Solving fixed-point equations on semirings

Javier Esparza

Technische Universität München

Joint work with

Stefan Kiefer and Michael Luttenberger

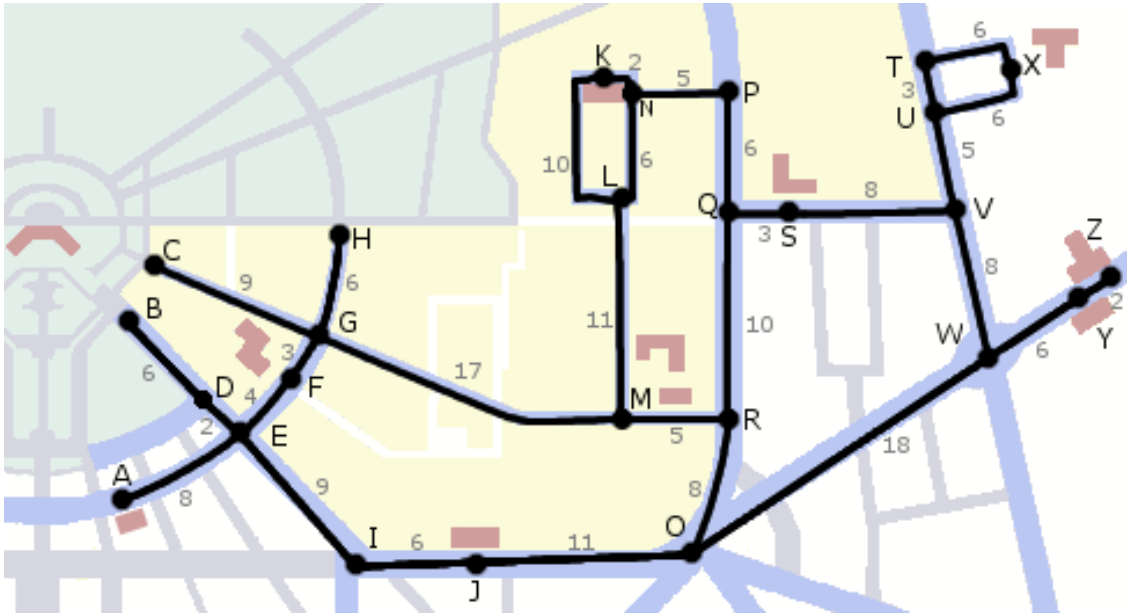
Fixed-point equations

We study systems of equations of the form

$$\begin{aligned} X_1 &= f_1(X_1, \ldots, X_n) \\ X_2 &= f_2(X_1, \ldots, X_n) \\ &\ldots \\ X_n &= f_n(X_1, \ldots, X_n) \end{aligned}$$

where the f_i's are "polynomial expressions".

Shortest paths



Lengths d_i of shortest paths from vertex 0 to vertex i in graph $G = (V, E)$ are the largest solution of

$$d_i = \min_{(i,j) \in E} (d_i, d_j + w_{ji})$$

where w_{ij} is the distance from i to j.

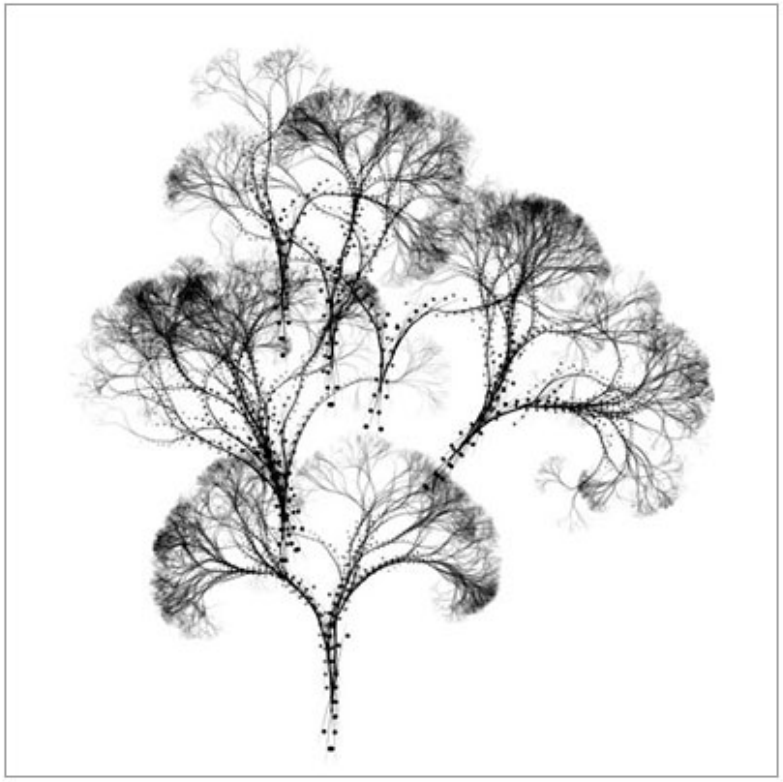
Context-free languages

Context-free grammar

$$\begin{aligned} X &\to ZY \mid Z \\ Y &\to aYa \mid ZX \\ Z &\to b \mid aYa \end{aligned}$$

Languages generated from X, Y, Z are the least solution of

$$\begin{aligned} L_X &= (L_Z \cdot L_Y) \cup L_Z \\ L_Y &= (\{a\} \cdot L_Y \cdot \{a\}) \cup (L_Z \cdot L_X) \\ L_Z &= \{b\} \cup (\{a\} \cdot L_Y \cdot \{a\}) \end{aligned}$$



Nuclear chain reaction

^{235}U ball of radius D, spontaneous fission.

Probability of a chain reaction is $(1 - p_0)$,

where p_α for $0 \leq \alpha \leq D$ is least solution of

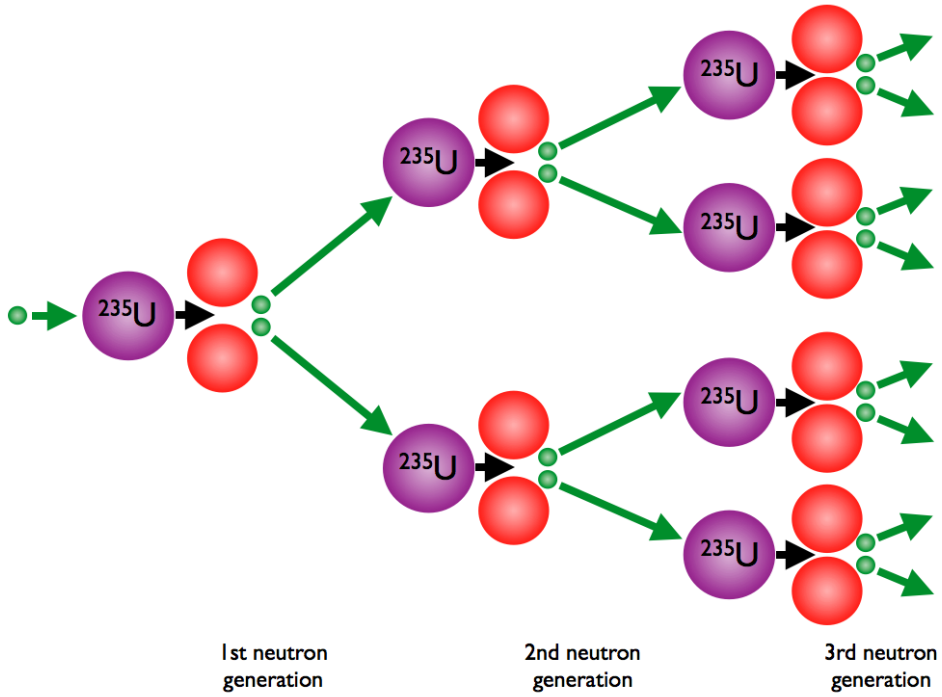
$$p_\alpha = k_\alpha + \int_0^D R_{\alpha,\beta}\, f(p_\beta)\, d\beta$$

for constants k_α, $R_{\alpha,\beta}$ and polynomial $f(x)$.

Discretizing the interval $[0, D]$ we get

$$p_i = k_i + \sum_{j=1}^{n} r_{i,j}\, f(p_j)$$

for constants k_i, $r_{i,j}$.



And many others . . .

Stochastic theory:	Stationary distribution of Markov chains
	Extinction probability of branching processes
Physics:	Heat equation
	Electrostatic equilibrium
Biology:	RNA structure prediction
	Population dynamics
Computer science:	Dataflow equations (abstract interpretation)
	Reputation systems
	Provenance in databases

Underlying structure: ω-continuous semirings

Semiring $(C, +, \times, 0, 1)$:

$(C, +, 0)$ is a commutative monoid

$(C, \times, 1)$ is a monoid

$\times$ distributes over $+$

$0 \times a = a \times 0 = 0$

ω-continuity:

the relation $a \sqsubseteq b \Leftrightarrow \exists c : a + c = b$ is a partial order

$\sqsubseteq$-chains have limits

Examples: nonnegative integers and reals plus ∞, min-plus (tropical), languages, complete lattices, multisets, Viterbi . . .

In the rest of the talk: semiring $\equiv$ ω-continuous semiring.

Research program

Develop generic solution methods valid for all semirings, or at least for large classes.

- Generic implementations.
- Exchange of algorithms and proof techniques between numerical mathematics, algebraic computation and language theory.

Research program

Develop generic solution methods valid for all semirings, or at least for large classes.

- Generic implementations.
- Exchange of algorithms and proof techniques between numerical mathematics, algebraic computation and language theory.

In this talk: brief survey of our work on Newton's method.

THE generic solution method: Kleene iteration

Theorem [Klee 38, Tars 55, Kui 97]: A system of fixed-point equations over a semiring has a least solution μf w.r.t. the natural order $\sqsubseteq$.

This least solution is the supremum of the Kleene approximants, denoted by $\{k_i\}_{i \geq 0}$, and given by

