

On the Convergence of Newton's Method for Monotone Systems of Polynomial Equations

Javier Esparza Stefan Kiefer Michael Luttenberger

Technische Universität München

Grundlagen der Informatik, October 18, 2007



Monotone Systems of Polynomial Equations

We study equation systems of the form

$$\mathbf{x} = \mathbf{f}(\mathbf{x}),$$

where

- \mathbf{x} is a vector of n variables,
- $\mathbf{f}(\mathbf{x})$ is a vector of polynomials with **positive coefficients**.

Example ($n = 2$)

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.2 \cdot x \cdot y + 0.3 \cdot y^2 \\ 0.7 \cdot x^2 + 0.2 \end{pmatrix}$$

We call such systems

Monotone **S**ystems of **P**olynomial **E**quations.

We are interested in the **least solution** $\mu\mathbf{f}$.

(There is a least solution if there is any.)



- Stochastic Context-Free Grammars
 - for describing secondary structure of RNA
[Durbin, Eddy, Krogh, Mitchison, 1999]
 - for Natural Language Processing
[Manning, Schütze, 1999]

Example (Stochastic Context-Free Grammar)

$$X \xrightarrow{0.5} aXX \quad X \xrightarrow{0.5} b$$

What is the probability that X produces a word?

\Rightarrow Least solution of $x = 0.5x^2 + 0.5$



- Stochastic Context-Free Grammars
 - for describing secondary structure of RNA
[Durbin, Eddy, Krogh, Mitchison, 1999]
 - for Natural Language Processing
[Manning, Schütze, 1999]
- Verification of Probabilistic Pushdown Systems
[Esparza, Kučera, Mayr, LICS'04]
and Recursive Markov Chains
[Etesami, Yannakakis, STACS'05]
(probability that the system terminates)



- Stochastic Context-Free Grammars
 - for describing secondary structure of RNA
[Durbin, Eddy, Krogh, Mitchison, 1999]
 - for Natural Language Processing
[Manning, Schütze, 1999]
- Verification of Probabilistic Pushdown Systems
[Esparza, Kučera, Mayr, LICS'04]
and Recursive Markov Chains
[Etesami, Yannakakis, STACS'05]
(probability that the system terminates)

cf. regular Markov chains

stationary distribution is solution of $\mathbf{x} = \mathbf{f}(\mathbf{x})$,
where $\mathbf{f}(\mathbf{x})$ is linear: $\mathbf{f}(\mathbf{x}) = P \cdot \mathbf{x}$



- Stochastic Context-Free Grammars
 - for describing secondary structure of RNA
[Durbin, Eddy, Krogh, Mitchison, 1999]
 - for Natural Language Processing
[Manning, Schütze, 1999]
- Verification of Probabilistic Pushdown Systems
[Esparza, Kučera, Mayr, LICS'04]
and Recursive Markov Chains
[Etesami, Yannakakis, STACS'05]
(probability that the system terminates)
- Stochastic model of behavior of web users with
“back button” [Fagin, Karlin, Kleinberg et al., STOC'00]



- Stochastic Context-Free Grammars
 - for describing secondary structure of RNA
[Durbin, Eddy, Krogh, Mitchison, 1999]
 - for Natural Language Processing
[Manning, Schütze, 1999]
- Verification of Probabilistic Pushdown Systems
[Esparza, Kučera, Mayr, LICS'04]
and Recursive Markov Chains
[Etesami, Yannakakis, STACS'05]
(probability that the system terminates)
- Stochastic model of behavior of web users with
“back button” [Fagin, Karlin, Kleinberg et al., STOC'00]

⇒ All those problems **reduce to** computing
the least solution of **an MSPE** $\mathbf{x} = \mathbf{f}(\mathbf{x})$.



The Decision Problem

MSPE-BOUND problem

Given an MSPE $\mathbf{x} = \mathbf{f}(\mathbf{x})$ with rational coefficients and $k \in \mathbb{Q}$,
decide whether $(\mu\mathbf{f})_1 \leq k$.

MSPE-BOUND is in PSPACE by appealing to the
existential fragment of the 1st-order theory of the reals.



The Decision Problem

MSPE-BOUND problem

Given an MSPE $\mathbf{x} = \mathbf{f}(\mathbf{x})$ with rational coefficients and $k \in \mathbb{Q}$, decide whether $(\mu \mathbf{f})_1 \leq k$.

On the other hand it is at least as hard as this problem:

SQUARE-ROOT-SUM problem

Given $(d_1, \dots, d_n) \in \mathbb{N}^n$ and $k \in \mathbb{N}$, decide whether $\sum_{i=1}^n \sqrt{d_i} \leq k$.

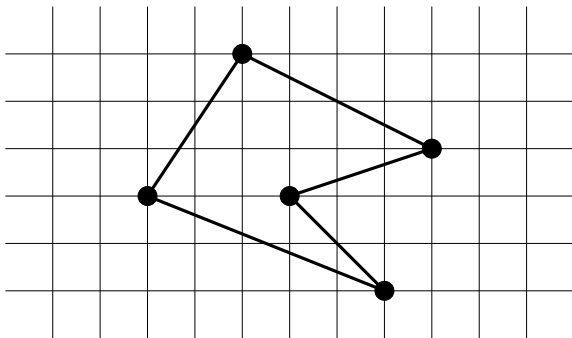
- SQUARE-ROOT-SUM is in PSPACE, but is **not known to be in NP**.
- For this reason, certain variants of the TSP problem (nodes on a grid, Euclidean distance metric) could only be proved NP-hard and not NP-complete.



Certain TSPs are not known to be in NP.

Is the route not longer than k ?

$$\Leftrightarrow \text{Is } \sqrt{2^2 + 3^2} + \sqrt{4^2 + 2^2} + \sqrt{3^2 + 1^2} + \sqrt{2^2 + 2^2} + \sqrt{5^2 + 2^2} \leq k?$$



Need for Approximation

In general μf cannot be computed exactly.

Example

Let $f(x) = \frac{1}{6}x^6 + \frac{1}{2}x^5 + \frac{1}{3}$.

Then μf is not expressible by roots.

\Rightarrow Need for approximation

$(0.3357037075 < \mu f < 0.3357037076)$



A Simple Approximation Method

Proposition (Kleene's fixed point theorem)

The Kleene sequence $\mathbf{0}, \mathbf{f}(\mathbf{0}), \mathbf{f}(\mathbf{f}(\mathbf{0})), \dots$ converges to $\mu\mathbf{f}$.



A Simple Approximation Method

Proposition (Kleene's fixed point theorem)

The Kleene sequence $\mathbf{0}, f(\mathbf{0}), f(f(\mathbf{0})), \dots$ converges to μf .

But very slowly!

Example

Let $f(x) = 0.5x^2 + 0.5$.

Then $\mu f = 1$, but $f^k(0) \leq 1 - \frac{1}{k+1}$,

so 2^i iterations needed to approximate μf within i bits.



A Simple Approximation Method

Proposition (Kleene's fixed point theorem)

The Kleene sequence $\mathbf{0}, f(\mathbf{0}), f(f(\mathbf{0})), \dots$ converges to μf .

But very slowly!

