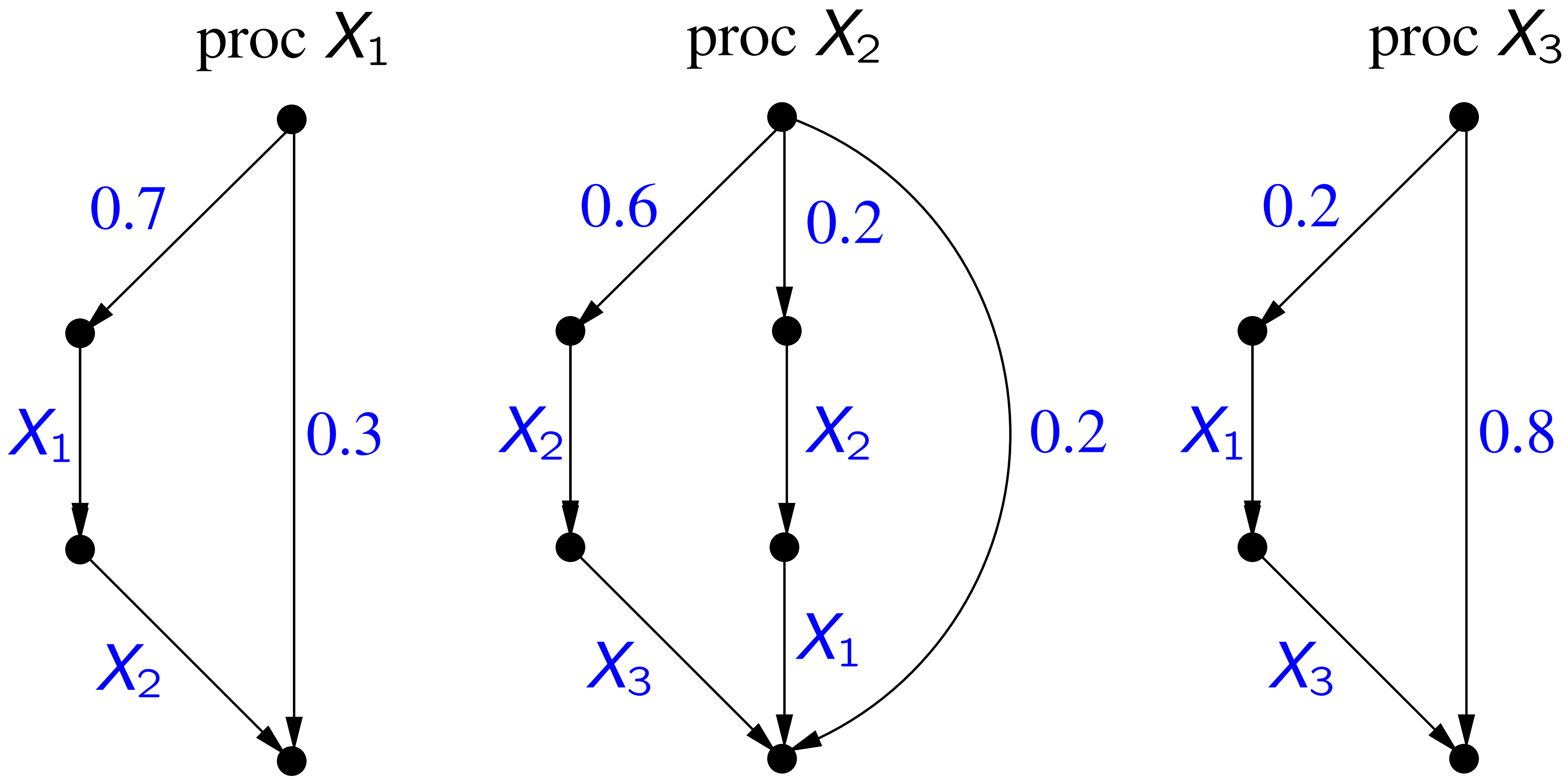
$$
\begin{aligned}
X &\rightarrow ZX \mid Z \\
Y &\rightarrow aYa \mid ZX \\
Z &\rightarrow b \mid aYa
\end{aligned}
$$

