Numerical Methods for Computational Science and Engineering

Autumn Semester 2018, Week 12

Prof. Rima Alaifari, SAM, ETH Zurich

Stiff IVPs

The problem $\dot{y} = My$ will probably be stiff if

$$\min\{\operatorname{Re}\lambda : \lambda \in \sigma(M)\} \ll 0$$

set of eigenvalues of M

Nonlinear IVPs

$$\dot{y} = f(y)$$

will probably be stiff if for substantial periods of time

$$\min\{\operatorname{Re}\lambda : \lambda \in \sigma(Df(y(t)))\} \ll 0$$

where $t \mapsto y(t)$ is the solution trajectory.

Df ... Jacobian

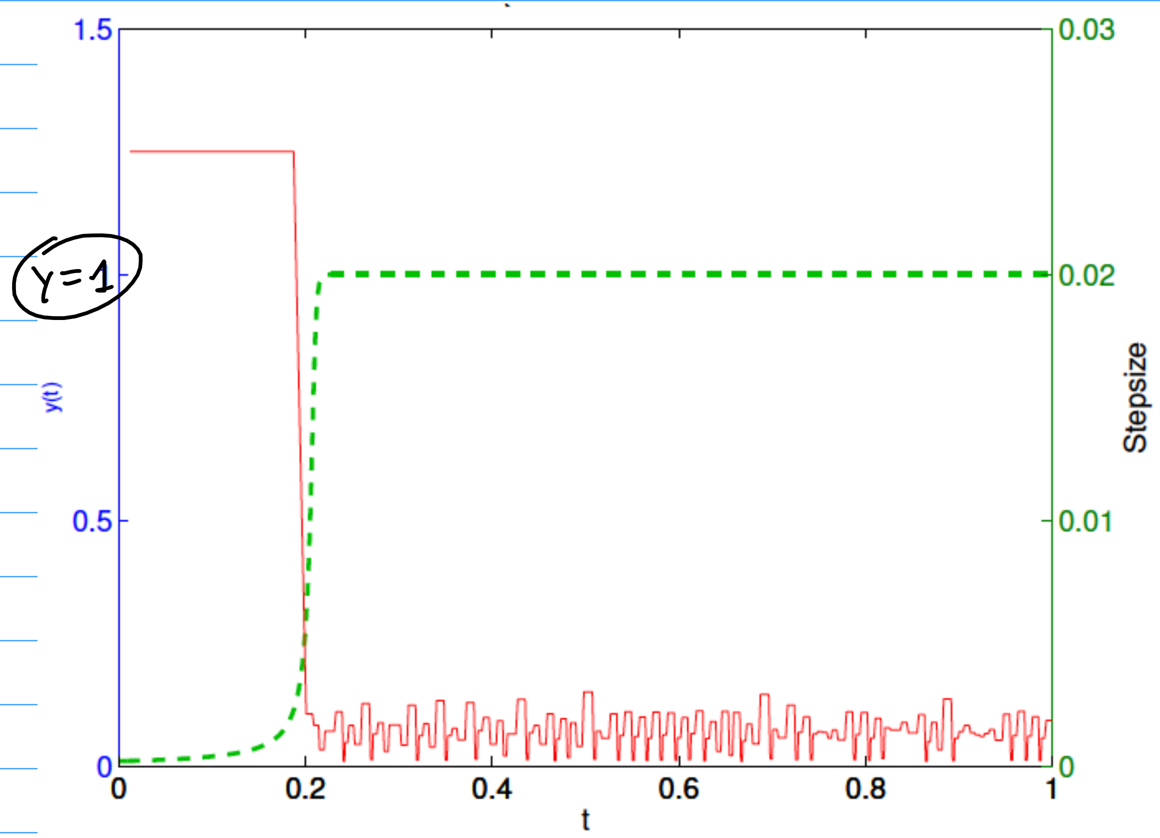
Example 1: logistic ODE

$$\dot{y} = f(y) = \lambda y^2(1-y) \qquad \lambda = 500$$

$$y(0) = \frac{1}{100}$$

$$f'(y) = \lambda\, 2y(1-y) - \lambda y^2 = 2\lambda y - 3\lambda y^2$$
$$= \lambda(2y - 3y^2)$$

behavior of ODE close to stationary state $y=1$



$$f'(1) = -\lambda$$

If $\lambda \gg 0$: stiff IVP close to the stationary state

As we reach $y=1$, we encounter stiffness!

<u>Example 2</u>:

$$\dot{\underline{y}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \underline{y} + \lambda\left(1 - \|\underline{y}\|^2\right)\underline{y} \qquad (*)$$

$$, \quad \|\underline{y}_0\|_2 = 1$$

If $\|\underline{y}_0\|_2 = 1$, we can find φ s.t. $\underline{y}_0 = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$

Then, $y(t) = \begin{pmatrix}\cos(t+\varphi)\\ \sin(t+\varphi)\end{pmatrix}$ is the unique solution

Check: $$\begin{pmatrix}-\sin(t+\varphi)\\ \cos(t+\varphi)\end{pmatrix} \underset{\substack{\uparrow\\ \|y(t)\|_2=1}}{=} \begin{pmatrix}-\sin(t+\varphi)\\ \cos(t+\varphi)\end{pmatrix} + \lambda\cdot \underline{0}$$

