Numerical Methods for Computational Science and Engineering

Fall Semester 2017 (HS17)

Prof. Rima Alaifari, SAM, ETH Zurich

8. Iterative Methods for Nonlinear Systems of Equations

So far: learned direct methods for solving linear systems of equations

Many models in real applications involve nonlinear systems of equations

In general: these systems can't be solved directly (nor exactly)!

$\rightsquigarrow$ iterative methods for finding approximations to the solution instead.

Example: liquid in spherical tank

r ... radius of the tank

ϱ ... rate of constant liquid flow

Full tank ($h_0 = 2r$) at time $t=0$.

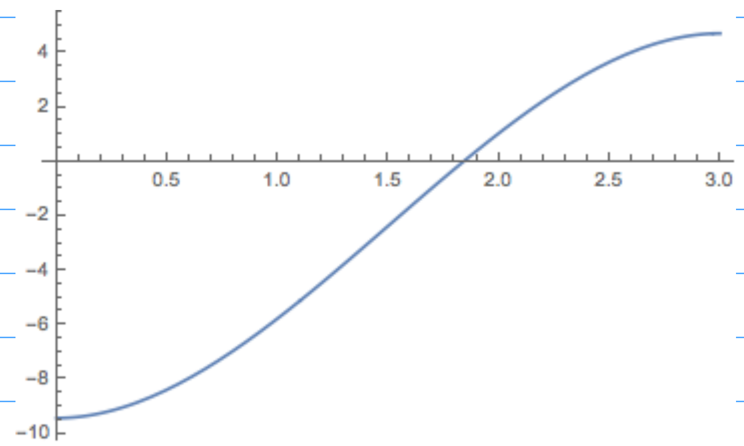
Height h of fluid at any time t:

$$-\frac{1}{3}\pi h^3 + \pi r h^2 + \left(\varrho t - \frac{4}{3}\pi r^3\right) = 0$$

Task: Given any time t, determine height h.

Define $f_t(h) := -\frac{1}{3}\pi h^3 + \pi r h^2 + \left(\varrho t - \frac{4}{3}\pi r^3\right)$

Find root h^*, i.e. solve for $f(h^*) = 0$.



$$t = \frac{1}{3} \cdot \underbrace{\frac{\frac{4}{3}\pi r^3}{S}}_{t_{end}}$$

$(r = 1.5m)$

Typical examples of nonlinear equations:

- thermodynamic models [involve equations of state for real gases]
- Colebrook equation for friction factor [= pressure drop in oil or gas pipeline]

$$\boxed{\text{How to solve a nonlinear equation } f(x) = 0 \text{ ?}}$$

One question before that:

When is $f(x)=0$ solvable?

Take $f(x) = e^{-\pi x^2}$ or $f(x) = \text{sign}(x)$

Both do not have roots.

First simple criterion: intermediate value theorem!

Theorem [IVT]: If $f : [a,b] \to \mathbb{R}$ is continuous and for some $t_l, t_r \in [a,b]$:

$$f(t_l) < u < f(t_r)$$

Then, there exists $z \in (t_l, t_r)$ s.t.

$$f(z) = u.$$

For root finding of $f \in C^0([a,b], \mathbb{R})$ f is real-val.

If for some $t_l, t_r \in [a,b]$: $f(t_l) < 0$ & $f(t_r) > 0$

$\Rightarrow f$ has root in (t_l, t_r).

→ Idea for our first algorithm for root finding:

Bisection algorithm: (for finding root x^*)

While $n \leq n_{max}$

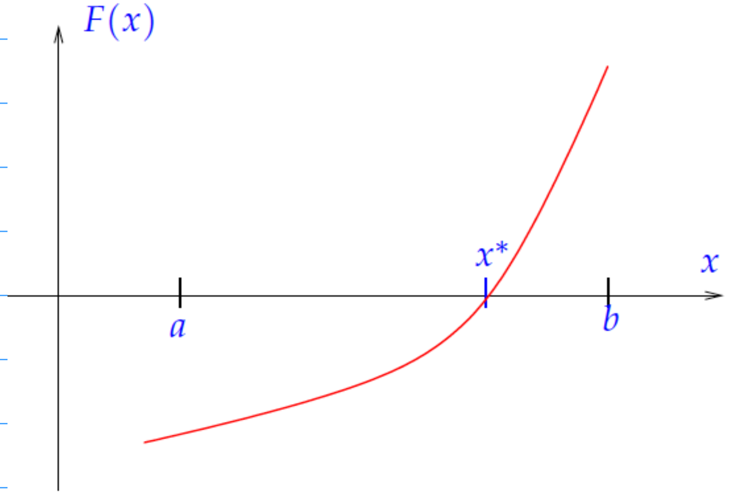
  Compute $b \leftarrow \frac{t_l + t_r}{2}$

  If $|f(b)| < TOL1$ or $|t_l - t_r| < TOL2$, return $x^* = b$.

  $n \leftarrow n+1$

  If $\text{sign}(f(t_l)) == \text{sign}(f(b))$, then $t_l \leftarrow b$

    Else $t_r \leftarrow b$



If f is cont: we always search in a region where a root exists

Region size is halved in every iteration step → convergence

Example [height of fluid in spherical tank]

Find height at time $t^* = \frac{1}{3} \cdot \underbrace{\frac{\frac{4}{3}\pi r^3}{\rho}}_{t_{end}}$ (no fluid in tank for first time)

$$f(h) = -\frac{1}{3}\pi h^3 + \pi r h^2 - \frac{8}{9}\pi r^3$$

We know: $f(0) < 0$

  $f(2r) > 0$

→ apply bisection method for $t_l = 0$ and $t_r = 2r$.

For iterates $(x^{(k)})_{k\in\mathbb{N}}$ of approximate solutions define iteration error $e^{(k)} = x^{(k)} - x^*$

solution

Rate of convergence $x^{(k)} \to x^*$?

For bisection method:

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

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

$$|x^{(k)} - x^*| \leq 2^{-k}|a-b| \qquad |e^{(k)}| \to 0 \text{ as } k\to\infty$$

"linear-type" convergence

error reduced by a fixed factor (here: $\frac{1}{2}$) in each step.

Bisection: ⊕ robustness, global convergence

⊖ rather slow convergence

no extension to higher dimensions

[n different quantities have to be zero simultaneously]

Can we find methods that

- extend to higher dimensions
- guarantee faster convergence (under additional assumptions)

Fixed Point Iterations:

Root finding of f, i.e. $f(x)=0$

$\Leftrightarrow$ finding fixed point of $\phi(x) := f(x) + x$

x^* is a fixed point (FP) of Φ $:\Leftrightarrow$ $\Phi(x^*) = x^*$

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

For bisection: only needed continuity of f

Now: Assume that Φ is Lipschitz cont. on $[a,b]$:

$$\exists\, L>0 \quad \forall\, x,y \in [a,b] : \quad |\Phi(x) - \Phi(y)| \le L \cdot |x-y|$$

notice order

Further assume $L < 1$ (Φ is a contractive mapping)

Idea of FPI:

- Start with initial guess $x^{(0)}$
- Iterate $x^{(k)} = \Phi(x^{(k-1)})$ [FPI]

$L < 1$ guarantees: If Φ has a fixed point, i.e. there exists x^* s.t. $\Phi(x^*) = x^*$, FPI will converge to x^*.

Derivation: We need to show $|e^{(k)}| \to 0$ as $k \to \infty$.

$$|e^{(k)}| = |x^{(k)} - x^*| = |\underbrace{\Phi(x^{(k-1)})}_{\text{def. of } x^{(k)}} - \underbrace{\Phi(x^*)}_{x^* \text{ is FP}}|$$

$$\le L \cdot |x^{(k-1)} - x^*| = L \cdot |e^{(k-1)}|$$

$$\Rightarrow |e^{(k)}| \le \underbrace{L^k}_{\to 0} \cdot |e^{(0)}| \to 0 \text{ as } k \to \infty. \quad \square$$

$L < 1$!!!