Languages generated from X, Y, Z are the least solution of

$$
\begin{aligned}
L_X &= (L_Z \cdot L_X) \cup L_Z \\
L_Y &= (\{a\} \cdot L_Y \cdot \{a\}) \cup (L_Z \cdot L_X) \\
L_Z &= \{b\} \cup (\{a\} \cdot L_Y \cdot \{a\})
\end{aligned}
$$



Probability of program termination



The probability that X_i terminates is the least solution of

$$\begin{aligned}
X_1 &= 0.7 \cdot X_1 \cdot X_2 + 0.3 \\
X_2 &= 0.6 \cdot X_2 \cdot X_3 + 0.2 \cdot X_2 \cdot X_1 + 0.2 \\
X_3 &= 0.2 \cdot X_1 \cdot X_3 + 0.8
\end{aligned}$$

Algorithms

Many specific algorithms for different cases:

Shortest paths: Dijkstra, Bellman-Ford, Floyd-warshall.

Right-linear grammars: Gauss elimination.

Probability of termination: Newton's method.

What do these problems have in common?

Underlying structure: ω-continuous semirings

Semiring $(C, +, \times, 0, 1)$:

$(C, +, 0)$ is a commutative monoid $\quad$ $\times$ distributes over $+$

$(C, \times, 1)$ is a monoid $\quad$ $0 \times a = a \times 0 = 0$

ω-continuity:

the relation $a \sqsubseteq b \Leftrightarrow \exists c : \; a + c = b$ is a partial order

$\sqsubseteq$-chains have limits

Theorem [Knaster-tarski]: A system of fixed-point equations over an ω-continuous semiring has a unique least solution (and an unique largest solution) w.r.t. $\sqsubseteq$.

In the rest of the talk: semiring $\equiv$ ω-continuous semiring.

Research program

Develop and implement generic solution or approximation methods valid for all semirings, or at least for large classes.

- Theoretical motivation: Exchange of algorithms and proof techniques between numerical mathematics, algebraic computation and language theory.
- Applications that require to solve the same system over many different semirings:
  - Authorization systems
  - Recommendation systems
  - Provenance computations in databases

A system for academic recommendations

Participants: researchers, universities, departments, conferences, papers

Relations: researcher-of, professor-at, student-of, author-of, . . .

- Notation: $p.r$
- Meaning: group of participants that are in relation r with p.

Particpants express group membership by adding rules or certificates to the system

Giessen.professor $\rightarrow$ Holzer

Giessen.researcher $\rightarrow$ Giessen.professor

Giessen.researcher $\rightarrow$ Giessen.researcher.Phd-student

Holzer.Phd-student $\rightarrow$ Jakobi

A system for academic recommendations

Membership explicitely determined by prefix-rewriting derivations

To find out that Jakbi is a researcher at Giessen:

Giessen.researcher

A system for academic recommendations

Membership explicitely determined by prefix-rewriting derivations

To find out that Jakbi is a researcher at Giessen:

$$\text{Giessen.researcher} \rightarrow \text{Giessen.researcher.Phd-student}$$

A system for academic recommendations

Membership explicitely determined by prefix-rewriting derivations

To find out that Jakbi is a researcher at Giessen:

$$\begin{aligned}\text{Giessen.researcher} &\to \text{Giessen.researcher.Phd-student}\\ &\to \text{Giessen.professor.Phd-student}\end{aligned}$$

A system for academic recommendations

Membership explicitely determined by prefix-rewriting derivations

To find out that Jakbi is a researcher at Giessen:

$$\begin{aligned}\text{Giessen.researcher} &\rightarrow \text{Giessen.researcher.Phd-student}\\ &\rightarrow \text{Giessen.professor.Phd-student}\\ &\rightarrow \text{Holzer.Phd-student}\end{aligned}$$

