NUMERICAL METHODS IN CIVIL ENGINEERING

LECTURE NOTES

Janusz ORKISZ

2007-09-09

l.888 Numerical Methods in Civil Engineering I

Introduction, errors in numerical analysis. Solution of nonlinear algebraic equations Solution of large systems of linear algebraic equations by direct and iterative methods. Introduction to matrix eigenvalue problems. Examples are drawn from structural mechanics.

Prep. Admission to Graduate School of Engineering.

l.889 Numerical Methods in Civil Engineering II

Continuation of l.888. Approximation of functions: interpolation, and least squares curve fitting; orthogonal polynomials. Numerical differentiation and integration. Solution of ordinary and partial differential equations, and integral equations; discrete methods of solution of initial and boundary-value problems. Examples are drawn from structural mechanics, geotechnical engineering, hydrology and hydraulics.

Prep. l.888, Numerical Methods in Civil Engineering I.

Table of contents

9. Numerical differentation

9.1. By means of the approximation and differentation
9.2. Generation of numerical derivatives by undetermined coefficients method

10. Numerical integration

10.1. Introduction
10.2. Newton – Cotes formulas
  10.2.1. Composite rules
10.3. Gaussian quadrature
  10.3.1. Composite Gaussian – Legendre integration
  10.3.2. Composite Gaussian – Legendre integration
  10.3.3. Summary of the Gaussian integration
  10.3.4. Special topics

11. Numerical solution of ordinary differential equations

11.1. Introduction
11.2. Classification
11.3. Numerical approach
11.4. The Euler Method
11.5. Runge Kutta method
11.6. Multistep formulas
  11.6.1. Open (Explicit) (Adams – Bashforth) formulas
  11.6.2. Closed (Implicit) formulas (Adams – Moulton)
  11.6.3. Predictor – corrector method

12. Boundary value problems

12.1. Finite difference solution approach

13. On solution of boundary value problems for partial differential equations by the finite difference approach (FDM)

13.1. Formulation
13.2. Classification of the second order problems
13.3. Solution approach for solution of elliptic equations by FDM

14. Parabolic equations

15. Hyperbolic equations

16. MFDM

16.1. MWLS Approximation

1. INTRODUCTION

1.1. NUMERICAL METHOD

– any method that uses only four basic arithmetic operations : + , – , : , *
– theory and art.

$$\boxed{x = \sqrt{a}} \quad \rightarrow \quad x^2 = a\,, \qquad a \geq 0$$

$$x = \frac{a}{x} \qquad , \; x \neq 0$$

$$x + x = x + \frac{a}{x}$$

$$x = \frac{1}{2}\left(x + \frac{a}{x} \right)$$

– numerical method

$$\boxed{x_n = \frac{1}{2}\left(x_{n-1} + \frac{a}{x_{n-1}} \right)}$$

1.2. ERRORS IN NUMERICAL COMPUTATION

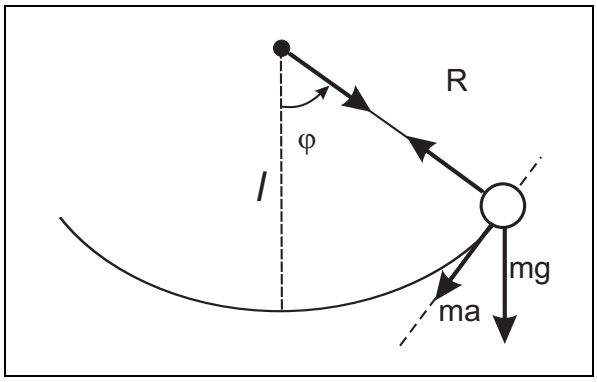
Types of errors :

Inevitable error

(i) Error arising from the inadequacy of the mathematical model

Example :

Pendulum



$a = l\dfrac{d^2\varphi}{dt^2}$ - acceleration

$\dfrac{d^2\varphi}{dt^2} + \alpha\left(\dfrac{d\varphi}{dt}\right)^n + \dfrac{g}{l}\sin\varphi = 0$ - nonlinear model including large displacements and friction

$\dfrac{d^2\varphi}{dt^2} + \dfrac{g}{l}\varphi = 0$ - simplified model - small displacement, linearized equation and no friction

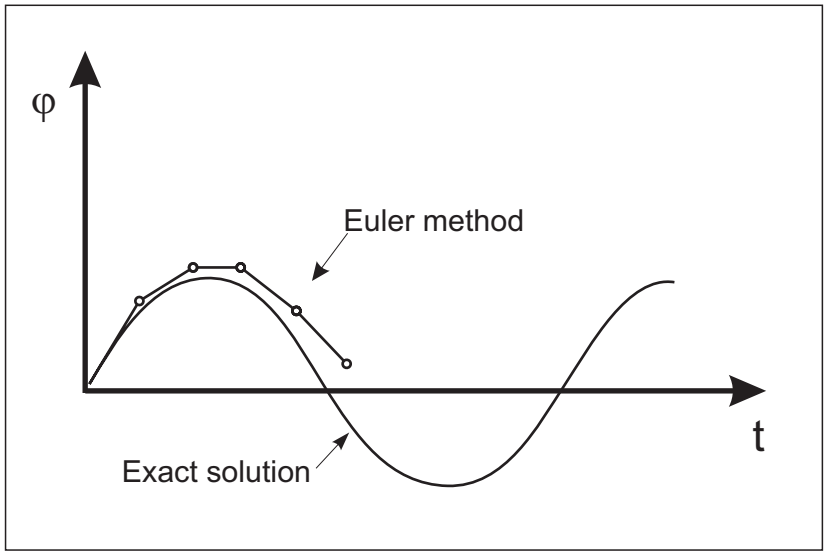
(ii) Error noise in the input data

l, g, $\varphi(0)$, $\dfrac{d\varphi}{dt}\Big|_{t=0}$, ...

Error of a solution method

Example:

$$\frac{d\varphi}{dt} = f(t, \varphi(t))$$



Numerical errors

(iii) Errors due to series truncation

Example

The temperature $u(x,t)$ in a thin bar:

$$u(x,t) = \sum_{n=1}^{\infty} C_n \exp\frac{\left(-n^2\pi^2 t\right)}{l^2} \sin\frac{n\pi x}{l} = \sum_{n=1}^{10} + \cancel{\sum_{n=11}^{\infty}} \approx \sum_{n=1}^{10}$$

(iv) Round off error

Example

$$x = \frac{2}{3} = 0.667$$

THE OTHER CLASSIFICATION OF ERRORS

(i) The absolute error

Let x - exact value, $\tilde{x}$ - approximate value

$$\varepsilon = |\tilde{x} - x|$$

(ii) The relative error

$$\delta = \left|\frac{\tilde{x} - x}{x}\right|$$

PRESENTATION OF RESULTS

$$x_{\text{expected}} = \tilde{x} \pm \varepsilon = \tilde{x}(1 \pm \delta)$$

Example

$$x_{\text{expected}} = 2.53 \pm 0.10 \approx 2.53(1 \pm 0.04) = 2.53 \pm 4\%$$

1.3. SIGNIFICANT DIGITS

Number of digits starting from the first non-zero on the left side

Example

	Number of significant digits		Number of significant digits
2345000	7	5	1
2.345000	7	5.0	2
0.023450	5	5.000	4
0.02345	4		

Example

Subtraction	Number of significant digits
2.3485302	8
-2.3485280	8
0.0000022	2

1.4. NUMBER REPRESENTATION

FIXED POINT		FLOATING POINT			
324.2500	: 1000	324.2500	=	3.2425×10^{2}	: 10^{3}
.3242	: 100			3.2425×10^{-1}	: 10^{2}
.0032	: 10			3.2425×10^{-3}	: 10
.0003				3.2425×10^{-4}	

1.5. ERROR BOUNDS

(iii) Summation and subtraction

Given: $a \pm \Delta a, \quad b \pm \Delta b$

Searched: $x = a + b = a \pm \Delta a + b \pm \Delta b$

error evaluation

$$|\Delta x| = |x - a - b| \leq |\Delta a| + |\Delta b|$$

(iv) Multiplication and division

$$x = \frac{ab}{cf} \rightarrow \ln x = \ln a + \ln b - \ln c - \ln f$$

$$\frac{dx}{x} = \frac{da}{a} + \frac{db}{b} - \frac{dc}{c} - \frac{df}{f}$$

error evaluation

$$|\Delta x| \leq |x| \left(\left|\frac{\Delta a}{a}\right| + \left|\frac{\Delta b}{b}\right| + \left|\frac{\Delta c}{c}\right| + \left|\frac{\Delta f}{f}\right| \right)$$

1.6. CONVERGENCE

Example

$$x_n = \frac{1}{2}\left(x_{n-1} + \frac{a}{x_{n-1}} \right), \qquad \lim_{n\to\infty} x_n = ?$$

let

$$\delta_n = \frac{x_n - x}{x} \rightarrow x_n = x\left(1 + \delta_n\right)$$

$$x\left(1+\delta_n\right) = \frac{1}{2}\left[x\left(1+\delta_{n-1}\right) + \frac{a}{x\left(1+\delta_{n-1}\right)} \right] \qquad \Big| \cdot \frac{x}{a}$$

$$1+\delta_n = \frac{1}{2}\left[1+\delta_{n-1} + \frac{1}{1+\delta_{n-1}} \right] = \frac{1}{2}\left[1+\delta_{n-1} + \frac{1+\delta_{n-1}-\delta_{n-1}}{1+\delta_{n-1}} \right] =$$

$$= \frac{1}{2}\left(2 + \frac{\delta_{n-1}^2}{1+\delta_{n-1}} \right)$$

for

$$x_0 = a \rightarrow \delta_0 > 0 \rightarrow \delta_{n-1} > 0 \rightarrow \frac{\delta_{n-1}}{1+\delta_{n-1}} < 1$$

one obtaines

$$\delta_n = \frac{\delta_{n-1}^2}{2\left(1+\delta_{n-1}\right)} = \frac{1}{2}\delta_{n-1}\left(\frac{\delta_{n-1}}{1+\delta_{n-1}}\right) < \frac{1}{2}\delta_{n-1}$$

$$\delta_n < \frac{1}{2}\delta_{n-1} \rightarrow \text{iteration is } \textit{convergent}$$

$$\lim_{n\to\infty} \delta_n = 0 \rightarrow \lim_{n\to\infty} x_n \rightarrow \sqrt{a}$$

In numerical calculations we deal with a number N. It describes a term that satisfy an admissible error B requirement
where

$$\varepsilon_n < B \quad \text{for} \quad n \geq N, \text{ where}$$

$$\varepsilon_n = \left| \frac{x_n - x_{n-1}}{x_n} \right| = \left| \frac{x\left(1+\delta_n\right) - x\left(1+\delta_{n-1}\right)}{x\left(1+\delta_n\right)} \right| = \left| \frac{\delta_n - \delta_{n-1}}{1+\delta_n} \right|$$

1.7. STABILITY

Solution is stable if it remains bounded despite truncation and round off errors.

Let

$$\tilde{x}_n = x_n\left(1+\gamma_n\right) = \frac{1}{2}\left(\tilde{x}_{n-1} + \frac{a}{\tilde{x}_{n-1}}\right)\left(1+\gamma_n\right) \rightarrow \delta_n = \frac{1}{2}\,\frac{\delta^2{}_{n-1}}{1+\delta_{n-1}}\left(1+\gamma_n\right)+\gamma_n$$

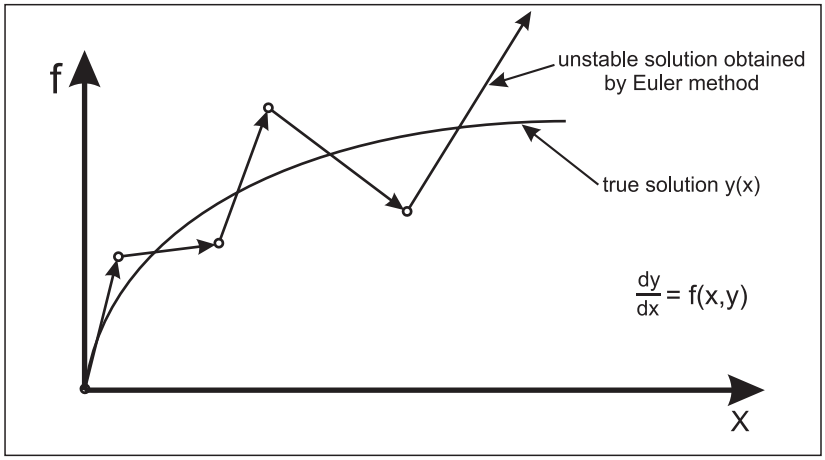
$\lim_{n\to\infty} \delta_n = \gamma_n$ $\rightarrow$ precision of the final result corresponds to the precision of the last step of calculations i.e.

$\tilde{x} \rightarrow x\left(1+\gamma_n\right)$

Example

Unstable calculations

(v) Time integration process



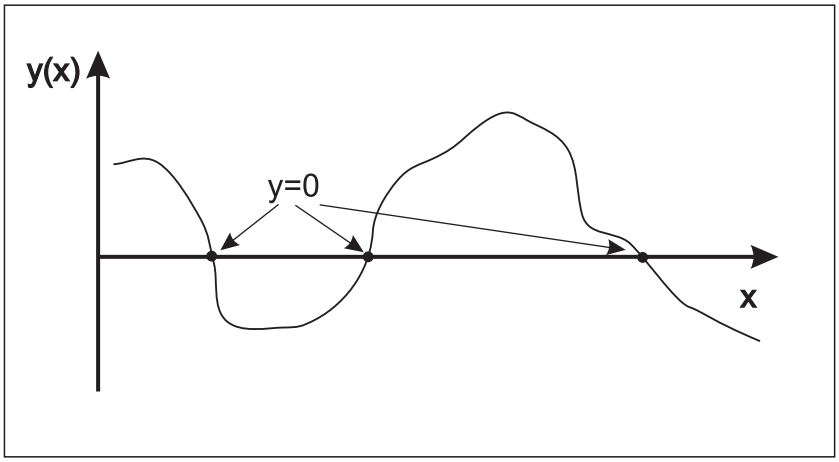
(vi) Ill-conditioned simultaneous algebraic equations

2. SOLUTION OF NON-LINEAR ALGEBRAIC EQUATIONS

2.1. INTRODUCTION

- source of algebraic equations
- multiple roots
- start from sketch
- iteration methods

equation to be solved $y(x) = 0 \rightarrow x = ...$



2.2. THE METHOD OF SIMPLE ITERATIONS

2.2.1. Algorithm

Algorithm

$x = f(x)$

$x_1 = f(x_0)$

$x_2 = f(x_1)$

.....

$x_n = f(x_{n-1})$

.....

Example

Let

$x_n = \frac{1}{2}\left(x_{n-1} + \frac{a}{x_{n-1}}\right)$

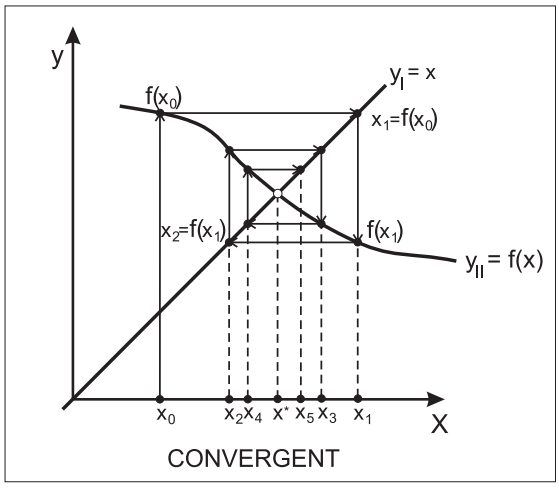
a=2, x_0=2

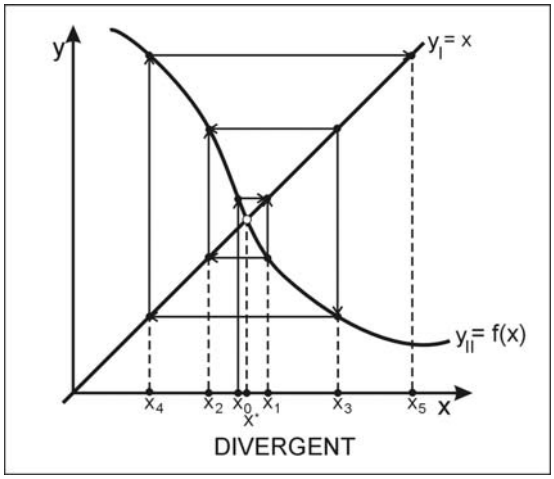
$x_1 = \frac{1}{2}\left(2 + \frac{2}{2}\right) = \frac{3}{2} = 1.5000$

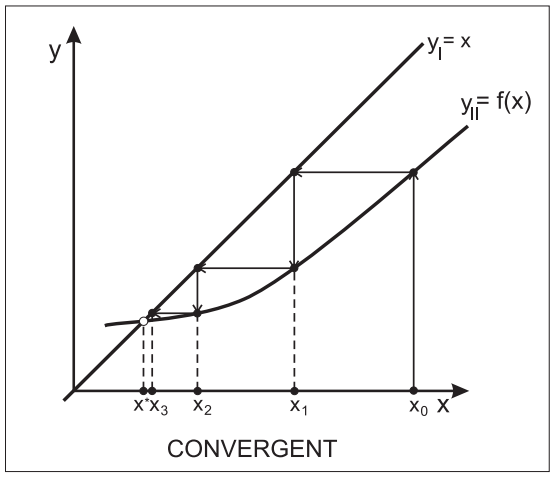
$x_2 = \frac{1}{2}\left(\frac{3}{2} + 2 \cdot \frac{2}{3}\right) = \frac{17}{12} = 1.4167$

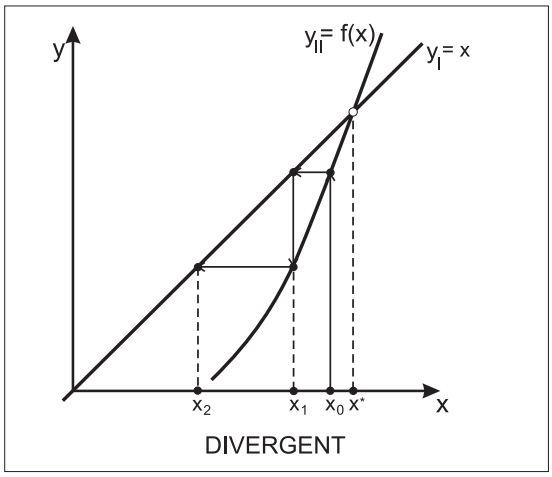
.....

Geometrical interpretation









Example :

$$x^2 - 4x + 2.3 = 0 \quad \rightarrow \quad x = f(x)$$

Algorithm

(i)

$$x = \frac{\left(x^2 + 2.3\right)}{4} \quad \rightarrow \quad x_n = \left(x_{n-1}^2 + 2.3\right)\frac{1}{4}$$

(ii)

$$x = \sqrt{4x - 2.3} \quad \rightarrow \quad x_n = \sqrt{4x_{n-1} - 2.3}$$

Let $x_0 = 0.6$

$$x_1 = \left(.6^2 + 2.3\right)\frac{1}{4} = .665$$

$$x_2 = \left(.665^2 + 2.3\right)\frac{1}{4} = .686$$

…….....

$$x_6 = \left(\underline{.696}^2 + 2.3\right)\frac{1}{4} = \underline{.696}$$

Solution converged within three digits : $\dfrac{x_6 - x_5}{x_6} = \dfrac{0.696 - 0.696}{0.696} = 0$

$$x_1 = 0.316$$

$$x_2 = \sqrt{1.264 - 2.3}$$

cannot be performed

2.2.2. Convergence theorems

Theorem 1

If

$$\left|f\left(x_1\right)-f\left(x_2\right)\right| \le L\left|x_1-x_2\right| \qquad \text{with} \qquad 0<L<1$$
$$\text{for } x_1,\, x_2 \in [a,\, b];$$

then the equation $x = f\left(x\right)$ has at most one root in $[a,\, b]$.

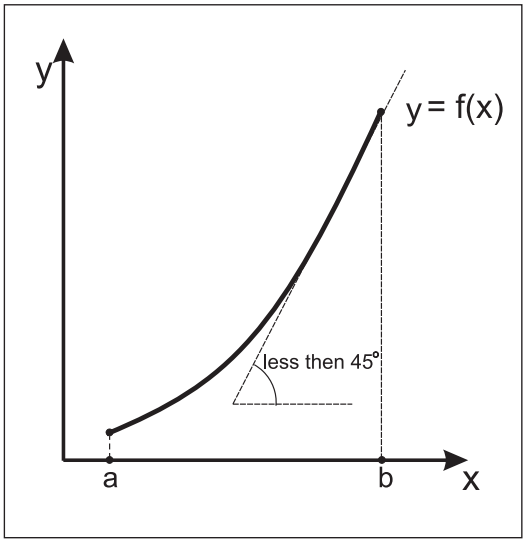
Theorem 2

If $f\left(x\right)$ satisfy conditions of *Theorem 1* then the iterative method

$$x_n = f\left(x_{n-1}\right)$$

converges to the unique solution $x \in [a,\, b]$; of $x = f\left(x\right)$ for any $x_0 \in [a,\, b]$;

Geometrical interpretation



2.2.3. Iterative solution criteria

Convergence

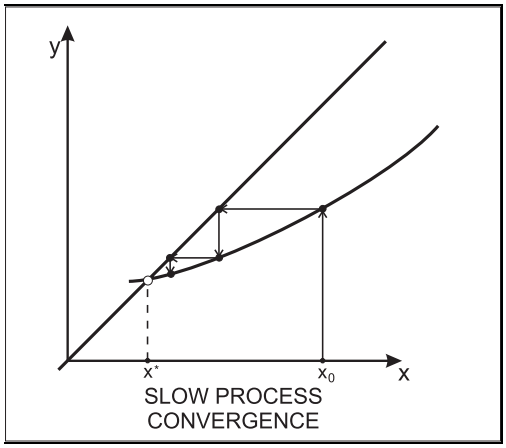
$$\delta_n = \left|\frac{x_n - x_{n-1}}{x_n}\right| < B$$

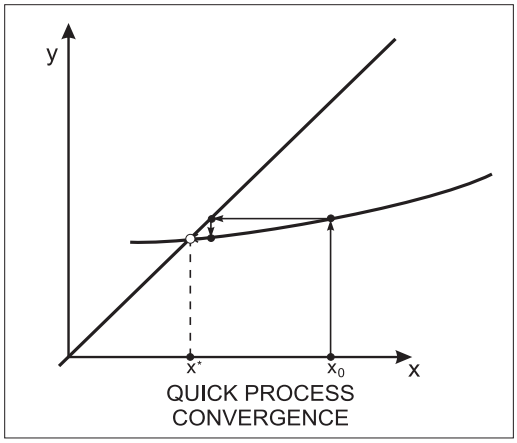
Residuum

$$r_n = \left|\frac{f(x_{n-1}) - x_{n-1}}{f(x_{n-1})}\right| = \left|\frac{x_n - x_{n-1}}{x_n}\right| = \delta_n < B$$

Notice : both criteria are the same for the simple iterations method

2.2.4. Acceleration of convergence by the relaxation technique





$$x = f(x)$$

$$\alpha x + x = \alpha x + f(x) \;\rightarrow\; x = \frac{\alpha}{1+\alpha} x + \frac{1}{1+\alpha} f(x) \equiv g(x)$$

The best situation if $g'(x) \approx 0$

$$g'(x) = \frac{\alpha}{1+\alpha} + \frac{1}{1+\alpha} f'(x)$$

let

$$g'(x^*) = 0 \rightarrow \alpha = -f'(x^*)$$

then

$$g(x) = \frac{1}{1 - f'(x^*)} f(x) - \frac{f'(x^*)}{1 - f'(x^*)} x$$

Example :

$$x^2 = a > 0 \;\rightarrow\; x = \frac{a}{x} \;\rightarrow\; f(x) = \frac{a}{x} \;\rightarrow\; f'(x) = -\frac{a}{x^2} = -1$$

then

$$g(x) = \frac{1}{1-(-1)} \frac{a}{x} - \frac{-1}{1-(-1)} x = \frac{1}{2}\left(x + \frac{a}{x} \right)$$

hence

$$\boxed{x_n = \frac{1}{2}\left(x_{n-1} + \frac{a}{x_{n-1}} \right)}$$

2.3. NEWTON – RAPHSON METHOD

2.3.1. Algorithm

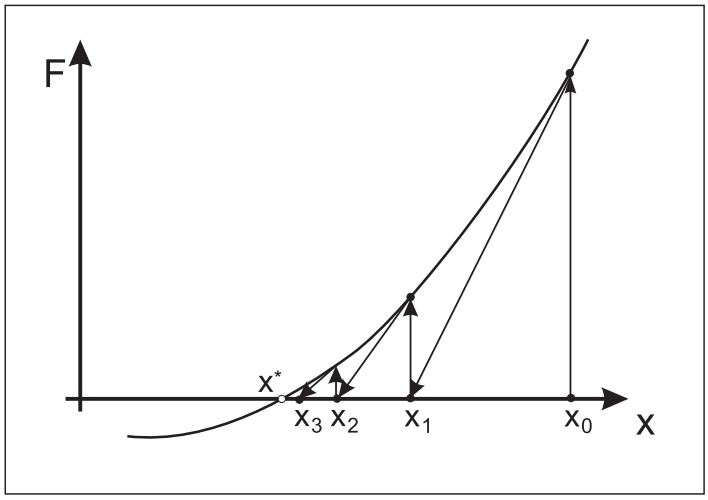
$$F(x) = 0$$

$$F(x+h) = F(x) + \left.\frac{dF}{dx}\right|_x h + \frac{1}{2}\left.\frac{d^2F}{dx^2}\right|_x h^2 + \ldots = F(x) + F'(x)h + R \approx F(x) + F'(x)h = 0$$

$$F(x) + F'(x)h = 0 \;\rightarrow\; h = -\frac{F(x)}{F'(x)}$$

$$\boxed{x_n = x_{n-1} + h \;=\; x_{n-1} - \frac{F(x_{n-1})}{F'(x_{n-1})}}$$

Geometrical interpretation



2.3.2. Convergence criteria

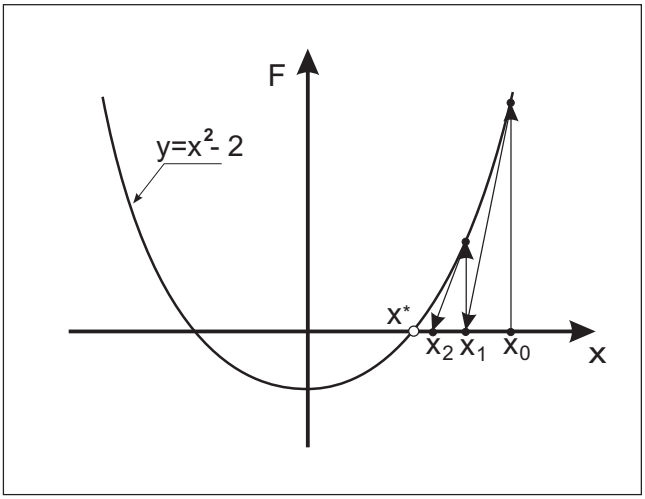
Solution convergence

$$\delta_n = \left|\frac{x_n - x_{n-1}}{x_n}\right| < B_1$$

Residuum

$$r_n = \left|\frac{F(x_n)}{F(x_0)}\right| < B_2, \qquad F(x_0) \neq 0$$

Example



$x^2 = 2 \;\rightarrow\; x^2 - 2 = 0$
$F(x) = x^2 - 2,$
$F'(x) = 2x$

$$\boxed{x_n = x_{n-1} - \frac{x_{n-1}^2 - 2}{2x_{n-1}}}$$

$x_0 = 2$

$$x_1 = 2 - \frac{2^2 - 2}{2 \cdot 2} = \frac{3}{2} = 1.500000$$

$$x_2 = \frac{3}{2} - \frac{\frac{9}{4} - 2}{2 \cdot \frac{3}{2}} = \frac{17}{12} = 1.416667$$

$$x_3 = \frac{577}{408} = 1.414216$$

……...

Convergence

$$\delta_1 = \left| \frac{\frac{3}{2} - 2}{\frac{3}{2}} \right| = 0.333333$$

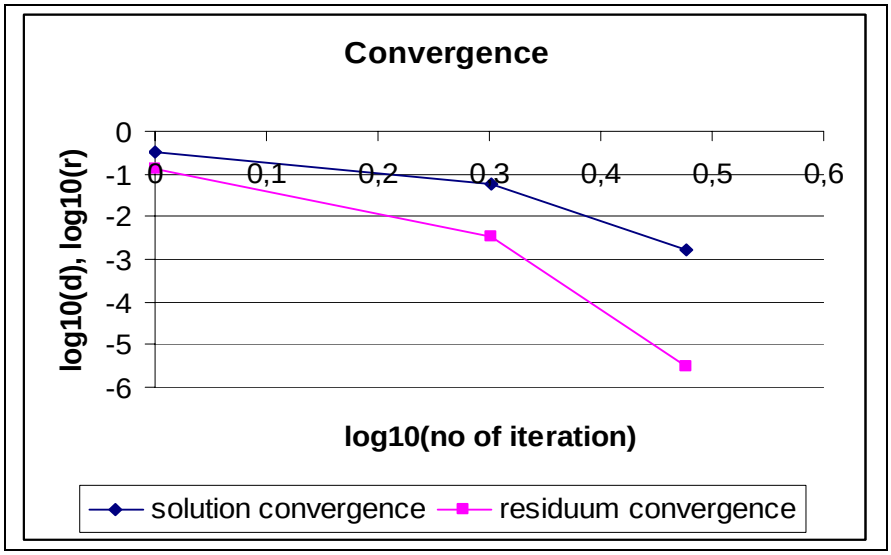
$$r_1 = \left| \frac{\left(\frac{3}{2}\right)^2 - 2}{2^2 - 2} \right| = 0.125000$$

$$\delta_2 = \left| \frac{\frac{17}{12} - \frac{3}{2}}{\frac{17}{12}} \right| = 0.058824$$

$$r_2 = \left| \frac{\left(\frac{17}{12}\right)^2 - 2}{2^2 - 2} \right| = 0.003472$$

$$\delta_3 = \left| \frac{\frac{577}{408} - \frac{17}{12}}{\frac{577}{408}} \right| = 0.001733$$

$$r_3 = \left| \frac{\left(\frac{577}{408}\right)^2 - 2}{2^2 - 2} \right| = 0.000003$$



2.3.3. Relaxation approach to the Newton – Raphson method

$$F(x) = 0$$

$$\alpha x + F(x) = \alpha x \;\rightarrow\; x = x + \frac{1}{\alpha} F(x) \;\equiv\; g(x)$$

$$g'(x) = 1 + \frac{1}{\alpha} F'(x)$$

$$g'(x^*) = 0 \;\rightarrow\; \alpha = -\frac{1}{F'(x^*)}$$

$$x = x - \frac{F(x)}{F'(x)} \;\rightarrow\; \boxed{x_n = x_{n-1} - \frac{F(x_{n-1})}{F'(x_{n-1})}}$$

2.3.4. Modification for multiple routs

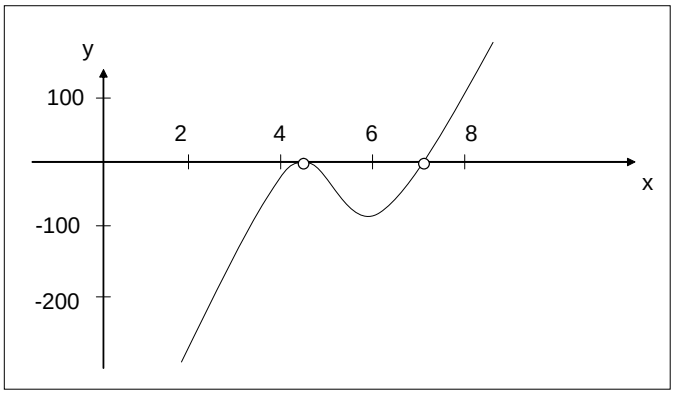
Let $x = c$ be a root of $F(x)$ multiplicity.
Then one way introduce

$$u(x) = \frac{F(x)}{F'(x)} \rightarrow u'(x) = 1 - \frac{F(x)F''(x)}{\left(F'(x)\right)^2}$$

Instead to $F(x)$ apply the Newton-Raphson method to $u(x)$

$$\boxed{x_n = x_{n-1} - \frac{u(x_{n-1})}{u'(x_{n-1})}}$$

Example



$F(x) = x^4 - 8.6x^3 - 35.51x^2 + 464.4x - 998.46 = 0$

$F'(x) = 4x^3 - 25.8x^2 - 71.02x + 464.4$

$F''(x) = 12x^2 - 51.6x - 71.02$

Let

$x_0 = 4.0$
$F(4.0) = -3.42;$
$F'(4.0) = 23.52$
$F''(4.0) = -85.42$

$$u(4) = \frac{F(4.0)}{F'(4.0)} = \frac{-3.42}{23.52} = -0.145408$$

$$u'(4) = 1.0 - \frac{F(4.0) \cdot F''(4.0)}{\left(F'(4.0)\right)^2} = 1.0 - \frac{-3.42 \cdot (-85.42)}{(23.52)^2} = 0.471906$$

$$x_1 = x_0 - \frac{u(4)}{u'(4)} = 4.0 - \frac{-.145408}{.471906} = 4.308129$$

$$x_2 = 4.308129 - .00812 = 4.300001$$

$$x_3 = 4.300000$$

conventional N – R method

$$x_{19} = 4.300000$$

…….....