$\Rightarrow$ The solution of $(*)$ satisfies $\|y(t)\|_2 = 1$ at any time t.

$$Df(y) \underset{\substack{\uparrow\\ \|y\|_2=1}}{=} \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} - 2\lambda y y^T =$$

$$= \underbrace{\begin{pmatrix}-2\lambda y_1^2 & -1-2\lambda y_1 y_2\\ 1-2\lambda y_1 y_2 & -2\lambda y_2^2\end{pmatrix}}_{=:A}$$

Eigenvalues of A:

$$\det(\mu I - A) = \det\begin{pmatrix}\mu+2\lambda y_1^2 & 1+2\lambda y_1 y_2\\ -1+2\lambda y_1 y_2 & \mu+2\lambda y_2^2\end{pmatrix}$$

$$= \mu^2 + 2\lambda\mu + \cancel{4\lambda^2 y_1^2 y_2^2} + 1 - \cancel{4\lambda^2 y_1^2 y_2^2}$$

$$= \mu^2 + 2\lambda\mu + 1 = 0$$

$$\mu_{1/2} = -\lambda \pm \sqrt{\lambda^2 - 1}$$

For large λ, $\lambda \gg 0$: $\mu_1 \approx 0$, $\mu_2 \ll 0$

$\rightsquigarrow$ typical for stiff IVPs

Applications involving chemical reactions:
often have large negative eigenvalue, and an eigenvalue close to zero

$\rightsquigarrow$ "large stiffness ratio"
$\rightarrow$ fast transients in the solution

8. Iterative Methods for Nonlinear Systems of Equations

So far: direct methods for solving linear systems of equations

Many models in applications involve nonlinear systems of equations
$\rightsquigarrow$ typically these systems can't be solved directly nor exactly!

This leads to iterative methods for finding approximations to the solution

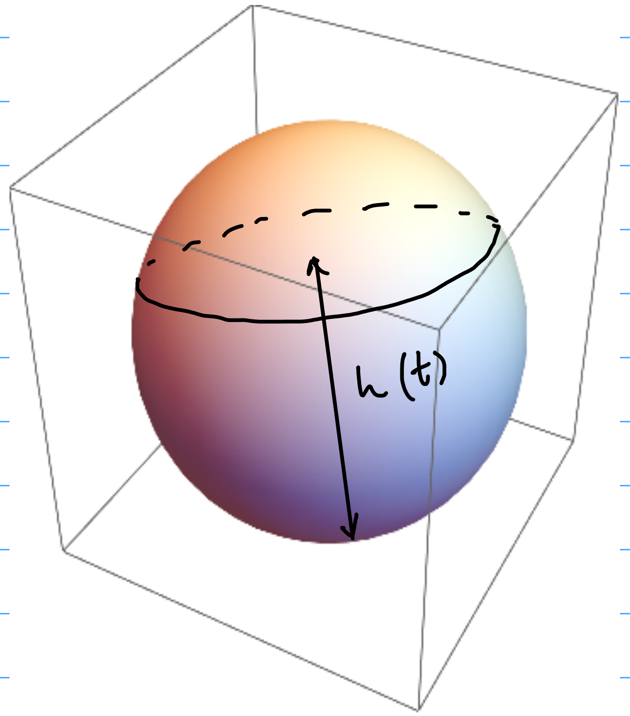
Example: liquid in a spherical tank

r ... radius of the tank

g ... rate of constant fluid flow

Full tank at $t=0$: $h_0 = 2r$

Task: Find height h of the fluid in the tank at any time t.



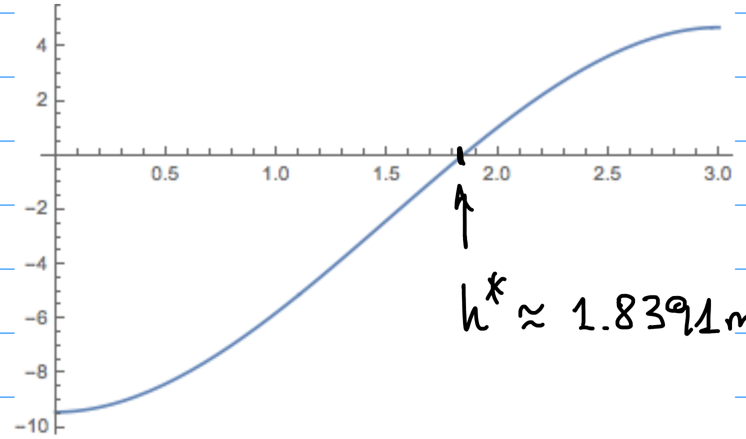
$$-\frac{1}{3}\pi h^3 + \pi r h^2 + \left(g t - \frac{4}{3}\pi r^3\right) = 0$$

Define $f_{t^*}(h) = -\frac{1}{3}\pi h^3 + \pi r h^2 + g t^* - \frac{4}{3}\pi r^3$

Find root h^* i.e. solve for $f_{t^*}(h) = 0$.

$$t^* = \frac{1}{3} \cdot \underbrace{\frac{\frac{4}{3}\pi r^3}{g}}_{= t_{end}} \qquad (r = 1.5\,m)$$

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



$h^* \approx 1.8391\,m$

$(h_0 = 3\,m)$

Typical examples of nonlinear equations:

- thermodynamic models [involve equations of state for real gases]
- Colebrook equation for the friction factor [= pressure drop in oil/gas pipeline]
- Time-stepping in implicit RK methods

One question before that:

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

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

$\rightsquigarrow$ does not have roots

First criterion: intermediate value theorem

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

$$f(t_\ell) < u < f(t_r)$$

$\Rightarrow$ Then there exists $z \in (t_\ell, t_r)$ s.t. $f(z) = u$.