Example

Let $f(x) = 0.5x^2 + 0.5$.

Then $\mu f = 1$, but $f^k(0) \leq 1 - \frac{1}{k+1}$,

so 2^i iterations needed to approximate μf within i bits.

Solution: Newton's Method

to approximate a zero of $f(x) - x$.



Illustration of Newton's Method (univariate)

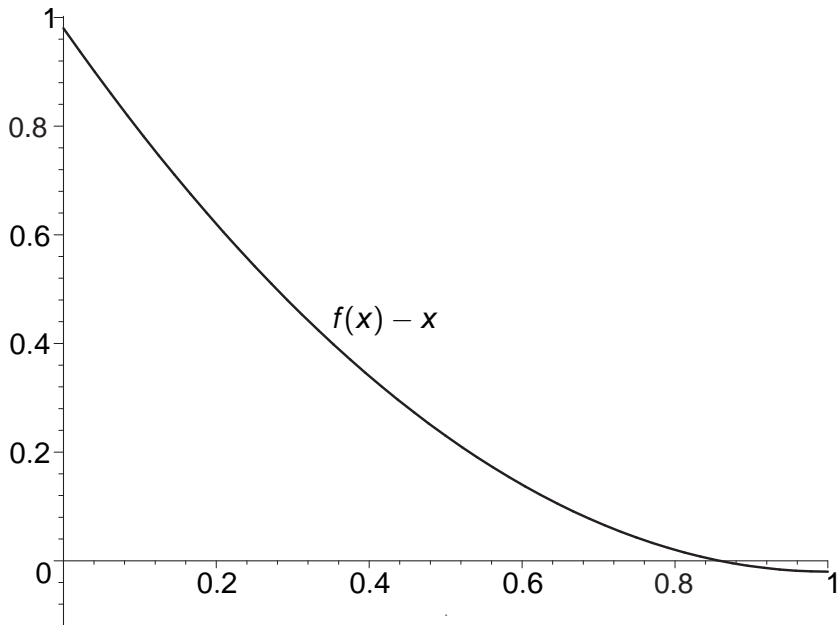


Illustration of Newton's Method (univariate)

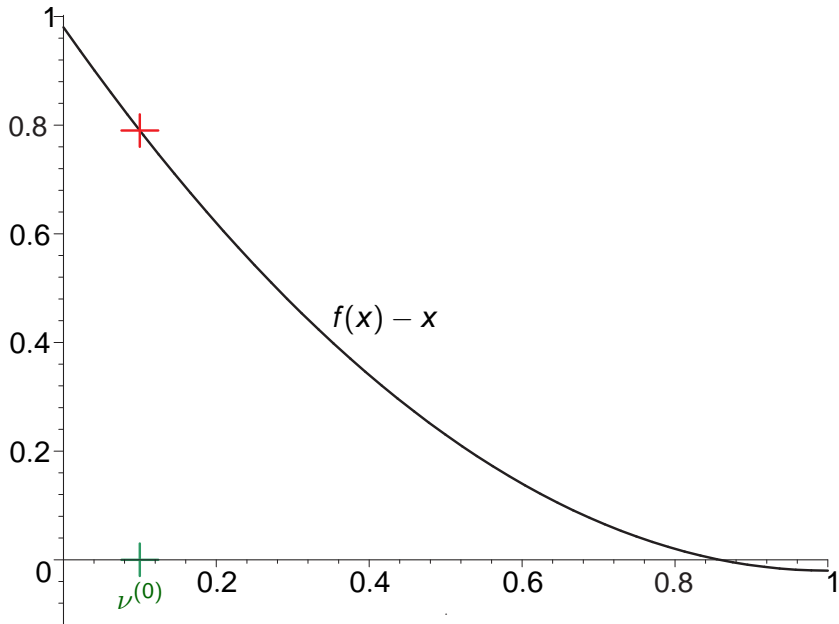


Illustration of Newton's Method (univariate)

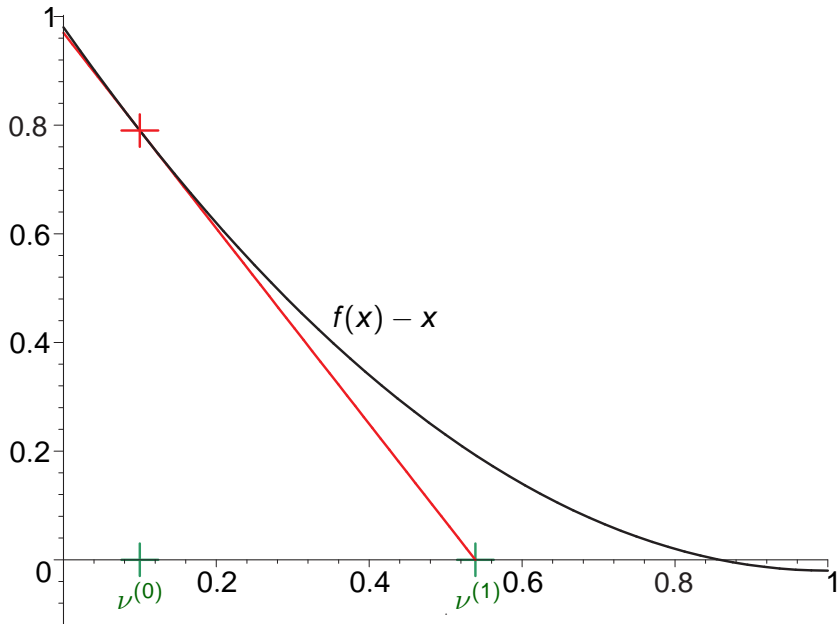
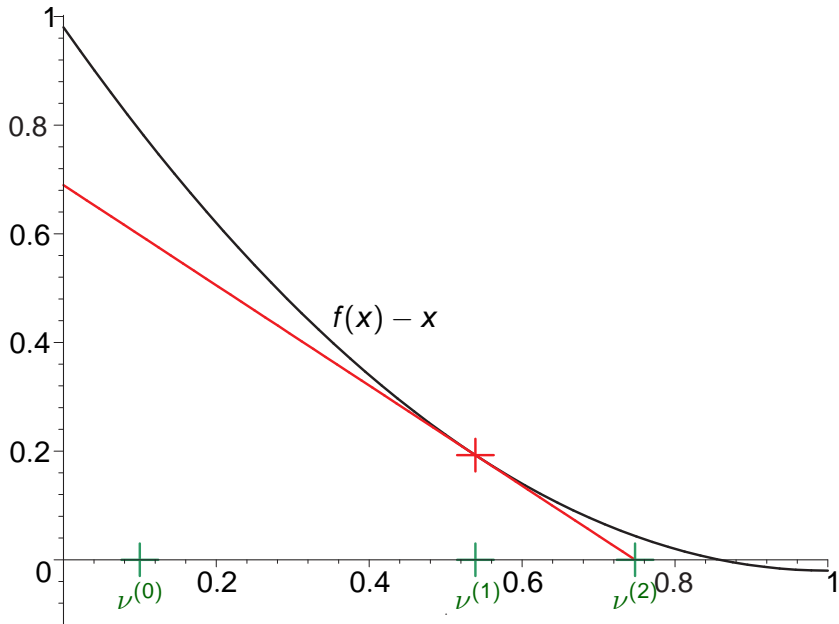


Illustration of Newton's Method (univariate)



Definition of Newton's Method

Let $\mathbf{g}(\mathbf{x})$ be a vector of polynomials.