2.4. THE SECANT METHOD

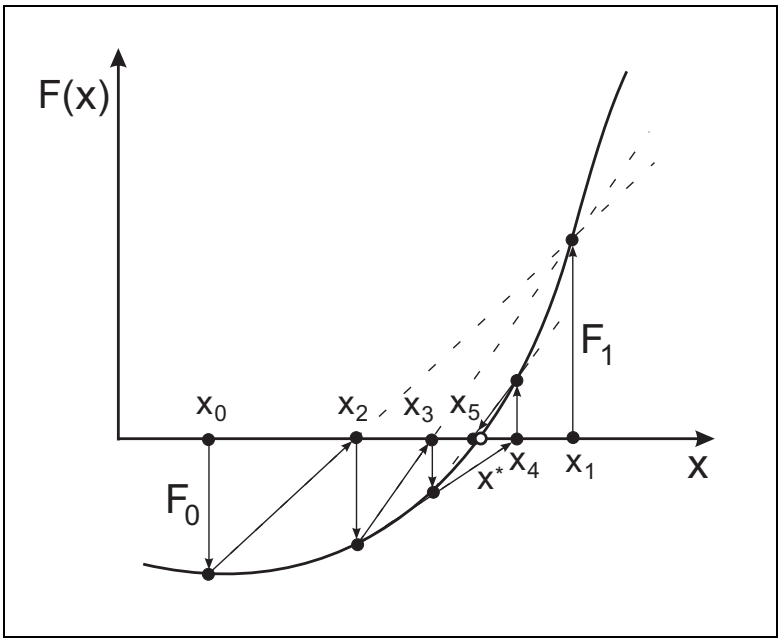
$$x_n = x_{n-1} - \frac{F_{n-1}}{F'_{n-1}} \approx x_{n-1} - F_{n-1}\frac{x_{n-1} - x_{n-2}}{F_{n-1} - F_{n-2}}$$

$$\boxed{x_n = x_{n-1} - \frac{F_{n-1}}{F_{n-1} - F_{n-2}}\left(x_{n-1} - x_{n-2}\right)}$$

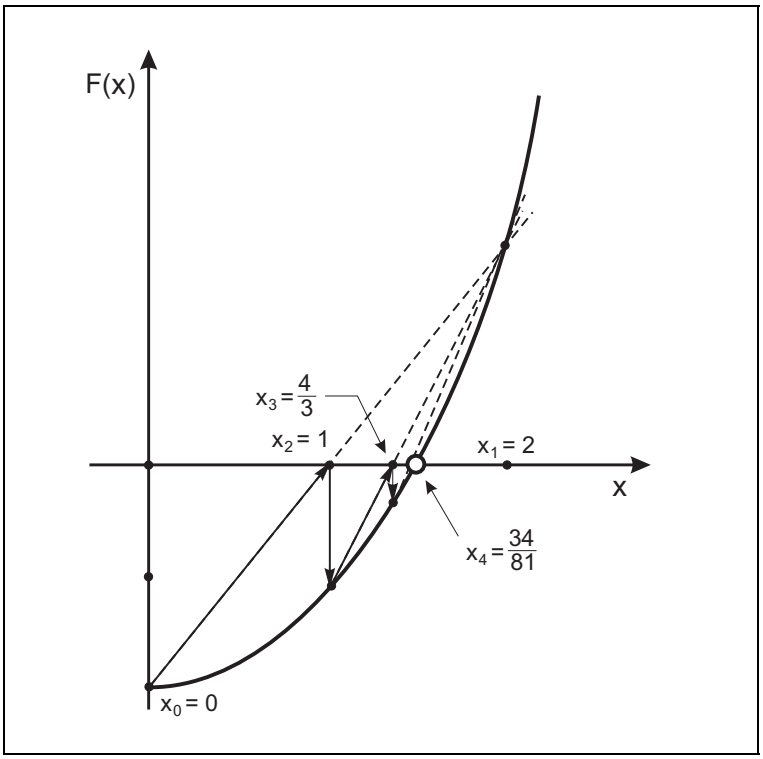
starting points should satisfy the inequality

$$F\left(x_0\right)F\left(x_1\right) < 0$$

Geometrical interpretation



Example



Algorithm

$$x^2 = 2 \rightarrow F(x) \equiv x^2 - 2 = 0$$

Let

$$x_0 = 0 \rightarrow F(0) = -2$$

and

$$x_1 = 2 \rightarrow F(2) = 4 - 2 = 2$$

then

$$x_2 = 2 - \frac{2}{2-(-2)}(2-0) = 1 \quad \rightarrow \quad F(1) = -1$$

$$x_3 = 1 - \frac{-1}{-1-2}(1-2) = \frac{4}{3} = 1.333333 \quad \rightarrow \quad F\left(\frac{4}{3}\right) = -\frac{2}{9} = -0.222222$$

$$x_4 = \frac{4}{3} - \frac{-\frac{2}{9}}{-\frac{2}{3}-(-1)}\left(\frac{4}{3}-1\right) = \frac{14}{9} = 1.555556 \quad \rightarrow \quad F\left(\frac{14}{9}\right) = \frac{34}{81} = 0.419753$$

$$x_5 = \frac{14}{9} - \frac{\frac{34}{81}}{\frac{34}{81}-\left(-\frac{2}{9}\right)}\left(\frac{14}{9}-\frac{4}{3}\right) = \frac{55}{39} = 1.410256$$

$$\rightarrow \quad F\left(\frac{55}{39}\right) = -\frac{17}{1521} = -0.011177$$

$x_T = \sqrt{2} \approx 1.414214$ – true solution

Convergence

$$\delta_1 = \left|\frac{2-0}{2}\right| = 1$$

$$r_1 = \left|\frac{2}{-2}\right| = 1$$

$$\delta_2 = \left|\frac{1-2}{1}\right| = 1$$

$$r_2 = \left|\frac{-1}{-2}\right| = \frac{1}{2} = 0.500000$$

$$\delta_3 = \left|\frac{\frac{4}{3}-1}{\frac{4}{3}}\right| = \frac{1}{4} = 0.250000$$

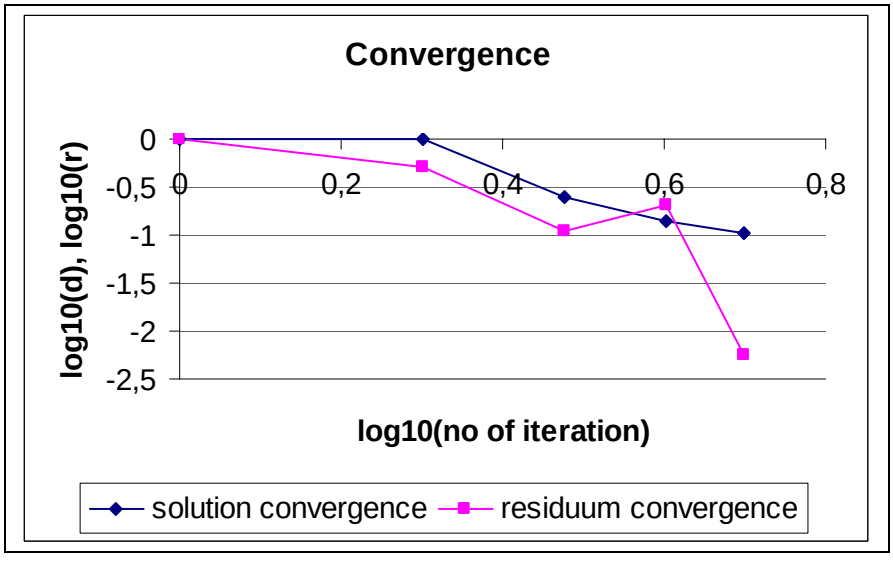
$$r_3 = \left|\frac{-\frac{2}{9}}{-2}\right| = \frac{1}{9} = 0.111111$$

$$\delta_4 = \left|\frac{\frac{14}{9}-\frac{4}{3}}{\frac{14}{9}}\right| = \frac{1}{7} = 0.142857$$

$$r_4 = \left|\frac{\frac{34}{81}}{-2}\right| = \frac{17}{81} = 0.209877$$

$$\delta_5 = \left|\frac{\frac{55}{39}-\frac{14}{9}}{\frac{55}{39}}\right| = \frac{17}{165} = 0.103030$$

$$r_5 = \left|\frac{-\frac{17}{1521}}{-2}\right| = \frac{17}{3042} = 0.005588$$

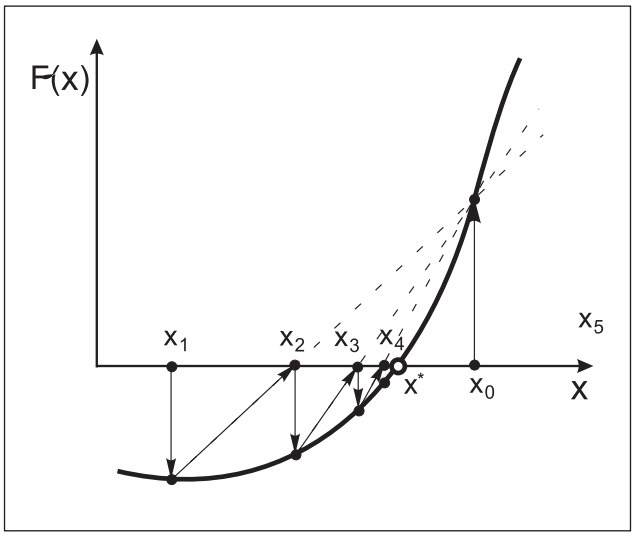


2.5. REGULA FALSI

Let fix one starting point e.g. $x = x_0$ in the secant method.
Then x_{n-2}, F_{n-2} in the secant method are replaced by x_0, F_0.

$$\boxed{x_n = x_{n-1} - \frac{F_{n-1}}{F_{n-1} - F_0}\left(x_{n-1} - x_0\right)}\ , \qquad F(x_0)F(x_1) < 0$$

Geometrical interpretation



Example

$$x^2 = 2 \rightarrow F(x) \equiv x^2 - 2 = 0$$

Let

$$x_0 = 2 \rightarrow F\left(2\right) = +2$$

and

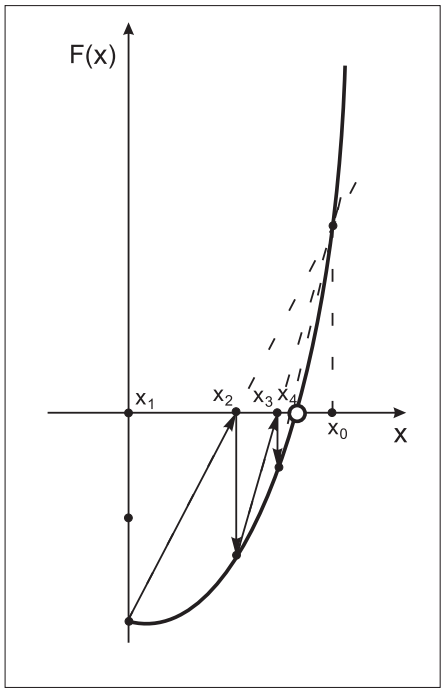
$$x_1 = 0 \rightarrow F\left(0\right) = -2$$

then

$$x_2 = 0 - \frac{-2}{-2-2}\left(0-2\right) = 1 \qquad \rightarrow F\left(1\right) = -1$$

$$x_3 = 1 - \frac{-1}{-1-2}\left(1-2\right) = \frac{4}{3} = 1.333333 \qquad \rightarrow F\left(\frac{4}{3}\right) = -\frac{2}{9}$$

$$x_4 = \frac{4}{3} - \frac{-\frac{2}{9}}{-\frac{2}{9}-2}\left(\frac{4}{3}-2\right) = \frac{7}{5} = 1.400000 \rightarrow F\left(\frac{7}{5}\right) = -\frac{1}{25}$$



$$x_5 = \frac{7}{5} - \frac{-\frac{1}{25}}{-\frac{1}{25} - 2}\left(\frac{7}{5} - 2\right) = \frac{24}{17} = 1.411769$$

$x_T = \sqrt{2} \approx\sim 1.414214$ – true solution

Convergence

$$\delta_1 = \left|\frac{0-2}{0}\right| \rightarrow \text{not exist}$$

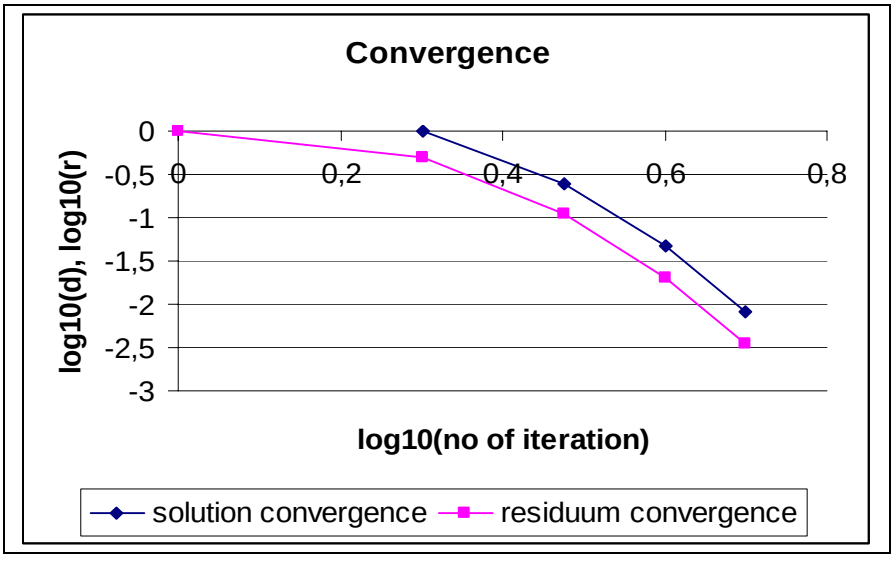
$$r_1 = \left|\frac{-2}{2}\right| = 1$$

$$\delta_2 = \left|\frac{1-0}{1}\right| = 1 \qquad r_2 = \left|\frac{-1}{2}\right| = \frac{1}{2} = 0.500000$$

$$\delta_3 = \left|\frac{\frac{4}{3}-1}{\frac{4}{3}}\right| = \frac{1}{4} = 0.250000 \qquad r_3 = \left|\frac{-\frac{2}{9}}{2}\right| = \frac{1}{9} = 0.111111$$

$$\delta_4 = \left|\frac{\frac{7}{5}-\frac{4}{3}}{\frac{7}{5}}\right| = \frac{1}{21} = 0.047619 \qquad r_4 = \left|\frac{-\frac{4}{100}}{2}\right| = \frac{2}{100} = 0.020000$$

$$\delta_5 = \left|\frac{\frac{24}{17}-\frac{7}{5}}{\frac{24}{17}}\right| = \frac{1}{120} = 0.008333 \qquad r_5 = \left|\frac{-\frac{2}{289}}{2}\right| = \frac{1}{289} = 0.003460$$

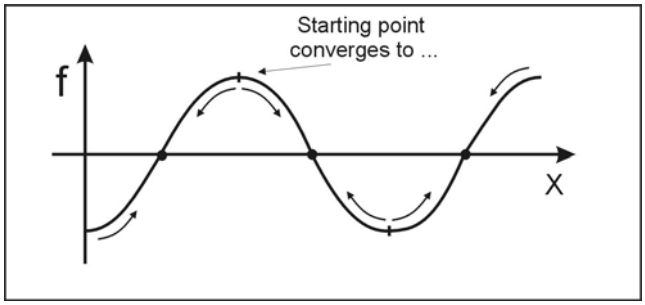


Remarks

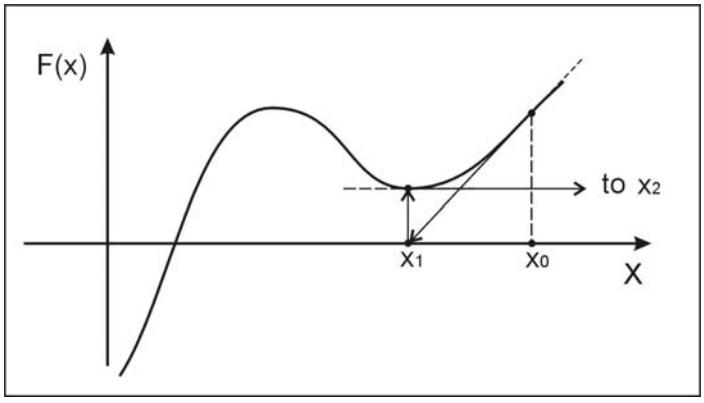
The regula falsi algorithm is more stable but slower than the one corresponding to the secant method.

2.6. GENERAL REMARKS

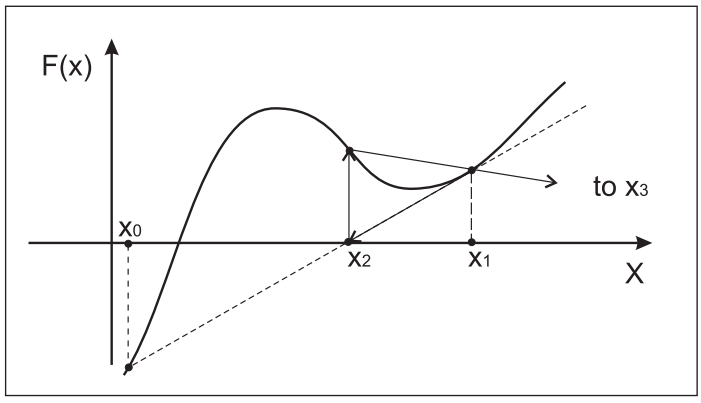
Rough preliminary evaluation of zeros (roots) is suggested



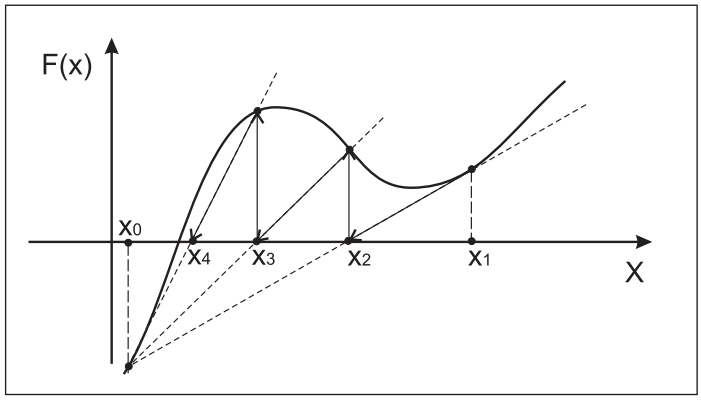
Traps



Newton Raphson
DIVERGENT



Secant Method
DIVERGENT



Regula Falsi
CONVERGENT

Rough evaluation of solution methods

REGULA FALSI	– the slowest but the safest
SECANT METHOD	– faster but less safe
NEWTON-RAPHSON	– the fastest but evaluation of the function derivative is necessary

3. VECTOR AND MATRIX NORM

Generalization of the modulus of a scalar function to a vector-valued function is called a vector norm, to a matrix-valued function is called a matrix norm.

3.1. VECTOR NORM

Vector norm $\|\mathbf{x}\|$ of the vector $\mathbf{x} \in \mathbf{V}$
where:

$\mathbf{V}$ is a linear N-dimensional vector space,
α is a scalar

satisfies the following conditions:

(i) $\|\mathbf{x}\| \geq 0$ $\forall\, \mathbf{x} \in \mathbf{V}$ and $\|\mathbf{x}\| = 0$ if $\mathbf{x} = \mathbf{0}$

(ii) $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$ $\forall$ scalars α and $\forall \mathbf{x} \in \mathbf{V}$

(iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ $\forall\, \mathbf{x}, \mathbf{y} \in \mathbf{V}$

Examples

(1) $\|\mathbf{x}\|_1 = \left[\sum_{i=1}^{N} |x_i|^2\right]^{\frac{1}{2}}$ $p = 2$ Euclidean norm

(2) $\|\mathbf{x}\|_2 = \max_i |x_i|$ $p = \infty$ maximum norm

(3) $\|\mathbf{x}\|_3 = \left[\sum_{i=1}^{N} |x_i|^p\right]^{\frac{1}{p}}$ $p \geq 1$

Examples

$\mathbf{x} = \{2,3,-6\}$

$\|\mathbf{x}\|_1 = \left(2^2 + 3^2 + 6^2\right)^{\frac{1}{2}} = 7$ $p = 2$

$\|\mathbf{x}\|_2 = |-6| = 6$ $p = \infty$

$\|\mathbf{x}\|_3 = |2| + |3| + |-6| = 11$ $p = 1$

3.2. MATRIX NORM

Matrix norm of the $(N \times N)$ matrix A must satisfy the following conditions:

(i) $\|\mathbf{A}\| \geq 0$ and $\|\mathbf{A}\| = 0$ if **A = 0**
(ii) $\|\alpha\mathbf{A}\| = |\alpha| \cdot \|\mathbf{A}\|$ $\forall$ scalar α
(iii) $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$
(iv) $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$

where **A** and **B** have to be of the same dimension.

Examples

$$\|\mathbf{A}\|_1 = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2\right]^{\frac{1}{2}} \quad \text{or} \quad \|\mathbf{A}\|_1 = \left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2\right]^{\frac{1}{2}} \quad \text{- average value}$$

$$\|\mathbf{A}\|_2 = \max_i \sum_{j=1}^{N} |a_{ij}| \quad \text{or} \quad \|\mathbf{A}\|_2 = \frac{1}{N}\max_i \sum_{j=1}^{N} |a_{ij}| \quad \text{- maximum value}$$

Example

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \rightarrow$$

$$\|\mathbf{A}\|_1 = \left[\frac{1}{3^2}\left(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2\right)\right]^{\frac{1}{2}} = 5.627314$$

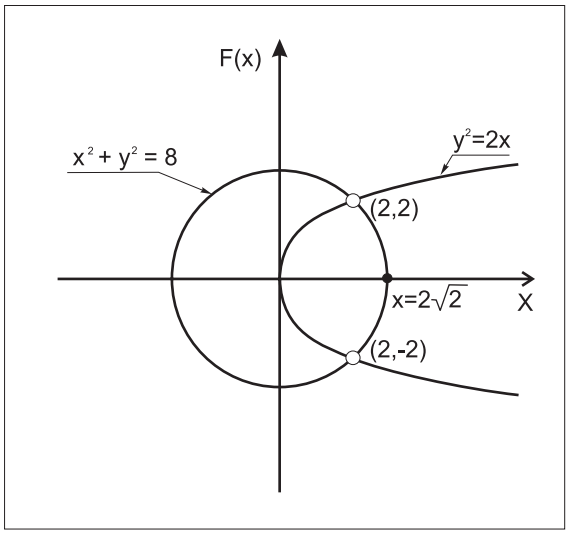
$$\|\mathbf{A}\|_2 = \frac{1}{3}\max\begin{Bmatrix} 1+2+3 \\ 4+5+6 \\ 7+8+9 \end{Bmatrix} = \frac{1}{3}\max\begin{Bmatrix} 6 \\ 15 \\ 24 \end{Bmatrix} = 8$$

4. SYSTEMS OF NONLINEAR EQUATIONS

Denotations

$$\mathbf{x} = \{x_1, x_2, x_3, ..., x_N\}$$
$$\mathbf{F}(\mathbf{x}) = \{F_1(\mathbf{x}), ...F_N(\mathbf{x})\}$$
$$\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

Example



$$\begin{cases} F_1(x,y) \equiv y^2 - 2x = 0 \\ F_2(x,y) \equiv x^2 + y^2 - 8 = 0 \end{cases}$$

4.1. THE METHOD OF SIMPLE ITERATIONS

Algorithm

$$\boxed{\mathbf{x}_n = \mathbf{f}(\mathbf{x}_{n-1})} \qquad \mathbf{f} = \{f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})\}, \qquad \mathbf{x} = \{x_1, \ldots, x_n\}$$

Example

$$\begin{cases} x = \dfrac{1}{2}y^2 \equiv f_1(\mathbf{x}) \\ y = \sqrt{8 - x^2} \equiv f_2(\mathbf{x}) \end{cases} \qquad \mathbf{x} = \{x, y\}$$
$$\mathbf{f}(\mathbf{x}) = \left\{\frac{1}{2}y^2, \sqrt{8 - x^2}\right\}$$

$$\begin{cases} x = y^2 - x \equiv f_1(\mathbf{x}) \\ y = x^2 + y^2 + y - 8 \equiv f_2(\mathbf{x}) \end{cases} \Rightarrow \qquad \mathbf{x} = \mathbf{f}(\mathbf{x})$$

Convergence criterion

$$\delta_n = \frac{\left\|\mathbf{x}_n - \mathbf{x}_{n-1}\right\|}{\left\|\mathbf{x}_n\right\|}, \quad \delta_n \le \delta_{amd}$$

δ_{amd} - admissible error

Theorem

Let $\Re$ denote the region $a_i \le x_i \le b_i$, $i = 1,2,\ldots,N$ in the Euclidean N-dimensional space.

Let **f** satisfy the conditions

- **f** is defined and continuous on $\Re$
- $\left\|\mathbf{J}_f(\mathbf{x})\right\| \le L < 1$
- For each $\mathbf{x} \in \Re$, $\mathbf{f}(\mathbf{x})$ also lies in $\Re$

Then for any $\mathbf{x_0}$ in $\Re$ the sequence of iterations $\mathbf{x}_n = \mathbf{f}(\mathbf{x}_{n-1})$ is convergent to the unique solution **x**

Jacobian matrix $$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & & \\ \cdots & \cdots & & \\ \cdots & \cdots & & \\ \frac{\partial f_N}{\partial x_1} & \cdots & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix}$$

Example

$$\begin{cases} f_1(\mathbf{x}) = y^2 - x \\ f_2(\mathbf{x}) = x^2 + y^2 + y - 8 \end{cases} \rightarrow \quad \mathbf{J} = \begin{bmatrix} -1 & 2y \\ 2x & 2y+1 \end{bmatrix}$$

4.2. NEWTON – RAPHSON METHOD

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

$$\mathbf{F}(\mathbf{x}+\mathbf{h}) = \mathbf{F}(\mathbf{x}) + \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}}\mathbf{h} + \frac{1}{2}\frac{\partial^2 \mathbf{F}(\mathbf{x})}{\partial^2 \mathbf{x}}\mathbf{h}^2 + \cdots$$

$$\mathbf{F}(\mathbf{x}+\mathbf{h}) \approx \mathbf{F}(\mathbf{x}) + \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}}\mathbf{h} \equiv \mathbf{F}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{h} = 0 \rightarrow \mathbf{h} = -\mathbf{J}^{-1}\mathbf{F}$$

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \mathbf{h}_{n-1} = \mathbf{x}_{n-1} - \mathbf{J}_{n-1}^{-1}\mathbf{F}_{n-1}$$

$$\boxed{\mathbf{x}_n = \mathbf{x}_{n-1} - \mathbf{J}_{n-1}\mathbf{F}_{n-1}} \rightarrow \boxed{\mathbf{J}_{n-1}\mathbf{x}_n = \mathbf{J}_{n-1}\mathbf{x}_{n-1} - \mathbf{F}_{n-1} = \mathbf{b}_{n-1}}$$

$$\boxed{\mathbf{J}_{n-1}\mathbf{x}_n = \mathbf{b}_{n-1}} \rightarrow \mathbf{x}_n$$

Solution of simultaneous linear algebraic equations on each iteration step

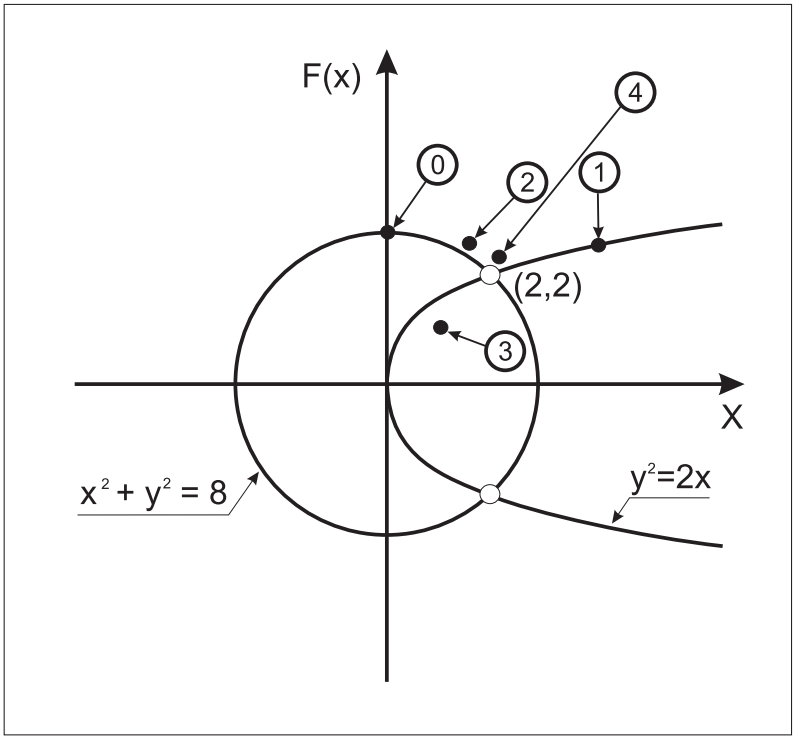
$$\mathbf{x} = \lim_{n\to\infty} \mathbf{x}_n$$

Relaxation factor μ may be also introduced

$$\mathbf{J}_{n-1}\mathbf{x}_n = \mathbf{J}_{n-1}\mathbf{x}_{n-1} - \mu\,\mathbf{F}_{n-1}$$

Example

$$\begin{cases} y^2 = 2x \\ x^2 + y^2 = 8 \end{cases} \rightarrow \begin{matrix} y^2 - 2x = 0 \\ x^2 + y^2 - 8 = 0 \end{matrix} \rightarrow \mathbf{F}(\mathbf{x}) = \begin{Bmatrix} y^2 - 2x \\ x^2 + y^2 - 8 \end{Bmatrix} \equiv \begin{Bmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \end{Bmatrix}, \quad \mathbf{x} = \begin{Bmatrix} x \\ y \end{Bmatrix}$$



$$J = \begin{bmatrix} \dfrac{\partial F_1}{\partial x} & \dfrac{\partial F_1}{\partial y} \\ \dfrac{\partial F_2}{\partial x} & \dfrac{\partial F_2}{\partial y} \end{bmatrix} = \begin{bmatrix} -2 & 2y \\ 2x & 2y \end{bmatrix}$$

Algorithm

$$\begin{bmatrix} -2 & 2y \\ 2x & 2y \end{bmatrix}_{n-1} \begin{Bmatrix} x \\ y \end{Bmatrix}_n = \begin{bmatrix} -2 & 2y \\ 2x & 2y \end{bmatrix}_{n-1} \begin{Bmatrix} x \\ y \end{Bmatrix}_{n-1} - \mu \begin{Bmatrix} y^2 - 2x \\ x^2 + y^2 - 8 \end{Bmatrix}$$

Let $\mu = 1$

$$x_0 = \begin{Bmatrix} 0 \\ 2\sqrt{2} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 2.8284 \end{Bmatrix}$$

$$\begin{bmatrix} -2 & 5.65685 \\ 0 & 5.65685 \end{bmatrix} \begin{Bmatrix} x_1 \\ y_1 \end{Bmatrix} = \begin{bmatrix} -2 & 5.65685 \\ 0 & 5.65685 \end{bmatrix} \begin{Bmatrix} 0 \\ 2.8284 \end{Bmatrix} - \begin{Bmatrix} 8 \\ 0 \end{Bmatrix}$$

$$x_1 = \begin{Bmatrix} 4.0000 \\ 2.8284 \end{Bmatrix}$$

Error estimation

(after the first step of iteration)

Estimated relative solution error $\delta_n = \dfrac{\|\mathbf{x}_n - \mathbf{x}_{n-1}\|}{\|\mathbf{x}_n\|}$

$$\delta_1 = \frac{\|\mathbf{x}_1 - \mathbf{x}_0\|}{\|\mathbf{x}_1\|}$$

$$\mathbf{x}_1 - \mathbf{x}_0 = \{4.0000 - 0.0000,\ 2.8284 - 2.8284\}$$
$$= \{4.0000\ ,\ 0.0000\}$$

Euclidean norm

$$\delta_1^E = \frac{\left[\frac{1}{2}\left(4.0000^2 + 0.0000^2\right)\right]^{1/2}}{\left[\frac{1}{2}\left(4.0000^2 + 2.8284^2\right)\right]^{1/2}} = \frac{2.8284}{3.4641} = 0.8165$$

Maximum norm

$$\delta_1^M = \frac{\sup\{4.0000\ ,\ 0.0000\}}{\sup\{4.0000\ , 2.8284\}} = \frac{4.0000}{4.0000} = 1.0000$$

Relative residual error $r_n = \dfrac{\|\mathbf{F}_n\|}{\|\mathbf{F}_0\|}$

$$r_1 = \frac{\|\mathbf{F}_1\|}{\|\mathbf{F}_0\|}$$

Euclidean norm $\|\mathbf{F}\|_E = \left\{\frac{1}{n}\sum_{j=1}^{n} F_j(\mathbf{x})^2\right\}^{1/2}$

$$\mathbf{F}_0 = \{8.0000\ ,\ 0.0000\}$$
$$\mathbf{F}_1 = \{0.0000, 16.0000\}$$
$$\|\mathbf{F}_0\|_E = \left\{\tfrac{1}{2}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\tfrac{1}{2}\left[0.0000^2 + 8.0000^2\right]\right\}^{1/2} = 5.6568$$
$$\|\mathbf{F}_1\|_E = \left\{\tfrac{1}{2}\left[0.0000^2 + 16.0000^2\right]\right\}^{1/2} = 11.3137$$

$$r_1^E = \frac{11.3137}{5.6568} = 2.0000$$

Maximum norm $\|\mathbf{F}\|_M = \sup_i |F_i|$

$$\|\mathbf{F}_0\|_M = \sup(8.0000\ ,\ 0.0000) = 8.0000$$
$$\|\mathbf{F}_1\|_M = \sup(0.0000, 16.0000) = 16.0000$$
$$r_1^M = \frac{16.0000}{8.0000} = 2.0000$$

Brake-off test

Assume admissible errors for convergence B_C and residuum B_R; check

$$\delta_1^E = 0.81649658 \quad > \quad B_C = 10^{-6}$$
$$\delta_1^M = 1.00000000 \quad > \quad B_C = 10^{-6}$$
$$r_1^E = 2.00000000 \quad > \quad B_R = 10^{-8}$$
$$r_1^M = 2.00000000 \quad > \quad B_R = 10^{-8}$$

Second step of iteration

$$\begin{bmatrix} -2.0000 & 5.6568 \\ 8.0000 & 5.6568 \end{bmatrix} \begin{Bmatrix} x_2 \\ y_2 \end{Bmatrix} = \begin{bmatrix} -2.0000 & 5.6568 \\ 8.0000 & 5.6568 \end{bmatrix} \begin{Bmatrix} 4.0000 \\ 2.8284 \end{Bmatrix} - \begin{Bmatrix} 0.0000 \\ 16.0000 \end{Bmatrix}$$

$$\mathbf{x}_2 = \begin{Bmatrix} 2.4000 \\ 2.2627 \end{Bmatrix}$$

Error estimation
(after the second step of iteration)

Estimated relative solution error

$$\delta_2 = \frac{\|\mathbf{x}_2 - \mathbf{x}_1\|}{\|\mathbf{x}_2\|}$$
$$\mathbf{x}_2 - \mathbf{x}_1 = \{2.4000 - 4.0000\ ,\ 2.2627 - 2.8284\} = \{-1.6000\ ,\ -0.5657\}$$

Euclidean norm

$$\delta_2^E = \frac{\left[\frac{1}{2}\left(1.6000^2 + 0.5657^2\right)\right]^{1/2}}{\left[\frac{1}{2}\left(2.4000^2 + 2.2627^2\right)\right]^{1/2}} = \frac{1.2000}{2.3324} = 0.5145$$

Maximum norm

$$\delta_2^M = \frac{\sup\{1.6000\ ,\ 0.5657\}}{\sup\{2.4000\ ,\ 2.2627\}} = \frac{1.6000}{2.4000} = 0.6667$$

Relative residual error

$$r_2 = \frac{\|\mathbf{F}_2\|}{\|\mathbf{F}_0\|}$$

Euclidean norm $\quad \|\mathbf{F}\|_E = \left\{\frac{1}{n}\sum_{j=1}^{n} F_j(\mathbf{x})^2\right\}^{1/2}$

$$\mathbf{F}_2 = \{0.3200\ ,\ 2.8800\}$$

$$\|\mathbf{F}_0\|_E = \left\{\frac{1}{2}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\frac{1}{2}\left[0^2 + \left(2\sqrt{2}\right)^2\right]\right\}^{1/2} = 5.6568$$

$$\|\mathbf{F}_2\|_E = \left\{\frac{1}{2}\left[0.3200^2 + 2.8800^2\right]\right\}^{1/2} = 2.0490$$

$$r_2^E = \frac{2.0490}{5.6568} = 0.3622$$

Maximum norm $\quad \|\mathbf{F}\|_M = \sup_i |F_i|$

$$\|\mathbf{F}_0\|_M = \sup(8.0000,\ 0.0000) = 8.0000$$

$$\|\mathbf{F}_2\|_M = \sup(0.3200\ ,\ 2.8800) = 2.8800$$

$$r_2^M = \frac{2.8800}{8.0000} = 0.3600$$

Brake-off test

$$\delta_2^E = 0.51449576 \quad > \quad B_C = 10^{-6}$$

$$\delta_2^M = 0.66666667 \quad > \quad B_C = 10^{-6}$$

$$r_2^E = 0.36221541 \quad > \quad B_R = 10^{-8}$$

$$r_2^M = 0.36000000 \quad > \quad B_R = 10^{-8}$$

Third step of iteration

$$\begin{bmatrix} -2 & 4.5255 \\ 4.8 & 4.5255 \end{bmatrix} \begin{Bmatrix} x_3 \\ y_3 \end{Bmatrix} = \begin{bmatrix} -2 & 4.5255 \\ 4.8 & 4.5255 \end{bmatrix} \begin{Bmatrix} 2.4000 \\ 2.2627 \end{Bmatrix} - \begin{Bmatrix} 0.3200 \\ 2.8800 \end{Bmatrix} = \begin{Bmatrix} 5.1200 \\ 18.8800 \end{Bmatrix}$$

$$x_3 = \begin{Bmatrix} 2.0235 \\ 2.0257 \end{Bmatrix}$$

Error estimation

(after five steps of iteration)

Estimated relative solution error

$$\delta_3 = \frac{\|\mathbf{x}_3 - \mathbf{x}_2\|}{\|\mathbf{x}_3\|}$$

$$\mathbf{x}_3 - \mathbf{x}_2 = \{2.0235 - 2.4000 \ , \ 2.0257 - 2.2627\} = \{-0.3765 \ , \ -0.2371\}$$

Euclidean norm

$$\delta_3^E = \frac{\left[\frac{1}{2}\left(0.3765^2 + 0.2371^2\right)\right]^{1/2}}{\left[\frac{1}{2}\left(2.0235^2 + 2.0257^2\right)\right]^{1/2}} = \frac{0.3146}{2.0259} = 0.1554$$

Maximum norm

$$\delta_3^M = \frac{\sup\{0.3765 \ , \ 0.2371\}}{\sup\{2.0235 \ , \ 2.0257\}} = \frac{0.3765}{2.0257} = 0.1859$$

