Set 9: Nonlinear Equations Part 1

Kyle A. Gallivan

Department of Mathematics

Florida State University

Foundations of Computational Math 1
Fall 2012

Overview

Iteration for nonlinear equations will be discussed:

- solving scalar nonlinear equations $f(x) = 0$ with $f : \mathbb{R} \to \mathbb{R}$
  - basic ideas and examples of methods
  - convergence and design of contraction mappings
- solving systems of nonlinear equations $F(x) = 0$ with $F : \mathbb{R}^n \to \mathbb{R}^n$
  - basic ideas and examples of methods
  - convergence and design of contraction mappings

The Problem

Let $f(x) : \mathbb{R} \to \mathbb{R}$ be a given function.

The problem is to find $x^* \in \mathbb{R}$ such that

$$f(x^*) = 0$$

The function $f(x)$ may be nonlinear and therefore may have many solutions or none.

Additional Reference

The book

> Analysis of Numerical Methods, E. Isaacson and H. Keller, Wiley.

is highly recommended. It contains rigorous proofs of most of the material presented in the textbook and notes.

Also the book

> Iterative Solution of Nonlinear Equations in Several Variables, J. M. Ortega and W. C. Rheinboldt, Academic Press, 1970.

is highly recommended. It is a classic text and contains rigorous proofs, in a very general setting, of most of the material presented in the textbook and notes.

Additional Reference

For nonlinear systems and related optimization algorithms the book

> Numerical Optimization, J. Nocedal and S. J. Wright, Springer, 2006, 2nd edition

is highly recommended.

Bisection Method

Theorem 9.1. *Let $f(x) : \mathbb{R} \to \mathbb{R}$ be continuous on the interval $[a, b]$. If $f(a)f(b) < 0$ then $\exists \alpha \in [a, b]$ such that $f(\alpha) = 0$*

- There may in fact be more than one root in the interval $[a, b]$. The theorem guarantees there is at least one.
- Given the interval $[a, b]$ the uncertainty associated with the root α is $|b - a|$.
- The bisection method, halves the uncertainty on each step.

Bisection Method

Bisection:

Given $\mathcal{I}_0 = [a^{(0)}, b^{(0)}] = [a, b]$, set $x^{(0)} = (a^{(0)} + b^{(0)})/2$ and $k = 0$

loop over k until uncertainty small enough

if $f(x^{(k)})f(a^{(k)}) < 0$ then

$a^{(k+1)} = a^{(k)}$ and $b^{(k+1)} = x^{(k)}$

$x^{(k+1)} = (a^{(k+1)} + b^{(k+1)})/2$

else if $f(x^{(k)})f(b^{(k)}) < 0$ then

$a^{(k+1)} = x^{(k)}$ and $b^{(k+1)} = b^{(k)}$

$x^{(k+1)} = (a^{(k+1)} + b^{(k+1)})/2$

else

stop since $f(x^{(k)}) = 0$

end if

end loop

Bisection Method

After k iterations we have

$$|\mathcal{I}_k| = \frac{|\mathcal{I}_0|}{2^k} = \frac{b-a}{2^k}$$

$$|e^{(k)}| = |x^{(k)} - \alpha| \leq \frac{b-a}{2^{k+1}}$$

$$\lim_{k\to\infty} |e^{(k)}| = 0$$

Given $b - a$ we can therefore determine how many iterations are required to have a particular error bound:

$$|e^{(k)}| \leq \frac{b-a}{2^{k+1}} = \epsilon$$

$$k \geq \log_2\left(\frac{b-a}{\epsilon}\right) - 1$$

Bisection Method

- To gain a decimal digit approximately 2.32 steps are required.
- Convergence guaranteed but slow.
- Convergence is not necessarily monotone.
- Therefore, the method does not have an "order" of convergence.
- Multisectioning can also be used.

Bisection Example

$f(x) = x^3 - 3x + 1$ and $\mathcal{I}_0 = [1, 2]$

k	x_k	$f(x_k)$
1	1.0000000000000	-1.0000000000000
2	2.0000000000000	3.0000000000000
3	1.5000000000000	-0.1250000000000
4	1.7500000000000	1.1093750000000
5	1.6250000000000	0.4160156250000
6	1.5625000000000	0.1271972656250
7	1.5312500000000	-0.0033874511719
8	1.5468750000000	0.0607719421387
9	1.5390625000000	0.0284104347229
10	1.5351562500000	0.0124412178993

Bisection Example

$f(x) = x^3 - 3x + 1$ and $\mathcal{I}_0 = [1, 2]$

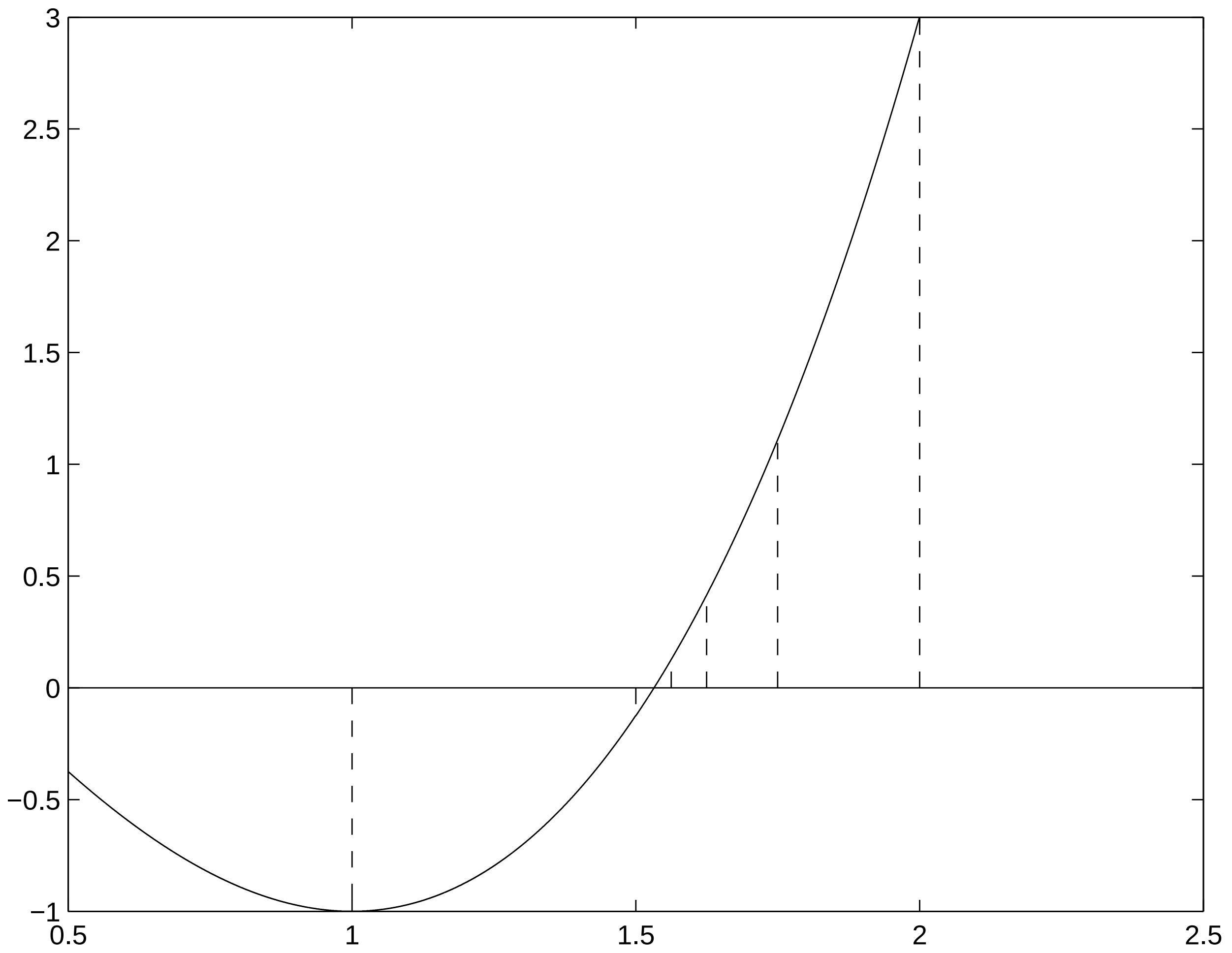
k	x_k	$f(x_k)$
11	1.5332031250000	0.0045093372464
12	1.5322265625000	0.0005565593019
13	1.5317382812500	-0.0014165415196
14	1.5319824218750	-0.0004302650486
15	1.5321044921875	0.0000630786362
16	1.5320434570312	-0.0001836103281
17	1.5320739746094	-0.0000602701265
18	1.5320892333984	0.0000014031847
19	1.5320816040039	-0.0000294337384
20	1.5320854187012	-0.0000140153437