Newton's Method is aimed at finding a zero of \mathbf{g} :

$$\begin{aligned}\boldsymbol{\nu}^{(0)} &= \mathbf{s} \\ \boldsymbol{\nu}^{(i+1)} &= \boldsymbol{\nu}^{(i)} - \left(\mathbf{g}'(\boldsymbol{\nu}^{(i)})\right)^{-1} \mathbf{g}(\boldsymbol{\nu}^{(i)}),\end{aligned}$$

where $\mathbf{g}'(\boldsymbol{\nu}^{(i)})$ is the **Jacobian** of \mathbf{g} .

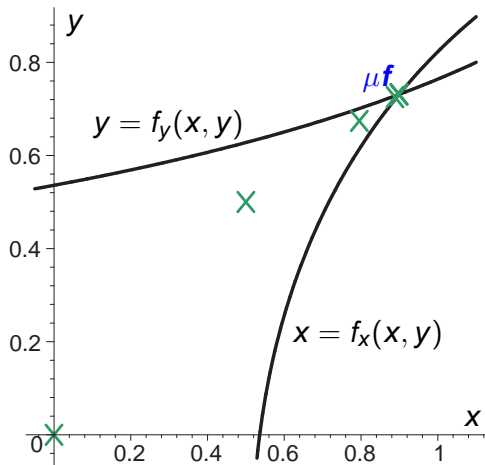
To solve $\mathbf{x} = \mathbf{f}(\mathbf{x})$ we can set $\mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{x}$ and get

$$\begin{aligned}\boldsymbol{\nu}^{(i+1)} &= \boldsymbol{\nu}^{(i)} - \left(\mathbf{f}'(\boldsymbol{\nu}^{(i)}) - \text{Id}\right)^{-1} \left(\mathbf{f}(\boldsymbol{\nu}^{(i)}) - \boldsymbol{\nu}^{(i)}\right) \\ &=: \mathcal{N}_{\mathbf{f}}(\boldsymbol{\nu}^{(i)})\end{aligned}$$



Illustration of Newton's Method (multivariate)

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{f}(x, y) = \begin{pmatrix} \frac{1}{8}x^2 + \frac{1}{4}xy + \frac{1}{4}y^2 + \frac{1}{2} \\ \frac{1}{4}xy + \frac{1}{8}y^2 + \frac{1}{2} \end{pmatrix}$$



Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be a **clean, feasible** and **strongly connected** MSPE. Let again

$$\mathcal{N}_{\mathbf{f}}(\mathbf{x}) = \mathbf{x} + (\text{Id} - \mathbf{f}'(\mathbf{x}))^{-1} (\mathbf{f}(\mathbf{x}) - \mathbf{x}).$$

Then

- (1) $(\text{Id} - \mathbf{f}'(\mathbf{x}))^{-1}$ exists for $\mathbf{0} \leq \mathbf{x} < \mu\mathbf{f}$;
- (2) the Newton sequence $\nu^{(0)}, \nu^{(1)}, \nu^{(2)}, \dots$
with $\nu^{(0)} = \mathbf{0}$ and $\nu^{(i+1)} = \mathcal{N}_{\mathbf{f}}(\nu^{(i)})$ is

- monotonically increasing: $\nu^{(i)} \leq \nu^{(i+1)}$
- bounded from above by $\mu\mathbf{f}$: $\nu^{(i)} \leq \mu\mathbf{f}$
- converges to $\mu\mathbf{f}$ $\nu^{(i)} \rightarrow \mu\mathbf{f}$
- faster than “Kleene”: $\nu^{(i)} \geq \mathbf{f}^i(\mathbf{0})$



a **strongly connected** MSPE:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0.869 \cdot x \cdot z + 0.131 \cdot z \\ 0.788 \cdot y + 0.212 \cdot x \cdot z \\ 0.105 \cdot y + 0.895 \end{pmatrix}$$



Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be a **clean, feasible** and **strongly connected** MSPE. Let again

$$\mathcal{N}_{\mathbf{f}}(\mathbf{x}) = \mathbf{x} + (\text{Id} - \mathbf{f}'(\mathbf{x}))^{-1} (\mathbf{f}(\mathbf{x}) - \mathbf{x}).$$

Then

- (1) $(\text{Id} - \mathbf{f}'(\mathbf{x}))^{-1}$ exists for $\mathbf{0} \leq \mathbf{x} < \mu\mathbf{f}$;
- (2) the Newton sequence $\nu^{(0)}, \nu^{(1)}, \nu^{(2)}, \dots$
with $\nu^{(0)} = \mathbf{0}$ and $\nu^{(i+1)} = \mathcal{N}_{\mathbf{f}}(\nu^{(i)})$ is

- monotonically increasing: $\nu^{(i)} \leq \nu^{(i+1)}$
- bounded from above by $\mu\mathbf{f}$: $\nu^{(i)} \leq \mu\mathbf{f}$
- converges to $\mu\mathbf{f}$ $\nu^{(i)} \rightarrow \mu\mathbf{f}$
- faster than “Kleene”: $\nu^{(i)} \geq \mathbf{f}^i(\mathbf{0})$



Convergence Example for Newton and Kleene Iterations

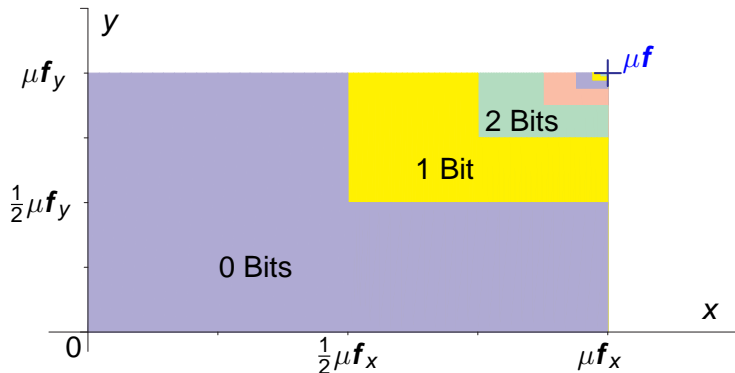
Example taken from [Knudsen, Hein, Bioinformatics'99] :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0.869 \cdot x \cdot z + 0.131 \cdot z \\ 0.788 \cdot y + 0.212 \cdot x \cdot z \\ 0.105 \cdot y + 0.895 \end{pmatrix}$$

i	$\nu_x^{(i)}$	$\nu_y^{(i)}$	$\nu_z^{(i)}$	$f^i(\mathbf{0})_x$	$f^i(\mathbf{0})_y$	$f^i(\mathbf{0})_z$
1	0.117	0.000	0.895	0.000	0.000	0.895
3	0.819	0.794	0.978	0.208	0.022	0.895
5	0.981	0.979	0.998	0.335	0.098	0.901
7	0.999	0.999	0.999	0.418	0.184	0.910



