Solving fixed-point equations on ω-continuous semirings

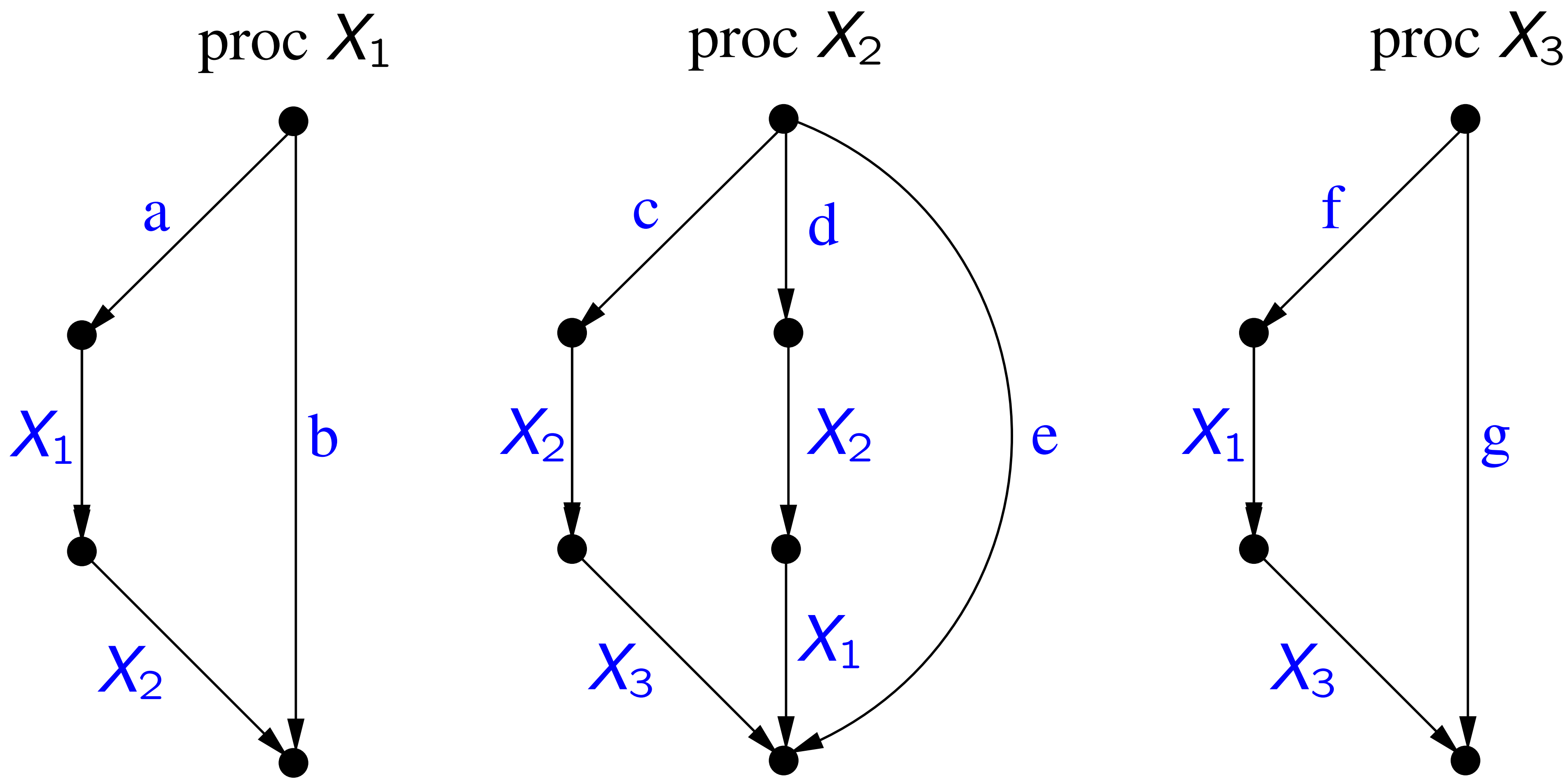
Javier Esparza

Technische Universität München

Joint work with

Stefan Kiefer and Michael Luttenberger

From programs to flowgraphs



From flowgraphs to equations

A syntactic transformation.

$$\begin{aligned} X_1 &= a \cdot X_1 \cdot X_2 + b \\ X_2 &= c \cdot X_2 \cdot X_3 + d \cdot X_2 \cdot X_1 + e \\ X_3 &= f \cdot X_1 \cdot X_3 + g \end{aligned}$$

But how should the equations be interpreted mathematically?

- What kind of objects are $a, \ldots, g$?
- What kind of operations are sum and product ?

From flowgraphs to equations

A syntactic transformation.

$$
\begin{aligned}
X_1 &= a \cdot X_1 \cdot X_2 + b \\
X_2 &= c \cdot X_2 \cdot X_3 + d \cdot X_2 \cdot X_1 + e \\
X_3 &= f \cdot X_1 \cdot X_3 + g
\end{aligned}
$$

But how should the equations be interpreted mathematically?

– What kind of objects are $a, \ldots, g$?
– What kind of operations are sum and product ?

It depends. Different interpretations lead to different semantics.

Input/output relational semantics

Interpret $a, \ldots, g$ as assignments or guards over a set of program variables V with set of valuations Val.

$R(X_i) = (v, v') \in Val \times Val$ such that X_i started at v, may terminate at v'.

Language semantics

Interpret the atomic actions as letters of an alphabet A.

$L(X_i) =$ words $w \in A^*$ such that X_i can execute w and terminate.

Language semantics

Interpret the atomic actions as letters of an alphabet A.

$L(X_i) =$ words $w \in A^*$ such that X_i can execute w and terminate.

$(L(X_1), L(X_2), L(X_3))$ is the least solution of the equations under the following interpretation:

- Universe: 2^{A^*} (languages over A).
- $a, \ldots, g$ are the singleton languages $\{a\}, \ldots, \{g\}$.
- sum is union of languages, product is concatenation:

$$L_1 \cdot L_2 = \{w_1 w_2 \mid w_1 \in L_1 \wedge w_2 \in l_2\}$$

Probabilistic termination semantics

Interpret $a, \ldots, g$ as probabilities.

$T(X_i) =$ probability that X_i terminates.

Probabilistic termination semantics

Interpret $a, \ldots, g$ as probabilities.

$T(X_i) =$ probability that X_i terminates.

$(T(X_1), T(X_2), T(X_3))$ is the least solution of the equations under the following interpretation:

- Universe: $\mathbb{R}^+$
- $a, \ldots, g$ are the probabilities of taking the transitions
- sum and product are addition and multiplication of reals

ω-continuous semirings

Underlying mathematical structure: ω-continuous semirings

Algebra $(C, +, \cdot, 0, 1)$

– $(C, +, 0)$ is a commutative monoid

– $(C, \cdot, 1)$ is a monoid

– $a \sqsubseteq a + b$ is a partial order

– $\cdot$ distributes over $+$

– $0 \cdot a = a \cdot 0 = 0$

– $\sqsubseteq$-chains have limits

System of (w.l.o.g. quadratic) equations $X = f(X)$ where

- $X = (X_1, \ldots, X_n)$ vector of variables,
- $f(X) = (f_1(X), \ldots, f_n(X))$ vector of terms over $C \cup \{X_1, \ldots, X_n\}$.

Notice: the f_i are polynomials

Kleenean program analysis

Theorem [Kleene]: The least solution μf is the supremum of $\{k_i\}_{i \geq 0}$, where

$$k_0 = f(0)$$
$$k_{i+1} = f(k_i)$$

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

Kleenean program analysis is slow

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

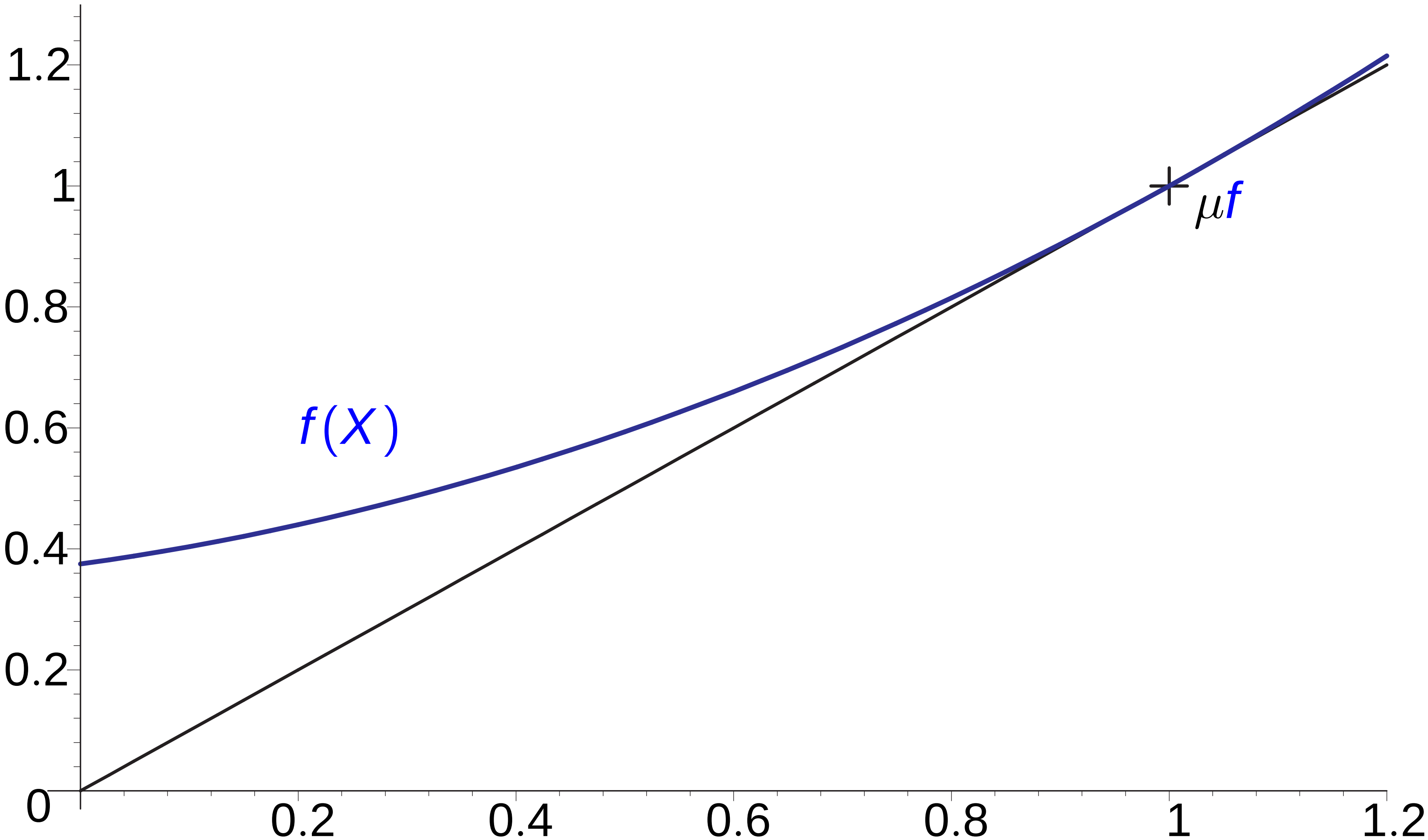
- $X = a \cdot X + b \qquad \mu f = a^* b$
- Kleene approximants are finite sets: $k_i = (\epsilon + a + \ldots + a^i)b$

Probabilistic interpretation: convergence can be very slow [EY STACS05].