Bisection Example

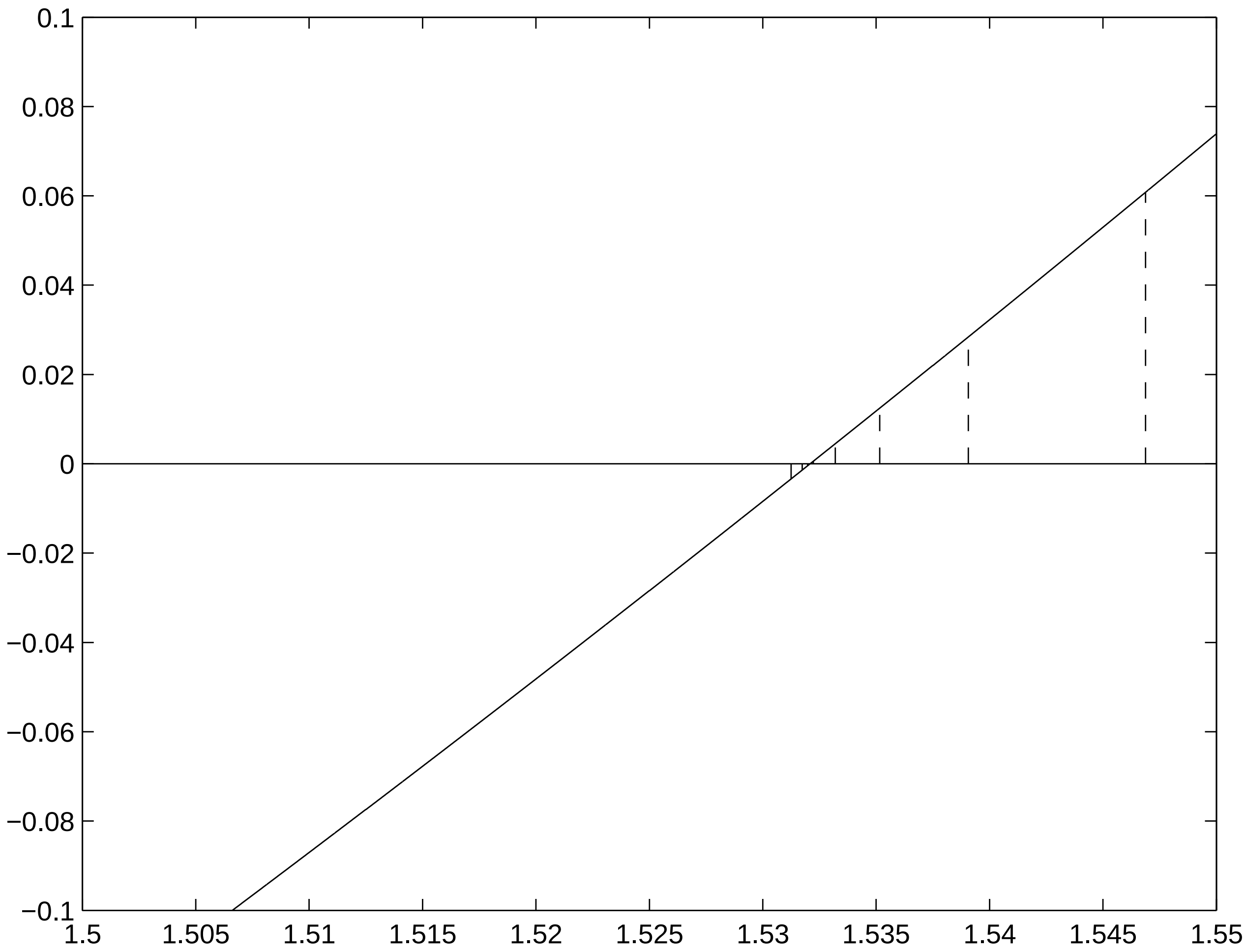
$f(x) = x^3 - 3x + 1$ and $\mathcal{I}_0 = [1, 2]$

k	x_k	$f(x_k)$
21	1.5320873260498	-0.0000063060962
22	1.5320882797241	-0.0000024514599
23	1.5320887565613	-0.0000005241387
24	1.5320889949799	0.0000004395228
25	1.5320888757706	-0.0000000423080
26	1.5320889353752	0.0000001986074
27	1.5320889055729	0.0000000781497

Iterations 1 to 6 for bisection



Iterations 7 to 15 for bisection



Order of a Method

Definition 9.1. A sequence $\{x^{(k)}\}$ is said to converge to α with order p if

$$\exists C > 0 \quad : \quad \frac{|x^{(k+1)} - \alpha|}{|x^{(k)} - \alpha|^p} \leq C, \;\; \forall k \geq k_0 \geq 0$$

Definition 9.2. A method that converges to a root, $\alpha \in \mathcal{I}$, of a function $f(x)$ for any $x^{(0)} \in \mathcal{I}$ is said to be globally convergent on the interval $\mathcal{I}$.

Taylor Series Idea

Approximate $f(x)$ around the root α

$$f(\alpha) = 0 = f(x) + (\alpha - x)f'(\xi)$$

ξ between α and x.

- Approximate $q_k \approx f'(\xi)$
- Set $x^{(k+1)}$ to root of $f(x^{(k)}) + (x - x^{(k)})q_k$

Secant Method

Approximate $f(x)$ around the root α

$$f(\alpha) = 0 = f(x) + (\alpha - x) f'(\xi)$$

ξ between α and x.

- Take $q_k = \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$
- Set $x^{(k+1)}$ to root of $f(x^{(k)}) + (x - x^{(k)}) q_k$

Secant Method

Secant:

Given $x^{(-1)}$ and $x^{(0)}$, set $k = 0$

loop over k until $|f(x^{(k)})|$ small enough

$$q_k = \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$$

$$x^{(k+1)} = x^{(k)} - q_k^{-1} f(x^{(k)})$$

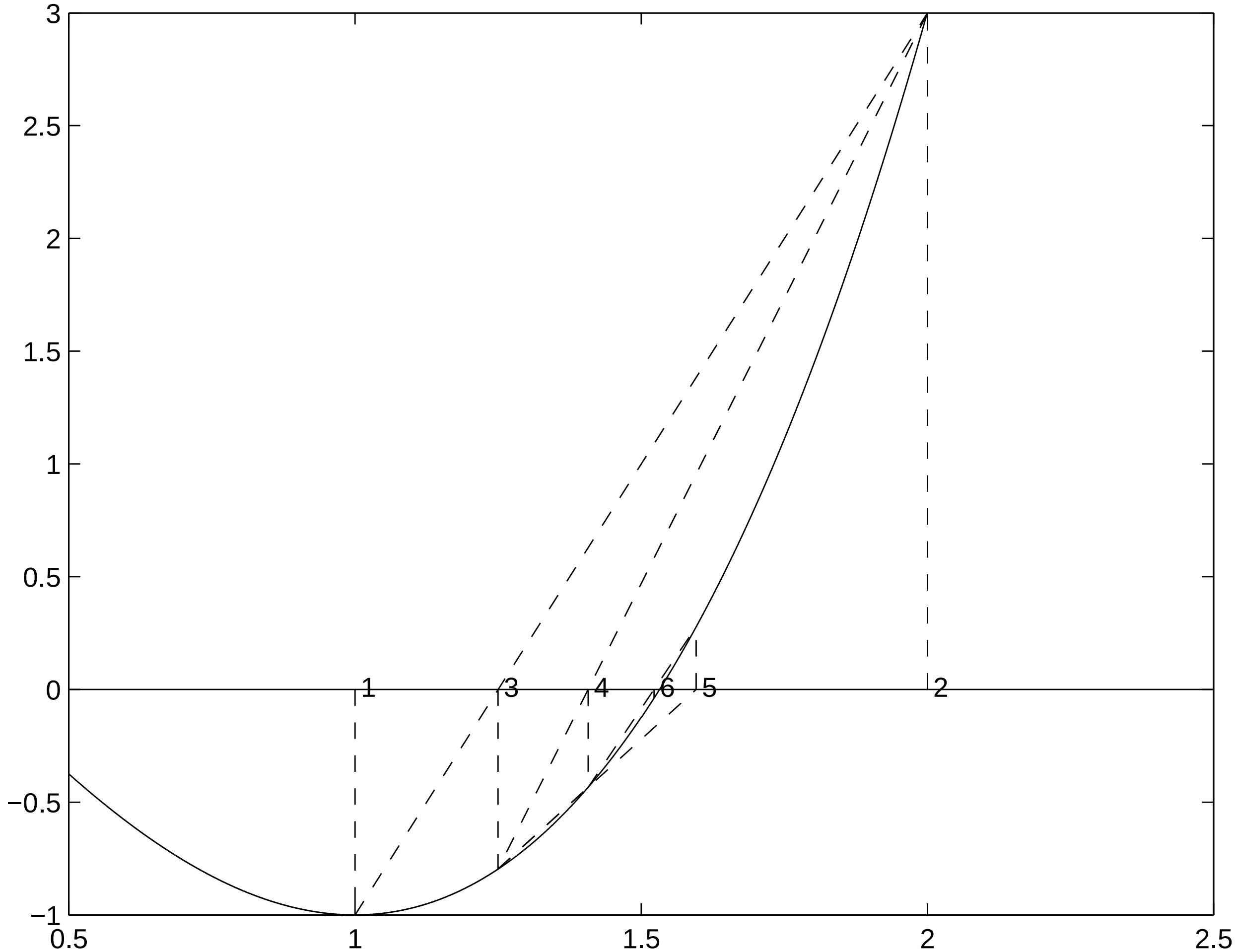
end loop

Secant Example

$f(x) = x^3 - 3x + 1$,$x^{(-1)} = 1$ and $x^{(0)} = 2$

k	x_k	$f(x_k)$
1	1.0000000000000	-1.0000000000000
2	2.0000000000000	3.0000000000000
3	1.2500000000000	-0.7968750000000
4	1.4074074074074	-0.4344358075497
5	1.5960829578881	0.2777418302669
6	1.5225014665094	-0.0383296857597
7	1.5314246225018	-0.0026828525395
8	1.5320961972127	0.0000295503946
9	1.5320888807121	-0.0000000223349

Iterations 1 to 6 for secant



Secant Example

- Converges faster than bisection on this example.
- More computations per step than bisection.
- Note the location of $x^{(5)}$ relative to $x^{(3)}$ and $x^{(4)}$
- Secant iterations do not stay within interval defined by previous two iterates.
- Secant is a locally convergent method. $x^{(0)}$ must be sufficiently close to α to guarantee convergence to α with order $p \approx 1.63$.
- This does not mean that it diverges if not close to α. It may converge to some other root.

Regula Falsi Method

- ideas from the bisection method and the secant method can be combined to give a series of iterates that are contained within the original interval.
- Two endpoints are maintained with $f(b^{(k)})f(a^{(k)}) < 0$
- Secant estimate of q_k used to get $x^{(k)}$ rather than bisection.
- The method is globally convergent on the interval $[a, b]$ containing α
- It has order $p = 1$ therefore faster than bisection but slower than the secant method.

Regula Falsi Method

Regula Falsi:

Given $\mathcal{I}_0 = [a^{(0)}, b^{(0)}] = [a, b]$, $k = 0$, $f(a)f(b) < 0$

$q_0 = \frac{f(b^{(0)}) - f(a^{(0)})}{b^{(0)} - a^{(0)}}$ and $x^{(0)} = a^{(0)} - f(a^{(0)})/q_0$

