A computer science look at stochastic branching processes

T. Brázdil J. Esparza S. Kiefer M. Luttenberger

Technische Universität München

For Thiagu, December 2008

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, and what is the probability that it goes extinct after r generations?*

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, and what is the probability that it goes extinct after r generations?*

Henry William Watson (1827-1903), vicar and mathematician:
The probability that the line goes extinct is the least solution of

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

Stochastic branching theory

Stochastic branching processes (SBPs)

Stochastic processes that model the behaviour of populations whose individuals die and reproduce.

Models of:

- reproduction of biological species,
- evolution of gene pools,
- chemical and nuclear reactions.

Stochastic branching theory

Stochastic branching processes (SBPs)

Stochastic processes that model the behaviour of populations whose individuals die and reproduce.

Models of:

- reproduction of biological species,
- evolution of gene pools,
- chemical and nuclear reactions.

Work in progress

Investigate SBPs as models of execution threads, OS tasks, computer viruses, information spread in social networks . . .

A classification of SBPs

Two classical dimensions

Single-type/Multi-type
(one/several "subspecies" with different offspring probabilities).

Untimed/Timed
(no time information/stochastic lifetime).

A classification of SBPs

Two classical dimensions

Single-type/Multi-type
(one/several "subspecies" with different offspring probabilities).

Untimed/Timed
(no time information/stochastic lifetime).

A new dimension for CS systems

Synchronous/Asynchronous
(generation moves/individuals move).

Random variables of interest

Past research has studied mostly . . .

- untimed synchronous systems: probability of extinction, population of the n-th generation.
- timed asynchronous systems: population at time t.

Random variables of interest

Past research has studied mostly . . .

- untimed synchronous systems: probability of extinction, population of the n-th generation.
- timed asynchronous systems: population at time t.

CS is interested in ressource consumption

- Time to termination (time to extinction).
- Maximal population size (memory or hardware consumption).

The fundamental equation (system)

A multi-type, untimed system

$$X \xrightarrow{0.4} X \parallel Y \qquad X \xrightarrow{0.6} \epsilon$$
$$Y \xrightarrow{0.3} X \parallel Y \qquad Y \xrightarrow{0.4} Y \parallel Z \qquad Y \xrightarrow{0.3} \epsilon$$
$$Z \xrightarrow{0.3} X \parallel Z \qquad Z \xrightarrow{0.7} \epsilon$$

The fundamental equation (system)

A multi-type, untimed system

$$X \xrightarrow{0.4} X \parallel Y \qquad X \xrightarrow{0.6} \epsilon$$
$$Y \xrightarrow{0.3} X \parallel Y \qquad Y \xrightarrow{0.4} Y \parallel Z \qquad Y \xrightarrow{0.3} \epsilon$$
$$Z \xrightarrow{0.3} X \parallel Z \qquad Z \xrightarrow{0.7} \epsilon$$

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

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

Probability of extinction

Observe

The probability of extinction is the same, independently of whether the system is synchronous or asynchronous.

Probability of extinction

Observe

The probability of extinction is the same, independently of whether the system is synchronous or asynchronous.

Theorem (well known)

The probability of extinction of the process types is equal to the least solution of the fundamental equation $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$.

Solving the fundamental equation

The least solution of the equation may be irrational:

Example

The least solution of

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

is irrational and not expressible by radicals.

We have $0.3357037075 < \mu f < 0.3357037076$

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 the least solution of $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$.

Example

For our multi-type system we get:

k	$\boldsymbol{f}_X^k(\mathbf{0})$	$\boldsymbol{f}_Y^k(\mathbf{0})$	$\boldsymbol{f}_Z^k(\mathbf{0})$
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

Convergence order

Definition: Convergence order

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 accurate digits of $a^{(k)}$.

Convergence order

Definition: Convergence order

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 accurate digits of $a^{(k)}$.

Example

Let: $a = 34,7815\cdots \quad a^{(0)} = 07,013\cdots \quad a^{(1)} = 34,7804\cdots$.
Then $\beta(0) = 0$ and $\beta(1) = 4$.

Convergence order

Definition: Convergence order

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 accurate digits of $a^{(k)}$.

Example

Let: $a = 34,7815 \cdots \quad a^{(0)} = 07,013 \cdots \quad a^{(1)} = 34,7804 \cdots$.
Then $\beta(0) = 0$ and $\beta(1) = 4$.

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

Convergence order

Definition: Convergence order

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 accurate digits of $a^{(k)}$.

Example

Let: $a = 34,7815\cdots \quad a^{(0)} = 07,013\cdots \quad a^{(1)} = 34,7804\cdots$.
Then $\beta(0) = 0$ and $\beta(1) = 4$.

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

Kleene Iteration is slow

The Kleene sequence may have logarithmic convergence order.

Example

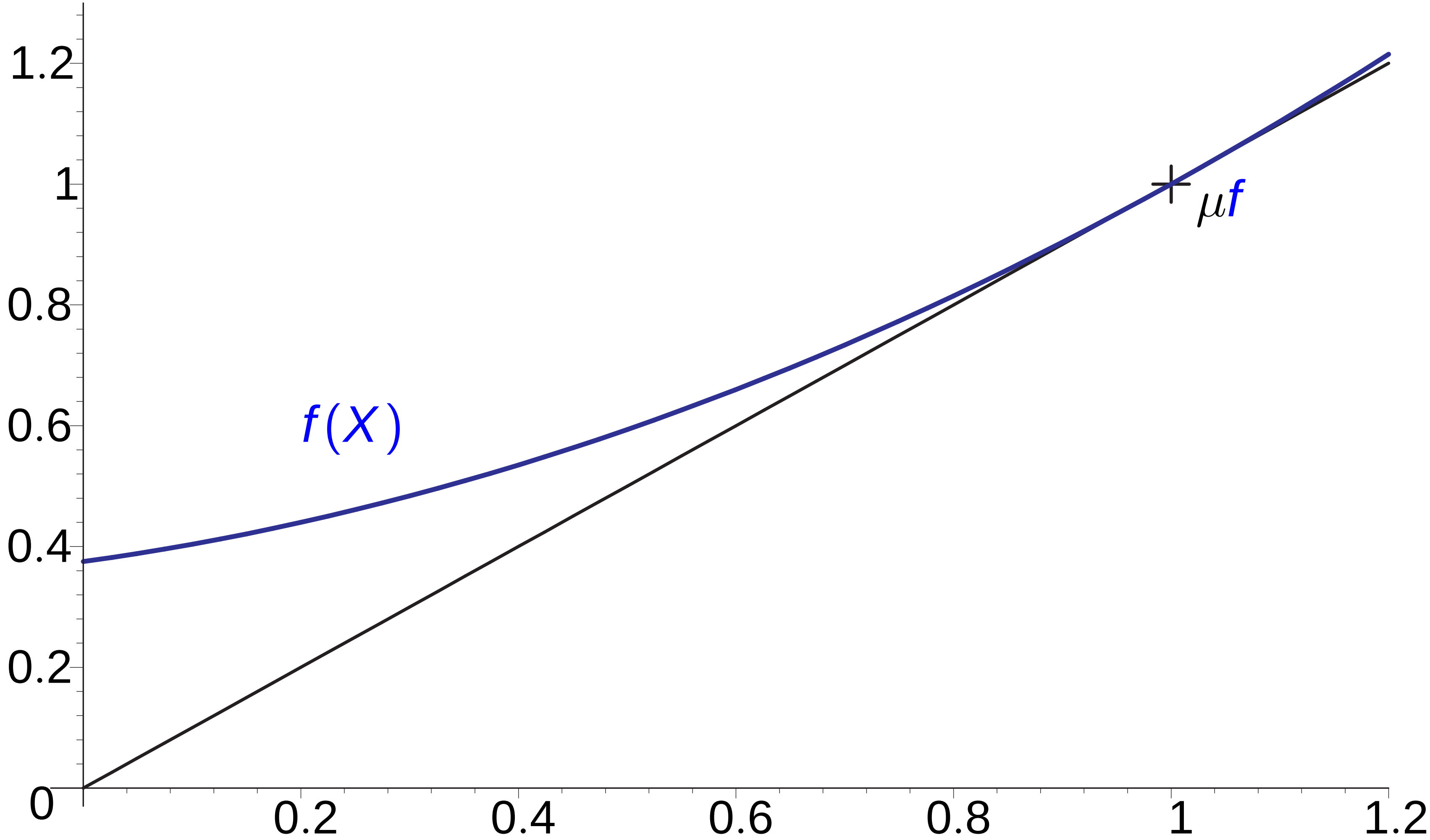
The least solution of $X = 0.5X^2 + 0.5$ is $1 = 0.999\cdots$.
The Kleene sequence needs k iterations for about $\log k$ digits:

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

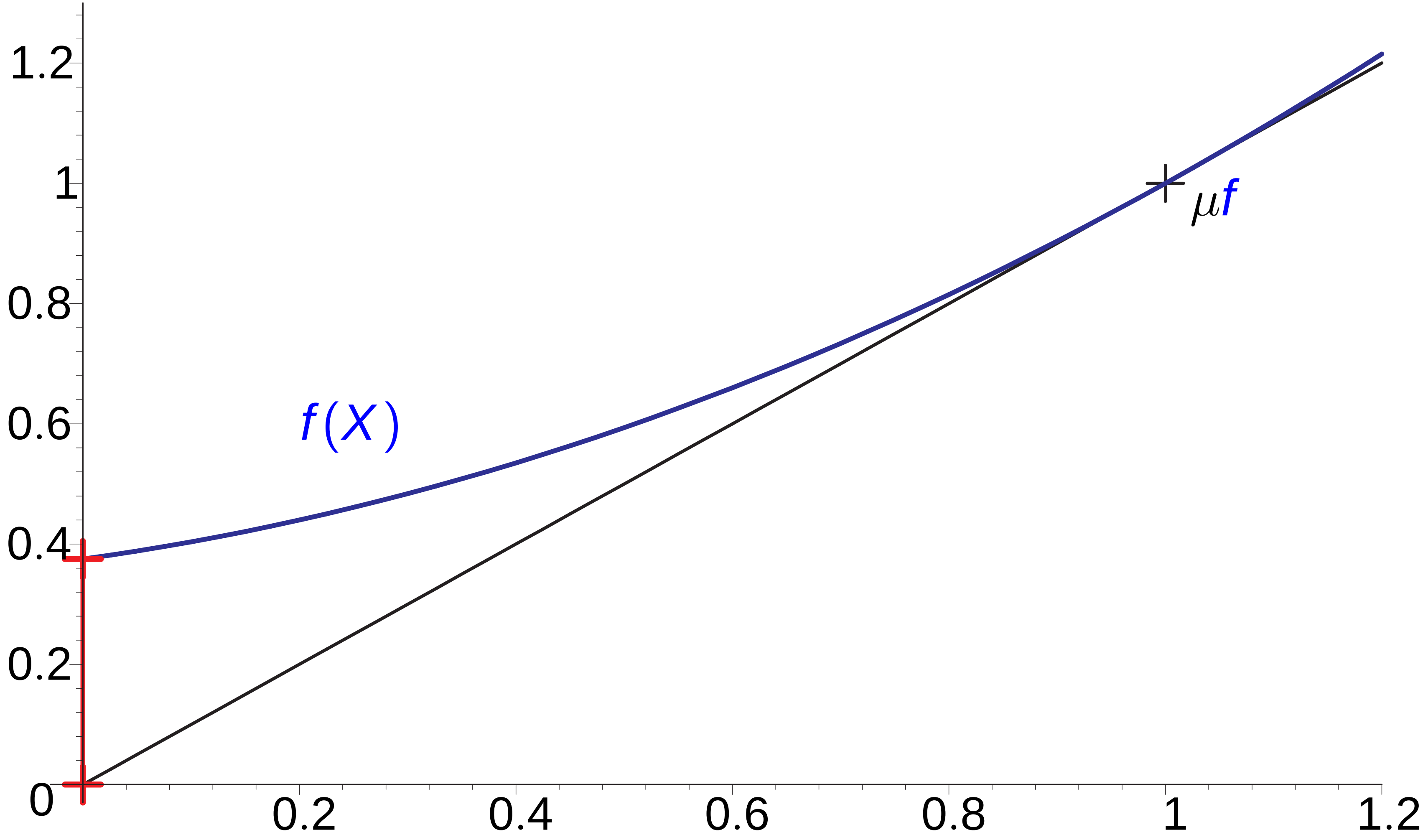
Kleene Iteration (univariate case)