Notion of Valid Bits

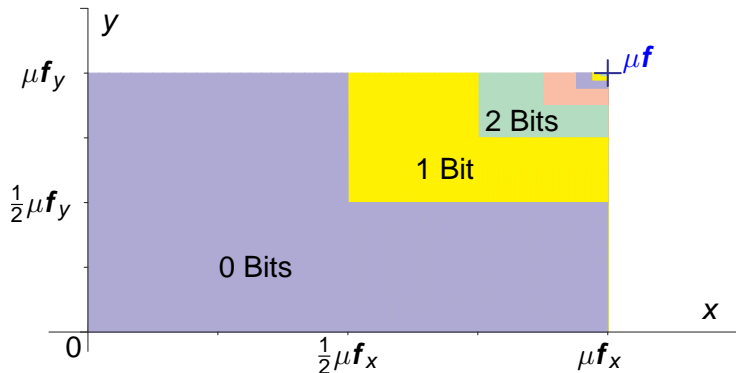


Notion of Valid Bits

Let $\mathbf{x} = \mathbf{f}(\mathbf{x})$ be an MSPE. A vector ν has i valid bits of $\mu\mathbf{f}$ if

$$\max_{m=1\dots n} \left\{ \frac{|\mu\mathbf{f}_m - \nu_m|}{|\mu\mathbf{f}_m|} \right\} \leq 2^{-i}.$$

Linear Convergence of Newton's Method

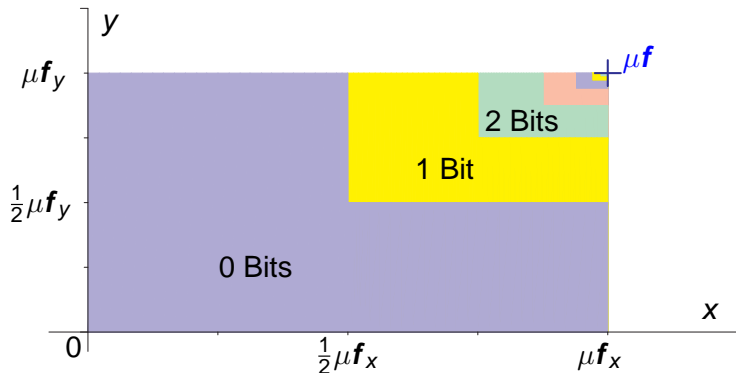


We proved: Newton's Method converges **at least linearly**:

Theorem (Esparza, K., Luttenberger, STOC'07)

*For any strongly connected, clean and feasible MSPE $\mathbf{x} = \mathbf{f}(\mathbf{x})$:
There is a threshold k_f such that
the Newton iterate $\nu^{(k_f+i)}$ has at least i valid bits of $\mu \mathbf{f}$.*

Linear Convergence of Newton's Method



We proved: Newton's Method converges **at least linearly**:

Theorem (Esparza, K., Luttenberger, STOC'07)

For any strongly connected, clean and feasible MSPE $\mathbf{x} = \mathbf{f}(\mathbf{x})$:

There is a threshold k_f such that

the Newton iterate $\nu^{(k_f+i)}$ has at least i valid bits of $\mu \mathbf{f}$.

In the Mean Time We Know More!

Theorem (Esparza, K., Luttenberger, STOC'07)

*For any strongly connected, clean and feasible MSPE $\mathbf{x} = \mathbf{f}(\mathbf{x})$:
There is a threshold k_f such that
the Newton iterate $\nu^{(k_f+i)}$ has at least i valid bits of $\mu\mathbf{f}$.*

The following is still unpublished.

We can give strong bounds on the threshold k_f .



In the Mean Time We Know More!

Theorem (Esparza, K., Luttenberger, STOC'07)

*For any strongly connected, clean and feasible MSPE $\mathbf{x} = \mathbf{f}(\mathbf{x})$:
There is a threshold k_f such that
the Newton iterate $\nu^{(k_f+i)}$ has at least i valid bits of $\mu\mathbf{f}$.*

The following is still unpublished.

We can give strong bounds on the threshold k_f .

Theorem (Esparza, K., Luttenberger, submitted)

You can choose $k_f \leq 3n^2(m + |\log \mu_{\min}|)$, where

- n is the number of equations (= number of variables),
- m is the size of the system (coefficients in binary),
- μ_{\min} is the minimal component of $\mu\mathbf{f}$.

For back-button processes even better: $k_f \leq 3nm$.

μ

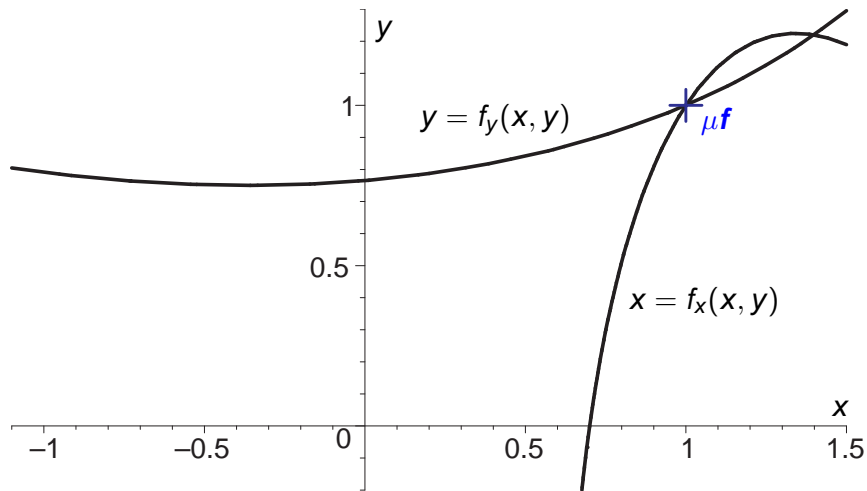
In the Mean Time We Know More!

We know much more than in the non-monotonic case:

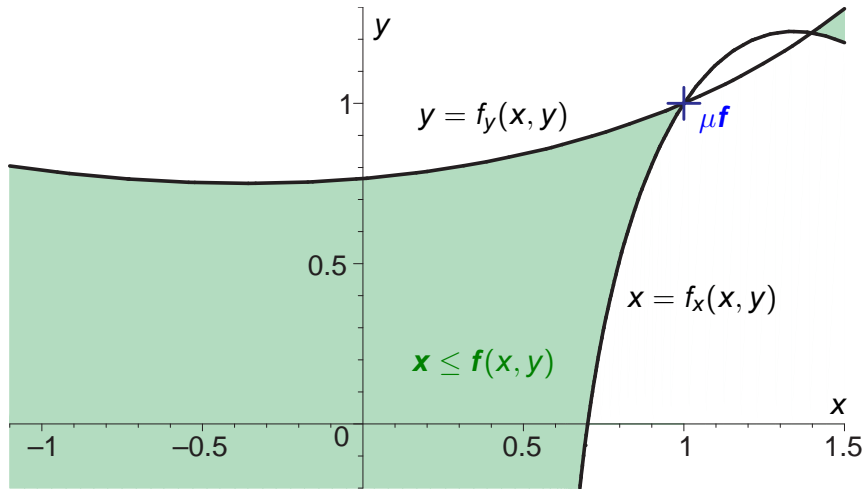
- ultimately linear convergence (1 bit per iteration)
- We know **when** this phase “**kicks in**”.