loop over k until $|f(x^{(k)})|$ small enough

if $f(x^{(k)})f(a^{(k)}) < 0$ then

$a^{(k+1)} = a^{(k)}$ and $b^{(k+1)} = x^{(k)}$

else if $f(x^{(k)})f(b^{(k)}) < 0$ then

$a^{(k+1)} = x^{(k)}$ and $b^{(k+1)} = b^{(k)}$

else stop since $f(x^{(k)}) = 0$

end if

$q_{k+1} = \frac{f(b^{(k+1)}) - f(a^{(k+1)})}{b^{(k+1)} - a^{(k+1)}}$

$x^{(k+1)} = a^{(k+1)} - f(a^{(k+1)})/q_{k+1}$

end loop

Regula Falsi Example

$f(x) = x^3 - 3x + 1$,$a^{(0)} = 1$ and $b^{(0)} = 2$

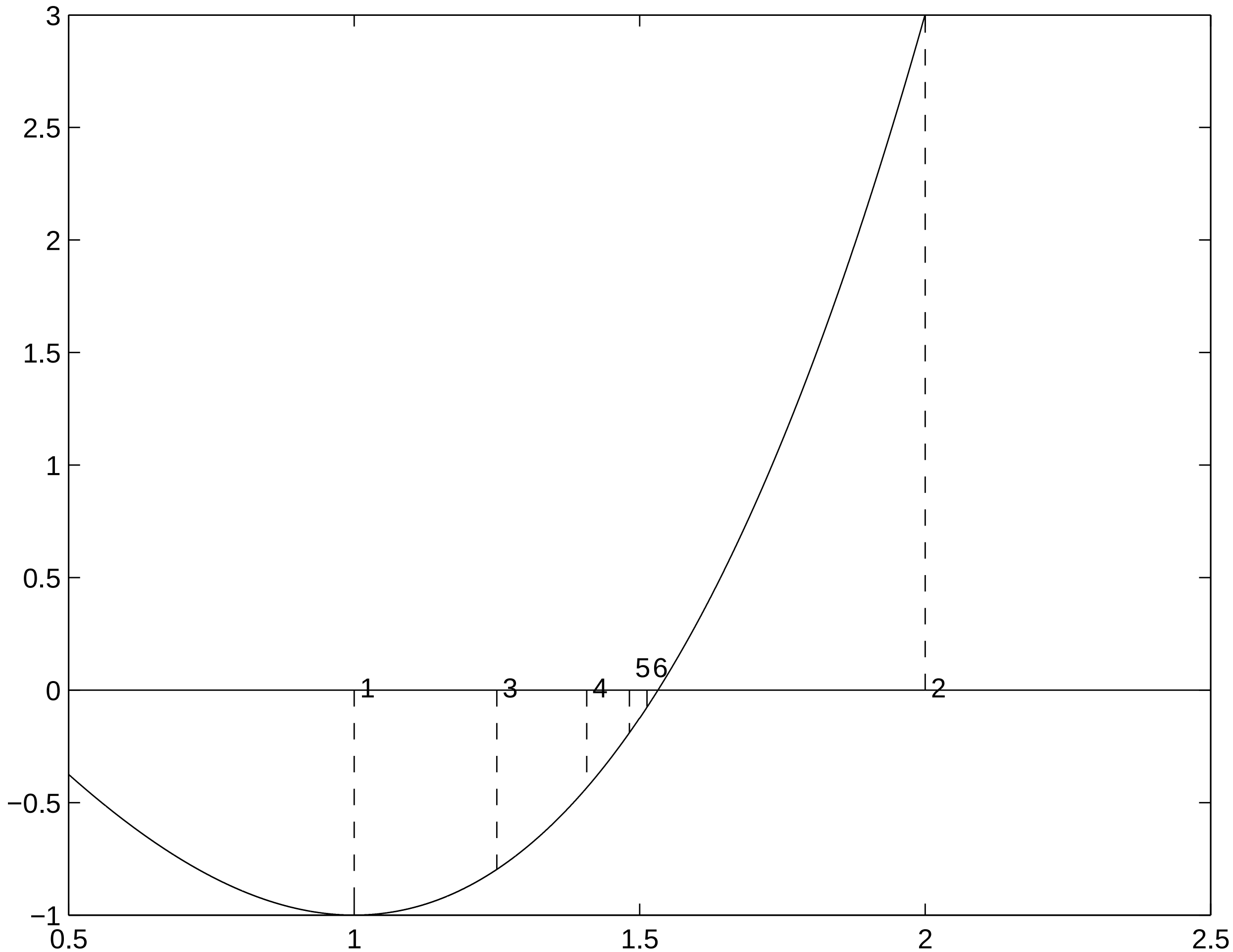
k	x_k	$f(x_k)$
1	1.0000000000000	-1.0000000000000
2	2.0000000000000	3.0000000000000
3	1.2500000000000	-0.7968750000000
4	1.4074074074074	-0.4344358075497
5	1.4823668639053	-0.1897305692834
6	1.5131565583507	-0.0748817012259
7	1.5250125153219	-0.0283721024127
8	1.5294625607933	-0.0105836310724
9	1.5311167233320	-0.0039250316037
10	1.5317293823232	-0.0014524809532

Regula Falsi Example

$f(x) = x^3 - 3x + 1$, $a^{(0)} = 1$ and $b^{(0)} = 2$

k	x_k	$f(x_k)$
11	1.5319559906595	-0.0005370680123
12	1.5320397661503	-0.0001985268554
13	1.5320707316591	-0.0000733772791
14	1.5320821765046	-0.0000271197908
15	1.5320864064109	-0.0000100231576
16	1.5320879697296	-0.0000037044214
17	1.5320885475100	-0.0000013691004
18	1.5320887610491	-0.0000005059993
19	1.5320888399700	-0.0000001870098
20	1.5320888691380	-0.0000000691160

Iterations 1 to 6 for Regula Falsi



Comparison

$$f(x) = x^3 - 3x + 1$$

$$a^{(0)} = 1, \quad b^{(0)} = 2, \quad \alpha \approx 1.532088$$

- Bisection: $k = 23$, $x^{(23)} = 1.53208875$ and $f(x^{(23)}) = -0.5241387 \times 10^{-6}$
- Regula Falsi: $k = 18$, $x^{(18)} = 1.53208876$ and $f(x^{(18)}) = -0.505993 \times 10^{-6}$
- Secant $k = 9$, $x^{(9)} = 1.53208888$ and $f(x^{(9)}) = -0.0223349 \times 10^{-7}$
- The secant shows the significant drop between $f(x^{(8)})$ and $f(x^{(9)})$ due to superlinear convergence.

Newton's Method

Approximate $f(x)$ around the root α

$$f(\alpha) = 0 = f(x) + (\alpha - x) f'(\xi)$$

ξ between α and x.

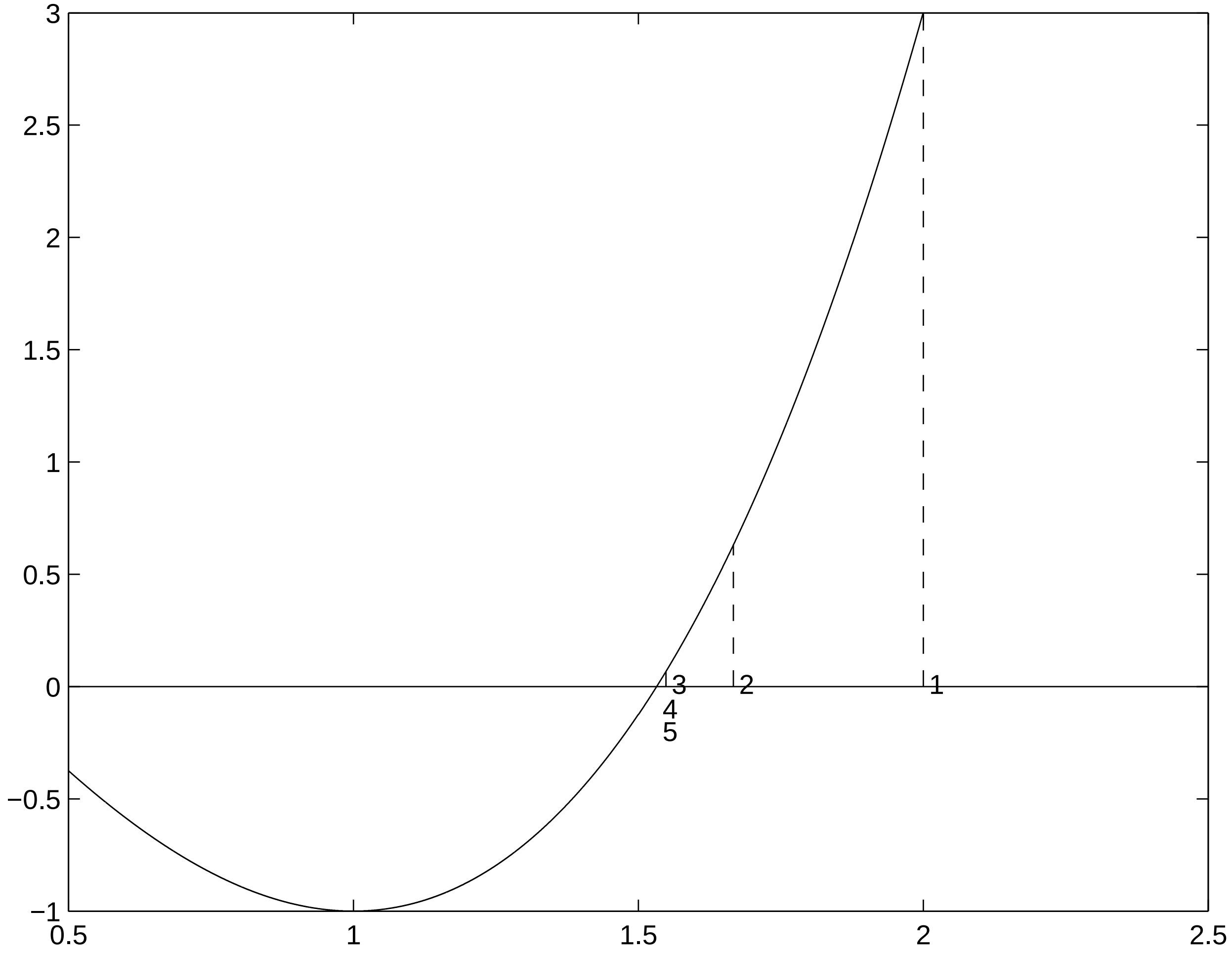
- Take $q_k = f'(x^{(k)})$
- Set $x^{(k+1)}$ to root of $f(x^{(k)}) + (x - x^{(k)}) q_k$

Newton Example

$f(x) = x^3 - 3x + 1$, $x^{(0)} = 2$

k	x_k	$f(x_k)$
1	2.0000000000000	3.0000000000000
2	1.6666666666667	0.6296296296296
3	1.5486111111111	0.0680402172282
4	1.5323901618654	0.0012181398810
5	1.5320889893972	0.0000004169584

Iterations 1 to 6 for Newton



Newton's Method

- Newton is a locally convergent method with order $p = 2$.
- Note the rapid convergence near the root in the example.
- It requires knowledge of the derivative.
- It requires evaluation of the derivative on each step.
- Usually worth it to get faster convergence unless the cost of evaluating $f'(x^{(k)})$ is significantly more than evaluating $f(x^{(k)})$

Fixed Point Methods

Solve

$$f(x) : \mathbb{R} \to \mathbb{R}$$

to find a real scalar x^* such that $f(x^*) = 0$.