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



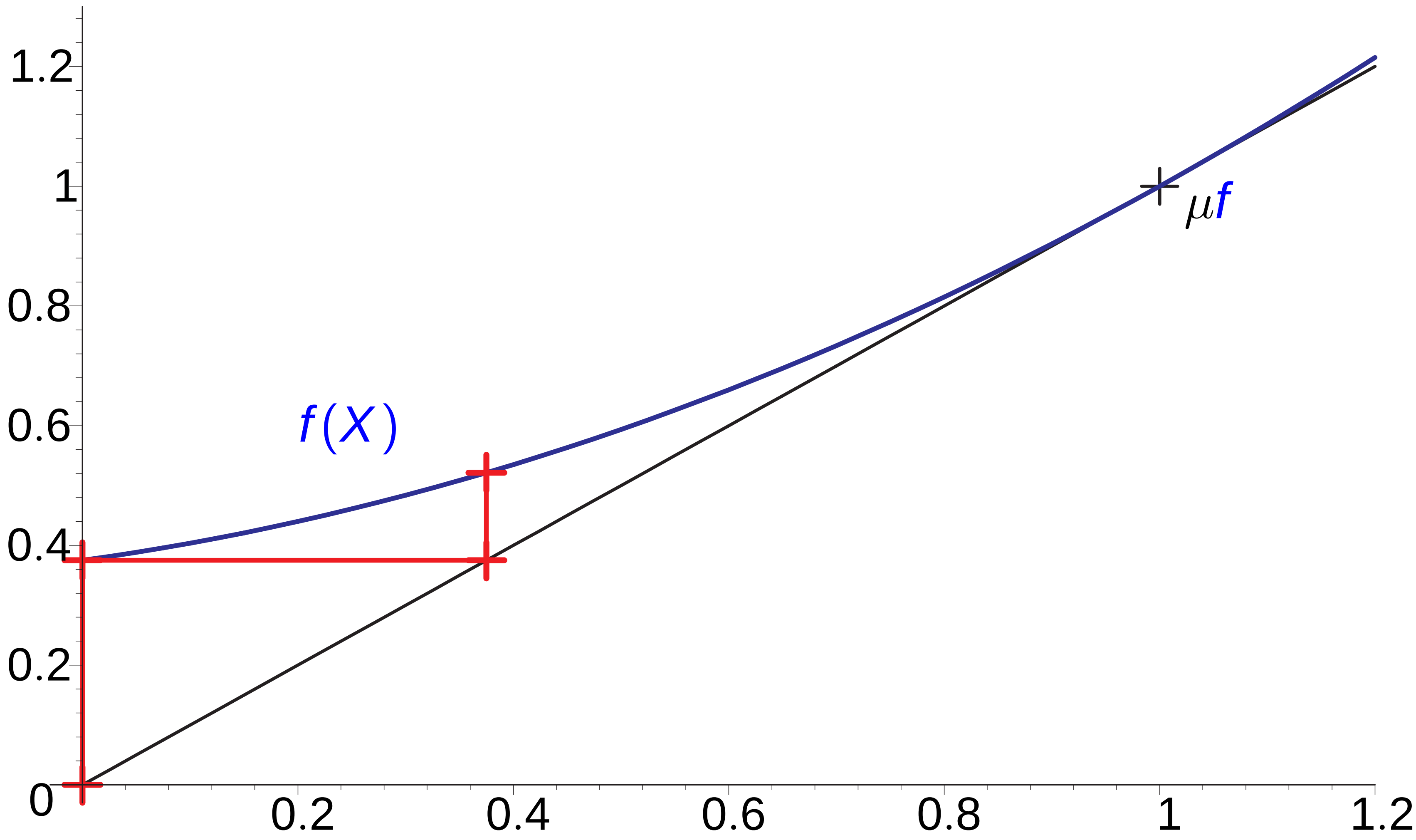
Kleene Iteration (univariate case)

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



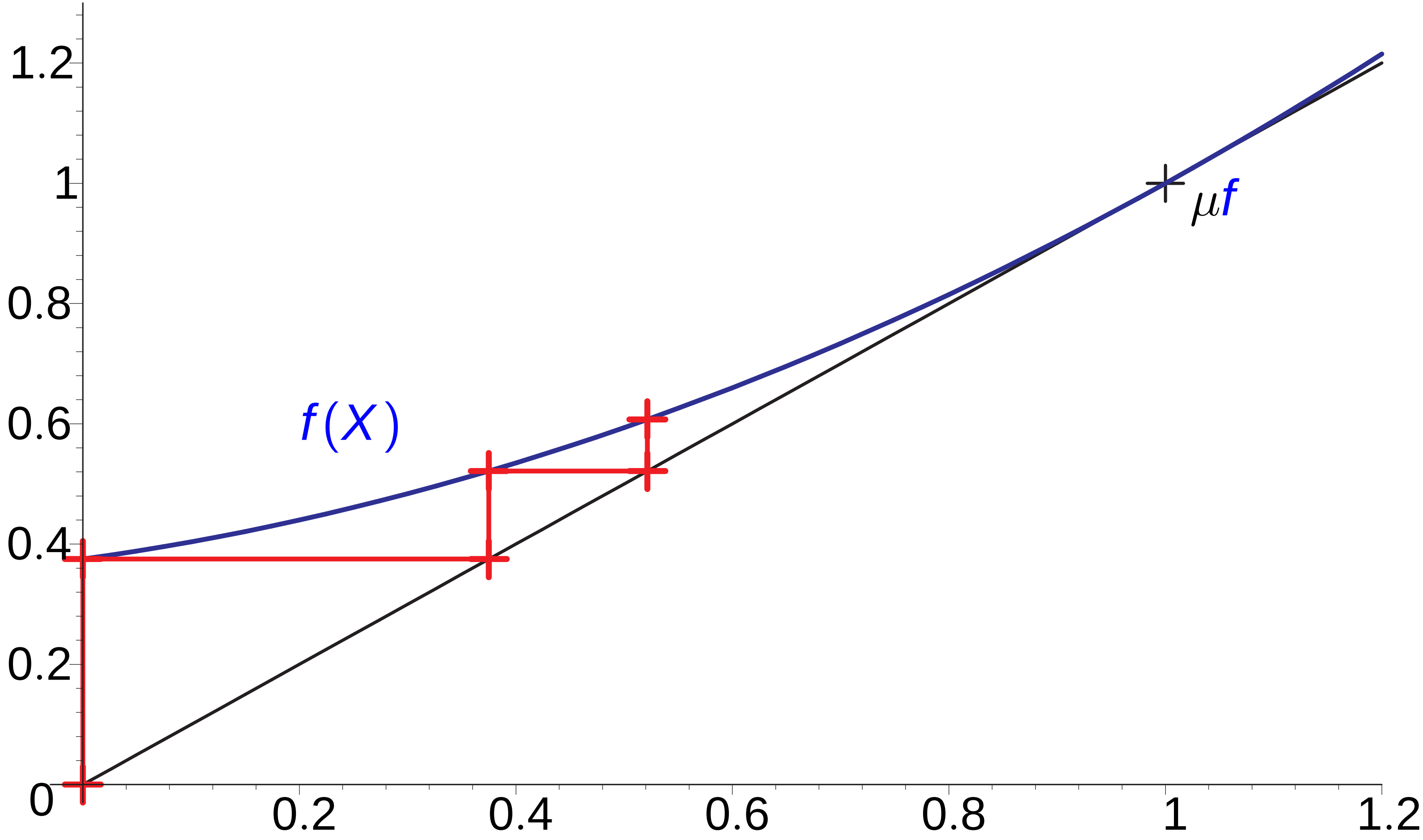
Kleene Iteration (univariate case)