Remarks: • It would suffice to have Φ Lipschitz with $L < 1$ on $[x^* - \delta, x^* + \delta]$ and $x^{(0)} \in [x^* - \delta, x^* + \delta]$

- If $\Phi \in C^1([a,b])$ and $|\Phi'(x^*)| < 1$

Convergence of FPI in a neighborhood of x^*:

There are $\delta > 0$, $\varepsilon > 0$ s.t. $|\Phi'(x)| < 1-\varepsilon$

for all $x \in [x^*-\delta, x^*+\delta] =: I^*$

Then, for all $x, y \in I^*$ $\exists\, \theta \in [x,y]$ s.t.

$$|\Phi(x) - \Phi(y)| = |\Phi'(\theta)| \cdot |x-y| \quad \text{[mean value thm]}$$
$$< (1-\varepsilon)\,|x-y|$$

$\Rightarrow$ Φ is Lipschitz with $L = 1-\varepsilon < 1$ on I^*.

- Convergence rate is (at least) linear:

Definition 8.1.9. Linear convergence

A sequence $\mathbf{x}^{(k)}$, $k = 0,1,2,\ldots$, in $\mathbb{R}^n$ converges linearly to $\mathbf{x}^* \in \mathbb{R}^n$,

$$\exists 0 < L < 1: \quad \underbrace{\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\right\|}_{\|e^{(k+1)}\|} \le L \underbrace{\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|}_{\|e^{(k)}\|} \quad \forall k \in \mathbb{N}_0 .$$

[Note: bisection was not linear, only of "linear-type"

because there $|e^{(k)}| \le L^k |a-b|$

but $|e^{(k)}| > L\,|e^{(k-1)}|$ was possible]

- If $\Phi'(x^*) = 0$ and Φ is C^2 in a neighborhood of x^*:

Taylor expansion of Φ at $x^{(k)}$ around x^*:

$$\Phi(x^{(k)}) = \Phi(x^*) + \underbrace{e^{(k)} \cdot \Phi'(x^*)}_{=0} + \frac{1}{2}(e^{(k)})^2\, \Phi''(x^*) + O\big((e^{(k)})^3\big)$$

$$|e^{(k)}| = |x^{(k)} - x^*| = |\Phi(x^{(k-1)}) - \Phi(x^*)|$$
$$= \frac{1}{2}\,(e^{(k-1)})^2\,|\Phi''(x^*)| + O\big((e^{(k-1)})^3\big)$$

if $|x^{(k-1)} - x^*| \le \delta < 1$ [guaranteed by conv.]:

$$|e^{(k)}| \le \frac{1}{2}\,|e^{(k-1)}|^2\,\big(|\Phi''(x^*)| + \delta\big)$$

$$|e^{(k)}| \le C \cdot |e^{(k-1)}|^2$$

$\uparrow$ quadratic convergence

Definition 8.1.17. Order of convergence → [?, Sect. 17.2], [?, Def. 5.14], [?, Def. 6.1]

A **convergent** sequence $\mathbf{x}^{(k)}$, $k = 0, 1, 2, \ldots$, in $\mathbb{R}^n$ with limit $\mathbf{x}^* \in \mathbb{R}^n$ converges with order p, if

$$\exists C > 0: \quad \left\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\right\| \le C\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|^p \quad \forall k \in \mathbb{N}_0 , \tag{8.1.18}$$

and, in addition, $C < 1$ in the case $p = 1$ (linear convergence → Def. 8.1.9).

higher p = faster convergence (fewer iterations to reach same accuracy)

Note: If Φ is not globally Lipschitz (on $[a,b]$) but only locally (e.g. $\Phi \in C^1$, $|\Phi'(x^*)| < 1$) then FPI is only locally convergent (i.e. need $x^{(0)}$ suff. close to x^*)

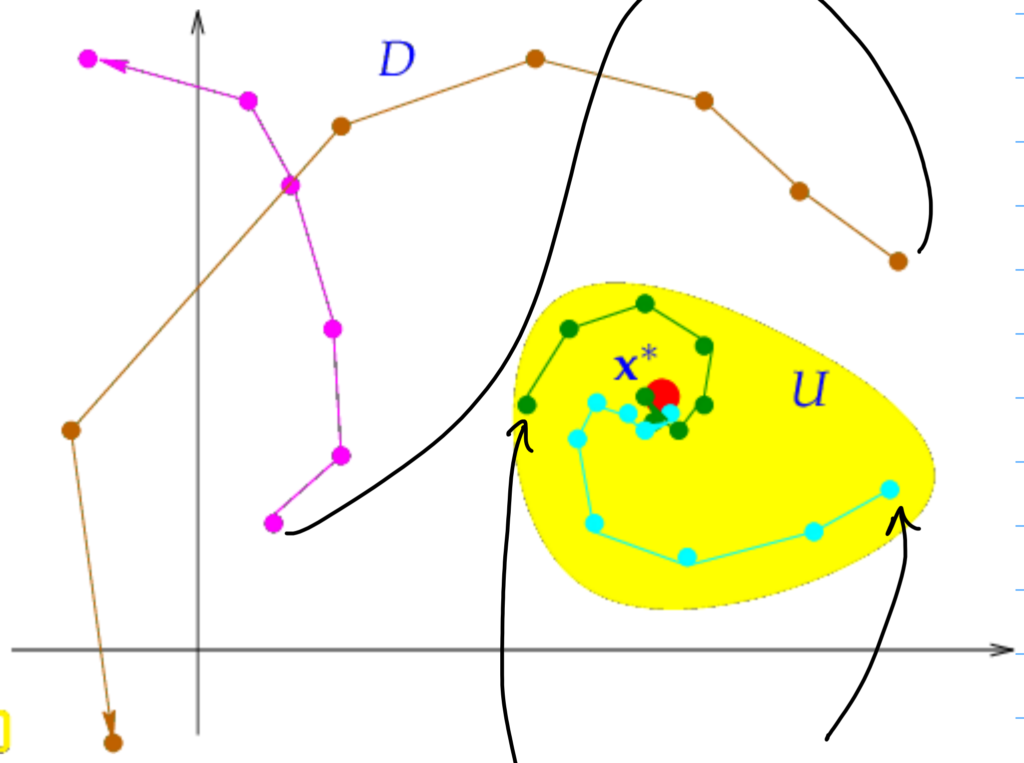


Fig. 283

$x^{(0)}$ outside of U

(illustration in 2D)

$x^{(0)}$ chosen in region where convergence is guaranteed

Algorithm for root-finding with quadratic convergence

Assumption needed: $f \in C^1$
$\uparrow$
gives first order Taylor appr.

Intuition: Approximation around $x^{(k)}$:

$$f(x) \approx f(x^{(k)}) + (x - x^{(k)}) \cdot f'(x^{(k)})$$

To find x^* with $f(x^*) = 0$ take next iterate $x^{(k+1)}$

s.t. $\quad f(x^{(k)}) + (x^{(k+1)} - x^{(k)}) f'(x^{(k)}) = 0$

$\Rightarrow$ Newton iteration:

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

Quadratic convergence when $f \in C^2$?

Reformulate as FPI: $\Phi(x) := x - \dfrac{f(x)}{f'(x)}$

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

Newton's method on f $\iff$ FPI on $\Phi(x)$

$$\Phi'(x) = 1 - \frac{(f'(x))^2 - f''(x) \cdot f(x)}{(f'(x))^2}$$

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

If $f'(x^*) \neq 0$ then $\boxed{\Phi'(x^*) = 0}$

$\uparrow$ simple root

$\exists$ neighborhood around x^* $|\Phi'| < 1$

For $x^{(0)}$ in a neighborhood I^* of x^*

[s.t. $\forall x \in I^* \quad |\Phi'(x)| \leq 1-\varepsilon$]

Newton's method converges quadratically if

$f'(x^*) \neq 0$.

