Solving Monotone Polynomial Equations

Javier Esparza   Stefan Kiefer   Michael Luttenberger

Technische Universität München

TCS 08, September 8

Back in victorian Britain . . .

There was concern amongst the Victorians that aristocratic families were becoming extinct.

Back in victorian Britain . . .

There was concern amongst the Victorians that aristocratic families were becoming extinct.

Francis Galton (1822-1911), anthropologist and polymath: Are families of English peers more likely to die out than the families of ordinary men?

> *Let $p_0, p_1, \ldots, p_n$ be the respective probabilities that a man has 0, 1, 2, . . . n sons, let each son have the same probability for sons of his own, and so on. What is the probability that the male line goes extinct?*

Back in victorian Britain . . .

There was concern amongst the Victorians that aristocratic families were becoming extinct.

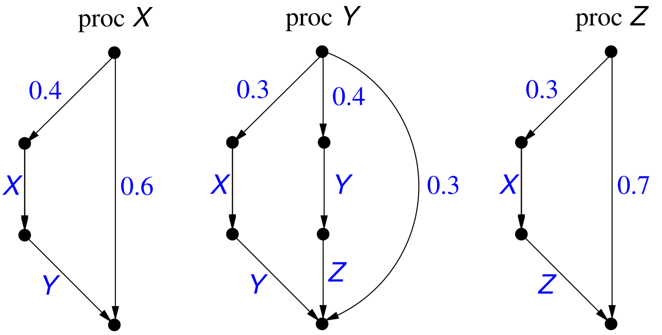
Francis Galton (1822-1911), anthropologist and polymath:
Are families of English peers more likely to die out than the families of ordinary men?

> *Let $p_0, p_1, \ldots, p_n$ be the respective probabilities that a man has 0, 1, 2, . . . n sons, let each son have the same probability for sons of his own, and so on. What is the probability that the male line goes extinct?*

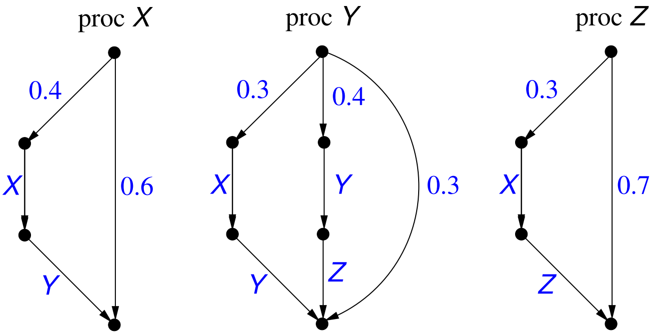
Henry William Watson (1827-1903), priest and mathematician:
The probability is the least solution of

$$X = p_0 + p_1 X + p_2 X^2 + \ldots + p_n X^n$$

Termination of probabilistic programs

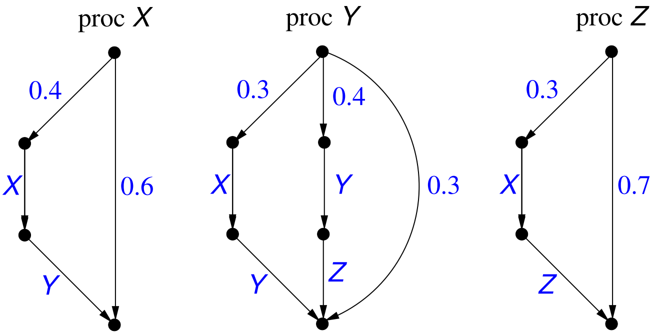


Termination of probabilistic programs



Does the program terminate with probability 1 ?

Termination of probabilistic programs



Does the program terminate with probability 1 ?
The probabilities of termination are the least solution of

$$\begin{aligned} X &= 0.4XY + 0.6 \\ Y &= 0.3XY + 0.4YZ + 0.3 \\ Z &= 0.3XZ + 0.7 \end{aligned}$$

Monotone Systems of Polynomial Equations

These are examples of equation systems of the form

$$\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$$

where

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

We call them Monotone Systems of Polynomial Equations.

Monotone Systems of Polynomial Equations

MSPEs appear in the

- analysis of stochastic branching processes
  – biology populations, chemical and nuclear reactions
- analysis of stochastic context-free grammars
  – Natural Language Processing, computational biology
- verification of probabilistic programs
- computation of reputations in reputation systems

We assume in this talk that there exists a non-negative solution.
Then there is a least one, denoted by $\mu \boldsymbol{f}$.

Monotone Systems of Polynomial Equations

MSPEs appear in the

- analysis of stochastic branching processes
  – biology populations, chemical and nuclear reactions
- analysis of stochastic context-free grammars
  – Natural Language Processing, computational biology
- verification of probabilistic programs
- computation of reputations in reputation systems

We assume in this talk that there exists a non-negative solution. Then there is a least one, denoted by $\mu \boldsymbol{f}$.

This talk surveys what is known about computing (approximating, gaining information on) $\mu \boldsymbol{f}$.

Comparing MSPEs and linear equations

The least solution of a linear system of equations with rational coefficients is rational.

Comparing MSPEs and linear equations

The least solution of a linear system of equations with rational coefficients is rational.

This property fails for MSPEs:

Example

The least solution of

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

is irrational and not expressible by radicals.

We have $0.3357037075 < \mu f < 0.3357037076$

Comparing MSPEs and linear equations

The least solution of a linear system of equations with rational coefficients can be written using polynomially many bits.

Comparing MSPEs and linear equations

The least solution of a linear system of equations with rational coefficients can be written using polynomially many bits.

This again fails for MSPEs:

Example

The *n*-th component of the least solution of

$$X_1 = 2, \quad X_2 = X_1^2, \quad \ldots \quad X_n = X_{n-1}^2$$

is $2^{2^{(n-1)}}$ and so needs $2^{(n-1)}$ bits.

Comparing MSPEs and linear equations

The least solution of a linear system of equations with rational coefficients can be computed in polynomial time (non-trivial).

Does this hold for MSPEs?