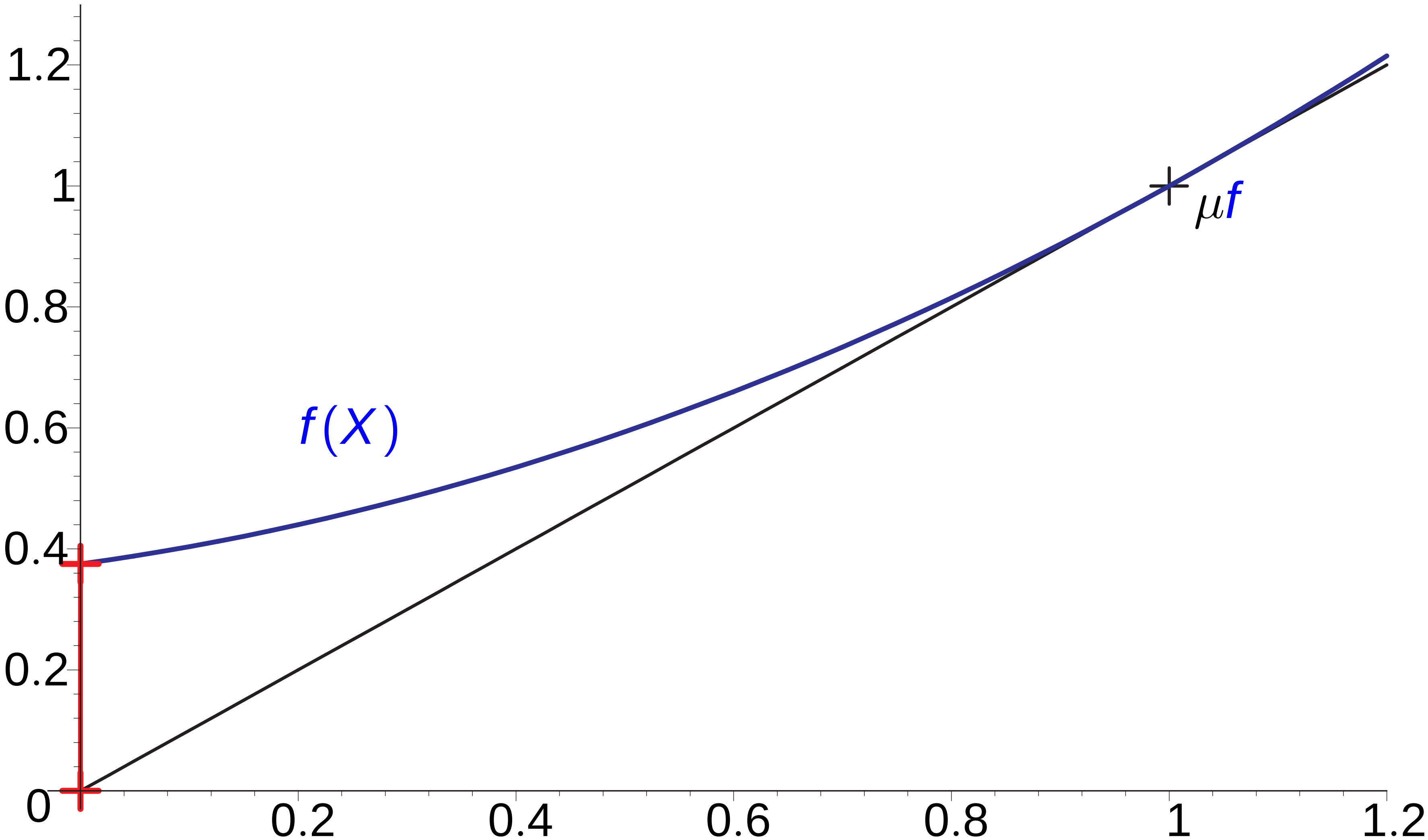
- $X = \frac{1}{2} X^2 + \frac{1}{2} \qquad \mu f = 1 = 0.99999\ldots$
- "Logarithmic convergence": k iterations to get $\log k$ bits of accuracy.

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

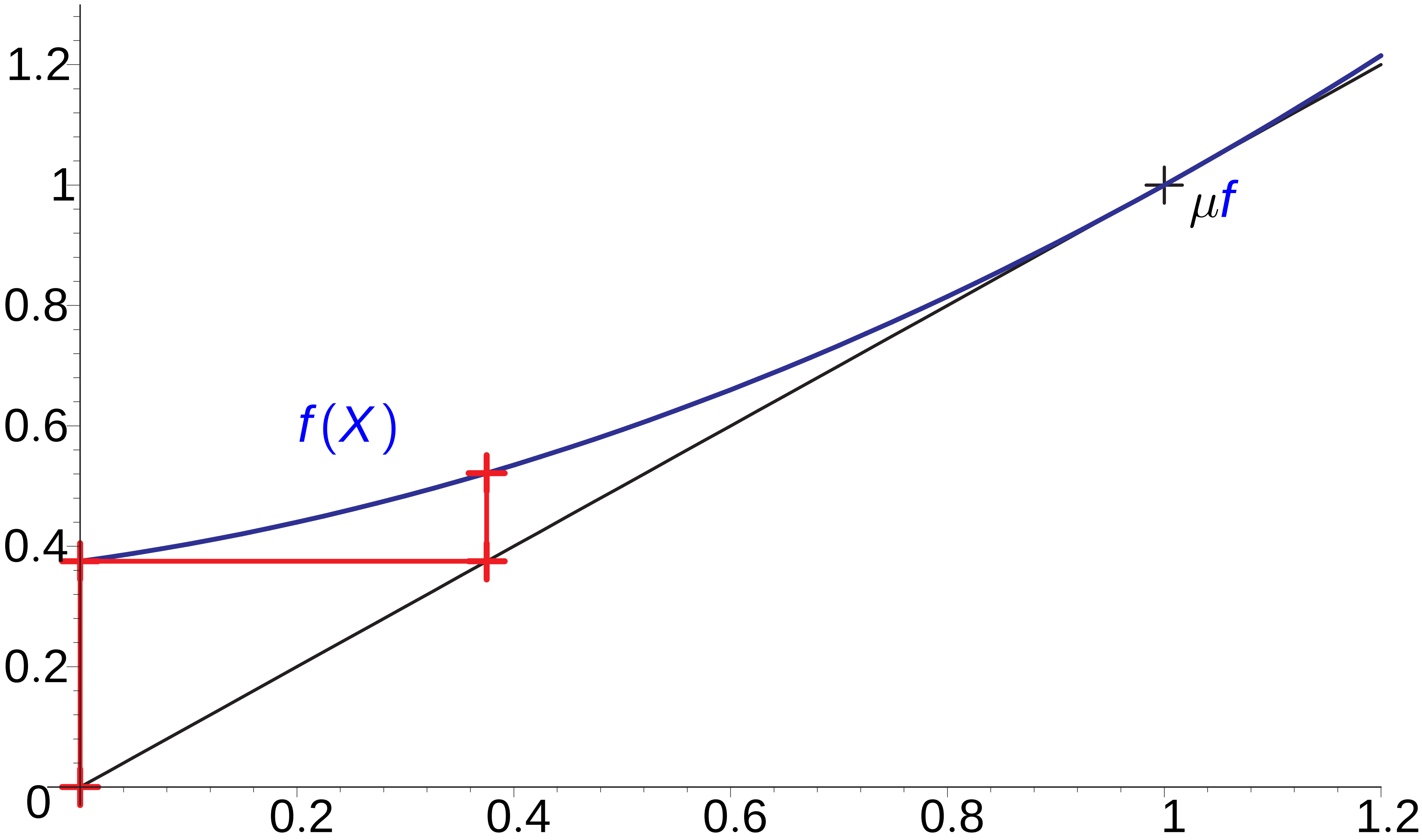
Kleene Iteration for $X = f(X)$ (univariate case)



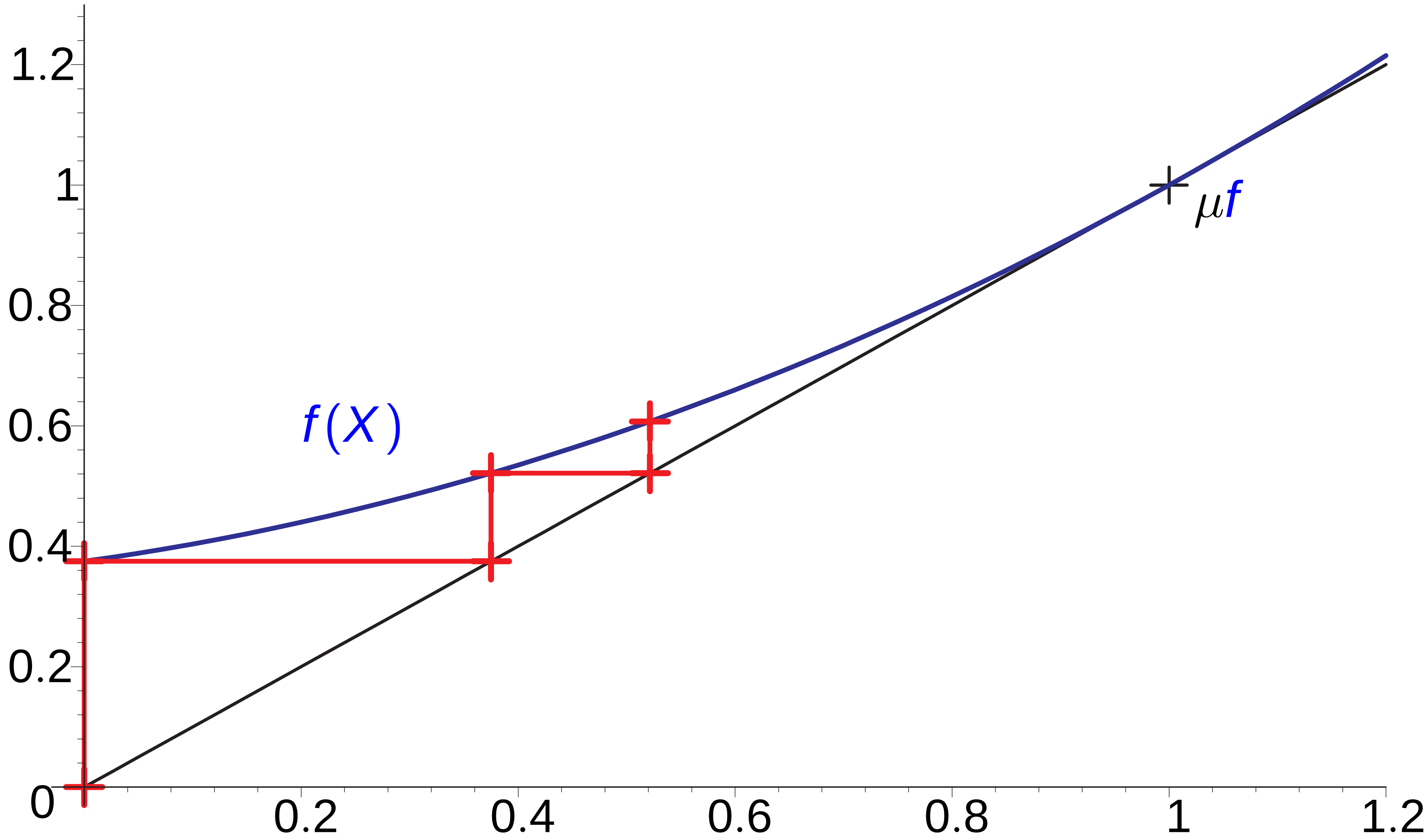
Kleene Iteration for $X = f(X)$ (univariate case)



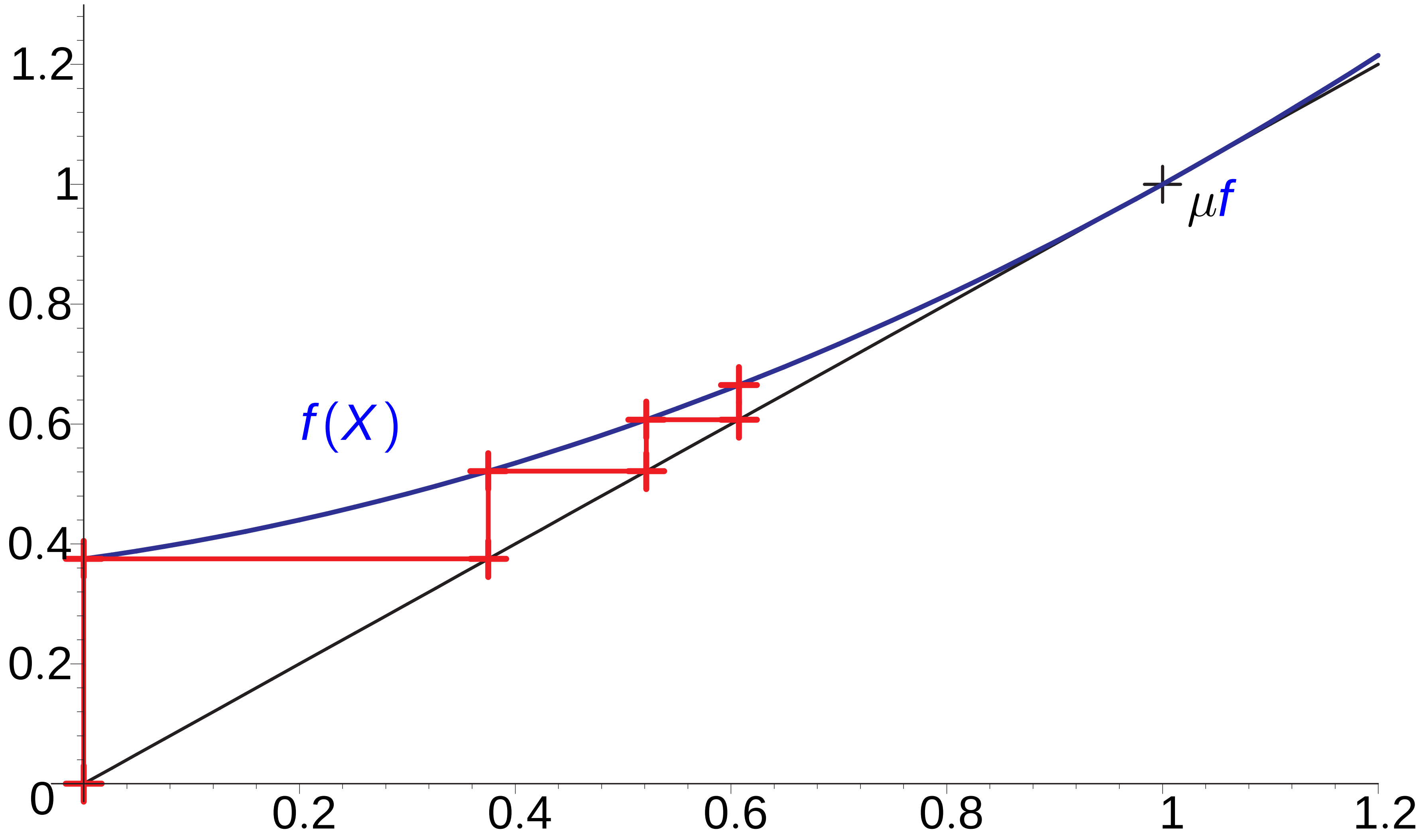
Kleene Iteration for $X = f(X)$ (univariate case)