Relative residual error

$$r_3 = \frac{\|\mathbf{F}_3\|}{\|\mathbf{F}_0\|}$$

Euclidean norm $\quad \|\mathbf{F}\|_E = \left\{ \frac{1}{n} \sum_{j=1}^{n} F_j(\mathbf{x})^2 \right\}^{1/2}$

$$\mathbf{F}_3 = \{+0.0562 \ , \ 0.1979\}$$

$$\|\mathbf{F}_0\|_E = \left\{ \frac{1}{2} \left[F_1\left(\mathbf{x}_0\right)^2 + F_2\left(\mathbf{x}_0\right)^2 \right] \right\}^{1/2} = \left\{ \frac{1}{2} \left[0.0000^2 + \left(8.0000\right)^2 \right] \right\}^{1/2} = 5.6568$$

$$\|\mathbf{F}_3\|_E = \left\{ \frac{1}{2} \left[0.0562^2 + 0.1979^2 \right] \right\}^{1/2} = 0.1455$$

$$r_3^E = \frac{0.1455}{5.6568} = 0.0257$$

Maximum norm $\|\mathbf{F}\|_M = \sup_i |F_i|$

$\|\mathbf{F}_0\|_M = \sup(8.0000\ ,\ 0.0000) = 8.0000$

$\|\mathbf{F}_3\|_M = \sup(0.0562\ ,\ 0.1979) = 0.1979$

$r_3^M = \dfrac{0.1979}{8.0} = 0.0247$

Brake-off test

$\delta_3^E = 0.15538736 \quad > \quad B_C = 10^{-6}$

$\delta_3^M = 0.18585147 \quad > \quad B_C = 10^{-6}$

$r_3^E = 0.02572098 \quad > \quad B_R = 10^{-8}$

$r_3^M = 0.02474265 \quad > \quad B_R = 10^{-8}$

Fourth step of iteration

$$\begin{bmatrix} -2 & 4.0513 \\ 4.0471 & 4.0513 \end{bmatrix} \begin{Bmatrix} x_4 \\ y_4 \end{Bmatrix} = \begin{bmatrix} -2 & 4.0513 \\ 4.0471 & 4.0513 \end{bmatrix} \begin{Bmatrix} 2.0235 \\ 2.0257 \end{Bmatrix} - \begin{Bmatrix} 0.0562 \\ 0.1979 \end{Bmatrix} = \begin{Bmatrix} 4.1033 \\ 16.1979 \end{Bmatrix}$$

$$x_4 = \begin{Bmatrix} 2.0001 \\ 2.0002 \end{Bmatrix}$$

Error estimation

(after four steps of iteration)

Estimated solution error

$$\delta_4 = \frac{\|\mathbf{x}_4 - \mathbf{x}_3\|}{\|\mathbf{x}_4\|}$$

$$\mathbf{x}_4 - \mathbf{x}_3 = \{2.0001 - 2.0235\ ,\ 2.0002 - 2.0257\} = \{-0.0236\ ,\ -0.0254\}$$

Euclidean norm

$$\delta_4^E = \frac{\left[\frac{1}{2}\left(0.0234^2 + 0.0254^2\right)\right]^{1/2}}{\left[\frac{1}{2}\left(2.0001^2 + 2.0002^2\right)\right]^{1/2}} = \frac{0.0247}{2.0002} = 0.0122$$

Maximum norm

$$\delta_4^M = \frac{\sup\{0.0234\ ,\ 0.0254\}}{\sup\{2.0001,\ 2.0002\}} = \frac{0.0254}{2.0002} = 0.0127$$

Relative residual error

$$r_4 = \frac{\|\mathbf{F}_4\|}{\|\mathbf{F}_0\|}$$

Euclidean norm $\|\mathbf{F}\|_E = \left\{ \frac{1}{n} \sum_{j=1}^{n} F_j(\mathbf{x})^2 \right\}^{1/2}$

$$\mathbf{F}_4 = \{+0.0007 \, , \, 0.0012\}$$

$$\|\mathbf{F}_0\|_E = \left\{ \tfrac{1}{2} \left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 \right] \right\}^{1/2} = \left\{ \tfrac{1}{2} \left[0.0000^2 + (8.0000)^2 \right] \right\}^{1/2} = 5.6568$$

$$\|\mathbf{F}_4\|_E = \left\{ \tfrac{1}{2} \left[0.0007^2 + 0.0012^2 \right] \right\}^{1/2} = 0.0010$$

$$r_4^E = \frac{0.0010}{5.6568} = 0.0002$$

Maximum norm $\|\mathbf{F}\|_M = \sup_i |F_i|$

$$\|\mathbf{F}_0\|_M = \sup(8.0000 \, , \, 0.0000) = 8.0000$$

$$\|\mathbf{F}_4\|_M = \sup(0.0007 \, , \, 0.0012) = 0.0012$$

$$r_4^M = \frac{0.0012}{8} = 0.0002$$

Brake-off test

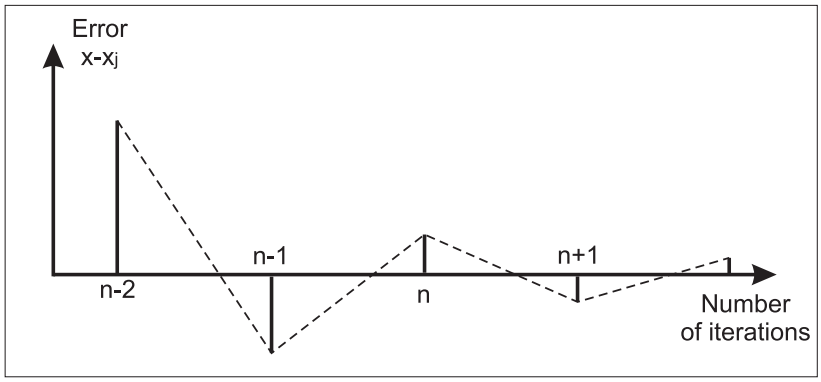
$$\delta_4^E = 0.01223018 \quad > \quad B_C = 10^{-6}$$

$$\delta_4^M = 0.01272134 \quad > \quad B_C = 10^{-6}$$

$$r_4^E = 0.00017009 \quad > \quad B_R = 10^{-8}$$

$$r_4^M = 0.0001460 \quad > \quad B_R = 10^{-8}$$

Aitken acceleration process



$$x - x_n = \alpha_n\left(x - x_{n-1}\right)$$

ASSUME $\alpha_n = \alpha$ constant

then

$$\left.\begin{aligned} x - x_n &= \alpha\left(x - x_{n-1}\right) \\ x - x_{n-1} &= \alpha\left(x - x_{n-2}\right) \end{aligned}\right\} \rightarrow \frac{x - x_n}{x - x_{n-1}} = \frac{x - x_{n-1}}{x - x_{n-2}} \rightarrow \boxed{x = \frac{x_{n-2}x_n - x_{n-1}^2}{x_n - 2x_{n-1} + x_{n-2}}}$$

Example

$$x_4^{NEW} = \frac{x_2 x_4^{OLD} - x_3^2}{x_4^{OLD} - 2x_3 + x_2} = \frac{2.400 \times 2.0001 - 2.0235^2}{2.0001 - 2 \times 2.0235 + 2.400} = 1.9985$$

$$y_4^{NEW} = \frac{y_2 y_4^{OLD} - y_3^2}{y_4^{OLD} - 2y_3 + y_2} = \frac{2.2627 \times 2.0002 - 2.0257^2}{2.0002 - 2 \times 2.0257 + 2.2627} = 1.9972$$

Hence continuing $N - R$ iteration

$$\begin{bmatrix} -2 & 3.9943 \\ 3.9971 & 3.9943 \end{bmatrix}\begin{Bmatrix} x_5 \\ y_5 \end{Bmatrix} = \begin{bmatrix} -2 & 3.9943 \\ 3.9971 & 3.9943 \end{bmatrix}\begin{Bmatrix} 1.9985 \\ 1.9972 \end{Bmatrix} - \begin{Bmatrix} -0.0085 \\ -0.0173 \end{Bmatrix} = \begin{Bmatrix} 3.9886 \\ 15.9828 \end{Bmatrix}$$

$$x_5 = \begin{Bmatrix} 2.0000 \\ 2.0000 \end{Bmatrix}$$

Error estimation

(after five steps of iteration)

Estimated solution error

$$\delta_5 = \frac{\left\|\mathbf{x}_5 - \mathbf{x}_4\right\|}{\left\|\mathbf{x}_5\right\|}$$

$$\mathbf{x}_5 - \mathbf{x}_4 = \{2.0000 - 1.9985\ ,\ 2.0000 - 1.9972\} = \{0.0015\ ,\ 0.0028\}$$

Euclidean norm

$$\delta_5^E = \frac{\left[\frac{1}{2}\left(0.0015^2 + 0.0028^2\right)\right]^{1/2}}{\left[\frac{1}{2}\left(2.0000^2 + 2.0000^2\right)\right]^{1/2}} = \frac{0.0023}{2.0000} = 0.0011$$

Maximum norm

$$\delta_5^M = \frac{\sup\{0.0015\ ,\ 0.0028\}}{\sup\{2.0000\ ,\ 2.0000\}} = \frac{0.0028}{2.0000} = 0.0014$$

Relative residual error

$$r_5 = \frac{\left\|\mathbf{F}_5\right\|}{\left\|\mathbf{F}_0\right\|}$$

Euclidean norm $\quad \left\|\mathbf{F}\right\|_E = \left\{\frac{1}{n}\sum_{j=1}^{n} F_j(\mathbf{x})^2\right\}^{1/2}$

$$\mathbf{F}_5 = \{0.00001\ ,\ 0.00001\}$$

$$\left\|\mathbf{F}_0\right\|_E = \left\{\frac{1}{2}\left[F_1\left(\mathbf{x}_0\right)^2 + F_2\left(\mathbf{x}_0\right)^2\right]\right\}^{1/2} = \left\{\frac{1}{2}\left[0.0000^2 + \left(8.0000\right)^2\right]\right\}^{1/2} = 5.6568$$

$$\left\|\mathbf{F}_5\right\|_E = \left\{\frac{1}{2}\left[0.00001^2 + 0.00001^2\right]\right\}^{1/2} = 0.00001$$

$$r_5^E = \frac{0.00001}{5.6568} = 0.000002$$

Maximum norm $\quad \left\|\mathbf{F}\right\|_M = \sup_i \left|F_i\right|$

$$\left\|\mathbf{F}_0\right\|_M = \sup\left(8.0000\ ,\ 0.0000\right) = 8.0000$$

$$\left\|\mathbf{F}_5\right\|_M = \sup\left(0.00001\ ,\ 0.00001\right) = 0.00001$$

$$r_5^M = \frac{0.00001}{8.0000} = 0.000002$$

Brake-off test

$\delta_5^E = 0.00113413 \quad > \quad B_C = 10^{-6}$

$\delta_5^M = 0.00142690 \quad > \quad B_C = 10^{-6}$

$r_5^E = 0.00000164 \quad > \quad B_R = 10^{-8}$

$r_5^M = 0.00000129 \quad > \quad B_R = 10^{-8}$

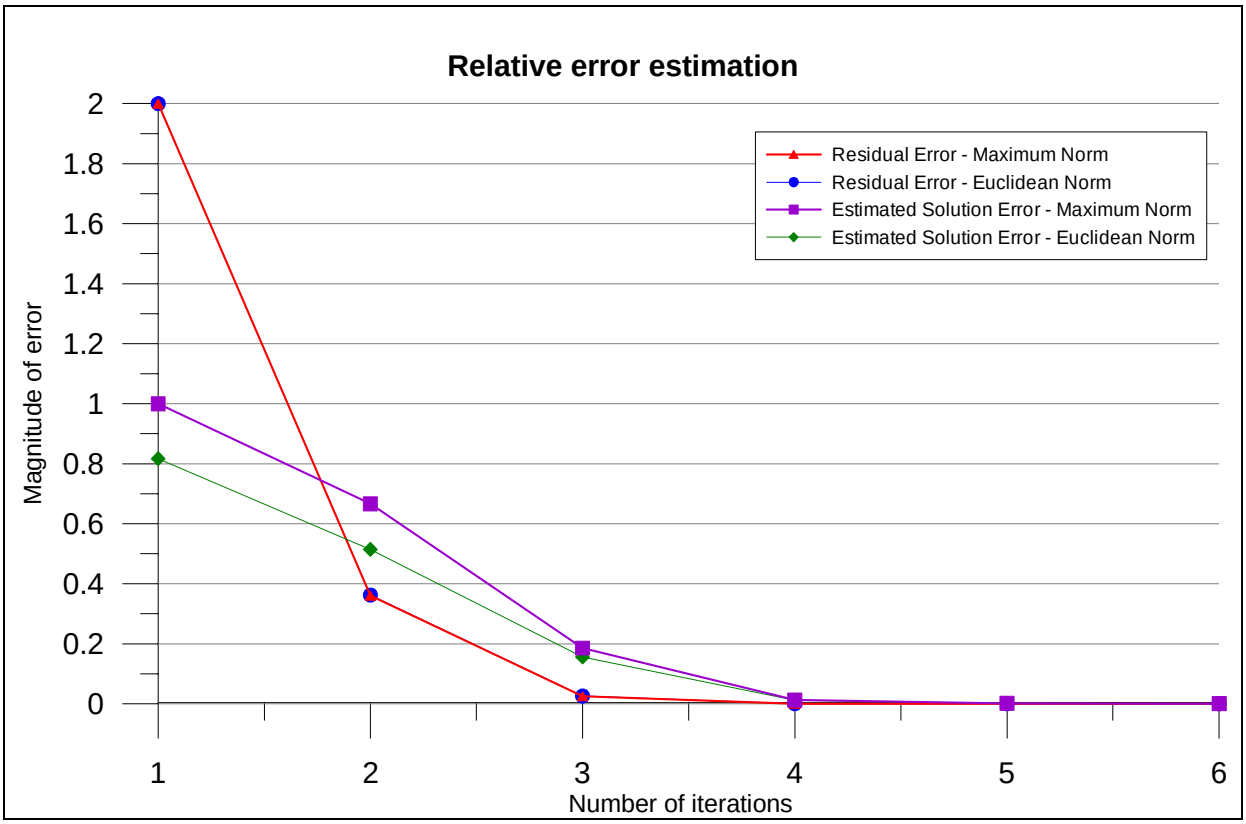
SOLUTION SUMMARY

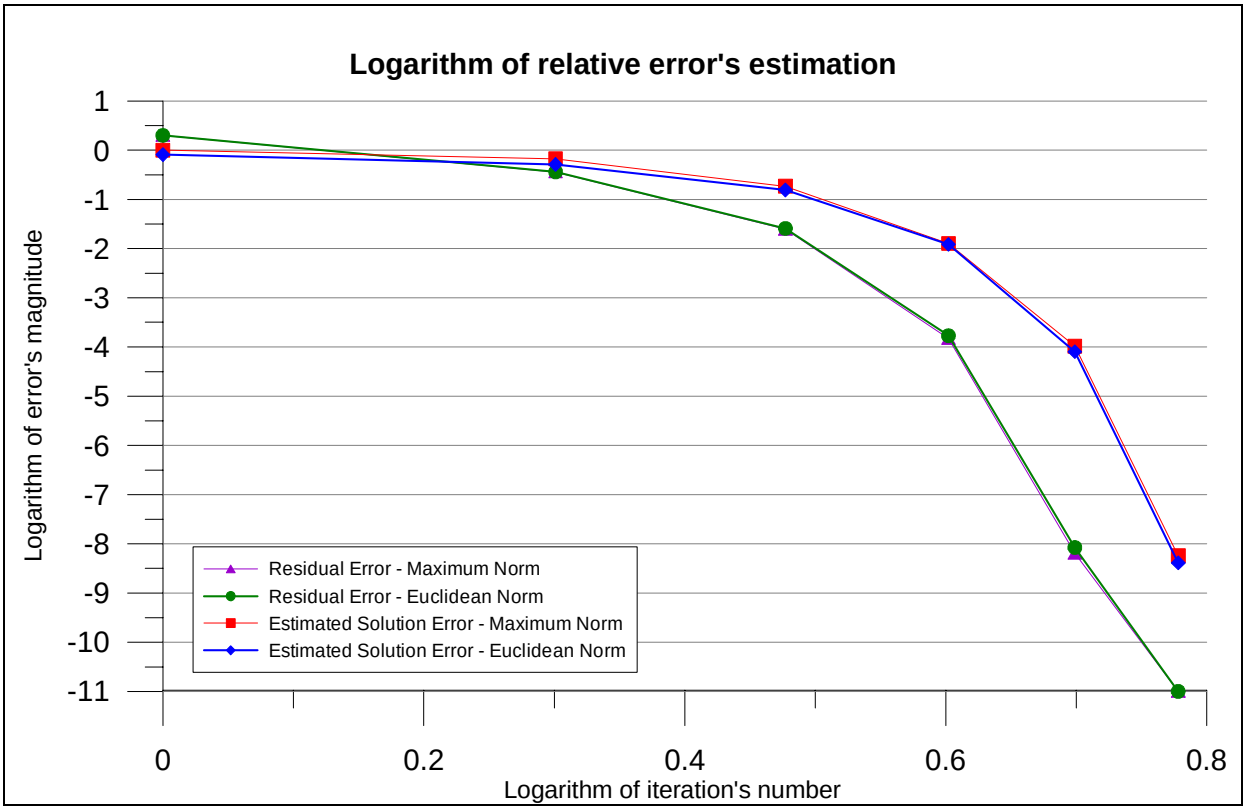
Standard case – no acceleration

Iteration Number	solution		Relative solution error		Relative residual error	
	x	y	Euclidean norm δ^E	Maximum norm δ^M	Euclidean norm r^E	Maximum norm r^M
1	4.0000000000	2.8284271247	0.8164965809	1.0000000000	2.0000000000	2.0000000000
2	2.4000000000	2.2627416998	0.5144957554	0.6666666667	0.3622154055	0.3600000000
3	2.0235294118	2.0256529555	0.1553873552	0.1858514743	0.0257209770	0.0247426471
4	2.0000915541	2.0002076324	0.0122301810	0.0127213409	0.0001700889	0.0001495997
5	2.0000000014	2.0000000115	0.0000802250	0.0001038105	0.0000000084	0.0000000064
6	2.0000000000	2.0000000000	0.0000000041	0.0000000057	0.0000000000	0.0000000000

Aitken Acceleration included from the fourth iteration

Iteration Number	solution		Relative solution error		Relative residual error	
	x	y	Euclidean norm δ^E	Maximum norm δ^M	Euclidean norm r^E	Maximum norm r^M
1	4.0000000000	2.8284271247	0.8164965809	1.0000000000	2.0000000000	2.0000000000
2	2.4000000000	2.2627416998	0.5144957554	0.6666666667	0.3622154055	0.3600000000
3	2.0235294118	2.0256529555	0.1553873552	0.1858514743	0.0257209770	0.0247426471
4	1.9985355138	1.9971484092	0.0134178544	0.0142627172	0.0024025701	0.0021567540
5	2.0000003576	2.0000022149	0.0011341275	0.0014269013	0.0000016404	0.0000012862
6	2.0000000000	2.0000000000	0.0000007932	0.0000011074	0.0000000000	0.0000000000
7	2.0000000000	2.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000





The same in the log-log scale

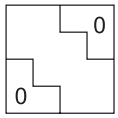
5. SOLUTION OF SIMULTANEOUS LINEAR ALGEBRAIC EQUATIONS (SLAE)

5.1. INTRODUCTION

- Sources of S.L.A.E.
- Features

$$\underset{n\times n}{\mathbf{A}}\ \underset{n\times 1}{\mathbf{x}} = \underset{n\times 1}{\mathbf{b}}$$

mostly $\mathbf{A}$:
- $\mathbf{A}^T = \mathbf{A} \rightarrow$ symmetric
- $\mathbf{x}^T\mathbf{A}\mathbf{x} > \mathbf{0} \quad \forall \quad \mathbf{x} \in R_n$ positive definite (energy >0)
- banded (or sparse)



- $n >> 1$

Solution methods

- = elimination : *Gauss – Jordan* $\left(\det \mathbf{A} \neq 0 - \text{non singular}\right)$
  Cholesky $\left(\mathbf{A}^T = \mathbf{A}, \quad \mathbf{x}^T\mathbf{A}\mathbf{x} > 0, \quad as \quad above\right)$
- = iterative : *Jacobi*
  Gauss – Seidel
- = combined *(iteration and elimination)*
- = special methods: *frontal solution*
  methods for sparse matrices

5.2. GAUSSIAN ELIMINATION

Example

$$\begin{bmatrix} 6 & 2 & 2 & 4 \\ -1 & 2 & 2 & -3 \\ 0 & 1 & 1 & 4 \\ 1 & 0 & 2 & 3 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{Bmatrix} = \begin{Bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{Bmatrix}$$

Assume Table

$$[\mathbf{A}\vdots\mathbf{b}] \quad \rightarrow \quad [\mathbf{I}\vdots\mathbf{x}]$$

$$\left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & 1 \\ -1 & 2 & 2 & -3 & -1 \\ 0 & 1 & 1 & 4 & 2 \\ 1 & 0 & 2 & 3 & 1 \end{array}\right] \rightarrow \left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & 1 \\ 0 & 7/3 & 7/3 & -7/3 & -5/6 \\ 0 & 1 & 1 & 4 & 2 \\ 0 & -1/3 & 5/3 & 7/3 & 5/6 \end{array}\right] \rightarrow$$

$$\left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & 1 \\ 0 & 7/3 & 7/3 & -7/3 & -5/6 \\ 0 & 0 & \boxed{0} & 5 & 33/14 \\ 0 & 0 & 2 & 2 & 5/7 \end{array}\right] \begin{array}{c} \textit{partial} \\ \textit{pivoting} \\ \rightarrow \\ \textit{interchange} \\ \textit{of rows } 3,4 \end{array} \left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & 1 \\ 0 & 7/3 & 7/3 & -7/3 & -5/6 \\ 0 & 0 & 2 & 2 & 5/7 \\ 0 & 0 & 0 & 5 & 33/14 \end{array}\right] \rightarrow$$

There are several ways how to proceed now.

(i)

$$\rightarrow \left[\begin{array}{cccc:c} 6 & 2 & 2 & 0 & -31/35 \\ 0 & 7/3 & 7/3 & 0 & 4/15 \\ 0 & 0 & 2 & 0 & -8/35 \\ 0 & 0 & 0 & 5 & 33/14 \end{array}\right] \rightarrow \left[\begin{array}{cccc:c} 6 & 2 & 0 & 0 & -23/35 \\ 0 & 7/3 & 0 & 0 & 8/15 \\ 0 & 0 & 2 & 0 & -8/35 \\ 0 & 0 & 0 & 5 & 33/14 \end{array}\right] \rightarrow$$

$$\rightarrow \left[\begin{array}{cccc:c} 6 & 0 & 0 & 0 & -39/35 \\ 0 & 7/3 & 0 & 0 & 8/15 \\ 0 & 0 & 2 & 0 & -8/35 \\ 0 & 0 & 0 & 5 & 33/14 \end{array}\right] \rightarrow \left[\begin{array}{cccc:c} 1 & 0 & 0 & 0 & -13/70 \\ 0 & 1 & 0 & 0 & 8/35 \\ 0 & 0 & 1 & 0 & -4/35 \\ 0 & 0 & 0 & 1 & 33/70 \end{array}\right]$$

final solution

(ii)

$$\rightarrow \left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & -1 \\ 0 & 7/3 & 7/3 & -7/3 & -5/6 \\ 0 & 0 & 2 & 2 & 5/7 \\ 0 & 0 & 0 & 1 & 33/70 \end{array}\right] \rightarrow \left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & 1 \\ 0 & 7/3 & 7/3 & -7/3 & -5/6 \\ 0 & 0 & 1 & 0 & -4/35 \\ 0 & 0 & 0 & 1 & 33/70 \end{array}\right] \rightarrow$$

$$\rightarrow \left[\begin{array}{cccc:c} 6 & 2 & 2 & 4 & 1 \\ 0 & 1 & 0 & 0 & 8/35 \\ 0 & 0 & 1 & 0 & -4/35 \\ 0 & 0 & 0 & 0 & 33/70 \end{array}\right] \rightarrow \left[\begin{array}{cccc:c} 1 & 0 & 0 & 0 & -13/70 \\ 0 & 1 & 0 & 0 & 8/35 \\ 0 & 0 & 1 & 0 & -4/35 \\ 0 & 0 & 0 & 1 & 33/70 \end{array}\right]$$

final solution

General algorithm

$$\mathbf{Ax} = \mathbf{b} \leftrightarrow \sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad i = 1,2,...,n$$

where

$$\underset{n\times n}{\mathbf{A}} \equiv \left[a_{ij} \right] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ . & . & \cdots & . \\ . & . & ..a_{ij} & . \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ i \end{matrix} \quad j$$

I *steps forward (without pivoting)*

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - m_{ik} a_{kj}^{(k-1)} \qquad \textit{where} \qquad m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \qquad a_{ij}^{(0)} = a_{ij}, \quad b_i^{(0)} = b_i$$

$$b_i^{(k)} = b_i^{(k-1)} - m_{ik} b_k^{(k-1)} \qquad k = 1,2,...,n-1; \quad j = k+1,...,n; \quad i = k+1,...,n$$

II *steps back*

$$x_i = \left[b_i^{(n-1)} - \sum_{j=i+1}^{n} a_{ij}^{(n-1)} x_j \right] \frac{1}{a_{ii}^{(n-1)}} \qquad i = n-1,...,2,1$$

Number of operations:

$\frac{1}{3} N^3 + N^2 + 0(N)$ - for Gauss procedure (not bounded)

$N^4 + 0\left(N^3\right)$ - for Cramer's formulas

Multiple right hand side

$$\left[\mathbf{A} \vdots \mathbf{b_1} \cdots \mathbf{b_k} \right] \rightarrow \left[\mathbf{I} \vdots \mathbf{x_1} \cdots \mathbf{x_k} \right]$$

5.3. MATRIX FACTORIZATION LU

Simultaneous equations in matrix notation:

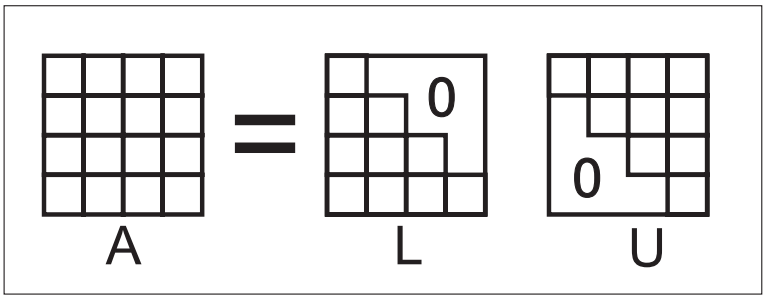
$$\mathbf{A}\,\mathbf{x} = \mathbf{b}\,, \qquad \det \mathbf{A} \neq 0$$

Matrix factorization

$$\mathbf{A} = \mathbf{L}\mathbf{U}$$

L – lower triangle matrix

U – upper triangle matrix



Given

$$\mathbf{L}\underbrace{\mathbf{U}\mathbf{x}}_{\mathbf{y}} = \mathbf{b} \quad \rightarrow \quad \boxed{\begin{array}{l}\mathbf{L}\mathbf{y} = \mathbf{b} \\ \mathbf{U}\mathbf{x} = \mathbf{y}\end{array}} \quad \begin{array}{l}\rightarrow \mathbf{y} \\ \rightarrow \mathbf{x}\end{array} \quad \begin{array}{l}\text{step foreward} \\ \text{step back}\end{array}$$

Gauss elimination method

I. Obtain $\mathbf{L}\mathbf{y} = \mathbf{b} \;\rightarrow\; \mathbf{y} = \mathbf{L}^{-1}\mathbf{b}$

II. Solve $\mathbf{U}\mathbf{x} = \mathbf{y} \;\rightarrow\; \mathbf{x} = \mathbf{U}^{-1}\mathbf{y}$

5.4. CHOLESKI ELIMINATION METHOD

Assumptions

$$\begin{cases} \mathbf{A}^{\mathbf{T}} = \mathbf{A} & \textit{symetric matrix} \\ \mathbf{x}^t\mathbf{A}\mathbf{x} > 0 \rightarrow \det \mathbf{A} \neq 0 & \textit{nonsingular and positive definite matrix} \\ a_{ij} = 0 \quad \textit{for} \quad |i-j| > m, \ m \leq n & \textit{banded matrix} \end{cases}$$

Definition

A matrix is said to be strictly diagonally dominant if

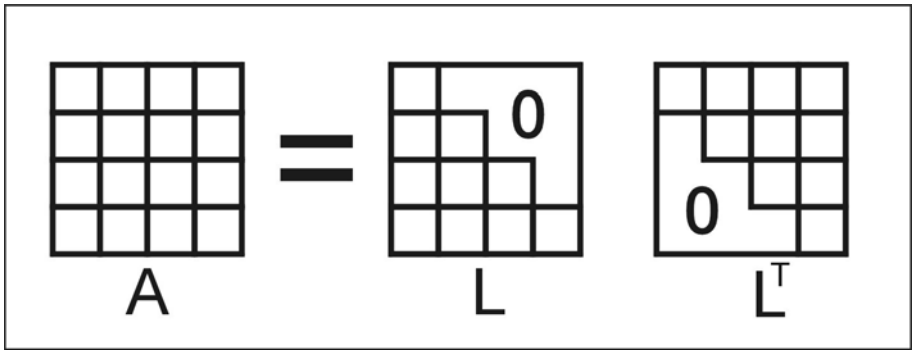
$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \qquad i = 1,2,\ldots,n$$

Theorem

If a real matrix $\mathbf{A}$ is symmetric, strictly diagonally dominant, and has positive diagonally elements, then $\mathbf{A}$ is positive definite.

Matrix factorization

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}, \qquad \mathbf{U} = \mathbf{L}^{\mathbf{T}}$$



$$\mathbf{A}\mathbf{x} = \mathbf{b} \rightarrow \begin{cases} \mathbf{L}\mathbf{y} = \mathbf{b} \rightarrow \mathbf{y} & \textit{step foreward} \\ \mathbf{L}^{\mathbf{T}}\mathbf{x} = \mathbf{y} \rightarrow \mathbf{x} & \textit{step back} \end{cases}$$

Remark

here $\mathbf{U} \equiv \mathbf{L}^{\mathbf{T}}$

Solution algorithm

Initial step: Choleski factorization of matrix

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2} \qquad \textit{diagonal elements}$$

$$l_{ij} = \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right) \frac{1}{l_{jj}} \qquad \textit{off diagonal elements}$$

$$\textit{where} \quad j = 1, 2, ..., n \quad , \quad i = j+1, ..., n$$

I step foreword

$$y_i = \left[b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right] \frac{1}{l_{ii}} \qquad i = 1, 2, ..., n$$

II step back – similar as in the Gauss-Jordan algorithm

$$x_i = \left[y_i - \sum_{j=i+1}^{n} l_{ji} x_j \right] \frac{1}{l_{ii}} \qquad i = n, ..., 2, 1$$

Example

Cholesky factorization of the given matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 5 & -2 \\ 0 & -2 & 5 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Column 1:

$$l_{11}^2 = a_{11} \quad \rightarrow \; l_{11} = \sqrt{a_{11}} \quad \rightarrow \quad l_{11} = \sqrt{4} = 2$$

$$l_{11} l_{21} = a_{21} \rightarrow \; l_{21} = \frac{a_{21}}{l_{11}} \quad \rightarrow \quad l_{21} = \frac{-2}{2} = -1$$

$$l_{11} l_{31} = a_{31} \rightarrow \; l_{31} = \frac{a_{31}}{l_{11}} \quad \rightarrow \quad l_{31} = \frac{0}{2} = 0$$

Column 2:

$$l_{21}^2 + l_{22}^2 = a_{22} \quad \rightarrow \quad l_{22} = \sqrt{a_{22} - l_{21}^2} \quad \rightarrow \quad l_{22} = \sqrt{5 - (-1)^2} = 2$$

$$l_{31} l_{21} + l_{32} l_{22} = a_{32} \quad \rightarrow \quad l_{32} = (a_{32} - l_{31} l_{21}) \frac{1}{l_{22}} \quad \rightarrow \quad l_{32} = \left[-2 - 0 \times (-1) \right] \frac{1}{2} = -1$$

Column 3:

$$l_{31}^2 + l_{32}^2 + l_{33}^2 = a_{33} \quad \rightarrow \quad l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2} \quad \rightarrow \quad l_{33} = \sqrt{5 - 0^2 - (-1)^2} = 2$$

Final result:

$$A = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 5 & -2 \\ 0 & -2 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$

5.5. ITERATIVE METHODS

Example

$$\left.\begin{aligned} 20x_1 + 2x_2 - x_3 &= 25 \\ 2x_1 + 13x_2 - 2x_3 &= 30 \\ x_1 + x_2 + x_3 &= 2 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= -\tfrac{1}{10}x_2 + \tfrac{1}{20}x_3 + \tfrac{5}{4} \\ x_2 &= -\tfrac{2}{13}x_1 + \tfrac{2}{13}x_3 + \tfrac{30}{13} \\ x_3 &= -x_1 - x_2 + 2 \end{aligned}$$

The method of simple iterations may be applied, using one of the following algorithms:

JACOBI ITERATION SCHEME

$$\begin{aligned} x_1^{(n)} &= -\tfrac{1}{10}x_2^{(n-1)} + \tfrac{1}{20}x_3^{(n-1)} + \tfrac{5}{4} \\ x_2^{(n)} &= -\tfrac{2}{13}x_1^{(n-1)} + \tfrac{2}{13}x_3^{(n-1)} + \tfrac{30}{13} \\ x_3^{(n)} &= -x_1^{(n-1)} - x_2^{(n-1)} + 2 \end{aligned}$$

GAUSS - SEIDEL ITERATION SCHEME

$$\begin{aligned} x_1^{(n)} &= -\tfrac{1}{10}x_2^{(n-1)} + \tfrac{1}{20}x_3^{(n-1)} + \tfrac{5}{4} \\ x_2^{(n)} &= -\tfrac{2}{13}x_1^{(n)} + \tfrac{2}{13}x_3^{(n-1)} + \tfrac{30}{13} \\ x_3^{(n)} &= -x_1^{(n)} - x_2^{(n)} + 2 \end{aligned}$$

Let $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$

$$\begin{aligned} x_1^{(1)} &= -\tfrac{1}{10}\cdot 0 + \tfrac{1}{20}\cdot 0 + \tfrac{5}{4} = 1.250000 \\ x_2^{(1)} &= -\tfrac{2}{13}\cdot 0 + \tfrac{2}{13}\cdot 0 + \tfrac{30}{13} = 2.307692 \\ x_3^{(1)} &= -0 - 0 + 2 = 2.000000 \end{aligned}$$

$$\begin{aligned} x_1^{(1)} &= -\tfrac{1}{10}\cdot 0 + \tfrac{1}{20}\cdot 0 + \tfrac{5}{4} = 1.250000 \\ x_2^{(1)} &= -\tfrac{2}{13}\cdot 1.250000 + \tfrac{2}{13}\cdot 0 + \tfrac{30}{13} = 2.115385 \\ x_3^{(1)} &= -1.250000 - 2.115385 + 2 = -1.365385 \end{aligned}$$

$$\begin{aligned} x_1^{(2)} &= -\tfrac{1}{10}\cdot 2.307692 + \tfrac{1}{20}\cdot 2.000000 + \tfrac{5}{4} = 1.119231 \\ x_2^{(2)} &= -\tfrac{2}{13}\cdot 1.250000 + \tfrac{2}{13}\cdot 2.000000 + \tfrac{30}{13} = 2.423077 \\ x_3^{(2)} &= -1.250000 - 2.307692 + 2 = -1.557692 \end{aligned}$$

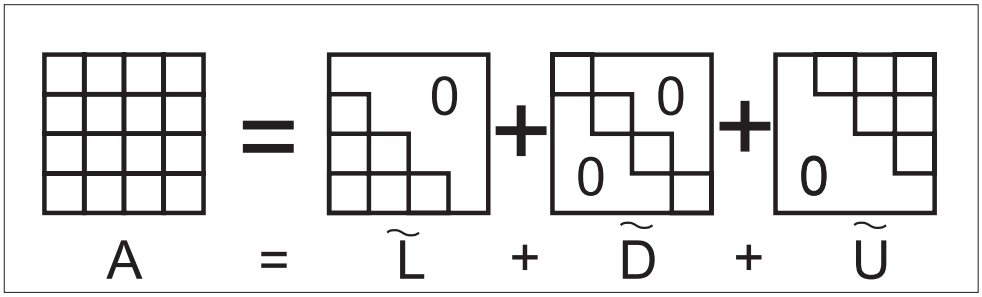
$$\begin{aligned} x_1^{(2)} &= -\tfrac{1}{10}\cdot 2.115385 - \tfrac{1}{20}\cdot 1.365385 + \tfrac{5}{4} = 0.970192 \\ x_2^{(2)} &= -\tfrac{2}{13}\cdot 0.970192 - \tfrac{2}{13}\cdot 1.365385 + \tfrac{30}{13} = 1.948373 \\ x_3^{(2)} &= -0.970192 - 1.948373 + 2 = -0.918565 \end{aligned}$$