Choose $\phi(x) : [a, b] \to \mathbb{R}$ such that $\phi(\alpha) = \alpha$ when $f(\alpha) = 0$.

Given $x^{(0)}$, iterate until convergence

$$x^{(k+1)} = \phi(x^{(k)})$$

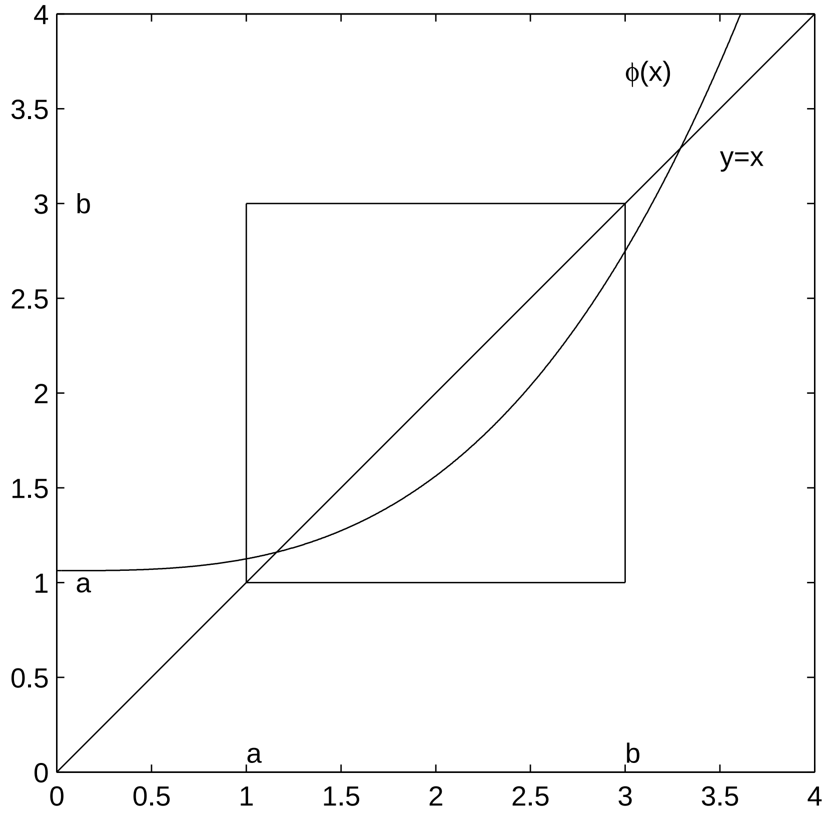
Fixed Point Methods

- $\phi(x) = x + F(f(x))$ where $F(y)$ is any continuous function with $F(0) = 0$
- $\phi(x) = x + f(x)$ simplest choice (Picard iteration)
- $f(x) = b - Ax$ is linear system simple form, i.e., simple Richardson's
- Given choice need to determine
  - Sufficient condition for convergence.
  - Rate of convergence.
  - Computational cost.

Geometric View

- $\phi(x) : [a, b] \to [a, b] \to$ remains in “box”.
- intersection points $y = \phi(x)$ with $y = x$
- Convergence?

Geometric View



Convergence

Definition 9.3. A continuous function $\phi(x) : [a, b] \to \mathbb{R}$ is a contraction mapping on $[a, b]$ if $\exists 0 < L < 1$ such that $\forall x,\ y \in [a, b]$

$$|\phi(x) - \phi(y)| \leq L|x - y|$$

i.e., $\phi(x)$ is Lipschitz continuous with constant strictly less than 1.

Definition 9.4. A real sequence $\{x^{(k)}\}$ is a Cauchy sequence if $\forall \epsilon > 0$ and integer $n > 0$, $\exists M > 0$ such that

$$\forall m > M \quad |x^{(m)} - x^{(m+n)}| < \epsilon$$

Convergence

Lemma 9.2. *If the real sequence* $\{x^{(k)}\}$ *is a Cauchy sequence then* $\lim_{k\to\infty} x^{(k)} = p$ *for some* $|p| < \infty$*, i.e., the sequence converges.*

Theorem 9.3. *If the continuous function* $\phi(x) : [a, b] \to \mathbb{R}$ *is a contraction mapping on* $[a, b]$ *then*

- $\exists p \in [a, b]$ *such that* $p = \phi(p)$ *and* p *is unique.*
- $\forall x^{(0)} \in [a, b], \;\; p = \lim_{k\to\infty} x^{(k)} = \lim_{k\to\infty} \phi(x^{(k-1)})$

Convergence

Proof. It can be shown that $x^{(k)} = \phi(x^{(k-1)})$ is a Cauchy sequence when $\phi(x)$ is a contraction mapping. Therefore, $\exists p \in [a, b]$ such that $p = \lim_{k\to\infty} x^{(k)}$.

Since $\phi(x)$ is continuous we have

$$p = \lim_{k\to\infty} x^{(k)} = \lim_{k\to\infty} \phi(x^{(k-1)}) = \phi(\lim_{k\to\infty} x^{(k-1)}) = \phi(p)$$

So p is a fixed point.

Suppose $p = \phi(p)$ and $q = \phi(q)$ with $q \neq p$. We have

$$|p - q| = |\phi(p) - \phi(q)| \leq L|p - q| < |p - q|$$

contradiction. $\square$

Convergence

Corollary 9.4. *If* $p = \phi(p)$ *then* $\forall x^{(0)} \in [a, b], \;\; x^{(k)} = \phi(x^{(k-1)})$ *converges to* p *on any interval* $[a, b]$ *containing* p *over which* $\phi(x)$ *is a contraction mapping.*

Corollary 9.5. *If the interval* $[a, b]$ *contains more than one fixed point of* $\phi(x)$ *then* $\phi(x)$ *cannot be a contraction mapping on the entire interval.*

Convergence

Lemma 9.6. *If* $|\phi'(x)| < 1$ *on* $[a, b]$ *then* $\phi(x)$ *is a contraction mapping on* $[a, b]$.