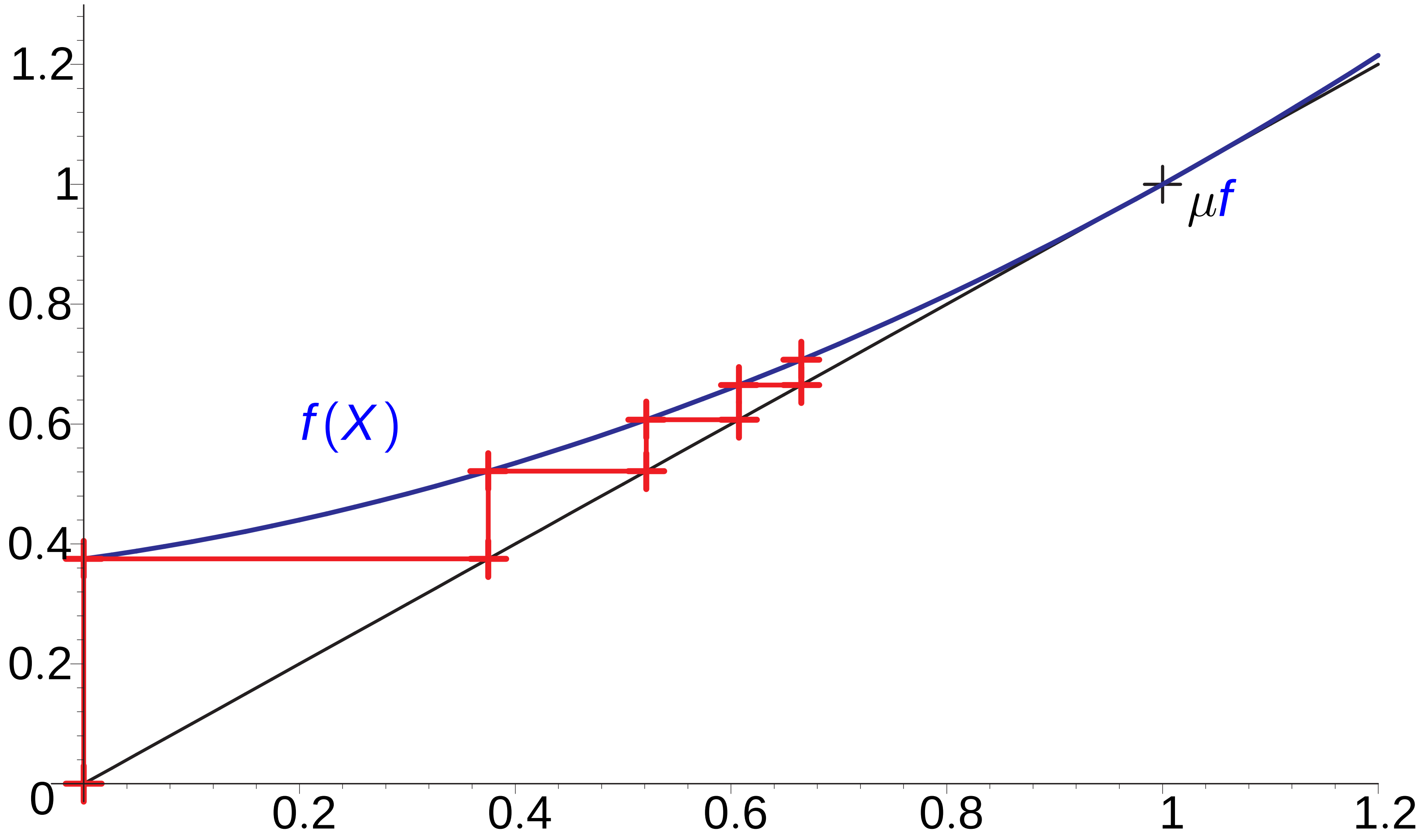
Kleene Iteration for $X = f(X)$ (univariate case)



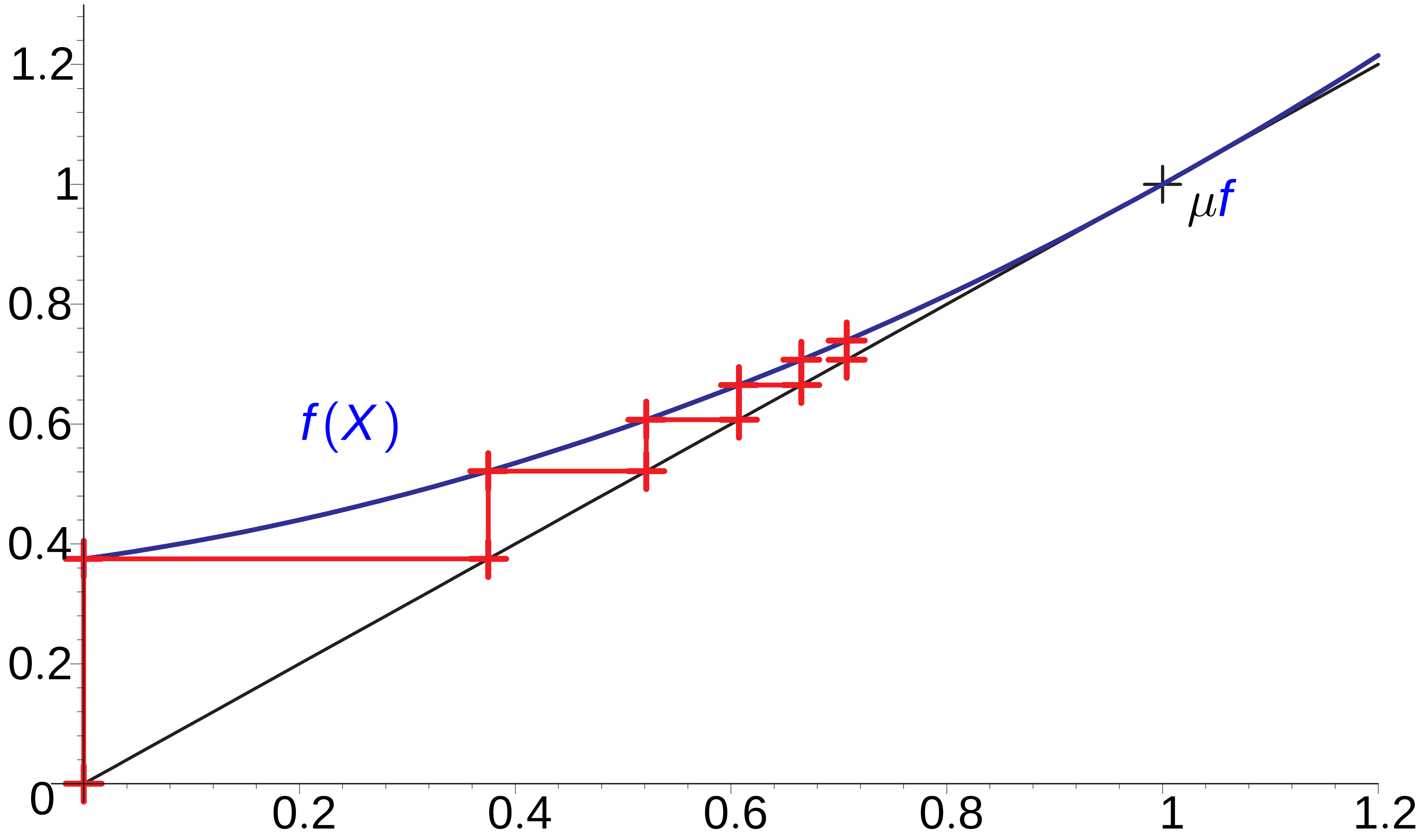
Kleene Iteration for $X = f(X)$ (univariate case)



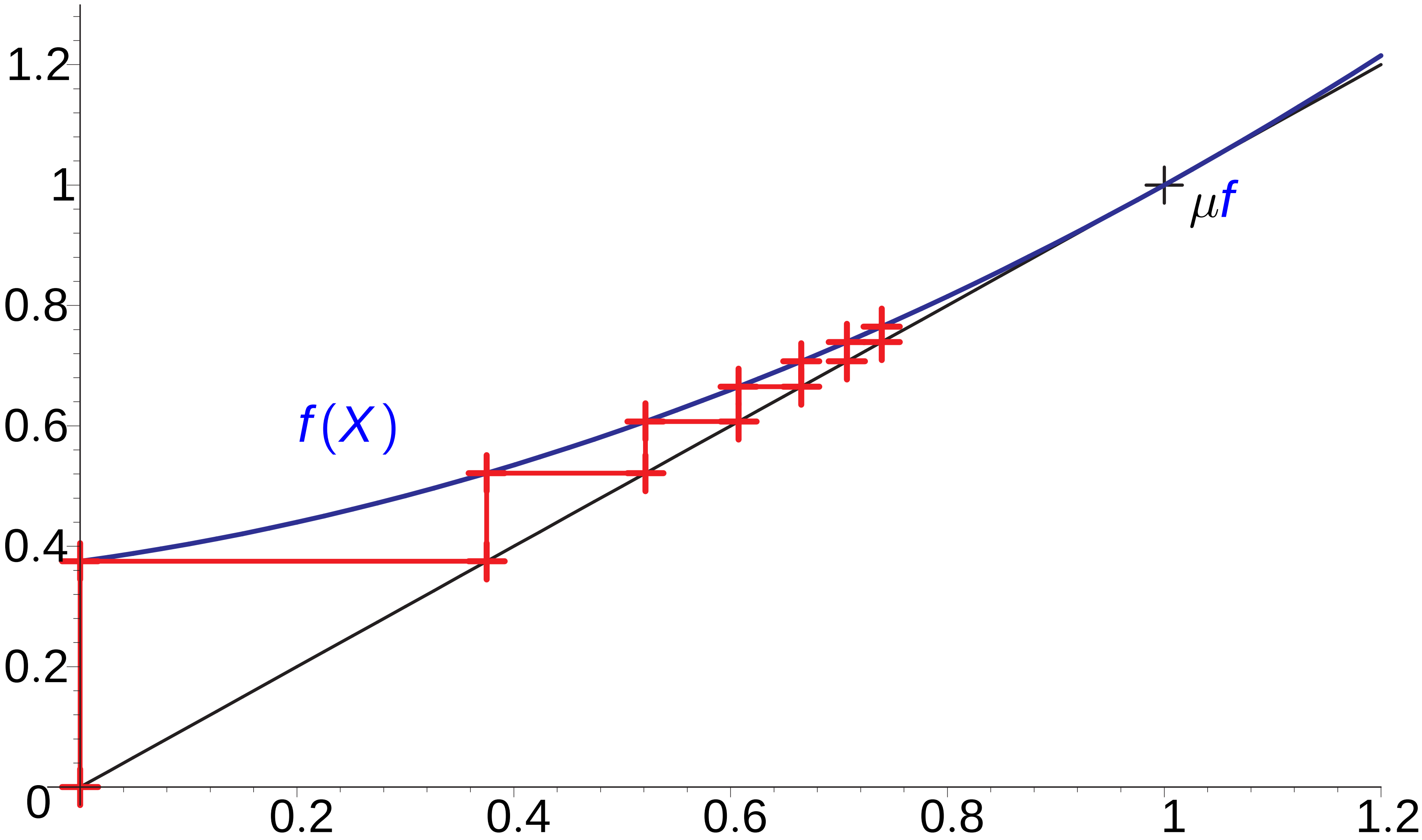
Kleene Iteration for $X = f(X)$ (univariate case)