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



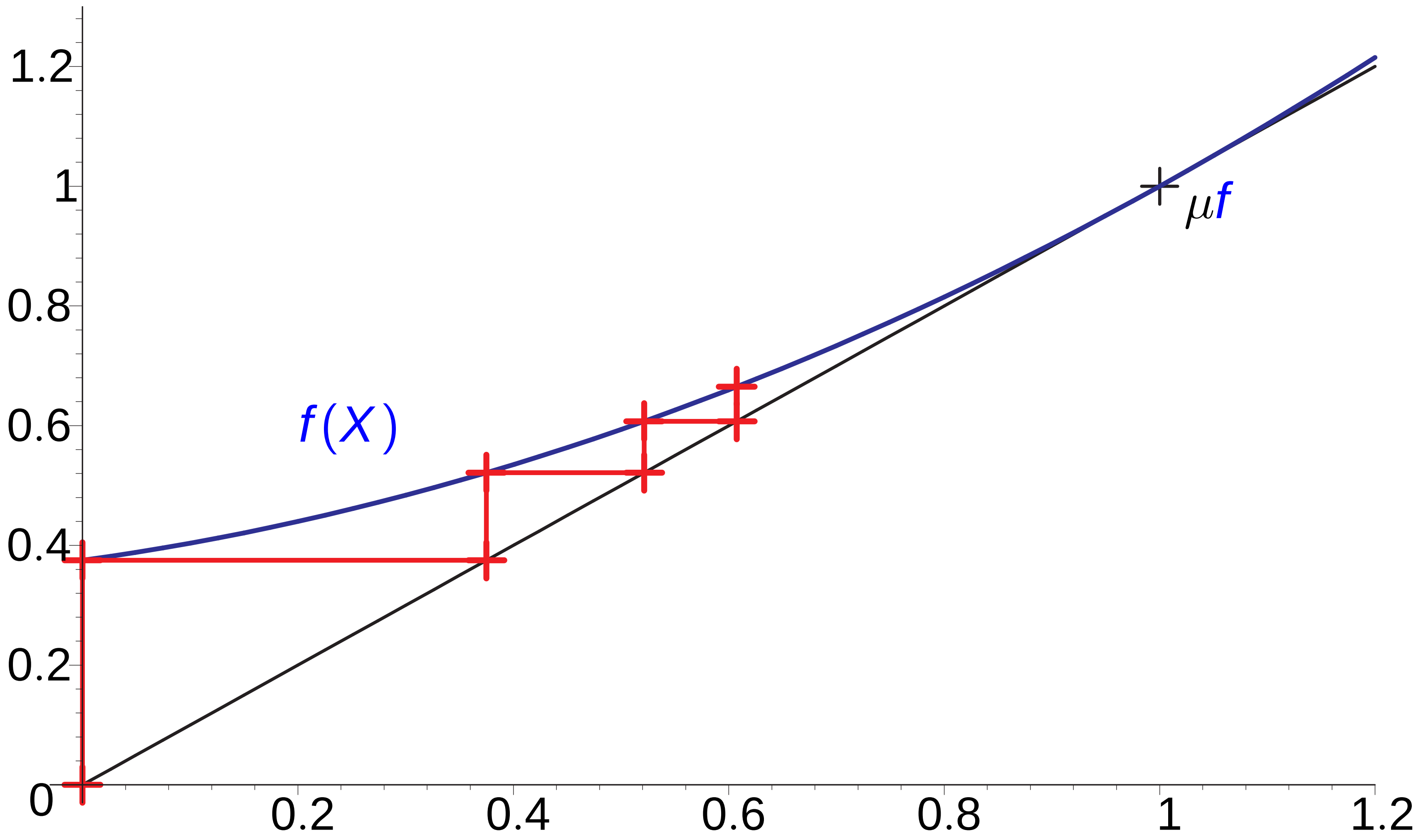
Kleene Iteration (univariate case)

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



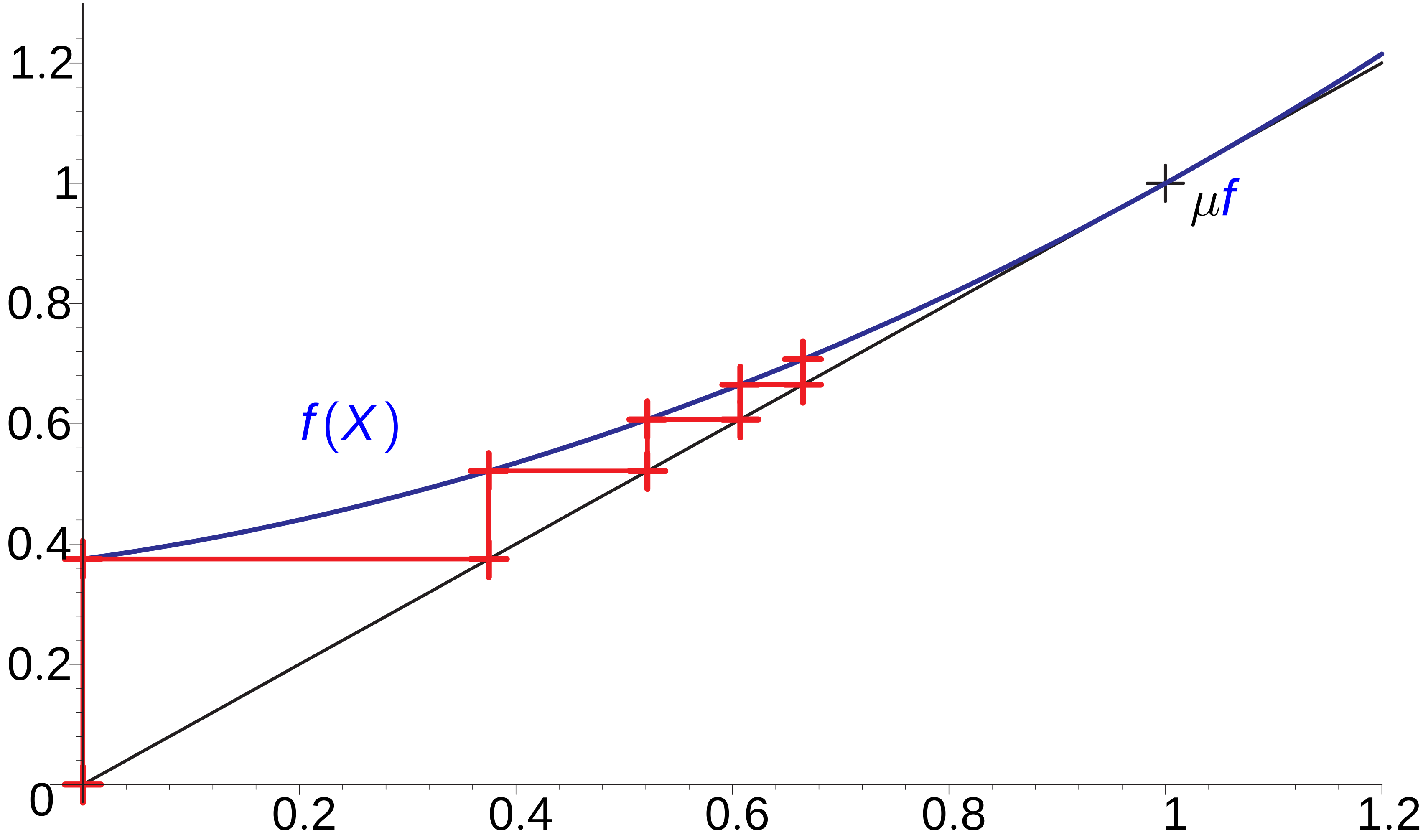
Kleene Iteration (univariate case)

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



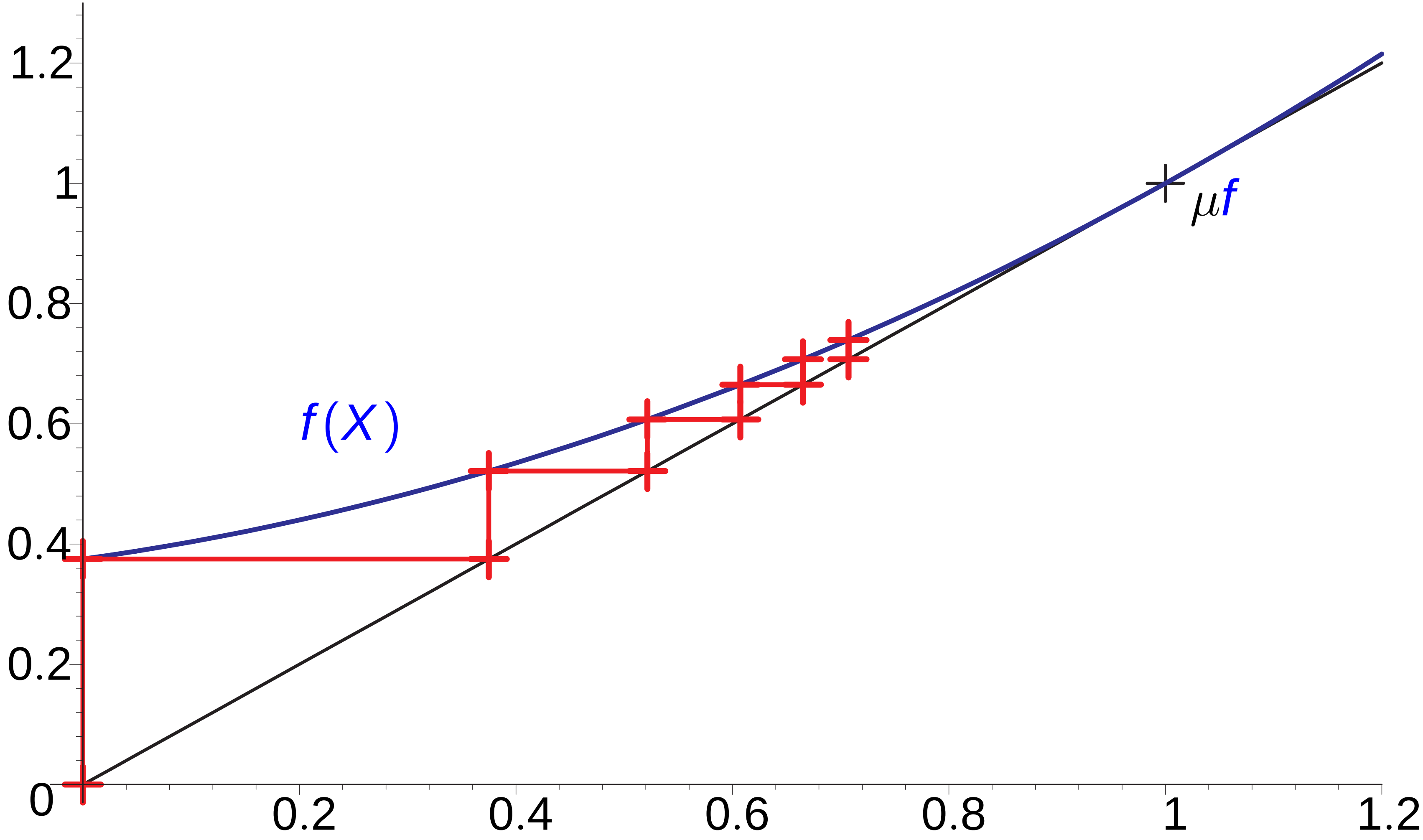
Kleene Iteration (univariate case)