(due to $\Phi'(x^*)=0$)

Note: • $x^{(0)} \in I^*$ guarantees $f'(x^{(k)}) \neq 0$ for all the iterates. This has to be guaranteed

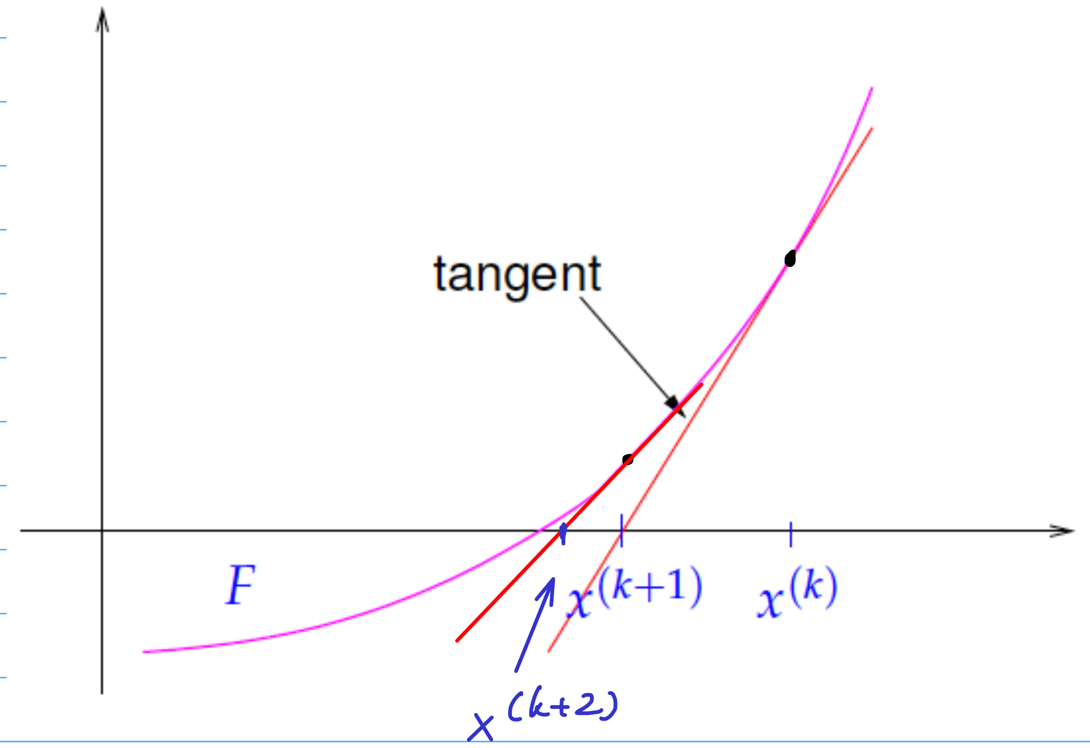
• For quadratic convergence of Newton's method $f \in C^2(I^*)$ suffices instead of $\Phi \in C^2(I^*)$

In summary: we need a neighborhood I^* of x^* s.t.

• I^* suff. small

• $f'(x\ \ ) \neq 0$ on I^*

• $f \in C^2(I^*)$

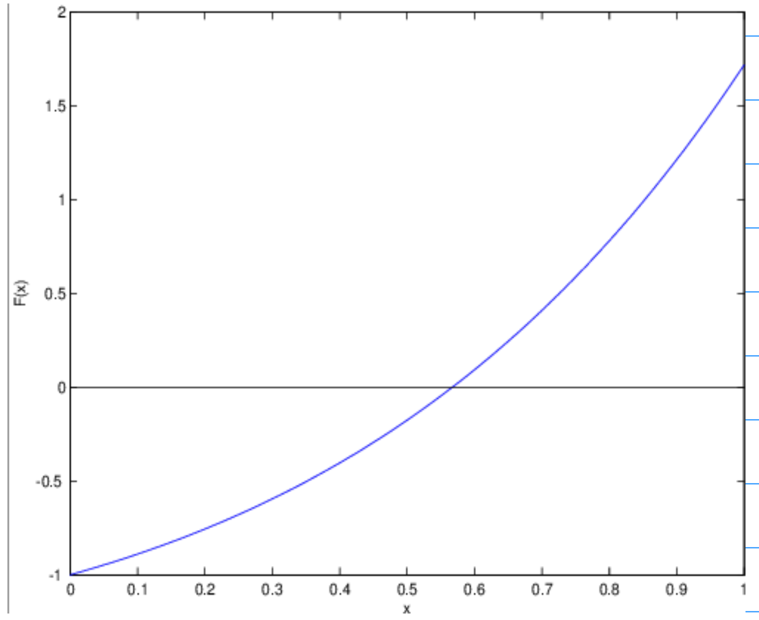


Example:

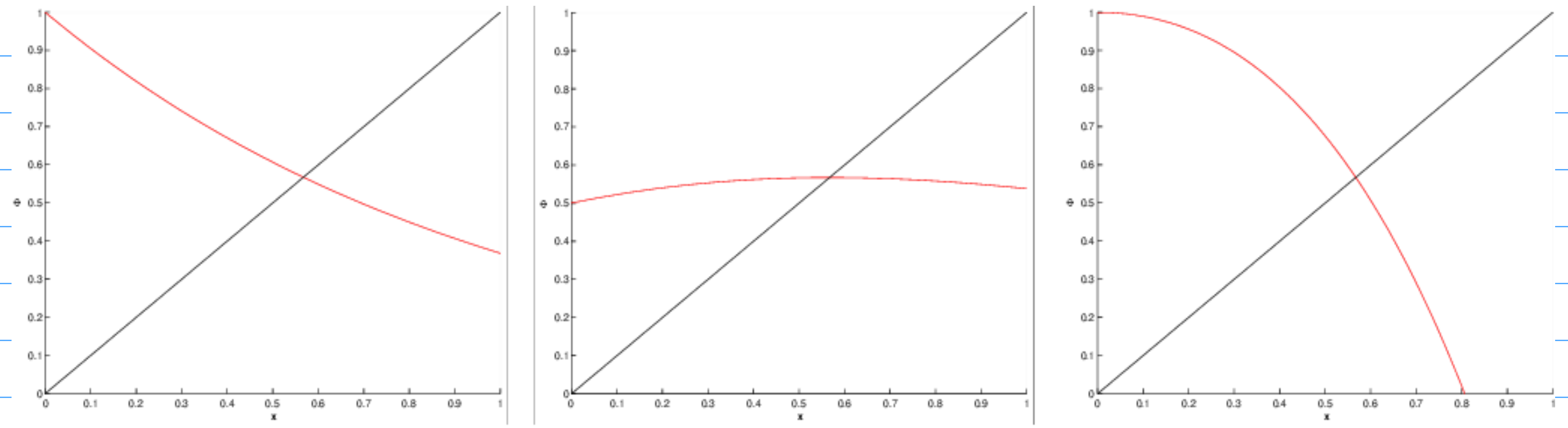
$$F(x) = xe^x - 1\,, \quad x \in [0,1]\,.$$

Different fixed point forms:

$$\Phi_1(x) = e^{-x}\,,$$
$$\Phi_2(x) = \frac{1+x}{1+e^x}\,,$$
$$\Phi_3(x) = x + 1 - xe^x\,.$$



$$x^* = e^{-x^*} \iff x^* e^* = 1$$



function Φ_1 function Φ_2 function Φ_3

FPI with $x^{(0)} = 0.5$:

k	$\Phi_1(x^*) = x^*$ $\lvert x_1^{(k+1)} - x^* \rvert$	$\Phi_2(x^*) = x^*$ $\lvert x_2^{(k+1)} - x^* \rvert$	$\Phi_3(x^*) = x^*$ $\lvert x_3^{(k+1)} - x^* \rvert$
0	0.067143290409784	0.067143290409784	0.067143290409784
1	0.039387369302849	0.000832287212566	0.108496074240152
2	0.021904078517179	0.000000125374922	0.219330611898582
3	0.012559804468284	0.000000000000003	0.288178118764323
4	0.007078662470882	0.000000000000000	0.723649245792953
5	0.004028858567431	0.000000000000000	0.410183132337935
6	0.002280343429460	0.000000000000000	1.186907542305364
7	0.001294757160282	0.000000000000000	0.146569797006362
8	0.000733837662863	0.000000000000000	0.310516641279937
9	0.000416343852458	0.000000000000000	0.357777386500765
10	0.000236077474313	0.000000000000000	0.974565695952037

roughly: FPI for Φ_1 : linearly conv.