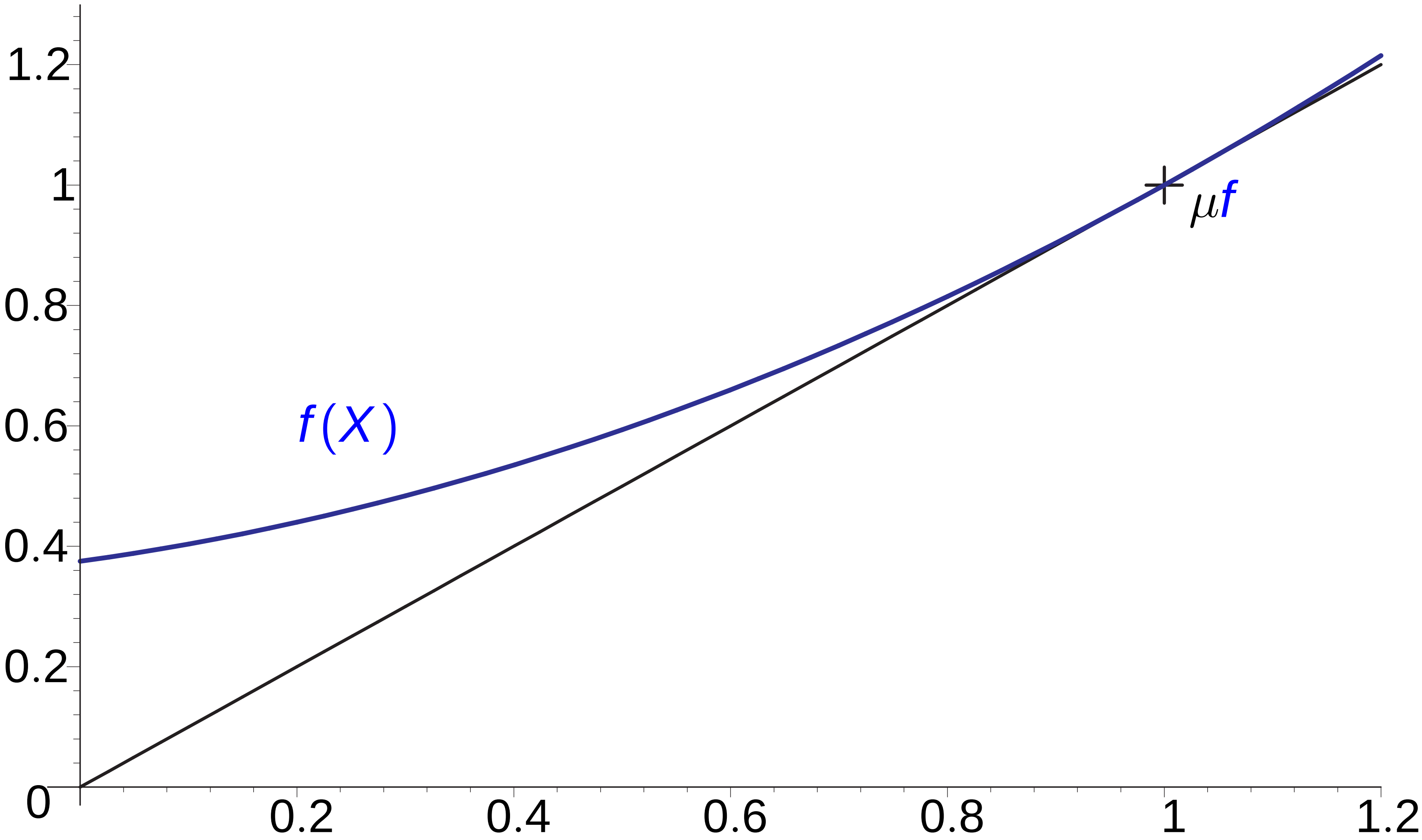
Kleene Iteration for $X = f(X)$ (univariate case)



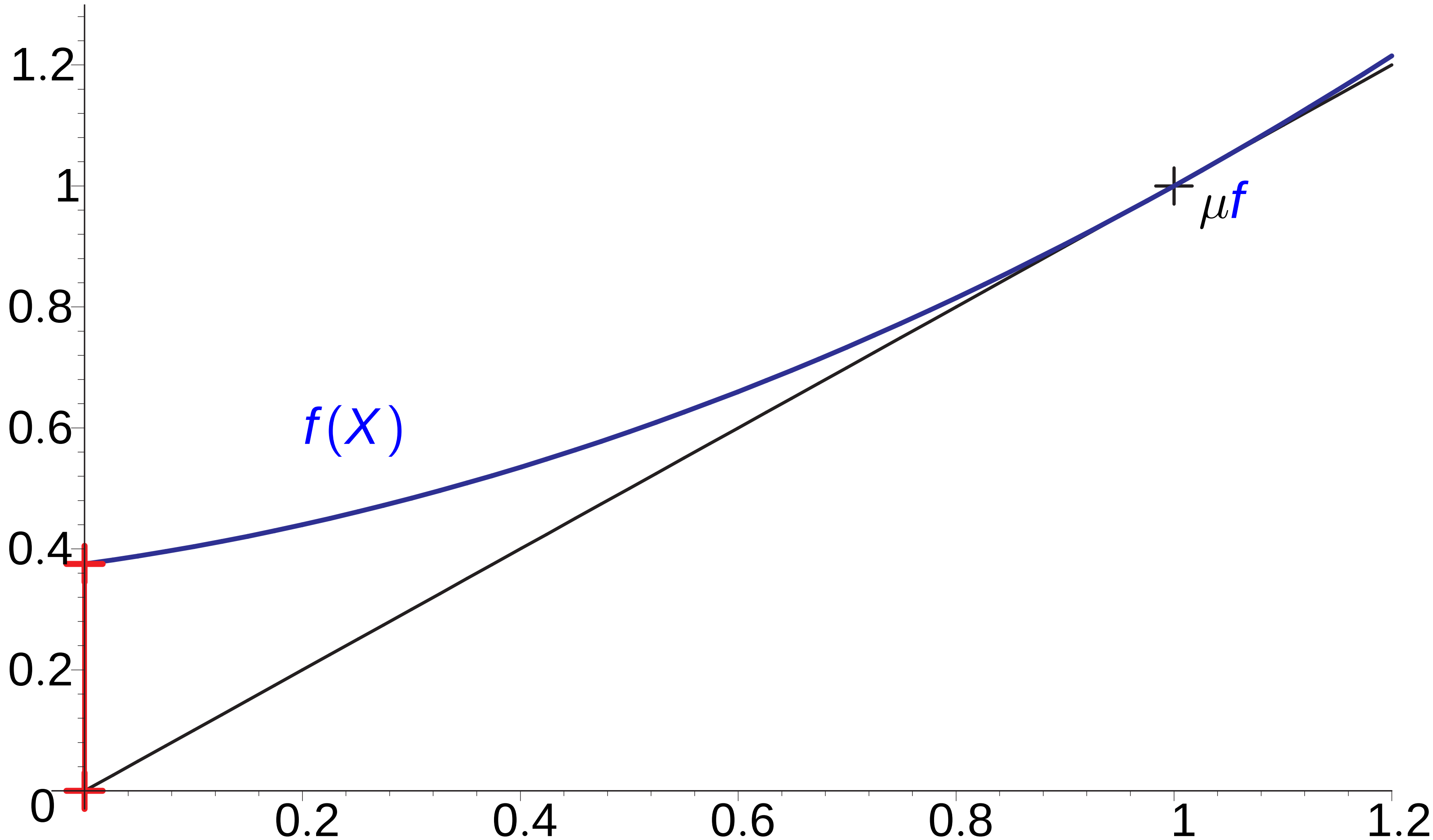
Kleene Iteration for $X = f(X)$ (univariate case)



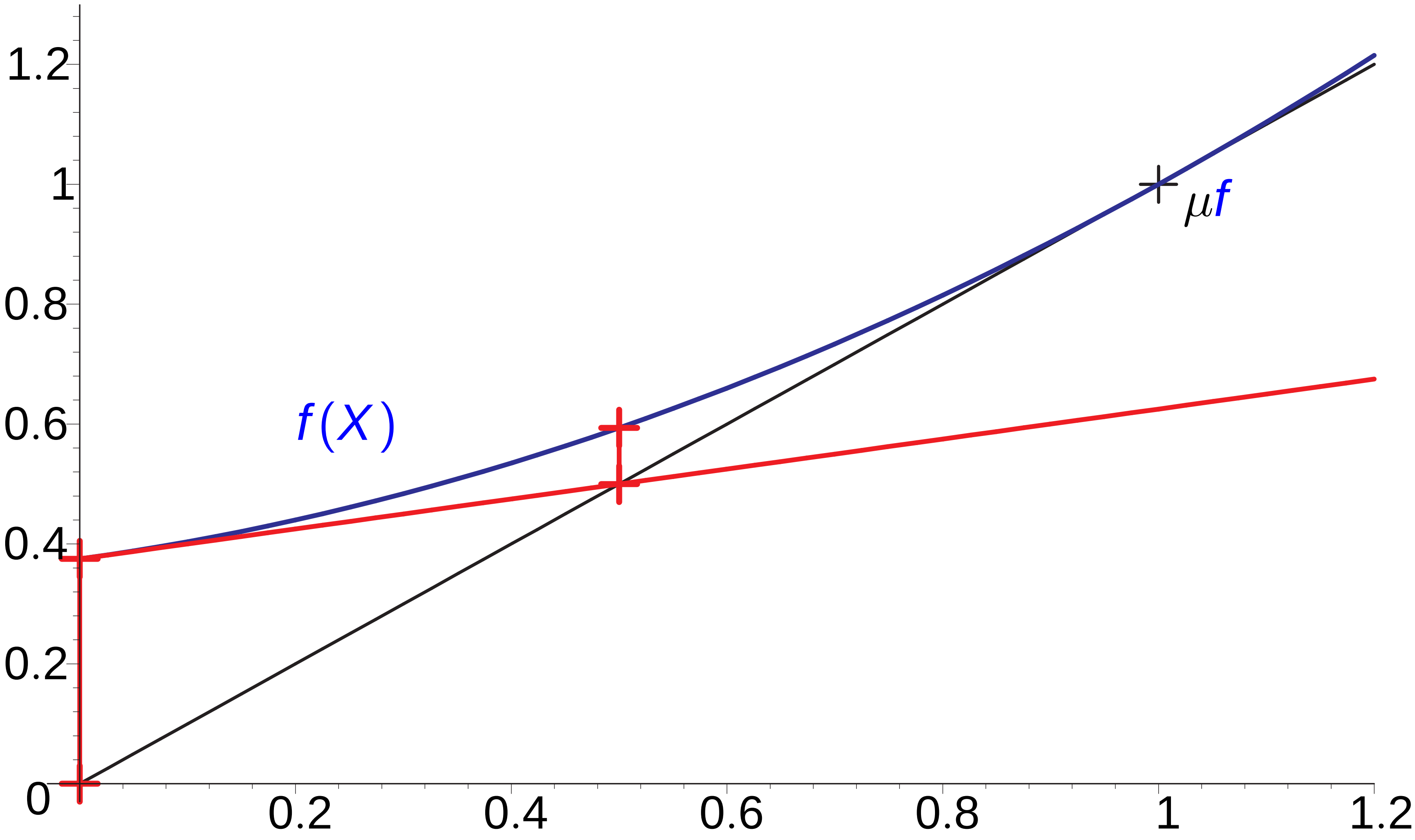
Newton's Method for $X = f(X)$ (univariate case)