Comparing MSPEs and linear equations

The least solution of a linear system of equations with rational coefficients can be computed in polynomial time (non-trivial).

Does this hold for MSPEs?

Since in general there is no closed form for the solution of a MSPE, we reformulate the question:

MSPE-DECISION

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

An upper bound on MSPE-DECISION

Proposition

MSPE-DECISION is in PSPACE.

An upper bound on MSPE-DECISION

Proposition

MSPE-DECISION is in PSPACE.

Proof.

For $X_1 = f_1(X_1, X_2), X_2 = f_2(X_1, X_2)$, we have $(\mu f)_1 \leq a$ iff the following formula is true over the reals:

$$\exists x_1, x_2 : x_1 = f_1(x_1, x_2) \ \wedge \ x_2 = f_2(x_1, x_2) \ \wedge x_1, x_2 \geq 0 \ \wedge x_1 \leq a$$

The first-order theory of the reals is decidable [Tarski 48], and its existential fragment is in PSPACE [Canny 88]. □

An upper bound on MSPE-DECISION

Proposition

MSPE-DECISION is in PSPACE.

Proof.

For $X_1 = f_1(X_1, X_2), X_2 = f_2(X_1, X_2)$, we have $(\mu f)_1 \leq a$ iff the following formula is true over the reals:

$$\exists x_1, x_2 : x_1 = f_1(x_1, x_2) \ \wedge \ x_2 = f_2(x_1, x_2) \ \wedge x_1, x_2 \geq 0 \ \wedge x_1 \leq a$$

The first-order theory of the reals is decidable [Tarski 48], and its existential fragment is in PSPACE [Canny 88]. □

However: current algorithms limited to 5 or 6 variables.

Possibly enough for our program example, but for little more ...

Lower bounds on MSPE-DECISION [EY]

SQUARE-ROOT-SUM

Given natural numbers $d_1, \ldots, d_n \in \mathbb{N}$ and a bound $k \in \mathbb{N}$, decide whether $\sum_{i=1}^{n} \sqrt{d_i} \leq k$.

(a "subproblem" of euclidean TSP with coordinates as input)

SQUARE-ROOT-SUM is in PSPACE, but it is not known to be in NP (despite rather intense efforts).

Lower bounds on MSPE-DECISION [EY]

SQUARE-ROOT-SUM

Given natural numbers $d_1, \ldots, d_n \in \mathbb{N}$ and a bound $k \in \mathbb{N}$, decide whether $\sum_{i=1}^{n} \sqrt{d_i} \leq k$.

(a "subproblem" of euclidean TSP with coordinates as input)

SQUARE-ROOT-SUM is in PSPACE, but it is not known to be in NP (despite rather intense efforts).

PosSLP (Positive Straight Line Program) [Allender et al 06]

Given an arithmetic circuit with integer inputs and gates $+, *, -$, does the circuit output a positive number?.

Hard for the problems that can be solved with a polynomial number of arithmetic operations. Unlikely to be in P.

Lower bounds on MSPE-DECISION [EY]

Proposition [EY]

SQUARE-ROOT-SUM $\leq$ PosSLP $\leq$ MSPE-DECISION.

Lower bounds on MSPE-DECISION [EY]

Proposition [EY]

SQUARE-ROOT-SUM $\leq$ PosSLP $\leq$ MSPE-DECISION.

Conclusion:

- MSPE-DECISION is in PSPACE and unlikely to be in P.
- It might be solvable using a polynomial number of arithmetic operations; a proof of this would be a sensational result.

A simple approximation method

Proposition (Kleene's fixed point theorem)

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

Example

For our probabilistic program we get:

k	$(\boldsymbol{f}^k(0))_1$	$(\boldsymbol{f}^k(0))_2$	$(\boldsymbol{f}^k(0))_3$
0	0.000	0.000	0.000
4	0.753	0.600	0.887
8	0.834	0.738	0.926
12	0.873	0.802	0.944
16	0.897	0.839	0.955

Is the solution $\mu\boldsymbol{f} = (1, 1, 1)$?

A simple approximation method

Proposition (Kleene's fixed point theorem)

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

Example

For our probabilistic program we get:

k	$(\boldsymbol{f}^k(0))_1$	$(\boldsymbol{f}^k(0))_2$	$(\boldsymbol{f}^k(0))_3$
0	0.000	0.000	0.000
4	0.753	0.600	0.887
8	0.834	0.738	0.926
12	0.873	0.802	0.944
16	0.897	0.839	0.955

Is the solution $\mu\boldsymbol{f} = (1, 1, 1)$?
For a proof we need a guarantee on the convergence speed.

Convergence order

Definition

Let $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ satisfying $\lim_{k \to \infty} a^{(k)} = a < \infty$.

The convergence order of $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ is the function $\beta \colon \mathbb{N} \to \mathbb{N}$ where $\beta(k)$ is the number of bits of $a^{(k)}$ that coincide with the corresponding bits of a.

Informally, $\beta(k)$ is the number of accurate bits of $a^{(k)}$.

Convergence order

Definition

Let $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ satisfying $\lim_{k\to\infty} a^{(k)} = a < \infty$.

The convergence order of $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ is the function $\beta\colon \mathbb{N} \to \mathbb{N}$ where $\beta(k)$ is the number of bits of $a^{(k)}$ that coincide with the corresponding bits of a.

Informally, $\beta(k)$ is the number of accurate bits of $a^{(k)}$.

Example

If $a = 101,0110\ldots$, $a^{(0)} = 010,01$, and $a^{(1)} = 100,0101\ldots$, then $\beta(0) = 0$ and $\beta(1) = 2$.

Convergence order

Definition

Let $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ satisfying $\lim_{k \to \infty} a^{(k)} = a < \infty$.

The convergence order of $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ is the function $\beta \colon \mathbb{N} \to \mathbb{N}$ where $\beta(k)$ is the number of bits of $a^{(k)}$ that coincide with the corresponding bits of a.

Informally, $\beta(k)$ is the number of accurate bits of $a^{(k)}$.