	JACOBI			***GAUSS - SEIDEL***		
n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$
3	*0.929808*	*1.895858*	*-1.542308*	*1.009234*	*2.011108*	*-1.020342*
4	*0.983299*	*1.927367*	*-0.825666*	*0.997872*	*1.997198*	*-0.995070*
5	*1.015980*	*2.029390*	*-0.910666*	*1.000527*	*2.000677*	*-1.001204*
10	*0.999906*	*2.000106*	*-1.002296*	*0.999999*	*1.999999*	*-0.999999*
11	*0.999875*	*1.999661*	*-1.000013*	*1.000000*	*2.000000*	*-1.000000*

General algorithm

Matrix notation

Matrix decomposition



Simultaneous algebraic equations to be analyzed

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \rightarrow \quad \tilde{\mathbf{L}}\,\mathbf{x} + \tilde{\mathbf{D}}\,\mathbf{x} + \tilde{\mathbf{U}}\,\mathbf{x} = \mathbf{b}$$

Iteration algorithms

Jacobi

$$\mathbf{x}^{(n)} = -\,\tilde{\mathbf{D}}^{-1}\,(\tilde{\mathbf{L}} + \tilde{\mathbf{U}})\,\mathbf{x}^{(n-1)} + \tilde{\mathbf{D}}^{-1}\mathbf{b}$$

Gauss - Seidel

$$\mathbf{x}^{(n)} = -(\tilde{\mathbf{L}} + \tilde{\mathbf{D}})^{-1}\,\tilde{\mathbf{U}}\,\mathbf{x}^{(n-1)} + (\tilde{\mathbf{L}} + \tilde{\mathbf{D}})^{-1}\mathbf{b}$$

Remark : Inversion of the whole matrix $(\tilde{\mathbf{L}} + \tilde{\mathbf{D}})$ is not required

Index notation

$$\mathbf{A} = \{a_{ij}\}, \qquad \mathbf{b} = \{b_i\}, \qquad \mathbf{x} = \{x_i\}; \qquad i,j = 1, 2, \ldots, n$$

Simultaneous algebraic equations to be analyzed

$$\sum_{i=1}^{n} a_{ij} x_j = b_i$$

Iteration algorithms

Jacobi

$$x_i^{(n)} = \frac{1}{a_{ii}}\left[-\sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j^{(n-1)} + b_i \right]$$

Gauss – Seidel

$$x_i^{(n)} = \frac{1}{a_{ii}}\left[-\sum_{j=1}^{i-1} a_{ij} x_j^{(n)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(n-1)} + b_i \right]$$

$i = 1, 2, \ldots, n$

Theorem

When **A** is a positive definite matrix the Jacobi and Gauss – Seidel methods are convergent. (It is a sufficient but not necessary condition)

Relaxation technique

$$x_i^{(n)} = x_i^{(n-1)} + \mu \left[\hat{x}_i^{(n)} - x_i^{(n-1)} \right] = (1-\mu)\, x_i^{(n-1)} + \mu\, \hat{x}_i^{(n)}$$

$\hat{x}_i^{(n)}$ - Direct Gauss – Seidel result, *n*-th iteration

$x_i^{(n)}$ - relaxed solution, *n*-th iteration

$\mu > 0$ - relaxation parameter

Variable relaxation parameter $\mu^{(n-1)}$

Residuum

$$\hat{\mathbf{r}}^{(n-1)} = \hat{\mathbf{x}}^{(n)} - \mathbf{x}^{(n-1)}, \qquad \Delta\hat{\mathbf{r}}^{(n-1)} = \hat{\mathbf{r}}^{(n)} - \hat{\mathbf{r}}^{(n-1)}$$

let

$$\mathbf{r}^{(n)} = \hat{\mathbf{r}}^{(n-1)} + \mu^{(n-1)} \left(\hat{\mathbf{r}}^{(n)} - \hat{\mathbf{r}}^{(n-1)} \right) = \hat{\mathbf{r}}^{(n-1)} + \mu^{(n-1)}\ \Delta\hat{\mathbf{r}}^{(n-1)}$$

and

$$\mathbf{I} = \left(\mathbf{r}^{(n)}\right)^t \mathbf{r}^{(n)} = \left(\mathbf{r}^{(n-1)}\right)^t \hat{\mathbf{r}}^{(n-1)} + 2\mu^{(n-1)} \left(\mathbf{r}^{(n-1)}\right)^t \Delta\hat{\mathbf{r}}^{(n-1)} + \left(\mu^{(n-1)}\right)^2 \left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \Delta\hat{\mathbf{r}}^{(n-1)}$$

hence using the condition

$$\min_{\mu^{(n-1)}} \mathbf{I} \rightarrow \frac{\mathrm{d}\mathbf{I}}{\mathrm{d}\mu^{(n-1)}} = 2\left(\hat{\mathbf{r}}^{(n-1)}\right)^t \Delta\hat{\mathbf{r}}^{(n-1)} + 2\left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \Delta\hat{\mathbf{r}}^{(n-1)} + 2\mu^{(n-1)} \left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \left(\Delta\hat{\mathbf{r}}^{(n-1)}\right) = 0$$

find the optimal relaxation coefficient

$$\mu^{(n-1)} = \frac{\left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \hat{\mathbf{r}}^{(n-1)}}{\left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)} = 1 - \frac{\left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \hat{\mathbf{r}}^{(n)}}{\left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)^t \left(\Delta\hat{\mathbf{r}}^{(n-1)}\right)}$$

hence

$$\mathbf{x}_i^{(n)} = \hat{\mathbf{x}}_i^{(n-1)} + \mu^{(n-1)}\ \hat{\boldsymbol{r}}^{(n-1)}$$

Example continuation : Relaxation (using Gauss – Seidel)

Let $\mu = 0.8 = const$

$$x_1^{(2)} = 1.250000 + 0.8 \cdot (0.970192 - 1.250000) = 1.026154$$
$$x_2^{(2)} = 2.115385 + 0.8 \cdot (1.948373 - 2.115385) = 1.981775 \Rightarrow$$
$$x_3^{(2)} = -1.365385 + 0.8 \cdot (-0.918565 - 1.365385) = -1.007929$$

Further iterations

Gauss – Seidel followed by relaxation

$$\begin{Bmatrix} 1.001426 \\ 1.998561 \\ -0.999987 \end{Bmatrix} \overset{G.S.\ (3)}{\Rightarrow} \begin{Bmatrix} 1.006372 \\ 1.995204 \\ -1.001575 \end{Bmatrix} \overset{RELAX\ (3)}{\Rightarrow} \begin{Bmatrix} 1.000401 \\ 1.999695 \\ -1.000097 \end{Bmatrix} \overset{G.S.\ (4)}{\Rightarrow} \begin{Bmatrix} 1.001595 \\ 1.998798 \\ -1.000393 \end{Bmatrix} \overset{RELAX\ (4)}{\Rightarrow}$$

Example continuation : Relaxation (using Gauss – Seidel)

	Gauss – Seidel iteration			Relaxation (3')	**Gauss – Seidel iteration**		Relaxation (5')
	1	2	3		4	5	
x_1	*1.250000*	*0.970192*	*1.009234*	*1.002145*	*0.999935*	*1.000045*	*1.000032*
x_2	*2.115385*	*1.948373*	*2.011108*	*1.999716*	*1.999724*	*2.000046*	*2.000007*
x_3	*-1.365385*	*-0.918565*	*-1.020342*	*-1.001861*	*-0.999659*	*-1.000090*	*-1.000038*
$\tilde{r}_1$		*-0.279808*	*0.039042*		*-0.002210*	*0.000109*	
$\tilde{r}_2$		*-0.167012*	*0.062735*		*0.000008*	*0.000322*	
$\tilde{r}_3$		*0.446820*	*-0.101777*		*0.002202*	*-0.000431*	
$\Delta\tilde{r}_1$			*0.318850*			*0.002319*	
$\Delta\tilde{r}_2$			*0.229747*			*0.000314*	
$\Delta\tilde{r}_3$			*-0.548597*			*-0.002633*	
μ			*0.818412*			*0.879922*	

Error estimation

after the first step of iteration

Estimated relative solution error

$$\delta_1 = \frac{\|\mathbf{x}_1 - \mathbf{x}_0\|}{\|\mathbf{x}_1\|}$$
$$\mathbf{x}_1 - \mathbf{x}_0 = \{1.250000 - 0.0000,\ 2.115385 - 0.000000, -1.365385 - 0\}$$
$$= \{1.250000\ , 2.115385, -1.365385\}$$

Euclidean norm

$$\delta_1^E = \frac{\left[\frac{1}{3}\left(1.250000^2 + 2.115385^2 + (-1.365385)^2\right)\right]^{1/2}}{\left[\frac{1}{3}\left(1.250000^2 + 2.115385^2 + (-1.365385)^2\right)\right]^{1/2}} = \frac{1.622922}{1.622922} = 1.000000$$

Maximum norm

$$\delta_1^M = \frac{\sup\left\{1.250000\ , 2.115385, \left|-1.365385\right|\right\}}{\sup\left\{1.250000\ , 2.115385, \left|-1.365385\right|\right\}} = \frac{2.115385}{2.115385} = 1.000000$$

Relative residual error $r_n = \frac{\|\mathbf{F}_n\|}{\|\mathbf{F}_0\|}$

$$r_1 = \frac{\|\mathbf{F}_1\|}{\|\mathbf{F}_0\|}$$

Euclidean norm $\|\mathbf{F}\|_E = \left\{\frac{1}{n}\sum_{j=1}^{n} F_j(\mathbf{x})^2\right\}^{1/2}$

$$\mathbf{F}_1 = \{\text{-}5.596155\ ,\ \text{-}2.730775, 0.000000\}$$

$$\|\mathbf{F}_0\|_E = \left\{\frac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\frac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798$$

$$\|\mathbf{F}_1\|_E = \left\{\frac{1}{3}\left[(\text{-}5.596155)^2 + (\text{-}2.730775)^2 + (0.000000)^2\right]\right\}^{1/2} = 3.595093$$

$$r_1^E = \frac{3.595093}{22.575798} = 0.159245$$

Maximum norm $\|\mathbf{F}\|_M = \sup_i |F_i|$

$$\|\mathbf{F}_0\|_M = \sup(25, 30, 2) = 30$$

$$\|\mathbf{F}_1\|_M = \sup\left(\left|\text{-}5.596155\right|\ ,\ \left|\text{-}2.730775\right|, 0.000000\right) = 5.596155$$

$$r_1^M = \frac{5.596155}{30} = 0.186539$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\delta_1^E = 1.000000 \quad > \quad B_C = 10^{-6}$$
$$\delta_1^M = 1.000000 \quad > \quad B_C = 10^{-6}$$
$$r_1^E = 0.159245 \quad > \quad B_R = 10^{-8}$$
$$r_1^M = 0.186539 \quad > \quad B_R = 10^{-8}$$

after second step of iteration

Estimated relative solution error

$$\delta_2 = \frac{\|\mathbf{x}_2 - \mathbf{x}_1\|}{\|\mathbf{x}_2\|}$$

$$\mathbf{x}_2 - \mathbf{x}_1 = \{0.970192 - 1.250000, 1.948373 - 2.115385, -0.918565 + 1.365385\}$$
$$= \{-0.279808, -0.167012, 0.446820\}$$

Euclidean norm

$$\delta_2^E = \frac{\left[\frac{1}{3}\left((-0.279809)^2 + 0.167012^2 + 0.446820^2\right)\right]^{1/2}}{\left[\frac{1}{3}\left(0.970192^2 + 1.948373^2 + (-0.918565)^2\right)\right]^{1/2}} = \frac{0.319288}{1.363934} = 0.234088$$

Maximum norm

$$\delta_2^M = \frac{\sup\{|-0.279808|, |-0.167012|, 0.446820\}}{\sup\{0.970192, 1.948373, |-0.918565|\}} = \frac{0.446820}{1.948373} = 0.229330$$

Relative residual error

$$r_2 = \frac{\|\mathbf{F}_2\|}{\|\mathbf{F}_0\|}$$

Euclidean norm

$$\mathbf{F}_2 = \{0.780849\ ,\ 0.893637, 0.000000\}$$

$$\|\mathbf{F}_0\|_E = \left\{\frac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\frac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798$$

$$\|\mathbf{F}_2\|_E = \left\{\frac{1}{3}\left[(0.780849)^2 + (0.893637)^2 + (0.000000)^2\right]\right\}^{1/2} = 0.685155$$

$$r_2^E = \frac{0.685155}{22.575798} = 0.030349$$

Maximum norm

$$\|\mathbf{F}_0\|_M = \sup(25, 30, 2) = 30$$

$$\|\mathbf{F}_2\|_M = \sup(0.780849\ ,\ 0.893637, 0.000000) = 0.893637$$

$$r_2^M = \frac{0.893637}{30} = 0.029788$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\begin{aligned}
\delta_2^E &= 0.234088 \quad > \quad B_C = 10^{-6} \\
\delta_2^M &= 0.229330 \quad > \quad B_C = 10^{-6} \\
r_2^E &= 0.030349 \quad > \quad B_R = 10^{-8} \\
r_2^M &= 0.029788 \quad > \quad B_R = 10^{-8}
\end{aligned}$$

after third step of iteration

Estimated relative solution error

$$\delta_3 = \frac{\|\mathbf{x}_3 - \mathbf{x}_2\|}{\|\mathbf{x}_3\|}$$

$$\begin{aligned}
\mathbf{x}_3 - \mathbf{x}_2 &= \{1.009234 - 0.970192, 2.011108 - 1.948373, -1.020342 + 0.918565\} \\
&= \{0.039042, 0.062735, -0.101777\}
\end{aligned}$$

Euclidean norm

$$\delta_3^E = \frac{\left[\frac{1}{3}\left(0.039042^2 + 0.062735^2 + (-0.101777)^2\right)\right]^{1/2}}{\left[\frac{1}{3}\left(1.009234^2 + 2.011108^2 + (-1.020342)^2\right)\right]^{1/2}} = \frac{0.072614}{1.426442} = 0.050906$$

Maximum norm

$$\delta_3^M = \frac{\sup\{0.039042, 0.062735, |-0.101777|\}}{\sup\{1.009234, 2.011108, |-1.020342|\}} = \frac{0.101777}{2.011108} = 0.050608$$

Relative residual error

$$r_3 = \frac{\|\mathbf{F}_3\|}{\|\mathbf{F}_0\|}$$

Euclidean norm

$$\mathbf{F}_3 = \{-0.227238, -0.203556, 0.000000\}$$

$$\|\mathbf{F}_0\|_E = \left\{\frac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\frac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798$$

$$\|\mathbf{F}_3\|_E = \left\{\frac{1}{3}\left[(-0.227247)^2 + (-0.203554)^2 + 0.000000^2\right]\right\}^{1/2} = 0.176140$$

$$r_3^E = \frac{0.176140}{22.575798} = 0.007802$$

Maximum norm

$$\|\mathbf{F}_0\|_M = \sup(25, 30, 2) = 30$$
$$\|\mathbf{F}_3\|_M = \sup(|-0.227238|, |-0.203556|, 0.000000) = 0.227238$$
$$r_3^M = \frac{0.227238}{30} = 0.007575$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\delta_3^E = 0.050906 > B_C = 10^{-6}$$
$$\delta_3^M = 0.050608 > B_C = 10^{-6}$$
$$r_3^E = 0.007802 > B_R = 10^{-8}$$
$$r_3^M = 0.007575 > B_R = 10^{-8}$$

after third step of iteration and relaxation

Estimated relative solution error

$$\delta_{3'} = \frac{\|\mathbf{x}_{3'} - \mathbf{x}_2\|}{\|\mathbf{x}_{3'}\|}$$
$$\mathbf{x}_{3'} - \mathbf{x}_2 = \{1.002145 - 0.970192,\ 1.999716 - 1.948373,\ -1.001861 + 0.918565\}$$
$$= \{0.031953, 0.051343, -0.083296\}$$

Euclidean norm

$$\delta_{3'}^E = \frac{\left[\frac{1}{3}\left(0.031953^2 + 0.051343^2 + (-0.083296)^2\right)\right]^{1/2}}{\left[\frac{1}{3}\left(1.002145^2 + 1.999716^2 + (-1.001861)^2\right)\right]^{1/2}} = \frac{0.059429}{1.415025} = 0.041998$$

Maximum norm

$$\delta_{3'}^M = \frac{\sup\{0.031953, 0.051343, |-0.083296|\}}{\sup\{1.002145, 1.999716, |-1.001861|\}} = \frac{0.083296}{1.999716} = 0.041654$$

Relative residual error

$$r_{3'} = \frac{\|\mathbf{F}_{3'}\|}{\|\mathbf{F}_0\|}$$

Euclidean norm

$$\mathbf{F}_{3'} = \{0.044190,\ 0.004317,\ 0.000000\}$$

$$\|\mathbf{F}_0\|_E = \left\{\tfrac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\tfrac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798$$

$$\|\mathbf{F}_{3'}\|_E = \left\{\tfrac{1}{3}\left[0.044190^2 + 0.004317^2 + 0.000000^2\right]\right\}^{1/2} = 0.025635$$

$$r_{3'}^E = \frac{0.025635}{22.575798} = 0.001135$$

Maximum norm

$$\|\mathbf{F}_0\|_M = \sup(25, 30, 2) = 30$$

$$\|\mathbf{F}_{3'}\|_M = \sup(0.044190\ ,\ 0.004317, 0.000000) = 0.044190$$

$$r_{3'}^M = \frac{0.044190}{30} = 0.001473$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\delta_{3'}^E = 0.041998 \quad > \quad B_C = 10^{-6}$$

$$\delta_{3'}^M = 0.041654 \quad > \quad B_C = 10^{-6}$$

$$r_{3'}^E = 0.001135 \quad > \quad B_R = 10^{-8}$$

$$r_{3'}^M = 0.001473 \quad > \quad B_R = 10^{-8}$$

after fourth step of iteration

Estimated relative solution error

$$\delta_4 = \frac{\|\mathbf{x}_4 - \mathbf{x}_{3'}\|}{\|\mathbf{x}_4\|}$$

$$\mathbf{x}_4 - \mathbf{x}_{3'} = \{0.999935 - 1.002145,\ 1.999724 - 1.999716,\ -0.9996590 + 1.001861\}$$

$$= \{-0.002210,\ 0.000008,\ 0.002202\}$$

Euclidean norm

$$\delta_4^E = \frac{\left[\tfrac{1}{3}\left((-0.002210)^2 + 0.000008^2 + 0.002202)^2\right)\right]^{1/2}}{\left[\tfrac{1}{3}\left(0.999935^2 + 1.999724^2 + (-0.999659)^2\right)\right]^{1/2}} = \frac{0.001801}{1.413988} = 0.001274$$

Maximum norm

$$\delta_4^M = \frac{\sup\{|-0.002210|, 0.000008, 0.002202\}}{\sup\{0.999935\ ,\ 1.999724, |-0.999659|\}} = \frac{0.002210}{1.999724} = 0.001105$$

Relative residual error

$$r_4 = \frac{\|\mathbf{F}_4\|}{\|\mathbf{F}_0\|}$$

Euclidean norm

$$\mathbf{F}_4 = \{-0.002186,\ -0.004403,\ 0.000000\}$$

$$\|\mathbf{F}_0\|_E = \left\{\tfrac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\tfrac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798$$

$$\|\mathbf{F}_4\|_E = \left\{\tfrac{1}{3}\left[(-0.002186)^2 + (-0.004403)^2 + 0.000000^2\right]\right\}^{1/2} = 0.002838$$

$$r_4^E = \frac{0.002838}{22.575798} = 0.000126$$

Maximum norm

$$\|\mathbf{F}_0\|_M = \sup(25, 30, 2) = 30$$

$$\|\mathbf{F}_4\|_M = \sup(|-0.002193|\ ,\ |-0.004400|, 0.000000) = 0.004400$$

$$r_4^M = \frac{0.004403}{30} = 0.000147$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\delta_4^E = 0.001274 \quad > \quad B_C = 10^{-6}$$

$$\delta_4^M = 0.001105 \quad > \quad B_C = 10^{-6}$$

$$r_4^E = 0.000126 \quad > \quad B_R = 10^{-8}$$

$$r_4^M = 0.000147 \quad > \quad B_R = 10^{-8}$$

after the fifth step of iteration

Estimated relative solution error

$$\delta_5 = \frac{\left\|\mathbf{x}_5 - \mathbf{x}_4\right\|}{\left\|\mathbf{x}_5\right\|}$$

$$\mathbf{x}_5 - \mathbf{x}_4 = \{1.000045 - 0.999935,\ 2.000046 - 1.999724,\ -1.000090 + 0.999659\}$$
$$= \{0.000109,\ 0.000322, -0.000431\}$$

Euclidean norm

$$\delta_5^E = \frac{\left[\frac{1}{3}\left(0.000109^2 + 0.000322^2 + (-0.000431)^2\right)\right]^{1/2}}{\left[\frac{1}{3}\left(1.000045^2 + 2.000046^2 + (-1.000090)^2\right)\right]^{1/2}} = \frac{0.000317}{1.414267} = 0.000224$$

Maximum norm

$$\delta_5^M = \frac{\sup\{|0.000109|, |0.000322|, |-0.000431|\}}{\sup\{1.000045, 2.000046, |-1.000090|\}} = \frac{0.000431}{2.000046} = 0.000215$$

Relative residual error

$$r_5 = \frac{\left\|\mathbf{F}_5\right\|}{\left\|\mathbf{F}_0\right\|}$$

Euclidean norm

$$\mathbf{F}_5 = \{0.001075,\ 0.000862,\ 0.000000\}$$

$$\left\|\mathbf{F}_0\right\|_E = \left\{\frac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\frac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798$$

$$\left\|\mathbf{F}_5\right\|_E = \left\{\frac{1}{3}\left[0.001075^2 + 0.000862^2 + 0.000000^2\right]\right\}^{1/2} = 0.000796$$

$$r_5^E = \frac{0.000796}{22.575798} = 0.000035$$

Maximum norm

$$\left\|\mathbf{F}_0\right\|_M = \sup(25, 30, 2) = 30$$

$$\left\|\mathbf{F}_5\right\|_M = \sup(0.001075,\ 0.000862, 0.000002) = 0.001075$$

$$r_5^M = \frac{0.001075}{30} = 0.000036$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\begin{aligned}
\delta_5^E &= 0.000224 \quad > \quad B_C = 10^{-6} \\
\delta_5^M &= 0.000215 \quad > \quad B_C = 10^{-6} \\
r_5^E &= 0.000035 \quad > \quad B_R = 10^{-8} \\
r_5^M &= 0.000036 \quad > \quad B_R = 10^{-8}
\end{aligned}$$

after fifth step of iteration and relaxation

Estimated relative solution error

$$\delta_{5'} = \frac{\left\|\mathbf{x}_{5'} - \mathbf{x}_4\right\|}{\left\|\mathbf{x}_{5'}\right\|}$$

$$\begin{aligned}
\mathbf{x}_{5'} - \mathbf{x}_4 &= \{1.000032 - 0.999935,\ 2.000007 - 1.999724,\ -1.000038 + 0.999659\} \\
&= \{0.000097,\ 0.000283,\ -0.000379\}
\end{aligned}$$

Euclidean norm

$$\delta_{5'}^E = \frac{\left[\frac{1}{3}\left(0.000097^2 + 0.00028^2 + (-0.000379)^2\right)\right]^{1/2}}{\left[\frac{1}{3}\left(1.000032^2 + 2.000007^2 + (-1.0000038)^2\right)\right]^{1/2}} = \frac{0.000279}{1.414233} = 0.000197$$

Maximum norm

$$\delta_{5'}^M = \frac{\sup\{0.000097\ , 0.000283, |-0.000379|\}}{\sup\{1.000032\ , 2.000007, |-1.000038|\}} = \frac{0.000379}{2.000007} = 0.000190$$

Relative residual error

$$r_{5'} = \frac{\left\|\mathbf{F}_{5'}\right\|}{\left\|\mathbf{F}_0\right\|}$$

Euclidean norm

$$\begin{aligned}
\mathbf{F}_{5'} &= \{0.000683,\ 0.000230,\ 0.000000\} \\
\left\|\mathbf{F}_0\right\|_E &= \left\{\frac{1}{3}\left[F_1(\mathbf{x}_0)^2 + F_2(\mathbf{x}_0)^2 + F_3(\mathbf{x}_0)^2\right]\right\}^{1/2} = \left\{\frac{1}{3}\left[25^2 + 30^2 + 2^2\right]\right\}^{1/2} = 22.575798 \\
\left\|\mathbf{F}_{5'}\right\|_E &= \left\{\frac{1}{3}\left[0.000683^2 + 0.000230^2 + 0.000000^2\right]\right\}^{1/2} = 0.000416
\end{aligned}$$

$$r_{5'}^E = \frac{0.000416}{22.575798} = 0.000018$$

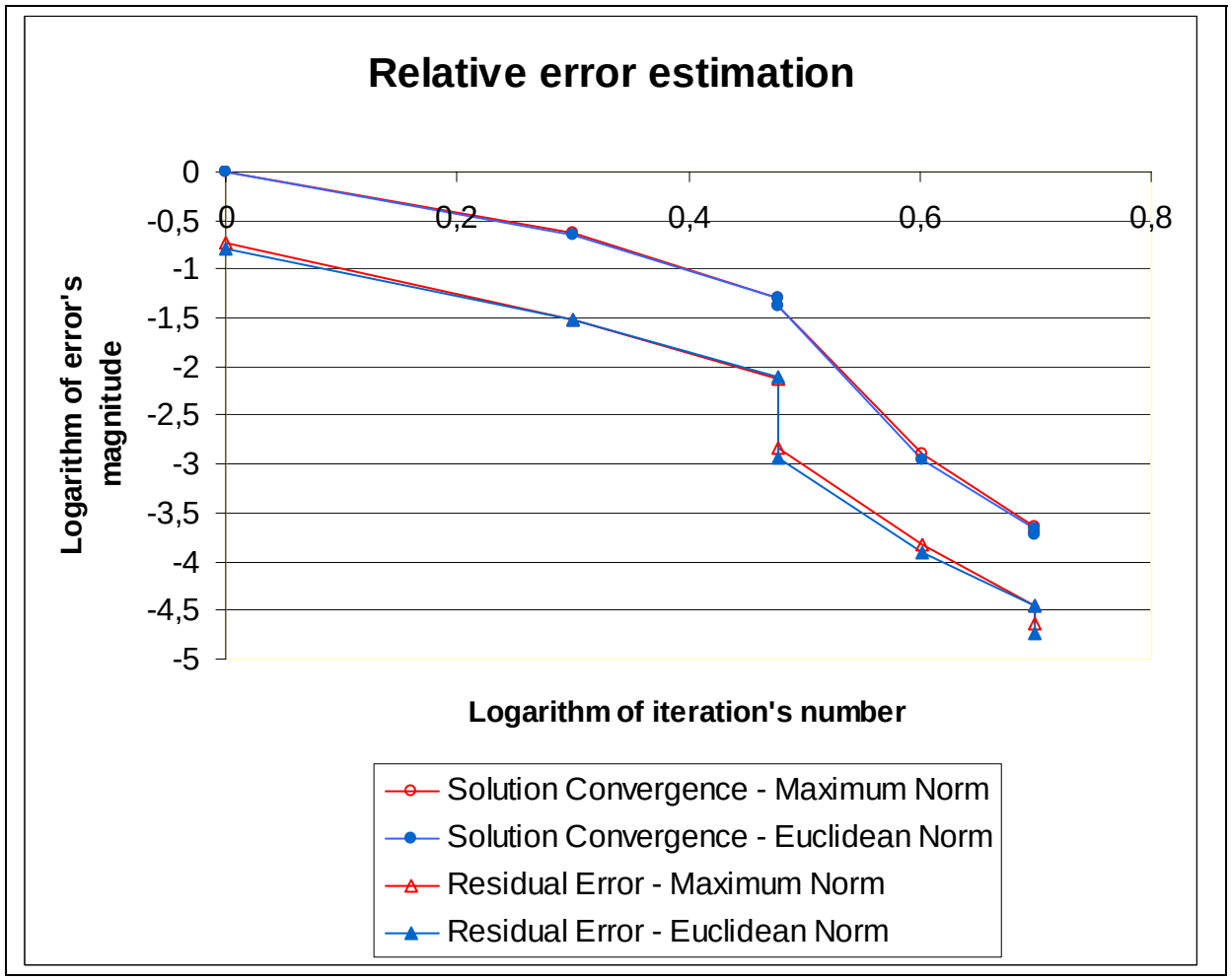
Maximum norm

$$\|\mathbf{F}_0\|_M = \sup(25, 30, 2) = 30$$
$$\|\mathbf{F}_{5'}\|_M = \sup(0.00683\ ,\ 0.000230, 0.000000) = 0.000683$$
$$r_{5'}^M = \frac{0.000683}{30} = 0.000023$$

Brake-off test

Assume admissible errors for convergence $\mathbf{B_C}$ and residuum $\mathbf{B_R}$; check

$$\delta_5^E = 0.000197 \quad > \quad B_C = 10^{-6}$$
$$\delta_5^M = 0.000190 \quad > \quad B_C = 10^{-6}$$
$$r_5^E = 0.000018 \quad > \quad B_R = 10^{-8}$$
$$r_5^M = 0.000023 \quad > \quad B_R = 10^{-8}$$



5.6. MATRIX FACTORIZATION LU BY THE GAUSSIAN ELIMINATION

A = LU

The **LU** factorization of matrix **A m**ay be done by the Gaussian elimination approach. The main difference between the Gauss procedures of the solution of the SLAE and matrix factorization LU is that in the last case we have to store the multipliers $\{m_{ij}\}$. Application:

- solution of problems with multiple right hand side
- matrix inversion

Example

$$\begin{bmatrix} 1 & 1 & 2 & -4 \\ 2 & -1 & 3 & 1 \\ 3 & 1 & -1 & 2 \\ 1 & -1 & -1 & -1 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 1 & 2 & -4 \\ \textcircled{2} & -3 & -1 & 9 \\ \textcircled{3} & -2 & -7 & 14 \\ \textcircled{1} & -2 & -3 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 & -4 \\ \textcircled{2} & -3 & -1 & 9 \\ \textcircled{3} & \textcircled{2/3} & -19/3 & 8 \\ \textcircled{1} & \textcircled{2/3} & -7/3 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 & -4 \\ \textcircled{2} & -3 & -1 & 9 \\ \textcircled{3} & \textcircled{2/3} & -19/3 & 8 \\ \textcircled{1} & \textcircled{2/3} & \textcircled{7/19} & -75/19 \end{bmatrix}$$

m_{21}, m_{31}, m_{41} — m_{32}, m_{42} — m_{43}

Then

$$\begin{bmatrix} 1 & 1 & 2 & -4 \\ 2 & -1 & 3 & 1 \\ 3 & 1 & -1 & 2 \\ 1 & -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2/3 & 1 & 0 \\ 1 & 2/3 & 7/19 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & -4 \\ 0 & -3 & -1 & 9 \\ 0 & 0 & -19/3 & 8 \\ 0 & 0 & 0 & -75/19 \end{bmatrix}$$

A = L U

Generally

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ m_{21} & 1 & \dots & 0 \\ \vdots & & & \vdots \\ m_{n1} & m_{n2} & \dots & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & & & \vdots \\ 0 & 0 & \dots & u_{nn} \end{bmatrix}$$

5.7. MATRIX INVERSION

5.7.1. Inversion of squared matrix using Gaussian Elimination

$$[\mathbf{A} \vdots \mathbf{I}] \rightarrow [\mathbf{I} \vdots \mathbf{A}^{-1}]$$

Example

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \rightarrow$$

$$\rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 3/2 & 1/2 & -1/2 & 1 & 0 \\ 0 & 1/2 & 3/2 & -1/2 & 0 & 1 \end{array}\right] \rightarrow$$

$$\rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 3/2 & 1/2 & -1/2 & 1 & 0 \\ 0 & 0 & 4/3 & -1/3 & -1/3 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 3/2 & 1/2 & -1/2 & 1 & 0 \\ 0 & 0 & 1 & -1/4 & -1/4 & 3/4 \end{array}\right] \rightarrow$$

$$\rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 3/2 & 0 & -3/8 & 7/8 & 0 \\ 0 & 0 & 1 & -1/4 & -1/4 & 3/4 \end{array}\right] \rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 0 & 5/4 & 1/4 & -3/4 \\ 0 & 3/2 & 0 & -3/8 & 9/8 & -3/8 \\ 0 & 0 & 1 & -1/4 & -1/4 & 3/4 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc:ccc} 2 & 1 & 0 & 5/4 & 1/4 & -3/4 \\ 0 & 1 & 0 & -1/4 & 3/4 & -1/4 \\ 0 & 0 & 1 & -1/4 & -1/4 & 3/4 \end{array}\right] \rightarrow \left[\begin{array}{ccc:ccc} 2 & 0 & 0 & 3/2 & -1/2 & -1/2 \\ 0 & 1 & 0 & -1/4 & 3/4 & -1/4 \\ 0 & 0 & 1 & -1/4 & -1/4 & 3/4 \end{array}\right] \rightarrow$$

$$\rightarrow \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 3/4 & -1/4 & -1/4 \\ 0 & 1 & 0 & -1/4 & 3/4 & -1/4 \\ 0 & 0 & 1 & -1/4 & -1/4 & 3/4 \end{array}\right]$$

Algorithm

$$\mathbf{A\,C = I} \quad , \quad \mathbf{A} = [a_{ij}] \quad , \quad \mathbf{C} = [c_{ij}], \quad \text{where } \mathbf{C_0 = I}$$

I. Step forward

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - m_{ik} a_{kj}^{(k-1)}, \qquad m_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \qquad k = 1, 2, \ldots, n\text{-}1; \quad i, j = k+1, \ldots, n;$$

$$c_{ij}^{(k)} = c_{ij}^{(k-1)} - m_{ik} c_{kj}^{(k-1)} \qquad\qquad j = 1, 2, \ldots, n;$$

II. Step back

$$a_{kk}^{(k-1)} = 1, \quad a_{ik}^{(k-1)} = 0, \qquad k = n, n-1, \ldots, 2; \quad i = k-1, k-2, \ldots, 1;$$

$$c_{kj}^{(k-1)} = c_{kj}^{(k)} \frac{1}{a_{kk}^{(k)}}$$

$$c_{ij}^{(k-1)} = c_{ij}^{(k)} - c_{kj}^{(k)} \underbrace{\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}}_{m_{ik}} \qquad j = 1, 2, \ldots, n;$$

5.7.2. Inversion of the lower triangular matrix

Example

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix}, \qquad \mathbf{C} \overset{?}{=} \begin{bmatrix} c_{11} & 0 & 0 \\ c_{21} & c_{22} & 0 \\ c_{31} & c_{32} & c_{33} \end{bmatrix} = \mathbf{L}^{-1} \quad , \quad \mathbf{LC=I}$$

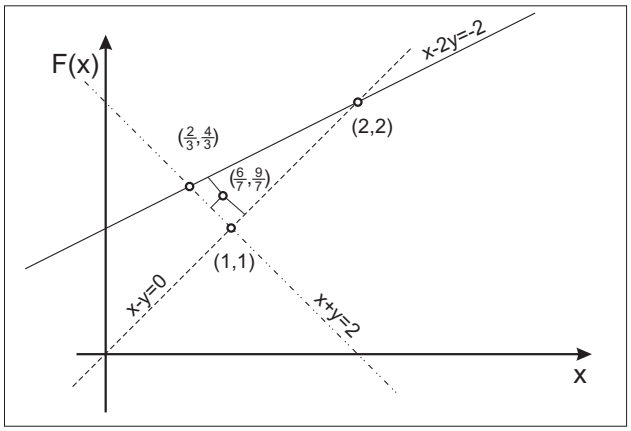
$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix} \begin{bmatrix} c_{11} & 0 & 0 \\ c_{21} & c_{22} & 0 \\ c_{31} & c_{32} & c_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\left.\begin{array}{lllllllll}
c_{11}\times 1 & + & c_{21}\times 0 & + & c_{31}\times 0 & = & 1 & \rightarrow & c_{11} = 1 \\
c_{11}\times 2 & + & c_{21}\times 4 & + & c_{31}\times 0 & = & 0 & \rightarrow & c_{21} = -\frac{1}{2} \\
c_{11}\times 3 & + & c_{21}\times 5 & + & c_{31}\times 6 & = & 0 & \rightarrow & c_{31} = -\frac{1}{12} \\
0\times 2 & + & c_{22}\times 4 & + & c_{32}\times 0 & = & 1 & \rightarrow & c_{22} = \frac{1}{4} \\
0\times 3 & + & c_{22}\times 5 & + & c_{32}\times 6 & = & 0 & \rightarrow & c_{32} = -\frac{5}{24} \\
0\times 3 & + & 0\times 5 & + & c_{33}\times 6 & = & 1 & \rightarrow & c_{33} = \frac{1}{6}
\end{array}\right\} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1/4 & 0 \\ -1/12 & -5/24 & 1/6 \end{bmatrix}$$

Algorithm

$$\boxed{\begin{array}{l} c_{ii} = \dfrac{1}{l_{ii}} \\ c_{ij} = -\dfrac{1}{l_{ii}} \sum_{k=j}^{i-1} l_{ik} c_{kj} \end{array}} \qquad \begin{array}{l} i = 1,2,\ldots,n \\ j = 1,2,\ldots,i-1; \qquad k = j, j+1,\ldots,i-1 \end{array}$$

5.8. OVERDETERMINED SIMULTANEOUS LINEAR EQUATIONS



$$x + y = 2$$
$$x - y = 0$$
$$x - 2y = -2$$

THE LEAST SQUARES METHOD

Use of the Euclidean error norm

Let $$B = (x + y - 2)^2 + (x - y)^2 + (x - 2y + 2)^2$$

$$\left.\begin{aligned} \frac{\partial B}{\partial x} &= 2(x + y - 2) + 2(x - y) + 2(x - 2y + 2) = 0 \quad \rightarrow \\ \frac{\partial B}{\partial y} &= 2(x + y - 2) - 2(x - y) - 2 \cdot 2(x - 2y + 2) = 0 \rightarrow \end{aligned}\right\} \quad \begin{aligned} 3x - 2y &= 0 \\ -2x + 6y &= 6 \end{aligned}$$

solution $x = 6/7$, $y = 9/7$ $\rightarrow$ $B = \left(\frac{1}{7}\right)^2 + \left(-\frac{3}{7}\right)^2 + \left(\frac{2}{7}\right)^2 = \frac{2}{7}$

General approach

Index notation

$$\sum_{j=1}^{m} a_{ij} x_j = b_i , \qquad i = 1,2,\cdots,n ; \qquad j = 1,2,\cdots,m \qquad m < n$$

where m – number of unknows
n – number of equations

In the above example m=2, n=3.