$$\begin{aligned} k_0 &= f(0) \\ k_{i+1} &= f(k_i) \ . \end{aligned}$$

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

Kleene iteration is slow

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

- $X = \{a\} \cdot X \cup \{b\} \qquad \mu f = a^* b$

  Kleene approximants are finite sets: $k_i = (\epsilon + a + \ldots + a^i)b$

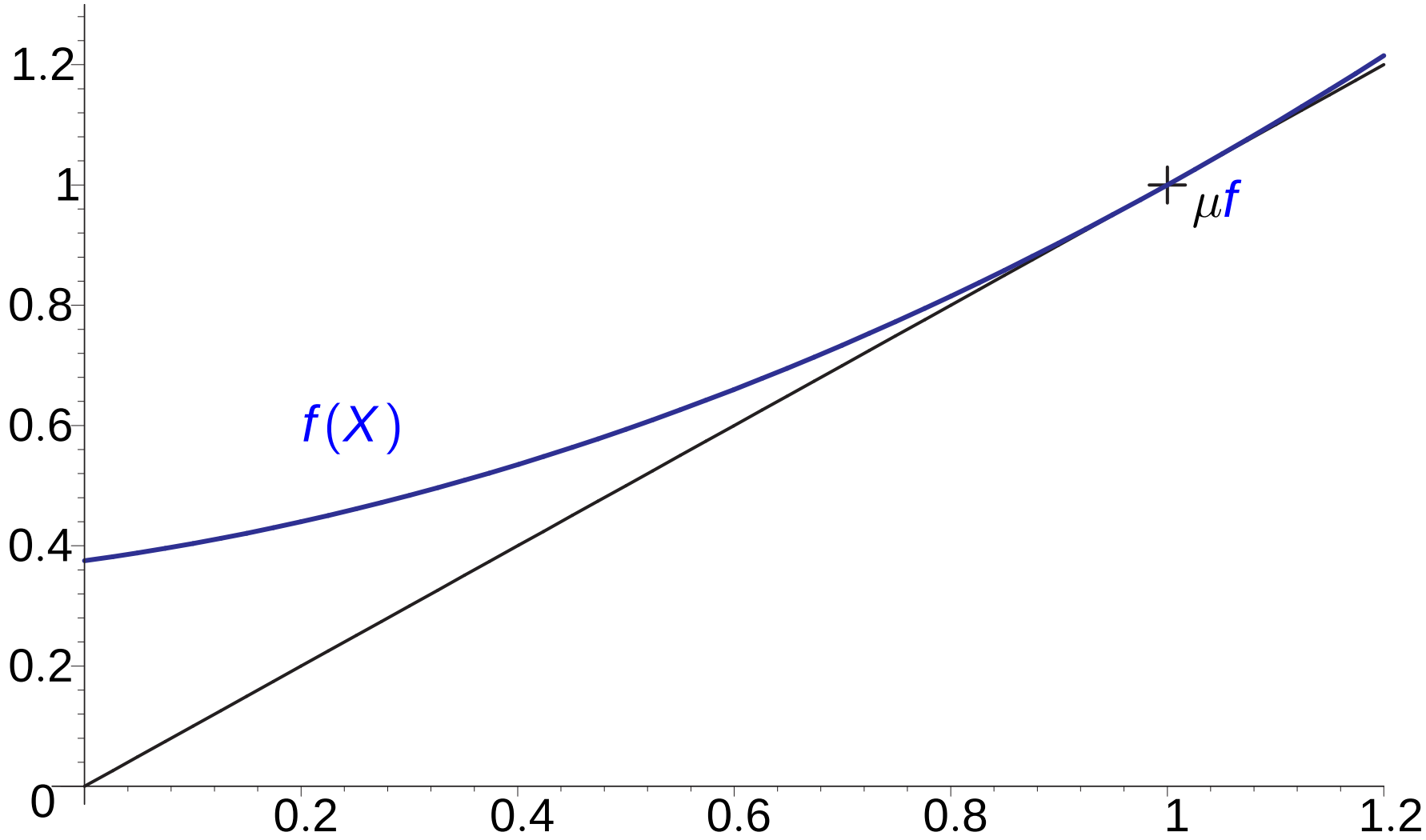
Real semiring: convergence can be very slow.

- $X = \frac{1}{2} X^2 + \frac{1}{2} \qquad \mu f = 1 = 0.99999\ldots$

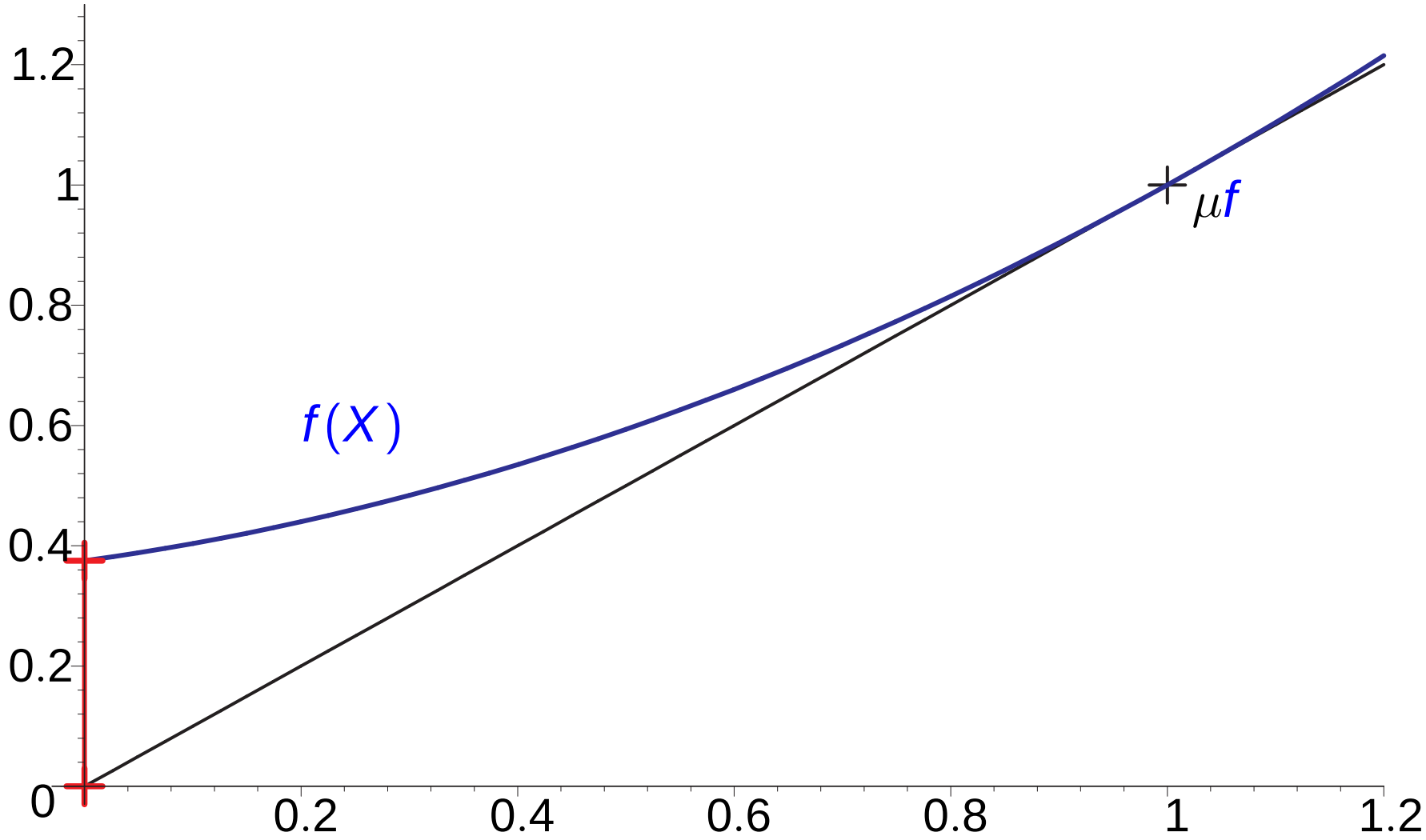
  "Logarithmic convergence": k iterations give $O(\log k)$ correct digits.

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

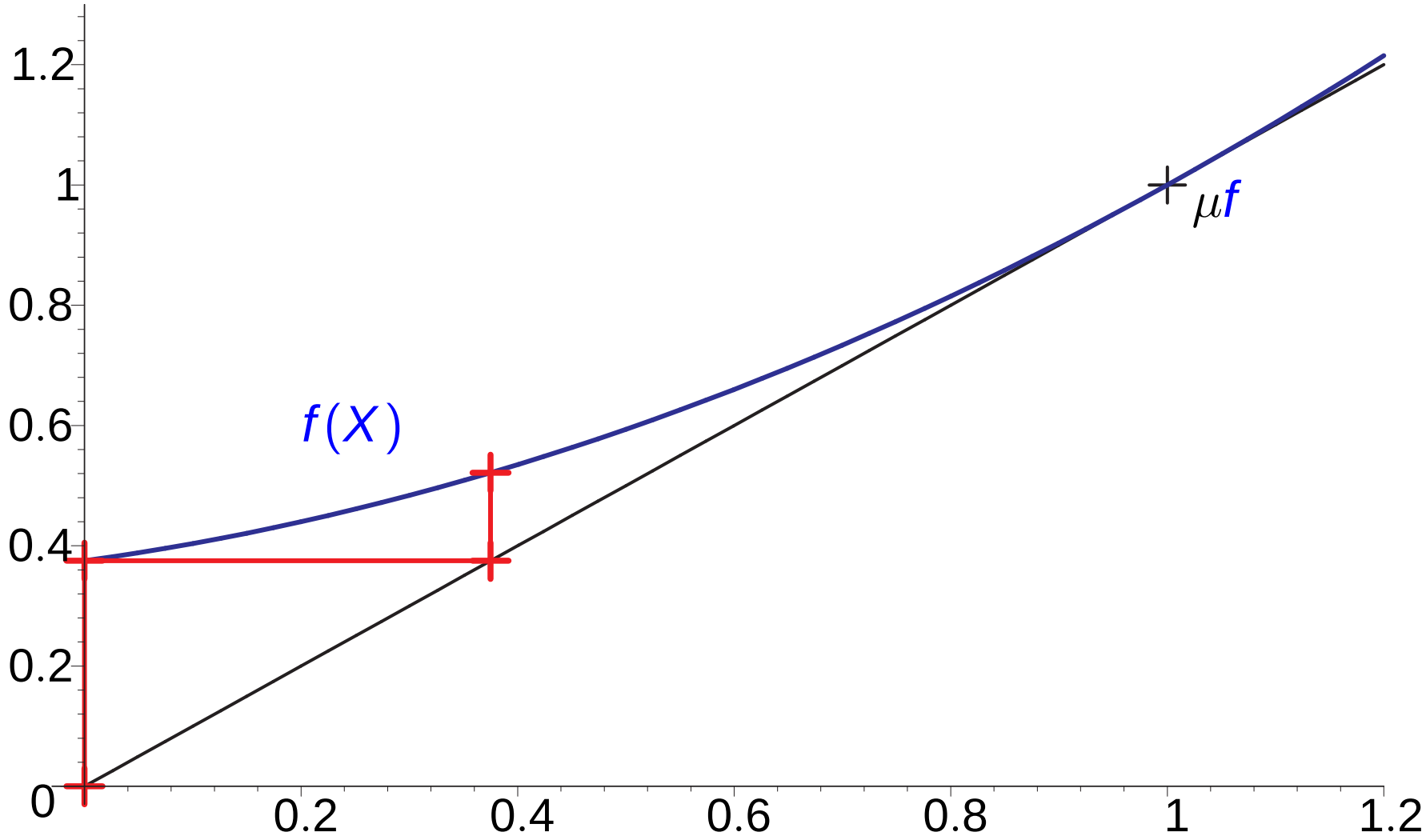
Kleene iteration for $X = f(X)$ (univariate case)