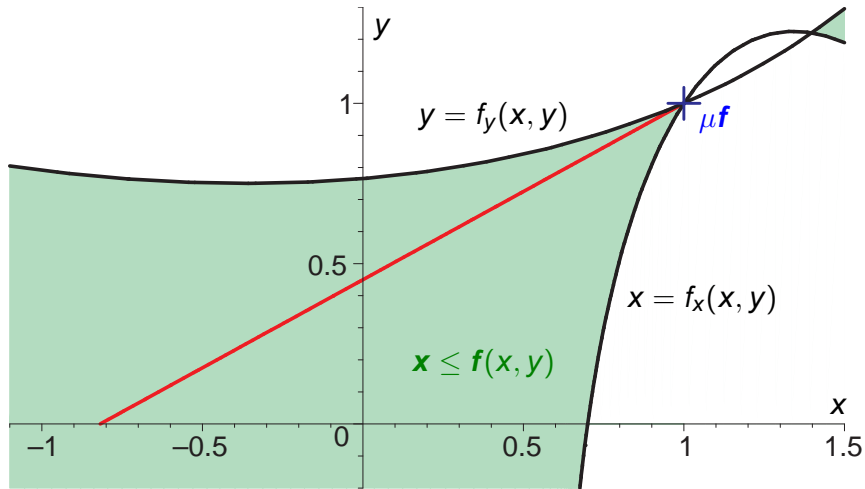
Proof Sketch



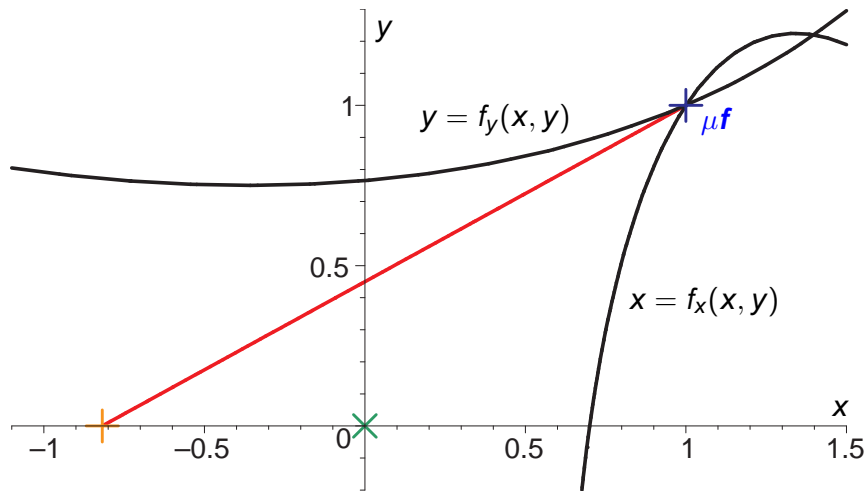
Proof Sketch



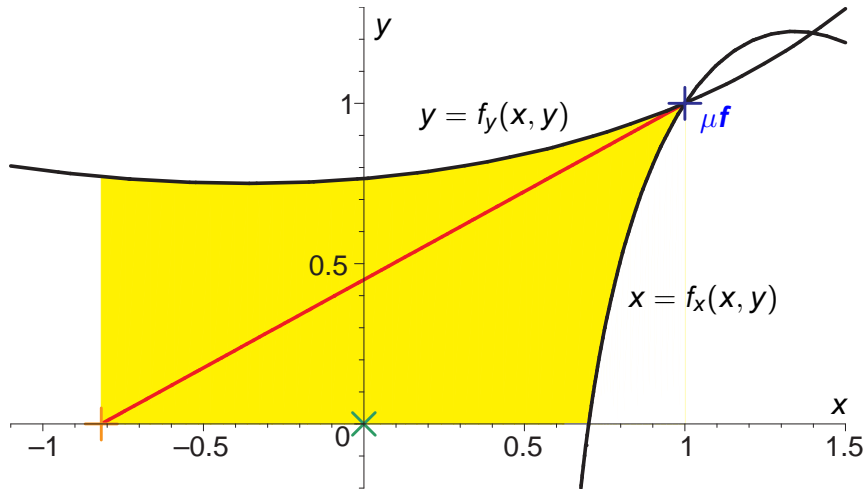
Proof Sketch



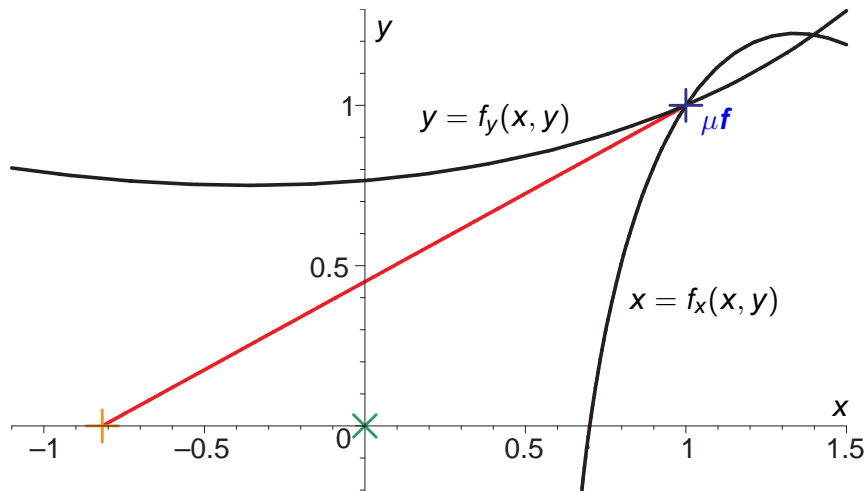
Proof Sketch



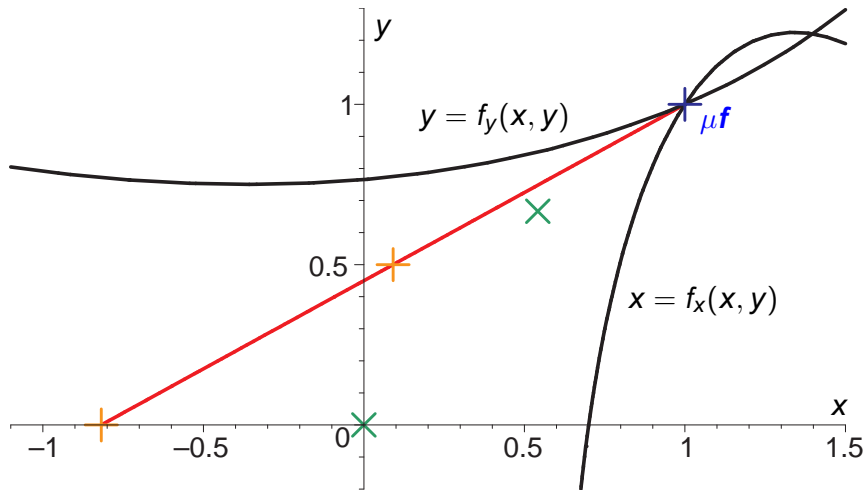
Proof Sketch



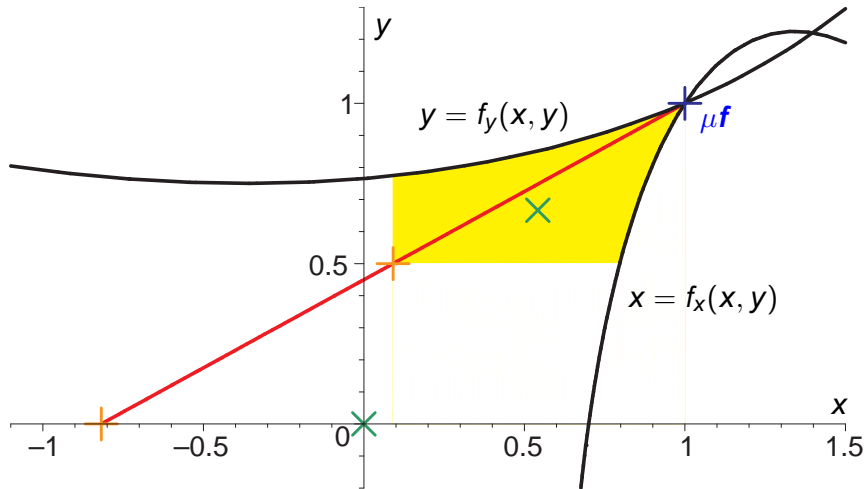
Proof Sketch



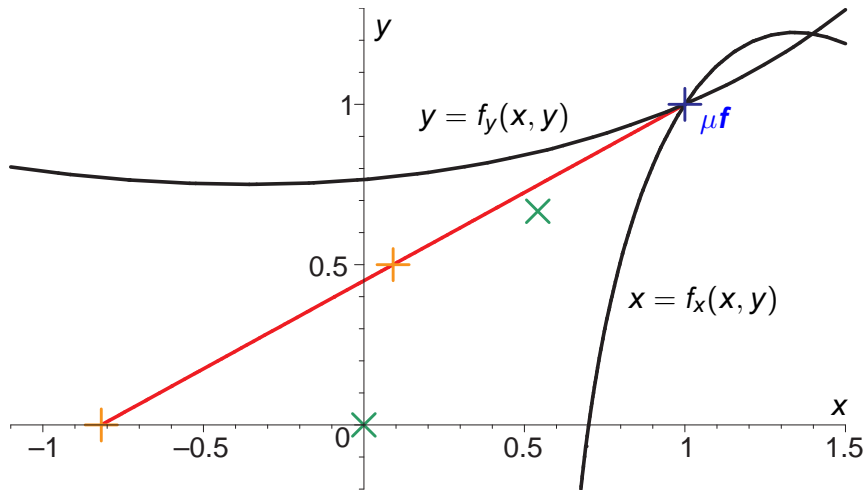