Example

If $a = 101, 0110 \ldots$, $a^{(0)} = 010, 01$, and $a^{(1)} = 100, 0101 \ldots$, then $\beta(0) = 0$ and $\beta(1) = 2$.

Extension to sequences of vectors: take for $\beta(k)$ the minimum of the number of accurate bits of the vector components.

Convergence order

Definition

Let $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ satisfying $\lim_{k \to \infty} a^{(k)} = a < \infty$.

The convergence order of $a^{(0)} \leq a^{(1)} \leq a^{(2)} \ldots$ is the function $\beta \colon \mathbb{N} \to \mathbb{N}$ where $\beta(k)$ is the number of bits of $a^{(k)}$ that coincide with the corresponding bits of a.

Informally, $\beta(k)$ is the number of accurate bits of $a^{(k)}$.

Example

If $a = 101,0110\ldots$, $a^{(0)} = 010,01$, and $a^{(1)} = 100,0101\ldots$, then $\beta(0) = 0$ and $\beta(1) = 2$.

Extension to sequences of vectors: take for $\beta(k)$ the minimum of the number of accurate bits of the vector components.
We speak of linear, exponential, or logarithmic orders.

Kleene Iteration is slow

The Kleene sequence may have logarithmic convergence order.

Example

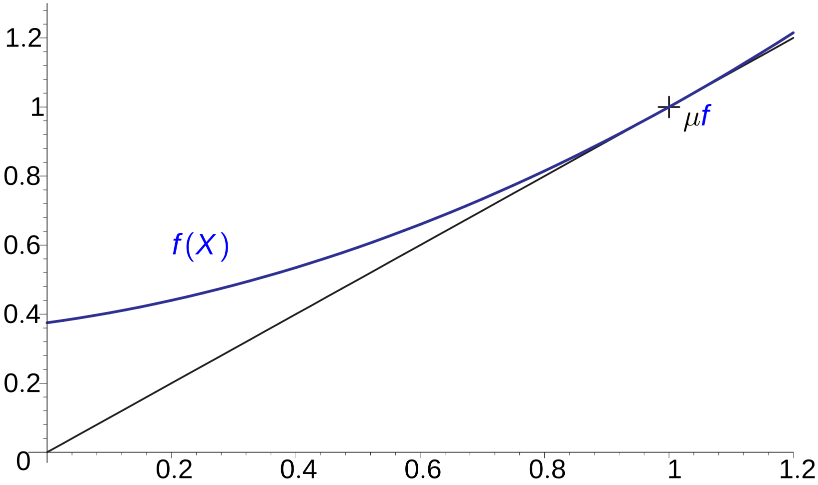
The least solution of $X = 0.5X^2 + 0.5$ is $1 = 0.999\cdots$.
The Kleene sequence needs k iterations for about $\log k$ bits:

k	$f^k(0)$
0	0.0000
1	0.5000
2	0.6250
3	0.6953
4	0.7417

k	$f^k(0)$
20	0.9200
200	0.9900
2000	0.9990

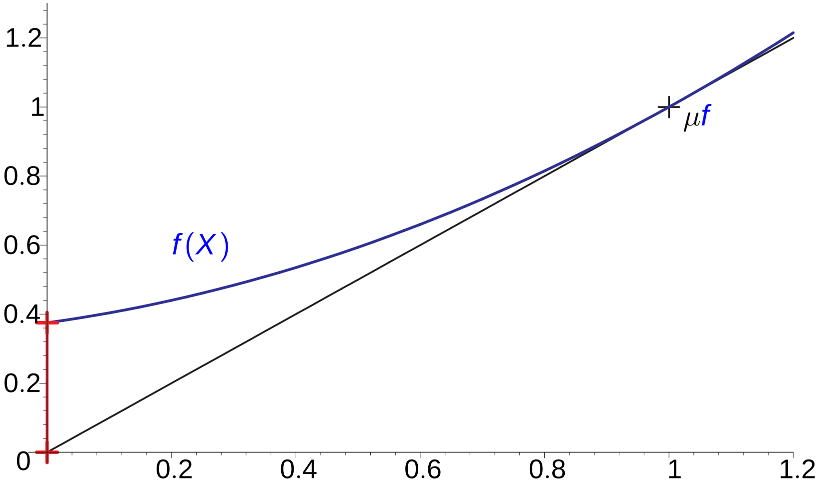
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



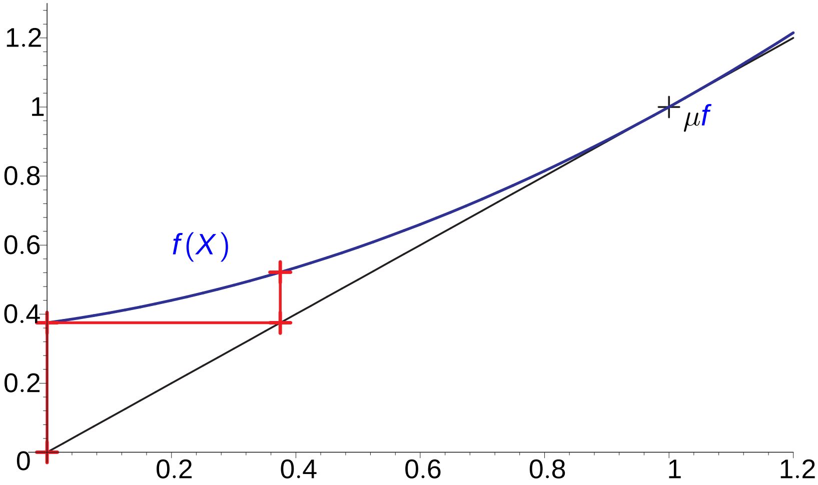
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