FPI for Φ_2 : quadr. conv.

Φ_3 : no conv.

Why? $\Phi_1'(x) = -e^{-x}$

$\Rightarrow |\Phi_1'(x)| < 1$ for $x \in [\delta, 1] = I_\delta^* \quad \forall \delta > 0$

$\rightarrow$ conv. of FPI with Φ_1 in I^*

$$\Phi_2'(x) = \frac{1 - xe^x}{(1+e^x)^2}$$

$$|\Phi_2'(x)| < \frac{|1-e|}{4} < \frac{1}{2}$$

$\rightarrow$ conv. of FPI of Φ_2 on $[0,1]$

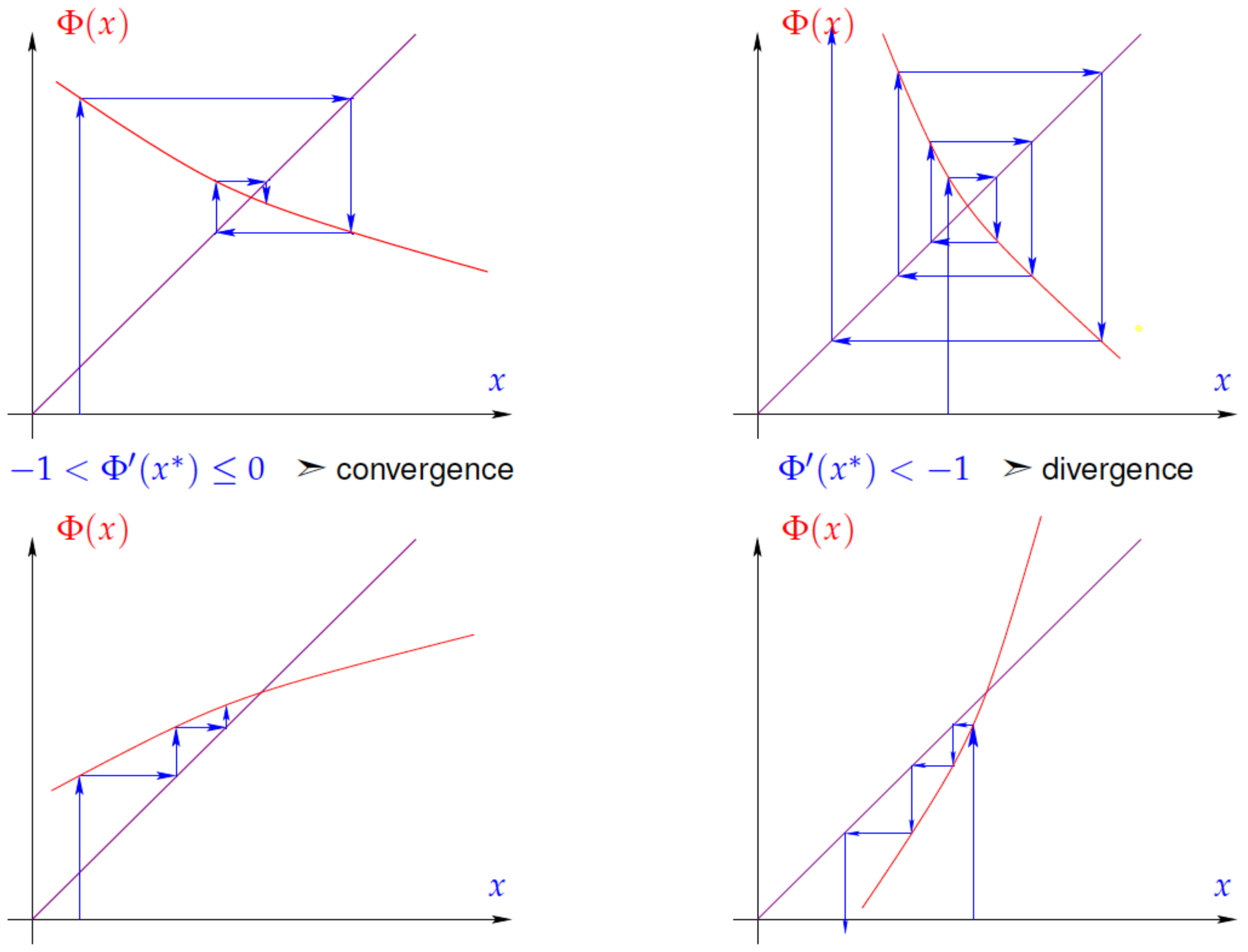
Furthermore $\Phi_2'(x^*) = 0 \rightarrow$ quadr convergence

$$\Phi_3'(x) = 1 - e^{x} - x e^{x} \qquad e^{-x^*} = x^*$$

$$\Phi_3'(x^*) = -e^{x^*} = -\frac{1}{x^*}$$

$x^* \in (0,1)$: $|\Phi_3'(x^*)| > 1$

$\Rightarrow$ FPI not contractive around x^*



$-1 < \Phi'(x^*) \leq 0$ ➢ convergence

$\Phi'(x^*) < -1$ ➢ divergence

$0 \leq \Phi'(x^*) < 1$ ➢ convergence

$1 < \Phi'(x^*)$ ➢ divergence

Remark on Newton's method

Requires computation of $f'(x^{(k)})$ for each iteration!

can be costly

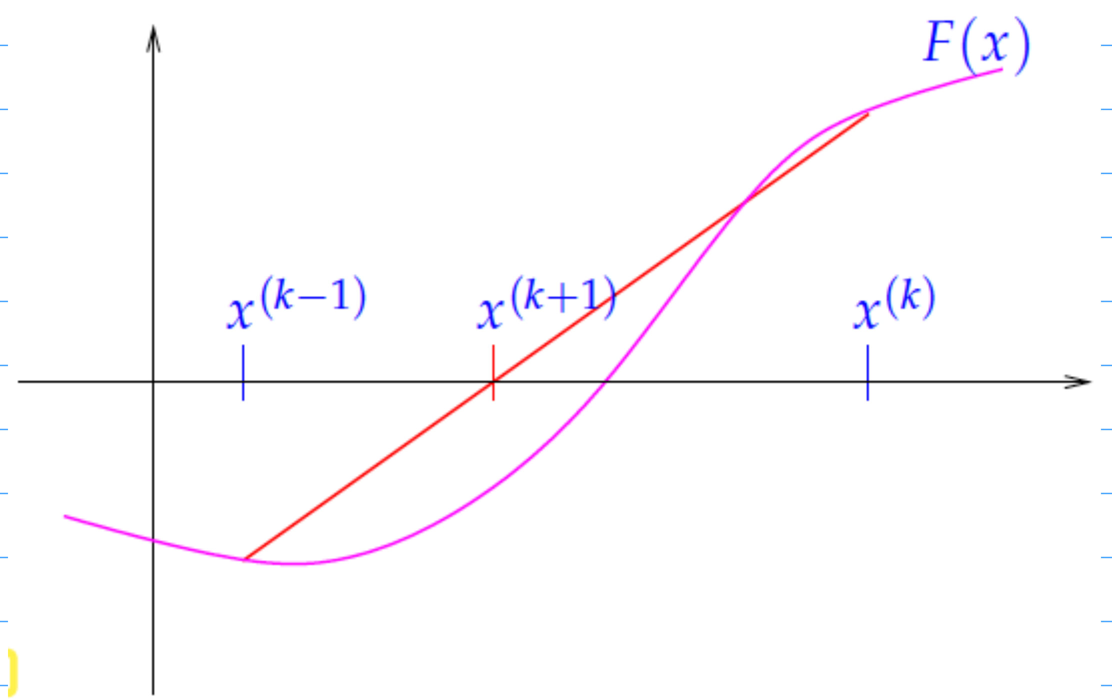
Alternative: Secant method

replace $f'(x^{(k)})$ by approximation

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

in Newton iteration to obtain

$$\boxed{x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})\,(x^{(k)} - x^{(k-1)})}{f(x^{(k)}) - f(x^{(k-1)})}}$$



Convergence of secant method:

- again local
- need $f'(x^*) \neq 0$ [simple root]
- f locally C^2

(The first three points: as for Newton's method.)