Proof. By the mean value theorem,

$$\phi(x) - \phi(y) = (x - y)\phi'(\xi)$$

for some ξ between x and y.

$$|\phi(x) - \phi(y)| = |(x - y)\phi'(\xi)| = |(x - y)||\phi'(\xi)| < |(x - y)|$$

$\square$

Convergence

Theorem 9.7. *Suppose p is a fixed point of the continuous function $\phi(x)$. If $\phi(x)$ is differentiable in a neighborhood of p and $|\phi'(p)| < 1$ then there exits $\delta > 0$ such that for any $x^{(0)}$ with $|x^{(0)} - p| < \delta$ the sequence $x^{(k)} = \phi(x^{(k-1)})$ converges to p.*

Note. Note that if $|\phi'(p)| > 1$ then

$$|x^{(k+1)} - p| > |x^{(k)} - p|$$

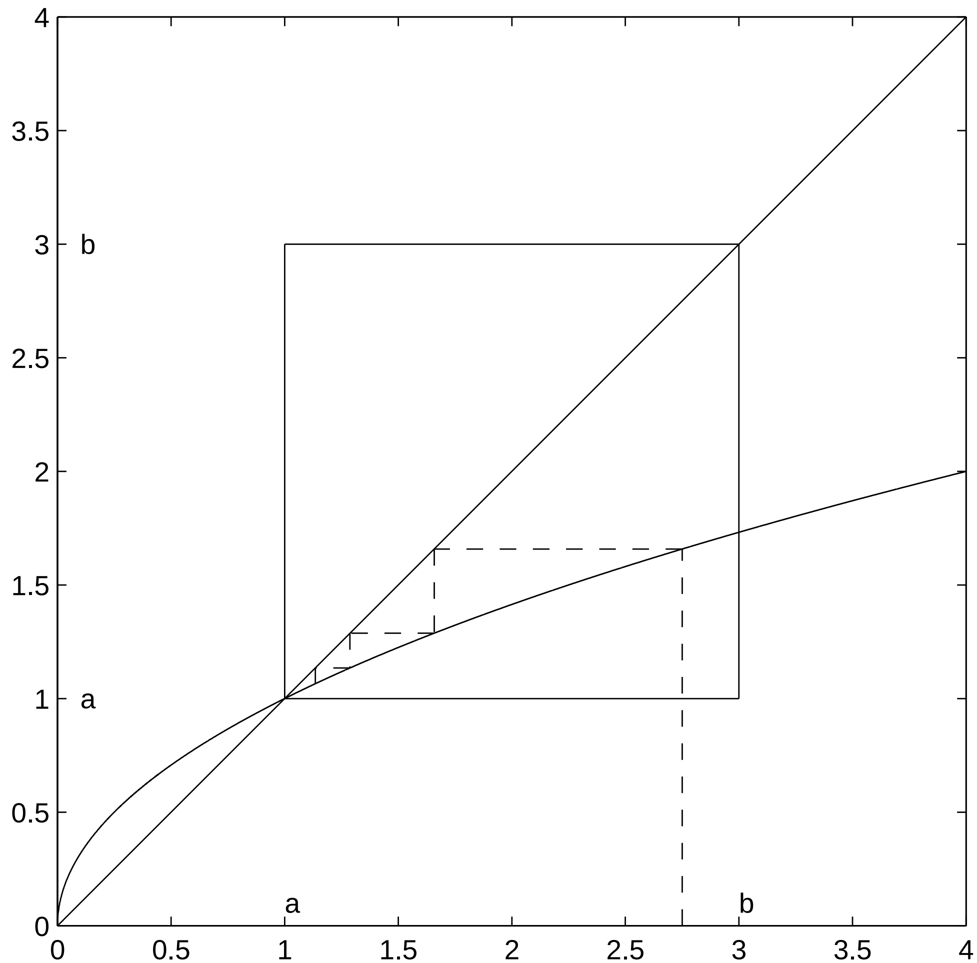
and p repels the sequence once some $x^{(k)}$ is close enough.

Example

$f(x) = x^2 - x$ has roots $\alpha_0 = 0$ and $\alpha_1 = 1$. Take $\phi_1(x) = \sqrt{x}$.

- $\phi_1'(x) = 1/(2\sqrt{x})$
- $1/4 < x \to |\phi_1'(x)| < 1$ and contraction map
- $0 < x \leq 1/4 \to |\phi_1'(x)| \geq 1$
- $0 < x < 1 \to 0 < \phi_1(x) < 1$ and $\phi_1(x) > x$; moving toward 1.
- $x > 1 \to 1 < \phi_1(x) < x$; moving toward 1.
- $0 < x \leq 1/4$ we have $\phi_1'(x) \geq 1$ but since $\phi_1(x) > x$ convergence to 1 occurs once $x^{(k)} > 1/4$; convergence from outside contracting interval.
- $\phi_1'(0) = \infty$ so we do not expect convergence to $\alpha_0 = 0$.

$f(x) = x^2 - x$, $\phi_1(x) = \sqrt{x}$, $x^{(0)} = 2.75$, $x^{(k)} \to 1$

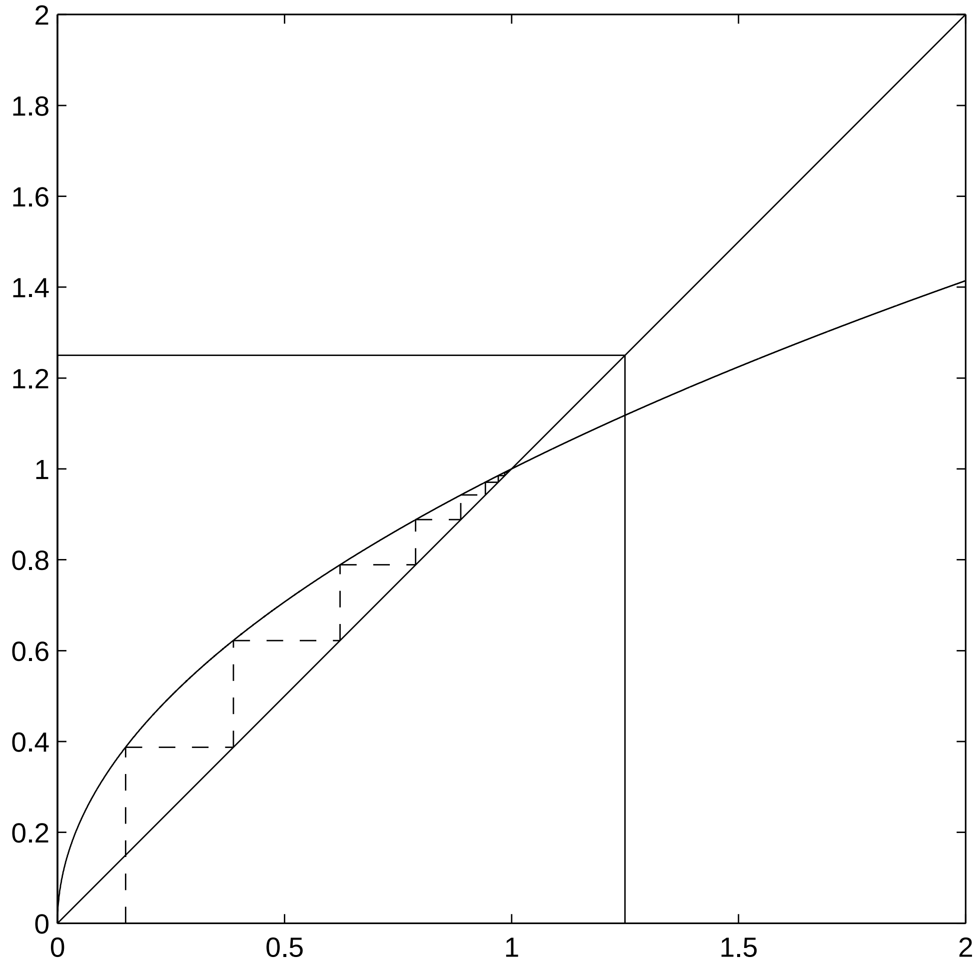


Example

$f(x) = x^2 - x$, $\phi_1(x) = \sqrt{x}$ and $x^{(0)} = 2.75$

k	$x^{(k)}$	$f(x^{(k)})$
10	1.0019777361760	0.0019816476164
11	1.0009883796409	0.0009893565352
12	1.0004940677689	0.0004943118719
13	1.0002470033791	0.0002470643898
14	1.0001234940642	0.0001235093150
15	1.0000617451259	0.0000617489383
16	1.0000308720864	0.0000308730395
17	1.0000154359241	0.0000154361623
18	1.0000077179322	0.0000077179918
19	1.0000038589587	0.0000038589736