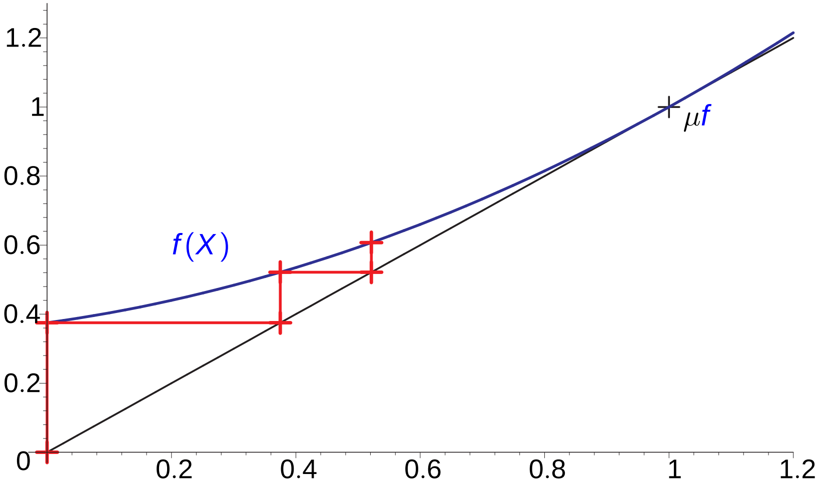
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



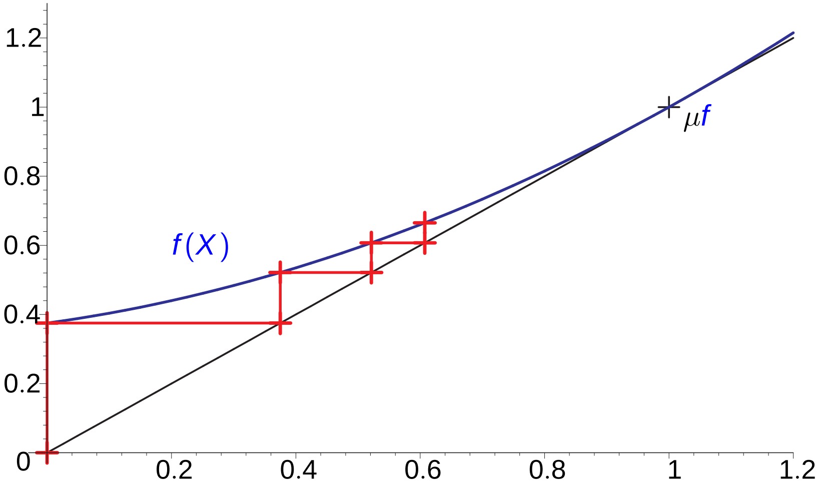
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



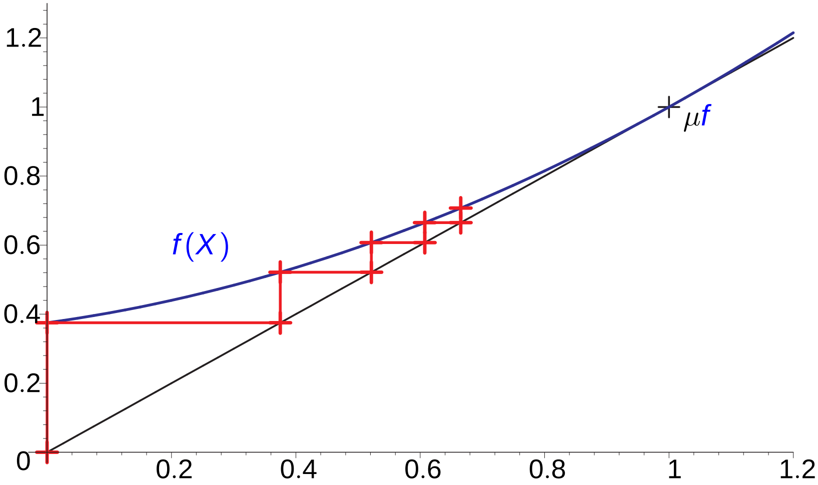
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



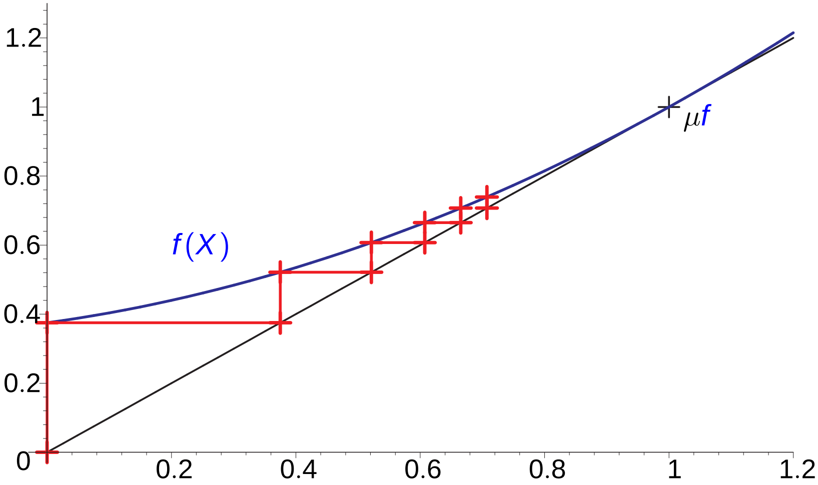
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



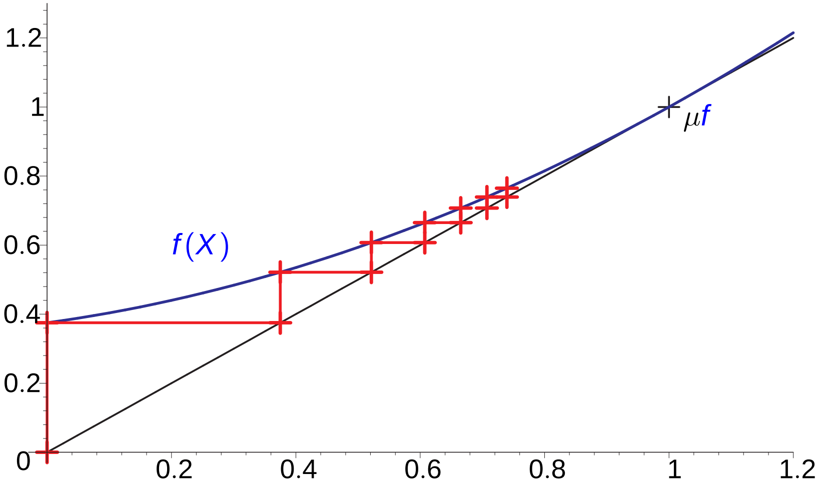
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



