RelationalAI

Dagstuhl Seminar on
Semirings in Databases, Automata, and Logic

Polytime Convergence of Datalogo over p-stable Semirings

Sungjin Im, Ben Moseley, Hung Ngo, Kirk Pruhs

Problem and Results

Problem in a Nutshell

Given

- a commutative p-stable semiring $\boldsymbol{S} = (S, \oplus, \otimes, \mathbf{0}, \mathbf{1})$
- a polynomial function $\boldsymbol{f} : S^n \to S^n$

Solve $\boldsymbol{x} = f(\boldsymbol{x})$ via the iterative method:

$$f^{(0)}(\mathbf{0}) = \mathbf{0} \qquad f^{(i+1)}(\mathbf{0}) = f(f^{(i)}(\mathbf{0}))$$

Problem

What is the number of iterations until convergence? (i.e. $f^{(t)}(\mathbf{0}) = f^{(t+1)}(\mathbf{0})$)

p-Stable Semirings

Given a commutative semiring $\boldsymbol{S} = (S, \oplus, \otimes, \mathbf{0}, \mathbf{1})$ is p-stable iff

$$a^{(p)} = a^{(p+1)} \quad \forall a \in S \qquad \text{where } a^{(p)} = \mathbf{1} \oplus a \oplus a^2 \cdots \oplus a^p$$

Examples

- Absorptive semirings $\equiv 0$-stable semirings $\qquad \mathbf{1} = \mathbf{1} \oplus a$
- The tropical semiring Trop_p is p-stable $\qquad$ (Definition later)

Previously Best Known Results

n = # of variables in f (i.e. output size)

KNPSW PODS 22,

S	f	bound on stability(f)	
0-stable		n	subsumes traditional Datalog
Trop_p	linear	$(p+1)n$	
p-stable		$\sum_{i=0}^{n}(p+2)^i$	
p-stable	linear	$\sum_{i=0}^{n}(p+1)^i$	

Our Results

σ = # of different constants in f δ = degree(f)

S	f	bound on stability(f)	
0-stable		n	subsumes traditional Datalog
Trop_p	linear	$(p+1)n$	
p-stable		$O(\sigma p n^2 (n^2 \log\delta + \lg\sigma))$	Roughly $O(pn^4)$ IMNP PODS 25
p-stable	linear	$O((p+1)n^3)$	IMNP ICDT 24
	linear	$O((p+1)n \log\lvert S\rvert)$	IMNP ICDT 24

Still to be confirmed: $\tilde{O}((p+1)n)$ for general f.

Proof Sketch

- Reformulate the problem as a question about a CFL and its Parikh image
- A strengthened Parikh theorem
- A couple of combinatorial lemmas about linear sets and p-stability

Stability and Context-Free Languages

Iterating a Polynomial Map and Context-Free Languages

Polynomial Map Iteration

$$\begin{bmatrix} X \\ Y \end{bmatrix} \xrightarrow{\boxed{f}} \begin{bmatrix} aXY + bY + c \\ uXY + vX + w \end{bmatrix}$$

$$\xrightarrow{\boxed{f}} \begin{bmatrix} a(aXY + bY + c)(uXY + vX + w) + b(uXY + vX + w) + c \\ u(aXY + bY + c)(uXY + vX + w) + v(aXY + bY + c) + w \end{bmatrix}$$

The Kleene Sequence for f:

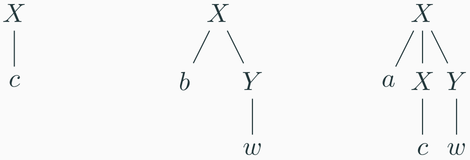
$$f^{(0)}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad f^{(1)}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} c \\ w \end{bmatrix} \qquad f^{(2)}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} acw + bw + c \\ ucw + vc + w \end{bmatrix}$$

Iterating a Polynomial Map and Context-Free Languages

Corresponding CFG (There must be a constant coefficient in every rule)

$$X \to aXY \mid bY \mid c \qquad\qquad Y \to uXY \mid vX \mid w$$

X-derivation trees of depth ≤ 2 for the grammar



$$f^{(2)}(\mathbf{0}) \mid_X = \; acw + bw + c$$

Iterating a Polynomial Map and Context-Free Languages

Lemma (Esparza, Kiefer, Luttenberger – JACM 2010)