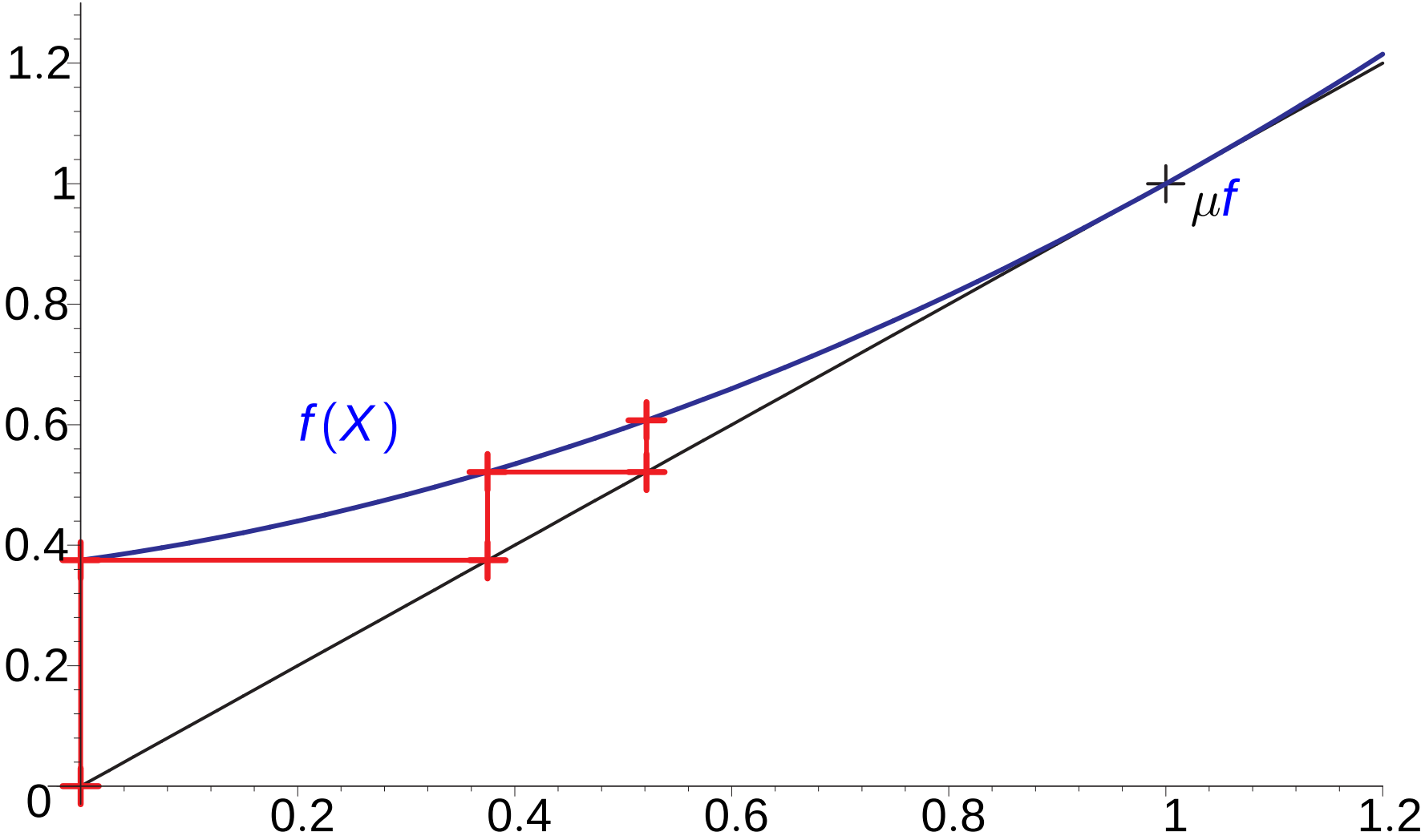
Kleene iteration for $X = f(X)$ (univariate case)



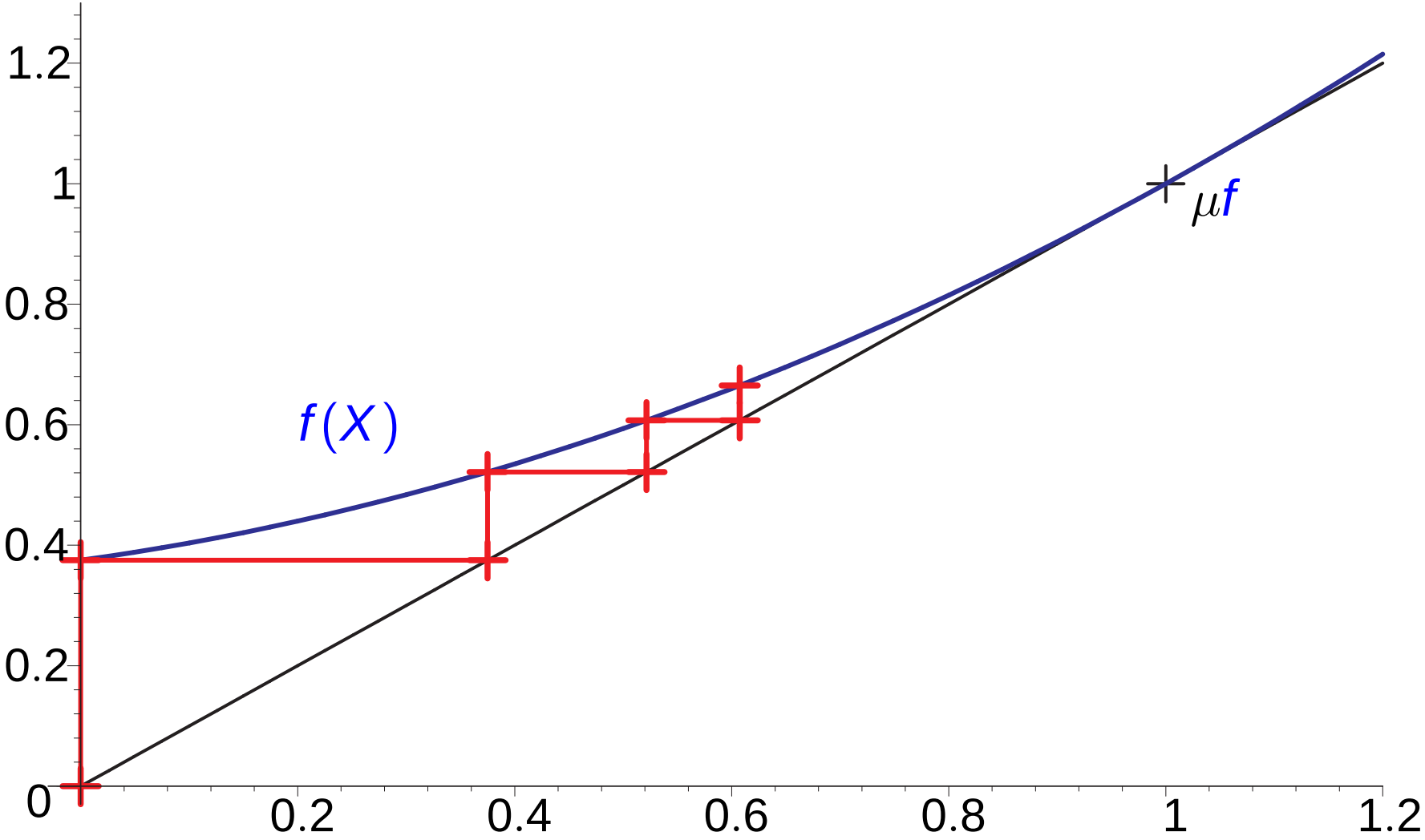
Kleene iteration for $X = f(X)$ (univariate case)



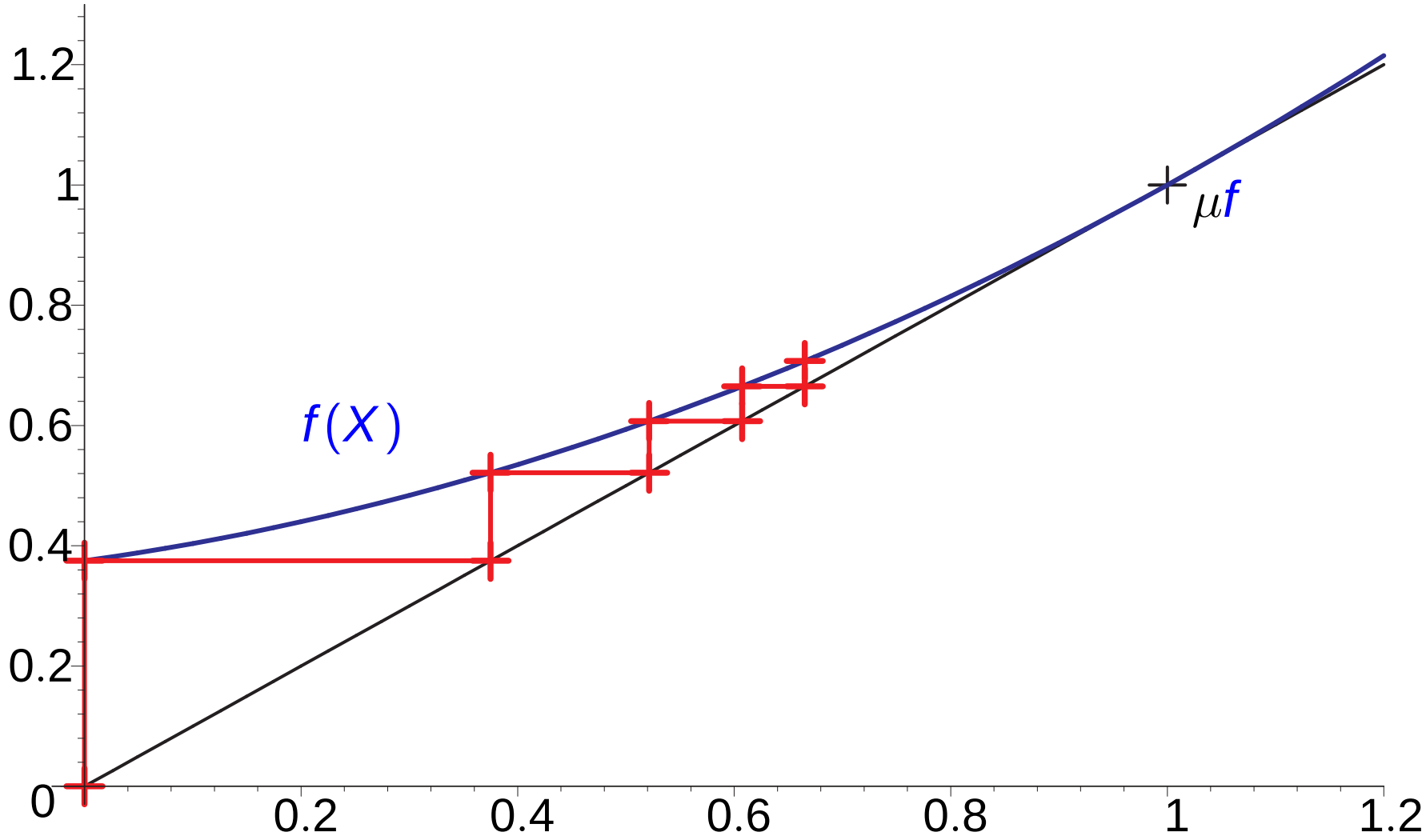
Kleene iteration for $X = f(X)$ (univariate case)