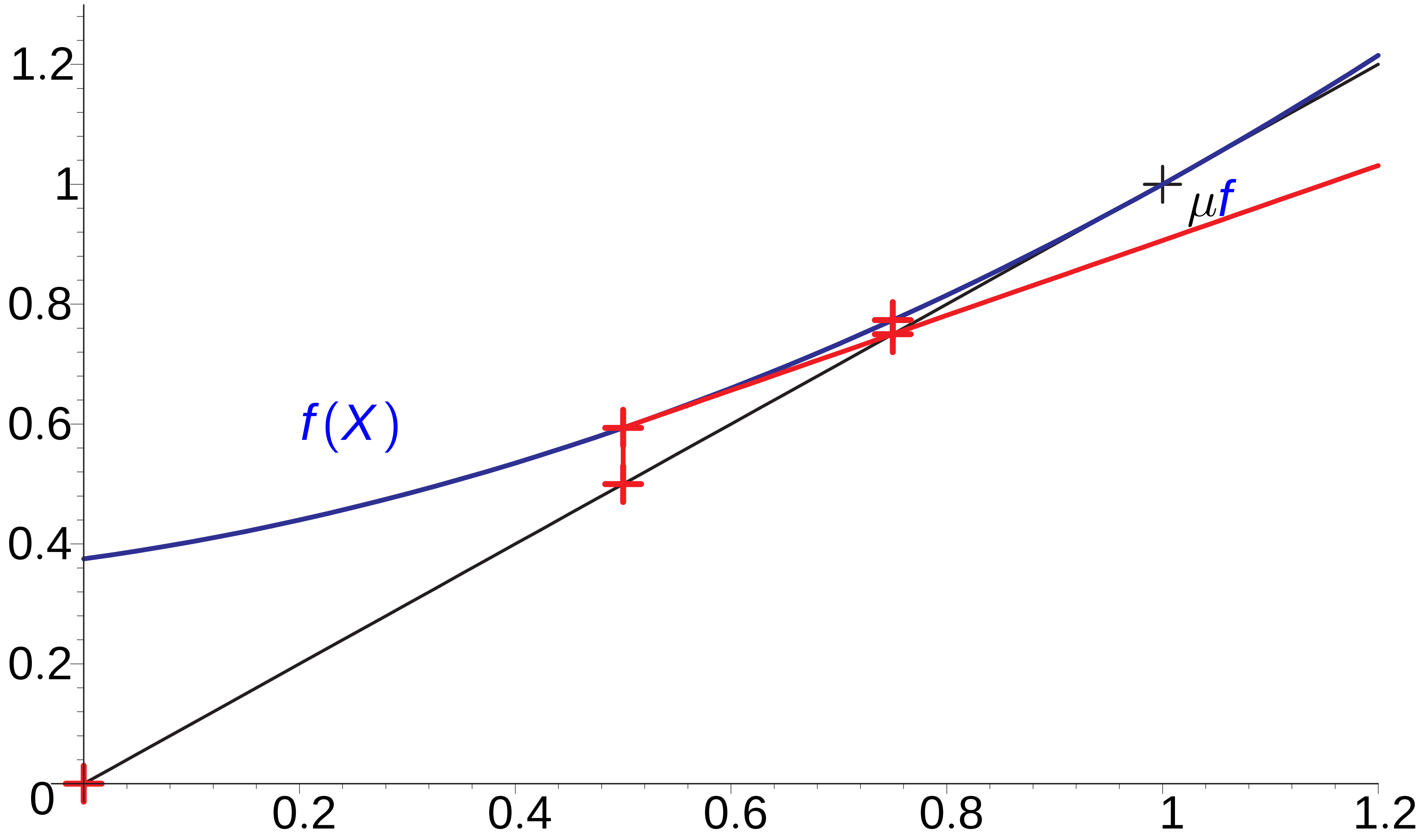
Newton's Method for $X = f(X)$ (univariate case)



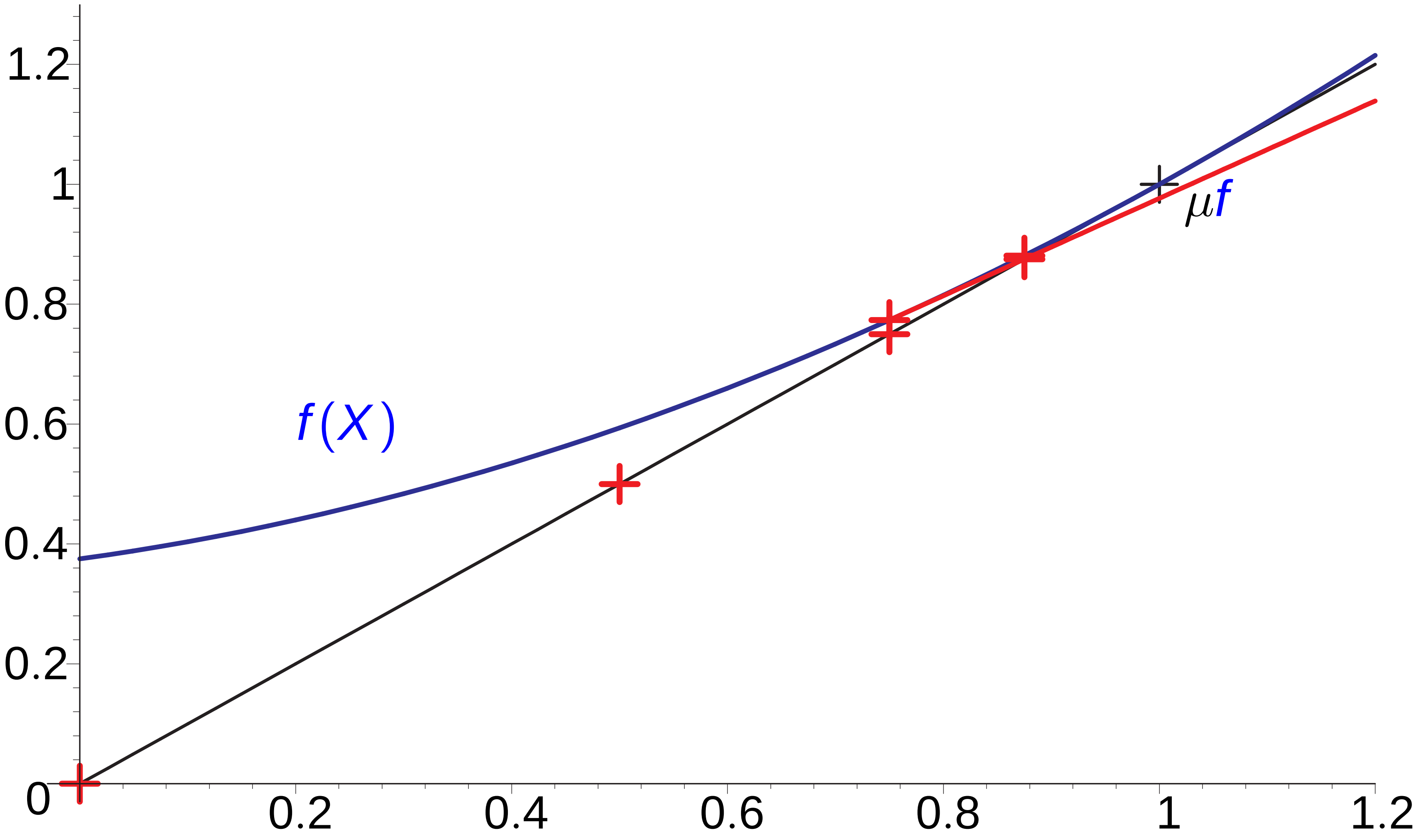
Newton's Method for $X = f(X)$ (univariate case)



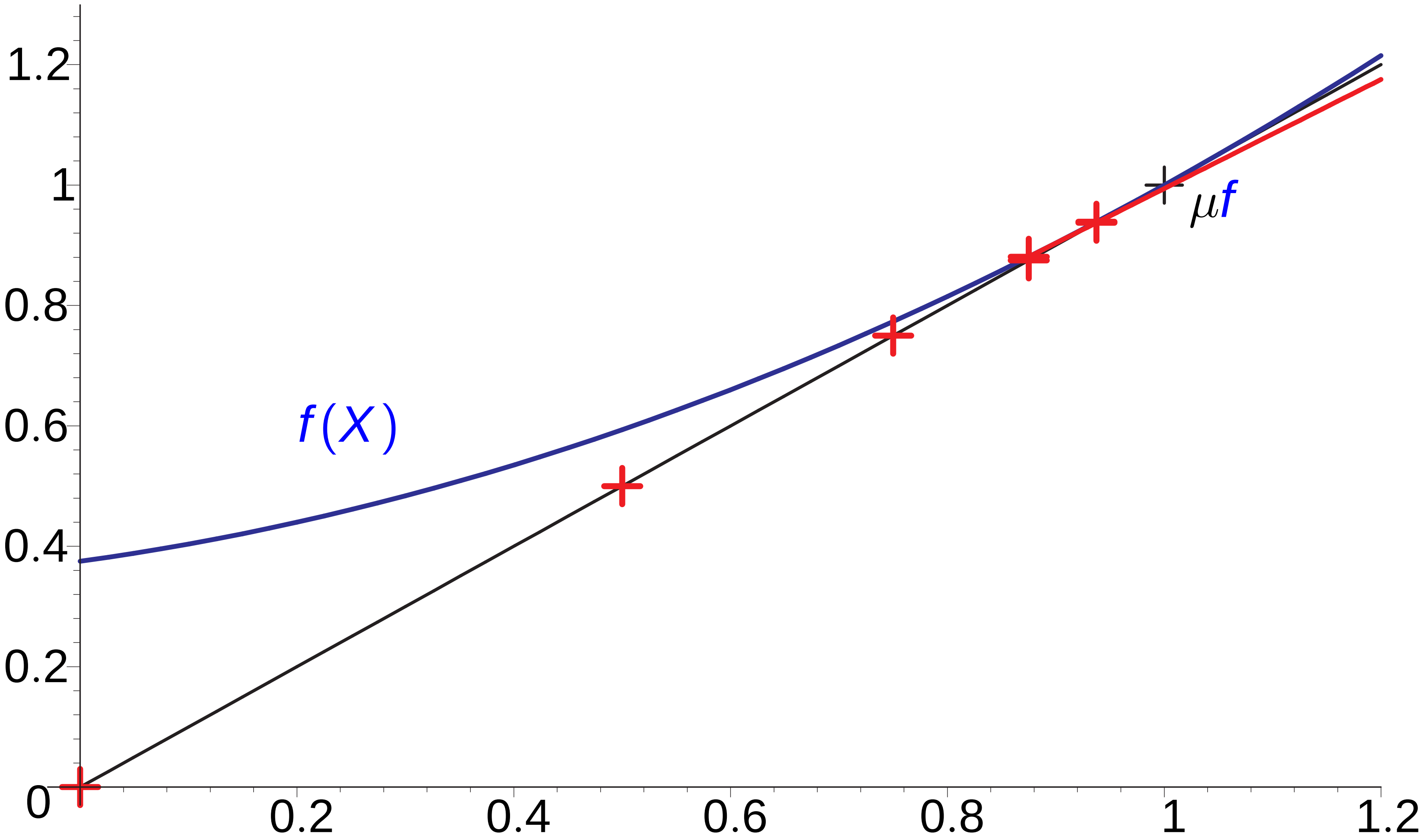
Newton's Method for $X = f(X)$ (univariate case)



Newton's Method for $X = f(X)$ (univariate case)



Newton's Method for $X = f(X)$ (univariate case)



Evaluation of Newton’s method

Newton’s Method is usually very efficient

- often exponential convergence

. . . but not robust:

- may not converge, converge only locally (in some neighborhood of the least fixed-point), or converge very slowly.

A puzzling mismatch

Program analysis:

- General domain: arbitrary ω-continuous semirings
- Kleene Iteration is robust and generally applicable
- . . . but converges slowly.

Numerical mathematics:

- Particular domain: the real field
- Newton's Method converges fast
- . . . but is not robust

Our main result

- Newton's Method can be defined for arbitrary ω-continuous semirings, and becomes as robust as Kleene's method.

Mathematical formulation of Newton's Method

Let ν be some approximation of μf. (We start with $\nu = f(0)$.)

- Compute the function $T_\nu(X)$ describing the tangent to $f(X)$ at ν
- Solve $X = T_\nu(X)$ (instead of $X = f(X)$), and take the solution as the new approximation

Elementary analysis: $T_\nu(X) = Df_\nu(X) + f(\nu) - \nu$

where $Df_{x_0}(X)$ is the differential of f at x_0

So:
$$\begin{aligned} \nu_0 &= 0 \\ \nu_{i+1} &= \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } X = Df_{\nu_i}(X) + f(\nu_i) - \nu_i \end{aligned}$$

Generalizing Newton's method

Key point: generalize $X = Df_\nu(X) + f(\nu) - \nu$

In an arbitrary ω-continuous semiring

- neither the differential $Df_\nu(X)$, nor
- the difference $f(\nu) - \nu$

are defined.

Differentials in semirings

Standard solution: take the algebraic definition

$$Df_\nu(X) = \begin{cases} 0 & \text{if } f(X) = c \\ X & \text{if } f(X) = X \\ Dg_\nu(X) + Dh_\nu(X) & \text{if } f(X) = g(X) + h(X) \\ Dg_\nu(X) \cdot h(\nu) + g(\nu) \cdot Dh_\nu(X) & \text{if } f(X) = g(X) \cdot h(X) \\ \sum_{i \in I} Df_\nu(X) & \text{if } f(X) = \sum_{i \in I} f_i(X). \end{cases}$$

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$$\nu_{i+1} = \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } \quad X = Df_{\nu_i}(X) + \delta_i$$

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$\nu_{i+1} = \nu_i + \Delta_i$ where Δ_i solution of $X = Df_{\nu_i}(X) + \delta_i$

But does δ_i always exist?

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$$\nu_{i+1} = \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } \quad X = Df_{\nu_i}(X) + \delta_i$$

But does δ_i always exist? Proposition: Yes.

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$\nu_{i+1} = \nu_i + \Delta_i$ where Δ_i solution of $X = Df_{\nu_i}(X) + \delta_i$

But does δ_i always exist? Proposition: Yes.