- rate is superlinear but not quadratic

  order $p = \frac{1+\sqrt{5}}{2} \approx 1.618$

Remark: Secant method is a 2-point method (computing $x^{(k+1)}$ involves $x^{(k)}, x^{(k-1)}$)

Definition: Stationary m-point iterative method

$x^{(k)}$ depends on m most recent iterates

$x^{(k-1)}, \dots, x^{(k-m)}$ iteration function

$$x^{(k)} = \Phi_F\left(x^{(k-1)}, \dots, x^{(k-m)}\right)$$

for solving $F(x) = 0$.

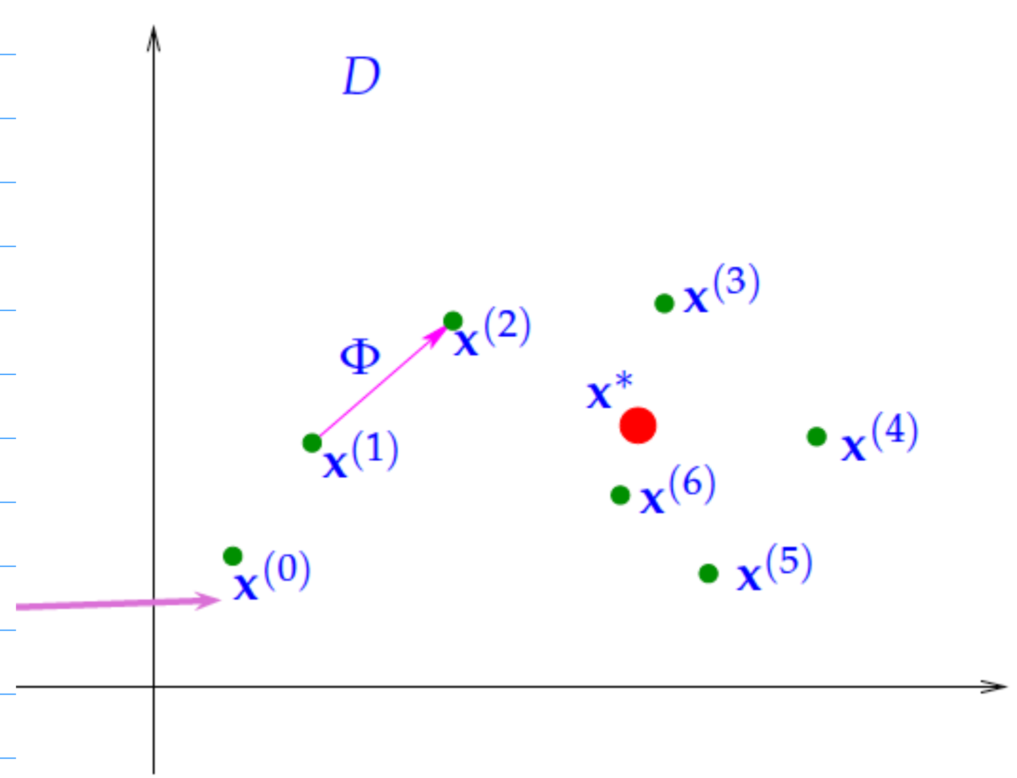
Note: m-point method requires m initial guesses $x^{(0)}, \dots, x^{(m-1)}$.

Nonlinear systems of equations

$F : D \subset \mathbb{R}^n \to \mathbb{R}^n$ (nonlinear system of equs, n equs, n unknowns)

Find x^* s.t. $F(x^*) = 0$.

initial guess



$$x^{(k)} = \Phi_F\left(x^{(k-1)}, \dots, x^{(k-m)}\right)$$

Aspects of iterative methods:

- Convergence: $(x^{(k)})_{k\in\mathbb{N}}$ convergent, $\lim_{k\to\infty} x^{(k)} = x^*$
- Consistency: $\Phi_F(x^*, \dots, x^*) = x^* \iff F(x^*) = 0$
- Rate of convergence $\|x^{(k)} - x^*\| \to 0$ with which order?

Note: $\|\cdot\|$ can be any norm for $\mathbb{R}^n$ in $\mathbb{R}^n$ (finite dim. vector space): all norms are equivalent

Definition 8.1.11. Equivalence of norms

Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on a vector space V are equivalent if

$$\exists \underline{C}, \overline{C} > 0: \quad \underline{C}\|v\|_a \le \|v\|_b \le \overline{C}\|v\|_a \quad \forall v \in V.$$

This implies: convergence in $\mathbb{R}^n$ independent of choice of norm

But in general: convergence rate depends on chosen norm

Local vs. global convergence:

Definition 8.1.8. Local and global convergence $\to$ [?, Def. 17.1]

As stationary m-point iterative method converges locally to $\mathbf{x}^* \in \mathbb{R}^n$, if there is a neighborhood $U \subset D$ of $\mathbf{x}^*$, such that

$$\mathbf{x}^{(0)}, \dots, \mathbf{x}^{(m-1)} \in U \quad \Rightarrow \quad \mathbf{x}^{(k)} \text{ well defined} \quad \wedge \quad \lim_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x}^*$$

where $(\mathbf{x}^{(k)})_{k\in\mathbb{N}_0}$ is the (infinite) sequence of iterates.
If $U = D$, the iterative method is globally convergent.

Fixed Point Iterations in $\mathbb{R}^n$

Definition 8.2.1.

A fixed point iteration $x^{(k+1)} = \Phi(x^{(k)})$ is consistent with $F(x) = 0$ if for $x \in U \cap D$

$$F(x) = 0 \iff \Phi(x) = x.$$

Definition 8.2.6. [Contractive mapping]

$\Phi : U \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ is contractive (w.r.t. norm $\|\cdot\|$ on $\mathbb{R}^n$) if

$$\exists\, L < 1 \qquad \|\Phi(x) - \Phi(y)\| \le L \cdot \|x - y\| \quad \forall\, x, y \in U$$

① contractivity of Φ $\Rightarrow$ if $\Phi(x^*) = x^*$ then FPI will converge to x^*.

$$\underbrace{\|x^{(k+1)} - x^*\|}_{\|e^{(k+1)}\|} = \|\Phi(x^{(k)}) - \Phi(x^*)\|$$

$$\le \underbrace{L}_{<1} \cdot \underbrace{\|x^{(k)} - x^*\|}_{\|e^{(k)}\|}$$

$$\|e^{(k+1)}\| \le L^k \|e^{(0)}\|$$

② Convergence is at least linear.

③ If Φ is contractive $\Rightarrow$ Φ has at most one FP.

Why? Suppose 2 FPs x_1^*, x_2^*:

$$\|x_1^* - x_2^*\| \underset{\text{FP}}{=} \|\Phi(x_1^*) - \Phi(x_2^*)\| \underset{\text{contractive}}{\le} L \cdot \|x_1^* - x_2^*\|$$

with $L < 1$ $\Rightarrow$ $x_1^* = x_2^*$.

Existence of a FP?

Theorem 8.2.9. Banach's fixed point theorem

If $D \subset \mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}, \mathbb{C}$) closed and bounded and $\Phi : D \mapsto D$ satisfies

$$\exists L < 1: \quad \|\Phi(\mathbf{x}) - \Phi(\mathbf{y})\| \le L\|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in D,$$

then there is a unique fixed point $\mathbf{x}^* \in D$, $\Phi(\mathbf{x}^*) = \mathbf{x}^*$, which is the limit of the sequence of iterates $\mathbf{x}^{(k+1)} := \Phi(x^{(k)})$ for any $\mathbf{x}^{(0)} \in D$.