Use this for root finding of $f \in C^0([a,b], \mathbb{R})$

If for some $t_\ell, t_r \in [a,b]$: $f(t_\ell) < 0$ and $f(t_r) > 0$

$\Rightarrow$ f must have a root in (t_ℓ, t_r)

This leads to idea for first root finding algorithm:

Bisection algorithm: (for finding root x^*)

While $n \leq nmax$

Compute $b \leftarrow \frac{t_\ell + t_r}{2}$

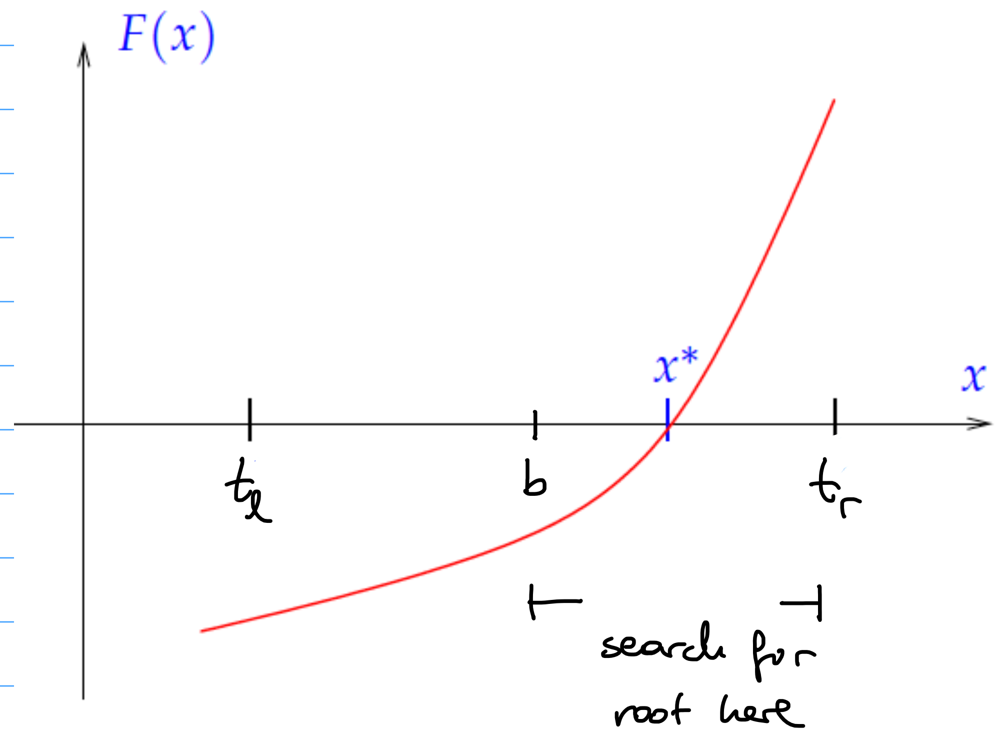
If $|f(b)| < TOL1$ or $|t_\ell - t_r| < TOL2$,

return $x^* = b$

$n \leftarrow n+1$

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

Else $t_r \leftarrow b$



If f is continuous: we always search in a region in which a root exists

At each step: size of interval is halved

$\rightarrow$ convergence

Example [spherical tank]

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

Check: $f_{t^*}(0) < 0$

$$f_{t^*}(2r) = -\frac{1}{3} 8 r^3 \pi + 4 r^3 \pi - \frac{8}{9} r^3 \pi > 0$$

$\rightsquigarrow$ apply bisection starting with $t_\ell = 0$ and $t_r = 2r$.

For iterates $(x^{(k)})_{k\in\mathbb{N}}$ of approximate solution

define $e^{(k)} = x^{(k)} - x^*$ [iteration error]

root/solution

We can ask: $x^{(k)} \to x^*$ at what rate?

$|e^{(1)}| \leq \frac{1}{2} |a-b|$ starting interval $[a,b]$

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

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

error is reduced by a fixed factor (here: $\frac{1}{2}$) at each iteration step: "linear-type convergence"

<u>Bisection:</u> ⊕ robustness & global convergence

⊖ only linear-type convergence

no extension to higher dimensions

Goal: Methods that • extend to higher dimensions

• converge faster (quadratically)

under additional assumptions

Fixed Point Iterations:

Root finding: find x^* s.t. $f(x^*) = 0$

$\Leftrightarrow$ Define $\Phi(x) := f(x) + x$

x^* is a FP of Φ: $\Phi(x^*) = x^*$

Before: bisection only required continuity of f

Now: Assume

• Φ is Lipschitz continuous on $[a,b]$:

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

• $L < 1$ i.e. Φ is a contractive mapping

Idea of FPI (fixed point iteration):

• start with initial guess $x^{(0)}$

• Iterate $x^{(k)} = \Phi(x^{(k-1)})$ FPI

$L < 1$ guarantees: If Φ has a fixed point then FPI will converge to it.