$f(x) = x^2 - x$, $\phi_1(x) = \sqrt{x}$, $x^{(0)} = 0.15$, $x^{(k)} \to 1$



Example

$f(x) = x^2 - x$, $\phi_1(x) = \sqrt{x}$ and $x^{(0)} = 0.15$

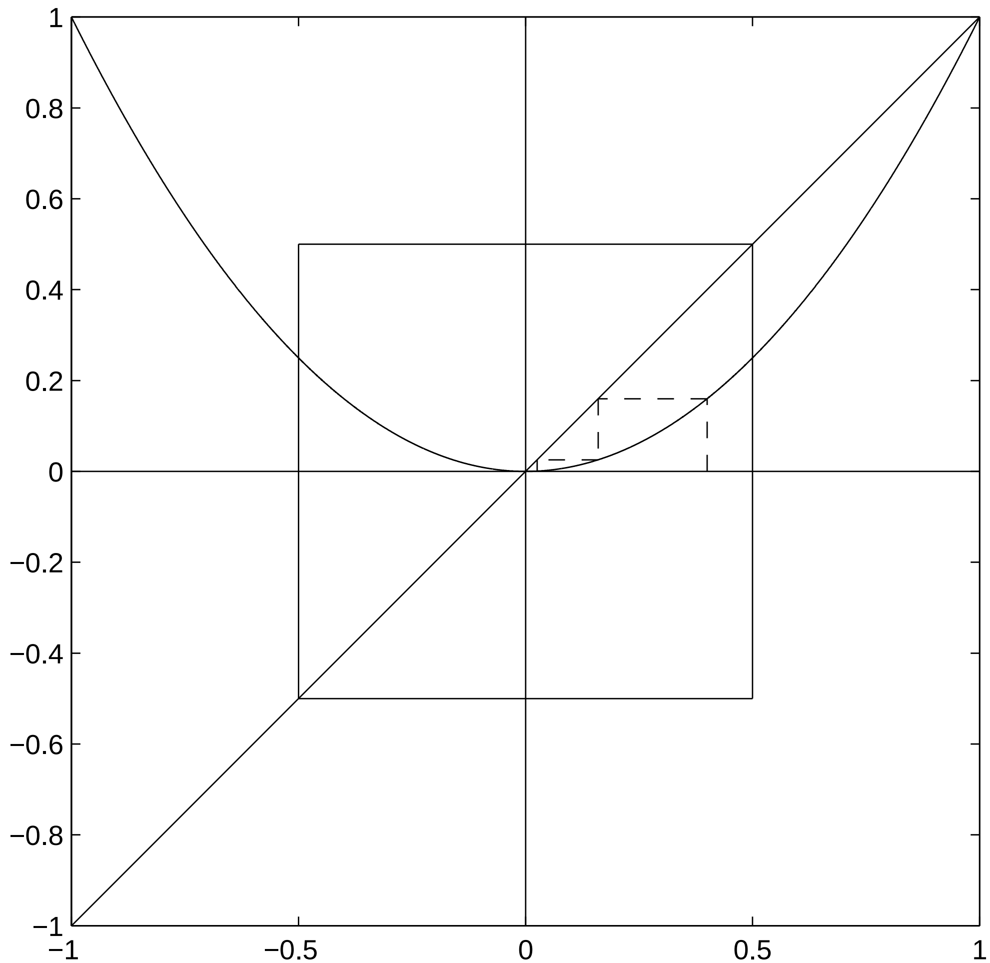
k	$x^{(k)}$	$f(x^{(k)})$
7	0.9707925300779	-0.0283543936228
8	0.9852880442175	-0.0144955141396
9	0.9926167660369	-0.0073287218194
10	0.9963015437290	-0.0036847776922
11	0.9981490588730	-0.0018475151440
12	0.9990741007918	-0.0009250419188
13	0.9995369431851	-0.0004628423933
14	0.9997684447836	-0.0002315015985
15	0.9998842156888	-0.0001157709052
16	0.9999421061686	-0.0000578904797

Example

$f(x) = x^2 - x$ has roots $\alpha_0 = 0$ and $\alpha_1 = 1$. Take $\phi_2(x) = x^2$.

- $\phi_2'(x) = 2x$.
- $-1/2 < x < 1/2 \rightarrow |\phi_2'(x)| < 1, \;\; \therefore$ convergence to $\alpha_0 = 0$ with $x^{(0)} \in (-1/2, 1/2)$ due to contraction mapping.
- $1/2 \leq |x| < 1 \rightarrow |\phi_2(x^{(k)})| < |x^{(k)}|$ and $|\phi_2'(x)| \geq 1, \;\; \therefore$ convergence to to $\alpha_0 = 0$ with $x^{(0)} \in (-1, 1)$; convergence from outside contracting interval.
- $\phi_2'(1) = 2$ and $|\phi_2(x)| > |x|, \;\; \therefore$ we do not expect convergence to $\alpha_1 = 1$ when $x^{(0)} > 1$.
- $\phi_2(1) = 1$ so convergence if you start at root $\alpha_1 = 1$. In general, however, starting at an "unstable" fixed point is not a reliable numerical situation.

$f(x) = x^2 - x$, $\phi_2(x) = x^2$, $x^{(0)} = 0.4$, $x^{(k)} \to 0$

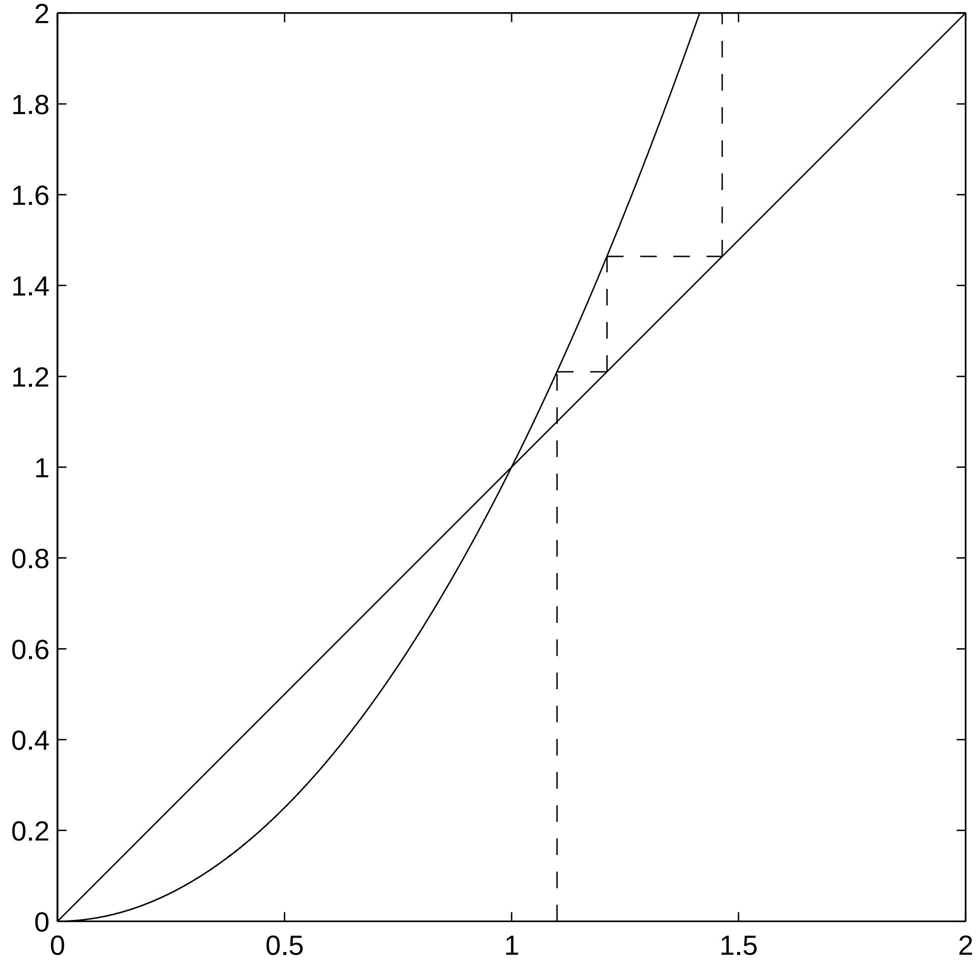


Example

$f(x) = x^2 - x$, $\phi_2(x) = x^2$ and $x^{(0)} = 0.4$

k	$x^{(k)}$	$f(x^{(k)})$
1	0.4000000000000	-0.2400000000000
2	0.1600000000000	-0.1344000000000
3	0.0256000000000	-0.0249446400000
4	0.0006553600000	-0.0006549305033
5	0.0000004294967	-0.0000004294965