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



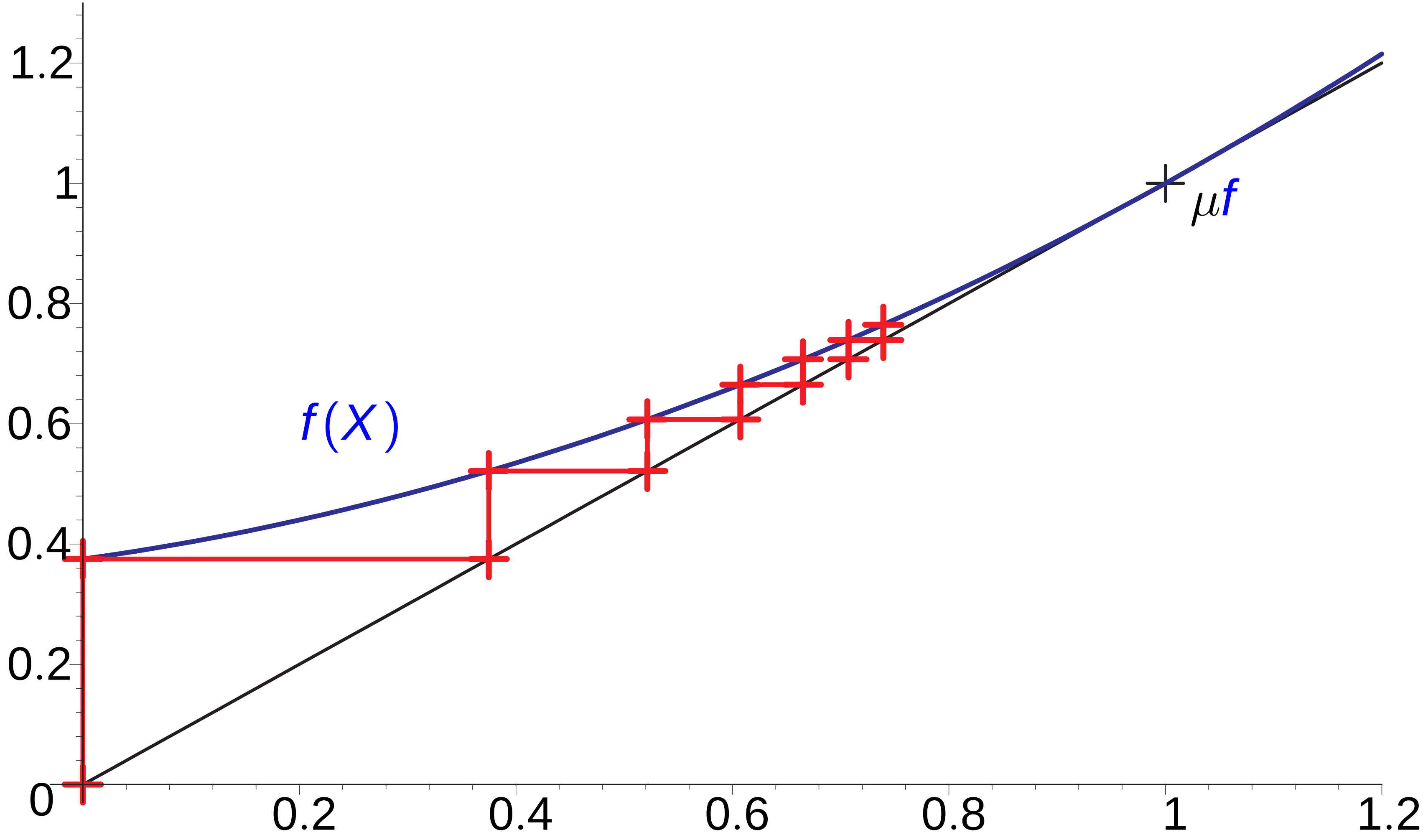
Kleene Iteration (univariate case)

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



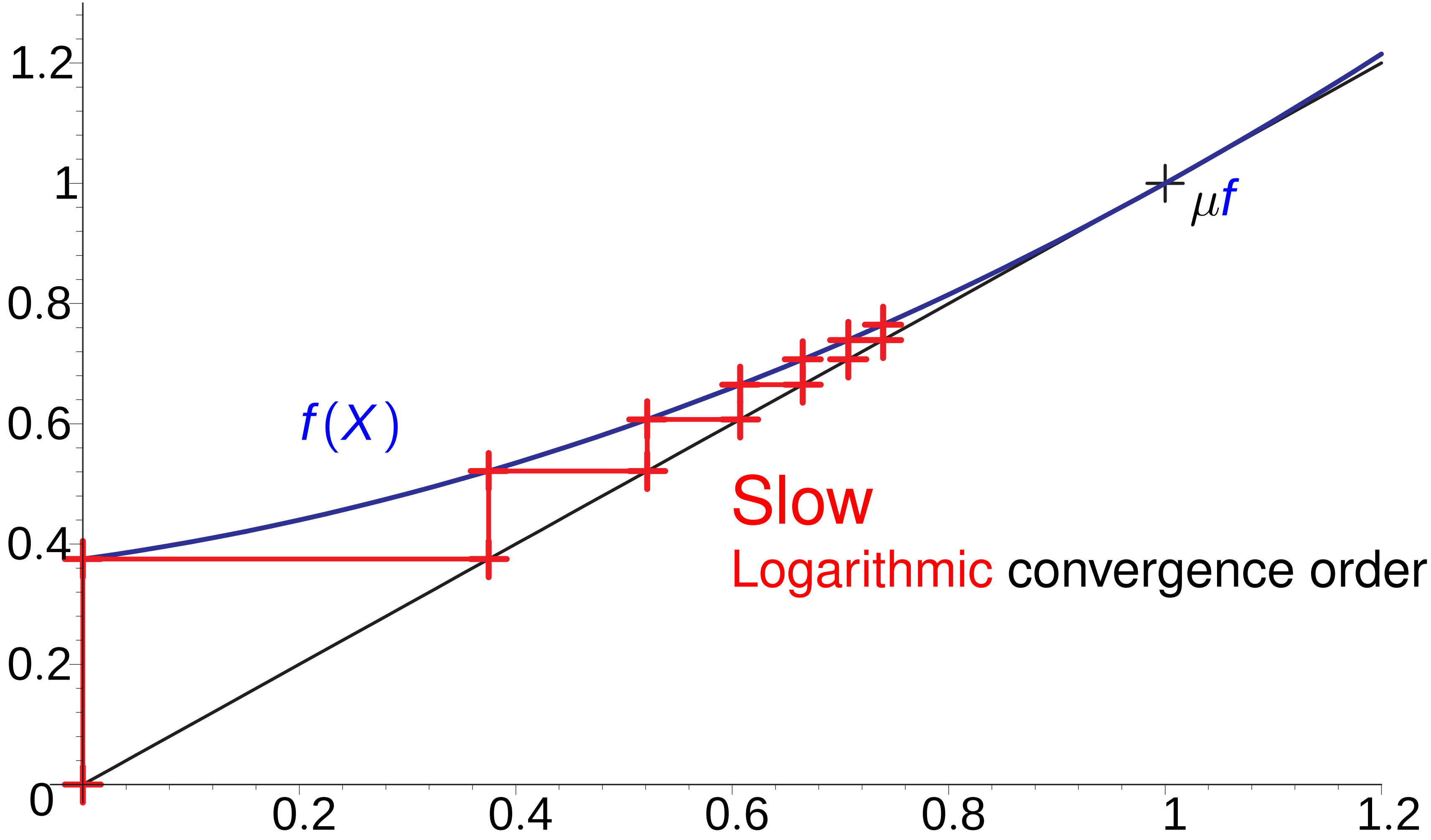
Kleene Iteration (univariate case)

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



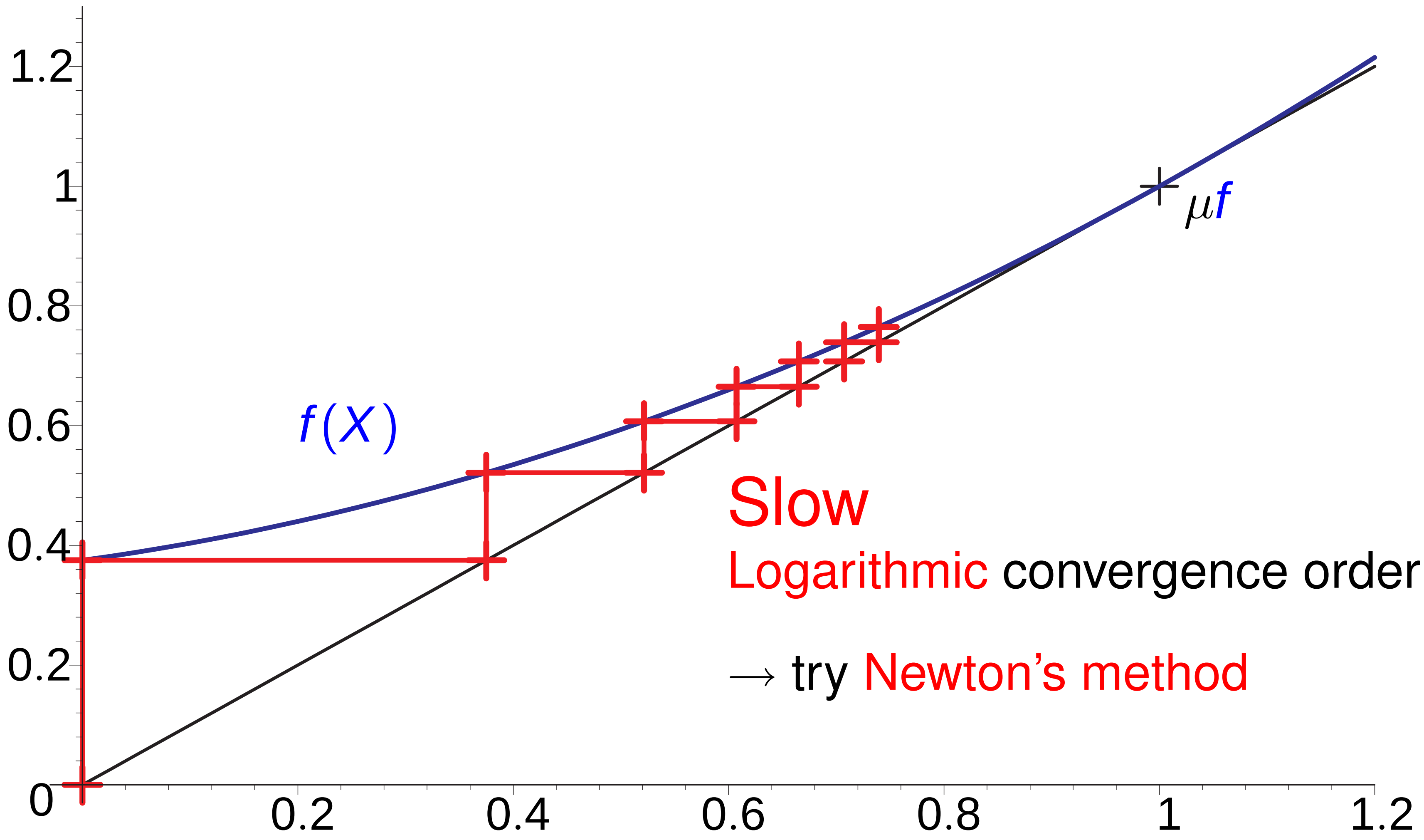
Kleene Iteration (univariate case)

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



Kleene Iteration (univariate case)