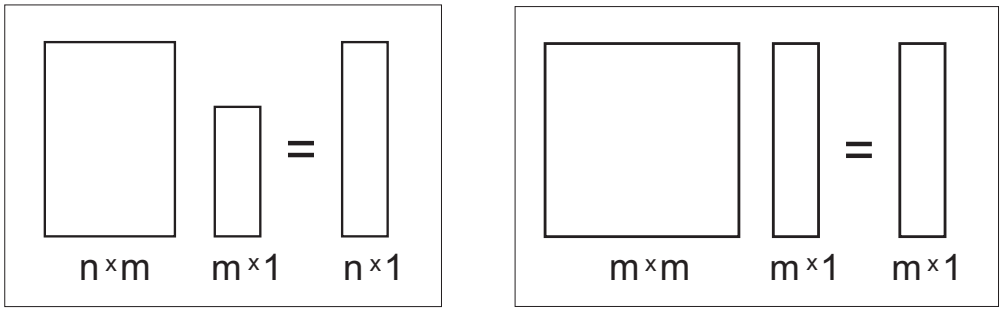
$$B = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} x_j - b_i \right)^2$$

$$\frac{\partial B}{\partial x_k} = 2\sum_{i=1}^{n} a_{ik}\left(\sum_{j=1}^{m} a_{ij}x_j - b_i\right) = 0 \rightarrow \sum_{i=1}^{n} a_{ik}\sum_{j=1}^{m} a_{ij}x_j = \sum_{i=1}^{n} a_{ik}b_i\,, \quad k=1, \ldots, m$$

Matrix notation

$$\underset{n\times m}{\mathbf{A}}\ \underset{m\times 1}{\mathbf{x}} = \underset{n\times 1}{\mathbf{b}} \ \rightarrow \ B = \left(\mathbf{Ax}-\mathbf{b}\right)^t\left(\mathbf{Ax}-\mathbf{b}\right)$$

$$\frac{\partial B}{\partial \mathbf{x}} = 2\mathbf{A}^t\left(\mathbf{Ax}-\mathbf{b}\right) = \mathbf{0} \ \rightarrow \ \boxed{\mathbf{A}^t\mathbf{Ax} = \mathbf{A}^t\mathbf{b}}$$



Example

Once more the same example as before, but posed now in the matrix notation

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -2 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix},$$

$$\mathbf{A}^t\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -2 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

Pseudo solution by means of the least squares method (LSM)

$$\begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \rightarrow \mathbf{x} = \begin{bmatrix} 6/7 \\ 9/7 \end{bmatrix}$$

A weighted LSM may be also considered.

Matrix notation

$$B = \left(\mathbf{Ax} - \mathbf{b}\right)^{\mathbf{t}} \mathbf{W} \left(\mathbf{Ax} - \mathbf{b}\right)$$

hence

$$\mathbf{A^t W A x = A^t W b}$$

where

$$\mathbf{W} = diag\left(w_1, w_2, \ldots, w_n\right) = \lfloor w_1, \ldots, w_n \rceil$$

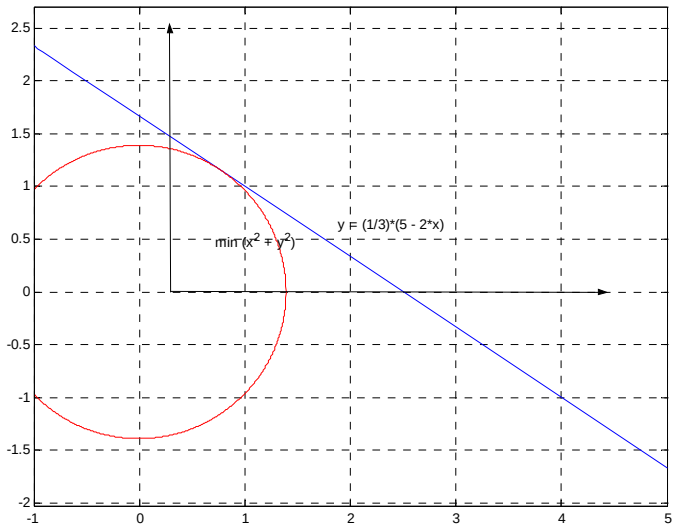
Index notation

$$B = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} - b_i \right)^2 w_i$$

$$\frac{\partial B}{\partial x_k} = 0 \quad , \quad k = 1, \ldots, m \quad \rightarrow \quad \sum_{i=1}^{n} a_{ik} w_i \sum_{j=1}^{m} a_{ij} x_j = \sum_{i=1}^{n} a_{ik} w_i b_i$$

5.9 UNDERDETERMINED SLAE – MINIMUM LENGTH METHOD

CASE: LESS EQUATIONS THAN UNKNOWNS



SOLUTION APPROACH: MINIMUM LENGTH METHOD (MLM)

Introductory example

Find

$$\min_{x,y} \rho^2 \qquad , \qquad \rho^2 = x^2 + y^2$$

when

$$2x + 3y = 5$$

(i) Elimination

$$y = \frac{1}{3}(5 - 2x)$$

hence

$$\rho^2 = x^2 + \frac{1}{9}(5 - 2x)^2$$

find

$$\min_{x} \rho^2 \qquad , \qquad \rho^2 = x^2 + \frac{1}{9}(5 - 2x)^2$$

$$\frac{d}{dx}\rho^2 = 2x - \frac{4}{9}(5 - 2x) = 0 \quad \rightarrow \quad x = \frac{10}{13} \quad , \quad y = \frac{15}{13}$$

(ii) Lagrange multipliers approach

$$I = (x^2 + y^2) - \lambda(2x + 3y - 5)$$

$$\left.\begin{aligned} \frac{\partial I}{\partial x} &= 2x - 2\lambda = 0 \\ \frac{\partial I}{\partial y} &= 2y - 3\lambda = 0 \\ \frac{\partial I}{\partial \lambda} &= -(2x + 3y - 5) = 0 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} x &= \frac{10}{13} \\ y &= \frac{15}{13} \\ \lambda &= \frac{10}{13} \end{aligned}\right.$$

GENERAL LINEAR CASE

Find

$$\min_{x_1,\ldots,x_n} \rho^2 \qquad , \qquad \rho^2 = \sum_{i=1}^{n} x_i^2$$

when

$$\underset{mxn}{\mathbf{A}}\ \underset{nx1}{\mathbf{x}} = \underset{mx1}{\mathbf{b}} \quad , \quad m < n \quad , \qquad \text{linear constraints}$$

SOLUTION BY THE ELIMINATION APPROACH

let

$$\underset{m\times n}{\mathbf{A}} = \left[\underset{m\times m}{\mathbf{A}}, \quad \underset{m\times(n-m)}{\mathbf{A}}\right] \quad , \quad \underset{n\times 1}{\mathbf{x}} = \left\{\begin{matrix} \underset{m\times 1}{\mathbf{x}} \\ \text{eliminated unknowns} \end{matrix} \quad , \quad \begin{matrix} \underset{(n-m)\times 1}{\mathbf{x}} \\ \text{remaining unknowns} \end{matrix}\right\}$$

hence

$$\underset{m\times n}{\mathbf{A}}\ \underset{n\times 1}{\mathbf{x}} = \underset{m\times m}{\mathbf{A}}\ \underset{m\times 1}{\mathbf{x}} + \underset{m\times(n-m)}{\mathbf{A}}\ \underset{(n-m)\times 1}{\mathbf{x}} = \underset{m\times 1}{\mathbf{b}}$$

$$\underset{m\times 1}{\mathbf{x}} = \underset{m\times m}{\mathbf{A}}^{-1}\left(\underset{m\times 1}{\mathbf{b}} + \underset{m\times(n-m)}{\mathbf{A}}\ \underset{(n-m)\times 1}{\mathbf{x}}\right) \qquad \text{eliminated unknowns,} \qquad (*)$$

and

$$\rho^2 = \sum_{i=1}^{m} x_i^2(x_{n-m},\ldots,x_n) + \sum_{i=n-m+1}^{n} x_i^2 = \rho^2(x_{n-m+1},\ldots,x_n)\,.$$

Finally we find in two steps the solution of the minimization problem

$$\min_{x_{n-m+1},\ldots,x_n} \rho^2(x_{n-m+1},\ldots,x_n)$$

- *step 1* – use of the optimality conditions

$$\frac{\partial \rho^2}{\partial x_k} = 0 \quad , \qquad \text{for} \quad k = n-m+1, n-m+2, \ldots, n$$

hence we obtain the first part of the unknowns $x_{n-m+1},\ldots,x_n$

- *step 2* – use of the elimination formulas (*); they provide the remaining unknowns $x_1, \ldots, x_m$

Example

Given undeterminen SLAE

$$2x + 3y - z = 4$$
$$-x + 4y - 2z = -4$$

Solution by the minimum length approach

Find

$$\min_{x,y,z} \rho^2 \quad , \quad \rho^2 = x^2 + y^2 + z^2$$

when

$$\begin{bmatrix} 2 & 3 & -1 \\ -1 & 4 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$$

Hence

$$\underset{2x3}{\mathbf{A}} = \begin{bmatrix} 2 & 3 & |-1 \\ -1 & 4 & |-2 \end{bmatrix} \rightarrow \underset{2x2}{\mathbf{A}} = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix} \quad , \quad \underset{2x1}{\mathbf{A}} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

and

$$\mathbf{x} = \{x \quad y \quad | \quad z\} \quad , \quad \mathbf{b} = \{4 \quad -4\}$$

Solution process

$$\underset{2\,x\,2}{\mathbf{A}}^{-1} = \frac{1}{11} \begin{bmatrix} 4 & -3 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 4 & -3 \\ 1 & 2 \end{bmatrix} \left(\begin{bmatrix} 4 \\ -4 \end{bmatrix} - \begin{bmatrix} -1 \\ -2 \end{bmatrix} [z] \right) = \frac{1}{11} \left(\begin{bmatrix} 28 \\ -4 \end{bmatrix} + \begin{bmatrix} -2 \\ 5 \end{bmatrix} [z] \right)$$

elimination and minimization

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 28 \\ -4 \\ 0 \end{bmatrix} + \frac{1}{11} \begin{bmatrix} -2 \\ 5 \\ 11 \end{bmatrix} [z]$$

$$\rho^2 = [x \quad y \quad z] \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{11^2} [(28 - 2z)^2 + (-4 + 5z)^2 + (11z)^2]$$

$$\frac{d\rho^2}{dz} = \frac{2}{121}[-2(28-2z)+5(-4+5z)+121z] = 0 \quad \rightarrow \quad z = \frac{1.52}{3}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 28 \\ -4 \end{bmatrix} + \frac{1}{11}\begin{bmatrix} -2 \\ 5 \end{bmatrix}\frac{1.52}{3} = \frac{1}{825}\begin{bmatrix} 2024 \\ -110 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 7.36 \\ -0.40 \end{bmatrix}$$

Finally

$$\{x \quad y \quad z\} = \frac{1}{3}\{7.36\,, -0.40\,, 1.52\} = \begin{bmatrix} 2.453333 \\ -0.133333 \\ 0.506667 \end{bmatrix}$$

6. THE ALGEBRAIC EIGENVALUE PROBLEM

6.1. INTRODUCTION

Application in mechanics : principal stresses, principal strains, dynamics, buckling, ...

Formulation

$$\mathbf{Ax} = \lambda\mathbf{x} \qquad \text{where} \qquad \mathbf{A} \quad \text{real } n \times n \quad \text{matrix}, \quad \mathbf{x} \in \Re^n$$

Find non-trivial solution $\rightarrow \det(\mathbf{A} - \lambda\mathbf{I}) = 0 \rightarrow \lambda$

$$\mathbf{A}\,\mathbf{x}_j = \lambda_j\,\mathbf{x}_j, \quad j=1, 2, 3, \ldots, n$$

where

$\lambda_1, \ldots, \lambda_n$ - eigenvalues

$\mathbf{x}_1, \ldots, \mathbf{x}_n$ - eigenvectors

Example

$$\begin{array}{rcrcl} 2\mathrm{x} & + & \mathrm{y} & = & \lambda\mathrm{x} \\ \mathrm{x} & + & 2\mathrm{y} & = & \lambda\mathrm{y} \end{array} \rightarrow \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ \mathrm{y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

eigenvalues evaluation

$$\det \mathbf{A} = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = 0 \quad \rightarrow \quad \lambda_1 = 1, \;\; \lambda_2 = 3$$

Let

$\lambda_1 = 3$

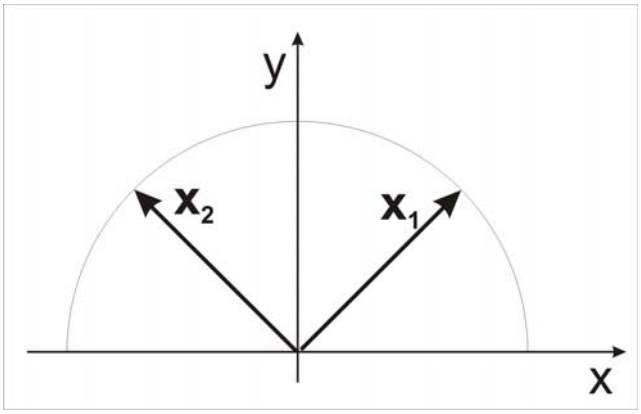
$$\begin{array}{l} 2x_1 + y_1 = 3x_1 \\ x_1 + 2y_1 = 3y_1 \end{array} \rightarrow \begin{array}{l} -x_1 + y_1 = 0 \\ x_1 - y_1 = 0 \end{array} \quad \textit{Let} \quad \sqrt{x_1^2 + y_1^2} = 1 \quad \begin{cases} x_1 = \dfrac{1}{\sqrt{2}} \\ y_1 = \dfrac{1}{\sqrt{2}} \end{cases}$$

$\lambda_2 = 1$

$$\begin{array}{l} 2x_2 + y_2 = x_2 \\ x_2 + 2y_2 = y_2 \end{array} \rightarrow \begin{array}{l} x_2 + y_2 = 0 \\ x_2 + y_2 = 0 \end{array} \quad \textit{Let} \quad \sqrt{x_2^2 + y_2^2 = 1} \quad \begin{cases} x_2 = -\dfrac{1}{\sqrt{2}} \\ y_2 = +\dfrac{1}{\sqrt{2}} \end{cases}$$

$$\mathbf{x} = \{x, y\}$$

$$\mathbf{x}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ +1 \end{bmatrix}$$



Eigenvectors

6.2. CLASSIFICATION OF NUMERICAL SOLUTION METHODS

- methods oriented on evaluation of all eigenvalues and eigenvectors (eg. Jacobi method)
- methods oriented on evaluation of a selected group of eigenvalues and eigenvectors
- methods oriented on evaluation of a single eigenvalue and eigenvector (usually extremal, eg. power method, reverse iteration method)

6.3. THEOREMS

Theorem 1 If **A** has distinct all eigenvalues, then there exists a complete set of linearly independent eigenvectors, unique up to a multiplicative constant.

Theorem 2 (Cayley-Hamilton theorem)
The matrix **A** satisfies its own characteristic equation, i.e. if $p(x)$ is a polynomial in x then

$$p(\mathbf{A}) = p(\lambda)$$

where λ is an eigenvalue of **A**.

Theorem 3 If $g(x)$ is a polynomial in x and λ is an eigenvalue of a matrix **A** then $g(\lambda)$ is an eigenvalue of the matrix $g(\mathbf{A})$.

Example

Let

$$\mathbf{C} = 3\mathbf{A}^2 - 2\mathbf{A} + 4\mathbf{I}$$

then

$$\lambda_C = 3\lambda_A^2 - 2\lambda_A + 4$$

Theorem 4 The eigenvalues (but not eigenvectors) are preserved under the similarity transformation.

Definition 1 The similarity transformation $\mathbf{R}^{-1}\mathbf{A}\mathbf{R}$ of the matrix **A,** where **R** is a non-singular matrix, does not change the eigenvalue λ.

Let $\mathbf{Ax} = \lambda\mathbf{x}$

$$\mathbf{x} = \mathbf{Ry} \;\rightarrow\; \mathbf{y} = \mathbf{R}^{-1}\mathbf{x}\,, \qquad \det \mathbf{R} \neq 0$$

$\mathbf{R}^{-1}\big|$ $\quad \mathbf{ARy} = \lambda\mathbf{Ry}$

$$\mathbf{R}^{-1}\mathbf{ARy} = \lambda\mathbf{y}$$

Thus eigenvalues for $\mathbf{A}$ and $\mathbf{R^{-1}AR}$ matrices are the same

Theorem 5 (Gerschgorin's theorem).

Let $\mathbf{A}$ be a given $n \times n$ matrix and let C_i, i = 1,2,...,n be the discs with centers a_{ii} and radii

$$R_i = \sum_{\substack{k=1 \\ k \neq i}}^{n} |a_{ik}|$$

Let D denote the union of the discs C_i. Then all the eigenvalues of $\mathbf{A}$ lie within D.

Conclusion:

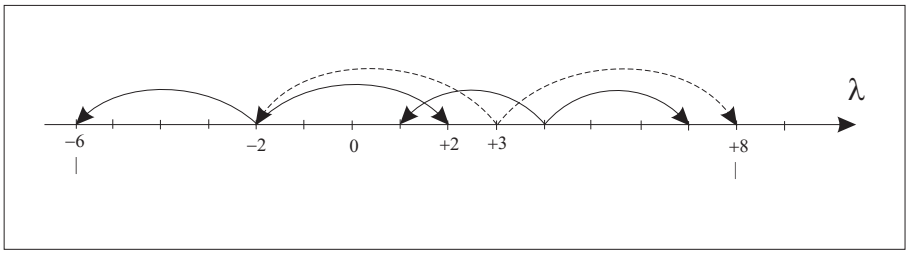
$$\lambda_{min} \geq \min_i \left(a_{ii} - R_i\right), \qquad \lambda_{max} \leq \max_i \left(a_{ii} + R_i\right)$$

Example

$$A = \begin{bmatrix} -2 & 1 & 3 \\ -1 & 4 & 2 \\ 3 & -2 & 3 \end{bmatrix}$$

$$a_{11} = -2 \qquad a_{22} = 4 \qquad a_{33} = 3$$

$$R_1 = 1 + 3 = 4$$
$$R_2 = |-1| + 2 = 3$$
$$R_3 = 3 + |-2| = 5$$



$$\lambda_{min} > \min \begin{Bmatrix} -2 - 4 \\ 4 - 3 \\ 3 - 5 \end{Bmatrix} = -6 \qquad \lambda_{max} < \max \begin{Bmatrix} -2 + 4 \\ 4 + 3 \\ 3 + 5 \end{Bmatrix} = 8$$

Remarks

- Theorem is useful for a rough evaluation of the eigenvalues spectrum
- Theorem holds also for complex matrices
- The quality of the Gerschgorin's evaluation depends on how much dominant are the diagonal terms of the matrix A considered. Evaluation is exact for diagonal matrices.

Theorem 6 (Sturm sequence property). The number of agreements in sign of successive numbers of the sequence $p_r(\hat{\lambda})$ for any given $\hat{\lambda}$, in a symmetric tridiagonal matrix $\mathbf{T}$, is equal to the number of eigenvalues of this matrix, which are strictly greater than $\hat{\lambda}$.

Here $p_r(\hat{\lambda})$ are the principal minors of the matrix $\mathbf{T}-\lambda\mathbf{I}$

$$\mathbf{T}=\begin{bmatrix} b_1 & c_1 & & & & & \\ c_1 & b_2 & c_2 & & & & \\ & c_2 & b_3 & c_3 & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & & & c_{n-1} & \\ & & & & c_{n-1} & b_n & \end{bmatrix}, \qquad \mathbf{T}-\lambda\mathbf{I}=\begin{bmatrix} b_1-\lambda & c_1 & & & & & \\ c_1 & b_2-\lambda & c_2 & & & & \\ & c_2 & b_3-\lambda & c_3 & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & & & c_{n-1} & \\ & & & & c_{n-1} & b_n-\lambda & \end{bmatrix}$$

and

$$\begin{aligned} &p_0(\lambda)=1 \\ &p_1(\lambda)=b_1-\lambda \\ &p_k(\lambda)=(b_k-\lambda)\,p_{k-1}(\lambda)-c_{k-1}^2\,p_{k-2}(\lambda), \qquad k=2,3,\cdots,n \end{aligned}$$

Remark:

If $p_j(\hat{\lambda})=0$ then we record the sign opposite to the sign $p_{j-1}(\hat{\lambda})$.

Example

Determine intervals containing at most one eigenvalue of the matrix

$$\mathbf{T}=\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

Solution

$$\mathbf{T}-\lambda\mathbf{I}=\begin{bmatrix} (2-\lambda) & -1 & & & \\ -1 & (2-\lambda) & -1 & & \\ & -1 & (2-\lambda) & -1 & \\ & & -1 & (2-\lambda) & -1 \\ & & & -1 & (2-\lambda) \end{bmatrix}$$

hence

$$p_0(\lambda)=1$$
$$p_1(\lambda)=2-\lambda$$
$$p_2(\lambda)=(2-\lambda)(2-\lambda)-(-1)^2*1=(2-\lambda)^2-1=\lambda^2-4\lambda+3$$
$$p_3(\lambda)=(2-\lambda)(\lambda^2-4\lambda+3)-(-1)^2(2-\lambda)=(2-\lambda)(\lambda^2-4\lambda+2)$$
$$p_4(\lambda)=(2-\lambda)(2-\lambda)(\lambda^2-4\lambda+3)-(-1)^2(\lambda^2-4\lambda+3)=(2-\lambda)^2(\lambda^2-4\lambda+2)-(\lambda^2-4\lambda+3)$$
$$p_5(\lambda)=(2-\lambda)\left[(2-\lambda)^2(\lambda^2-4\lambda+2)-(\lambda^2-4\lambda+3)\right]-(-1)^2(2-\lambda)(\lambda^2-4\lambda+2)=$$
$$=(2-\lambda)\left[(2-\lambda)^2(\lambda^2-4\lambda+2)-2(\lambda^2-4\lambda+2)-1\right]=(2-\lambda)\left\{\left[(2-\lambda)^2-2\right](\lambda^2-4\lambda+2)-1\right\}$$
$$\equiv \det(\mathbf{T}-\lambda\mathbf{I})$$

We select now values of $\hat{\lambda}$, and we record for each value of $\hat{\lambda}$ the sign of the polynomials $p_j(\hat{\lambda})$. For $p_j(\hat{\lambda})=0$ we record the sign opposite to the sign $p_{j-1}(\hat{\lambda})$

Table no. 1

k \ $\hat{\lambda}$	Sign of $p_k(\hat{\lambda})$								
	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
0	+	+	+	+	+	+	+	+	+
1	+	+	+	+	-	-	-	-	-
2	+	+	-	-	-	-	-	+	+
3	+	+	-	-	+	+	+	-	-
4	+	-	-	+	+	+	-	-	+
5	+	-	+	+	-	-	+	+	-
Number of eigenvalues $> \hat{\lambda}$	5	4	3	3	2	2	1	1	0

Examining the final row of this Table 1 we find a single eigenvalue in each of the intervals [0, 0.5], [0.5, 1.0], [1.5, 2.0], [2.5, 3.0], [3.5, 4.0]. Moreover a conclusion can be drawn from the first column of signs in Table 1 - the matrix **T** has 5 positive eigenvalues i.e. it is positive definite. The same conclusion may be drawn from the Gershgorin's Theorem:

$$\lambda_{\min} > 2-(|-1|+|-1|)=0 .$$

Theorem 7 Positive definite matrix has all eigenvalues positive. Symmetric positive definite matrix $(n\times n)$ has all *n* linearly independent eigenvectors.

Example

Positive definite matrix

$$\mathbf{A}=\begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix} \rightarrow (\lambda-1)^3=0 \rightarrow \lambda=1 \rightarrow \text{ one eigenvector } \{1\ \ 0\ \ 0\}$$

Symmetric positive definite matrix

$$\mathbf{B}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} \rightarrow (\lambda-1)^3=0 \rightarrow \lambda=1 \rightarrow \text{ three eigenvectors } \{1\ \ 0\ \ 0\},\ \{0\ \ 1\ \ 0\},\ \{0\ \ 0\ \ 1\}$$

Definition 2 Rayleigh Quotient.

$$\mathbf{Ax}=\lambda\mathbf{x} \rightarrow \Lambda=\frac{\mathbf{x}^t\mathbf{Ax}}{\mathbf{x}^t\mathbf{x}}$$

$$\lambda_{\min} \le \Lambda \le \lambda_{\max} \qquad \text{for arbitrary } \mathbf{x}\in\Omega$$

Remarks:

If **x** is eigenvector of a matrix **A**, the Rayleigh Quotient (RQ) Λ constitute the corresponding exact eigenvalue of this matrix.

However, for a vector **x,** being only an approximation of an eigenvector, the RQ presents an evaluation of the relevant fine eigenvalue.

Theorem 8 Orthogonal transformation preserve matrix symmetry as well as its eigenvalues

Given

$$\mathbf{Ax}=\lambda\mathbf{x},$$

is an orthogonal matrix.

Let

$$\mathbf{x}=\mathbf{Qy}, \quad \text{where } \mathbf{Q}$$

$$\mathbf{Q^TQ}=\mathbf{I} \rightarrow \mathbf{Q}^{-1}=\mathbf{Q^T}$$

Then

$$\mathbf{AQy}=\lambda\mathbf{Qy}$$

$$\mathbf{Q^TAQy}=\lambda\mathbf{y}$$

Remark:

The orthogonal transformation of the matrix is a particular case of the similarity transformation. Therefore, the statement of this theorem is obvious.

6.4. THE POWER METHOD

6.4.1. Concept of the method and its convergence

This is a method to find the *unique* dominant eigenvalue: $\max\left(\lambda_{max}, \left|-\lambda_{min}\right|\right)$

Let

$$\underset{n\times n}{\mathbf{A}}\ \underset{n\times 1}{\mathbf{u}_j} = \lambda_j \mathbf{u}_j \qquad \mathbf{u}_j - eigenvector \qquad j = 1,2,...,n$$
$$\lambda_j - eigenvalue$$

Let $|\lambda_1| > |\lambda_2| \geq |\lambda_3| ... \geq |\lambda_n|$

Let $\mathbf{x}_0 = \sum_{j=1}^{n} \alpha_j \mathbf{u}_j$

$\alpha_j, \mathbf{u}_j$ - are not known

Let $\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0$

$\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{A}\mathbf{x}_0 = \mathbf{A}^2\mathbf{x}_0$

.....

$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ k = 1, 2, ...

$$\mathbf{x}_{k+1} = \mathbf{A}^{k+1}\mathbf{x}_0 = \mathbf{A}^{k+1}\sum_{j=1}^{n} \alpha_j \mathbf{u}_j = \sum_{j=1}^{n} \alpha_j \mathbf{A}^{k+1}\mathbf{u}_j = \sum_{j=1}^{n} \alpha_j \lambda_j^{k+1}\mathbf{u}_j$$

$$\mathbf{x}_{k+1} = \lambda_1^{k+1}\left[\sum_{j=1}^{n} \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^{k+1} \mathbf{u}_j\right]$$

Notice:

Effect of multiplication the vector **x** by the matrix **A** is equivalent to multiplication by λ

Let $x_{S(k+1)}$ is the s-th vector component. Then

$$\frac{x_{s(k+1)}}{x_{s(k)}} = \frac{\lambda_1^{k+1}\left[\alpha_1 u_{s1} + \sum_{j=2}^{n} \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^{k+1} u_{sj}\right]}{\lambda_1^{k}\left[\alpha_1 u_{s1} + \sum_{j=2}^{n} \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^{k} u_{sj}\right]} = \lambda_1 + O\left(\frac{\lambda_2}{\lambda_1}\right)^k$$

Thus magnitude of $\frac{\lambda_2}{\lambda_1}$ decides about the convergence rate

Finally

$$\lim_{k\to\infty} \frac{x_{s(k+1)}}{x_{s(k)}} = \lambda_1$$

Remark:

If the dominant eigenvalue has multiplicity r, say we get

$$\mathbf{x}_{k+1} = \lambda_1^{k+1}\left[\sum_{j=1}^{r}\alpha_j\mathbf{u}_j + \sum_{j=r+1}^{n}\alpha_j\left(\frac{\lambda_j}{\lambda_1}\right)^{k+1}\mathbf{u}_j\right]$$

and

$$\frac{x_{s(k+1)}}{x_{s(k)}} = \lambda_1 + O\left(\frac{\lambda_{r+1}}{\lambda_1}\right)^k$$

The dominant eigenvalue λ_1 is found but $\mathbf{x}_{k-1}$ converges to a linear combination of the first r eigenvectors.

For real symmetric matrices the Rayleigh Quotient provides means of accelerating the convergence rate over the $\frac{x_{s(k+1)}}{x_{s(k)}}$ ratio.

6.4.2. Procedure using the Rayleigh quotient

POWER METHOD

GIVEN PROBLEM $\mathbf{Ax} = \lambda\mathbf{x}$

RAYLEIGH QUOTIENT $\Lambda = \dfrac{\mathbf{x}^T\mathbf{Ax}}{\mathbf{x}^T\mathbf{x}}$

0.	ASSUMPTION	$\mathbf{x}_0$
1.	NORMALIZATION	$\mathbf{v}_k = \dfrac{\mathbf{x}_k}{\left(\mathbf{x}_k^T\mathbf{x}_k\right)^{\frac{1}{2}}}$
2.	POWER STEP	$\mathbf{x}_{k+1} = \mathbf{Av}_k$
3.	RAYLEIGH QUOTIENT	$\Lambda_{k+1} = \dfrac{\mathbf{v}_k^T\mathbf{Av}_k}{\mathbf{v}_k^T\mathbf{v}_k} = \mathbf{v}_k^T\mathbf{Av}_k = \mathbf{v}_k^T\mathbf{x}_{k+1}$
4.	ERROR ESTIMATION	$\varepsilon_{k+1}^{(\Lambda)} = \left\lvert\dfrac{\Lambda_{k+1}-\Lambda_k}{\Lambda_{k+1}}\right\rvert, \quad \varepsilon_{k+1}^{(v)} = \lvert\mathbf{v}_{k+1}-\mathbf{v}_k\rvert$
5.	BRAKE OFF TEST	$\varepsilon_{k+1}^{(\Lambda)} \overset{?}{<} B_\Lambda$, $\varepsilon_{k+1}^{(v)} \overset{?}{<} B_v$ if No – go to 1, if Yes – go to 6.
6.	FINAL RESULTS	$\lambda_{\max} \approx \Lambda_{k+1}$, $\mathbf{x}_{\max} \approx \mathbf{x}_{k+1}$

Example

$$\mathbf{A}=\begin{bmatrix} 4 & 1/2 & 0 \\ 1/2 & 4 & 1/2 \\ 0 & 1/2 & 4 \end{bmatrix} \quad \rightarrow \quad \det(\mathbf{A}-\lambda\mathbf{I})=\begin{bmatrix} 4-\lambda & 1/2 & 0 \\ 1/2 & 4-\lambda & 1/2 \\ 0 & 1/2 & 4-\lambda \end{bmatrix}=0 \quad \rightarrow$$

$$\rightarrow \quad \lambda_1=4+\frac{1}{\sqrt{2}}, \qquad \lambda_2=4, \qquad \lambda_3=4-\frac{1}{\sqrt{2}} \quad \text{exact eigenvalues}$$

$$\lambda_1=4.7071067,\ \lambda_2=4,\ \lambda_3=3.2928933$$

$$\mathrm{v}_1=\left\{\frac{1}{2}\ \ \frac{1}{\sqrt{2}}\ \ \frac{1}{2}\right\},\ \mathrm{v}_2=\left\{\frac{1}{\sqrt{2}}\ \ 0\ \ -\frac{1}{\sqrt{2}}\right\},\ \mathrm{v}_3=\left\{-\frac{1}{2}\ \ \frac{1}{\sqrt{2}}\ \ -\frac{1}{2}\right\},$$

EIGENSPECTRUM EVALUATION BY THE GERSCHGORIN'S THEOREM

$$\lambda_{\min}>4-\frac{1}{2}-\frac{1}{2}=3 \qquad\qquad \lambda_{\max}<4+\frac{1}{2}+\frac{1}{2}=5$$

Matrix is positive definite and symmetric $\rightarrow$ 3 different eigenvalues.