Convergence criteria for FPI for Φ differentiable & knowing $\Phi(x^*) = x^*$.

Lemma 8.2.10. Sufficient condition for local linear convergence of fixed point iteration → **[?, Thm. 17.2], [?, Cor. 5.12]**

If $\Phi : U \subset \mathbb{R}^n \mapsto \mathbb{R}^n$, $\Phi(\mathbf{x}^*) = \mathbf{x}^*$, Φ differentiable in $\mathbf{x}^*$, and $\|\mathrm{D}\,\Phi(\mathbf{x}^*)\| < 1$, then the fixed point iteration

$$\mathbf{x}^{(k+1)} := \Phi(\mathbf{x}^{(k)}), \tag{8.2.2}$$

converges locally and at least linearly.

matrix norm, Def. 1.5.76 !

$$D\Phi(x) := \left[\frac{\partial \Phi_j}{\partial x_i}(x)\right]_{j,i=1}^{n} \in \mathbb{R}^{n,n}$$

Jacobian

Lemma 8.2.12. Sufficient condition for linear convergence of fixed point iteration

Let U be convex and $\Phi : U \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ be continuously differentiable with

$$L := \sup_{\mathbf{x} \in U} \|D\Phi(\mathbf{x})\| < 1 .$$

If $\Phi(\mathbf{x}^*) = \mathbf{x}^*$ for some interior point $\mathbf{x}^* \in U$, then the fixed point iteration $\mathbf{x}^{(k+1)} = \Phi(\mathbf{x}^{(k)})$ converges to $\mathbf{x}^*$ at least linearly with rate L.

$\Rightarrow$ Locally contractive Φ $\Rightarrow$ iteration converges locally around FP (at least linear conv.)

Termination criterion for contractive FPI:

When to stop iterating (for finding x^* s.t. $F(x^*)=0$)?

① residual based

Stop when $\|F(x^{(k)})\| \leq \tau$

② Correction based

Stop when $\|x^{(k+1)} - x^{(k)}\| \leq \tau$

or $\|x^{(k+1)} - x^{(k)}\| \leq \tau_{rel} \|x^{(k+1)}\|$

Recall discussion about cond. number:

$$\underbrace{\|F(x^{(k)}) - \underbrace{F(x^*)}_{=0}\|}_{\text{compute}} \text{ small} \not\Rightarrow \underbrace{\|x^{(k)} - x^*\|}_{\text{can't compute}} \text{ small}$$

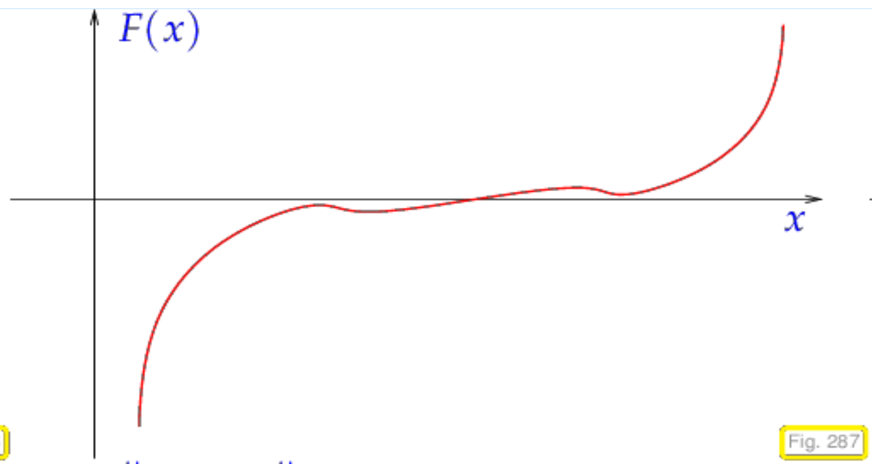


Fig. 286

$\|F(\mathbf{x}^{(k)})\|$ small $\not\Rightarrow$ $|x - x^*|$ small

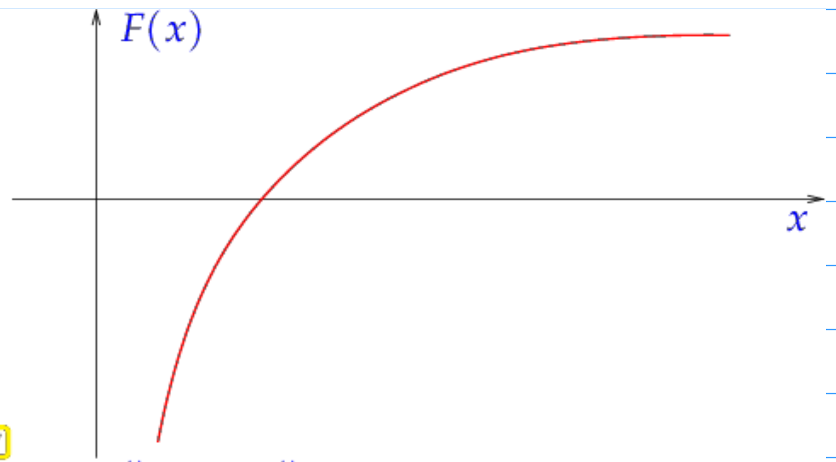


Fig. 287

$\|F(\mathbf{x}^{(k)})\|$ small $\Rightarrow$ $|x - x^*|$ small

Ultimate goal: guarantee of the form $\|x^{(k)} - x^*\| \leq \tau$

If iteration is linearly convergent:

$$\|x^{(k)} - x^*\| \leq \|x^{(k+1)} - x^{(k)}\| + \|x^{(k+1)} - x^*\|$$
$$\leq \|x^{(k+1)} - x^{(k)}\| + L\,\|x^{(k)} - x^*\|$$

$$\Rightarrow \quad (1-L)\,\|x^{(k)}-x^*\| \leq \|x^{(k+1)}-x^{(k)}\|$$

$$\Rightarrow \quad \underbrace{\|x^{(k+1)}-x^*\|}_{\text{not computable}} \leq L\,\|x^{(k)}-x^*\|$$

$$\leq \frac{L}{1-L}\,\underbrace{\|x^{(k+1)}-x^{(k)}\|}_{\text{computable!}}$$

Suggests to ask for

$$\frac{L}{1-L}\,\|x^{(k+1)}-x^{(k)}\| \leq \tau$$

as stopping criterion. It guarantees

$$\|x^{(k+1)}-x^*\| \leq \tau .$$

Note: Estimating L can be difficult.

But: pessimistic estimate $\tilde{L} > L$ is still reliable!

More generally:

$$\left\|\mathbf{x}^*-\mathbf{x}^{(k)}\right\| \leq \frac{L^{k-l}}{1-L}\left\|\mathbf{x}^{(l+1)}-\mathbf{x}^{(l)}\right\| . \qquad (8.2.21)$$

Set $l=0$ in (8.2.21)	Set $l=k-1$ in (8.2.21)
a priori termination criterion	a posteriori termination criterion
$\left\|\mathbf{x}^*-\mathbf{x}^{(k)}\right\| \leq \frac{L^{k}}{1-L}\left\|\mathbf{x}^{(1)}-\mathbf{x}^{(0)}\right\|$ (8.2.22)	$\left\|\mathbf{x}^*-\mathbf{x}^{(k)}\right\| \leq \frac{L}{1-L}\left\|\mathbf{x}^{(k)}-\mathbf{x}^{(k-1)}\right\|$ (8.2.23)

8.4. Newton's method (in higher dimensions)

Extend idea from 1D:

First order approximation (linearization)

$$F(x) \approx F(x^{(k)}) + \underbrace{DF(x^{(k)})}_{\in \mathbb{R}^{n,n}}(x-x^{(k)}) =: \tilde{F}_k(x)$$

Jacobian of F at $x^{(k)}$

▶ Newton iteration: (generalizes (8.3.4) to $n > 1$)

$$\boxed{\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} - \mathrm{D}F(\mathbf{x}^{(k)})^{-1}F(\mathbf{x}^{(k)})} \quad , \quad [\text{ if } \mathrm{D}F(\mathbf{x}^{(k)}) \text{ regular }] \tag{8.4.1}$$

Terminology: $-\mathrm{D}F(\mathbf{x}^{(k)})^{-1}F(\mathbf{x}^{(k)})$ = Newton correction

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

needed $f'(x^{(k)}) \neq 0$

Now : Invertibility of Jacobian

To compute Newton correction : solve LSE

$$DF(x^{(k)})\, y = -F(x^{(k)})$$

Convergence of Newton's method :

If $F(x^*) = 0$ and $DF(x^*)$ is regular, then it is locally quadratically convergent.