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



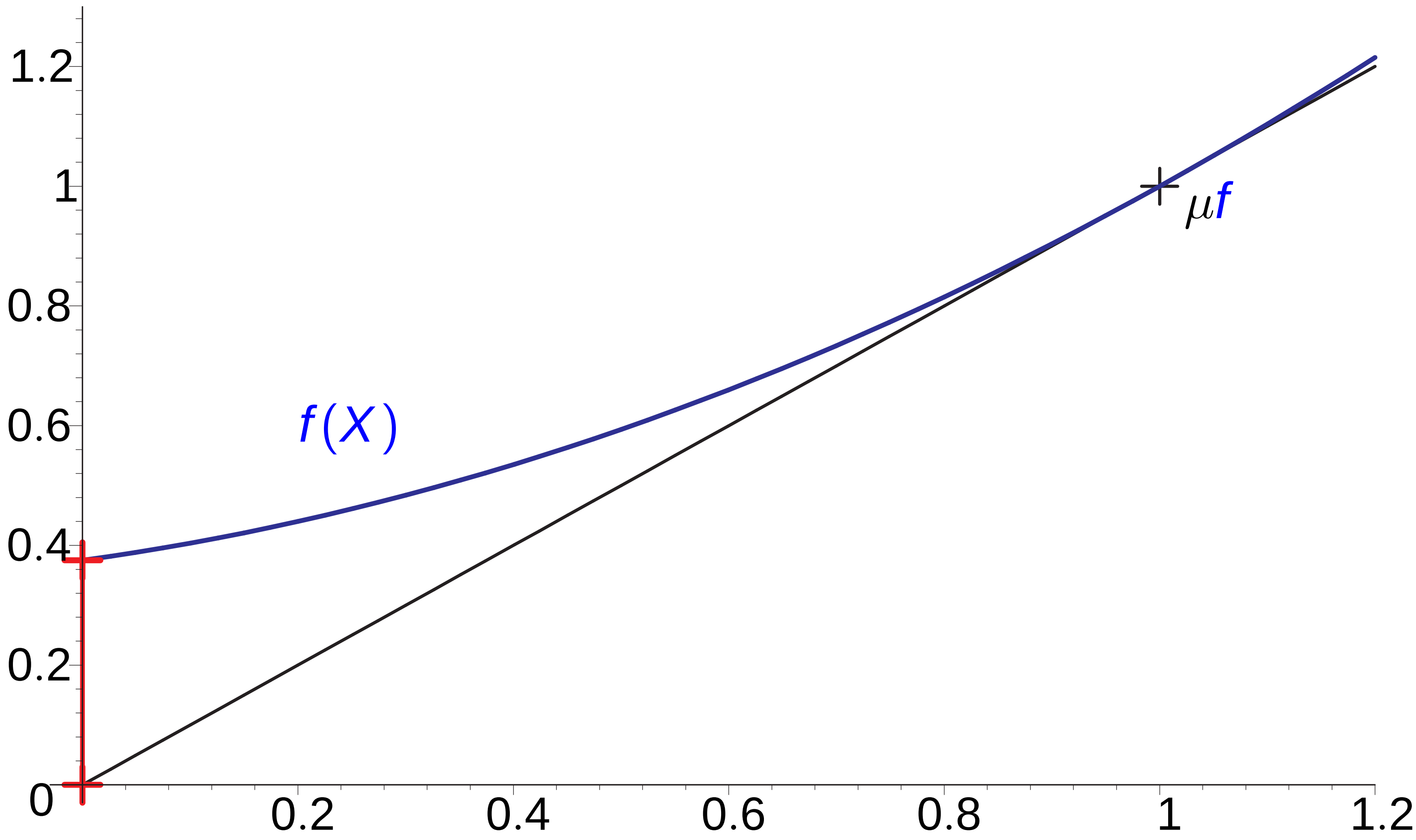
Slow

Logarithmic convergence order

$\longrightarrow$ try Newton's method

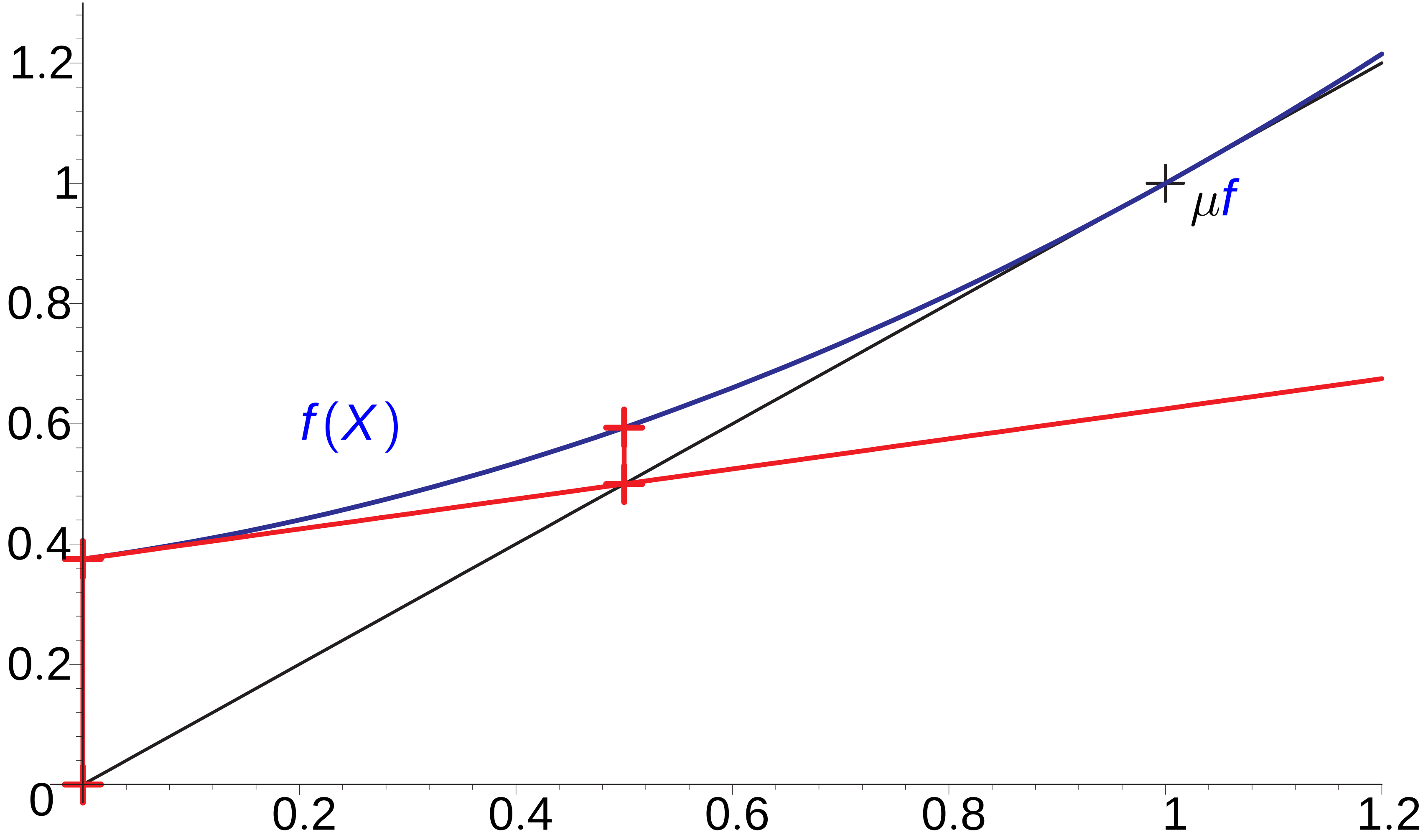
Newton's Method (univariate case)

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



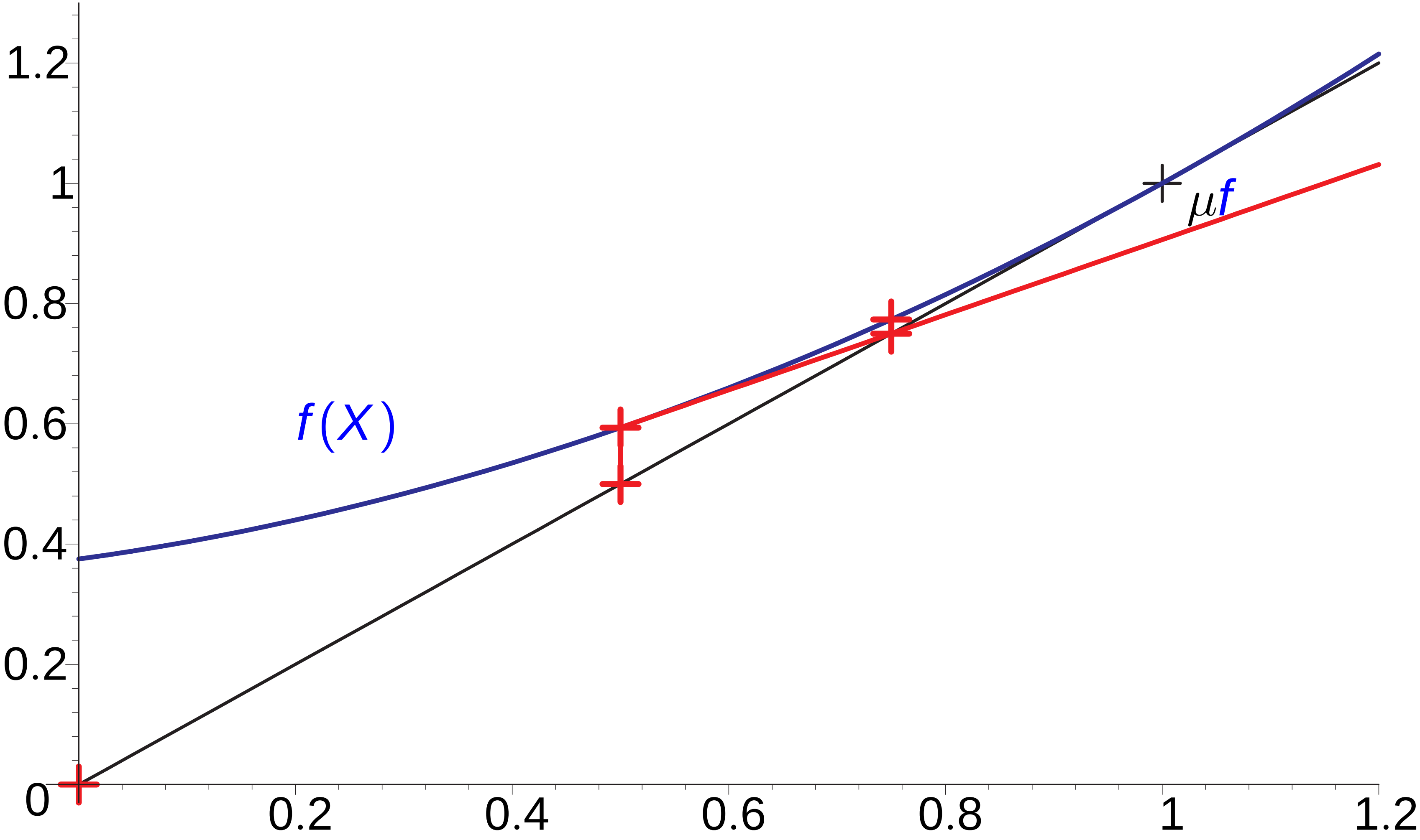
Newton's Method (univariate case)