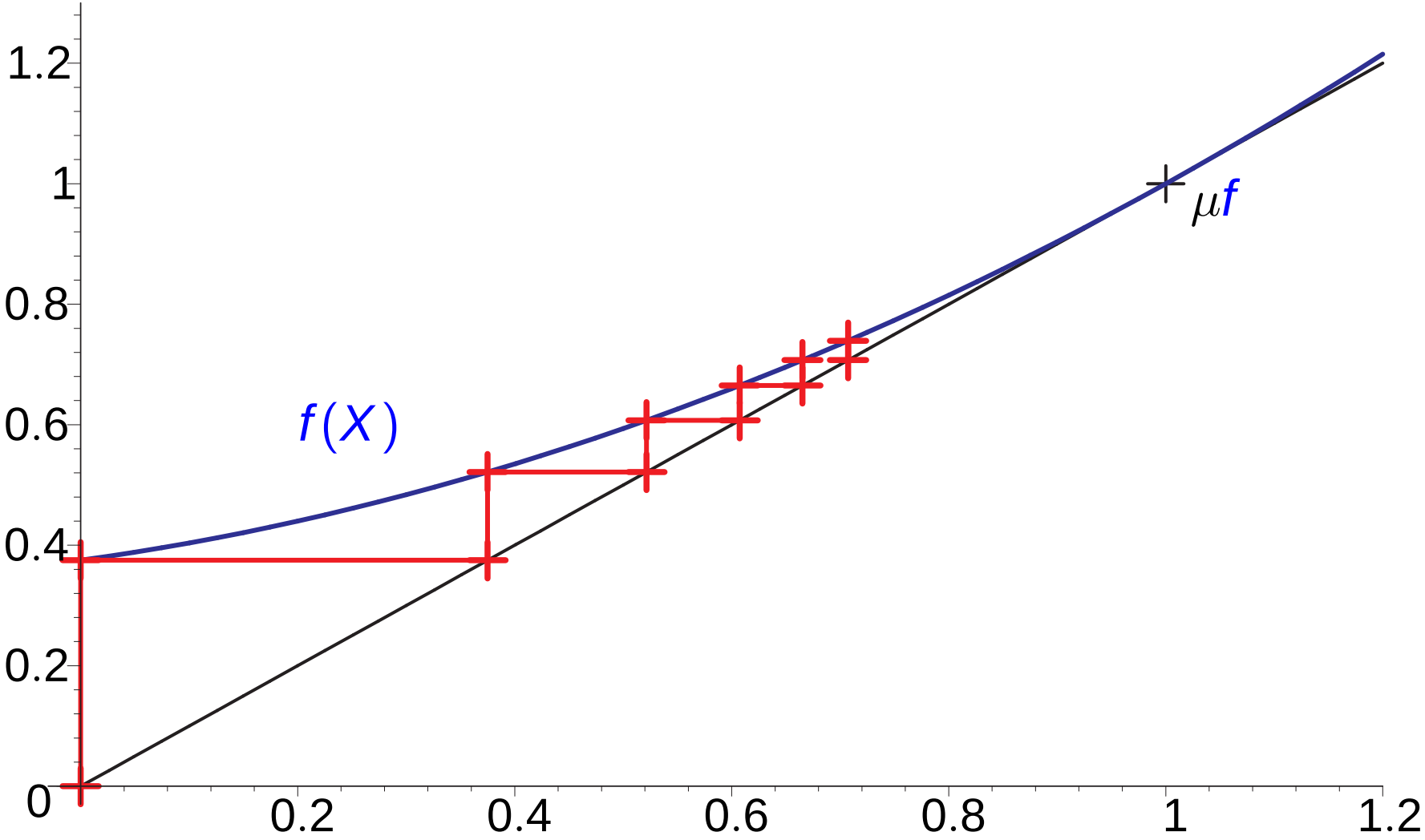
Kleene iteration for $X = f(X)$ (univariate case)



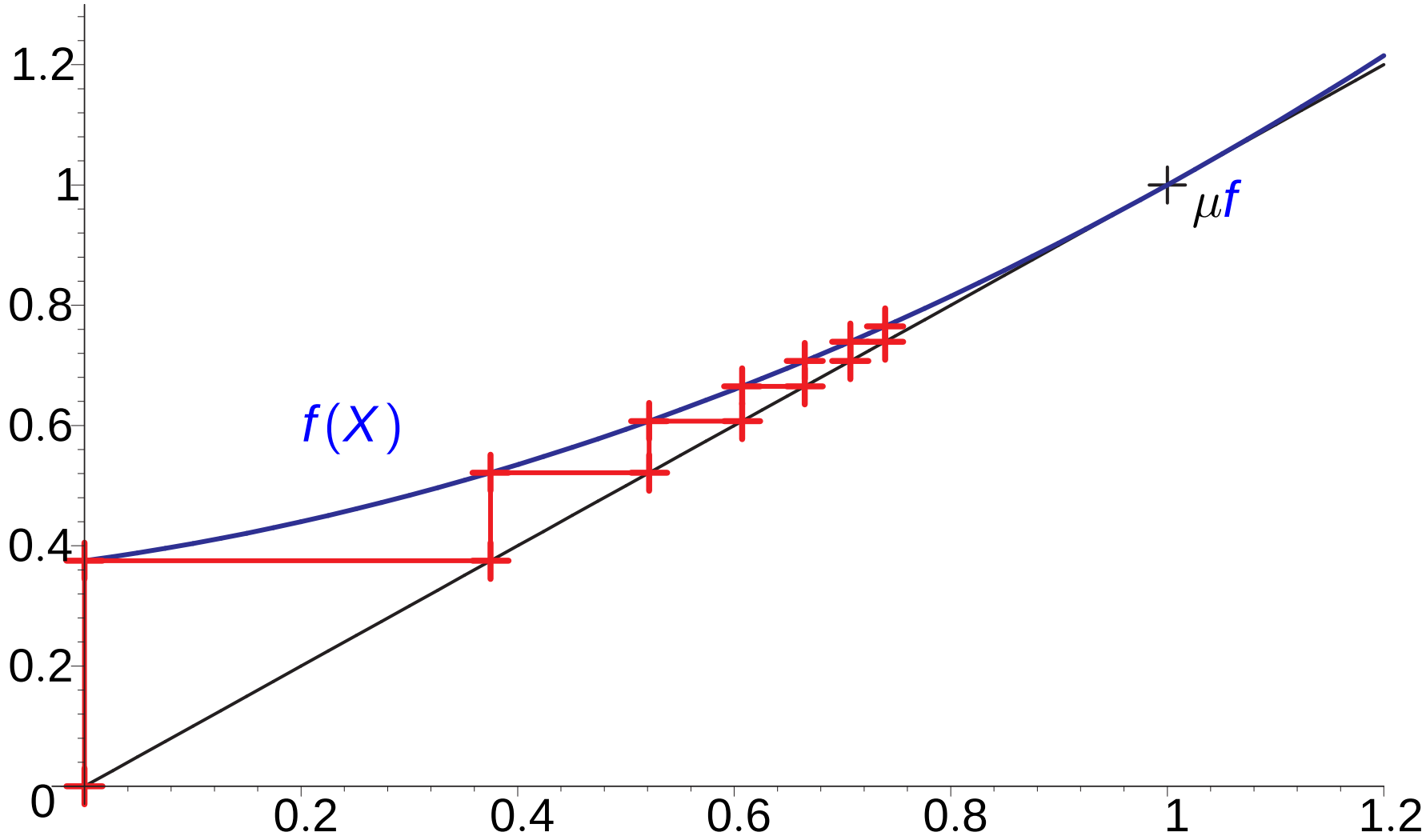
Kleene iteration for $X = f(X)$ (univariate case)



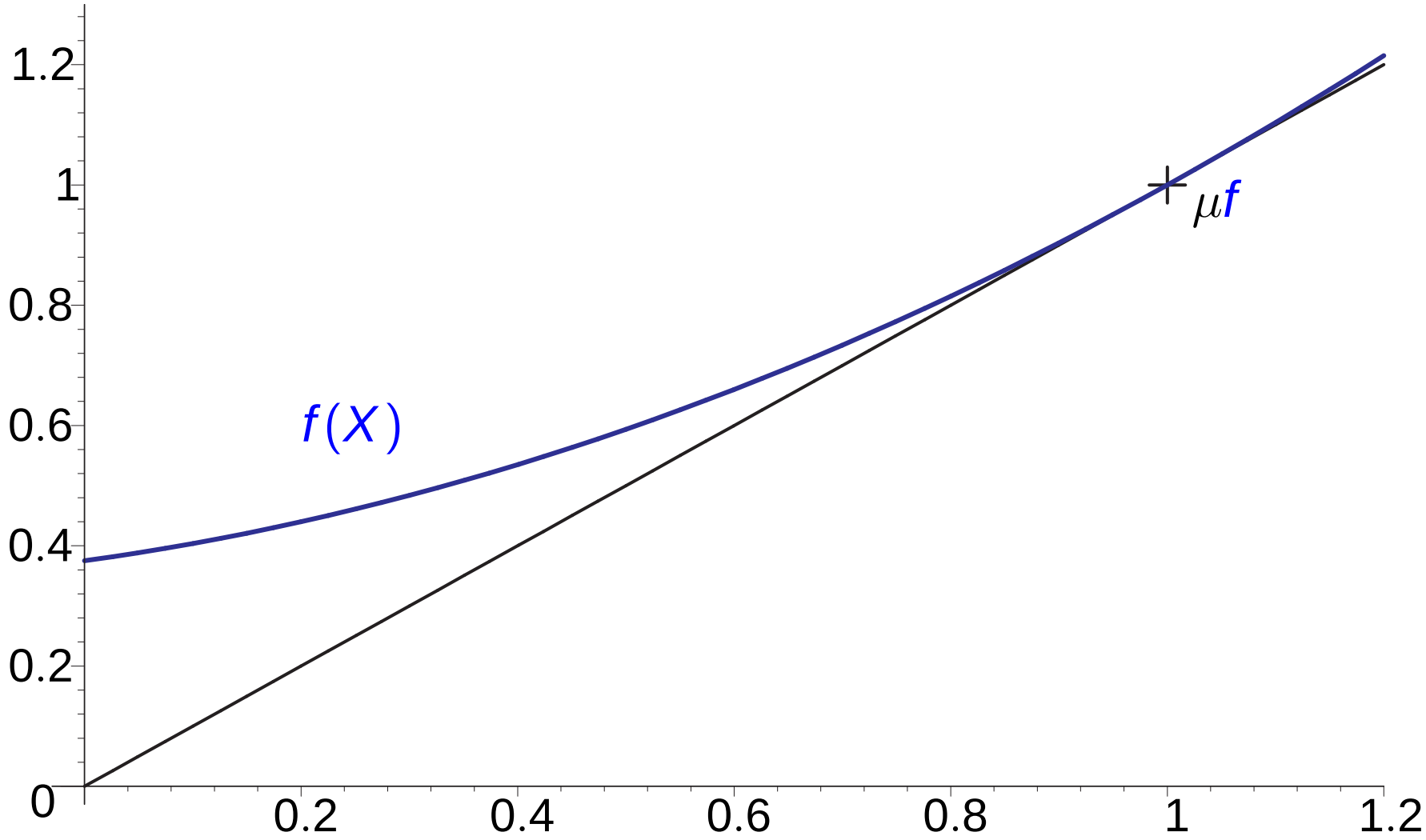
Kleene iteration for $X = f(X)$ (univariate case)