Why ?

Can we verify $|e^{(k)}| \to 0$? as $k \to \infty$

$$|e^{(k)}| = |x^{(k)} - x^*| \underset{\substack{\uparrow \\ \text{FPI}}}{=} |\Phi(x^{(k-1)}) - x^*|$$

$$\underset{\substack{\uparrow \\ \Phi(x^*)=x^* \\ \text{fixed point}}}{=} |\Phi(x^{(k-1)}) - \Phi(x^*)| \underset{\substack{\uparrow \\ \text{Lipschitz} \\ \text{const.}}}{\leq} L \cdot |x^{(k-1)} - x^*|$$

$$\leq L \cdot |e^{(k-1)}|$$

$$\Rightarrow \quad |e^{(k)}| \leq L^k \cdot |e^{(0)}| \longrightarrow 0 \quad \text{as } k \to \infty$$

provided $\boxed{L < 1}$

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 $|\Phi'(x^*)| < 1$

We get: Convergence of FPI in a neighborhood of x^*

There are $\delta > 0$, $\varepsilon > 0$ s.t. $|\Phi'(x)| < 1-\varepsilon$ for $x \in [x^*-\delta, x^*+\delta] =: I^*$

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

$$|\Phi(x) - \Phi(y)| = |\Phi'(\theta)| \cdot |x-y| \qquad \text{[mean value thm]}$$

$$< (1-\varepsilon)|x-y|$$

$\Rightarrow$ Lipschitz const. on I^* with $L = 1-\varepsilon$.

$\rightsquigarrow$ If Φ is only locally contractive then FPI is also only locally convergent

- Convergence rate is (at least) linear:

$$\exists\ 0<L<1 \qquad |x^{(k+1)} - x^*| \leq L \cdot |x^{(k)} - x^*|$$

$$|e^{(k+1)}| \leq L\,|e^{(k)}|$$

[Note: bisection method

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

"linear-type" because $|e^{(k)}| > L\,|e^{(k-1)}|$ is possible]

Example:

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

search for x^* s.t. $F(x^*) \stackrel{!}{=} 0$

$\Leftrightarrow$ $\quad \Phi(x) :=$

search for x^* s.t. $\Phi(x^*) = x^*$

$$\Phi(x) = xe^x - 1 + x$$

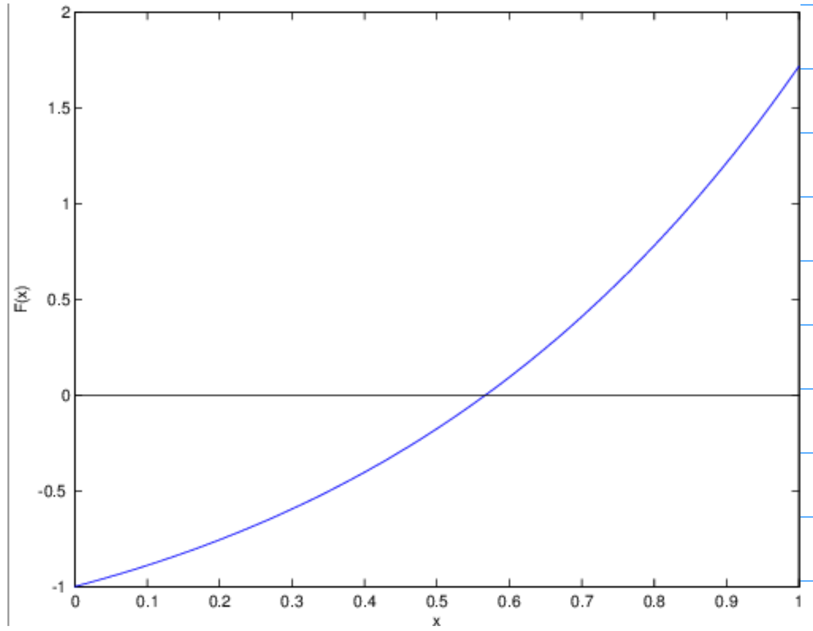
$$\tilde{\Phi}(x) = c \cdot (xe^x - 1) + x$$