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



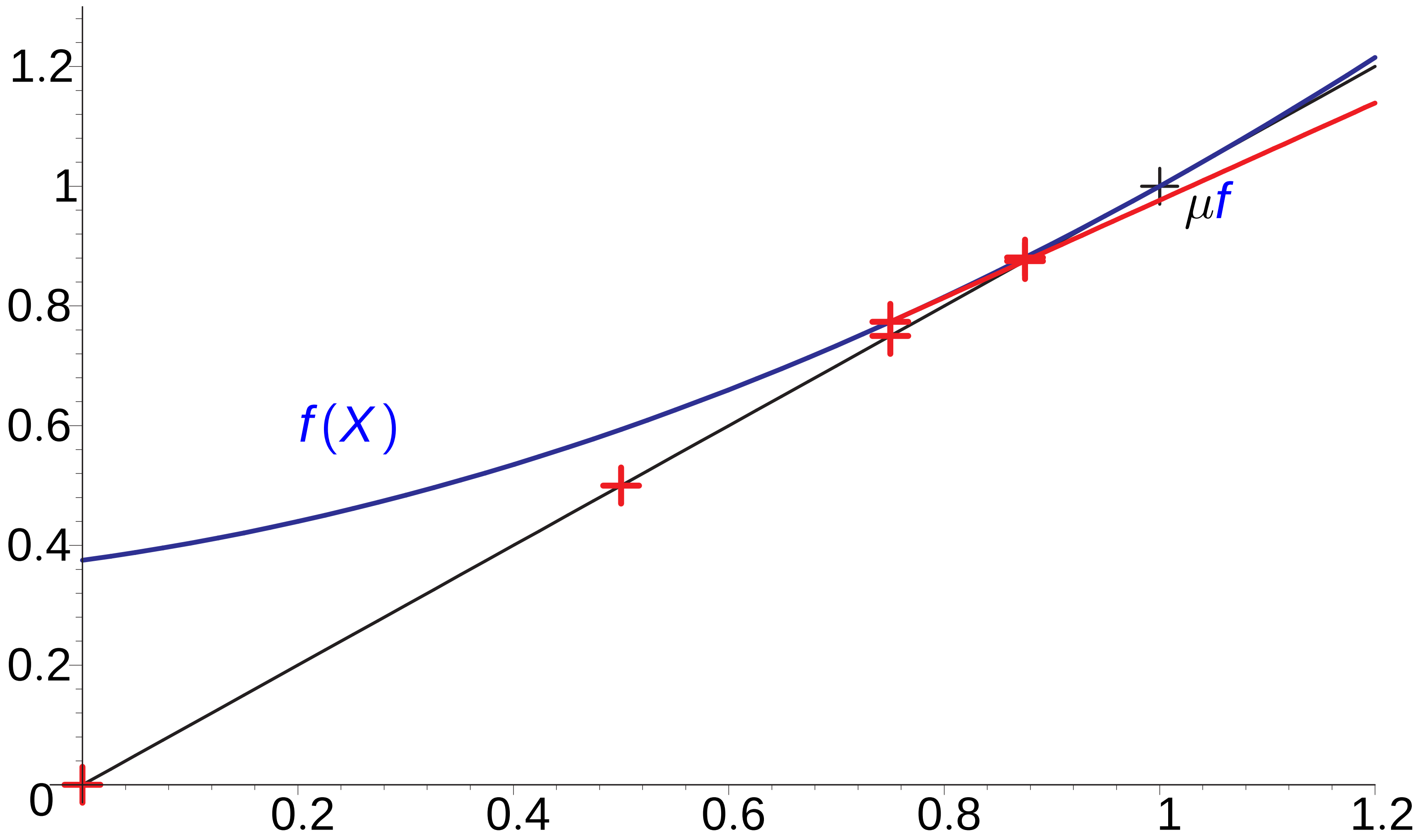
Newton's Method (univariate case)

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



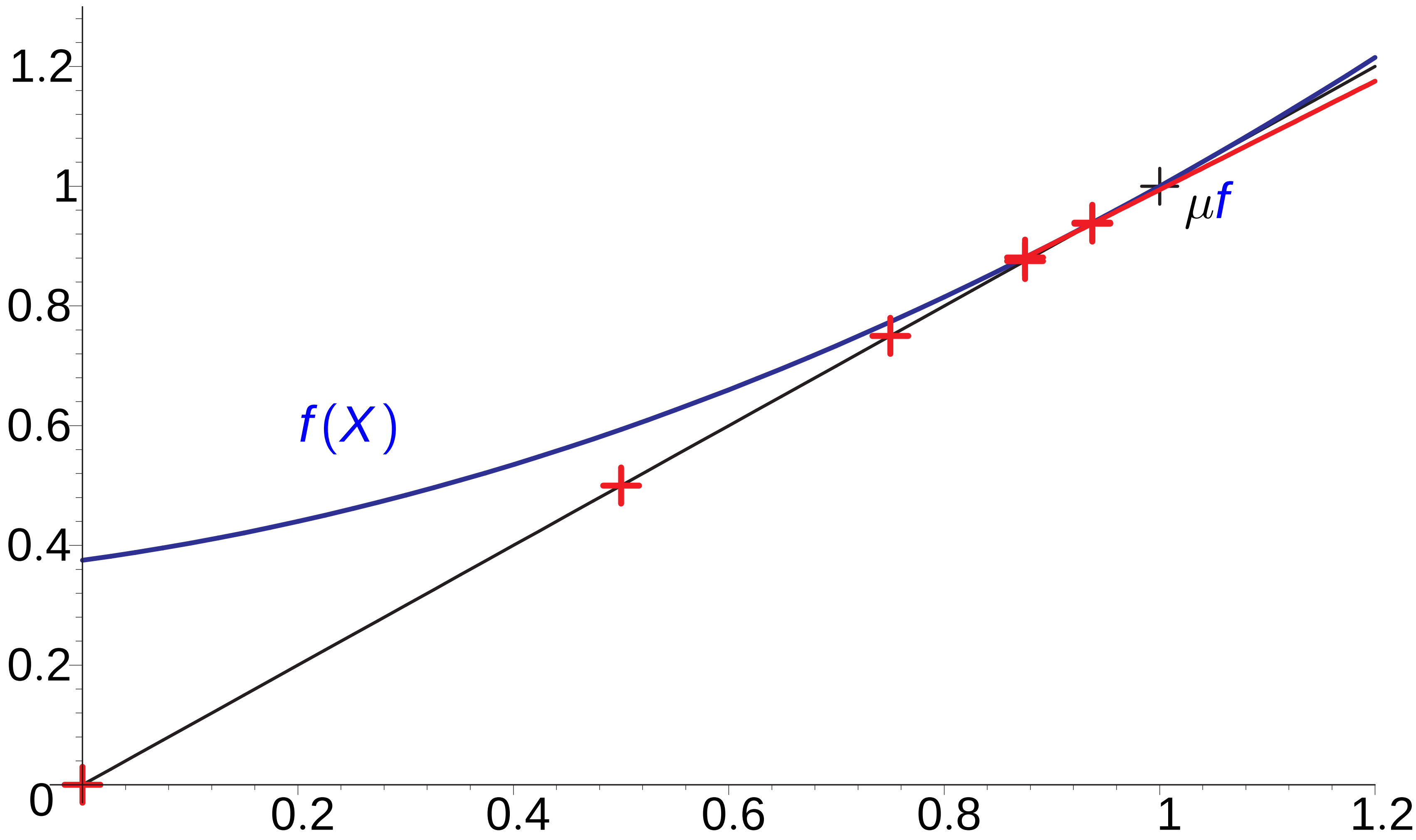
Newton's Method (univariate case)

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



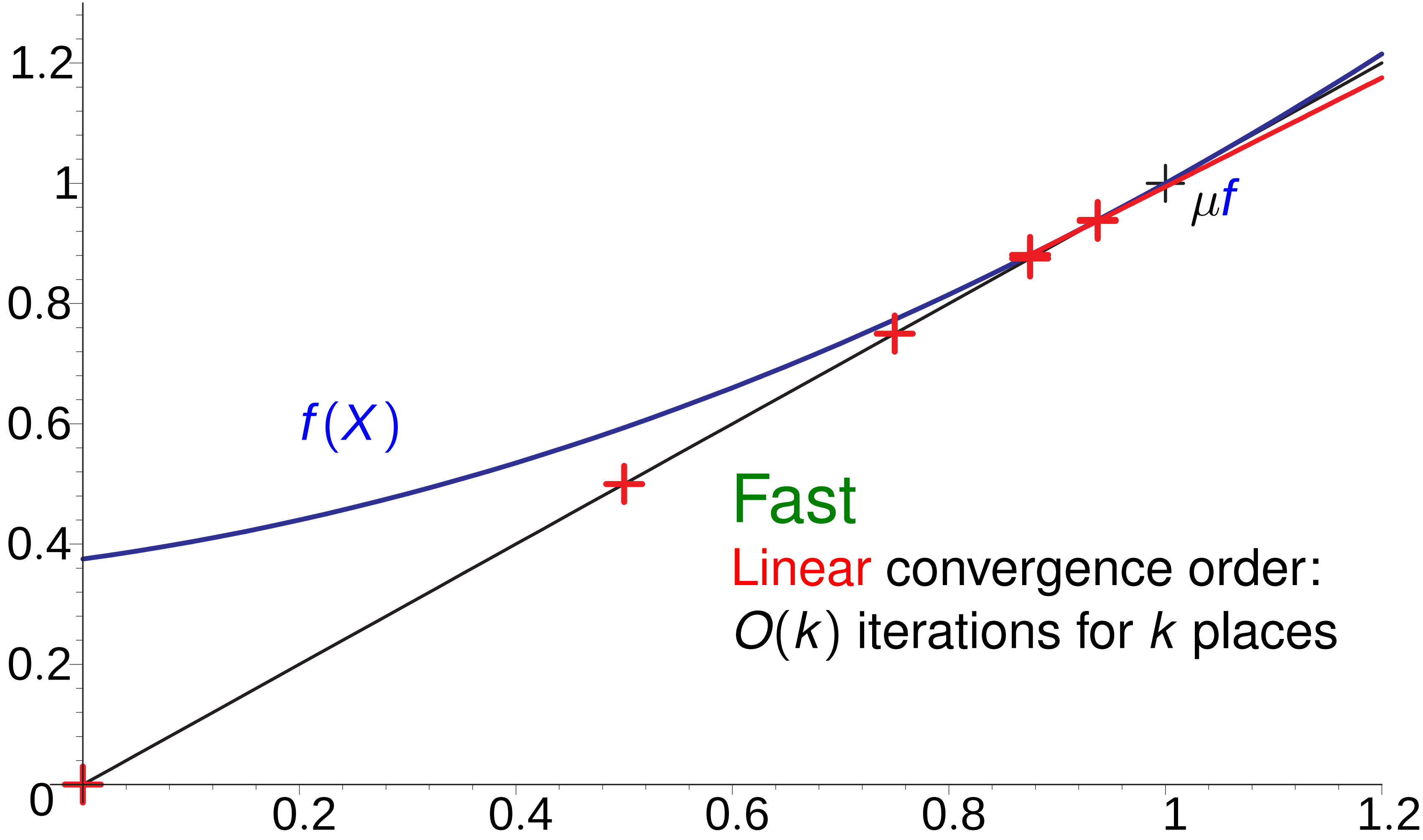
Newton's Method (univariate case)

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:

$O(k)$ iterations for k places

Mathematical formulation

Let $X = f(X)$ be the fundamental equation (one dimension). The Newton sequence $\nu^{(i)}$ is given by:

$$\nu^{(0)} = 0 \qquad \nu^{(i+1)} = \nu^{(i)} + \frac{f\left(\nu^{(i)}\right) - \nu^{(i)}}{1 - f'\left(\nu^{(i)}\right)}$$

In the multivariate case $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$:

$$\boldsymbol{\nu}^{(0)} = \mathbf{0} \qquad \boldsymbol{\nu}^{(i+1)} = \boldsymbol{\nu}^{(i)} + \left(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(i)})\right)^{-1} \left(\boldsymbol{f}(\boldsymbol{\nu}^{(i)}) - \boldsymbol{\nu}^{(i)}\right)$$

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 multitype system 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	$\boldsymbol{f}_X^k(0)$	$\boldsymbol{f}_Y^k(0)$	$\boldsymbol{f}_Z^k$	$\nu_X^{(k)}$	$\nu_Y^{(k)}$	$\nu_Z^{(k)}$
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

Synchronous case: Time to termination

Proposition

The probability of termination in at most k generations is the k-th Kleene approximant $\boldsymbol{f}^k(\mathbf{0})$.