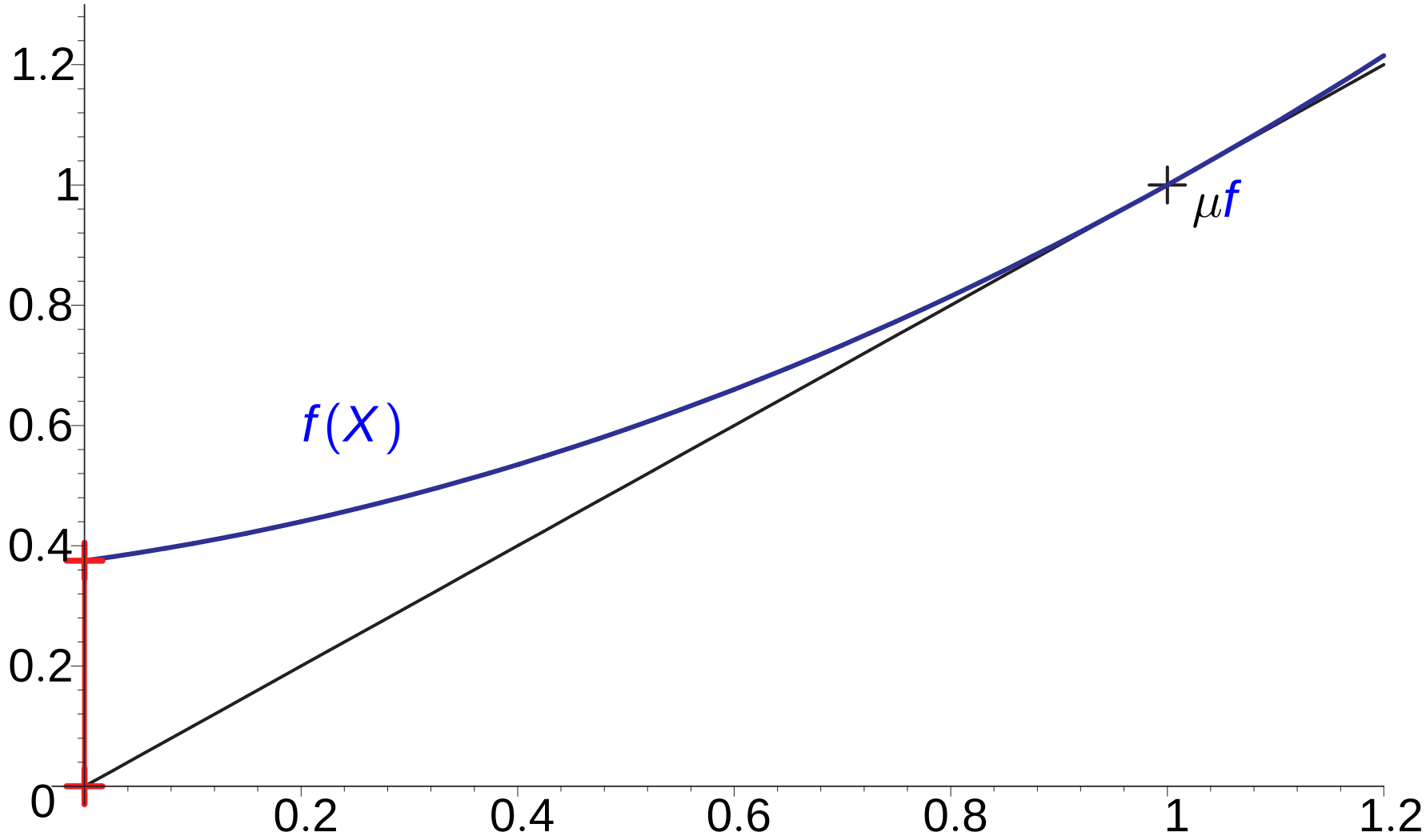
Kleene iteration for $X = f(X)$ (univariate case)



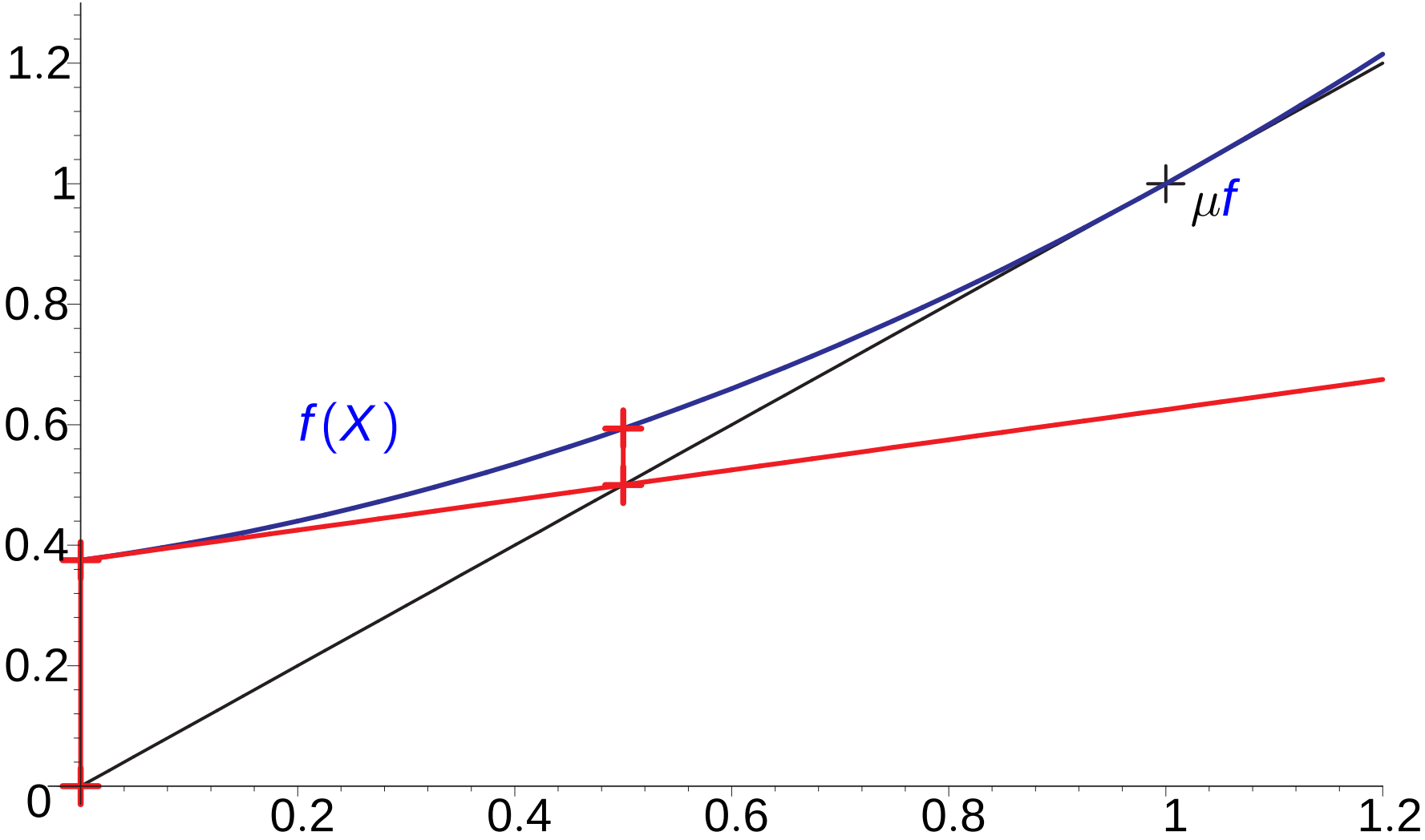
Newton's method for $X = f(X)$ (univariate case)



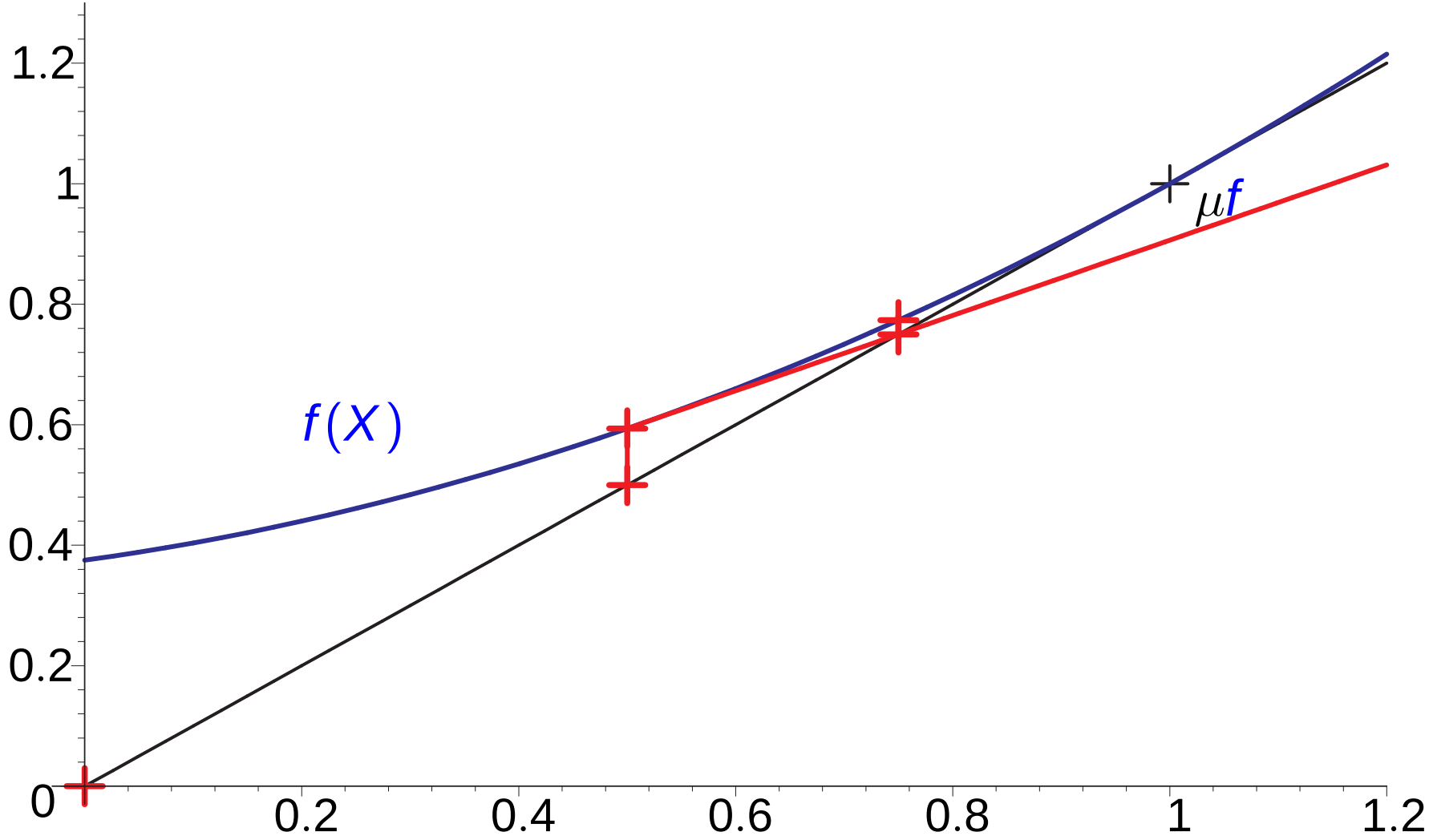
Newton's method for $X = f(X)$ (univariate case)