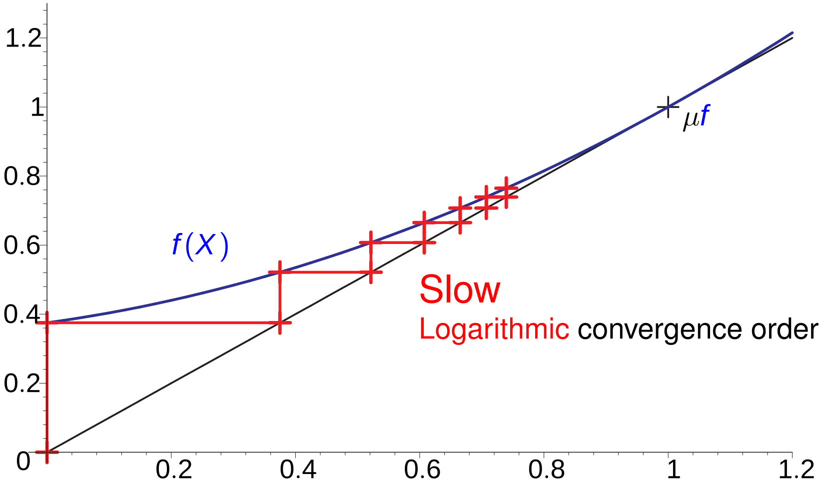
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



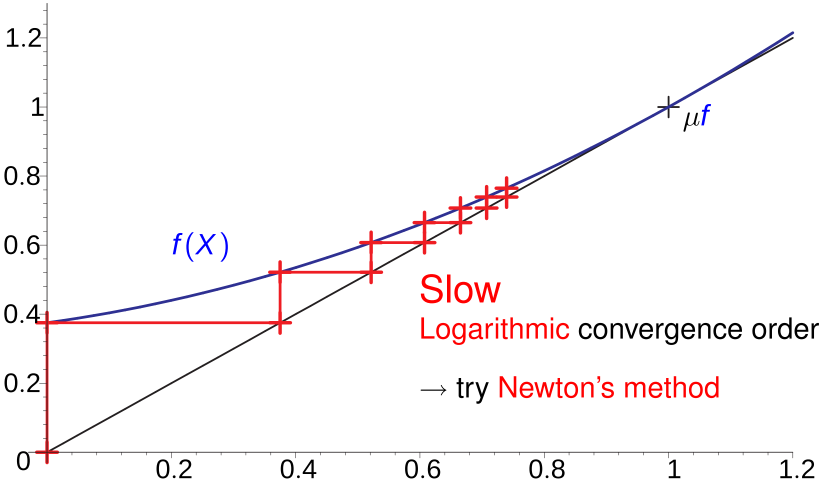
Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



Kleene Iteration (univariate case)

Consider $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



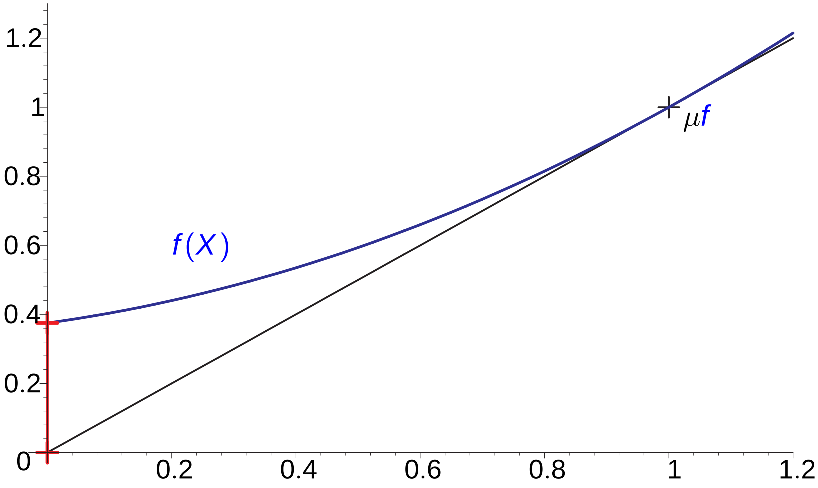
Slow

Logarithmic convergence order

→ try Newton's method

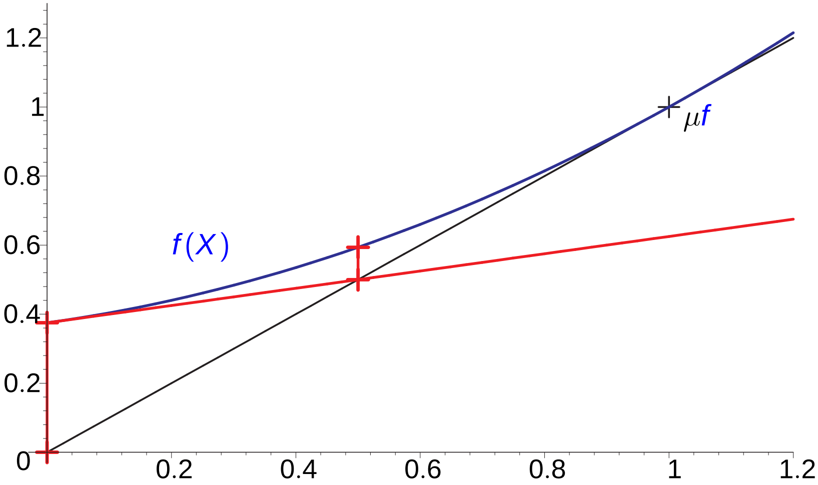
Newton's Method (univariate case)