Proof Sketch



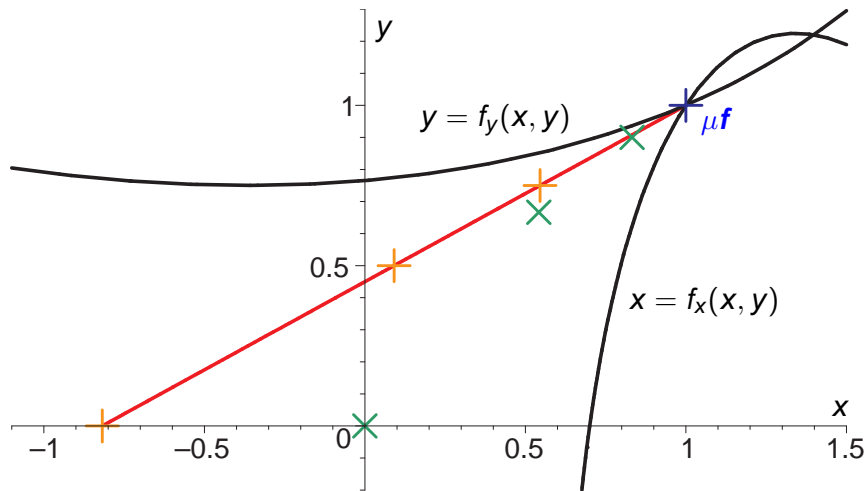
Proof Sketch



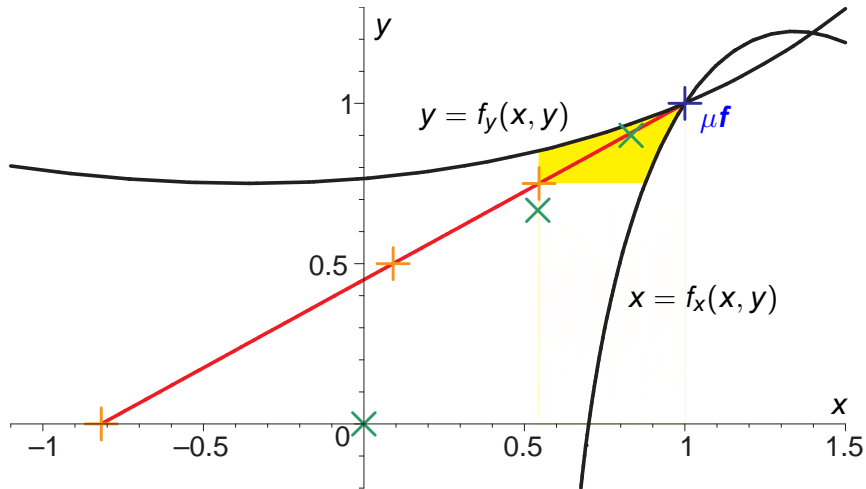
Proof Sketch



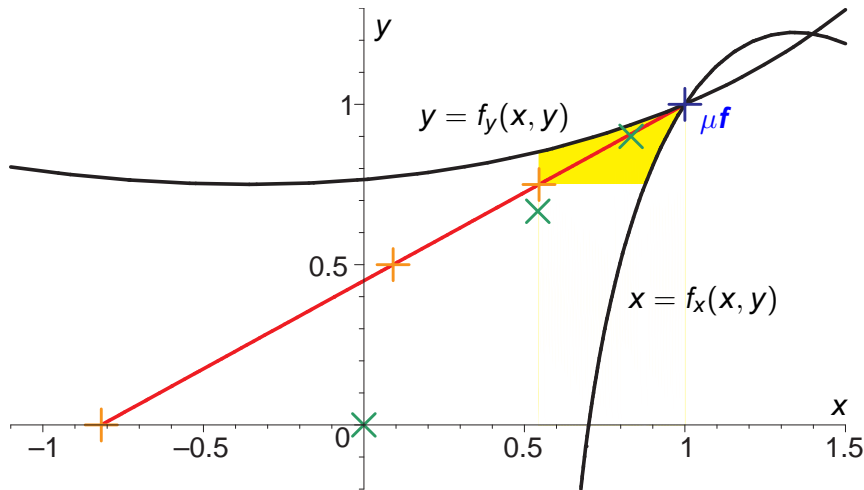
Proof Sketch



Proof Sketch

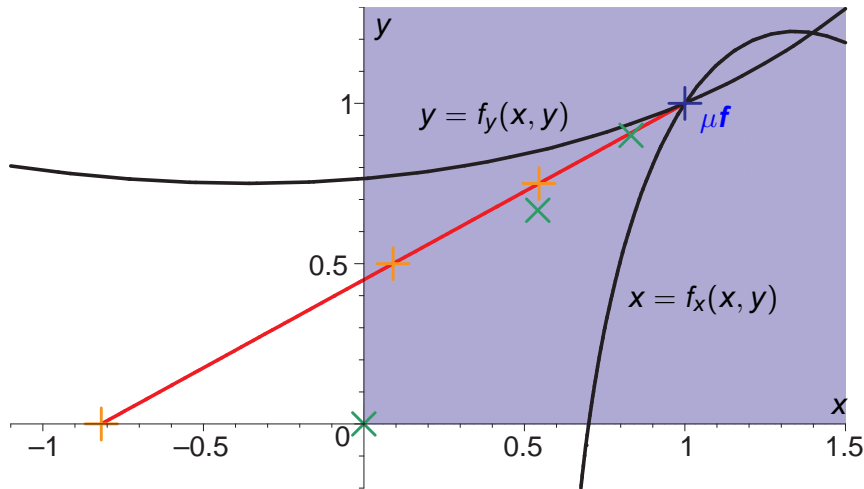


Proof Sketch



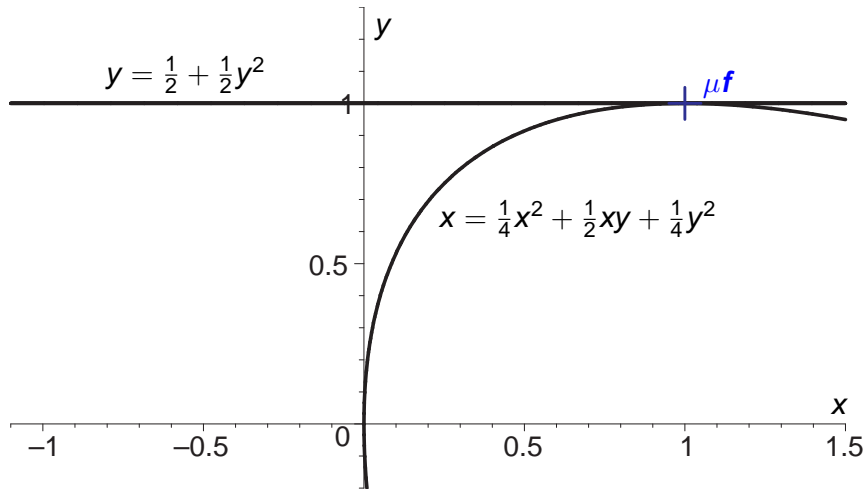
\Rightarrow Asymptotic Convergence: 1 bit per iteration