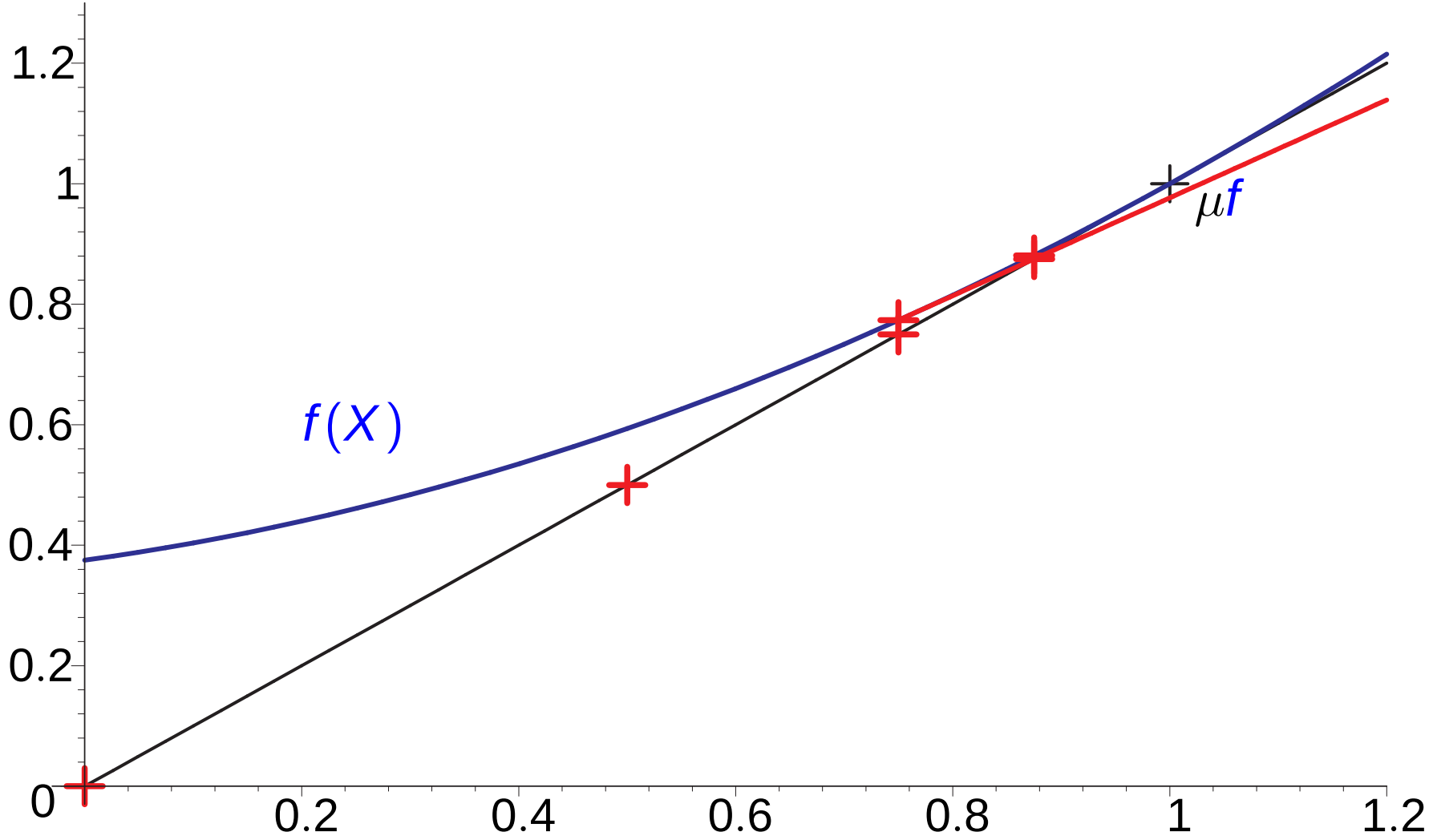
Newton's method for $X = f(X)$ (univariate case)



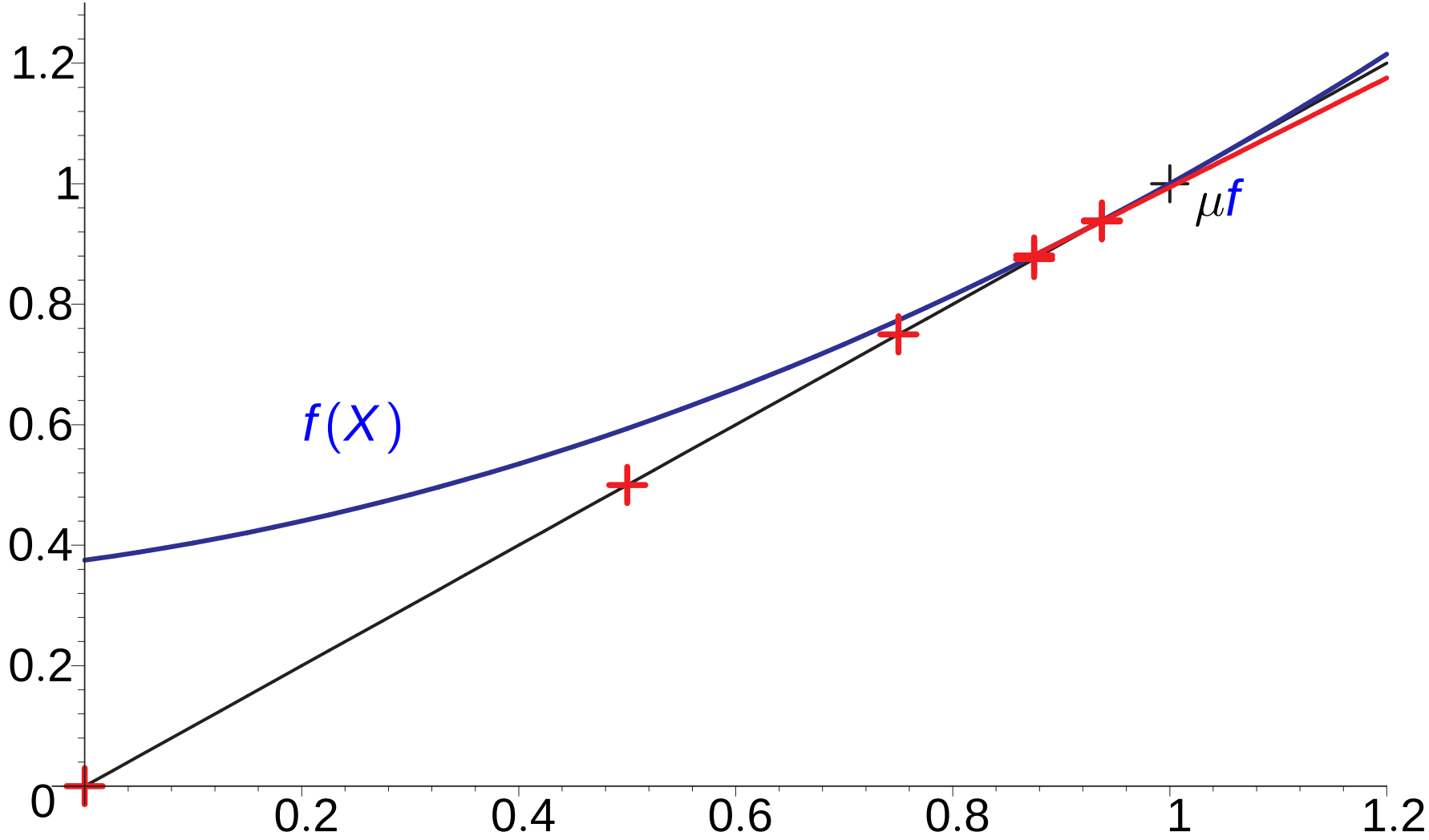
Newton's method for $X = f(X)$ (univariate case)



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



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



Newton's Method is very efficient . . .

- Often exponential convergence*.

. . . if it works.

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

* Called quadratic convergence in numerical mathematics.

A puzzling mismatch

- Kleene iteration is robust and applicable to every semiring, but converges slowly.
- Newton's method converges fast, but it is not robust and can only be applied to the reals.

A puzzling mismatch and our solution

- Kleene iteration is robust and applicable to every semiring, but converges slowly.
- Newton's method converges fast, but it is not robust and can only be applied to the reals.

Our results:

- Newton's method can be generalized to arbitrary semirings, and becomes as robust as Kleene iteration [EKL JACM 10].
- Newton's method converges globally and at least linearly in the real semiring [EKL SICOMP 10].

Generalizing Newton's method

Mathematical formulation of Newton's method

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

- Compute the linear function $T_\nu(X)$ for 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 $Df_\nu(X)$ is the differential of f at ν.

So: $\nu_0 = f(0)$

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

Two obstacles

Generalize $X = Df_\nu(X) + f(\nu) - \nu$

In an arbitrary semiring

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

are defined.

Overcoming the obstacles

For the differential: take the algebraic definition

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

For $f(\nu) - \nu$: generic (and nontrivial ...) construction of a δ such that $\nu + \delta = f(\nu)$.

The theorem

Theorem [EKL JACM 10]: Let $X = f(X)$ be an equation over an arbitrary semiring. The sequence

$$\begin{aligned} \nu_0 &= f(0) \\ \nu_{i+1} &= \nu_i + \Delta_i \end{aligned}$$

where Δ_i is the least solution of $X = Df_{\nu_i}(X) + \delta_i$ satisfies

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

for every $i \geq 0$.

(This theorem and all subsequent results extend to systems of equations)

Exploring Newton's method with language theory

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

Example: $X = 0.3X^2 + 0.7$ induces $X \to 0.3\,X\,X \mid 0.7$

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

$Y(t)$ = (ordered) product of the leaves of t

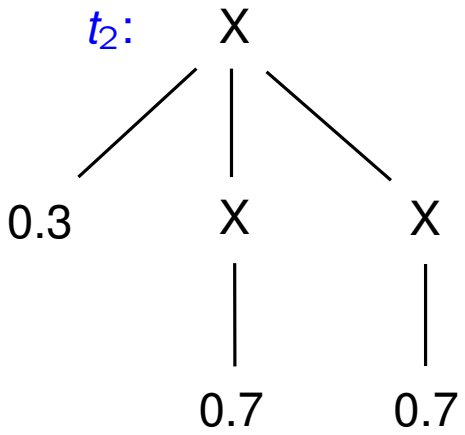
Assign to a set T of derivation trees its yield $Y(T)$

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

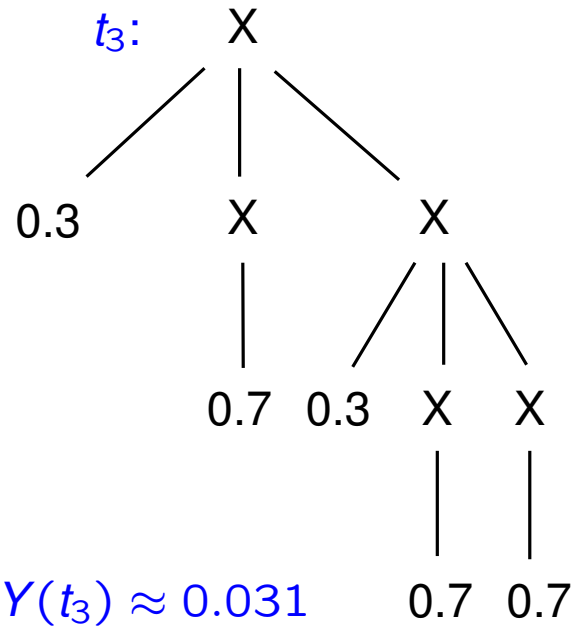
$$X \rightarrow 0.3\,X\,X \mid 0.7$$

t_1:

t_2:



t_3:



$Y(t_1) = 0.7$

$Y(t_2) = 0.147$

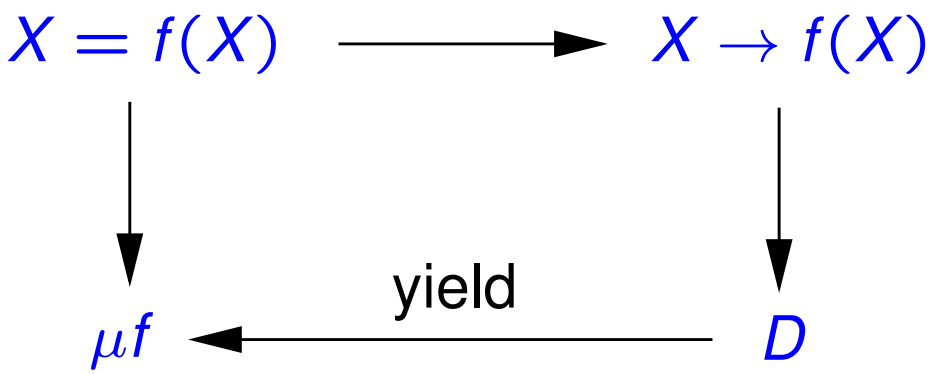
$Y(t_3) \approx 0.031$

$$Y(\{t_1, t_2, t_3\}) \approx 0.7 + 0.147 + 0.031 = 0.878$$

Derivation trees and μf

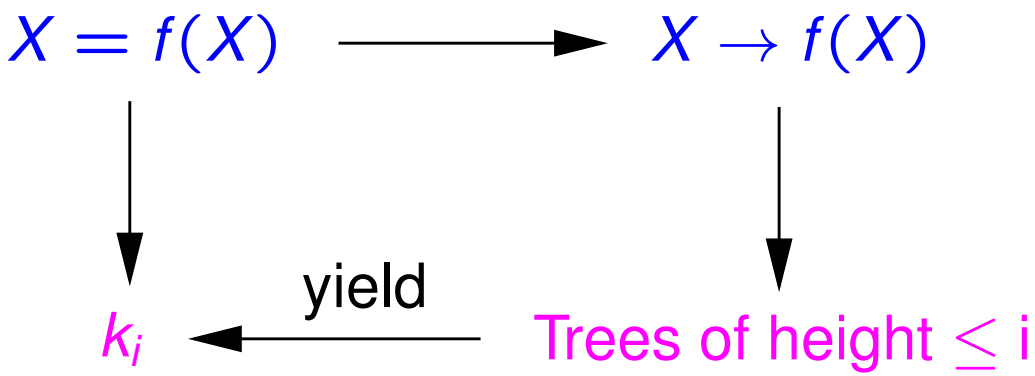
Proposition: Let D be the set of all derivation trees of G. Then