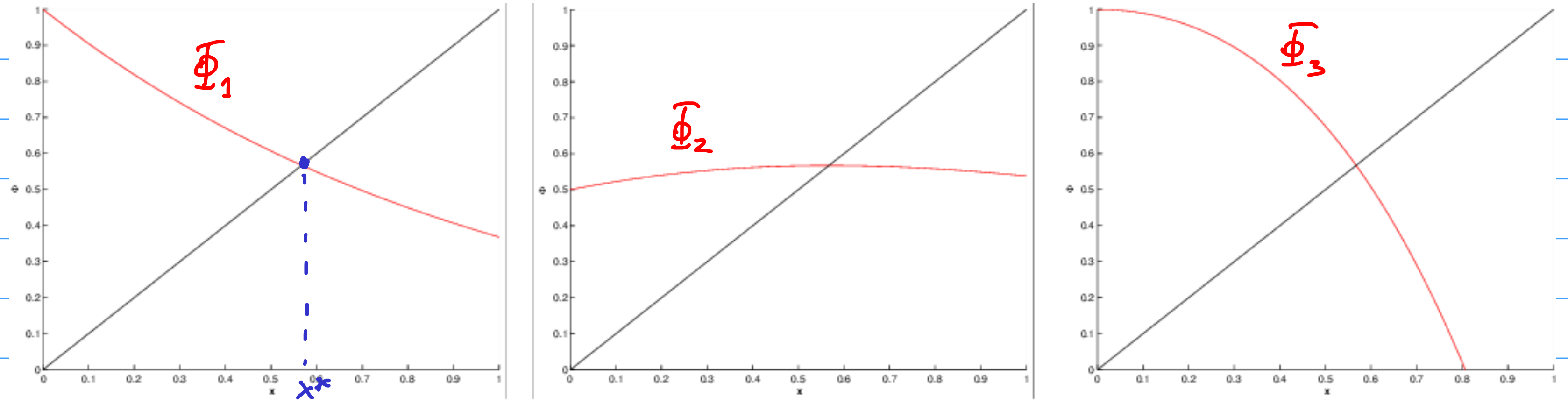
$$\tilde{\tilde{\Phi}}(x) = x - xe^x + 1$$

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

Different fixed point forms:

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





function Φ_1 function Φ_2 function Φ_3

(1)
$$xe^x - 1 = 0$$
$$xe^x = 1$$
$$x = e^{-x}$$

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

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

$\Phi_1(x^{(k-1)}) = x^{(k)}$ $\quad \Phi_2(x^{(k-1)}) = x^{(k)}$ $\quad \Phi_3(x^{(k-1)}) = x^{(k)}$

k	$\lvert x_1^{(k+1)} - x^*\rvert$	$\lvert x_2^{(k+1)} - x^*\rvert$	$\lvert x_3^{(k+1)} - x^*\rvert$
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

$\lvert e^{(k+1)} \rvert$

- Derive largest region of convergence for
$$\Phi_1(x^{(k)}) = x^{(k+1)}$$
$$\Phi_2(x^{(k)}) = x^{(k+1)}$$

- Why is
$$\Phi_3(x^{(k)}) = x^{(k+1)}$$
not converging

1.) $\Phi_1(x) = -e^{-x}$

$$|\Phi_1'(x)| = e^{-x} < 1 \quad \text{for} \quad x \in [\delta, 1] = I^*$$
$$\forall \delta > 0$$

$\Rightarrow$ convergence of FPI with Φ_1 on I^*.

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

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

$\Rightarrow$ convergence of FPI with Φ_2 on $[0, 1]$

$\rightsquigarrow$ global convergence

3.) $$\Phi_3'(x) = 1 - xe^x - e^x$$

Least : $|\Phi_3'(x^*)| < 1$???