Consider $X = f(X)$ with $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



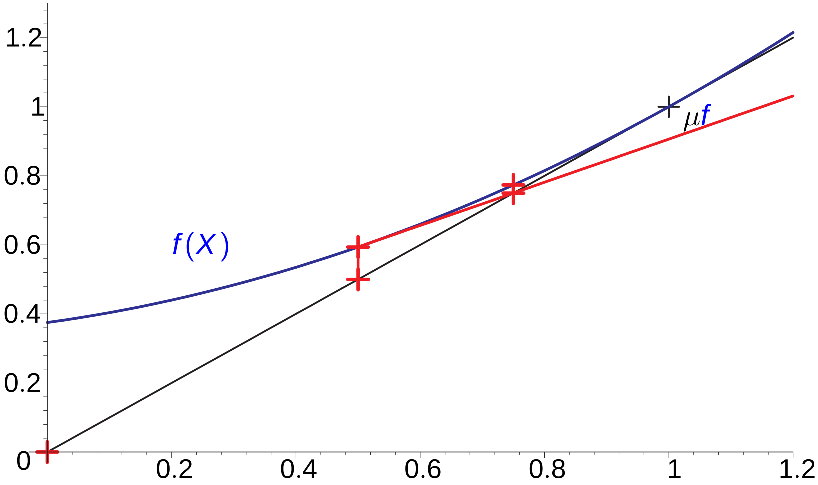
Newton's Method (univariate case)

Consider $X = f(X)$ with $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



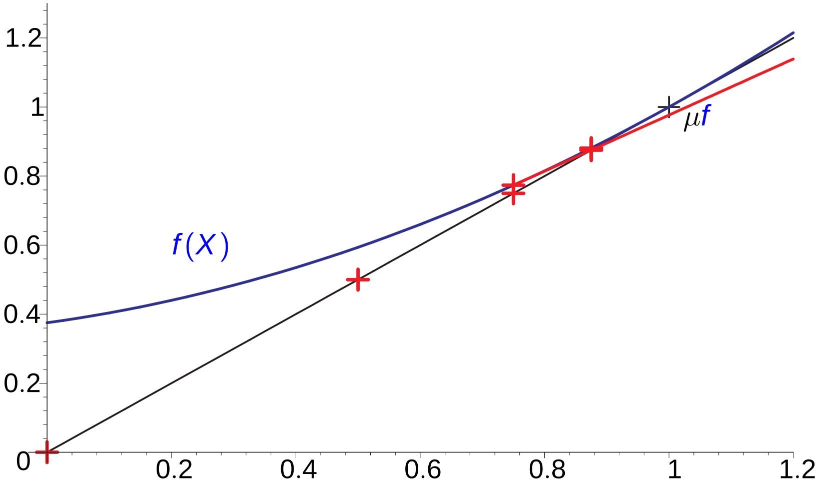
Newton's Method (univariate case)

Consider $X = f(X)$ with $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



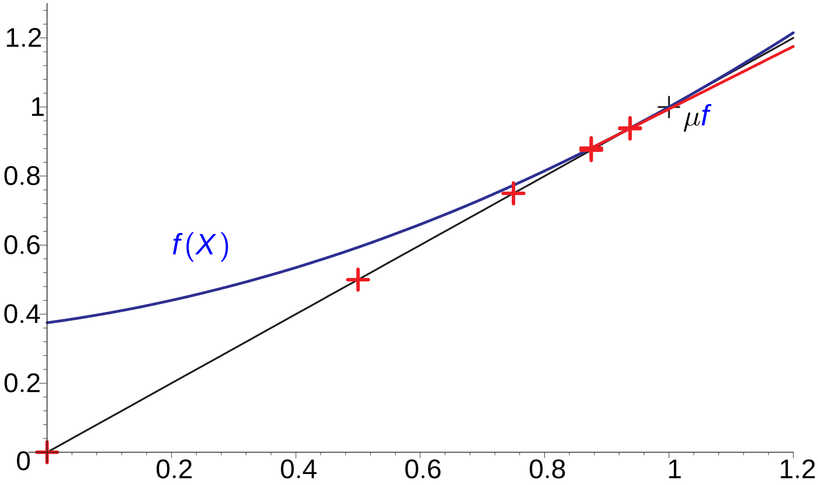
Newton's Method (univariate case)

Consider $X = f(X)$ with $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



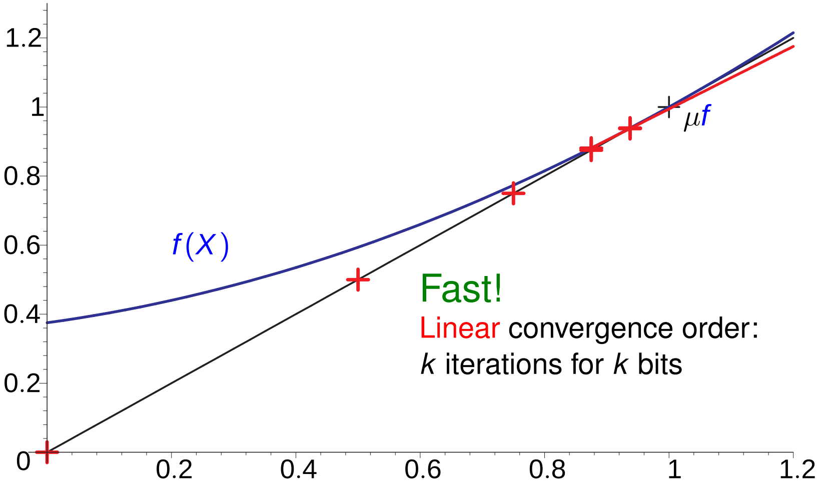
Newton's Method (univariate case)

Consider $X = f(X)$ with $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



Newton's Method (univariate case)

Consider $X = f(X)$ with $f(X) = \frac{3}{8}X^2 + \frac{1}{4}X + \frac{3}{8}$



Fast!

Linear convergence order:

k iterations for *k* bits

Mathematical formulation (univariate case)

Let $X = f(X)$ be a monotonic equation and let ν be some approximation of μf.