A system for academic recommendations

Membership explicitely determined by prefix-rewriting derivations

To find out that Jakbi is a researcher at Giessen:

$$\begin{aligned}\text{Giessen.researcher} &\rightarrow \text{Giessen.researcher.Phd-student}\\ &\rightarrow \text{Giessen.professor.Phd-student}\\ &\rightarrow \text{Holzer.Phd-student}\\ &\rightarrow \text{Jakobi}\end{aligned}$$

A system for academic recommendations

Group membership qualified by weights

$$\text{CIAA.author} \xrightarrow{11/2400} \text{Holzer}$$

$$\text{CIAA.author} \xrightarrow{1/2400} \text{Jakobi}$$

A system for academic recommendations

Group membership qualified by weights

$$\text{CIAA.author} \xrightarrow{11/2400} \text{Holzer}$$

$$\text{CIAA.author} \xrightarrow{1/2400} \text{Jakobi}$$

$$\text{Holzer.co-author} \xrightarrow{15/175} \text{Jakobi}$$

$$\text{Jakobi.co-author} \xrightarrow{15/16} \text{Holzer}$$

A system for academic recommendations

Group membership qualified by weights

$$\text{CIAA.author} \xrightarrow{11/2400} \text{Holzer}$$

$$\text{CIAA.author} \xrightarrow{1/2400} \text{Jakobi}$$

$$\text{Holzer.co-author} \xrightarrow{15/175} \text{Jakobi}$$

$$\text{Jakobi.co-author} \xrightarrow{15/16} \text{Holzer}$$

Recursive group definitions with damping weights.

$$\text{Holzer.community} \xrightarrow{1} \text{Holzer.co-author}$$

$$\text{Holzer.community} \xrightarrow{0.5} \text{Holzer.community.co-author}$$

A system for academic recommendations

Recommendations expressed and qualified in the same way

$$\text{Holzer} \xrightarrow{10} \text{Jakobi}$$

$$\text{Holzer} \xrightarrow{8} \text{Holzer.co-author}$$

$$\text{Holzer} \xrightarrow{6} \text{CIAA.author}$$

Quantitative prefix-rewriting derivations

$$\text{Holzer} \xrightarrow{10} \text{Jakobi}$$

$$\text{Holzer} \xrightarrow{8} \text{Holzer.co-author} \xrightarrow{15/175} \text{Jakobi}$$

$$\text{Holzer} \xrightarrow{6} \text{CIAA-author} \xrightarrow{1/2400} \text{Jakobi}$$

Questions: Weight of a recommendation path?

Aggregate weight of different paths?

A system for academic reputation

'Agnostic" solution: introduce two operations $\otimes$ and $\oplus$

$$\text{Holzer} \xrightarrow{10} \text{Jakobi}$$

$$\text{Holzer} \xrightarrow{8} \text{Holzer.co-author} \xrightarrow{15/175} \text{Jakobi}$$

$$\text{Holzer} \xrightarrow{10 \oplus (0.8 \otimes 1/2400)} \text{Jakobi}$$

We only require: the operations must satisfy the semiring axioms.

Semantics

A set of rules and recommendations is equivalent to a weighted pushdown system.