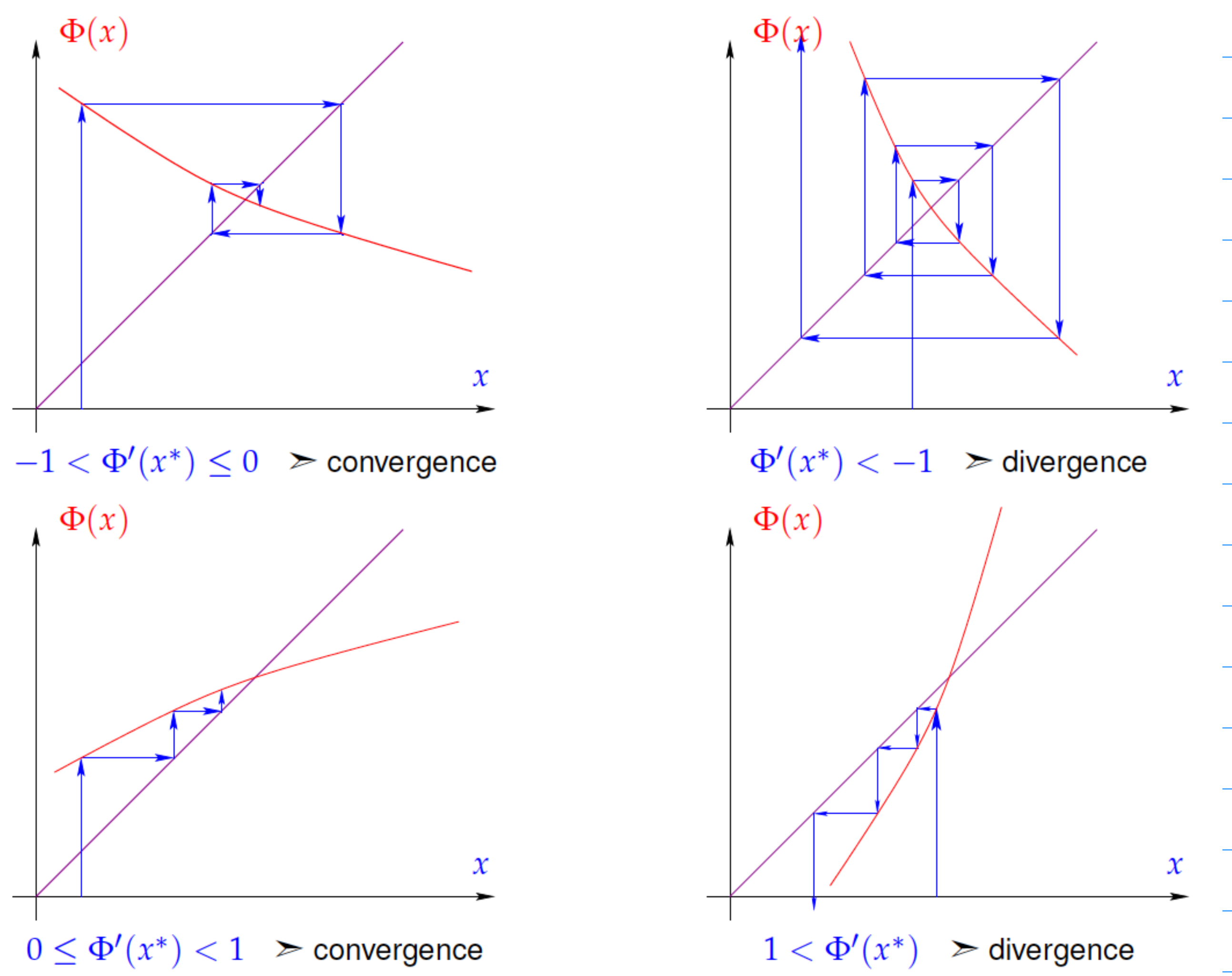
We know: $x^* = e^{-x^*}$

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

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

$\rightsquigarrow$ No region around x^* for which FPI with Φ_3 is contractive

$\rightsquigarrow$ no convergence



Why is FPI with Φ_2 converging faster than the one with Φ_1?

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

For x^* FP: $\quad 1 - x^* e^{x^*} = 0$

$$\Rightarrow \Phi_2'(x^*) = 0.$$

In general, if $\Phi'(x^*) = 0$ and Φ is C^2 in a neighborhood of x^*:

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

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

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

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

Taylor exp.

If $|x^{(k-1)} - x^*| \leq \delta < 1$ [holds if we have convergence] :

$$|e^{(k)}| \leq \frac{1}{2} |e^{(k-1)}|^2 \cdot \left[\Phi''(x^*) + C\delta\right]$$

$\Rightarrow$ We can expect faster convergence if $\Phi'(x^*) = 0$.

"quadratic" convergence

If $\exists\ C > 0$ s.t.

$$|e^{(k)}| \leq C \cdot |e^{(k-1)}|^p \qquad \forall k \in \mathbb{N}_0$$

then we speak of convergence with order p.

Can we find an algorithm for root-finding with quadratic convergence?

Assumption: $f \in C^1$

Intuition:

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

1st order Taylor approximation around $x^{(k)}$

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

s.t.

$$f(x^{(k)}) + (x^{(k+1)} - x^{(k)}) f'(x^{(k)}) = 0$$

Newton iteration:

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

When $f \in C^2$ $\rightsquigarrow$ quadratic convergence

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

Newton's method on f $\Longleftrightarrow$ FPI on Φ

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

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

If $\underline{f'(x^*) \neq 0}$ then $\Phi'(x^*) = \dfrac{f''(x^*) \overbrace{f(x^*)}^{=0}}{(f'(x^*))^2}$

$= 0$

$\Rightarrow$ convergence in a neighborhood of x^*

$|\Phi'| < 1$

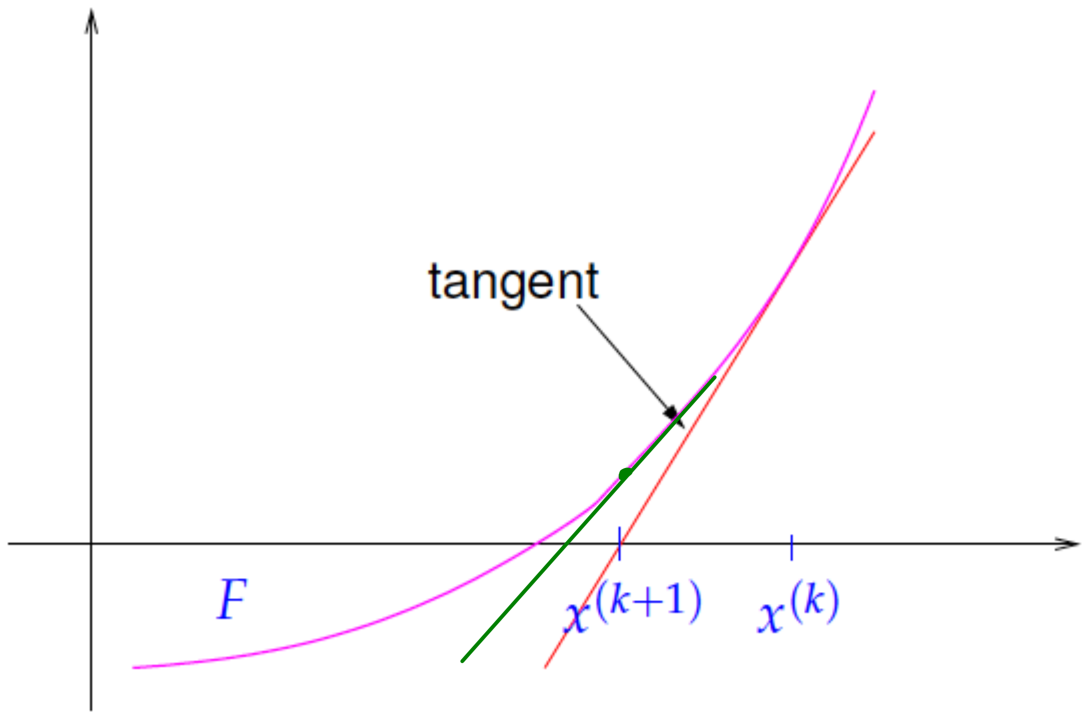
$\Rightarrow$ quadr. conv.