Newton's method gets a better approximation ν' as follows:

1. Compute the tangent of f at ν:

$$Y = f(\nu) + f'(\nu) \cdot (X - \nu)$$

2. Take ν' as its intersection with the straight line $Y = X$:

$$\nu' := \nu + \frac{f(\nu) - \nu}{1 - f'(\nu)}$$

Generalization to the multivariate case

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be an MSPE and let ν be some approximation of $\mu\boldsymbol{f}$.

We get a better approximation ν' as follows:

$$\nu' := \nu + (\mathrm{Id} - \boldsymbol{f}'(\nu))^{-1}(\boldsymbol{f}(\nu) - \nu)$$

where

- $\boldsymbol{f}'$ is the Jacobian of $\boldsymbol{f}$, i.e., the matrix of partial derivatives of $\boldsymbol{f}$, and
- Id is the identity matrix.

Our probabilistic program again . . .

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4XY + 0.6 \\ 0.3XY + 0.4YZ + 0.3 \\ 0.3XZ + 0.7 \end{pmatrix}$$

k	$(\boldsymbol{f}^k(0))_X$	$(\boldsymbol{f}^k(0))_Y$	$(\boldsymbol{f}^k(0))_Z$	$\nu_X^{(k)}$	$\nu_Y^{(k)}$	$\nu_Z^{(k)}$
0	0.000	0.000	0.000	0.000	0.000	0.000
4	0.753	0.600	0.887	0.933	0.899	0.972
8	0.834	0.738	0.926	0.983	0.974	0.993
12	0.873	0.802	0.944	0.983	0.974	0.993
16	0.897	0.839	0.955	0.983	0.974	0.993

Our probabilistic program again . . .

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4XY + 0.6 \\ 0.3XY + 0.4YZ + 0.3 \\ 0.3XZ + 0.7 \end{pmatrix}$$

k	$(\boldsymbol{f}^k(0))_X$	$(\boldsymbol{f}^k(0))_Y$	$(\boldsymbol{f}^k(0))_Z$	$\nu_X^{(k)}$	$\nu_Y^{(k)}$	$\nu_Z^{(k)}$
0	0.000	0.000	0.000	0.000	0.000	0.000
4	0.753	0.600	0.887	0.933	0.899	0.972
8	0.834	0.738	0.926	0.983	0.974	0.993
12	0.873	0.802	0.944	0.983	0.974	0.993
16	0.897	0.839	0.955	0.983	0.974	0.993

Is the solution $\mu\boldsymbol{f} = (1, 1, 1)$? Probably no, but we don't have a proof.

And perhaps we've just been lucky with the example!

Mathematicians on Newton's method

Studied by mathematicians for general systems $\boldsymbol{f}(\boldsymbol{X}) = \mathbf{0}$.

Mathematicians on Newton's method

Studied by mathematicians for general systems $\boldsymbol{f}(\boldsymbol{X}) = \mathbf{0}$.

The method may perform brilliantly (exponential convergence order), but it is fragile. It may:

- be ill defined ($(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(i)}))$ may be singular);
- diverge;
- converge only in a small neighbourhood of the solution (local convergence); or
- converge as slowly as Kleene iteration.

Mathematicians on Newton's method

Studied by mathematicians for general systems $\boldsymbol{f}(\boldsymbol{X}) = \mathbf{0}$.

The method may perform brilliantly (exponential convergence order), but it is fragile. It may:

- be ill defined ($(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(i)}))$ may be singular);
- diverge;
- converge only in a small neighbourhood of the solution (local convergence); or
- converge as slowly as Kleene iteration.

Numerical mathematics has provided

- a few, restrictive sufficient conditions for global exponential convergence (Kantorovich's theorem), and
- miscellaneous conditions for local exponential convergence (often expensive or impossible to check!).

Computer scientists on Newton's method

MSPEs are important in computer science.

Is Newton's method robust for MPSEs?

Can we find guarantees on the convergence order?

Computer scientists on Newton's method

MSPEs are important in computer science.

Is Newton's method robust for MPSEs?

Can we find guarantees on the convergence order?

Next slides: results on this question obtained by Etessami and Yannakakis and by us since 2005.

Global convergence for MSPEs [EY'05,EKL'07]

Proposition

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be an MSPE. The Newton sequence $\mathbf{0} = \boldsymbol{\nu}^{(0)}, \boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}, \ldots$ is

- *well defined (the inverses exist);*
- *monotonically increasing, i.e., $\boldsymbol{\nu}^{(i)} \leq \boldsymbol{\nu}^{(i+1)}$;*
- *bounded from above by $\mu\boldsymbol{f}$, i.e, $\boldsymbol{\nu}^{(i)} \leq \mu\boldsymbol{f}$;*
- *converges to $\mu\boldsymbol{f}$; and*
- *converges at least as fast as the Kleene sequence, i.e., $\boldsymbol{f}^i(\mathbf{0}) \leq \boldsymbol{\nu}^{(i)}$.*

Best case

Theorem (easy to prove)

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be a MSPE.
If the matrix $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))$ is non-singular, then the Newton sequence has exponential convergence order.

Best case

Theorem (easy to prove)

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be a MSPE.
If the matrix $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))$ is non-singular, then the Newton sequence has exponential convergence order.

However, since $\mu\boldsymbol{f}$ is what we wish to compute, the condition is not very useful!

A guarantee of linear convergence

Theorem (KLE STOC'07)

The Newton sequence has linear convergence order for arbitrary MSPEs.

A guarantee of linear convergence

Theorem (KLE STOC'07)

The Newton sequence has linear convergence order for arbitrary MSPEs.

But: this only shows $\beta(k) = a \cdot k + b$ for some a and b.
It says nothing about how big or small a and b are!

Definitions

An MSPE is called

- strongly connected
  if every variable depends transitively on every variable.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4XY + 0.6 \\ 0.3XY + 0.4YZ + 0.3 \\ 0.3XZ + 0.7 \end{pmatrix}$$

- fully inhomogeneous if $\boldsymbol{f}(\mathbf{0}) > \mathbf{0}$ (in all components).