POWER METHOD SOLUTION PROCESS

Assume

$$\mathbf{x}_0=\{1\ \ 1\ \ 1\}$$

hence

$$\mathbf{v}_0=\frac{\mathbf{x}_0}{\left(\mathbf{x}_0^{\mathrm{T}}\mathbf{x}_0\right)^{\frac{1}{2}}}=\frac{\mathbf{x}_0}{(1\ \ 1\ \ 1)^{\frac{1}{2}}}=\left\{\frac{1}{\sqrt{3}}\ \ \frac{1}{\sqrt{3}}\ \ \frac{1}{\sqrt{3}}\right\}$$

$$\mathbf{x}_1=\mathbf{A}\mathbf{v}_0=\begin{bmatrix} 4 & 1/2 & 0 \\ 1/2 & 4 & 1/2 \\ 0 & 1/2 & 4 \end{bmatrix}\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}=1/\sqrt{3}\begin{bmatrix} 9/2 \\ 5 \\ 9/2 \end{bmatrix}$$

$$\Lambda_1=\mathbf{v}_0^t\mathbf{x}_1=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}1/\sqrt{3}\begin{bmatrix} 9/2 \\ 5 \\ 9/2 \end{bmatrix}=\frac{14}{3}=4.666667$$

$$\mathbf{v}_1=\frac{1}{\sqrt{3}}\left\{\frac{9}{2}\ \ 5\ \ \frac{9}{2}\right\}\frac{1}{\frac{1}{\sqrt{3}}\left[\left(\frac{9}{2}\right)^2+5^2+\left(\frac{9}{2}\right)^2\right]^{\frac{1}{2}}}=\frac{6}{13\sqrt{3}}\left\{\frac{9}{2}\ \ 5\ \ \frac{9}{2}\right\}=\begin{bmatrix} 0.556022 \\ 0.617802 \\ 0.556022 \end{bmatrix}$$

$$\mathbf{x}_2 = \mathbf{A}\mathbf{v}_1 = \begin{bmatrix} 4 & 1/2 & 0 \\ 1/2 & 4 & 1/2 \\ 0 & 1/2 & 4 \end{bmatrix} \begin{bmatrix} 0.556022 \\ 0.617802 \\ 0.556022 \end{bmatrix} = \begin{bmatrix} 2.532988 \\ 3.027230 \\ 2.532988 \end{bmatrix}$$

$$\Lambda_2 = \begin{bmatrix} 0.556022, 0.617802, 0.556022 \end{bmatrix} \begin{bmatrix} 2.532988 \\ 3.027230 \\ 2.532988 \end{bmatrix} = 4.687023$$

$$\mathbf{v}_2 = \frac{\mathbf{x}_2}{\left(\mathbf{x}_2^T \mathbf{x}_2\right)^{1/2}} = \begin{bmatrix} 0.540082 \\ 0.645464 \\ 0.540082 \end{bmatrix}$$

Error estimation

$$\varepsilon_2^{(\Lambda)} = \left| \frac{4.687023 - 4.666667}{4.687023} \right| = \underline{0.004342}$$

$$\mathbf{v}_2 - \mathbf{v}_1 = \{0.5400818 - 0.5560218, 0.6454636 - 0.6178020, 0.5400818 - 0.5560218\} =$$

$$= \{-0.015940, 0.027662, -0.015940\}$$

$$\varepsilon^{(v)} = \left| \mathbf{v}_2 - \mathbf{v}_1 \right| = \underline{0.035684}$$

Notice greater accuracy of Λ_2 than $\mathbf{v}_2$ i.e. $\varepsilon_2^{(\lambda)} < \varepsilon_2^{(v)}$

$$\mathbf{v}_3 = \{0.528458, 0.664428, 0.528458\}$$

$$\Lambda_3 = 4.697206$$

Error estimation

$$\varepsilon_3^{(\Lambda)} = \left| \frac{4.69721 - 4.68702}{4.69721} \right| = 0.002169$$

$$\varepsilon_3^{(v)} = 0.016438$$

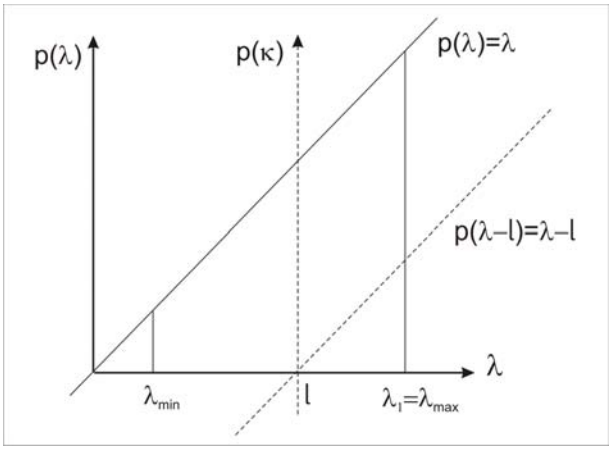
.....

$$\begin{cases} \mathbf{v}_{11} = \{0.501681, 0.704721, 0.501681\} \\ \Lambda_{11} = 4.707074 \end{cases}$$

$$\begin{cases} \mathbf{v}_{25} = \{0.500056, 0.707028, 0.500056\} & \text{- result exact within } 3 \div 4 \text{ digits} \\ \Lambda_{25} = 4.707107 & \text{- result exact within 7 digits} \end{cases}$$

$$\begin{cases} \mathbf{v}_{\infty} = \{0.500000, 0.707107, 0.500000\} & \text{- result exact} \\ \Lambda = 4.707107 \end{cases}$$

6.4.3. Shift of the eigenspectrum



Given eigenvalue problem

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

Let

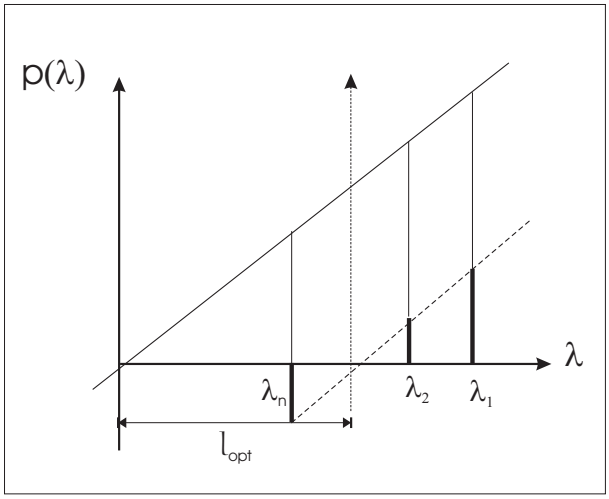
$$\lambda = \kappa + l \quad \rightarrow \quad \mathbf{A}\mathbf{x} = \kappa\mathbf{x} + l\mathbf{x}$$

$$\left(\mathbf{A} - l\mathbf{I}\right)\mathbf{x} = \kappa\mathbf{x} \quad \rightarrow \quad \mathbf{x}, \kappa \rightarrow \lambda = \kappa + l$$

$$\mathbf{A} \rightarrow p(\lambda) = \lambda$$

$$\mathbf{A} - l\mathbf{I} \rightarrow p(\kappa) = p(\lambda - l) = \lambda - l$$

6.4.4. Application of shift to acceleration of convergence to $\lambda_{\max} = \lambda_1$



The optimal shift

$$l_{opt} = \frac{\lambda_2 + \lambda_n}{2}$$

in order to speed-up the convergence rate while evaluating λ_1

Example

$$l_{opt} = \frac{\lambda_2 + \lambda_3}{2} = \frac{4 + 4 - \frac{1}{\sqrt{2}}}{2} = 4 - \frac{1}{2\sqrt{2}} = 3.646447$$ - optimal shift for λ_1 evaluation

$$\mathbf{A}-l_{opt}\mathbf{I}=\begin{bmatrix}\left(4-4+\frac{1}{2\sqrt{2}}\right) & \frac{1}{2} & 0\\ \frac{1}{2} & \left(4-4+\frac{1}{2\sqrt{2}}\right) & \frac{1}{2}\\ 0 & \frac{1}{2} & \left(4-4+\frac{1}{2\sqrt{2}}\right)\end{bmatrix}=\frac{1}{2}\begin{bmatrix}\frac{1}{\sqrt{2}} & 1 & 0\\ 1 & \frac{1}{\sqrt{2}} & 1\\ 0 & 1 & \frac{1}{\sqrt{2}}\end{bmatrix}$$

Let

$$\mathbf{x_0}=\{1,\quad 1,\quad 1\}$$

$$\mathbf{v}_0=\left\{\frac{1}{\sqrt{3}},\quad \frac{1}{\sqrt{3}},\quad \frac{1}{\sqrt{3}}\right\}$$

$$\mathbf{x}_1=\frac{1}{2}\begin{bmatrix}\frac{1}{\sqrt{2}} & 1 & 0\\ 1 & \frac{1}{\sqrt{2}} & 1\\ 0 & 1 & \frac{1}{\sqrt{2}}\end{bmatrix}\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}=\frac{1}{2\sqrt{3}}\begin{bmatrix}1+\frac{1}{\sqrt{2}}\\ 2+\frac{1}{\sqrt{2}}\\ 1+\frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}0.492799\\0.781474\\0.492799\end{bmatrix}$$

$$\kappa_1=\frac{1}{3}[1,\quad 1,\quad 1]\frac{1}{2\sqrt{3}}\begin{bmatrix}1+\frac{1}{\sqrt{2}}\\ 2+\frac{1}{\sqrt{2}}\\ 1+\frac{1}{\sqrt{2}}\end{bmatrix}=\frac{1}{6}\left(4+\frac{3}{\sqrt{2}}\right)=1.020220\ \rightarrow$$

$$\rightarrow\quad \lambda_{(1)}=1.020220+3.646447=4.666667$$

$$\mathbf{v_1}=\{0.470635,\ 0.746326,\ 0.470635\}$$

$$\mathbf{x_2}=\{0.539558,\ 0.734501,\ 0.539558\}$$

$$\kappa_2=1.056047\rightarrow\ \lambda_{(2)}=4.702493$$

$$\mathbf{v_2}=\{0.509439,\ 0.693501,\ 0.509439\}$$

Error estimation

$$\varepsilon_2^{(\lambda)}=\left|\frac{4.702493-4.666667}{4.702493}\right|=\underline{0.007619}$$

$$\varepsilon_2^{(v)}=\left|\mathbf{v_2}-\mathbf{v_1}\right|=\underline{0.076171}$$

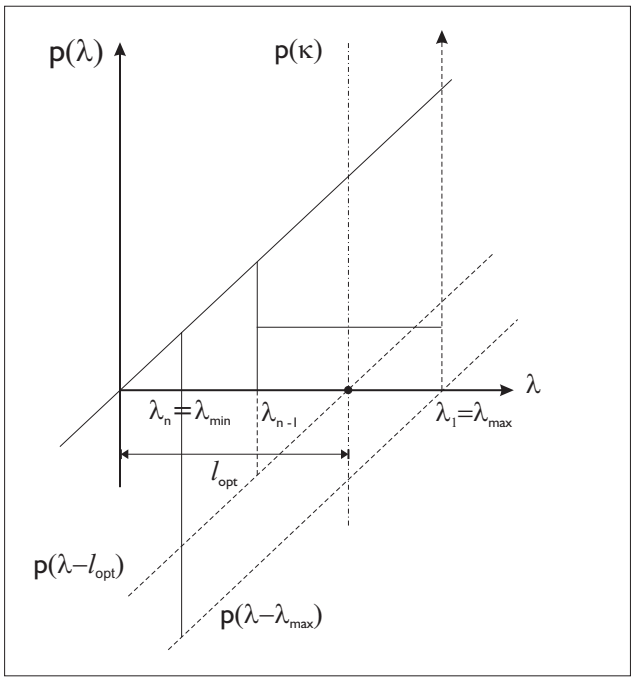
$\mathbf{v}_8=\{0.500073,\ 0.707173,\ 0.500073\}$ - result exact within 3-4 digits

$\lambda_{(8)}=4.707107$ - result exact within 7 digits

Remark:

Due to shift number of iterations was reduced from 25 to 8.

6.4.5. Application of a shift to acceleration of convergence to λ_{min}



When $l = \lambda_{\max}$ the solution procedure is convergent to $\lambda_{\min}$

The optimal convergence to $\lambda_{\min}$ is obtained when:

$$l_{opt} = \frac{\lambda_{n-1} + \lambda_{\max}}{2}$$

Examples

(i) Let

$$l = \lambda_3 = 4 + \frac{1}{\sqrt{2}} = 4.707107 \rightarrow \lambda = \kappa + 4.707107$$

$$\mathbf{A} - \lambda_3 \mathbf{I} = \begin{bmatrix} \left(4 - 4 - \frac{1}{\sqrt{2}}\right) & \frac{1}{2} & 0 \\ \frac{1}{2} & \left(4 - 4 - \frac{1}{\sqrt{2}}\right) & \frac{1}{2} \\ 0 & \frac{1}{2} & \left(4 - 4 - \frac{1}{\sqrt{2}}\right) \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Let

$$\mathbf{x_0} = \{1,\ 1,\ 1\}$$

$$\mathbf{v_0} = \frac{\mathbf{x_0}}{\left(\mathbf{x_0^t} \cdot \mathbf{x_0}\right)^{\frac{1}{2}}} = \left\{\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}\right\}$$

$$\mathbf{x_1} = \left(\mathbf{A} - \lambda_3 \mathbf{I}\right)\mathbf{v_0} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{\sqrt{2}} \end{bmatrix} \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{2\sqrt{3}} \begin{bmatrix} 1 - \sqrt{2} \\ 2 - \sqrt{2} \\ 1 - \sqrt{2} \end{bmatrix} = \begin{bmatrix} -0.119573 \\ 0.169102 \\ -0.119573 \end{bmatrix}$$

$$\kappa_1 = \mathbf{v_0^t} \cdot \mathbf{x_1} = \frac{1}{\sqrt{3}}[1, 1, 1] \cdot \frac{1}{2\sqrt{3}} \left\{1-\sqrt{2},\quad 2-\sqrt{2},\quad 1-\sqrt{2}\right\} = \frac{2}{3} - \frac{\sqrt{2}}{2} = -0.040440$$

$$\lambda_{(1)} = \kappa_1 + l = -0.040440 + 4.707107 = 4.666667$$

$$\mathbf{v_1} = \frac{\mathbf{x_1}}{\left(\mathbf{x_1^t} \cdot \mathbf{x_1}\right)^{\frac{1}{2}}} = \{-0.500000,\ 0.707107,\ -0.500000\} \approx \left\{-\frac{1}{2},\ \frac{1}{\sqrt{2}},\ -\frac{1}{2}\right\}$$

$$\mathbf{x_2} = \left(\mathbf{A} - \lambda_3 \mathbf{I}\right)\mathbf{v_1} = \left\{\frac{1}{\sqrt{2}},\ -1,\ \frac{1}{\sqrt{2}}\right\}$$

$$\kappa_2 = \mathbf{v}_1^{\mathrm{T}} \mathbf{x_2} = -\sqrt{2}$$

$$\lambda_{(2)} = \kappa_2 + l = -\sqrt{2} + 4 + \frac{1}{\sqrt{2}} = 4 - \frac{1}{\sqrt{2}} \equiv \lambda_3$$

here $\mathbf{x_2}$, κ_2 and $\lambda_{(2)}$ are the exact results.

(ii) Let

$$l_{opt} = \frac{\lambda_1 + \lambda_2}{2} = \left(4 + \frac{1}{\sqrt{2}} + 4\right)\frac{1}{2} = 4 + \frac{1}{2\sqrt{2}} = 4.353553$$

$$\mathbf{A} - l_{opt}\mathbf{I} = \begin{bmatrix} -\frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -\frac{1}{\sqrt{2}} & 1 & 0 \\ 1 & -\frac{1}{\sqrt{2}} & 1 \\ 0 & 1 & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Let

$$\mathbf{x_0} = \{1, 1, 1\}$$

$$\mathbf{v_0} = \left\{\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right\}$$

$$\mathbf{x_1} = \left(\mathbf{A} - l_{opt}\mathbf{I}\right)\mathbf{v_0} = \frac{1}{2\sqrt{3}}\left\{1-\frac{1}{\sqrt{2}},\ 2-\frac{1}{\sqrt{2}},\ 1-\frac{1}{\sqrt{2}}\right\} = \{0.084551,\ 0.373226,\ 0.084551\}$$

$$\kappa_1 = \mathbf{v_0^t} \cdot \mathbf{x_1} = 0.313113$$

$$\lambda_{(1)} = \kappa_1 + l_{opt} = 0.313113 + 4.353554 = 4.666667$$

$$\mathbf{v_1} = \frac{\mathbf{x_1}}{\left(\mathbf{x_1^t} \cdot \mathbf{x_1}\right)^{\frac{1}{2}}} = \{0.215739,\ 0.952320,\ 0.215739\}$$

……

6.5. INVERSE ITERATION METHOD

6.5.1. The basic algorithm

$$\mathbf{Ax} = \lambda\mathbf{x} \rightarrow \mathbf{A}^{-1}\mathbf{Ax} = \lambda\mathbf{A}^{-1}\mathbf{x}\ , \qquad \text{where } \mathbf{A} \text{ is a non-singular matrix}$$

$$\mathbf{A}^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

Let

$$\mu = \frac{1}{\lambda} \qquad \rightarrow \qquad \lambda_c = \frac{1}{\mu_{\max}}, \qquad \lambda_{\max} = \frac{1}{\mu_c}$$

hence

$$\mathbf{A}^{-1}\mathbf{x} = \mu\mathbf{x}$$

Notice:

Here λ_c and μ_c mean eigenvalues closest to zero

INVERSE METHOD

0. ASSUMPTION $\mathbf{x}_0$

1. NORMALIZATION $\mathbf{v}_K = \dfrac{\mathbf{x}_k}{\left(\mathbf{x}_k^T \cdot \mathbf{x}_k\right)^{\frac{1}{2}}}$

2. POWER STEP

$$\mathbf{x}_{k+1} = \mathbf{A}^{-1}\mathbf{v}_k \rightarrow \overbrace{\mathbf{A}\mathbf{x}_{k+1} = \mathbf{v}_k}^{\substack{\text{solution of linear algebraic simultaneous equations}}} \rightarrow \mathbf{x}_{k+1} \ = \ \begin{cases} \mathbf{A} = \mathbf{LU} & \textit{LU decomposition} \\ \mathbf{L}\mathbf{y}_k = \mathbf{v}_k \rightarrow \mathbf{y}_k & \textit{step foreward} \\ \mathbf{U}_\mathbf{k}\mathbf{x}_{k+1} = \mathbf{y}_k \rightarrow \mathbf{x}_{k+1} & \textit{step back} \end{cases}$$

No

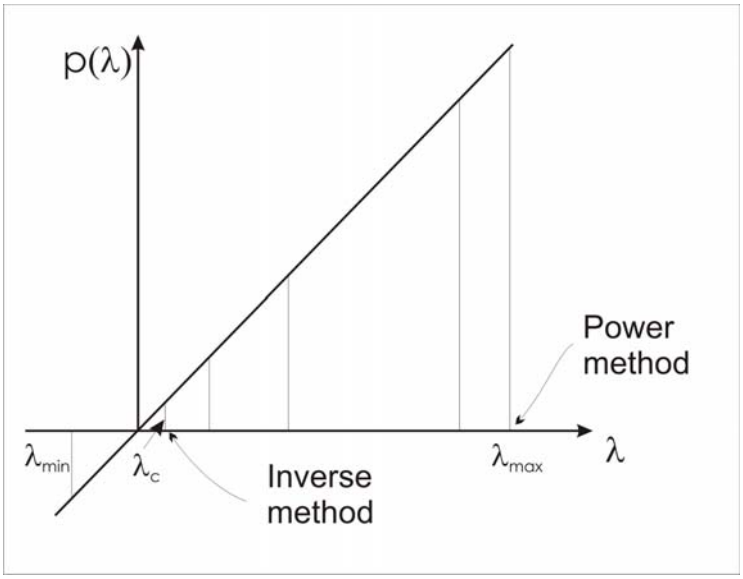
3. RAYLEIGH QUOTIENT $\Lambda_{k+1} = \dfrac{\mathbf{v}_k^T\mathbf{A}^{-1}\mathbf{v}_k}{\mathbf{v}_k^T\mathbf{v}_k} = \mathbf{v}_k^T\mathbf{x}_{k+1}$, $\lambda_{(k+1)} = \Lambda_{k+1}^{-1}$

4. ERROR ESTIMATION $\varepsilon_k^{(\lambda)} = \left|\dfrac{\Lambda_{k+1} - \Lambda_k}{\Lambda_{k+1}}\right|$, $\varepsilon_{k+1}^{(v)} = \left|\mathbf{v}_{k+1} - \mathbf{v}_k\right|$

5. BRAKE OFF TEST $\varepsilon^{(\lambda)}{}_{k+1} \overset{?}{<} \mathrm{B}_\lambda$, $\varepsilon^{(v)}{}_{k+1} < \mathrm{B}_v$

Yes

6. FINAL RESULTS $\lambda_c \approx \Lambda_{k+1}^{-1}$ $\mathbf{x}_c \approx \mathbf{x}_{k+1}$



Example

Let

$$\mathbf{A} = \begin{bmatrix} 4 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & \frac{1}{2} \\ 0 & \frac{1}{2} & 4 \end{bmatrix}$$

Matrix decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{bmatrix} 2 & 0 & 0 \\ \frac{1}{4} & \frac{\sqrt{63}}{4} & 0 \\ 0 & \frac{2}{\sqrt{63}} & 2\sqrt{\frac{62}{63}} \end{bmatrix} \begin{bmatrix} 2 & \frac{1}{4} & 0 \\ 0 & \frac{\sqrt{63}}{4} & \frac{2}{\sqrt{63}} \\ 0 & 0 & 2\sqrt{\frac{62}{63}} \end{bmatrix}$$

Let

$$\mathbf{x}_0 = \{1,\ 1,\ 1\}$$

$$\mathbf{v}_0 = \frac{\mathbf{x_0}}{\left(\mathbf{x_0^t} \cdot \mathbf{x_0}\right)^{\frac{1}{2}}} = \left\{\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}\right\}$$

$$\mathbf{A}\mathbf{x}_1 = \begin{bmatrix} 4 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & \frac{1}{2} \\ 0 & \frac{1}{2} & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \rightarrow \mathbf{x}_1 = \left\{ \frac{7}{3\sqrt{3}}, \quad \frac{6}{3\sqrt{3}}, \quad \frac{7}{3\sqrt{3}} \right\} = \begin{bmatrix} 0.130369 \\ 0.111745 \\ 0.130369 \end{bmatrix}$$

$$\Lambda_1 = \mathbf{v}_0^T \mathbf{x}_1 = 0.215054 \rightarrow \lambda_{(1)} = \Lambda_1^{-1} = 4.649995$$

$$\mathbf{v}_1 = \frac{\mathbf{x}_1}{\left(\mathbf{x}_1^{\mathbf{t}} \cdot \mathbf{x}_1\right)^{\frac{1}{2}}} = \left\{ \frac{7}{\sqrt{134}}, \ \frac{6}{\sqrt{134}}, \ \frac{7}{\sqrt{134}} \right\} = \{0.604708, \ 0.518321, \ 0.604708\}$$

$$\mathbf{A}\mathbf{x}_2 = \begin{bmatrix} 4 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & \frac{1}{2} \\ 0 & \frac{1}{2} & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{134}} \begin{bmatrix} 7 \\ 6 \\ 7 \end{bmatrix} \rightarrow \mathbf{x}_2 = \frac{1}{31\sqrt{134}} \{50, \quad 34, \quad 50\} = \begin{bmatrix} 0.139334 \\ 0.094747 \\ 0.139334 \end{bmatrix}$$

$$\Lambda_2 = \mathbf{v}_1^T \mathbf{x}_2 = 0.217622 \rightarrow \lambda_{(2)} = \Lambda_2^{-1} = 4.591342$$

$$\mathbf{v}_2 = \frac{\mathbf{x}_2}{\left(\mathbf{x}_2^{\mathbf{t}} \cdot \mathbf{x}_2\right)^{\frac{1}{2}}} = \frac{1}{9\sqrt{19}} \{25, \ 17, \ 25\} = \{0.637266, \ 0.433341, \ 0.6372659\}$$

$$\mathbf{A}\mathbf{x}_3 = \begin{bmatrix} 4 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & \frac{1}{2} \\ 0 & \frac{1}{2} & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{9\sqrt{19}} \begin{bmatrix} 25 \\ 17 \\ 25 \end{bmatrix} \rightarrow \mathbf{x}_3 = \begin{bmatrix} 0.150477 \\ 0.070716 \\ 0.150477 \end{bmatrix}$$

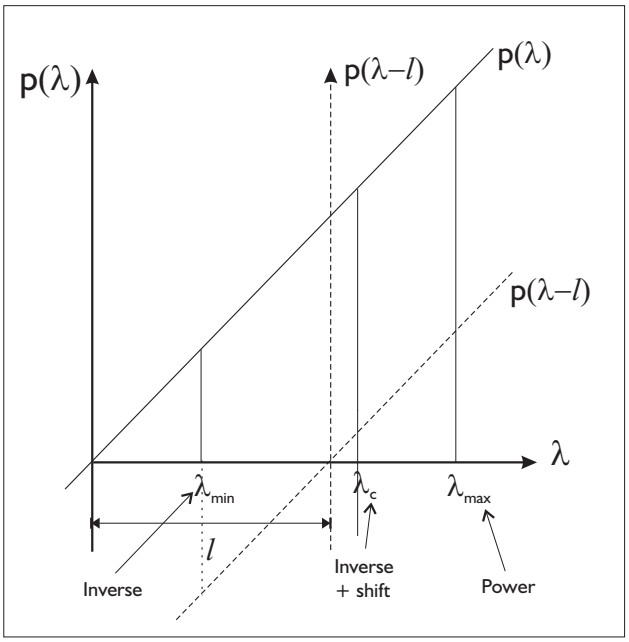
.....

$$\Lambda_{final} = 0.303684 \ \rightarrow \ \lambda_{final} = \Lambda_{final}^{-1} = 3.292893 = \lambda_3$$
$$\mathbf{v}_{final} = \{-0.50000, \ 0.707107, \ -0.50000\}$$

Remark:

Convergence is initially slow because of unhappy choice of $\mathbf{x}_0$.

6.5.2. Use of inverse and shift In order to find the eigenvalue λ_c closest to a given one



Let

$$\lambda = \kappa + l$$

CONCEPT

The same like in the case of inverse method but:

- $\mathbf{A}$ is replaced now by $\mathbf{A} - l\mathbf{I}$
- $\lambda = \Lambda^{-1} + l$

Example

$$\mathbf{A} = \begin{bmatrix} 4 & \frac{1}{2} & 0 \\ \frac{1}{2} & 4 & \frac{1}{2} \\ 0 & \frac{1}{2} & 4 \end{bmatrix}$$

Let

$$l = 3.75$$

Thus

$$\tilde{\mathbf{A}} = \mathbf{A} - 3.75\mathbf{I} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

Error estimation

$$\varepsilon_k^{\lambda} = \left| \frac{\Lambda_{k+1} - \Lambda_k}{\Lambda_{k+1}} \right|, \qquad \varepsilon_k^{(\mathrm{v})} = \left| \mathbf{v}_{k+1} - \mathbf{v}_k \right|$$

Let us assume first a symmetric starting vector

$$\mathbf{x}_0 = \{1,\ 1,\ 1\}$$

then

$$\mathbf{v}_0 = \frac{\mathbf{x}_0}{\left(\mathbf{x}_0^{\mathbf{t}} \cdot \mathbf{x}_0\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{3}} \{1,\ 1,\ 1\}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \rightarrow \mathbf{x}_1 = \frac{4}{7\sqrt{3}} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.329914 \\ 0.989743 \\ 0.329914 \end{bmatrix}$$

$$\mathbf{v}_1 = \frac{\mathbf{x}_1}{\left(\mathbf{x}_1^{\mathbf{t}} \cdot \mathbf{x}_1\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{11}} \{1,\ 3,\ 1\} = \begin{bmatrix} 0.301511 \\ 0.904534 \\ 0.301511 \end{bmatrix}$$

$$\Lambda_1 = \mathbf{v}_0^T \mathbf{x}_1 = \frac{20}{21} = 0.952381 \ \rightarrow\ \lambda_{(1)} = \Lambda_1^{-1} + l = \frac{21}{20} + 3.75 = 4.800000$$

error estimation

$$\varepsilon_1^{(\lambda)} = 1.00\,, \qquad \varepsilon_1^{(\mathrm{v})} = 2.54 \times 10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_2 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{11}} \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \rightarrow\ \mathbf{x}_2 = \frac{4}{7\sqrt{11}} \begin{bmatrix} 5 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.861461 \\ 0.172292 \\ 0.861461 \end{bmatrix}$$

$$\mathbf{v}_2 = \frac{\mathbf{x}_2}{\left(\mathbf{x}_2^{\mathbf{t}} \cdot \mathbf{x}_2\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{51}} \{5,\ 1,\ 5\} = \begin{bmatrix} 0.700140 \\ 0.140028 \\ 0.700140 \end{bmatrix}$$

$$\Lambda_2 = \mathbf{v}_1^t \mathbf{x}_2 = \frac{52}{77} = 0.675325, \quad \lambda_{(2)} = \Lambda_2^{-1} + l = \frac{77}{52} + 3.75 = 5.230769$$

error estimation

$$\varepsilon_2^{(\lambda)} = 8.24 \times 10^{-2}\,, \qquad \varepsilon_2^{(\mathrm{v})} = 5.48 \times 10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_3 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{51}} \begin{bmatrix} 5 \\ 1 \\ 5 \end{bmatrix} \rightarrow \mathbf{x}_3 = \frac{4}{7\sqrt{51}} \begin{bmatrix} -3 \\ 19 \\ -3 \end{bmatrix} = \begin{bmatrix} -0.240048 \\ 1.520304 \\ -0.240048 \end{bmatrix}$$

$$\mathbf{v}_3 = \frac{\mathbf{x}_3}{\left(\mathbf{x}_3^{\mathbf{t}} \cdot \mathbf{x}_3\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{379}} \{-3,\ 19,\ -3\} = \begin{bmatrix} -0.154100 \\ 0.975964 \\ -0.154100 \end{bmatrix}$$

$$\Lambda_3 = \mathbf{v}_2^t \mathbf{x}_3 = -\frac{44}{7 \cdot 51} = -0.123249, \quad \lambda_{(3)} = \Lambda_3^{-1} + l = -\frac{357}{44} + 3.75 = -4.363636$$

error estimation

$$\varepsilon_3^{(\lambda)} = 2.20\,, \qquad \varepsilon_3^{(\mathrm{v})} = 6.57 \times 10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_4 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{379}} \begin{bmatrix} -3 \\ 19 \\ -3 \end{bmatrix} \rightarrow \mathbf{x}_4 = \frac{4}{7\sqrt{379}} \begin{bmatrix} 41 \\ -31 \\ 41 \end{bmatrix} = \begin{bmatrix} 1.203444 \\ -0.909922 \\ 1.203444 \end{bmatrix}$$

$$\mathbf{v}_4 = \frac{\mathbf{x}_4}{\left(\mathbf{x}_4^{\mathbf{t}} \cdot \mathbf{x}_4\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{4323}} \{41,\ -31,\ 41\} = \begin{bmatrix} 0.623579 \\ -0.471486 \\ 0.623579 \end{bmatrix}$$

$$\Lambda_4 = \mathbf{v}_3^t \mathbf{x}_4 = -1.258952, \quad \lambda_4 = \Lambda_4^{-1} + l = -\frac{1}{1.258952} + 3.75 = 2.955689$$

error estimation

$$\varepsilon_4^{(\lambda)} = 2.48\,, \qquad \varepsilon_4^{(\mathrm{v})} = 4.81 \times 10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_5 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{4323}} \begin{bmatrix} 41 \\ -31 \\ 41 \end{bmatrix} \rightarrow \mathbf{x}_5 = \frac{4}{7\sqrt{4323}} \begin{bmatrix} -103 \\ 195 \\ 103 \end{bmatrix} = \begin{bmatrix} -0.895172 \\ 1.694743 \\ -0.895172 \end{bmatrix}$$

$$\mathbf{v}_5 = \frac{\mathbf{x}_5}{\left(\mathbf{x}_5^{\mathbf{t}} \cdot \mathbf{x}_5\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{59243}} \{-103,\ 195,\ -103\} = \begin{bmatrix} -0.423174 \\ 0.801154 \\ -0.423174 \end{bmatrix}$$

$$\Lambda_5 = \mathbf{v}_4^t \mathbf{x}_5 = -1.915469, \quad \lambda_5 = \Lambda_5^{-1} + l = -\frac{1}{1.915469} + 3.75 = 3.227935$$

error estimation

$$\varepsilon_5^{(\lambda)} = 8.43 \times 10^{-2}, \qquad \varepsilon_5^{(\mathrm{v})} = 2.51 \times 10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_6 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{59243}} \begin{bmatrix} -103 \\ 195 \\ -103 \end{bmatrix} \rightarrow \mathbf{x}_6 = \frac{4}{7\sqrt{59243}} \begin{bmatrix} 493 \\ -607 \\ 493 \end{bmatrix} = \begin{bmatrix} 1.157418 \\ -1.425057 \\ 1.157418 \end{bmatrix}$$

$$\mathbf{v}_6 = \frac{\mathbf{x}_6}{\left(\mathbf{x}_6^{\mathbf{t}} \cdot \mathbf{x}_6\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{854547}} \{493, -607,\ 493\} = \begin{bmatrix} 0.533309 \\ -0.656630 \\ 0.533309 \end{bmatrix}$$

$$\Lambda_6 = \mathbf{v}_5^t \mathbf{x}_6 = -2.121268, \quad \lambda_6 = \Lambda_6^{-1} + l = -\frac{1}{2.121268} + 3.75 = 3.278584$$

error estimation

$$\varepsilon_6^{(\lambda)} = 1.54 \times 10^{-2}, \qquad \varepsilon_6^{(\mathrm{v})} = 1.23 \times 10^{-1}$$

.....

$$\tilde{\mathbf{A}}\,\mathbf{x}_{10} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.496219 \\ 0.712414 \\ -0.496219 \end{bmatrix} \rightarrow \mathbf{x}_{10} = \begin{bmatrix} 1.097741 \\ -1.541308 \\ 1.097741 \end{bmatrix}$$

$$\mathbf{v}_{10} = \frac{\mathbf{x}_{10}}{\left(\mathbf{x}_{10}^{\mathbf{t}} \cdot \mathbf{x}_{10}\right)^{\frac{1}{2}}} = \begin{bmatrix} 0.501796 \\ -0.704558 \\ 0.501796 \end{bmatrix}$$

$$\Lambda_{10} = \mathbf{v}_9^t \mathbf{x}_{10} = -2.187631, \quad \lambda_{10} = \Lambda_{10}^{-1} + l = -\frac{1}{2.187631} + 3.75 = 3.292884$$

error estimation

$$\varepsilon_{10}^{(\lambda)} = 3.94 \times 10^{-5}, \quad \varepsilon_{10}^{(\mathrm{v})} = 6.43 \times 10^{-3}$$

.....

$$\tilde{\mathbf{A}}\,\mathbf{x}_{28} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.500000 \\ 0.707107 \\ -0.500000 \end{bmatrix} \rightarrow \mathbf{x}_{28} \rightarrow \mathbf{v}_{28} = \begin{bmatrix} 0.500000 \\ -0.707107 \\ 0.500000 \end{bmatrix}$$

$$\Lambda_{28} = \mathbf{v}_{27}^t \mathbf{x}_{28} = -2.187673, \quad \lambda_{28} = \Lambda_{28}^{-1} + l = -\frac{1}{2.187673} + 3.75 = 3.292893$$

error estimation

$$\varepsilon_{28}^{(\lambda)} = 1.35 \times 10^{-16}, \quad \varepsilon_{28}^{(\mathrm{v})} = 1.35 \times 10^{-8}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_{29} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.500000 \\ -0.707107 \\ 0.500000 \end{bmatrix} \rightarrow \mathbf{x}_{29} \rightarrow \mathbf{v}_{29} = \begin{bmatrix} -0.500000 \\ 0.707107 \\ -0.500000 \end{bmatrix}$$

$$\Lambda_{29} = \mathbf{v}_{28}^t \mathbf{x}_{29} = -2.187673, \quad \lambda_{29} = \Lambda_{29}^{-1} + l = -\frac{1}{2.187673} + 3.75 = 3.292893$$

error estimation

$$\varepsilon_{29}^{(\lambda)} = 0, \quad \varepsilon_{29}^{(\mathrm{v})} = 1.58 \times 10^{-8}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_{30}=\begin{bmatrix}\frac{1}{4}&\frac{1}{2}&0\\ \frac{1}{2}&\frac{1}{4}&\frac{1}{2}\\ 0&\frac{1}{2}&\frac{1}{4}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-0.500000\\0.707107\\-0.500000\end{bmatrix}\rightarrow\ \mathbf{x}_{30}\rightarrow\ \mathbf{v}_{30}=\begin{bmatrix}0.500000\\-0.707107\\0.500000\end{bmatrix}$$

$$\Lambda_{30}=\mathbf{v}_{29}^{t}\mathbf{x}_{30}=-2.187673,\quad \lambda_{30}=-\frac{1}{2.187673}+3.75=3.292893$$

error estimation

$$\varepsilon_{30}^{(\lambda)}=1.35\times10^{-16},\qquad \varepsilon_{30}^{(\mathrm{v})}=2.75\times10^{-8}$$

……

$$\tilde{\mathbf{A}}\,\mathbf{x}_{50}=\begin{bmatrix}\frac{1}{4}&\frac{1}{2}&0\\ \frac{1}{2}&\frac{1}{4}&\frac{1}{2}\\ 0&\frac{1}{2}&\frac{1}{4}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-0.500514\\0.707107\\-0.500514\end{bmatrix}\rightarrow\ \mathbf{x}_{50}=\begin{bmatrix}1.085564\\-1.546912\\1.102099\end{bmatrix}$$

$$\mathbf{v}_{50}=\frac{\mathbf{x}_{50}}{\left(\mathbf{x}_{50}^{\mathbf{t}}\cdot\mathbf{x}_{50}\right)^{\frac{1}{2}}}=\begin{bmatrix}0.496214\\-0.707097\\0.503772\end{bmatrix}$$

$$\Lambda_{50}=\mathbf{v}_{49}^{t}\mathbf{x}_{50}=-2.187662,\quad \lambda_{50}=\Lambda_{50}^{-1}+l=-\frac{1}{2.187662}+3.75=3.292882$$

error estimation

$$\varepsilon_{50}^{(\lambda)}=2.35\times10^{-6},\qquad \varepsilon_{50}^{(\mathrm{v})}=4.77\times10^{-3}$$

.....

$$\tilde{\mathbf{A}}\,\mathbf{x}_{58}=\begin{bmatrix}\frac{1}{4}&\frac{1}{2}&0\\ \frac{1}{2}&\frac{1}{4}&\frac{1}{2}\\ 0&\frac{1}{2}&\frac{1}{4}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-0.712203\\0.664212\\-0.227135\end{bmatrix}\rightarrow\ \mathbf{x}_{58}\rightarrow\ \mathbf{v}_{58}=\begin{bmatrix}0.023209\\-0.588084\\0.808467\end{bmatrix}$$

$$\Lambda_{58}=\mathbf{v}_{57}^{t}\mathbf{x}_{58}=-1.459722,\quad \lambda_{58}=\Lambda_{58}^{-1}+3.75=3.064938$$

error estimation

$$\varepsilon_{58}^{(\lambda)} = 5.62\times10^{-2}, \quad \varepsilon_{58}^{(\mathrm{v})} = 5.22\times10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_{59} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.023209 \\ -0.588084 \\ 0.808467 \end{bmatrix} \rightarrow \mathbf{x}_{59} \rightarrow \mathbf{v}_{59} = \begin{bmatrix} -0.863853 \\ 0.448093 \\ 0.230153 \end{bmatrix}$$

$$\Lambda_{59} = \mathbf{v}_{58}^{t}\mathbf{x}_{59} = -0.279919, \quad \lambda_{59} = \Lambda_{59}^{-1} + 3.75 = 0.177535$$

error estimation

$$\varepsilon_{59}^{(\lambda)} = 1.63\times10^{1}, \quad \varepsilon_{59}^{(\mathrm{v})} = 5.95\times10^{-1}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_{60} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.863853 \\ 0.448093 \\ 0.230153 \end{bmatrix} \rightarrow \mathbf{x}_{60} \rightarrow \mathbf{v}_{60} = \begin{bmatrix} -0.440870 \\ -0.289111 \\ 0.849734 \end{bmatrix}$$

$$\Lambda_{60} = \mathbf{v}_{59}^{t}\mathbf{x}_{60} = 1.515181, \quad \lambda_{60} = \Lambda_{60}^{-1} + 3.75 = 4.409987$$

error estimation

$$\varepsilon_{60}^{(\lambda)} = 9.60\times10^{-1}, \quad \varepsilon_{60}^{(\mathrm{v})} = 4.43\times10^{-1}$$

.....

$$\tilde{\mathbf{A}}\,\mathbf{x}_{70} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.708086 \\ 0.001387 \\ 0.706125 \end{bmatrix} \rightarrow \mathbf{x}_{70} \rightarrow \mathbf{v}_{70} = \begin{bmatrix} -0.706570 \\ -0.000759 \\ 0.707643 \end{bmatrix}$$

$$\Lambda_{70} = \mathbf{v}_{69}^{t}\mathbf{x}_{70} = 3.999885, \quad \lambda_{70} = \Lambda_{70}^{-1} + l = \frac{1}{3.999885} + 3.75 = 4.000001$$

error estimation

$$\varepsilon_{70}^{(\lambda)} = 8.72\times10^{-7}, \quad \varepsilon_{70}^{(\mathrm{v})} = 1.29\times10^{-3}$$

.....