Exact theorem : Thm 8.4.45 in lecture notes

$\rightsquigarrow$ hardly ever possible to verify in practice

Note : If $DF(x^*)$ is singular, Newton is only locally <u>linearly</u> convergent.

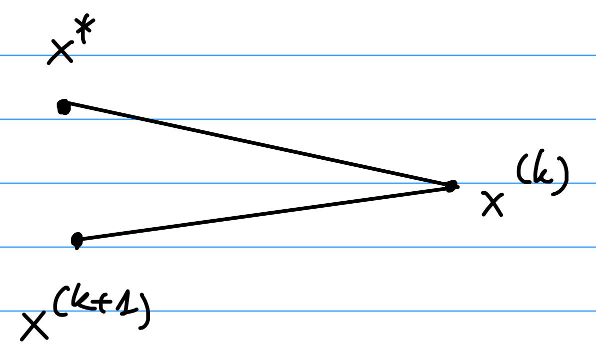
A few remaining questions :

1.) Stopping criterion for Newton's method ?
2.) Larger region of convergence possible ?
[At the cost of loosing quadratic convergence]
3.) Newton correction is costly [Solve LSE with different system matrix in <u>each</u> iteration]

$\to$ Any remedy ?

<u>Ad 1.)</u> Quadratic convergence

$$\|x^{(k+1)} - x^*\| \ll \|x^{(k)} - x^*\|$$



Roughly: $\|x^{(k)} - x^*\| \approx \underbrace{\|x^{(k+1)} - x^{(k)}\|}$

$$\|x^{(k+1)} - x^{(k)}\| = \underbrace{\|DF(x^{(k)})^{-1} F(x^{(k)})\| \leq \tau \|x^{(k)}\|}_{\text{computable stopping criterion}}$$

guarantee $\|x^{(k)} - x^*\| \lesssim \tau \|x^{(k)}\|$

BUT: If $x^{(k)}$ was a good approximation, we would have computed new Newton correction $DF(x^{(k)})^{-1} F(x^{(k)})$ but not used in iteration.

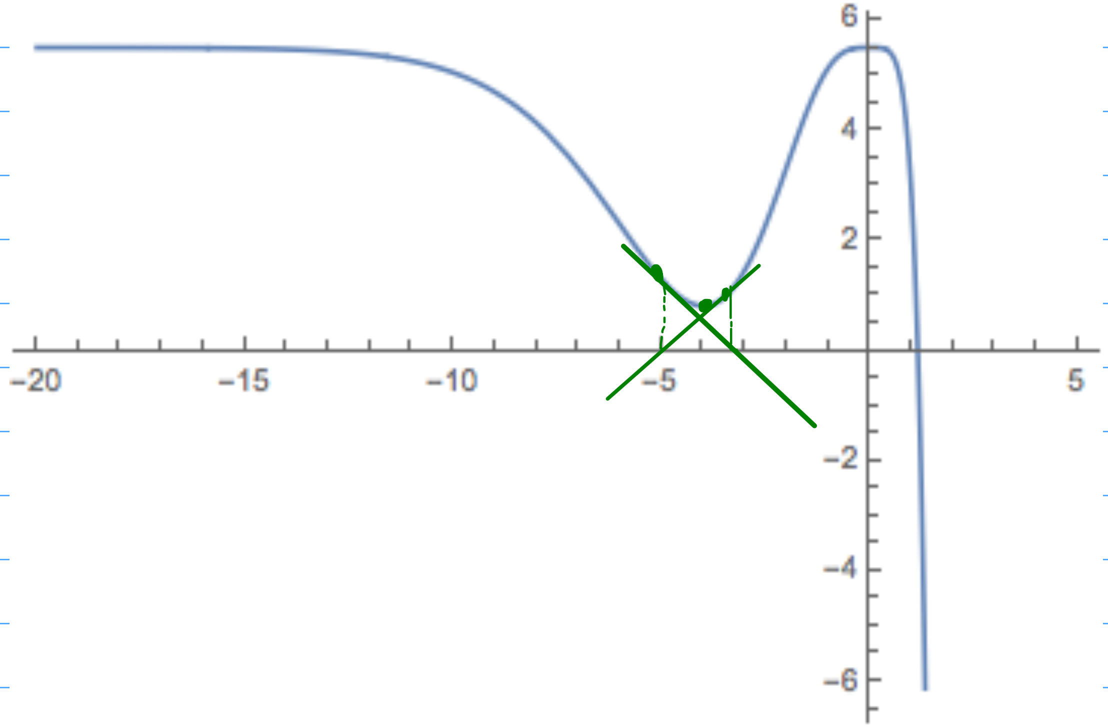
Idea: "Cheaper" stopping criterion:

$$\underbrace{\|DF(x^{(k-1)})^{-1} F(x^{(k)})\|}_{\text{simplified Newton correction}} \leq \tau \|x^{(k)}\|$$

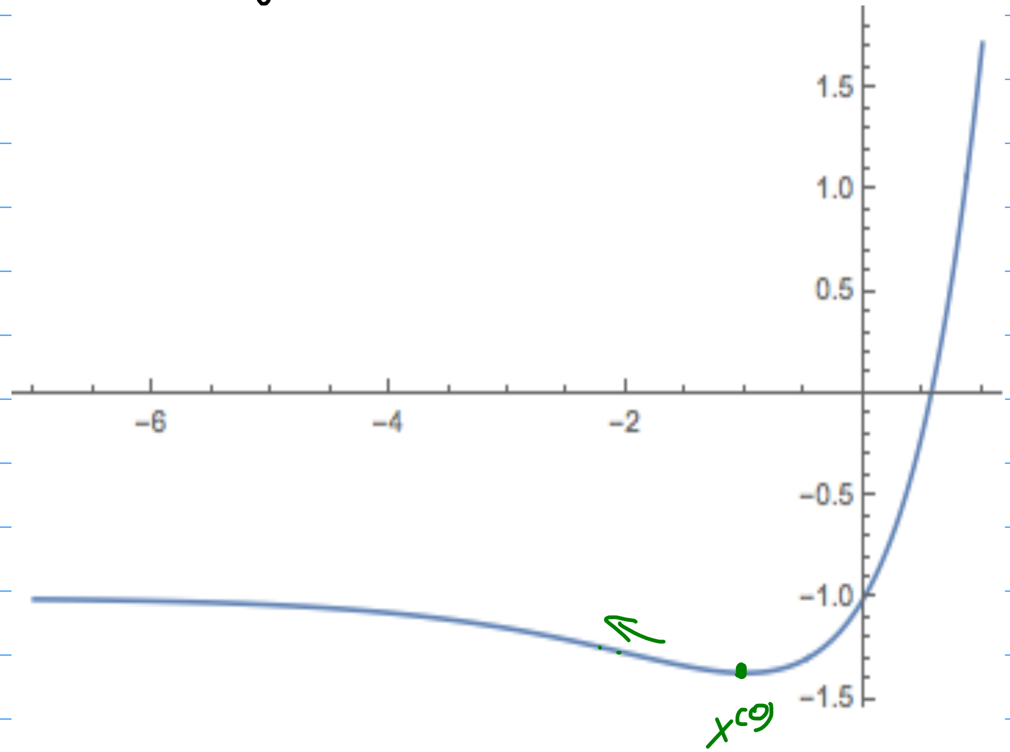
motivation:
- Due to fast convergence $DF(x^{(k)}) \approx DF(x^{(k-1)})$ during last steps of iteration
- We can reuse LU factorization of $DF(x^{(k-1)})$

Ad 2.) Examples of failures of Newton's method

① Local min. /max.