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



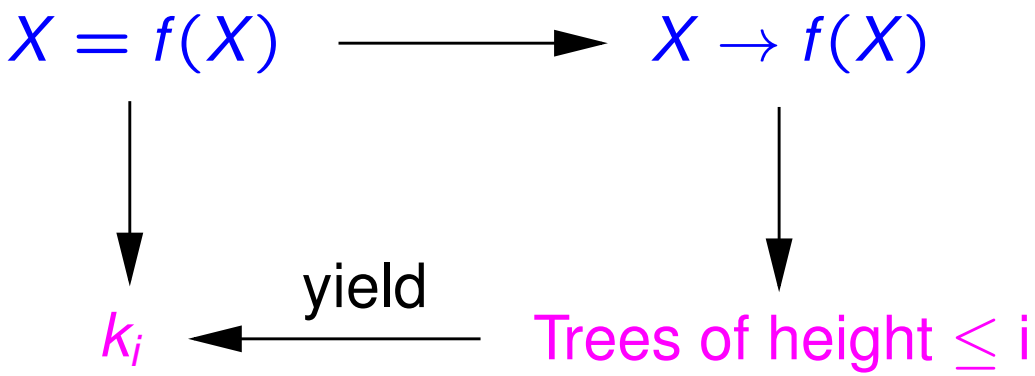
Approximants as yields: Kleene iteration

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



Approximants as yields: Kleene iteration

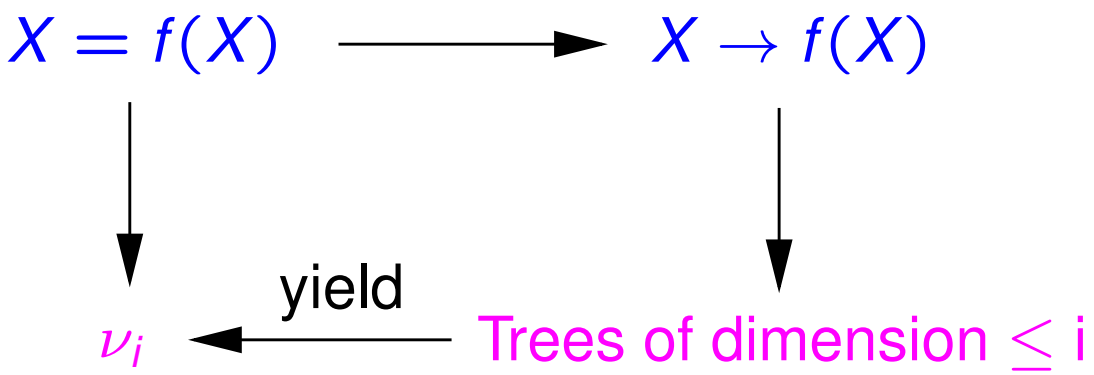
Proposition: The i-th Kleene approximant k_i is the yield of all derivation trees of height at most i.



I.e.: The Kleene approximants correspond to evaluating $Y(D)$ by increasing height.

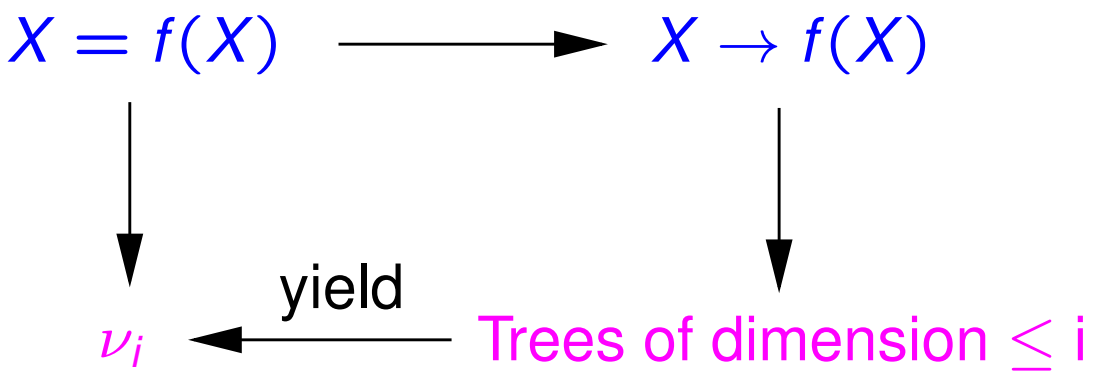
Approximants as yields: Newton iteration

Theorem [EKL JACM 10]: The i-th Newton approximant ν_i is the yield of all derivation trees of dimension at most i.



Approximants as yields: Newton iteration

Theorem [EKL JACM 10]: The i-th Newton approximant ν_i is the yield of all derivation trees of dimension at most i.

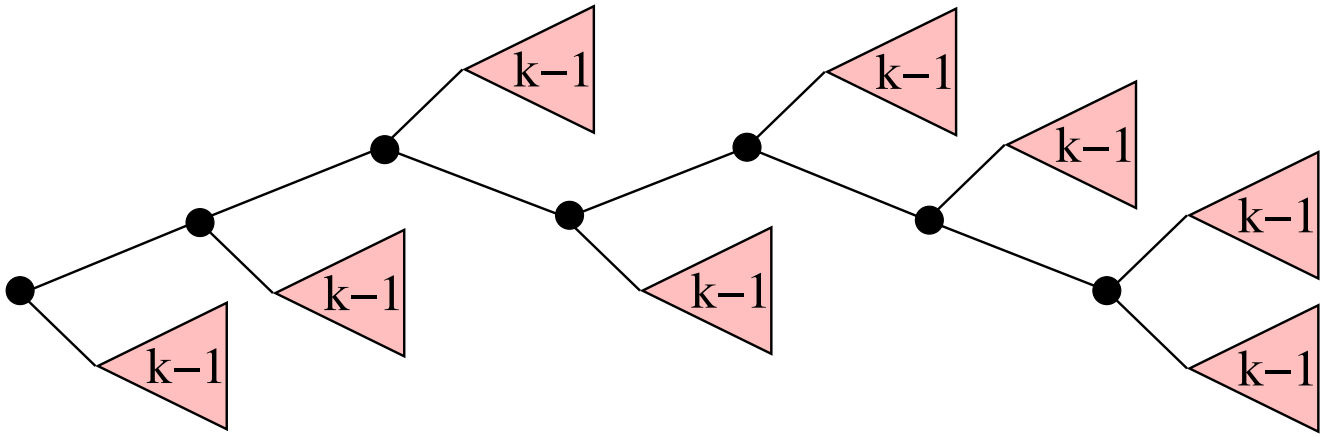


I.e.: The Newton approximants correspond to evaluating $Y(D)$ by increasing dimension.

Understanding dimension

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 $k - 1$).

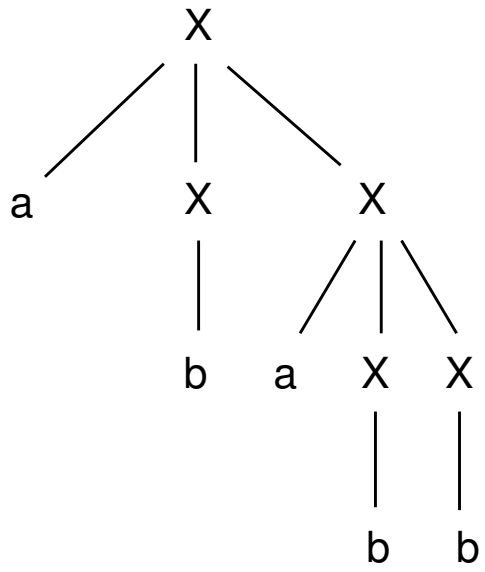


Understanding dimension

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 $w_1, \ldots, w_n$ contain at most k occurrences of non-terminals and some w_1 contains k occurrences.



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

Some applications

Stochastic thread creation

Threads can spawn new threads with known probabilities.

Execution by one processor. We assume termination with probability 1.

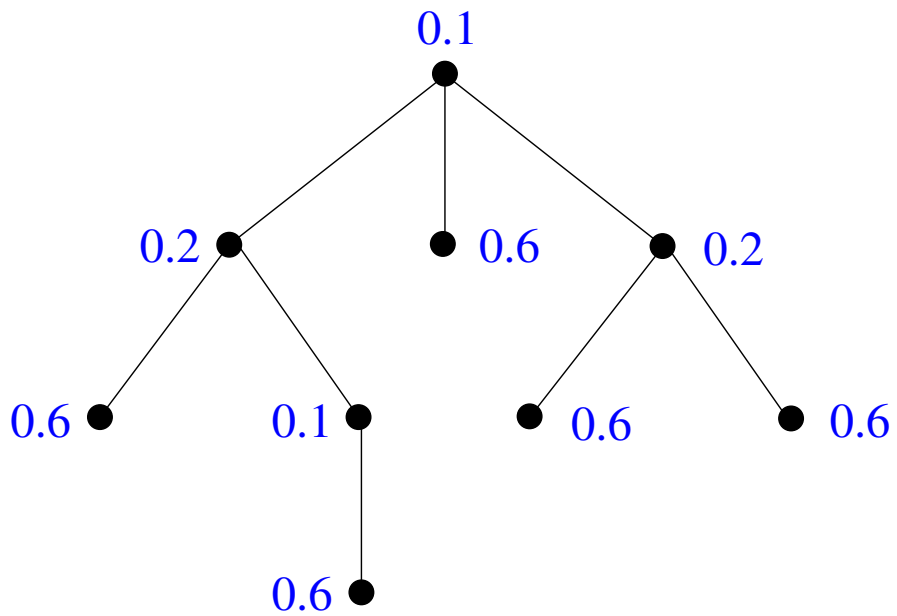
Example (only one type of thread):

$$X \xrightarrow{0.1} \langle X, X, X \rangle \quad X \xrightarrow{0.2} \langle X, X \rangle \quad X \xrightarrow{0.1} X \quad X \xrightarrow{0.6} \epsilon$$

Probability generating function

$$f(X) = 0.1X^3 + 0.2X^2 + 0.1X + 0.6$$

Describing executions: family trees



Probability of a family tree: product of the probabilities of its nodes.

Execution order depends on a scheduler that chooses a thread from the pool of inactive threads and executes it for one time unit.

Completion space S^σ for a scheduler σ: maximal size reached by the pool during execution.

Completion space of the optimal scheduler

Lemma: The family trees with completion space $S^{op} = k$ "are" the derivation trees of dimension k.

Theorem [BEKL I&C '11]: The probability $\Pr[S^{op} \leq k]$ of completing execution within space at most k is equal to the k-th Newton approximant of $X = f(X)$.