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4XY + 0.6 \\ 0.3XY + 0.4YZ + 0.3 \\ 0.3XZ + 0.7 \end{pmatrix}$$

A threshold for strongly connected MSPEs

Theorem (KLE STOC'07)

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be a strongly connected MSPE.
There is a threshold t (depending on $\boldsymbol{f}$ such that for every $i \geq 0$ the Newton sequence satisfies

$$\beta(t+i) \geq i \ .$$

That is: after t iterations we are guaranteed at least one bit of accuracy for each new iteration. We say that the method has linear convergence order with convergence rate 1.

However, the proof is based on a purely topological property of $\mathbb{R}^n$. Again, it only proves that t exists!

Bounds on the threshold

Theorem (KLE STOC'07)

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be a strongly connected MSPE.
There is a threshold t (depending on $\boldsymbol{f}$) such that for every $i \geq 0$ the Newton sequence satisfies:

$$\beta(t + i) \geq i \, .$$

Theorem (EKL STACS'08)

Above theorem holds with $t = 3n^2(m + |\log \mu_{min}|)$, 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 $\mu\boldsymbol{f}$.*

For fully inhomogeneous MSPEs even better: $t = 3nm$.

But you're cheating!

Isn't it still useless? We do not know μ_{min}!

But you're cheating!

Isn't it still useless? We do not know μ_{min}!

Yes, but we can compute bounds for it

- either syntactic ones, or, better,
- dynamic ones, updated as the computation progresses.

Our probabilistic program again . . .

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4XY + 0.6 \\ 0.3XY + 0.4YZ + 0.3 \\ 0.3XZ + 0.7 \end{pmatrix}$$

After 14 Newton steps we got earlier:

$$\boldsymbol{\nu}^{(14)} = (0.98283\cdots, 0.97380\cdots, 0.99269\cdots)$$

Is the solution $\mu\boldsymbol{f} = (1, 1, 1)$?

Our probabilistic program again . . .

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4XY + 0.6 \\ 0.3XY + 0.4YZ + 0.3 \\ 0.3XZ + 0.7 \end{pmatrix}$$

After 14 Newton steps we got earlier:

$$\boldsymbol{\nu}^{(14)} = (0.98283\cdots, 0.97380\cdots, 0.99269\cdots)$$

Is the solution $\mu \boldsymbol{f} = (1, 1, 1)$? No!
The MSPE is strongly connected, and $0.97380 \leq \mu_{\text{min}}$.
Our theorem proves that the error after 14 iterations is at most 0.004 (8 bits). So:

$$\mu \boldsymbol{f} \leq \boldsymbol{\nu}^{(14)} + \begin{pmatrix} 0.004 \\ 0.004 \\ 0.004 \end{pmatrix} \leq \begin{pmatrix} 0.987 \\ 0.978 \\ 0.997 \end{pmatrix} < \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Non-strongly-connected MSPEs

$$
\begin{aligned}
X_1 &= 1/2 + 1/2 \cdot X_1^2 \\
X_2 &= 1/4 \cdot X_1^2 + 1/2 \cdot X_1 X_2 + 1/4 \cdot X_2^2 \\
&\vdots \\
X_n &= 1/4 \cdot X_{n-1}^2 + 1/2 \cdot X_{n-1} X_n + 1/4 \cdot X_n^2
\end{aligned}
$$

The least fixed-point of the system is $(1, 1, \ldots, 1)$.

We have $\nu_n^{(2^{n-1})} \leq 1/2$, and so that at least 2^{n-1} iterations of Newton's method are needed to obtain the first bit of X_n [KLE STOC'07].

The method still has linear convergence order, but a worse convergence rate.

Non-strongly-connected MSPEs

Theorem (KLE STOC'07)

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be a MSPE.
There is a threshold t such that for every $i \geq 0$ the Newton sequence satisfies:

$$\beta(t + i \cdot (h+1) \cdot 2^h) \geq i \ .$$

where h is the height of the graph of strongly connected components.

Conclusions

- Solving MSPEs is central to several computer science problems.

Conclusions

- Solving MSPEs is central to several computer science problems.
- The associated decision problem can be very important to understand the unit-cost model.

Conclusions

- Solving MSPEs is central to several computer science problems.
- The associated decision problem can be very important to understand the unit-cost model.
- Newton's Method works very well
  for approximating least solutions of MSPEs.

Conclusions

- Solving MSPEs is central to several computer science problems.
- The associated decision problem can be very important to understand the unit-cost model.
- 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.

Conclusions

- Solving MSPEs is central to several computer science problems.
- The associated decision problem can be very important to understand the unit-cost model.
- 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.
- Thresholds give guarantee that linear convergence has kicked in.

Conclusions

- Solving MSPEs is central to several computer science problems.
- The associated decision problem can be very important to understand the unit-cost model.
- 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.
- Thresholds give guarantee that linear convergence has kicked in.
- Far stronger results than for general systems.

Conclusions

- Solving MSPEs is central to several computer science problems.
- The associated decision problem can be very important to understand the unit-cost model.
- 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.
- Thresholds give guarantee that linear convergence has kicked in.
- Far stronger results than for general systems.
- In the paper: extension to min-max MSPEs [EKL ICALP'08].

Back in victorian Britain . . .

There was concern amongst the Victorians that aristocratic families were becoming extinct.

Francis Galton (1822-1911), anthropologist and polymath:
Are families of English peers more likely to die out than the families of ordinary men?

> *Let $p_0, p_1, \ldots, p_n$ be the respective probabilities that a man has 0, 1, 2, . . . n sons, let each son have the same probability for sons of his own, and so on. What is the probability that the male line goes extinct?*

Henry William Watson (1827-1903), priest and mathematician:
The probability is the least solution of

$$X = p_0 + p_1 X + p_2 X^2 + \ldots + p_n X^n$$

English peers again . . .

Watson concluded wrongly (due to an algebraic error) that all families eventually die out.

But Galton found a fact, that, with hindsight, provides a possible explanation for the data:

- English peers tended to marry heiresses (daughters without brothers)
- Heiresses come from families without sons, and so perhaps, by inheritance, with lower fertility rates (lower probabilities $p_2, p_3, \ldots$).
- . . . which increases the probability of the family dying out.