② Asymptotes



$$F(x) = xe^x - 1$$

$$F'(x) = xe^x + e^x$$

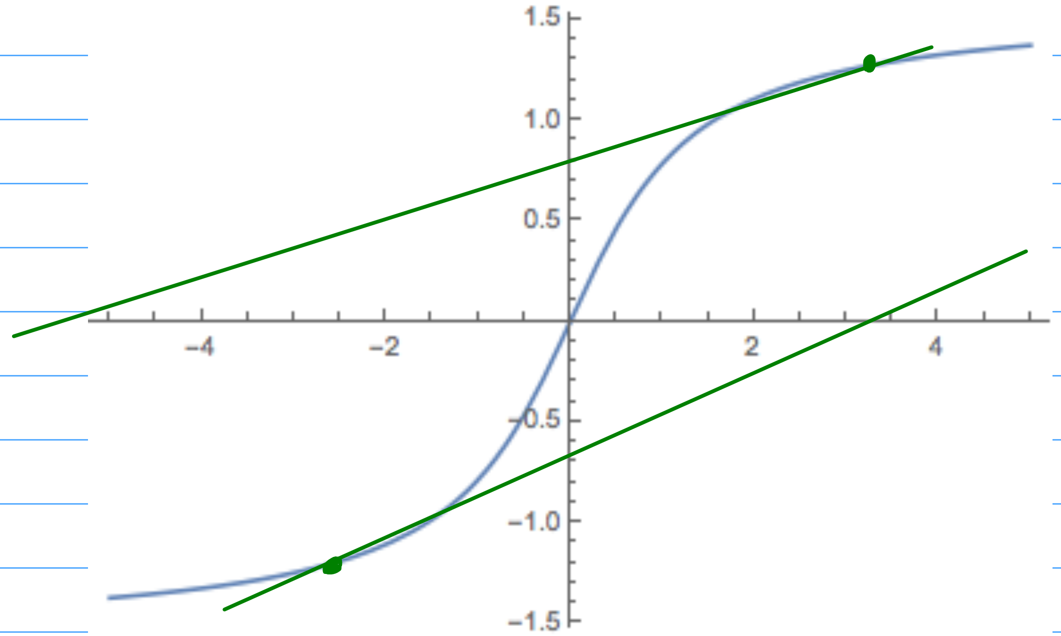
$$F'(-1) = 0$$

$$x^{(0)} < -1 \Rightarrow x^{(k)} \to -\infty$$

$$x^{(0)} > -1 \Rightarrow x^{(k)} \to x^*$$

③ Overshooting

$F(x) = \arctan(x) \qquad F(0) = 0$



A remedy for overshooting:

8.4.4 Damped Newton method

Idea: In each iteration step, check whether distance $\|x^{(k+1)} - x^{(k)}\|$ is decreasing

i.e. roughly whether at least

$$\|x^{(k+2)} - x^{(k+1)}\| \leq \frac{1}{2} \|x^{(k+1)} - x^{(k)}\|$$

If not, don't take a full Newton step!

→ Damp the Newton correction

▶ we observe "overshooting" of Newton correction

Idea: damping of Newton correction:

With $\lambda^{(k)} > 0$: $\quad \mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} - \lambda^{(k)} \, \mathrm{D}F(\mathbf{x}^{(k)})^{-1} F(\mathbf{x}^{(k)})$.

Terminology: $\lambda^{(k)}$ = damping factor

How to choose $\lambda^{(k)}$?

Strategy: largest possible $\lambda^{(k)}$ so that distance between iterates is decreasing

Affine invariant damping strategy

Choice of damping factor: affine invariant natural monotonicity test [?, Ch. 3]:

choose "maximal" $0 < \lambda^{(k)} \leq 1$: $\left\|\Delta\bar{\mathbf{x}}(\lambda^{(k)})\right\| \leq (1 - \frac{\lambda^{(k)}}{2})\left\|\Delta\mathbf{x}^{(k)}\right\|_2$ (8.4.57)

where $\Delta\mathbf{x}^{(k)} := \mathrm{D}\,F(\mathbf{x}^{(k)})^{-1}F(\mathbf{x}^{(k)})$ → current Newton correction ,

$\Delta\bar{\mathbf{x}}(\lambda^{(k)}) := \mathrm{D}\,F(\mathbf{x}^{(k)})^{-1}F(\underbrace{\mathbf{x}^{(k)} - \lambda^{(k)}\Delta\mathbf{x}^{(k)}}_{\tilde{x}^{(k+1)}\text{ with damping}})$ → tentative simplified Newton correction .

$x^{(k+1)} - x^{(k)}$ with no damping

Approximation for $x^{(k+2)} - \tilde{x}^{(k+1)}$

where $\tilde{x}^{(k+1)} = x^{(k)} - \lambda^{(k)}\Delta x^{(k)}$

Check whether $\|\Delta\bar{x}(\lambda^{(k)})\|$ is strictly smaller than $\|x^{(k+1)} - x^{(k)}\|$ where $x^{(k+1)} = x^{(k)} - \Delta x^{(k)}$.

In practice: $\lambda^{(k)} = 1$ and check NMT, repeatedly take $\lambda^{(k)} \leftarrow \frac{\lambda^{(k)}}{2}$ until NMT passes for the first time.

Ad 3.) Cheaper (approximate) Newton corrections?

Secant method in 1D:

$$f'(x^{(k)}) \approx \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$$

$n > 1$?

Approximation $\underset{\in \mathbb{R}^{n,n}}{J_k} \approx DF(x^{(k)})$ s.t.

$$J_k\left(x^{(k)}-x^{(k-1)}\right) = F(x^{(k)}) - F(x^{(k-1)}) \quad (*)$$

From Newton's method for $x^{(k)}$:

$$x^{(k)} = x^{(k-1)} - J_{k-1}^{-1} F(x^{(k-1)})$$

$$\Leftrightarrow \quad J_{k-1}\left(x^{(k)}-x^{(k-1)}\right) = -F(x^{(k-1)}) \quad (**)$$

$(*)-(**)$:

$$\underbrace{\left(J_k - J_{k-1}\right)\left(x^{(k)}-x^{(k-1)}\right) = F(x^{(k)})}_{\text{underdetermined}}$$

Possible cheap choice: outer product in $\mathbb{R}^n$

$$J_k - J_{k-1} = \underbrace{\frac{F(x^{(k)})\left(x^{(k)}-x^{(k-1)}\right)^T}{\left\|x^{(k)}-x^{(k-1)}\right\|_2^2}}_{\text{rank 1 matrix}}$$

Given initial J_0 (take $J_0 = DF(x^{(0)})$),

get J_k by rank-1 updates:

$$J_k = J_{k-1} + \frac{F(x^{(k)})\left(x^{(k)}-x^{(k-1)}\right)^T}{\left\|x^{(k)}-x^{(k-1)}\right\|_2^2}$$

Final form of Broyden's quasi-Newton method for solving $F(\mathbf{x}) = 0$:

$$\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} + \Delta\mathbf{x}^{(k)} \quad , \quad \Delta\mathbf{x}^{(k)} := -\mathbf{J}_k^{-1} F(\mathbf{x}^{(k)}),$$
$$\mathbf{J}_{k+1} := \mathbf{J}_k + \frac{F(\mathbf{x}^{(k+1)})(\Delta\mathbf{x}^{(k)})^\top}{\left\|\Delta\mathbf{x}^{(k)}\right\|_2^2} . \qquad (8.4.66)$$

Note: Can use Sherman-Morrison-Woodbury formula to calculate J_k^{-1} from J_{k-1}^{-1}.

Remark: In general, iterative methods for nonlinear systems should have convergence monitor [i.e. simple check at each iteration whether convergence to be expected or not.

Example: NMT for damped Newton:

if repeated failure: stop & report error