If I^* is a neighborhood of x^* s.t.

$\forall x \in I^* \quad |\Phi'(x)| \leq 1-\varepsilon$

- $f \in C^2(I^*)$.

For $x^{(0)} \in I^*$ Newton's method converges quadratically if $f'(x^*) \neq 0$.



Note: • Contractivity of Φ on I^* guarantees $f'(x^{(k)}) \neq 0$ for all iterates [if started with $x^{(0)} \in I^*$]

- For quadr. conv. of Newton, $f \in C^2$ instead of $\Phi \in C^2$ suffices.

Altogether: We need I^* neighborhood of x^* s.t.

- I^* suff. small
- $f'(x) \neq 0$ on I^*

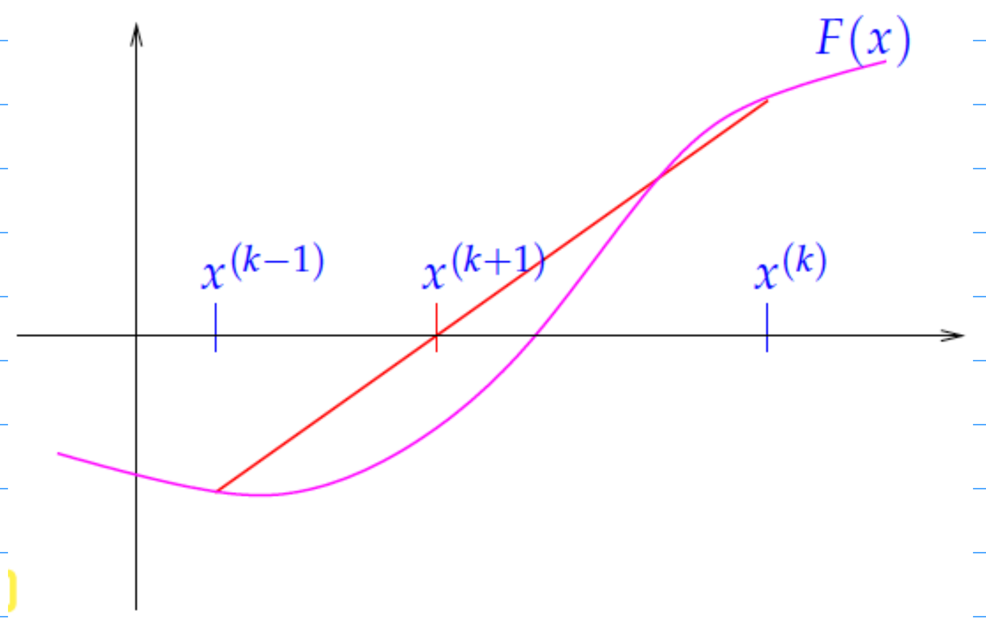
Newton's method requires the computation of $f'(x^{(k)})$ at each iteration step k. (can be costly)

Alternative: Secant method:

Instead of $f'(x^{(k)})$ use

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

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



Convergence of secant method:

As for Newton:
- local
- need $f'(x^*) \neq 0$
- f locally C^2

BUT: rate is now of order $p = \frac{1+\sqrt{5}}{2} \in (1,2)$

$\rightsquigarrow$ superlinear but not quadr.

Remark: Secant method is a 2-point method (uses the last 2 iterates at each step)