Let $\mathcal{T}_X^{(t)}$ *denote the set of* X-derivation trees *of depth* $\leq t$ *for the grammar. Let* $Y(T)$ *denote the product of the constants in the leaves of* T*. Then,*

$$f^{(t)}(\mathbf{0})\,|_X = \bigoplus_{T \in \mathcal{T}_X^{(t)}} Y(T) \qquad \forall t \geq 0$$

Iterating a Polynomial Map and Context-Free Languages

$$f^{(t)}(\mathbf{0}) = f^{(t+1)}(\mathbf{0}) \Leftrightarrow \bigoplus_{T \in \mathcal{T}_X^{(t)}} Y(T) = \bigoplus_{T \in \mathcal{T}_X^{(t+1)}} Y(T) \qquad \forall X$$

We will show that, for "large" t,

$$\text{depth}(\hat{T}) = t + 1 \Rightarrow \boxed{Y(\hat{T}) \oplus \bigoplus_{T \in \mathcal{T}_X^{(t)}} Y(T) = \bigoplus_{T \in \mathcal{T}_X^{(t)}} Y(T)}$$

A Strengthened Parikh Theorem

Parikh Images

- Let $\boldsymbol{c} := (c_1, \ldots, c_\sigma)$ be the vector of all constants appearing in $\boldsymbol{f}$
- For any derivation tree T, its Parikh image $\Pi(T)$ is

$$\Pi(T) := \begin{bmatrix} \#c_1(T) \\ \vdots \\ \#c_\sigma(T) \end{bmatrix} \in \mathbb{N}^\sigma, \qquad Y(T) = \boldsymbol{c}^{\Pi(T)}, \text{ where } \boldsymbol{c}^{\boldsymbol{\alpha}} = c_1^{\alpha_1} \cdots c_\sigma^{\alpha_\sigma}, \boldsymbol{\alpha} \in \mathbb{N}^\sigma$$

$$\Pi(T) = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \boldsymbol{c} = (a, b, c, u, v, w)$$

Reformulation of what we want to get to

If t is "large", for any tree $\hat{T}$ with depth$(\hat{T}) \geq t+1$,

$$\boldsymbol{c}^{\Pi(\hat{T})} \oplus \bigoplus_{T \in \mathcal{T}_X^{(t)}} \boldsymbol{c}^{\Pi(T)} = \bigoplus_{T \in \mathcal{T}_X^{(t)}} \boldsymbol{c}^{\Pi(T)}$$

Parikh's Theorem

A linear set is a set of the form

$$L_{\boldsymbol{v},\boldsymbol{M}} := \{\boldsymbol{v} + \boldsymbol{M}\boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathbb{N}^k\}$$

where $\boldsymbol{M} \in \mathbb{N}^{\sigma \times k}$, $\boldsymbol{v} \in \mathbb{N}^\sigma$, and $k \in \mathbb{N}$. $\boldsymbol{v}, \boldsymbol{M}$ are parameters of the linear set

Theorem (Parikh)

The Parikh image of a CFL is a union of finitely many linear sets.

How does Parikh's Theorem help?

$$\boldsymbol{c}^{\Pi(\hat{T})} = \boldsymbol{c}^{\boldsymbol{v}+\boldsymbol{M}\boldsymbol{\alpha}} = \boldsymbol{c}^{\boldsymbol{v}} \prod_{i=1}^{k} (\boldsymbol{c}^{\boldsymbol{m}_i})^{\alpha_i}$$

From p-stability, if $\|\boldsymbol{\alpha}\|_\infty > p$, then (will prove this later)

$$\boldsymbol{c}^{\boldsymbol{v}+\boldsymbol{M}\boldsymbol{\alpha}} \oplus \bigoplus_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_\infty \le p} \boldsymbol{c}^{\boldsymbol{v}+\boldsymbol{M}\beta} = \bigoplus_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_\infty \le p} \boldsymbol{c}^{\boldsymbol{v}+\boldsymbol{M}\beta}$$

It remains to prove a slightly *strengthen* Parikh's theorem:

- If depth$(\hat{T}) \ge t+1$, then $\exists(\boldsymbol{v}, \boldsymbol{M})$ s.t. $\Pi(\hat{T}) = \boldsymbol{v} + \boldsymbol{M}\boldsymbol{\alpha}$ where $\|\boldsymbol{\alpha}\|_\infty > p$
- $\boldsymbol{v} + \boldsymbol{M}\boldsymbol{\beta} = \Pi(T)$ for some distinct T with depth$(T) \le t$

The Strengthened Parikh's Theorem

Theorem (IMNP 2025)

For any CFL $\mathcal{L}$, $\exists$ a finite set $\mathcal{C}$ of parameters $(\boldsymbol{v}, \boldsymbol{M})$ s.t.