Participants $\approx$ Control states

Relations $\approx$ Stack alphabet

Weighted rules and recommendations $\approx$ Weighted transition rules

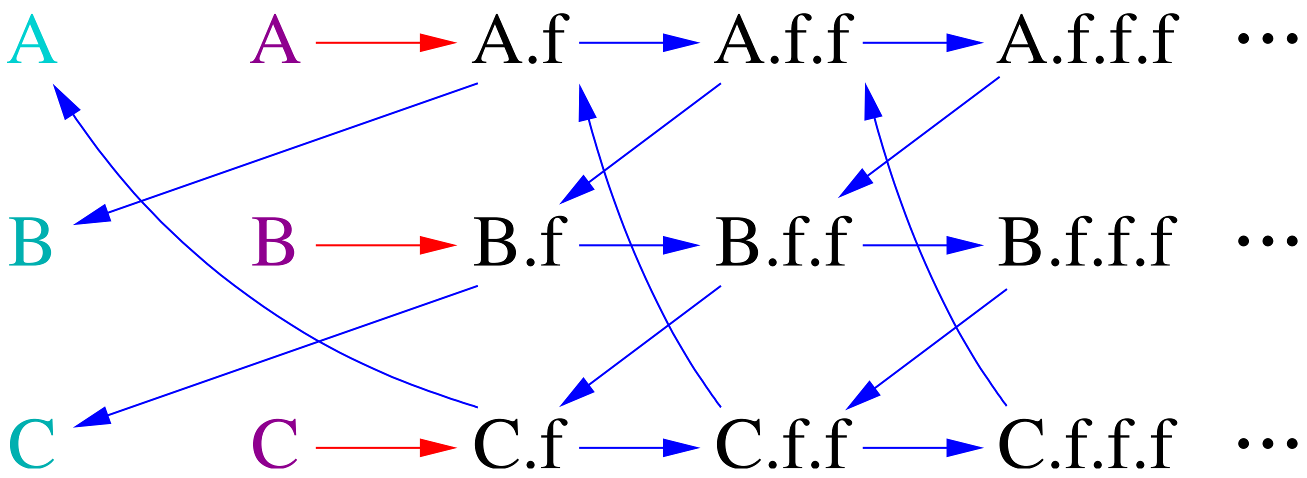
Problem: the weighted transition system associated to the automaton can be infinite.

An example

$\text{Alice.frs} \xrightarrow{0.7} \text{Bob}$ $\text{Alice.frs} \xrightarrow{0.3} \text{Alice.frs.frs}$ $\text{Alice} \xrightarrow{1} \text{Alice.frs}$

$\text{Bob.frs} \xrightarrow{0.9} \text{Charlie}$ $\text{Bob.frs} \xrightarrow{0.1} \text{Bob.frs.frs}$ $\text{Bob} \xrightarrow{1} \text{Bob.frs}$

$\text{Charlie.frs} \xrightarrow{0.5} \text{Alice}$ $\text{Charlie.frs} \xrightarrow{0.5} \text{Charlie.frs.frs}$ $\text{Charlie} \xrightarrow{1} \text{Charlie.frs}$



Alice's trust in Bob: total weight of the paths leading from A to B.

Equations

Define $[pXq]$ as the total weight of all paths from the set pX to q.

Theorem: The $[pXq]$'s are the least solution of the following system of equations:

$$\langle pXq \rangle = \bigoplus_{pX \xrightarrow{w} q} w \;\oplus \bigoplus_{pX \xrightarrow{w} rYZ} w \odot \bigoplus_{s \in P} \langle rYs \rangle \odot \langle sZq \rangle$$

where P is the set of participants.

The total weight of the paths from p to q is then given by $\bigoplus_X [pXq]$.

FPsolve: a generic solver

THE generic solution method: Kleene iteration

Theorem [Klee 38, Tars 55, Kui 97]: The least solution of a system f of fixed-point equations is the supremum of the Kleene approximants, denoted by $\{k_i\}_{i \geq 0}$, and given by

$$\begin{aligned} k_0 &= f(0) \\ k_{i+1} &= f(k_i) . \end{aligned}$$

Basic algorithm for calculation of μf : compute k_0, k_1, k_2, . . . until either $k_i = k_{i+1}$ or the approximation is considered adequate.

Implementation in FPsolve

Abstract base class `Semiring`

```
ViterbiSemiring operator *= (const ViterbiSemiring& elem){
// multiplication: times
value_ *= elem.value_;
return *this;
}

ViterbiSemiring operator += (const ViterbiSemiring& elem){
// addition: max
if (elem.value_ > value_)
value_ = elem.value_;
return *this;
}
```

Kleene iteration may be slow

Set interpretations: Kleene iteration never terminates if μf is an infinite set.