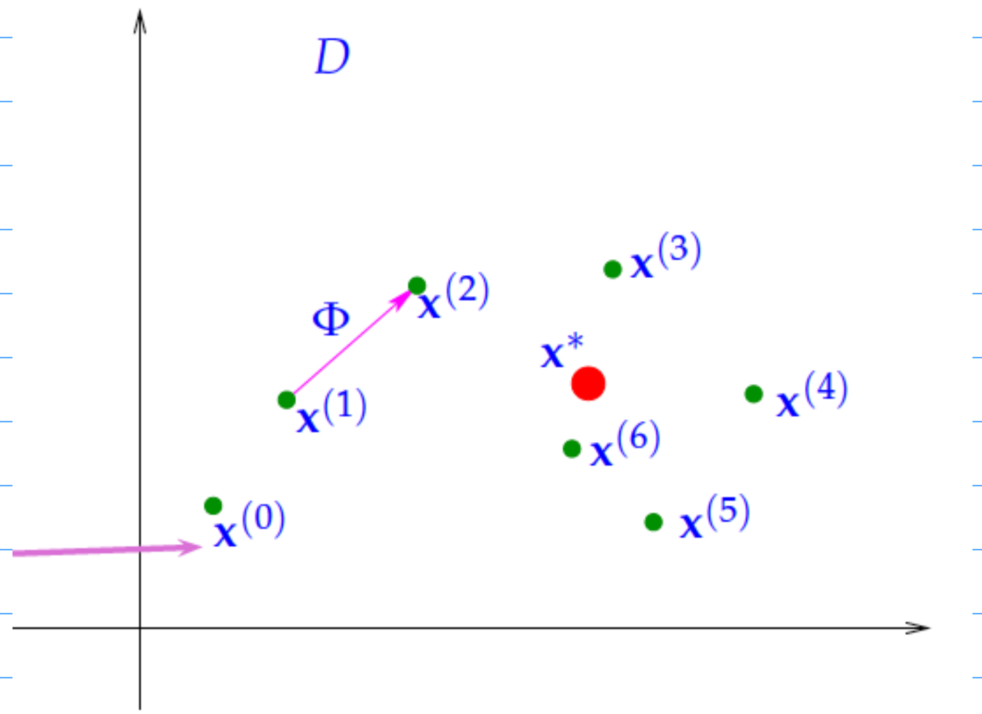
Nonlinear systems of equations

$F : D \subset \mathbb{R}^n \to \mathbb{R}^n$

nonlinear system of equations

n equations, n unknowns

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



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

m-point method

Aspects of iterative methods:

- Convergence : $\lim_{k \to \infty} x^{(k)} = x^*$?
- Consistency : $\Phi_F(x^*, \ldots, x^*) = x^*$
  $\Leftrightarrow F(x^*) = 0$?
- Rate of convergence:
  $\| x^{(k)} - x^* \| \to 0$ with which order?

Note: $\| \cdot \|$ can be any norm for $\mathbb{R}^n$

finite dim. vector space

$\Rightarrow$ All norms for $\mathbb{R}^n$ are equivalent

$$\left(\exists\, c_1, c_2 > 0: \quad c_1 \|v\|_a \le \|v\|_b \le c_2 \|v\|_a \quad \forall\, v \in V \right)$$

$\Rightarrow$ Convergence in $\mathbb{R}^n$ is a property independent of choice of norm

But: convergence rate is not independent of the choice of norm.

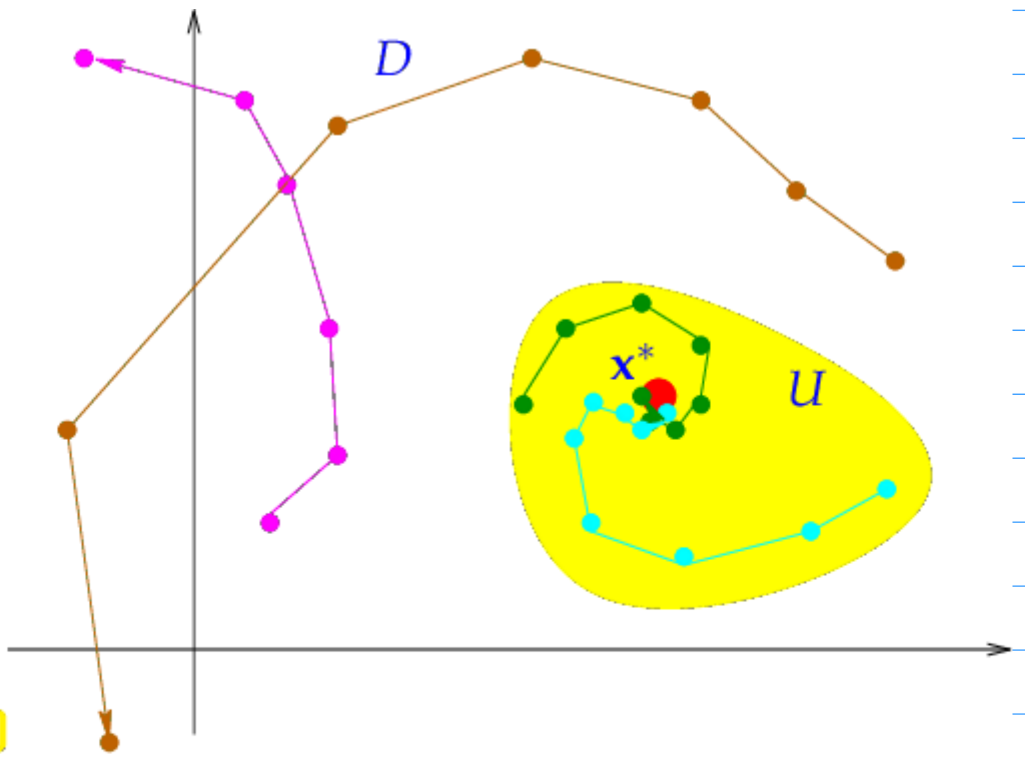


Fig. 283

An m-point iterative method converges <u>locally</u> to $x^* \in \mathbb{R}^n$ if there is a neighborhood $U \subset D$ of x^* s.t.

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

If $U = D$: the method is <u>globally</u> convergent.