But ν_{i+i} depends on your choice of δ_i !

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$$\nu_{i+1} = \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } \quad X = Df_{\nu_i}(X) + \delta_i$$

But does δ_i always exist? Proposition: Yes.

But ν_{i+i} depends on your choice of δ_i ! Theorem: No, it doesn't.

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$$\nu_{i+1} = \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } \quad X = Df_{\nu_i}(X) + \delta_i$$

But does δ_i always exist? Proposition: Yes.

But ν_{i+i} depends on your choice of δ_i ! Theorem: No, it doesn't.

Can't you give a closed form for ν_{i+1} ?

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$$\nu_{i+1} = \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } \quad X = Df_{\nu_i}(X) + \delta_i$$

But does δ_i always exist? Proposition: Yes.

But ν_{i+i} depends on your choice of δ_i ! Theorem: No, it doesn't.

Can't you give a closed form for ν_{i+1} ? Proposition: Yes.

The difference $f(\nu_i) - \nu_i$

Solution: Replace $f(\nu_i) - \nu_i$ by any δ_i such that $f(\nu_i) = \nu_i + \delta_i$

$$\nu_{i+1} = \nu_i + \Delta_i \quad \text{where } \Delta_i \text{ solution of } \quad X = Df_{\nu_i}(X) + \delta_i$$

But does δ_i always exist? Proposition: Yes.

But ν_{i+i} depends on your choice of δ_i ! Theorem: No, it doesn't.

Can't you give a closed form for ν_{i+1} ? Proposition: Yes.

The least solution of $X = Df_{\nu_i}(X) + \delta_i$ is $Df^*_{\nu_i}(\delta_i) := \sum_{j=0}^{\infty} Df^j_{\nu_i}(\delta_i)$

and so: $\nu_{i+1} = \nu_i + Df^*_{\nu_i}(\delta_i)$

Theorem [EKL DLT07]: Let $X = f(X)$ be an equation over an arbitrary ω-continuous semiring. The sequence

$$\begin{aligned} \nu_0 &= f(0) \\ \nu_{i+1} &= \nu_i + Df^*_{\nu_i}(\delta_i) \end{aligned}$$

where δ_i satisfies $f(\nu_i) = \nu_i + \delta_i$ exists, is unique and satisfies

$$k_i \sqsubseteq \nu_i \sqsubseteq \mu f$$

for every $i \geq 0$.

Multivariate case

Systems of equations:

- ν_i, Δ_i, δ_i become vectors (elements of S^n)
- The differential becomes a function $S^n \longrightarrow S^n$
  Geometric intuition: $Df_{\nu_i}(X_1, \ldots, X_n)$ is the hyperplane tangent to f at the (n-dimensional) point ν_i

Derivation trees I

An equation $X = f(X)$ induces a context-free grammar $G : X \to f(X)$

Examples:

$$X = 0.7X^2 + 0.3 \quad \text{induces} \quad X \to 0.7\ X\ X \mid 0.3$$

$$X = 0.2XY + 0.8 \quad \text{induces} \quad X \to 0.2\ X\ Y \mid 0.8$$

$$Y = 0.7XY + 0.3 \qquad\qquad Y \to 0.7\ X\ Y \mid 0.3$$

(Actually one grammar for each variable, differing only in the axiom.)

Derivation trees II

Assign to a derivation tree t its yield $Y(t)$:

$$Y(t) = \text{(ordered) product of } t\text{'s leaves}$$

Assign to a set T of derivation trees its yield $Y(T)$

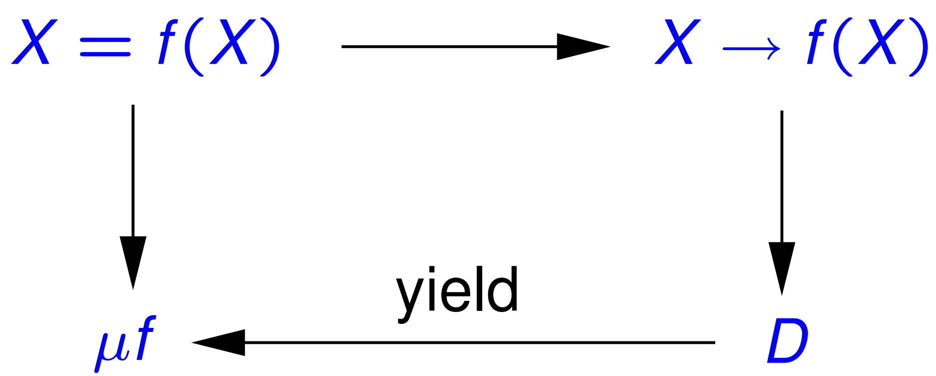
$$Y(T) = \sum_{t \in T} Y(t)$$

Example: $X \rightarrow 0.7\, X\, X \mid 0.3$

Derivation trees III

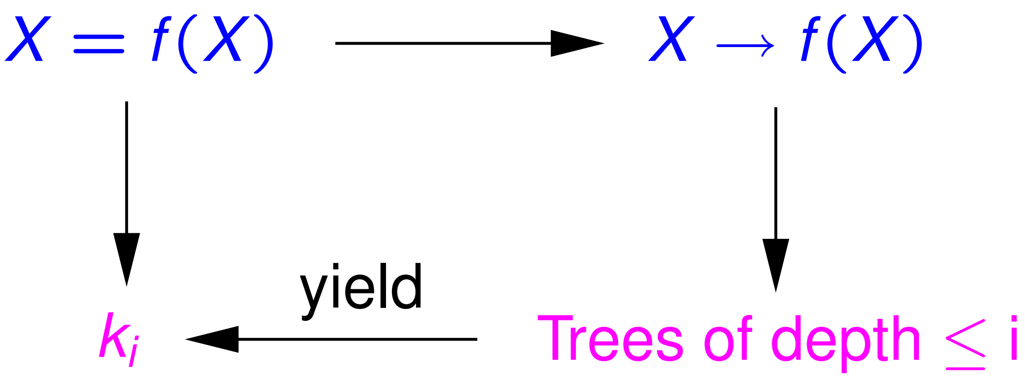
Proposition: Let D be the set of all derivation trees of G. Then

$$\mu f = Y(D)$$



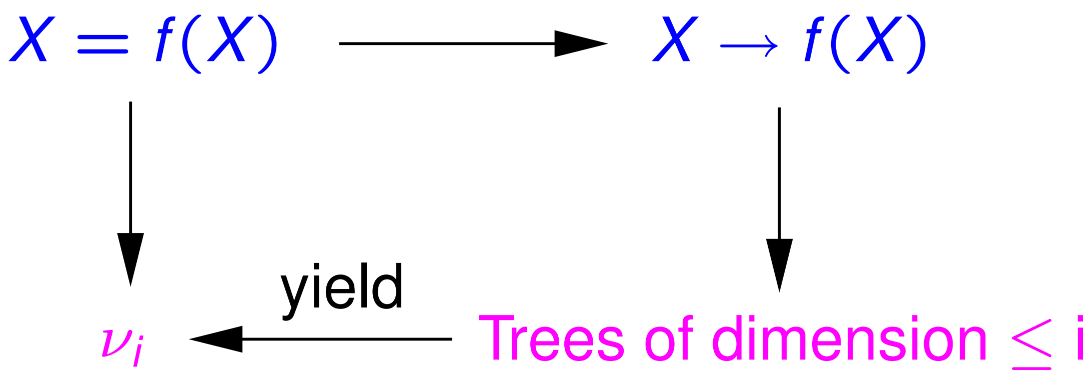
Approximants as yields: Kleene

Proposition: The i-th Kleene approximant k_i is the yield of all derivation trees of depth at most i.



Approximants as yields: Newton

Main Theorem: The *i*-th Newton approximant ν_i is the yield of all derivation trees of dimension at most *i*.

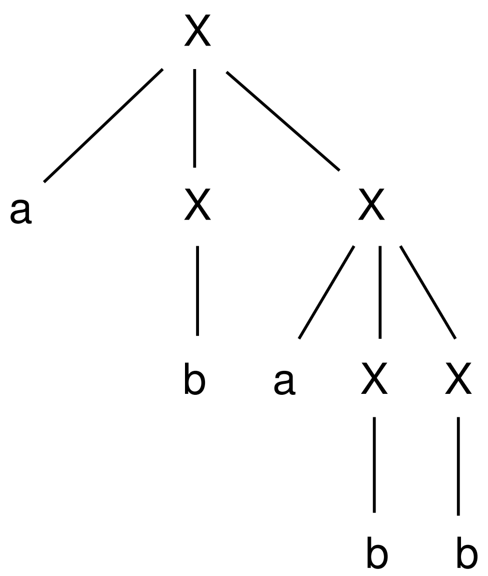


Understanding dimension I

A derivation tree has dimension k if at least one of its derivations

$$X \Rightarrow w_1 \Rightarrow w_2 \ldots \Rightarrow w_n \Rightarrow w$$

satisfies that all of $w_1, \ldots, w_n$ contain at most k occurrences of non-terminals (and at least one of them contains k occurrences).



$X \Rightarrow aXX \Rightarrow abX \Rightarrow abaXX \Rightarrow ababX \Rightarrow abaaa$

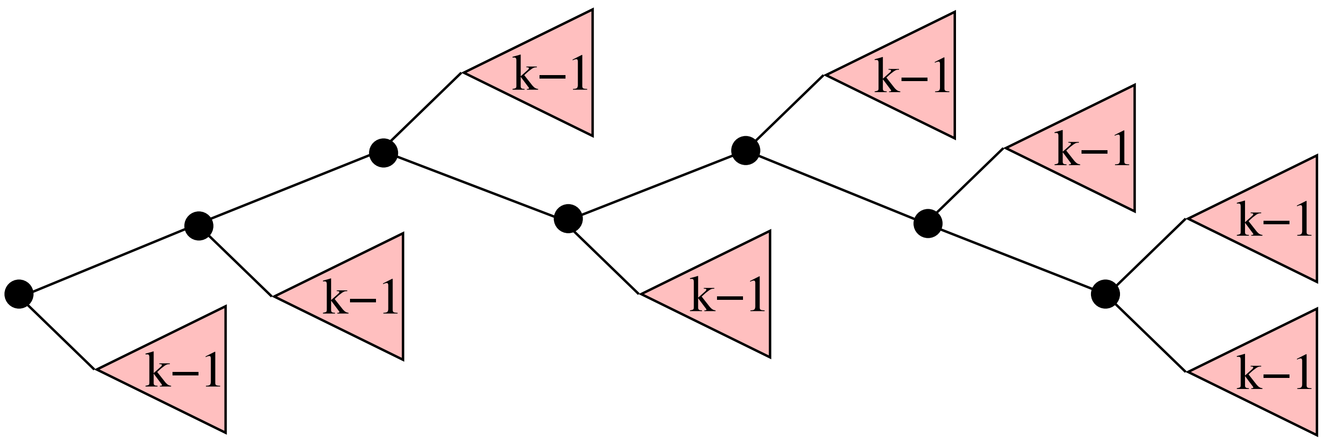
Understanding dimension II

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

Understanding dimension II

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 $k - 1$).



The proof

Theorem [EKL DLT07]: Let $X = f(X)$ be an equation over an arbitrary ω-continuous semiring. The Newton sequence $\{\nu_i\}_{i \geq 0}$ is unique and satisfies $k_i \sqsubseteq \nu_i \sqsubseteq \mu f$ for every $i \geq 0$.

The proof

Theorem [EKL DLT07]: Let $X = f(X)$ be an equation over an arbitrary ω-continuous semiring. The Newton sequence $\{\nu_i\}_{i\geq 0}$ is unique and satisfies $k_i \sqsubseteq \nu_i \sqsubseteq \mu f$ for every $i \geq 0$.

Proof:

Uniqueness: follows from tree characterization.

The proof

Theorem [EKL DLT07]: Let $X = f(X)$ be an equation over an arbitrary ω-continuous semiring. The Newton sequence $\{\nu_i\}_{i \geq 0}$ is unique and satisfies $k_i \sqsubseteq \nu_i \sqsubseteq \mu f$ for every $i \geq 0$.

Proof:

Uniqueness: follows from tree characterization.

$k_i \sqsubseteq \nu_i$: trees of depth i have dimension at most i.

The proof

Theorem [EKL DLT07]: Let $X = f(X)$ be an equation over an arbitrary ω-continuous semiring. The Newton sequence $\{\nu_i\}_{i \geq 0}$ is unique and satisfies $k_i \sqsubseteq \nu_i \sqsubseteq \mu f$ for every $i \geq 0$.

Proof:

Uniqueness: follows from tree characterization.

$k_i \sqsubseteq \nu_i$: trees of depth *i* have dimension at most *i*.