Proof Sketch

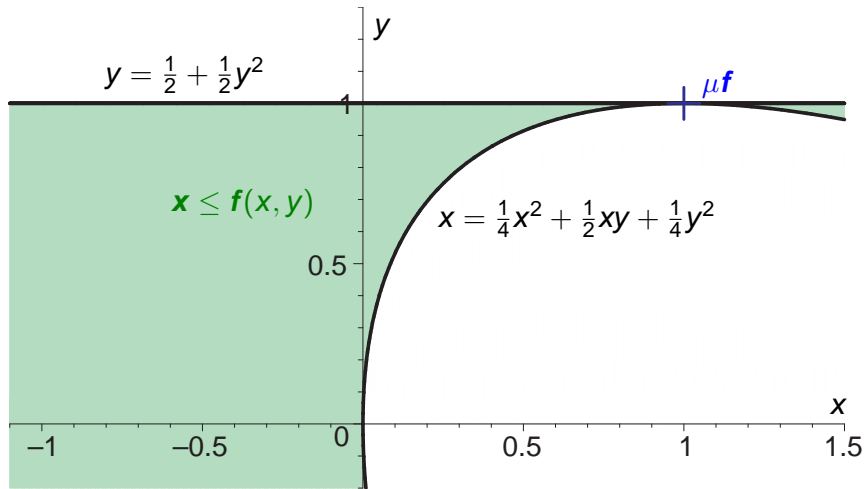


\Rightarrow Threshold: When does the **orange sequence** reach positive coordinates?

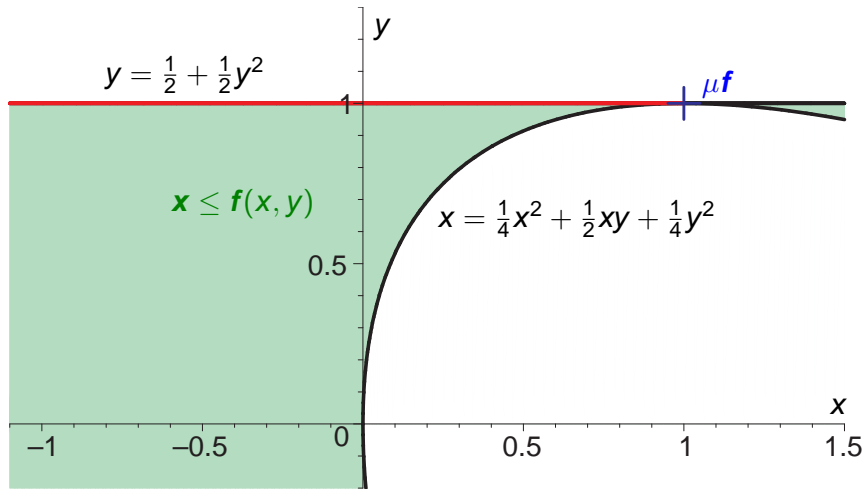
Proof Sketch: Threshold



Proof Sketch: Threshold

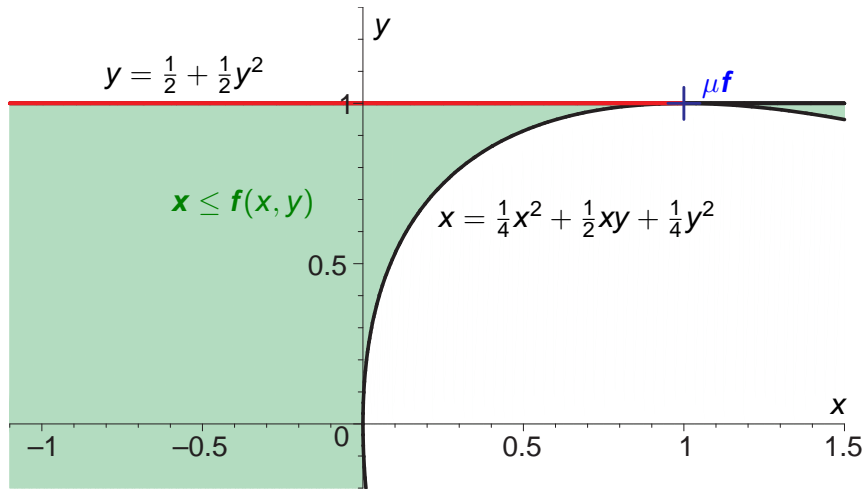


Proof Sketch: Threshold



The orange sequence is not even defined here. What's wrong?