- $X = \{a\} \cdot X \cup \{b\} \qquad \mu f = a^* b$

  Kleene approximants are finite sets: $k_i = (\epsilon + a + \ldots + a^i)b$

Real semiring: convergence can be very slow.

- $X = 0.5\, X^2 + 0.5 \qquad \mu f = 1 = 0.99999\ldots$

  "Logarithmic convergence": k iterations give $O(\log k)$ correct digits.

$$k_n \leq 1 - \frac{1}{n+1} \qquad k_{2000} = 0.9990$$

Language-theoretic characterization of μf

An equation $X = f(X)$ over a semiring induces a context-free grammar G and a valuation V

Language-theoretic characterization of μf

An equation $X = f(X)$ over a semiring induces a context-free grammar G and a valuation V

Example: $X = 0.25X^2 + 0.25X + 0.5$

Grammar: $X \rightarrow aXX \mid bX \mid c$

Valuation: $V(a) = 0.25, V(b) = 0.25, V(c) = 0.5$

Language-theoretic characterization of μf

An equation $X = f(X)$ over a semiring induces a context-free grammar G and a valuation V

Example: $X = 0.25X^2 + 0.25X + 0.5$

Grammar: $X \rightarrow a X X \mid b X \mid c$

Valuation: $V(a) = 0.25, V(b) = 0.25, V(c) = 0.5$

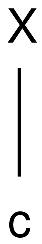
V extends to derivation trees and sets of derivation trees:

$$V(t) := \text{ordered product of the leaves of } t$$

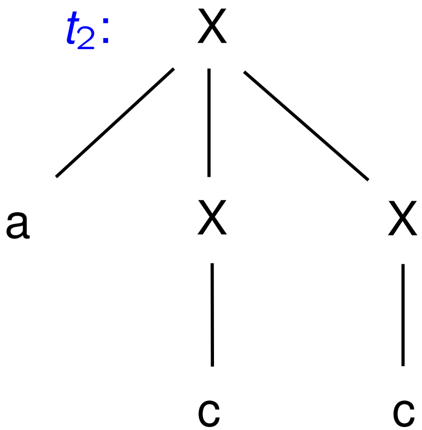
$$V(T) := \sum_{t \in T} V(t)$$

$X \rightarrow aXX \mid bX \mid c$ $\qquad V(a) = V(b) = 0.25, V(c) = 0.5$

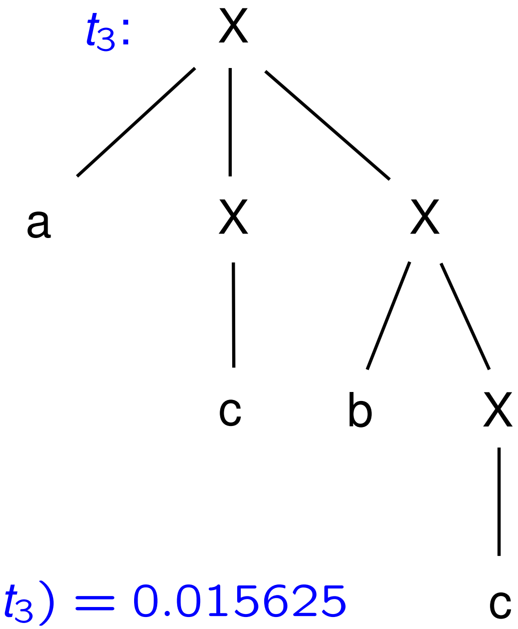
t_1:



t_2:



t_3:



$V(t_1) = 0.5$ $\qquad V(t_2) = 0.25 \cdot 0.5 \cdot 0.5 = 0.0625$ $\qquad V(t_3) = 0.015625$

$$V(\{t_1, t_2, t_3\}) = 0.5 + 0.0625 + 0.015625 = 0.578125$$

Language-theoretic characterization of μf

Fundamental Theorem [Boz99,EKL10]: Let G be the grammar for $X = f(X)$, and let $T(G)$ be the set of derivation trees of G . Then $\mu f = V(T(G)) \stackrel{def}{=} V(G)$

$$\begin{tikzcd}
X = f(X) \arrow[r] \arrow[d] & G \arrow[d] \\
\mu f = V(T(G)) & T(G) \arrow[l, "V"']
\end{tikzcd}$$

Approximating grammars

Let G be the grammar for $X = f(X)$.

An unfolding of G is a sequence $U^1, U^2, U^3, \ldots$ of grammars such that, $T(U^1), T(U^2), T(U^3)$ is a partition of $T(G)$.

Formally: the $T(U^i)$ are pairwise disjoint, and there is a yield-preserving bijection between $\bigcup_{i=1}^{\infty} T(U^i)$ and $T(G)$.

From $U^1, U^2, U^3, \ldots$ we get $G^1, G^2, G^3, \ldots$ such that $T(G^j) = \bigcup_{i=1}^{j} T(U^i)$.

μf is then the supremum of the sequence $V(G^1), V(G^2), V(G^3) \ldots$.

Approximating grammars by height

Goal: U^i (G^i) contain the derivation trees of G of height i (at most i).

$G: \; X \; \rightarrow \; a\,X\,X \;\mid\; b\,X \;\mid\; c$

.

Approximating grammars by height

Goal: U^i (G^i) contain the derivation trees of G of height i (at most i).

$$G: \ X \rightarrow aXX \mid bX \mid c$$

$$X^{\langle 1 \rangle} \rightarrow c$$

$$X^{[1]} \rightarrow X^{\langle 1 \rangle}$$

.

Approximating grammars by height

Goal: U^i (G^i) contain the derivation trees of G of height i (at most i).

$$G: \; X \;\to\; a\,X\,X \;\mid\; b\,X \;\mid\; c$$

$$X^{\langle 1 \rangle} \;\to\; c$$

$$X^{[1]} \;\to\; X^{\langle 1 \rangle}$$

$$X^{\langle k \rangle} \;\to\; aX^{\langle k-1 \rangle}X^{\langle k-1 \rangle} \mid aX^{[k-2]}X^{\langle k-1 \rangle} \mid aX^{\langle k-1 \rangle}X^{[k-2]} \mid bX^{\langle k-1 \rangle}$$

.

Approximating grammars by height

Goal: U^i (G^i) contain the derivation trees of G of height i (at most i).

$$G: \; X \rightarrow a\,X\,X \mid b\,X \mid c$$

$$X^{\langle 1 \rangle} \rightarrow c$$

$$X^{[1]} \rightarrow X^{\langle 1 \rangle}$$

$$X^{\langle k \rangle} \rightarrow aX^{\langle k-1 \rangle}X^{\langle k-1 \rangle} \mid aX^{[k-2]}X^{\langle k-1 \rangle} \mid aX^{\langle k-1 \rangle}X^{[k-2]} \mid bX^{\langle k-1 \rangle}$$

$$X^{[k]} \rightarrow X^{\langle k \rangle} \mid X^{[k-1]}$$

.

Approximating grammars by height

Goal: U^i (G^i) contain the derivation trees of G of height i (at most i).

$$G: X \rightarrow aXX \mid bX \mid c$$

$$X^{\langle 1\rangle} \rightarrow c$$

$$X^{[1]} \rightarrow X^{\langle 1\rangle}$$

$$X^{\langle k\rangle} \rightarrow aX^{\langle k-1\rangle}X^{\langle k-1\rangle} \mid aX^{[k-2]}X^{\langle k-1\rangle} \mid aX^{\langle k-1\rangle}X^{[k-2]} \mid bX^{\langle k-1\rangle}$$

$$X^{[k]} \rightarrow X^{\langle k\rangle} \mid X^{[k-1]}$$

U^i (G^i) is the grammar with $X^{\langle i\rangle}$ $(X^{[i]})$ as axiom.