Synchronous case: Time to termination

Proposition

The probability of termination in at most k generations is the k-th Kleene approximant $\mathbf{f}^k(\mathbf{0})$.

(by example).

Consider $X = 1/2\, XY + 1/4\, X + 1/4$. The probability P_X^k of termination of X in at most k-generations satisfies the equation

$$P_X^k = 1/2\, P_X^{(k-1)} P_Y^{(k-1)} + 1/4\, P_X^{(k-1)} + 1/4$$

□

Synchronous case: Time to termination

Proposition

- *The expected number of generations until termination of processes of type X is given by $E[T_X] = \sum_{k=0}^{\infty}(1 - \mathbf{f}(\mathbf{0})_X^k)$.*
- *It can be decided in polynomial time whether $E[T_X]$ is finite.*
- *If $E[T_X]$ is finite, then the partial sums converge to it with linear convergence order.*

Proposition

Consider the class of one-type systems

$$X \xrightarrow{p} XX \qquad X \xrightarrow{q} \epsilon$$

$E[T_X]$ is finite iff $p < 1/2$.

Asynchronous case: Scenario

- Threads are executed by a microprocessor.

Asynchronous case: Scenario

- Threads are executed by a microprocessor.
- A scheduler chooses the next thread to be executed from the current pool of active threads.

Asynchronous case: Scenario

- Threads are executed by a microprocessor.
- A scheduler chooses the next thread to be executed from the current pool of active threads.
- The chosen thread is executed for a (logical) time unit, after which it dies and generates 0,1,2, . . . new threads acording to the probability distribution.

Asynchronous case: Scenario

- Threads are executed by a microprocessor.
- A scheduler chooses the next thread to be executed from the current pool of active threads.
- The chosen thread is executed for a (logical) time unit, after which it dies and generates 0,1,2, . . . new threads acording to the probability distribution.
- The probability of termination/generation has been determined statistically
(probability as lack of information).

Standard semantics: Markov Decision Process (MDP)

States

Multisets of threads awaiting to be scheduled.

- Example: $X^4 Y^2 Z$.

Size of a state: number of threads in the multiset.

- Example: the size of $X^4 Y^2 Z$ is 7.

Nondeterministic/Stochastic choices

Nondeterministic choices model the options of the scheduler. Stochastic choices determine whether the thread terminates without offspring or reproduces.

Standard semantics: Markov Decision Process (MDP)

Scheduler

Function that maps a finite execution to the next thread to be executed.

- Resolves nondeterminism: only stochastic choices left.
- Depends in general on the complete past.

Analysis

- fix a class of schedulers;
- for each scheduler in the class analyze the corresponding Markov chain;
- take the maximum (minimum) of the values of the analysis. (angelical/demonical scheduler)

Time to termination

Proposition

The time to termination is independent of the choice of scheduler. (The random variable has the same distribution for all schedulers).

Corollary

The expected time to termination can be computed by solving a system of linear equations.

For the proof: Choose the LIFO scheduler, corresponding to a pushdown system, and apply the solution from [EKM LICS05].

Space consumption (memory, hardware)

Definition: Width of an execution

The width of an execution is the supremum of the sizes of the states visited along it.
The random variable W assigns to every execution its width

Space consumption (memory, hardware)

Definition: Width of an execution

The width of an execution is the supremum of the sizes of the states visited along it.
The random variable W assigns to every execution its width

Observe

W does not have the same distribution for every scheduler.

$$X \xrightarrow{0.5} XY \qquad X \xrightarrow{0.5} \epsilon$$

$$Y \xrightarrow{1} \epsilon$$

Competitive analysis for space consumption

- Well-known notion for online algorithms: comparison with an optimal offline algorithm.
  (Offline algorithm = online algorithm that knows the future.)

Competitive analysis for space consumption

- Well-known notion for online algorithms: comparison with an optimal offline algorithm.
  (Offline algorithm = online algorithm that knows the future.)
- Compare the performance of a given (class of) schedulers with respect to an optimal scheduler.

Competitive analysis for space consumption

- Well-known notion for online algorithms: comparison with an optimal offline algorithm.
  (Offline algorithm = online algorithm that knows the future.)
- Compare the performance of a given (class of) schedulers with respect to an optimal scheduler.
- The optimal scheduler has perfect information about the future sochastic evolution of the thread and its descendants.

Competitive analysis for space consumption

- Well-known notion for online algorithms: comparison with an optimal offline algorithm.
  (Offline algorithm = online algorithm that knows the future.)
- Compare the performance of a given (class of) schedulers with respect to an optimal scheduler.
- The optimal scheduler has perfect information about the future sochastic evolution of the thread and its descendants.
- Intuitively: Comparison with optimal scheduler indicates how much space can be spared by analyzing the code of the thread.

Family trees

Problem of the MDP semantics

Schedulers with information about the future cannot be formalized in the MDP semantics.

In the sequel we call MDP schedulers black-box schedulers (no access to the code).

Family trees

Problem of the MDP semantics

Schedulers with information about the future cannot be formalized in the MDP semantics.

In the sequel we call MDP schedulers black-box schedulers (no access to the code).

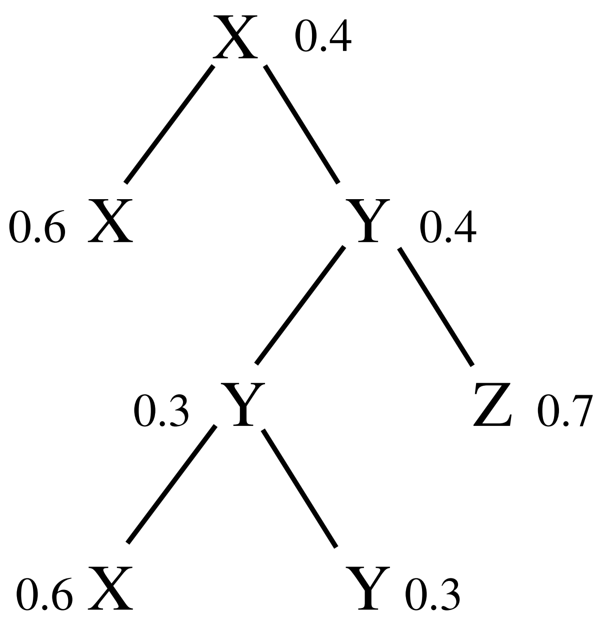
Solution: Family trees

A partial-order semantics of branching systems.

Trees obtained by resolving the stochastic choices, leaving nondeterminism unresolved.

A finite family tree and one of its linearizations

$$X \xrightarrow{0.4} X \parallel Y \qquad X \xrightarrow{0.6} \epsilon$$
$$Y \xrightarrow{0.3} X \parallel Y \qquad Y \xrightarrow{0.4} Y \parallel Z \qquad Y \xrightarrow{0.3} \epsilon$$
$$Z \xrightarrow{0.3} X \parallel Z \qquad Z \xrightarrow{0.7} \epsilon$$



$$X \to XY \to XYZ \to YZ \to XYZ \to XZ \to X \to \varepsilon$$

Family tree semantics

Probability space

- Elementary events: cilinders generated by the finite family trees.
- Probability of a cilinder: product of the probabilities associated to its nodes.

(White box) schedulers

(White box) Scheduler

Function that assigns to a family tree one of its linearizations.

Optimal width of a tree

Width of the linearization with smallest width.

(White box) schedulers

(White box) Scheduler

Function that assigns to a family tree one of its linearizations.

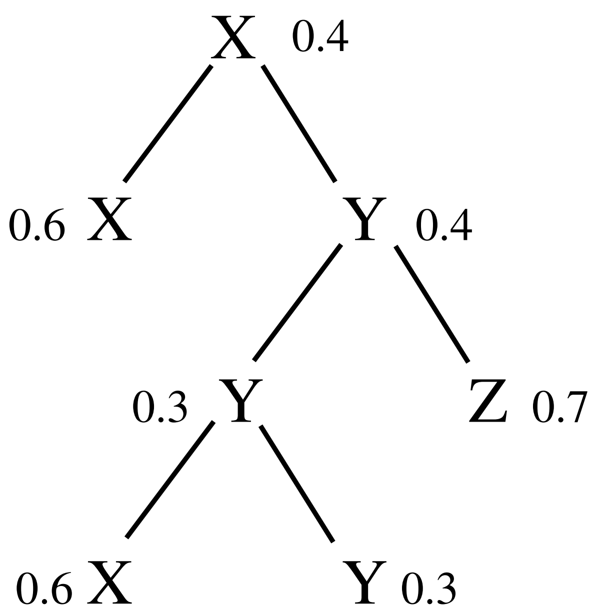
Optimal width of a tree

Width of the linearization with smallest width.

Optimal scheduler

At each point: Execute completely the subtree with smaller optimal width, and then execute completely the other one. (Notice: needs knowledge of the future!)

An example



- Optimal scheduler yields width 2:

$$X \rightarrow XY \rightarrow Y \rightarrow YZ \rightarrow Y \rightarrow XY \rightarrow Y \rightarrow \epsilon$$

- Pessimal scheduler yields width 4:

$$X \rightarrow XY \rightarrow XYZ \rightarrow XXYZ \rightarrow XYZ \rightarrow YZ \rightarrow Z \rightarrow \epsilon$$

Main result on the optimal scheduler

Theorem: prob. distribution of the optimal scheduler

Let $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ be the fundamental equation of a terminating system.
Let W_X^{opt} be the random variable that assigns to every finite family tree starting with variable X its optimal width.
Let $\nu^{(i)}$ be the i-th Newton approximant of the fundamental equation.
We have for every $i \geq 0$: $Pr(W_X^{opt} \leq i) = \nu_X^{(i)}$.