In our example:

$\Pr[S^{op} = 1]$	$= 2$	$= 3$	$= 4$	$= 5$
0.667	0.237	0.081	0.014	0.001

Commutative idempotent semirings

Theorem [HK LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an idempotent and commutative semiring is reached by

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

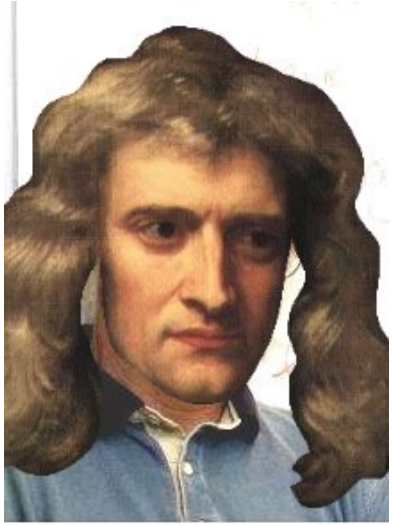
after at most $O(3^n)$ iterations.

Commutative idempotent semirings

Theorem [HK LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an idempotent and commutative semiring is reached by

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

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

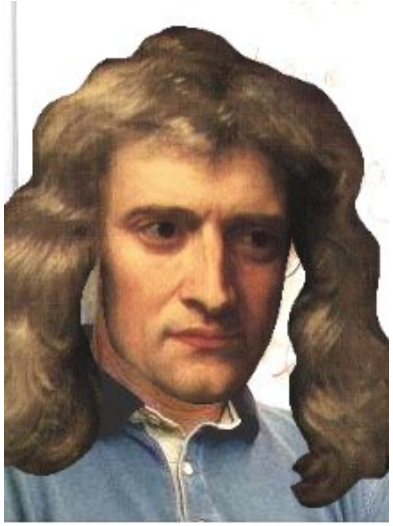


Commutative idempotent semirings

Theorem [HK LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an idempotent and commutative semiring is reached by

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

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



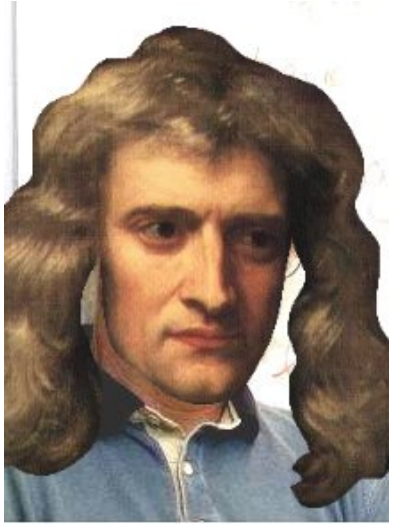
Theorem [EKL JACM '10]: This is exactly Newton's sequence.

Commutative idempotent semirings

Theorem [HK LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an idempotent and commutative semiring is reached by

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

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



Theorem [EKL JACM '10]: This is exactly Newton's sequence.

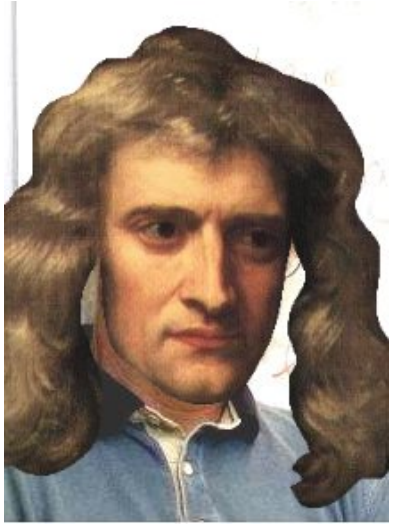
Lemma: Every derivation tree has the same yield as some other derivation tree of dimension at most n.

Commutative idempotent semirings

Theorem [HK LICS '99]: The least fixed point of a system $X = f(X)$ of n equations over an idempotent and commutative semiring is reached by

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

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



Theorem [EKL JACM '10]: This is exactly Newton's sequence.

Lemma: Every derivation tree has the same yield as some other derivation tree of dimension at most n.

Corollary: The fixed point is reached after at most n iterations.

Other applications: three new algorithms

$O(n^3)$ algorithm for computing the throughput of context-free grammars (improving $O(n^4)$ algorithm by Caucal et al.) [EKL TCS '11].

New algorithm for pattern-based verification of multithreaded procedural programs with fixed number of threads [GMM CAV '10, EG POPL '11].

Very simple algorithm for transforming a context-free grammar into a Parikh-equivalent NFA [EGKL IPL '11].

Newton's method on the real semiring: proving robustness

Newton's method on the real semiring

On the real field Newton's method may not converge, or converge only locally.

On the real semiring these problems disappear: Newton's method always converges and always exhibits linear or exponential convergence order [EKL SICOMP '10].

For an important subclass: efficient algorithms returning an arbitrarily small interval $[a, b]$ such that (provably) $a \leq \mu f \leq b$ [EGK STACS '10].

Nuclear Chain Reaction

Computing the probability p_{cr} of a chain reaction for radius D.
D measured in mean path length, ball discretized in 100 layers.
Critical radius (size from which $p_{cr} > 0$) in [2.981,2991].

Runtime for computing 10^{-4}-interval for p_{cr}:

D	2	3	6	10
safe?	yes	no	no	no
[EGK STACS '10] (sec.)	4	34	28	23
exact LP (sec.)	258	124	168	222

Conclusions and future work

New connections between analysis and numerical mathematics and TCS, leading to many new algorithms!

Conclusions and future work

New connections between analysis and numerical mathematics and TCS, leading to many new algorithms!

Numerical mathematics $\rightarrow$ TCS

- Connection between dimension and pathwidth
- Algorithmic language theory
- Applications to verification
- Can Newton's method separate unit-cost RAMs and Turing Machines? [Tiw JC '92]

Conclusions and future work

New connections between analysis and numerical mathematics and TCS, leading to many new algorithms!

Numerical mathematics $\rightarrow$ TCS

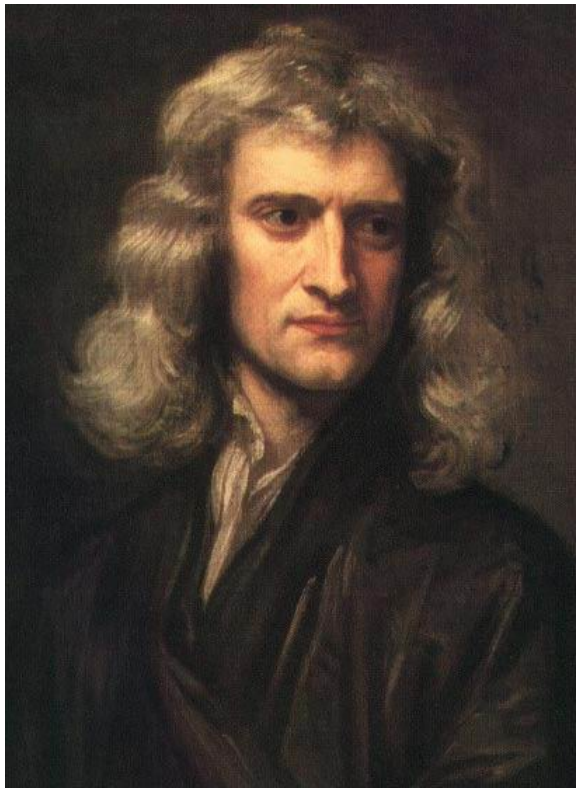
- Connection between dimension and pathwidth
- Algorithmic language theory
- Applications to verification
- Can Newton's method separate unit-cost RAMs and Turing Machines? [Tiw JC '92]

TCS $\rightarrow$ Numerical mathematics

- Generic convergence results (see recent breakthrough by M. Luttenberger)
- Applications to linear programming

. . . and the overall conclusion

Newton did it all!



Thermal equilibrium (2d)

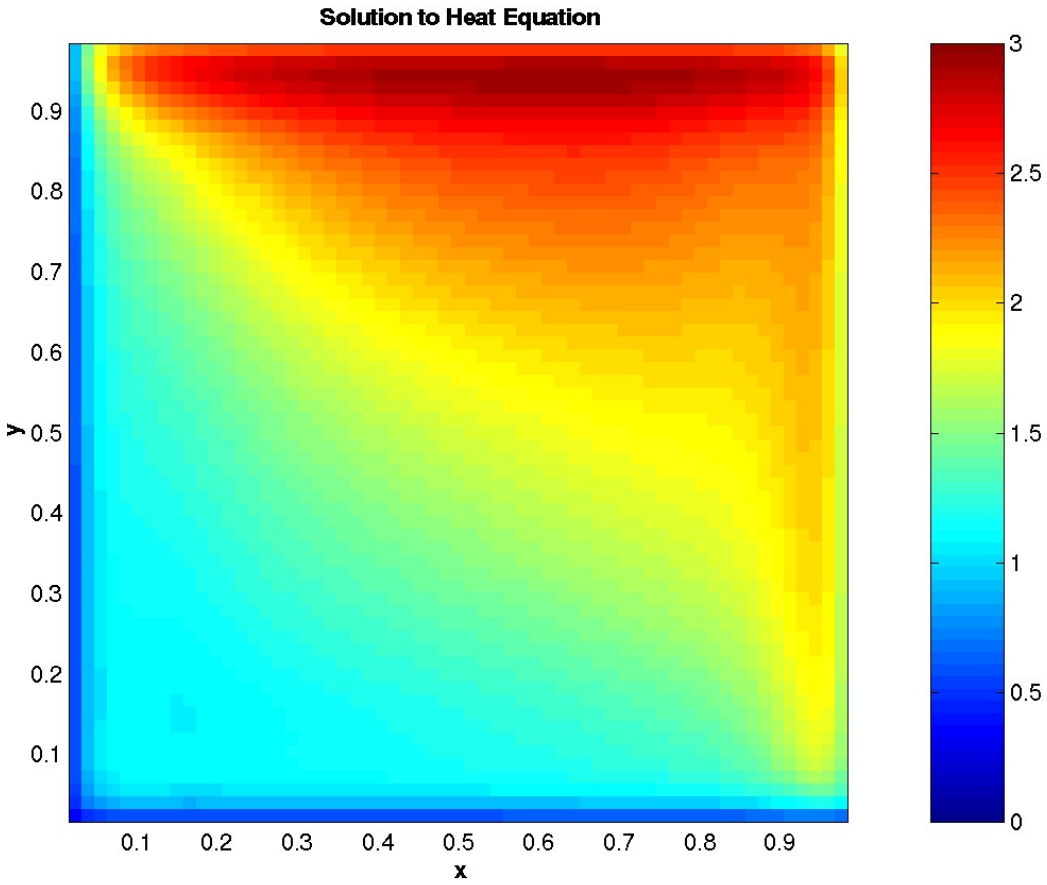
Heat equation in 2d

$$\frac{\partial u}{\partial t} = h^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right)$$

After discretization, temperature at thermal equilibrium is a solution of

$$u_{i,j} = k_{i,j}(u_{i-1,j} + u_{i+1,j} + u_{i,j+1} + u_{i,j-1})$$

for constants $k_{i,j}$ plus boundary conditions.



Abstract Interpretation: Collecting semantics

Collecting semantics of a program: assigns to each program point p the possible values of the memory when the program reaches p.

Solution of the equations

$$p_i\ Store = \bigsqcup_{p_j \in pred(p_i)} f_{ij}(p_j\ Store)$$

Basis of abstract interpretation

Idempotent semirings: derivation tree analysis

Idempotent semiring: $a + a = a$

Technique for computing μf algebraically:

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

(2) Show that for every $t \in D$ there is $t' \in T$ such that $Y(t) \sqsubseteq Y(t')$.

Then by idempotence we have $\mu f = Y(D) = Y(T)$