$\nu_i \sqsubseteq \mu f$: the yield of all trees of dimension at most *i* is smaller than or equal to the yield of all trees.

Idempotent semirings: derivation tree analysis

Idempotent semiring: $a + a = a$

Technique for computing μf algebraically:

(1) Identify a set T of derivation trees such that $Y(T)$ can be computed algebraically.

(2) Show that $Y(t) \sqsubseteq Y(T)$ holds for every derivation tree t.

Idempotent semirings: derivation tree analysis

Idempotent semiring: $a + a = a$

Technique for computing μf algebraically:

(1) Identify a set T of derivation trees such that $Y(T)$ can be computed algebraically.

(2) Show that $Y(t) \sqsubseteq Y(T)$ holds for every derivation tree t.

$$
\begin{aligned}
\mu f &= Y(D) && \text{(proposition above)} \\
&= \sum_{t \in D} Y(t) && \text{(definition of yield)} \\
&\subseteq \sum_{t \in D} Y(T) && (Y(t) \subseteq Y(T)) \\
&= Y(T) && \text{(idempotence)}
\end{aligned}
$$

Commutative idempotent semirings

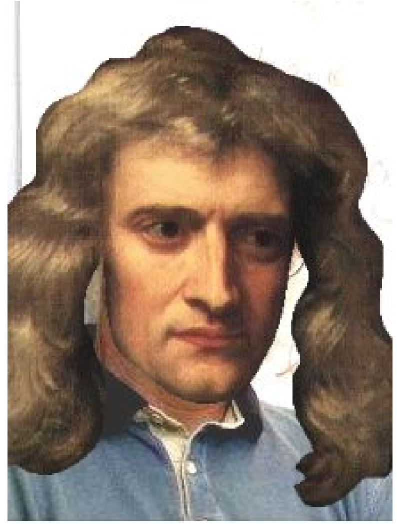
Theorem [Hopkins-Kozen LICS ’99]: The least fixed point of a system $X = f(X)$ of n equations over an ω-continuous idempotent and commutative semiring is reached by the sequence

$$\begin{aligned} \nu_0 &= f(0) \\ \nu_{i+1} &= J(\nu_i)^* \cdot f(\nu_i) \end{aligned}$$

after at most $O(3^n)$ iterations.

Commutative idempotent semirings

Theorem [Hopkins-Kozen LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an ω-continuous idempotent and commutative semiring is reached by the sequence

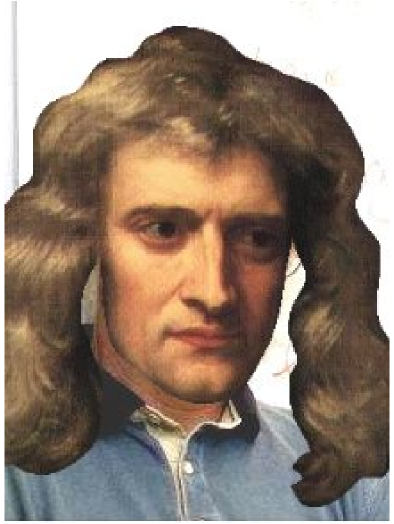


$$\begin{aligned} \nu_0 &= f(0) \\ \nu_{i+1} &= J(\nu_i)^* \cdot f(\nu_i) \end{aligned}$$

after at most $O(3^n)$ iterations.

Commutative idempotent semirings

Theorem [Hopkins-Kozen LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an ω-continuous idempotent and commutative semiring is reached by the sequence

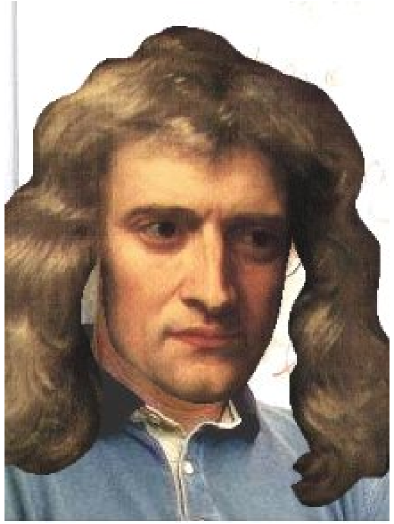


$$\begin{aligned} \nu_0 &= f(0) \\ \nu_{i+1} &= J(\nu_i)^* \cdot f(\nu_i) \end{aligned}$$

after at most $O(3^n)$ iterations.

Theorem [EKL STACS'07]: This is exactly Newton's sequence.

Commutative idempotent semirings



Theorem [Hopkins-Kozen LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an ω-continuous idempotent and commutative semiring is reached by the sequence

$$\begin{aligned} \nu_0 &= f(0) \\ \nu_{i+1} &= J(\nu_i)^* \cdot f(\nu_i) \end{aligned}$$

after at most $O(3^n)$ iterations.

Theorem [EKL STACS'07]: <u>This is exactly Newton's sequence</u>.

The fixed point is reached after at most n iterations, i.e. $\mu f = \nu_n$.

Proof with derivation tree analysis

Lemma: Let $X = f(X)$ be a system of n equations over an ω-continuous idempotent and commutative semiring.
For every derivation tree t there is another tree t' of dimension at most n such that $Y(t) = Y(t')$.

Theorem: $\mu f = \nu_n$.

Proof: Let T_n be the set of trees of dimension n. Then $Y(T_n) = \nu_n \sqsubseteq \mu f$.

$$\begin{aligned}
\mu f = \sum_{t \in D} Y(t) &= \sum_{t \in D} Y(t') && \text{(definition of yield, } Y(t) = Y(t')\text{)} \\
&= \sum_{t \in T_n} Y(t') && (t' \in T_n\text{, idempotence)} \\
&\sqsubseteq Y(T_n) = \nu_n
\end{aligned}$$

An example

The Newton sequence terminates for all idempotent and commutative analyses, the Kleene sequence does not.

$$
\begin{aligned}
X &= a \cdot X \cdot X + b \\
f'(X) &= a \cdot X + a \cdot X = a \cdot X
\end{aligned}
$$

For one equation: $\mu f = \nu_1 = f'(\nu_0)^* \cdot \nu_0$

We obtain:

$$
\begin{aligned}
\nu_0 &= b \\
\nu_1 &= (ab)^* b
\end{aligned}
$$

Other results proved by derivation tree analysis

Star-distributive commutative semirings: $(a + b)^* = a^* + b^*$.

$$\mu f = Df^*_{f^n(0)}(f(0)) \cdot f(0)$$

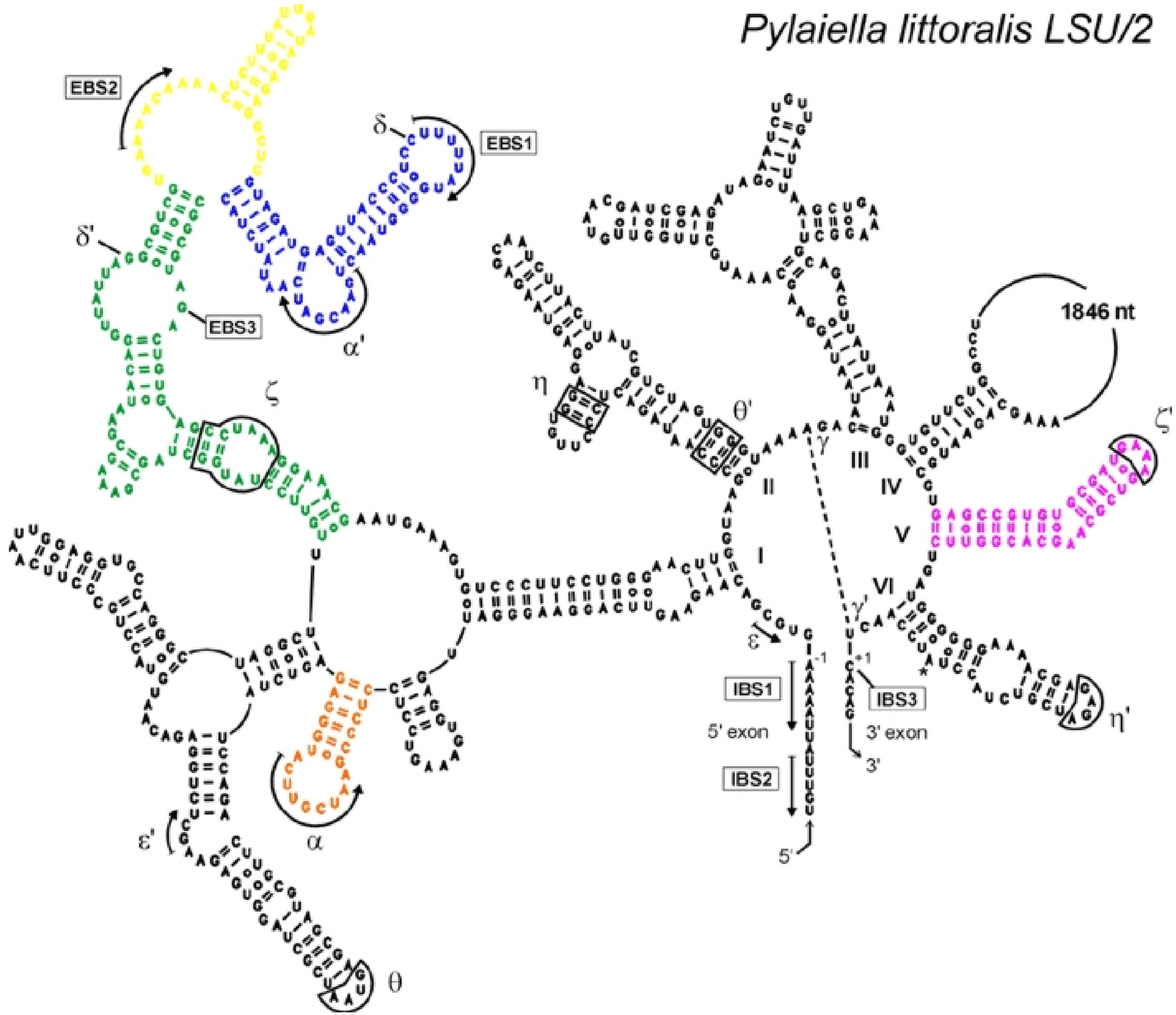
(improving the complexity of an algorithm for computing throughput of context free grammars due to Caucal et al.)

Lossy semirings: $a \sqsubseteq 1$ for every $a \neq 0$.

$$\mu f = Df^*_{f^n(0)}(f(0)) \cdot f(0)$$

(algebraic version of a result by Courcelle)

Having fun: Secondary structure of RNA



(image by Bassi, Costa, Michel; www.cgm.cnrs-gif.fr/michel/)

An stochastic context-free grammar

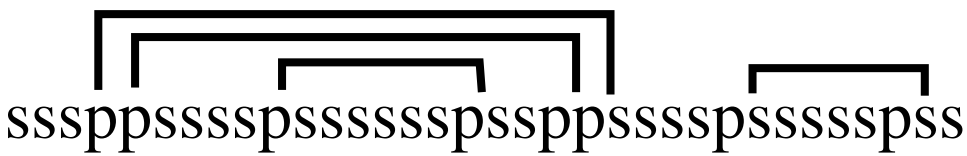
[Knudsen, Hein 99]: Model the distribution of secondary structures as the derivation trees of the following stochastic context-free grammar:

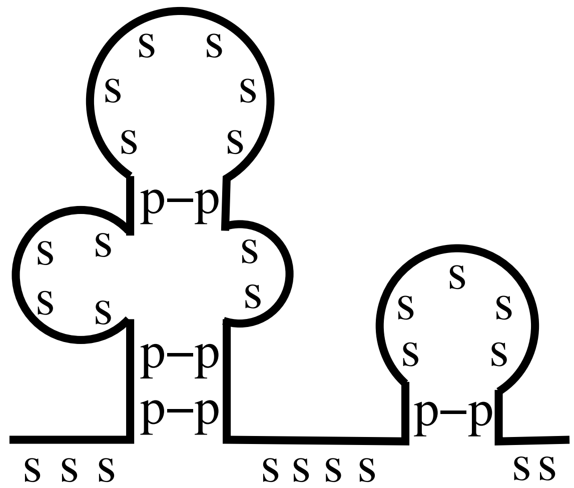
$$L \xrightarrow{0.869} CL \qquad L \xrightarrow{0.131} C$$

$$S \xrightarrow{0.788} pSp \qquad S \xrightarrow{0.212} CL$$

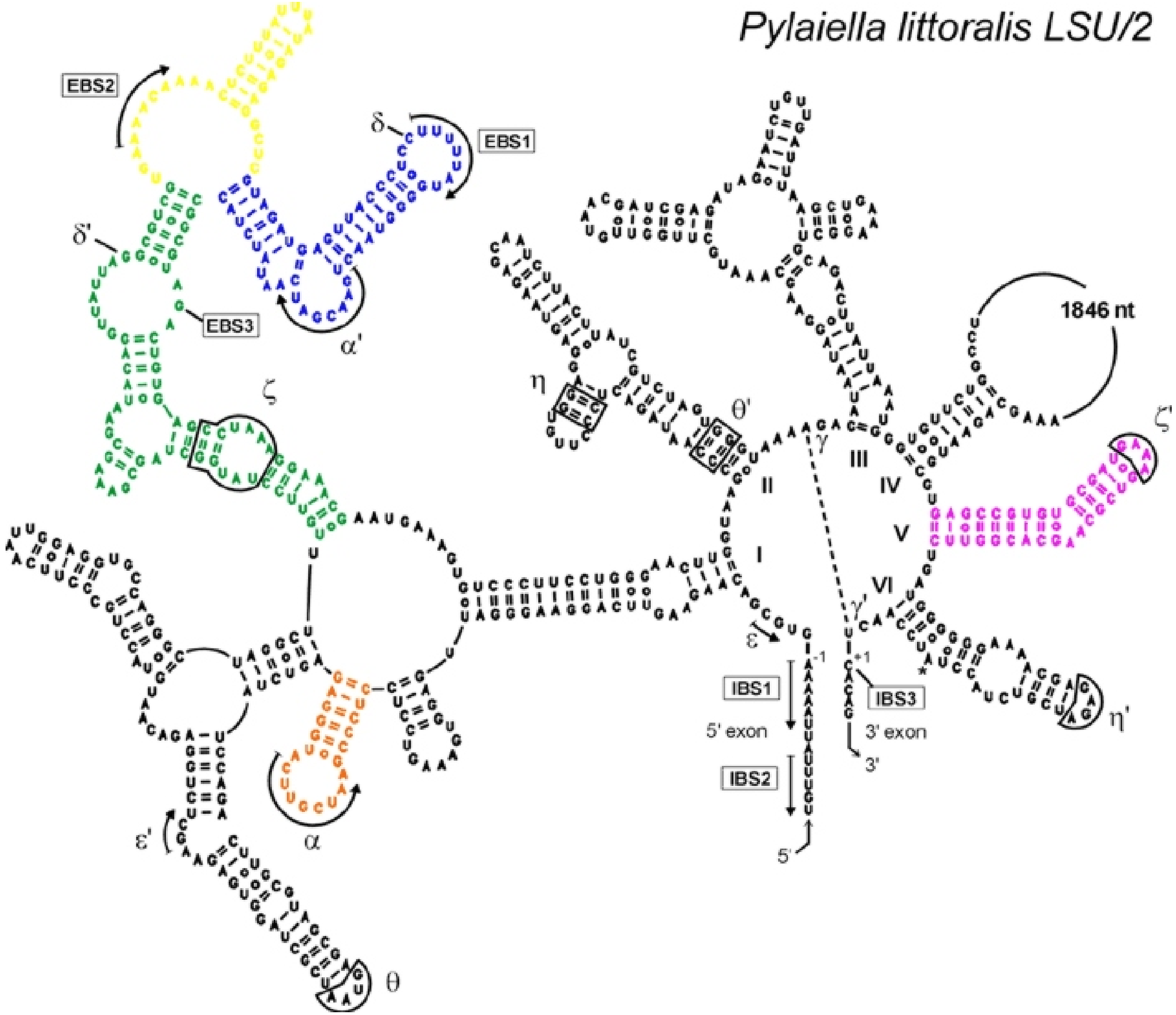
$$C \xrightarrow{0.895} s \qquad C \xrightarrow{0.105} pSp$$

Graphical interpretation:

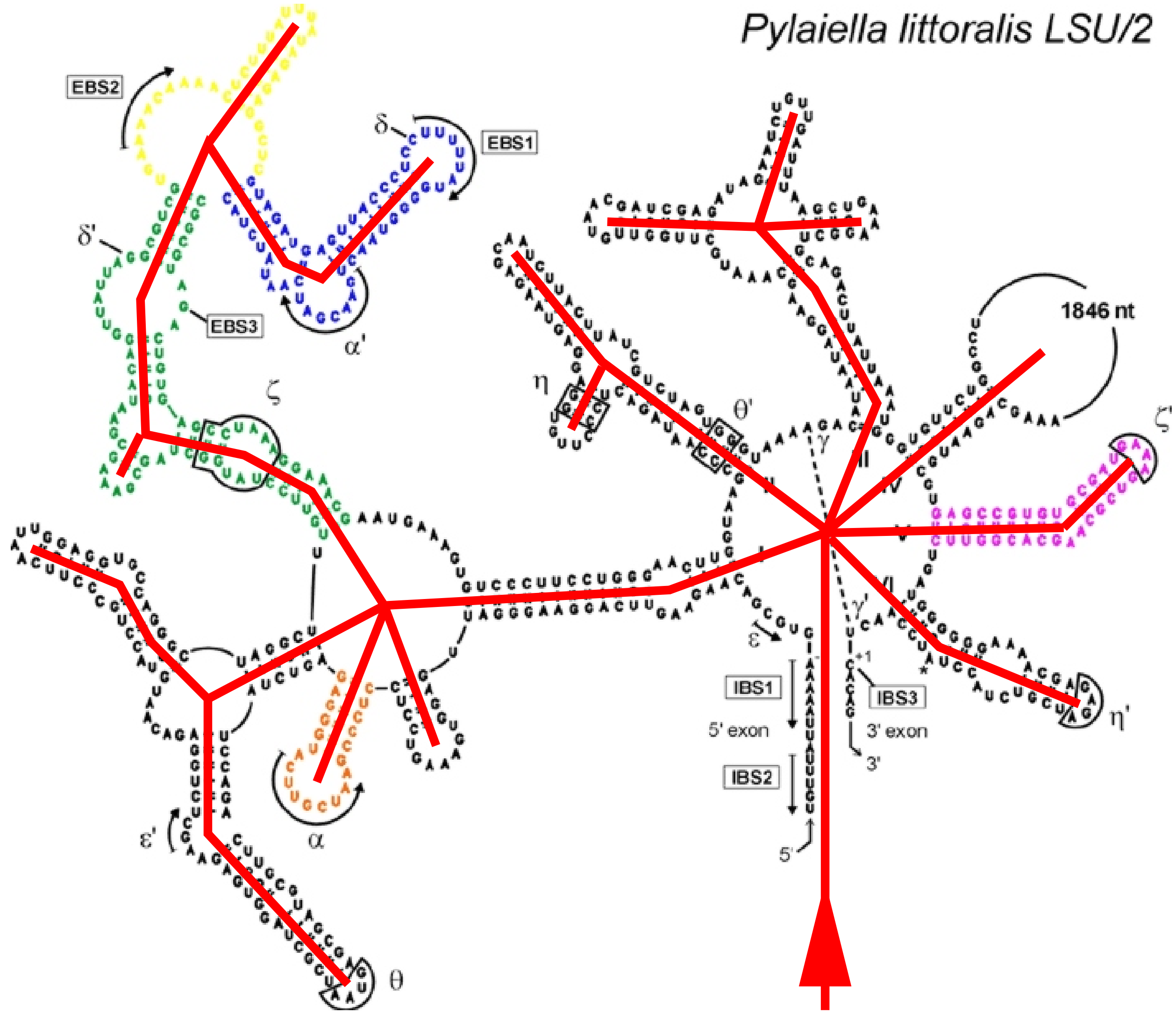




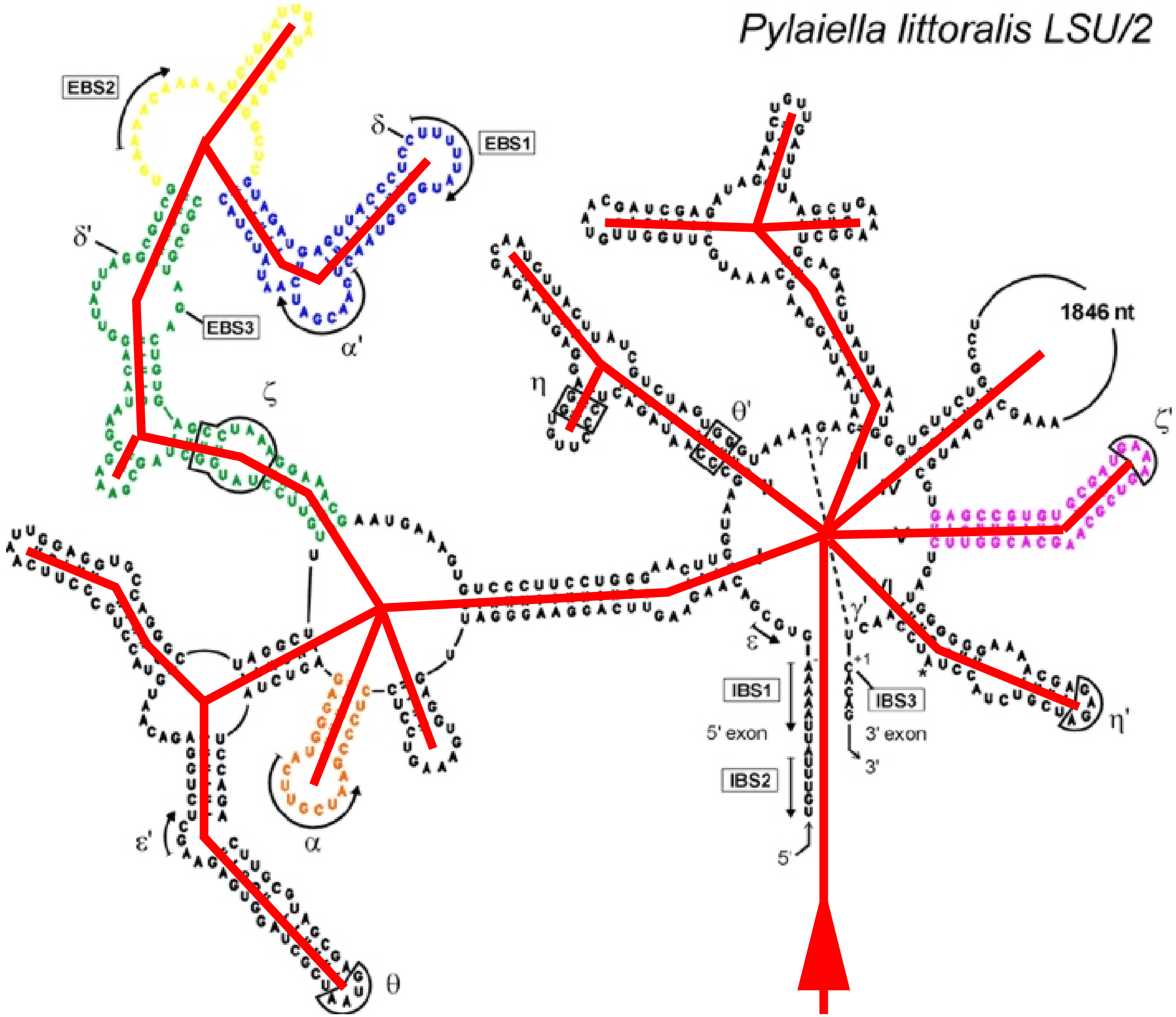
Visualizing the index of a derivation



Visualizing the index of a derivation



Visualizing the index of a derivation



Dimension = depth of the red tree + 1

Grammar leads to two equation systems:

$$
\begin{aligned}
L &= C \cdot L + C \\
S &= p \cdot S \cdot p + C \cdot L \\
C &= s + p \cdot S \cdot p
\end{aligned}
\qquad
\begin{aligned}
\widehat{L} &= 0.869 \cdot \widehat{C} \cdot \widehat{L} + 0.131 \cdot \widehat{C} \\
\widehat{S} &= 0.788 \cdot \widehat{S} + 0.212 \cdot \widehat{C} \cdot \widehat{L} \\
\widehat{C} &= 0.895 + 0.105 \cdot \widehat{S}
\end{aligned}
$$

$$
\begin{aligned}
\nu_0(L) &= \text{der. of dim.} \leq 1 \\
\nu_1(L) &= \text{der. of dim.} \leq 2 \\
\nu_2(L) &= \text{der. of dim.} \leq 3 \\
\nu_3(L) &= \text{der. of dim.} \leq 4 \\
\nu_4(L) &= \text{der. of dim.} \leq 5 \\
\nu_5(L) &= \text{der. of dim.} \leq 6
\end{aligned}
\qquad
\begin{aligned}
\widehat{\nu}_0(L) &= 0.5585 \\
\widehat{\nu}_1(L) &= 0.8050 \\
\widehat{\nu}_2(L) &= 0.9250 \\
\widehat{\nu}_3(L) &= 0.9789 \\
\widehat{\nu}_4(L) &= 0.9972 \\
\widehat{\nu}_5(L) &= 0.9999
\end{aligned}
$$

Conclusions

Newton did it all