Similar result for systems with a positive probability of non-termination.

Example

k	$\boldsymbol{f}_X^k(0)$	$\boldsymbol{f}_Y^k(0)$	$\boldsymbol{f}_Z^k$	$\nu_X^{(k)}$	$\nu_Y^{(k)}$	$\nu_Z^{(k)}$
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

Black-box schedulers

No positive results yet on optimal black-box schedulers.

Examples showing that the optimal scheduler requires unbounded memory.

Watch this space . . .

A particular case: One-type systems

Consider the family of systems

$$X \xrightarrow{p} XX \qquad X \xrightarrow{q} \epsilon \qquad \text{where } p < q.$$

Theorem: Distribution of black-box schedulers

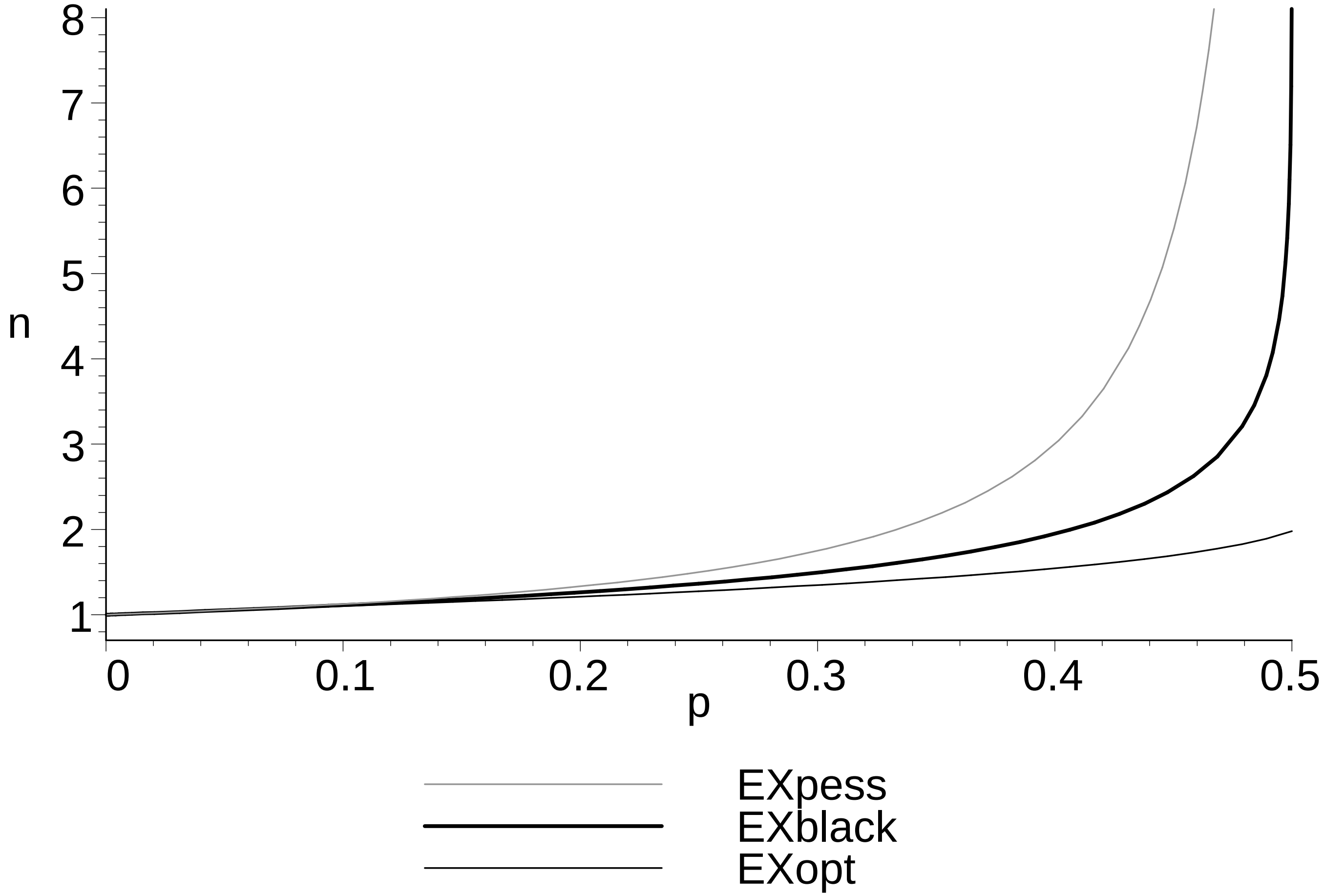
Let σ, τ be black-box schedulers, and let W^σ, W^τ be their associated widths. We have:

$$Pr(W^\sigma \le k) = Pr(W^\tau \le k) \quad \text{for every } k \ge 0\,.$$

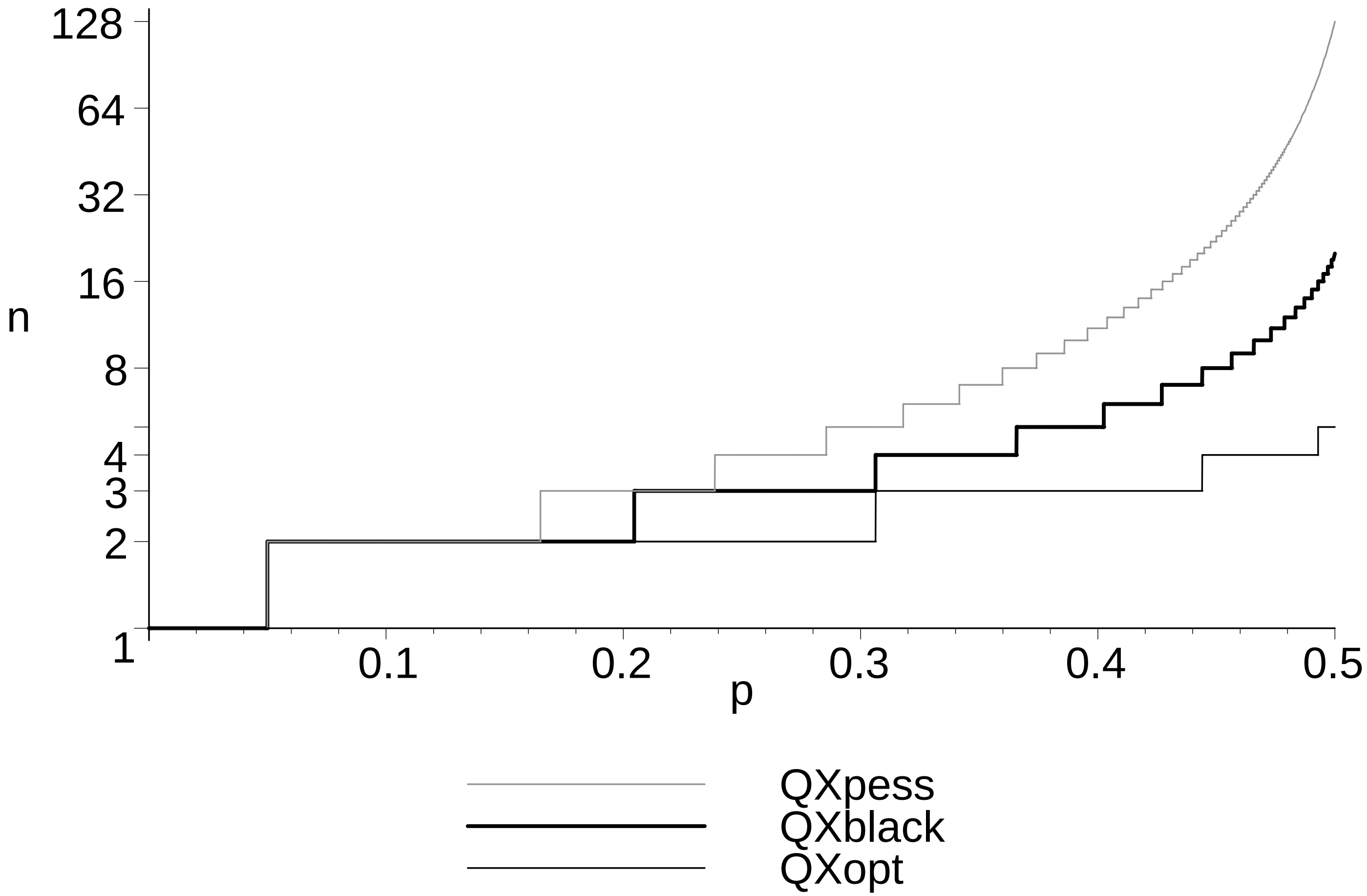
Let W^{bb} denote the random variable of an arbitrary black-box scheduler. Then

$$Pr[W^{bb} \ge k] = \frac{\left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^{k-1}}{1 - \left(\frac{p}{q}\right)^k}$$

Competitive analysis: Expectation of width



Competitive analysis: 95% Quantiles



Conclusions

- Stochastic branching processes naturally appear in computer science.

Conclusions

- Stochastic branching processes naturally appear in computer science.
- The distinction synchronous/asynchronous has not been yet considered.

Conclusions

- Stochastic branching processes naturally appear in computer science.
- The distinction synchronous/asynchronous has not been yet considered.
- Random variables associated to space consumption have not yet been studied.

Conclusions

- Stochastic branching processes naturally appear in computer science.
- The distinction synchronous/asynchronous has not been yet considered.
- Random variables associated to space consumption have not yet been studied.
- There is a surprising connection between optimal schedulers and Newton approximants.

Conclusions

- Stochastic branching processes naturally appear in computer science.
- The distinction synchronous/asynchronous has not been yet considered.
- Random variables associated to space consumption have not yet been studied.
- There is a surprising connection between optimal schedulers and Newton approximants.
- The one-type case is almost completely solved.

Back to 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), vicar 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 . . .

Due to an algebraic error, Watson concluded wrongly that all families eventually die out.

English peers again . . .

Due to an algebraic error, Watson concluded wrongly that all families eventually die out.

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

English peers again . . .

Due to an algebraic error, Watson concluded wrongly that all families eventually die out.

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

- English peers tended to marry heiresses (daughters without brothers)

English peers again . . .

Due to an algebraic error, Watson concluded wrongly that all families eventually die out.

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

- English peers tended to marry heiresses (daughters without brothers)
- Heiresses come from families with lower fertility rates (lower probabilities $p_1, p_2, p_3, \ldots$).

English peers again . . .

Due to an algebraic error, Watson concluded wrongly that all families eventually die out.

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

- English peers tended to marry heiresses (daughters without brothers)
- Heiresses come from families with lower fertility rates (lower probabilities $p_1, p_2, p_3, \ldots$).
- . . . which increases the probability of the family dying out.