Finally

$$\tilde{\mathbf{A}}\,\mathbf{x}_{89} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.707107 \\ 0.000000 \\ 0.707107 \end{bmatrix} \rightarrow \mathbf{x}_{89} \rightarrow \mathbf{v}_{89} = \begin{bmatrix} -0.707107 \\ 0.000000 \\ 0.707107 \end{bmatrix}$$

$$\Lambda_{89} = \Lambda_{\text{final}} = 4.000000 \quad \rightarrow \quad \lambda_{\text{final}} = \Lambda_{\text{final}}^{-1} + l = \frac{1}{3.999885} + 3.75 = 4.00000 = \lambda_{\text{second}}$$

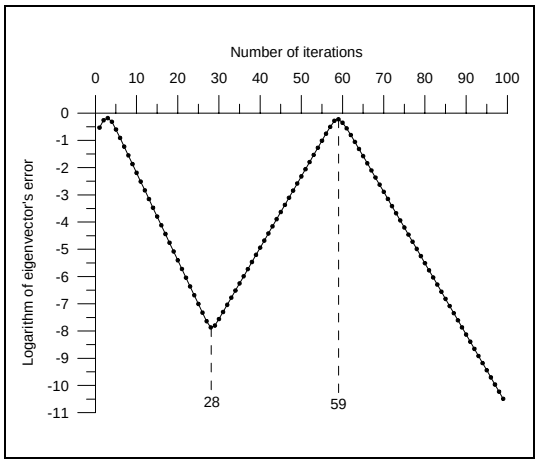
$$\mathbf{v}_{89} = \mathbf{v}_{\text{final}} = \{-0.707107,\ 0.000000,\ 0.707107\} = \mathbf{v}_{\text{second}}$$

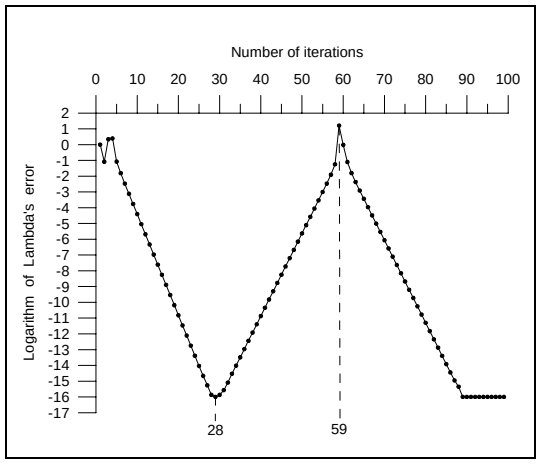
error estimation

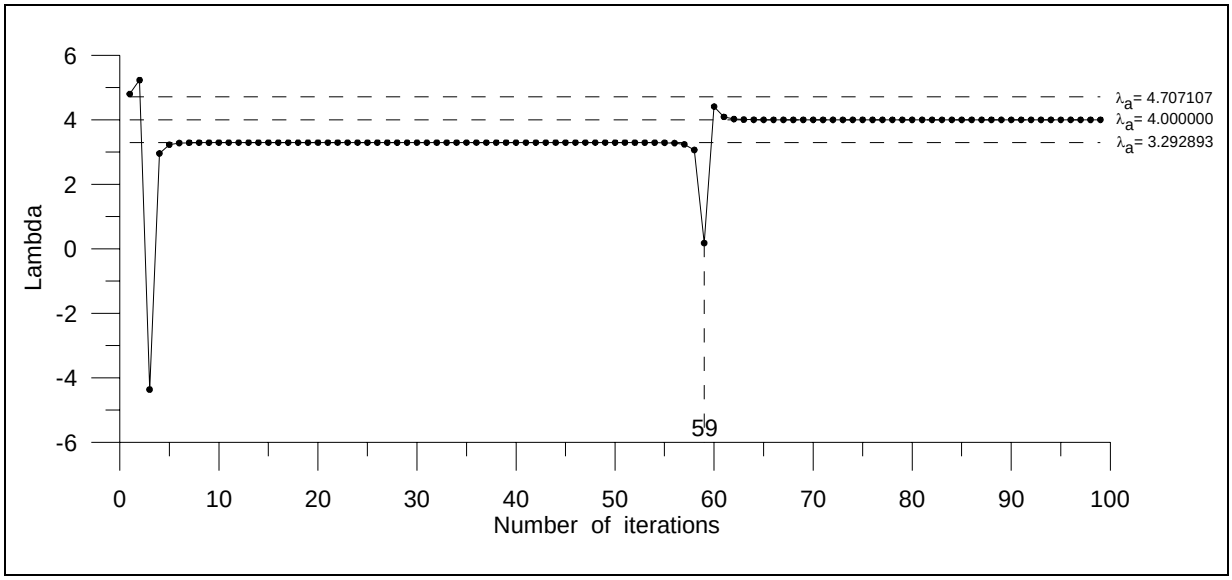
$$\varepsilon_{89}^{(\lambda)} = 0\,, \qquad \varepsilon_{89}^{(\text{v})} = 1.35\times 10^{-8}$$

Remarks

- for error treshold 10^{-7} only 86 iterations is sufficient because
  $\varepsilon_{86}^{(\lambda)} = 3.55\times 10^{-15}$,
  $\varepsilon_{86}^{(\text{v})} = 8.27\times 10^{-8}$
- notice: after 25 iterations we also have small estimation error
  $\varepsilon_{25}^{(\lambda)} = 9.17\times 10^{-15}$,
  $\varepsilon_{25}^{(\text{v})} = 9.86\times 10^{-8}$
  however, the results
  $\lambda_{25} = 3.292893 \approx \lambda_{\text{third}}$
  $\mathbf{v}_{25} = \{-0.500000,\ 0.707107,\ -0.500000\} \approx \mathbf{v}_{\text{third}}$
  correspond to a false solution, namely to the third eigenvalue λ_{third}, and eigenvector $\mathbf{v}_{\text{third}}$.







(ii) Let us use now a non-symetric starting vector

$$\mathbf{x}_0 = \{1,\ 1,\ -1\}$$

then

$$\mathbf{v}_0 = \frac{\mathbf{x}_0}{\left(\mathbf{x}_0^{\mathbf{t}} \cdot \mathbf{x}_0\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{3}}\{1,\ 1,\ -1\}$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \rightarrow \mathbf{x}_1 = \frac{4}{7\sqrt{3}} \begin{bmatrix} 9 \\ -1 \\ -5 \end{bmatrix}$$

$$\mathbf{v}_1 = \frac{\mathbf{x}_1}{\left(\mathbf{x}_1^{\mathbf{t}} \cdot \mathbf{x}_1\right)^{\frac{1}{2}}} = \frac{1}{\sqrt{107}}\ \{9,\ -1,\ -5\}$$

$$\Lambda_1 = \mathbf{v}_0^T \mathbf{x}_1 = \frac{52}{21} = 2.476191\ \rightarrow\ \lambda_{(1)} = \Lambda_1^{-1} + l = \frac{21}{52} + 3.75 = 4.153846$$

$$\tilde{\mathbf{A}}\,\mathbf{x}_2 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{\sqrt{107}} \begin{bmatrix} 9 \\ -1 \\ -5 \end{bmatrix} \rightarrow\ \mathbf{x}_2 = \frac{4}{7\sqrt{107}} \begin{bmatrix} 45 \\ 9 \\ -57 \end{bmatrix}$$

$$\Lambda_2 = \mathbf{v}_1^t \mathbf{x}_2 = 3.530040, \quad \lambda_{(2)} = \Lambda_2^{-1} + l = \frac{1}{3,530040} + 3.75 = 4.033283$$

…….

Finally

After 39 iterations we obtain as before

$$\Lambda_{final} = 4.000000 \rightarrow \lambda_{final} = \Lambda_{final}^{-1} + l = \frac{1}{4} + 3.75 = 4.000000 = \lambda_2$$

$$\mathbf{v} = \{0.707107,\ 0.000000,\ 0.707107\} \approx \left\{\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}\right\}$$

The error level is 10^{-10} then.

Remark

Due to different starting vector convergence in the *(ii)* case proved to be much faster than in the case *(i)*.

6.6. THE GENERALIZED EIGENVALUE PROBLEM

$$\mathbf{Ax} = \lambda \mathbf{Bx}$$

e.g. from FEM where $\mathbf{A} = \mathbf{K}$, $\mathbf{B} = \mathbf{M}$
K – stiffness matrix, M – inertia matrix

If

$$\underset{symmetric}{\mathbf{B} = \mathbf{B}^t} \quad \text{and} \quad \underset{positive\ \ definite}{\mathbf{x}^t\mathbf{Bx} > \mathbf{0}}$$

then

$$\mathbf{B} = \mathbf{LL}^t, \qquad \mathbf{L}^t\mathbf{x} = \mathbf{y} \quad \rightarrow \quad \mathbf{x} = \mathbf{L}^{-t}\mathbf{y}$$

$$\mathbf{Ax} = \lambda \mathbf{LL}^t\mathbf{x} \quad \rightarrow \quad \mathbf{AL}^{-t}\mathbf{y} = \lambda \mathbf{Ly}$$

$$\underbrace{\mathbf{L}^{-1}\mathbf{AL}^{-t}}\mathbf{y} = \lambda \mathbf{y}$$

$$\hat{\mathbf{A}}\boldsymbol{y} = \lambda \boldsymbol{y}, \qquad \hat{\mathbf{A}} = \mathbf{L}^{-1}\mathbf{AL}^{-t}$$

Remark

The generalized eigenvalue problem was transformed into the standard one with the same eigenvalues λ preserved.

SOLUTION ALGORITHM BASED ON THE POWER METHOD

(Search for the largest eigenvalue $\lambda_{\max}$)

ASSUMPTION $\mathbf{y_0}$

NORMALIZATION $\mathbf{u}_n \equiv \dfrac{\mathbf{y}_n}{\left(\mathbf{y}_n^t \cdot \mathbf{y}_n\right)^{1/2}}$

POWER STEP $\mathbf{L}^{-1}\mathbf{AL}^{-t}\mathbf{u}_n = \mathbf{y}_{n+1}$

$\mathbf{B} = \mathbf{LL}^t$ - matrix decomposition (only once)

How it is done: $\mathbf{L}^t\mathbf{x}_n = \mathbf{u}_n \rightarrow \mathbf{x}_n$ - step back

$\mathbf{Ly}_{n+1} = \mathbf{Ax}_n \rightarrow \mathbf{y}_{n+1}$ - step foreward

RAYLEIGH QUOTIENT $\Lambda_n = \dfrac{\mathbf{u}_n^t\left(\mathbf{L}^{-1}\mathbf{AL}^{-t}\right)\mathbf{u}_n}{\underbrace{\mathbf{u}_n^t \cdot \mathbf{u}_n}_{=1}} = \mathbf{u}_n^t\mathbf{y}_{n+1} = \mathbf{x}_n^t\mathbf{Ax}_n$

ERROR ESTIMATION $\varepsilon_n^{(\lambda)} = \left|\dfrac{\Lambda_n - \Lambda_{n-1}}{\Lambda_n}\right|$, $\varepsilon_n^{(u)} = \left|u_n - u_{n-1}\right|$

BRAKE OFF TEST $\varepsilon_n^{(\lambda)} \underset{\substack{yes \\ \downarrow \\ \lambda_{\max} \approx \Lambda_n}}{<} B_\lambda$, $\varepsilon_n^{(u)} < B_u$, B_λ, B_u imposed required precision of errors $\varepsilon^{(\lambda)}$ and $\varepsilon^{(u)}$ evaluation.

RESULTS $\lambda_{\max} \approx \Lambda_n$, $\mathbf{x}_1 \approx \mathbf{x}_n$

Example

$$\mathbf{Ax} = \lambda \mathbf{Bx}$$

$$\begin{bmatrix} 7 & 4 & 3 \\ 4 & 8 & 2 \\ 3 & 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \begin{bmatrix} 8 & 1 & 3 \\ 1 & 6 & 4 \\ 3 & 4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

(i) DIRECT APPROACH

1. Decompose $\mathbf{B} = \mathbf{L}\mathbf{L}^t$

$$\begin{bmatrix} 8 & 1 & 3 \\ 1 & 6 & 4 \\ 3 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 2.828427 & 0 & 0 \\ 0.353553 & 2.423840 & 0 \\ 1.060660 & 1.495561 & 0.798935 \end{bmatrix} \begin{bmatrix} 2.828427 & 0.353553 & 1.060660 \\ 0 & 2.423840 & 1.495561 \\ 0 & 0 & 0.798935 \end{bmatrix}$$

2. Find inverse matrices $\mathbf{L}^{-1}, \mathbf{L}^{-t}$

$$\mathbf{L}^{-1} = \begin{bmatrix} 0.353555 & 0 & 0 \\ -0.051571 & 0.412568 & 0 \\ -0.372836 & -0.772303 & 1.251663 \end{bmatrix}, \quad \mathbf{L}^{-t} = \begin{bmatrix} 0.353553 & -0.051571 & -0.372836 \\ 0 & 0.412568 & -0.772303 \\ 0 & 0 & 1.251663 \end{bmatrix}$$

3. Find $\hat{\mathbf{A}} = \mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-t}$

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.353553 & 0 & 0 \\ -0.051571 & 0.412568 & 0 \\ -0.372836 & -0.772303 & 1.251663 \end{bmatrix} \begin{bmatrix} 7 & 4 & 3 \\ 4 & 8 & 2 \\ 3 & 2 & 6 \end{bmatrix} \begin{bmatrix} 0.353553 & -0.051571 & -0.372836 \\ 0 & 0.412568 & -0.772303 \\ 0 & 0 & 1.251663 \end{bmatrix}$$

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.874995 & 0.455828 & -0.687333 \\ 0.455828 & 1.210105 & -2.031250 \\ -0.687333 & -2.031250 & 10.781520 \end{bmatrix}$$

4. Find eigenvalues λ and eigenvectors $\mathbf{u}$ of $\tilde{\mathbf{A}}$ by any standard method

$\lambda_1 = 11.251110$, $\lambda_2 = 1.116762$, $\lambda_3 = 0.498742$

$$\mathbf{u}_1 = \begin{bmatrix} -0.073535 \\ -0.200948 \\ 0.976838 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0.717447 \\ 0.669696 \\ 0.191774 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -0.692722 \\ 0.714931 \\ 0.094923 \end{bmatrix}$$

5. Find the eigenvectors of the original problem $\mathbf{x} = \mathbf{L}^{-t}\mathbf{u}$

$$\begin{bmatrix} 0.353553 & -0.051571 & -0.372836 \\ 0 & 0.412568 & -0.772303 \\ 0 & 0 & 1.251663 \end{bmatrix} \left[\begin{array}{ccc} -0.073535 & 0.717447 & -0.692722 \\ -0.200948 & 0.669696 & 0.714931 \\ 0.976838 & 0.191774 & 0.094923 \end{array}\right] =$$

$$= \begin{bmatrix} -0.379836 & 0.147619 & -0.317174 \\ -0.837319 & 0.128188 & 0.221647 \\ 1.222672 & 0.240036 & 0.118812 \end{bmatrix} \Rightarrow \left[\begin{array}{ccc} -0.248290 & 0.476833 & -0.775220 \\ -0.547337 & 0.414068 & 0.541741 \\ 0.799233 & 0.775357 & 0.290393 \end{array}\right]$$

In the last matrix all eigenvectors $\mathbf{x}$ are normalized

(ii) MORE EFFICIENT APPROACH

2a Solve simultaneous equations

Step back $\mathbf{L}^t\mathbf{x}_n = \mathbf{u}_n \quad \rightarrow \quad \mathbf{x_n}$

$$\begin{bmatrix} 2.820427 & 0.353553 & 1.060660 \\ 0 & 2.423840 & 1.495561 \\ 0 & 0 & 0.798935 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_n = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}_n \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \cdots$$

Step forward $\mathbf{L}\mathbf{y}_{n+1} = \mathbf{A}\mathbf{x}_n \quad \rightarrow \quad \mathbf{y_{n+1}}$

$$\begin{bmatrix} 2.828427 & 0 & 0 \\ 0.353553 & 2.423840 & 0 \\ 1.060660 & 1.495561 & 0.798935 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{n+1} = \begin{bmatrix} 7 & 4 & 3 \\ 4 & 8 & 2 \\ 3 & 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \cdots$$

Let

$$\mathbf{y}_1 = \{1,\ 1,\ 1\}$$

$$\mathbf{u}_1 = \frac{\mathbf{y_1}}{\left(\mathbf{y}_1^t\mathbf{y}_1\right)^{1/2}} = \left\{\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}\right\} = \{0.577350,\ 0.577350,\ 0.577350\}$$

$$\mathbf{L}^t\mathbf{x}_1 = \mathbf{u}_1 \rightarrow \mathbf{x}_1 = \{-0.040907,\ -0.207693,\ 0.722648\}$$

$$\mathbf{A}\mathbf{x}_1 = \{1.050824,\ -0.379873,\ 3.797781\}$$

$$\lambda_{(1)} = \mathbf{x}_1^t\mathbf{A}\mathbf{x}_1 = 2.780369$$

$$\mathbf{L}\mathbf{y}_2 = \mathbf{A}\mathbf{x}_1 \rightarrow \mathbf{y}_2 = \{0.371522,\ -0.210916,\ 4.655135\}$$

$$\mathbf{u}_2 = \frac{\mathbf{y}_2}{\left(\mathbf{y}_2^t\mathbf{y}_2\right)^{1/2}} = \{0.079475,\ -0.045119,\ 0.995815\}$$

$$\mathbf{L}^t\mathbf{x}_2 = \mathbf{u}_2 \rightarrow \mathbf{x}_2 = \{-0.340850, -0.787685, 1.246425\}$$
$$\mathbf{A}\mathbf{x}_2 = \{-1.797415, -5.172031, 4.880630\}$$
$$\lambda_{(2)} = \mathbf{x}_2^t\mathbf{A}\mathbf{x}_2 = 10.769220$$
.....

Finally

$$\mathbf{L}^t\mathbf{x} = \mathbf{u} \rightarrow \mathbf{x} = \{-0.379836, -0.837319, 1.222672\}$$
$$\mathbf{A}\mathbf{x} = \{-2.340110, \quad -5.772553, \quad 4.521886\}$$
$$\lambda_{\max} = \mathbf{x}^t\mathbf{A}\mathbf{x} = 11.251110$$

6.7. THE JACOBI METHOD

The Jacobi method for finding all of the eigenvalues and eigenvectors of a symmetric, positive definite matrix **A** is based on consecutive simple orthogonal transformations in order to convert this matrix A into the diagonal diagonal form.

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}\ , \qquad \mathbf{Q}^t\mathbf{A}\mathbf{Q} = \mathbf{D} \quad \text{diagonal}\ , \qquad \mathbf{A} = \left[a_{ij}\right]$$

Let

$$\mathbf{U} = \begin{array}{c} \\ p \\ \\ \\ q \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} & & & p & & & q \end{array} \\ \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & c & & & s & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & -s & & & c & \\ & & & & & & 1 \end{bmatrix} \end{array}$$

$$\mathbf{A}' = \mathbf{U}^t\mathbf{A}\mathbf{U}\ , \qquad \mathbf{A}' = \left[a'_{ij}\right]$$

$$\mathbf{A}' = \begin{array}{c} \\ p \\ \\ \\ q \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} & & & p & & & q \end{array} \\ \begin{bmatrix} & III & & \longleftarrow II \longrightarrow & & & \\ & & & \downarrow & & & \\ & & a'_{pp} & & & a'_{pq} & \\ & & & I & & & \\ & & a'_{qp} & & & a'_{qq} & \\ & & & & & & \end{bmatrix} \end{array}$$

After transformation $\mathbf{U}^t\mathbf{A}\mathbf{U}$ we have

$$\text{zone } III \begin{cases} a'_{ij} = a_{ij} \end{cases}$$

except:

$$\text{zone } I \begin{cases} pp,\ pq,\ qp,\ qq \ terms: \end{cases} \begin{cases} a'_{pp} = c^2 a_{pp} + s^2 a_{qq} - 2csa_{pq} \\ a'_{qq} = c^2 a_{qq} + s^2 a_{pp} + 2csa_{pq} \\ a'_{pq} = a'_{qp} = (c^2 - s^2)a_{pq} + cs(a_{pp} - a_{qq}) \end{cases}$$

$$\text{zone } II \begin{cases} pth \ \ and \ \ qth \ \ rows \ \ (j \neq p \ \ and \ \ j \neq q) \ \ : \begin{cases} a'_{pj} = ca_{pj} - sa_{qj} \\ a'_{qj} = sa_{pj} + ca_{qj} \end{cases} \\ pth \ \ and \ \ qth \ \ columns \ \ (i \neq p \ \ and \ \ i \neq q) : \begin{cases} a'_{ip} = ca_{ip} - sa_{iq} \\ a'_{iq} = sa_{ip} + ca_{iq} \end{cases} \end{cases}$$

6.7.1. Conditions imposed on transformation

1. orthogonality:

$$\mathbf{U}^{-1} = \mathbf{U}^t \quad \text{because} \quad \mathbf{U}^t\mathbf{U} = \mathbf{I} \rightarrow c^2 + s^2 = 1$$

2. a'_{pq} term anihilation

$$a'_{pq} = a'_{qp} = 0$$

In order to satisfy the first condition we assume

$$c = \cos\vartheta$$

$$s = \sin\vartheta$$

hence

$$\left(\cos^2\vartheta - \sin^2\vartheta\right)a_{pq} + \sin\vartheta\cos\vartheta\left(a_{pp} - a_{qq}\right) = 0$$

Since

$$\cos^2\vartheta - \sin^2\vartheta = \cos 2\vartheta, \quad \sin\vartheta\cos\vartheta = \frac{1}{2}\sin 2\vartheta$$

The second condition

$$a'_{pq} = a_{pq}\cos 2\vartheta + \frac{1}{2}\left(a_{pp} - a_{qq}\right)\sin 2\vartheta = 0$$

yield

$$\tan 2\vartheta = -\frac{2a_{pq}}{a_{pp} - a_{qq}} \rightarrow 2\vartheta$$

Remarks:

- Appropriately chosen rotation of the coordinate system about the angle ϑ provides annihilation of a'_{pq}
- In fact we need rather $\cos\vartheta$ and $\sin\vartheta$ than ϑ itself. They may be found from the following formulas:

$$\alpha = \frac{1}{2}\left(a_{pp} - a_{qq}\right)$$

$$\beta = \left(\alpha^2 + a_{pq}^2\right)^{1/2}$$

$$c = \cos\vartheta = \left(\frac{1}{2} + \frac{|\alpha|}{2\beta}\right)^{1/2}$$

$$s = \sin\vartheta = \frac{-a_{pq}}{2\beta\cos\vartheta}\,\mathrm{sgn}(\alpha), \quad \textit{where} \quad \mathrm{sgn}(\alpha) = \begin{cases} 1 & \alpha > 0 \\ -1 & \alpha < 0 \\ 0 & \alpha = 0 \end{cases}$$

SOLUTION PROCEDURE

We have to annihilate all off-diagonal terms, however, subsequent annihilation steps usually destroy the previously done. Fortunately such process is convergent to the final solution. The final matrix, through usually not strictly diagonal has off-diagonal terms as small as required when compared with diagonal ones.

Thus solution procedure is continued until brake-off test is satisfied. It is based either on the average or on the maximum error norms defined below:

Average norms

$${}_a v = \left(\frac{1}{n^2 - n}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}^2\right)^{1/2}$$ - The average norm of the off-diagonal elements

$${}_a w = \left(\frac{1}{n}\sum_{i=1}^{n} a_{ii}^2\right)^{1/2}$$ - The average norm of diagonal elements

Maximum norms

$${}_m v = \max_{\substack{i,j \\ i\neq j}} \left|a_{ij}\right|$$ - The maximum norm of off-diagonal elements

$${}_m w = \max_{i} \left|a_{ii}\right|$$ - The maximum norm of diagonal elements

BRAKE OFF TEST

$$v \le \varepsilon w$$

where: v, w are equal either to ${}_a v$, ${}_a w$ or to ${}_m v$, ${}_m w$

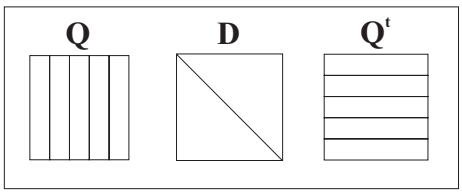
ε is a given threshold e.g. $\varepsilon = 10^{-6}$

Finally

$$\underbrace{\mathbf{U}_k^t \cdots \mathbf{U}_2^t \mathbf{U}_1^t}_{\mathbf{Q}^t} \mathbf{A} \underbrace{\mathbf{U}_1 \mathbf{U}_2 \cdots \mathbf{U}_k}_{\mathbf{Q}} \approx \mathbf{Q}^t \mathbf{A} \mathbf{Q} = \mathbf{D}$$

Hence matrix decomposition is

$$\mathbf{A} = \mathbf{Q}^{-\mathrm{T}} \mathbf{D} \mathbf{Q}^{-1} = \mathbf{Q} \mathbf{D} \mathbf{Q}^{\mathrm{T}}$$



Q - orthogonal matix - its columns are eigenvectors of the matrix **A**
D - diagonal matrix composed of eigenvalues of the matrix **A**

If it is desired to compute the eigenvectors along with the eingevalues, this can be accomplished by initializing the matrix $\mathbf{Q}$ as $\mathbf{I}$, and then modifying $\mathbf{Q}$ along with modifications to $\mathbf{A}$ in the following way

p-th column $q'_{ip} = cq_{ip} - sq_{iq}$

q-th column $q'_{iq} = sq_{ip} + cq_{iq}$

All other columns remain unchanged i.e. $q'_{ij} = q_{ij}$

Columns of the final $\mathbf{Q}$ are the eigenvectors of the original matrix **A**, while its eigenvalues are given by the diagonal terms of the matrix **D**.

Example

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 3 & 7 \\ 2 & 8 & 5 & 1 \\ 3 & 5 & 12 & 9 \\ 7 & 1 & 9 & 7 \end{bmatrix}$$

required precision is given by ε=10^{-6}

SOLUTION PROCEDURE

Average off-diagonal terms of the matrix **A**

$$ {}_a v_0 = \left(\frac{1}{4^2 - 4} \cdot 2 \cdot \left[2^2 + 3^2 + 7^2 + 5^2 + 1^2 + 9^2 \right] \right)^{\frac{1}{2}} = 5.307228 $$

In order to find a reasonable off-diagonal element of the matrix **A** we seek for the first element greater then ${}_a v_0$. Thus we have:

$$ a_{41} = a_{14} = 7 \quad > \quad {}_a v_0 = 5.307228 \qquad \rightarrow \quad p = 1, \quad q = 4 $$

and anihilated will be element $\mathrm{a}_{41} = \mathrm{a}_{14} = 7$. We find then:

$$ \alpha = \frac{1}{2}\left(a_{11} - a_{44} \right) = \frac{1}{2}\left(4 - 7 \right) = -1.500000 $$

$$ \beta = \left(a_{14}^2 + \alpha^2 \right)^{\frac{1}{2}} = \left(7^2 + \left(-\frac{3}{2} \right)^2 \right)^{\frac{1}{2}} = 7.158911 $$

$$ c = \left(\frac{1}{2} + \frac{|\alpha|}{2\beta} \right)^{\frac{1}{2}} = \left(\frac{1}{2} + \frac{1.5}{2\left(7.158911 \right)} \right)^{\frac{1}{2}} = 0.777666 $$

$$ s = \frac{\alpha\left(-a_{14} \right)}{2\beta |\alpha| c} = \frac{\left(-1.5 \right)\left(-7 \right)}{2\left(7.158911 \right)\left| -1.5 \right|\left(.777666 \right)} = 0.628678 $$

Now we find

$$ \begin{cases} a'_{11} = c^2 a_{11} + s^2 a_{44} - 2csa_{14} = \left(0.777666 \right)^2 \cdot 4 + \left(0.628678 \right)^2 \cdot 7 - 2\left(0.777666 \right)\left(0.628678 \right) \cdot 7 = \\ \qquad = -1.658911 \\ a'_{44} = c^2 a_{44} + s^2 a_{11} + 2csa_{14} = (0.777666)^2 \cdot 7 + (0.628678)^2 \cdot 4 + 2(0.777666)(0.628678) \cdot 7 = \\ \qquad = 12.658910 \end{cases} $$

$$ a'_{12} = a'_{21} = ca_{12} - sa_{42} = 0.777666 \cdot 2 - 0.628678 \cdot 1 = 0.926655 $$

$$ a'_{13} = a'_{31} = ca_{13} - sa_{43} = 0.777666 \cdot 3 - 0.628678 \cdot 9 = -3.325099 $$

$$ a'_{42} = a'_{24} = sa_{12} + ca_{42} = 0.628678 \cdot 2 + 0.777666 \cdot 1 = 2.035051 $$

$$ a'_{43} = a'_{34} = sa_{13} + ca_{43} = 0.628678 \cdot 3 + 0.777666 \cdot 9 = 8.885027 $$

$$ \mathbf{A}' = \begin{bmatrix} -1.658911 & 0.926655 & -3.325099 & 0 \\ 0.926655 & 8 & 5 & 2.035021 \\ -3.325099 & 5 & 12 & 8.885027 \\ 0 & 2.035021 & 8.885027 & 12.658910 \end{bmatrix} $$

Search for **Q** matrix:

initially $\mathbf{Q} = \mathbf{I}$

after the first iteration

$$\mathbf{q}'_{11} = cq_{11} - sq_{14} = 0.777666 \cdot 1 - 0.628678 \cdot 0 = 0.777666$$
$$\mathbf{q}'_{21} = cq_{21} - sq_{24} = c \cdot 0 - s \cdot 0 = 0$$
$$\mathbf{q}'_{31} = cq_{31} - sq_{34} = c \cdot 0 - s \cdot 0 = 0$$
$$\mathbf{q}'_{41} = cq_{41} - sq_{44} = c \cdot 0 - s \cdot 1 = -0.628678$$
$$\mathbf{q}'_{14} = sq_{11} + cq_{14} = s \cdot 1 + c \cdot 0 = 0.628678$$
$$\mathbf{q}'_{24} = sq_{21} + cq_{24} = s \cdot 0 + c \cdot 0 = 0$$
$$\mathbf{q}'_{34} = sq_{31} + cq_{34} = s \cdot 0 + c \cdot 0 = 0$$
$$\mathbf{q}'_{44} = sq_{41} + cq_{44} = s \cdot 0 + c \cdot 1 = 0.777666$$

$$\mathbf{Q} = \begin{bmatrix} 0.777666 & 0 & 0 & 0.628678 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -0.628678 & 0 & 0 & 0.777666 \end{bmatrix}$$

BRAKE OFF TESTS

Let required precision of the solution is

$$\varepsilon = 10^{-6}$$

(i) For the initial matrix

Average norms

$${}_a w_0 = \left[\frac{1}{4} \left(4^2 + 8^2 + 12^2 + 7^2 \right) \right]^{\frac{1}{2}} = 8.261356 \qquad \text{- diagonal}$$

$${}_a v_0 = \left[\frac{1}{4^2 - 4} \cdot 2 \left(2^2 + 3^2 + 7^2 + 5^2 + 1^2 + 9^2 \right) \right]^{\frac{1}{2}} = 5.307228 \qquad \text{- off-diagonal}$$

$$\frac{{}_a v_0}{{}_a w_0} = \frac{5.307228}{8.261356} = 0.642416 > \varepsilon = 10^{-6}$$

Maximum norms

$${}_m w_0 = \max_i \left| a_{ii} \right| = 12 \qquad \text{- diagonal terms}$$

$${}_m v_0 = \max_{\substack{i,j, \\ i \neq j}} a_{ij} = 9 \qquad \text{- off-diagonal terms}$$

$$\frac{{}_m v_0}{{}_m w_0} = \frac{9}{12} = 0.75 \; > \; \varepsilon = 10^{-6}$$

(ii) After the first iteration

Average norms

$${}_a w_1 = \left\{ \frac{1}{4} \left[\left(-1.658911\right)^2 + 8^2 + 12^2 + 12.65891^2 \right] \right\}^{\frac{1}{2}} = 9.63068$$

$${}_a v_1 = \left\{ \frac{1}{4^2 - 4} \cdot 2 \left[0.926655^2 + \left(-3.325099\right)^2 + 0^2 + 5^2 + 2.035021^2 + 8.885027^2 \right] \right\}^{\frac{1}{2}} = 4.472136$$

Maximum norms

$${}_m w_1 = 12.65891 \qquad \text{- diagonal terms}$$
$${}_m v_1 = 8.885027 \qquad \text{- off-diagonal terms}$$

$$\frac{{}_a v_1}{{}_a w_1} = \frac{4.4721357}{9.63068} = 0.464363 \; > \; \varepsilon \quad , \quad \frac{{}_m v_1}{{}_m w_1} = \frac{8.885027}{12.65891} = 0.701879 \; > \; \varepsilon$$

Finally after all iterations the following matrices are obtained:

$$\mathbf{A'} = \begin{bmatrix} -3.233881 & 4.89 \times 10^{-9} & -6.06 \times 10^{-15} & 1.72 \times 10^{-5} \\ 4.89 \times 10^{-9} & 3.739112 & 0 & -3.78 \times 10^{-10} \\ -6.06 \times 10^{-15} & 0 & 23.04466 & 5.14 \times 10^{-6} \\ 1.72 \times 10^{-5} & -3.78 \times 10^{-10} & 5.14 \times 10^{-6} & 7.450091 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} 0.580781 & 0.678728 & 0.345658 & 0.287300 \\ -0.203742 & 0.374957 & 0.311701 & -0.848963 \\ 0.365143 & -0.617460 & 0.688355 & -0.107607 \\ -0.698465 & 0.132200 & 0.556353 & 0.430280 \end{bmatrix}$$

Precision of the final solution

Average norms

$${}_a w = \left\{ \frac{1}{4} \left[\left(-3.233881\right)^2 + 3.739112^2 + 23.04466^2 + 7.450091^2 \right] \right\}^{\frac{1}{2}} = 12.359199$$

$$
{}_{a}v = \left\{ \frac{1}{4^2 - 4} \cdot 2[(4.89 \times 10^{-9})^2 + (-6.06 \times 10^{-15})^2 + (1.72 \times 10^{-5})^2 + \right.
$$

$$
\left. + 0^2 + (-3.78 \times 10^{-10})^2 + (5.14 \times 10^{-6})^2] \right\}^{\frac{1}{2}} = 7.33 \times 10^{-6}
$$

Brake-off test

$$
\frac{{}_{a}v}{{}_{a}w} = \frac{7.33 \times 10^{-6}}{12.36} = 5.93 \times 10^{-7} < \varepsilon = 10^{-6}
$$

Maximum norms

$$
{}_{m}w = 23.04466
$$

$$
{}_{m}v = 1.72 \times 10^{-5}
$$

Brake-off test

$$
\frac{{}_{m}v}{{}_{m}w} = \frac{1.72 \times 10^{-5}}{23.04466} = 7.46 \times 10^{-7} < \varepsilon = 10^{-6}
$$

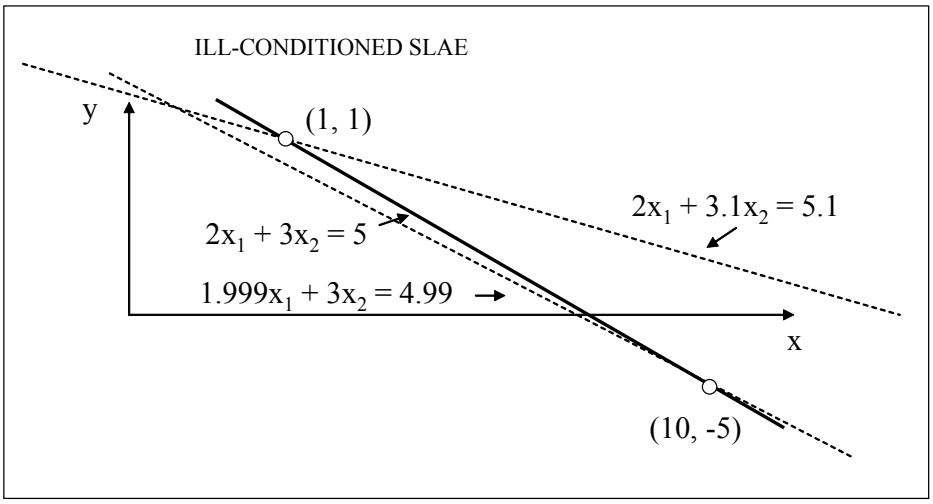
7. ILL-CONDITIONED SYSTEMS OF SIMULTANEOUS LINEAR EQUATIONS

7.1. INTRODUCTION

Example

$$\text{I} \quad \left.\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 2x_1 + 3.1x_2 &= 5.1 \end{aligned}\right\} \quad \rightarrow \quad x_1 = x_2 = 1$$

$$\text{II} \quad \left.\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 1.999x_1 + 3x_2 &= 4.99 \end{aligned}\right\} \quad \rightarrow \quad x_1 = 10, \quad x_2 = -5$$



7.2. SOLUTION APPROACH

Questions:

1. What is the phenomenon of ill-conditioning?
2. How ill-conditioning may be measured?
3. What may be done in order to overcome ill-conditioning problem?