- *For any derivation tree T of $\mathcal{L}$, $\exists(\boldsymbol{v}, \boldsymbol{M}) \in \mathcal{C}$ s.t.*

$$\Pi(T) = \boldsymbol{v} + \boldsymbol{M}\boldsymbol{\alpha} \tag{1}$$

$$\textit{depth}(T) \leq (\|\boldsymbol{\alpha}\|_1 + 1)n(n+1) \tag{2}$$

- *For any $(\boldsymbol{v}, \boldsymbol{M}) \in \mathcal{C}$, $\exists$ a derivation tree T of $\mathcal{L}$ s.t. (1) and (2) hold.*
- *For any $(\boldsymbol{v}, \boldsymbol{M}) \in \mathcal{C}$, $\|\boldsymbol{v}\|_\infty \leq \delta^{n(n+1)}$ and $\|\boldsymbol{M}\|_{\max} \leq \delta^{n(n+1)}$.*

$$\|\boldsymbol{\alpha}\|_\infty \geq \frac{\|\boldsymbol{\alpha}_1\|}{\|\boldsymbol{\alpha}\|_0} = \Omega\left(\frac{\text{depth}(\hat{T})}{n^2}\right) \cdot \frac{1}{\|\boldsymbol{\alpha}\|_0}$$

also want $\|\boldsymbol{\alpha}\|_0$ to be "small"

A Couple of Combinatorial Lemmas

Linear Sets with Small Supports

Lemma

For any $(\boldsymbol{v}, \boldsymbol{M}) \in \mathcal{C}$*, where* $\boldsymbol{M}$ *has* k *columns, and any* $\boldsymbol{\alpha} \in \mathbb{N}^k$*, there exists a* $\boldsymbol{\alpha}' \in \mathbb{N}^k$ *with* $\boldsymbol{v} + \boldsymbol{M}\boldsymbol{\alpha}' = \boldsymbol{v} + \boldsymbol{M}\boldsymbol{\alpha}$*, where* $\|\boldsymbol{\alpha}'\|_1 = \|\boldsymbol{\alpha}\|_1$ *and* $\|\boldsymbol{\alpha}'\|_0 = O(\sigma(n^2 \log \delta + \lg \sigma))$.

A Property of p-Stable Semirings

Lemma (KNPSW PODS 22, IMNP PODS 25)

For any $(\boldsymbol{v}, \boldsymbol{M}) \in \mathcal{C}$*, where* $\boldsymbol{M}$ *has* k *columns, and any* $\boldsymbol{\alpha} \in \mathbb{N}^k$*, where* $\|\boldsymbol{\alpha}\|_\infty > p$*. Define* $L_{\leq p} := \{\boldsymbol{v} + \boldsymbol{M}\boldsymbol{\beta} \mid \textit{support}(\boldsymbol{\beta}) \subseteq \textit{support}(\boldsymbol{\alpha}), \boldsymbol{\beta} \in \mathbb{N}^k, \|\boldsymbol{\beta}\|_\infty \leq p\}$*. Then,*

$$\boldsymbol{c}^{\boldsymbol{v}+\boldsymbol{M}\boldsymbol{\alpha}} \oplus \bigoplus_{\boldsymbol{u} \in L_{\leq p}} \boldsymbol{c}^{\boldsymbol{u}} = \bigoplus_{\boldsymbol{u} \in L_{\leq p}} \boldsymbol{c}^{\boldsymbol{u}}$$

As $\Pi(\hat{T}) = \boldsymbol{v} + \boldsymbol{M}\boldsymbol{\alpha}$ where $\|\boldsymbol{\alpha}\|_\infty > p$, and $L_{\leq p} \subseteq \{\Pi(T) \mid T \in \mathcal{T}_X^{(t)}\}$, proof is complete.

Application to Datalogo

Datalogo Convergence Problem

Given a Datalogo program under a p-stable semiring, how many steps does it take for the iterative method to converge?

Datalogo Examples

$$\mathsf{TC}(u,v) = E(u,v) \vee \exists w \; \mathsf{TC}(u,w) \wedge \mathsf{TC}(w,v) \qquad \text{quadratic}$$

$$\mathsf{APSP}(u,v) = \min\left(E(u,v),\; \min_w \{\mathsf{APSP}(u,w) + E(w,v)\}\right) \qquad \text{linear}$$

$$\mathsf{APSP}_p(u,v) = \mathsf{minp}\,(E(u,v),\; \mathsf{minp}_w\, \{\mathsf{APSP}(u,w) \oplus_p E(w,v)\}) \qquad \text{linear}$$

The Trop_p semiring captures the p-shortest paths problem.