Approximating grammars by height

$$X^{\langle k\rangle} \rightarrow aX^{\langle k-1\rangle}X^{\langle k-1\rangle} \mid aX^{[k-2]}X^{\langle k-1\rangle} \mid aX^{\langle k-1\rangle}X^{[k-2]} \mid bX^{\langle k-1\rangle}$$

$$X^{[k]} \rightarrow X^{\langle k\rangle} \mid X^{[k-1]}$$

"Taking values" we get:

$$\begin{aligned} V(U^k) &= V(a)\cdot V(U^{k-1})^2 + V(a)\cdot V(G^{k-2})\cdot V(U^{k-1}) \\ &+ V(a)\cdot V(U^{k-1})\cdot V(G^{k-2}) + V(b)\cdot V(U^{k-1}) \\ V(G^k) &= V(G^{k-1}) + V(U^k) \end{aligned}$$

and since $f(X) = V(a)\cdot X^2 + V(b)\cdot X + V(c)$

$$\begin{aligned} V(G^1) &= f(0) \\ V(G^{i+1}) &= f(V(G^i)) \quad \text{for every } i \geq 1 \end{aligned}$$

Kleene approximation corresponds to evaluating the derivation trees of G by increasing height.

A "faster" approximation

$$G: \; X \;\rightarrow\; a\,X\,X \;\mid\; b\,X \;\mid\; c \;.$$

Recall the approximation by height

$$X^{\langle k\rangle} \;\rightarrow\; aX^{\langle k-1\rangle}X^{\langle k-1\rangle} \mid aX^{[k-2]}X^{\langle k-1\rangle} \mid aX^{\langle k-1\rangle}X^{[k-2]} \mid bX^{\langle k-1\rangle}$$

To capture more trees we allow linear recursion.

$$X^{\langle k\rangle} \;\rightarrow\; aX^{\langle k-1\rangle}X^{\langle k-1\rangle} \mid aX^{[k-1]}X^{\langle k\rangle} \mid aX^{\langle k\rangle}X^{[k-1]} \mid bX^{\langle k-1\rangle}$$

U^i (G^i) defined as before.

Taking values

$$X^{\langle k\rangle} \;\to\; aX^{\langle k-1\rangle}X^{\langle k-1\rangle} \mid aX^{[k-1]}X^{\langle k\rangle} \mid aX^{\langle k\rangle}X^{[k-1]} \mid bX^{\langle k-1\rangle}$$

$V(U^i)$ is the least solution of the linear equation

$$\begin{aligned} X \;=\; & \quad V(a) \cdot V(U^{i-1})^2 + V(a) \cdot V(G^{i-1}) \cdot X \\ & + \; V(a) \cdot X \cdot V(G^{i-1}) + V(b) \cdot X \end{aligned}$$

Iterative approximation of $V(G)$:

- $V(G^1) =$ least solution of $X = V(b) \cdot X + V(c)$
- $V(G^{i+1}) = V(G^i) + V(U^{i+1})$ for every $i \geq 1$

Recipe to approximate μf by solving linear equations.

Interpreting the new approximation

Consider equations $X = f(X)$ on the real semiring

Let $g(X) = f(X) - X$. Then μf is a zero of $g(X)$.

Simple arithmetic yields

$$V(G^{i+1}) = V(G^i) - \frac{g(V(G^i))}{g'(V(G^i))}$$

where $g'(X)$ is the derivative of g.

This is Newton's method for approximating a zero of a differentiable function.