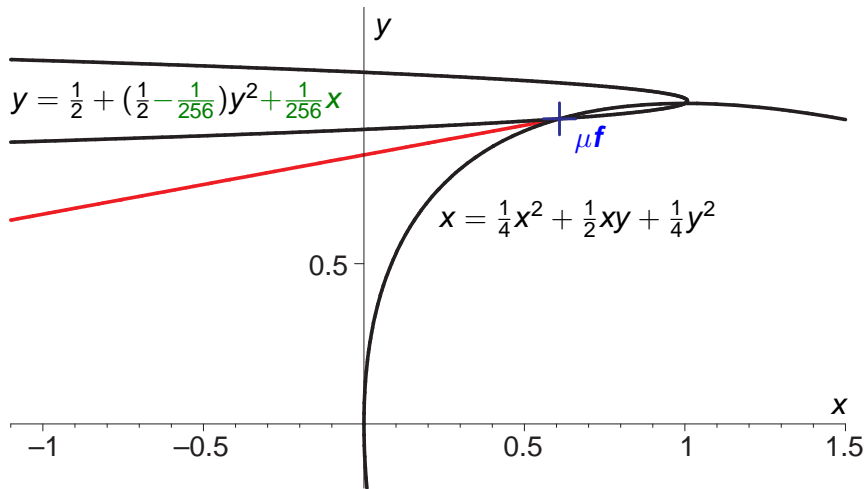
Proof Sketch: Threshold



The orange sequence is not even defined here. What's wrong?

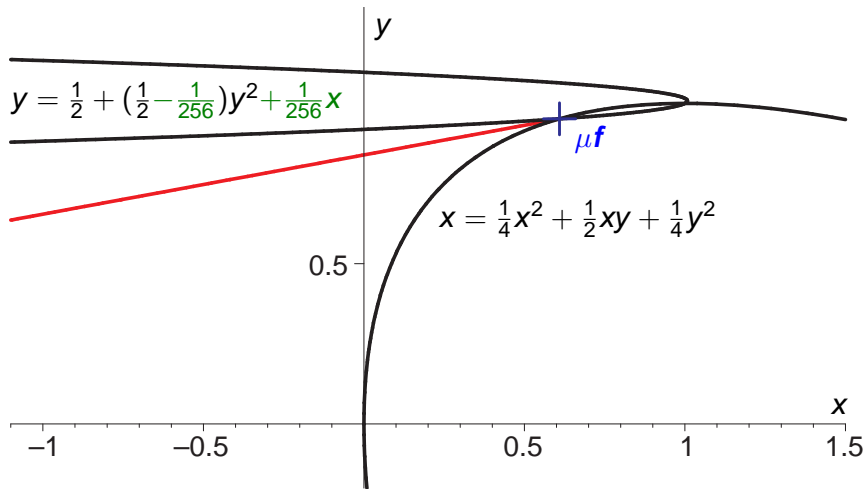
⇒ It's not a single SCC! But what if we make it one?

Proof Sketch: Threshold



Now the orange sequence is very long. What's wrong here?

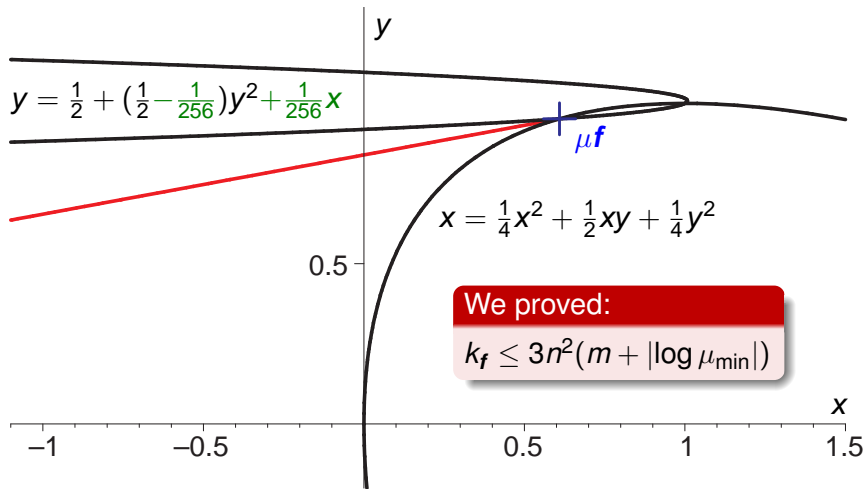
Proof Sketch: Threshold



Now the orange sequence is very long. What's wrong here?

⇒ Nothing. But now we have a very small coefficient.

Proof Sketch: Threshold



We proved:

$$k_f \leq 3n^2(m + |\log \mu_{\min}|)$$

Now the orange sequence is very long. What's wrong here?

⇒ Nothing. But now we have a very small coefficient.

An Application for Convergence Thresholds

Example

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0.4yx + 0.6 \\ 0.3xy + 0.4zy + 0.3 \\ 0.3xz + 0.7 \end{pmatrix}$$

We do 14 Newton steps (e.g. with Maple) and get:

$$\nu^{(14)} = \begin{pmatrix} 0.98283 \dots \\ 0.97380 \dots \\ 0.99269 \dots \end{pmatrix}$$

Is the solution $\mu \mathbf{f} = (1, 1, 1)^\top$?

μ

An Application for Convergence Thresholds

Example

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0.4yx + 0.6 \\ 0.3xy + 0.4zy + 0.3 \\ 0.3xz + 0.7 \end{pmatrix}$$

We do 14 Newton steps (e.g. with Maple) and get:

$$\boldsymbol{\nu}^{(14)} = \begin{pmatrix} 0.98283 \dots \\ 0.97380 \dots \\ 0.99269 \dots \end{pmatrix}$$

Is the solution $\boldsymbol{\mu f} = (1, 1, 1)^\top$? **No!** Our theorem implies:
The error after 14 iterations is at most 0.004 (8 bits). So:

$$\boldsymbol{\mu f} \leq \boldsymbol{\nu}^{(14)} + \begin{pmatrix} 0.004 \\ 0.004 \\ 0.004 \end{pmatrix} \leq \begin{pmatrix} 0.994 \\ 0.984 \\ 0.997 \end{pmatrix} < \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

μ

Some More Results

- If f is **not strongly connected** Newton's Method **still works** and converges to μf (improving a result of [EY05]).
- If f is **not strongly connected** the convergence is still linear, however worse than 1 bit per iteration: the convergence rate **depends on the height of the DAG** of SCCs.



Conclusions

- Solving MSPEs is central to several computer science problems.
- Newton's Method works very well for approximating least solutions of MSPEs.
- The convergence of Newton's Method for MSPEs can be sharply analyzed: ultimately 1 bit per iteration in the strongly connected case
- Recent work answers when linear convergence kicks in and studies the non-strongly-connected case.



Thank you!