Grounding Datalogo Programs

$$\mathsf{TC}(u,v) = E(u,v) \oplus \bigoplus_{w} \mathsf{TC}(u,w) \otimes \mathsf{TC}(w,v)$$

$$E = \{(1,2),(2,3),(3,4)\}$$

The grounded version

$$X_{12} = e_{12} \qquad X_{23} = e_{23}$$

$$X_{13} = X_{12} \otimes X_{23} \qquad X_{24} = X_{23} \otimes X_{34}$$

$$X_{14} = X_{12} \otimes X_{24} \oplus X_{13} \otimes X_{34} \qquad X_{34} = e_{34}.$$

Equivalent Formulation: Stability Index of Kleene Sequence

Given a commutative semiring $\boldsymbol{S} = (S, \oplus, \otimes, \mathbf{0}, \mathbf{1})$

Solve $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x})$

$$\boldsymbol{f} : S^n \to S^n$$

(f_i is a polynomial $\forall i$)

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

The stability index of f is the smallest t such that $f^{(t)}(\mathbf{0}) = f^{(t+1)}(\mathbf{0})$

Equivalent Problem

Given $\boldsymbol{f}$, and p-stable semiring $\boldsymbol{S}$, find the stability index of $\boldsymbol{f}$.

The Trop_p Semiring

For any bag $\boldsymbol{x} = \{\{x_0, x_1, \ldots, x_n\}\}$, where $x_0 \leq x_1 \leq \ldots \leq x_n$, and any $p \geq 0$, define:

$$\min_p(\boldsymbol{x}) := \{\{x_0, x_1, \ldots, x_{\min(p,n)}\}\}$$

Let $\mathcal{B}_{p+1}(\mathbb{R}_+ \cup \{\infty\})$ be the set of bags of size $p+1$.

The Trop_p semiring

$$\mathsf{Trop}_p^+ := (\mathcal{B}_{p+1}(\mathbb{R}_+ \cup \{\infty\}), \oplus_p, \otimes_p, \mathbf{0}_p, \mathbf{1}_p)$$

$$\boldsymbol{x} \oplus_p \boldsymbol{y} := \min_p(\boldsymbol{x} \uplus \boldsymbol{y}) \qquad \mathbf{0}_p := \{\{\infty, \infty, \ldots, \infty\}\}$$

$$\boldsymbol{x} \otimes_p \boldsymbol{y} := \min_p(\boldsymbol{x} + \boldsymbol{y}) \qquad \mathbf{1}_p := \{\{0, \infty, \ldots, \infty\}\}$$

Concluding Remarks

Open Problems

- $O(pn^4)$ is likely not tight. We likely will get $O(pn)$ soon
- Bound the runtime, not just the stability index.
  - If $\boldsymbol{f}$ is linear, then generalized Gaussian elimination can compute the fix point in $O(pn + n^3)$-time (Flloyd-Warshall). What about non-linear $\boldsymbol{f}$?
  - WLOG, we can assume $\boldsymbol{f}$ is quadratic (Chomsky normal form).
- Runtime in terms of some data parameters

References (and references thereof)

- Javier Esparza, An Introduction to Newton's Method on Semirings
  - https://simons.berkeley.edu/talks/javier-esparza-technische-universitat-munchen-2023-11-14
- Dan Suciu, Algebraic Structures in Logic and Query Evaluation 3 and 4
  - https://simons.berkeley.edu/workshops/logic-algorithms-database-theory-ai-boot-camp/videos

IMNP 25 Sungjin Im, Ben Moseley, Hung Q. Ngo, Kirk Pruhs, "Polynomial Time Convergence of the Iterative Evaluation of Datalogo Programs", PODS 2025.

IMNP 24 Sungjin Im, Benjamin Moseley, Hung Q. Ngo, Kirk Pruhs, "On the Convergence Rate of Linear Datalogo over Stable Semirings", ICDT 2024.

KNPSW 22 Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, and Yisu Remy Wang. "Convergence of Datalog over (Pre-) Semirings", PODS 2022.

EKL 10 Newtonian program analysis J Esparza, S Kiefer, M Luttenberger - Journal of the ACM (JACM), 2010

Thank You!