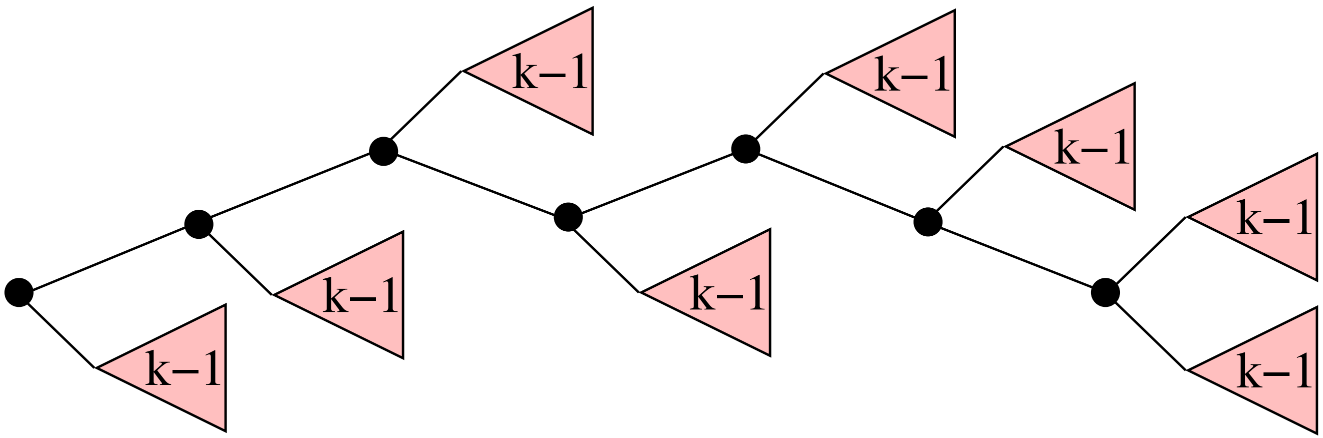
Language theoretic view of Newton's method

$$X^{\langle k \rangle} \;\to\; aX^{\langle k-1 \rangle}X^{\langle k-1 \rangle} \mid aX^{[k-1]}X^{\langle k \rangle} \mid aX^{\langle k \rangle}X^{[k-1]} \mid bX^{\langle k-1 \rangle}$$

Say a tree of G has dimension k if it is derived from U^k

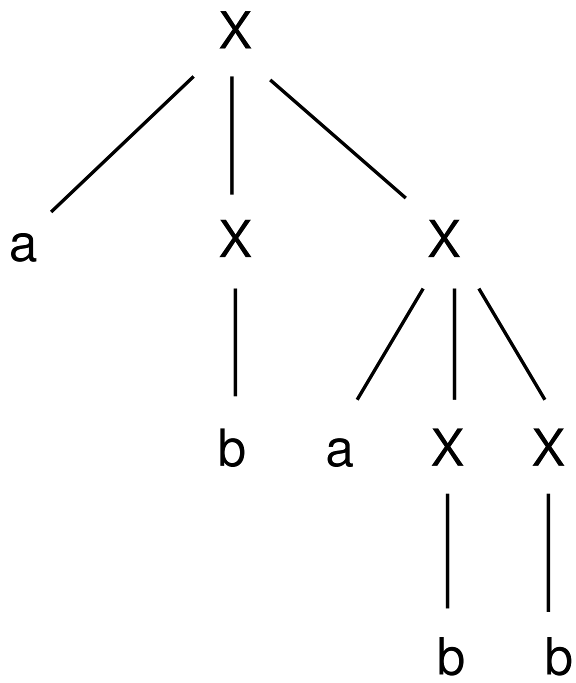
A derivation tree has dimension 0 if it has one node.

A derivation tree has dimension $k > 0$ if it consists of a spine with subtrees of dimension at most $k-1$ (and at least one subtree of dimension $k-1$).



Understanding dimension

The dimension of a derivation tree is the height of the largest full binary tree embeddable in it (ignoring terminals).



Newton approximation corresponds to evaluating the derivation trees of G by increasing dimension.