$f(x) = x^2 - x$, $\phi_2(x) = x^2$, $x^{(0)} = 1.1$, $x^{(k)} \to \infty$



Example

- $f(x) = x^2 - x$, roots $\alpha_0 = 0$ and $\alpha_1 = 1$
- Two different iterations are easily derived

$$\phi_1(x) = \sqrt{x} \quad \text{and} \quad \phi_2(x) = x^2$$

- Different convergence properites: in this case different limits
- Note that empirically $\phi_2(x^{(k)}) \to 0$ faster than $\phi_1(x^{(k)}) \to 1$ asymptotically.
- $\phi_1'(1) = 1/2$ and $\phi_2'(0) = 0$ is related to this observation, i.e., want $|\phi(x)|$ as small as possible as long as the work per step and total work is not increased too much.

Convergence of a Fixed Point Method

Theorem 9.8. *Suppose* $\phi \in \mathcal{C}^{p+1}$ *in an open interval around the fixed point* α *for some integer* $p \geq 1$. *If* $\phi^{(i)}(\alpha) = 0$ *for* $1 \leq i \leq p$ *and* $\phi^{(p+1)}(\alpha) \neq 0$ *then the iteration has order* $p + 1$ *convergence for* $x^{(0)}$ *in the interval and*

$$\lim_{k\to\infty} \frac{(x^{(k+1)} - \alpha)}{(x^{(k)} - \alpha)^{p+1}} = \frac{\phi^{(p+1)}(\alpha)}{(p+1)!}$$

Proof. See Textbook page 262. $\square$

Convergence of a Fixed Point Method

Corollary 9.9. *If $p = 0$ in Theorem 9.8 for $\phi(x)$ and $|\phi'(\alpha)| < 1$ then the iteration*

$$x^{(k+1)} = \phi(x^{(k)})$$

- *has linear convergence, i.e., order* 1,
- *has the asymptotic convergence factor*

$$\lim_{k\to\infty} \frac{(x^{(k+1)} - \alpha)}{(x^{(k)} - \alpha)} = \phi'(\alpha) < 1$$

- *has the asymptotic convergence rate*

$$R = -\log(|\phi'(\alpha)|)$$

Convergence of a Fixed Point Method

- $\phi_1'(1) = 1/2 \rightarrow$ linear convergence to $\alpha_1 = 1$
- $\phi_2'(0) = 0 \rightarrow$ quadratic convergence to $\alpha_0 = 0$
- Given two iterations of the same order then, generally, you should choose the one with smaller $|\phi^{(p+1)}(\alpha)|$
- may or may not know $|\phi^{(p+1)}(\alpha)|$
- If work per step is dramatically larger for iteration with smaller $|\phi^{(p+1)}(\alpha)|$ then the other iteration may be preferred if $x^{(0)}$ is known to be close enough to α.

Newton's Method

Given $f(x)$ with $f(\alpha) = 0$ the method is

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} \to \phi(x) = x - \frac{f(x)}{f'(x)}$$

$$\phi'(x) = \frac{f(x)f''(x)}{(f'(x))^2}$$

We therefore have, if α is a simple root

$$f'(\alpha) \neq 0 \to \phi'(\alpha) = \frac{f(\alpha)f''(\alpha)}{(f'(\alpha))^2} = 0$$

$$\exists\delta > 0, \quad \text{such that } \; |x^{(k)} - \alpha| < \delta \to \lim_{k\to\infty} x^{(k)} = \alpha$$

Newton's Method

Theorem 9.10. *If* $f(\alpha) = 0$, $f'(\alpha) \neq 0$, *and* $\lim_{k\to\infty} x^{(k)} = \alpha$, *i.e., Newton converges to* α *then*

$$\lim_{k\to\infty} \frac{x^{(k+1)} - \alpha}{(x^{(k)} - \alpha)^2} = \frac{\phi''(\alpha)}{2} = \frac{f''(\alpha)}{2f'(\alpha)}$$

and therefore convergence is asymptotically quadratic to a simple root.

Newton's Method

$f(x) = x^2 - 2$, $f'(x) = 2x$, $\alpha_{\pm} = \pm\sqrt{2}$, $x^{(0)} = 1.0$

k	$x^{(k)}$	$f(x^{(k)})$
0	1.0000000000000	-1.0000000000000
1	1.5000000000000	0.2500000000000
2	1.4166666666667	0.0069444444444
3	1.4142156862745	0.0000060073049
4	$\underline{1.414213562}3747$	0.0000000000045

$$e^{(4)} = |x^{(4)} - \sqrt{2}| < 10^{-9}$$

Newton's Method

Newton's method can converge to a multiple root. Convergence is linear with the constant growing as the mulitplicity increases.

$$f(x) = x^d \rightarrow f'(x) = dx^{d-1}$$

$$x^{(k+1)} = x^{(k)} - \frac{(x^{(k)})^d}{d(x^{(k)})^{d-1}} = (1 - \frac{1}{d})x^{(k)}$$

$$\frac{|x^{(k+1)}|}{|x^{(k)}|} = (1 - \frac{1}{d})$$

If the multiplicity is known using Modified Newton can restore quadratic convergence. If it must be estimated then Adaptive Newton can improve convergence.

Newton's Method

Newton's method can cycle.

$$f(x) = 4x^3 - 10x, \quad f'(x) = 12x^2 - 10$$

$$\phi(x) = x - \frac{(2x^2 - 5)x}{(6x^2 - 5)} = \left(\frac{4x^2}{6x^2 - 5} \right) x$$

$$x^{(k)} = \sqrt{\frac{1}{2}} \to \phi(x^{(k)}) = -\sqrt{\frac{1}{2}}$$

$$x^{(k+1)} = -\sqrt{\frac{1}{2}} \to \phi(x^{(k+2)}) = \sqrt{\frac{1}{2}}$$

Stagnation. Check for $\Delta x^{(k+1)} = -\Delta x^{(k)}$.

Termination Criteria

- Termination is based on a combination of checks.
- small changes to $x^{(k)}$ observed repeatedly
- small values of $f(x^{(k)})$, i.e., small residual
- using either alone is not reliable
- knowledge of $f(x)$ around roots is useful
- some combination of change in x and residual size are usually used
- read Section 6.5 in Text for a basic discussion

Other Possibilities

- the structure of $f(x)$ can be exploited, e.g., find the roots of a polynomial, see Text Section 6.4.
- acceleration techniques, e.g., Aitken acceleration.
- using additional evaluations of $f(x)$ rather than higher order derivatives, e.g., Steffensen vs. Newton
- using higher order recurrences often involving higher order polynomial approximations to estimate one or more derivatives of $f(x)$.