AD 1. ILL CONDITIONING PHENOMENON

Thus a very *small change* in the *coefficients* gives rise to a *large change* in the *solution*. Such behavior characterizes what is called an *ill-conditioned* system. Ill-conditioning causes problem since we cannot carry out computations with *infinite* precision.

AD 2. HOW TO MEASURE ILL-CONDITIONING

(i) The first approach

Let

$$\mathbf{Ax} = \mathbf{b} \quad \rightarrow \quad \mathbf{x_T} \equiv \mathbf{A}^{-1}\mathbf{b}$$
$$\mathbf{Ax_c} = \mathbf{b} + \mathbf{r} \quad \rightarrow \quad \mathbf{r} \equiv \mathbf{Ax}_c - \mathbf{b}$$
$$\mathbf{x}_T - \mathbf{x}_c = -\mathbf{A}^{-1}\mathbf{r} \equiv -\mathbf{e}$$

where

$\mathbf{x_T}$ – true solution
$\mathbf{x_c}$ – computed solution
$\mathbf{r}$ – denote a residuum, due to rounding-off errors r may differ from zero
$\mathbf{e}$ – solution error

From the properties of norm

$$\|\mathbf{By}\| \le \|\mathbf{B}\| \ \|\mathbf{y}\|$$

Also, since $\mathbf{y} = \mathbf{B}^{-1}\mathbf{By}$ we have

$$\|\mathbf{y}\| \le \|\mathbf{B}^{-1}\| \ \|\mathbf{By}\| \rightarrow \boxed{\frac{\|\mathbf{y}\|}{\|\mathbf{B}^{-1}\|} \le \|\mathbf{By}\| \le \|\mathbf{B}\|\|\mathbf{y}\|}$$

Hence we may write

$$\frac{\|\mathbf{r}\|}{\|\mathbf{A}\|} \le \overbrace{\|\mathbf{A}^{-1}\mathbf{r}\|}^{\|\mathbf{e}\|} \le \|\mathbf{A}^{-1}\| \ \|\mathbf{r}\| \quad ,$$

$$\frac{\|\mathbf{b}\|}{\|\mathbf{A}\|} \le \underbrace{\|\mathbf{A}^{-1}\mathbf{b}\|}_{\|\mathbf{x_T}\|} \le \|\mathbf{A}^{-1}\| \ \|\mathbf{b}\| \rightarrow \frac{1}{\|\mathbf{A}^{-1}\| \ \|\mathbf{b}\|} \le \frac{1}{\|\mathbf{x}_T\|} \le \frac{\|\mathbf{A}\|}{\|\mathbf{b}\|} \quad ,$$

since

$$\|\mathbf{A}^{-1}\mathbf{b}\| = \|\mathbf{x_T}\| \qquad \text{and} \qquad \|\mathbf{A}^{-1}\mathbf{r}\| = \|\mathbf{e}\|$$

Multiplying both inequalities we get

$$\frac{1}{\|\mathbf{A}\|\|\mathbf{A}^{-1}\|} \ \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \le \frac{\|\mathbf{e}\|}{\|\mathbf{x_T}\|} \le \|\mathbf{A}\|\|\mathbf{A}^{-1}\| \ \frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$$

Assuming

$$k = \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \quad \text{conditioning number}$$

we find the folowing evaluation

$$\frac{1}{k}\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \le \frac{\|\mathbf{e}\|}{\|\mathbf{x}_T\|} \le k\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$$

Here $\frac{\|\mathbf{e}\|}{\|\mathbf{x}_T\|}$ is the relative error of solution, and $\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$ is the relative residuum.

Thus the quality of the computed solution $\mathbf{x}_c$ is dependent on the value of

$$k = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|, \qquad k \ge 1$$

called the *condition number (cond A)*. Using the spectral norm of matrix $\| \|_S$

$$\|\mathbf{A}\|_S = \left\{ \rho\left(\mathbf{A}^*\mathbf{A}\right) \right\}^{\frac{1}{2}}$$

induced by the Euclidean vector norm

$$\|\mathbf{x}\| = \left\{ \sum_{i=1}^{N} |x_i|^2 \right\}^{\frac{1}{2}}$$

where

$$\rho(\mathbf{A}) = \max_k |\lambda_k|$$

denotes the special radius, we get

$$k = \|\mathbf{A}\|_S \|\mathbf{A}^{-1}\|_S = |\lambda_{max}| \left|\frac{1}{\lambda_{min}}\right| = \left|\frac{\lambda_{max}}{\lambda_{min}}\right|$$

Here $k \ge 1$ (conditioning number) is the measure of the system conditioning (ill- or well-).

(ii) The second approach

Consider the effects of rounding error of the coefficient matrix **A**

Let

$$\mathbf{A}_c = \mathbf{A} + \mathbf{A}_E \ ,$$

where

$\mathbf{A}_c$ = computed matrix, **A** = exact matrix, $\mathbf{A}_E$ =matrix of perturbations

Let

$$\mathbf{A}_c\mathbf{x}_c = \mathbf{b} \ \rightarrow \ \mathbf{x}_c$$

Then

$$\mathbf{x}_T = \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1}\mathbf{A}_c\mathbf{x}_c = \mathbf{A}^{-1}\left(\mathbf{A} + \mathbf{A}_E\right)\mathbf{x}_c \rightarrow \mathbf{x}_T - \mathbf{x}_c = \mathbf{A}^{-1}\mathbf{A}_E\mathbf{x}_c$$

Hence

$$\left\|\mathbf{x}_\mathrm{T} - \mathbf{x}_\mathrm{c}\right\| \ \le \ \left\|\mathbf{A}^{-1}\right\| \left\|\mathbf{A}_\mathrm{E}\right\| \left\|\mathbf{x}_\mathrm{c}\right\| \ = \ k\frac{\left\|\mathbf{A}_\mathrm{E}\right\|}{\left\|\mathbf{A}\right\|}\left\|\mathbf{x}_\mathrm{c}\right\| = \ \mathrm{cond}\left(\mathbf{A}\right) \ \frac{\left\|\mathbf{A}_\mathrm{E}\right\|}{\left\|\mathbf{A}\right\|} \ \left\|\mathbf{x}_\mathrm{c}\right\|$$

where

$$\mathrm{cond}\left(\mathbf{A}\right) \ \equiv \ k = \left\|\mathbf{A}\right\|\left\|\mathbf{A}^{-1}\right\|,$$

from which we have the following evaluation

$$\frac{\left\|\mathbf{x}_\mathrm{T} - \mathbf{x}_\mathrm{c}\right\|}{\left\|\mathbf{x}_\mathrm{c}\right\|} \le \mathrm{k}\frac{\left\|\mathbf{A}_\mathrm{E}\right\|}{\left\|\mathbf{A}\right\|}$$

Thus the computed solution can vary from the theoretical solution, as a result of round-off errors, by an amount directly proportional to the conditioning number k.

On computational precision

$$\boxed{q \approx p - \beta \log k}$$

p - introduced precision
q - obtained precision
β - norm dependent coefficient (for the spectral norm β=1)

Examples

(i)

$$\mathbf{A}_1 = \begin{bmatrix} 2 & 3 \\ 2 & 3.1 \end{bmatrix} \rightarrow k = \frac{\lambda_1}{\lambda_2} = \frac{5.060478}{0.039522} = 128.04$$

$$\log k = 2.107 \approx 2 \ \text{digits}$$

(ii)

$$\mathbf{A}_2 = \begin{bmatrix} 2 & 3 \\ 1.999 & 3 \end{bmatrix} \rightarrow k \approx \frac{4.9994}{0.0006} = 8331.33$$

$$\log k = 3.9207 \approx 4 \text{ digits}$$

How important are „small perturbations"? Compare given and inversed matrices.

$$\mathbf{A}_1 = \begin{bmatrix} 2 & 3 \\ 2 & 3.1 \end{bmatrix}, \qquad \mathbf{A}_1^{-1} = \begin{bmatrix} 15.5 & -15 \\ -10 & 10 \end{bmatrix}, \qquad \mathbf{A}_2 = \begin{bmatrix} 2 & 3 \\ 1.999 & 3 \end{bmatrix}, \qquad \mathbf{A}_2^{-1} = \begin{bmatrix} 1000 & 1000 \\ -666\frac{1}{3} & 666\frac{2}{3} \end{bmatrix}$$

AD 3. HOW TO DEAL WITH ILL CONDITIONED SYSTEMS

1. apply double precision
2. apply iterative residual approach

$$\mathbf{r}^{(1)} = \mathbf{A}\mathbf{x}^{(1)} - \mathbf{b} \qquad \text{- residuum}$$

Let

$$\boldsymbol{\varepsilon}^{(1)} = \mathbf{x}_T - \mathbf{x}^{(1)} \qquad \text{- solution error}$$

$$\underline{\mathbf{A}\boldsymbol{\varepsilon}^{(1)}} = \mathbf{A}\mathbf{x}_T - \mathbf{A}\mathbf{x}^{(1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(1)} = \underline{-\mathbf{r}^{(1)}} \quad \rightarrow \quad \varepsilon^{(1)}$$

The error vector $\boldsymbol{\varepsilon}^{(1)}$ is itself the solution of the given system but with a new right-hand side $-\mathbf{r}^{(1)}$. After solution we get from there

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \boldsymbol{\varepsilon}^{(1)} \rightarrow \mathbf{r}^{(2)} = \mathbf{A}\mathbf{x}^{(2)} - \mathbf{b} \qquad \text{etc.}$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\varepsilon}^{(k)} \rightarrow \mathbf{r}^{(k+1)} = \mathbf{A}\mathbf{x}^{(k+1)} - \mathbf{b}$$

until

$$\frac{\left\| \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \right\|}{\left\| \mathbf{x}^{(k+1)} \right\|} < B$$

3. change the method (e.g. the method of discretization)

4. use a regularization method – e.g. the Tikchonov method

Problems

Which of the following matrices gives rise to an ill-conditioned system; estimate loss of accuracy.

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \\ 1 & 1 & 0 \end{bmatrix} \quad \mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad , \qquad \mathbf{A}_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 8 \end{bmatrix} \quad \mathbf{b}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \mathbf{b}_3 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \quad , \qquad \mathbf{A}_4 = \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1/3 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad \mathbf{b}_4 = \begin{bmatrix} 1/2 \\ 1/4 \\ 3/4 \end{bmatrix}$$

7.3. TIKCHONOW SOLUTION APPROACH

Given

ILL-CONDITIONED SIMULTANEOUS ALGEBRAIC EQS

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad , \quad i, j = 1, 2, ..., n$$

FIND AN APPROXIMATE SOLUTION CLOSEST TO THE ORIGIN OF THE CARTESIAN COORDINATE SYSTEM $x_1, ..., x_n$

Find $$\min_{x_i} I \quad , \quad I = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2 + \alpha \sum_{i=1}^{n} x_i^2$$

THE RESULT DEPENDS ON α

Example

$$\begin{cases} 2x_1 + 3x_2 = 5 \\ 1.999x_1 + 3x_2 = 4.99 \end{cases}$$

$$I = \left(2x_1 + 3x_2 - 5\right)^2 + \left(1.999x_1 + 3x_2 - 4.99\right)^2 + \alpha\left(x_1^2 + x_2^2\right)$$

$$\frac{\partial}{\partial x_1} I = 0 \quad , \quad \frac{\partial}{\partial x_2} I = 0 \quad \rightarrow$$

$$x_1 = \frac{0.00009 + 19.975\alpha}{0.000009 + 25.996\alpha + \alpha^2}$$

$$x_2 = \frac{-0.000045 + 29.97\alpha}{0.000009 + 25.996\alpha + \alpha^2}$$

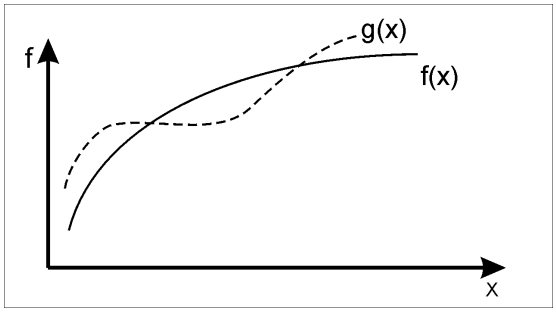
FAMILY OF SOLUTIONS

α	x_1	x_2	conditioning number $k = \left\| \frac{\lambda_{max}}{\lambda_{min}} \right\|$
0	10	-5	$7.51 \cdot 10^7$
0.1	0.7655	1.1484	260.96
1	0.7399	1.1102	27.00
10	0.5549	0.8326	3.60
100	0.1585	0.2379	1.26
10^6	$2 \cdot 10^{-5}$	$3 \cdot 10^{-5}$	1.00
"∞"	0	0	1.00

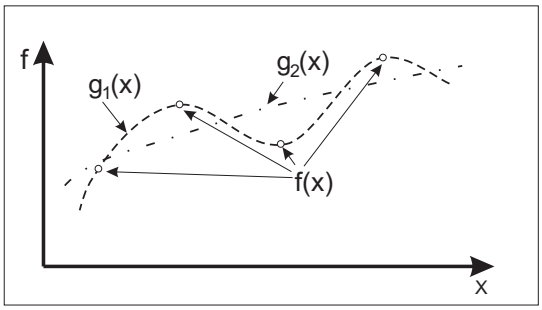
8. APPROXIMATION

8.1. INTRODUCTION

Approximation is a replacement of a function *f(x)*, given as a continuous or discrete one, by an other function *g(x)*.



f(x) continuos



f(x) discrete

Approximation is usually assumed in the form

$$f(\mathbf{x}) \approx g(\mathbf{x}) = \mathbf{a}^T \varphi \equiv P_n(\mathbf{x}) \rightarrow \varepsilon = f(\mathbf{x}) - P_n(\mathbf{x})$$

Where :

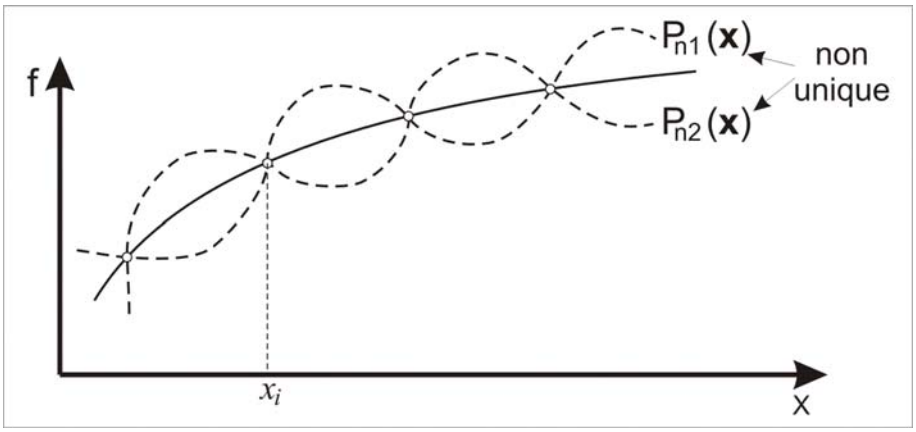
$\mathbf{a} = \{a_0 ... a_n\}$ - Unknown coefficients of approximation

$\varphi = \{\varphi_0(\mathbf{x}) ... \varphi_n(\mathbf{x})\}$ - Assumed bases functions

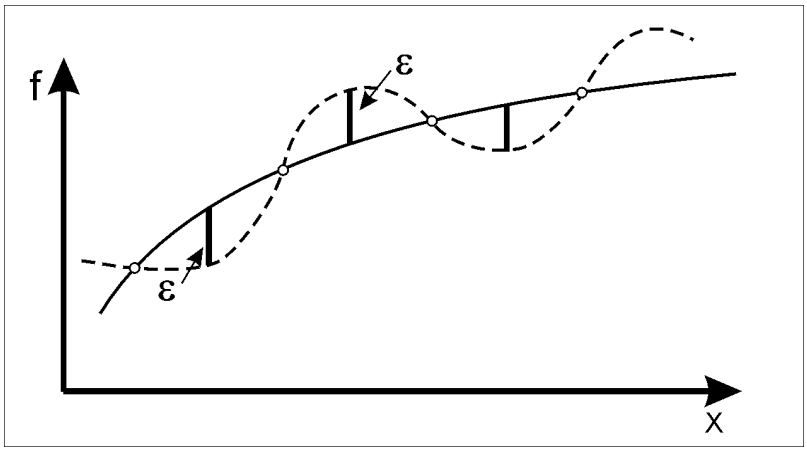
$\mathbf{x} = \{x^{(1)} ... x^{(m)}\}$ - Position vector of an arbitrary point in m-dimensional space

$\varepsilon(\mathbf{x}) = f(\mathbf{x}) - P_n(\mathbf{x})$ - approximation error

We distinguish : INTERPOLATION, BEST APPROXIMATION

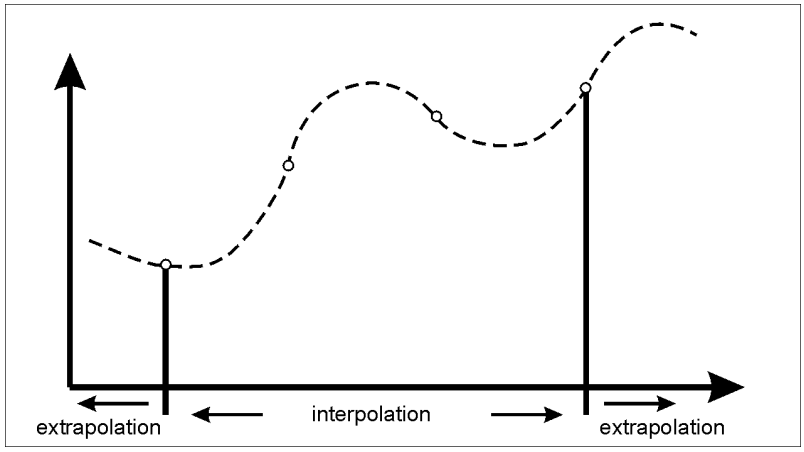


INTERPOLATION : $\varepsilon(\mathbf{x}_i) = 0$ *for* $i = 0,1,... ,n \rightarrow \mathbf{a}$



BEST APPROXIMATION : $\min_{\mathbf{a}} \|\varepsilon\| \rightarrow \mathbf{a}$

Interpolation – Extrapolation



Bases functions

Examples

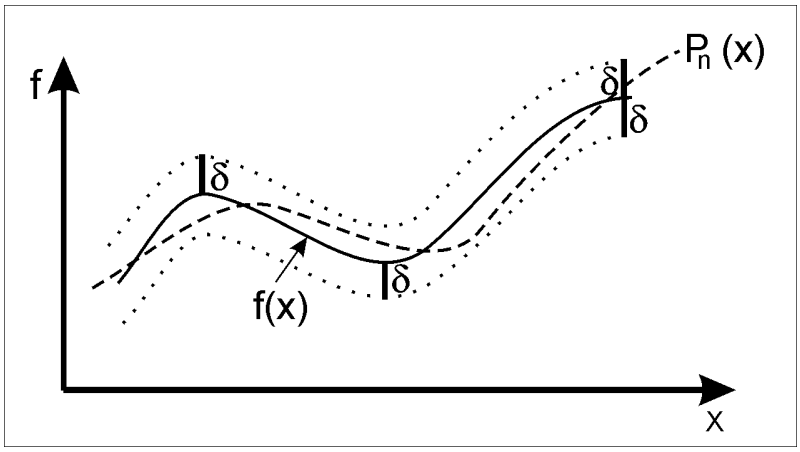
- Monomials

1, x, x^2, ..., x^n

Weierstrass Theorem

If continuous function *f(x)* is approximated by the polynomial $P_n(x)$ then for any given $\delta > 0$ such *n* may be found that

$$|f(x) - P_n(x)| < \delta$$



- Tschebychev polynomials

  $T_0(x), T_1(x), ..., T_n(x)$

- Trigonometric functions

  1, cos(x), sin(x), cos(2x), sin(2x), ..., cos(nx), sin(nx)

8.2. INTERPOLATION IN 1D SPACE

$$f(x_i) = P_n(x_i) = [\varphi_0(x_i)...\varphi_n(x_i)]\begin{Bmatrix} a_0 \\ ... \\ a_n \end{Bmatrix}, \qquad i = 0,1...n\ ;\ f(x_i) \equiv f_i$$

Let

$$\mathbf{F} = \{ f_0, ..., f_n \} \qquad \Phi = \begin{bmatrix} \varphi_0(x_0) & ... & \varphi_n(x_0) \\ ... & ... & ... \\ \varphi_0(x_n) & ... & \varphi_n(x_n) \end{bmatrix}$$

Required is solution of simultaneous linear algebraic equations

$$\boxed{\Phi \mathbf{a} = \mathbf{F} \rightarrow \mathbf{a} = \Phi^{-1}\mathbf{F}}$$

Example

For monomials x^n assumed as the bases functions we obtain Vandermonde determinant

$$\Phi = \begin{bmatrix} 1 & x_0 & ... & x_0^n \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ 1 & x_n & ... & x_n^n \end{bmatrix}$$

Here $\det \Phi \neq 0$ if $x_i \neq x_j$

1D Example

Let

$$f(x) \approx a_0 + a_1 x \qquad \text{-} \qquad \text{linear interpolation}$$

then

$$\begin{matrix} f(x_0) \equiv f_0 = a_0 + a_1 x_0 \\ f(x_1) \equiv f_1 = a_0 + a_1 x_1 \end{matrix} \Rightarrow \begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \rightarrow \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{x_1 - x_0} \begin{bmatrix} f_0 x_1 - f_1 x_0 \\ f_1 - f_0 \end{bmatrix}$$

hence

$$f(x) \approx P_1(x) = f_0 \frac{x - x_1}{x_0 - x_1} + f_1 \frac{x - x_0}{x_1 - x_0} = f_0 L_0^{(1)}(x) + f_1 L_1^{(1)}(x)$$

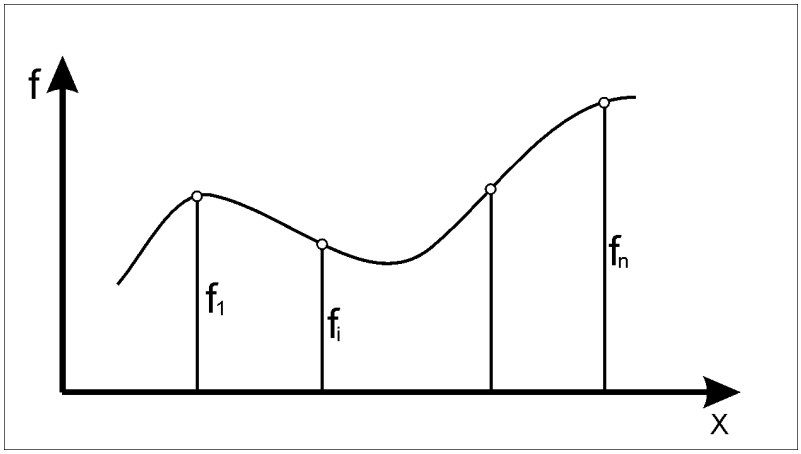
Generally

$$\mathbf{F} = \mathbf{\Phi}\,\mathbf{a} \rightarrow \mathbf{a} = \mathbf{\Phi}^{-1}\,\mathbf{F}$$

Solution of SLAE (simultaneous linear algebraic equations)

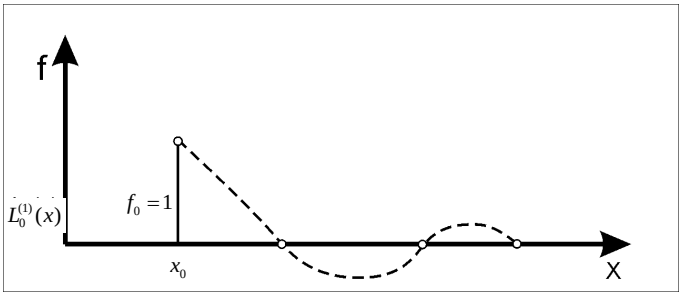
8.3. LAGRANGIAN INTERPOLATION (1D APPROXIMATION)

Given : f_i, for $i=0, 1, \ldots, n$



Required : An interpolating polynomial of degree n which passes through $n+1$ points (x_i, f_i)

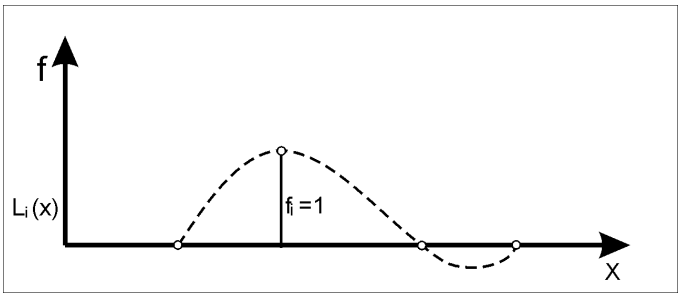
Construction of the solution



Nodal values

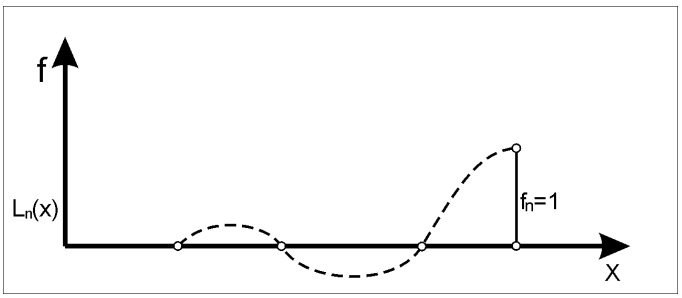
$$L_0^{(n)}(x_j) = \begin{cases} 1 & \text{if } x_j = x_0 \\ 0 & \text{if } x_j \neq x_0 \end{cases}$$

$$for \quad j = 0,1,\ldots,n$$



$$L_i^{(n)}(x_j) = \begin{cases} 1 & \text{if } x_j = x_i \\ 0 & \text{if } x_j \neq x_i \end{cases}$$

$$for \quad i, j = 0,1,...,n$$



$$L_n^{(n)}(x_j) = \begin{cases} 1 & \text{if } x_j = x_n \\ 0 & \text{if } x_j \neq x_n \end{cases}$$

$$for \quad , \quad j = 0,1,...,n$$

Generally

$$L_j^{(n)}(x_i) = \begin{cases} 1 & if \quad j = i \quad , \; j = 0,1,\ldots,n \\ 0 & if \quad j \neq i \quad , \; i = 0,1,\ldots,n \end{cases} \qquad (*)$$

The interpolating polynomial defined by :

$$P_n(x) = \sum_{j=0}^{n} L_j^{(n)}(x) f_j$$

satisfies equations

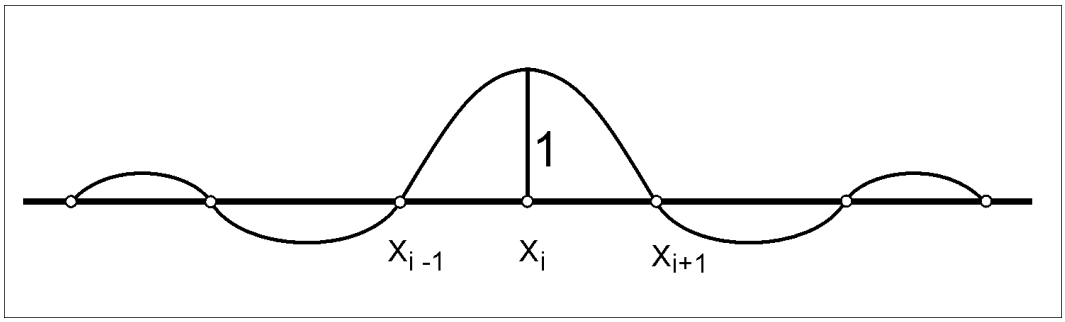
$$f(x_i) = P_n(x_i), \qquad i = 0,1,\ldots,n$$

How to find $L_j(x)$ *?*

Given

$$x_0,\ x_1,\ldots,\ x_{i-1},\ x_i,\ x_{i+1},\ldots,x_n$$

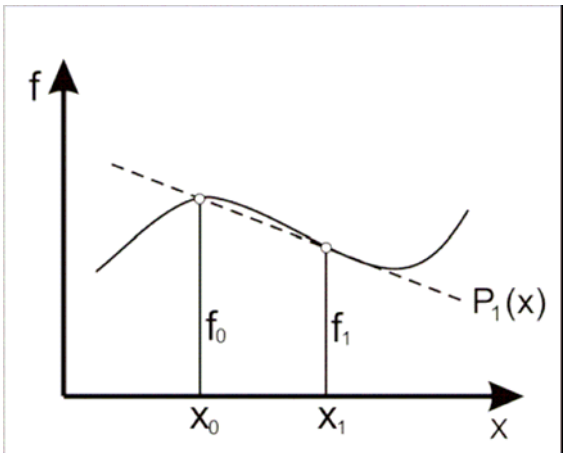
Required the lowest order polynomial $L_j^{(n)}(x)$ satisfying the conditions (*) posed above

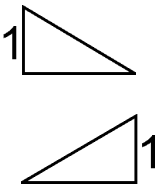


$$L_i^{(n)}(x) = \frac{(x - x_0)(x - x_1)\ldots(x - x_{i-1})(x - x_{i+1})\ldots(x - x_n)}{(x_i - x_0)(x_i - x_1)\ldots(x_i - x_{i-1})(x_i - x_{i+1})\ldots(x_i - x_n)} = \frac{\prod_{\substack{j=0 \\ j\neq i}}^{n}(x - x_j)}{\prod_{\substack{j=0 \\ j\neq i}}^{n}(x_i - x_j)}$$

Examples

n=1



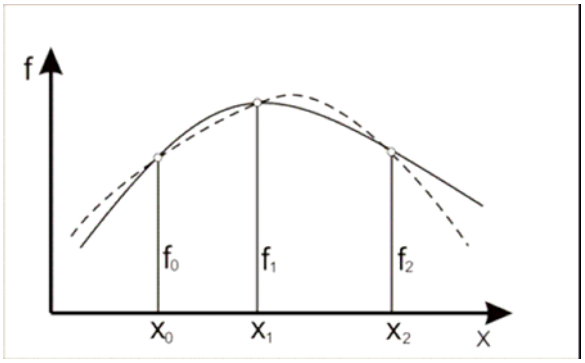


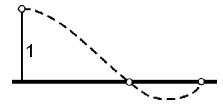
$$L_0^{(1)}(x) = \frac{x - x_1}{x_0 - x_1}$$

$$L_1^{(1)}(x) = \frac{x - x_0}{x_1 - x_0}$$

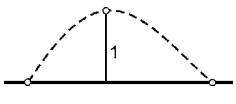
$$P_1(x) = \frac{x - x_1}{x_0 - x_1} f_0 + \frac{x - x_0}{x_1 - x_0} f_1$$

n=2





$$L_0^{(2)}(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}$$



$$L_1^{(2)}(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$$

$$L_2^{(2)}(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

$$P_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f_2$$

Example

Given the following data set

i	0	1	2	3
x_i	1	2	4	8
$f(x_i)$	1	3	7	11

Interpolate for $f(7)$ using the third order Lagrange polynomial. Repeat solution for the linear interpolation.

$$P_3(7) = 1 \cdot L_0^{(3)}(7) + 3 \cdot L_1^{(3)}(7) + 7 \cdot L_2^{(3)}(7) + 11 \cdot L_3^{(3)}(7)$$

$$L_0^{(3)}(7) = \frac{(7-2)(7-4)(7-8)}{(1-2)(1-4)(1-8)} = 0.71429$$

$$L_1^{(3)}(7) = \frac{(7-1)(7-4)(7-8)}{(2-1)(2-4)(2-8)} = -1.5$$

$$L_2^{(3)}(7) = \frac{(7-1)(7-2)(7-8)}{(4-1)(4-2)(4-8)} = 1.25$$

$$L_3^{(3)}(7) = \frac{(7-1)(7-2)(7-4)}{(8-1)(8-2)(8-4)} = 0.53571$$

$$f(7) \approx P_3(7) = 0.71429 + 3 \cdot (-1.5) + 7 \cdot (1.25) + 11 \cdot (0.53571) = 10.85710$$

Compare with the linear interpolation

$$P_1(7) = 7 \cdot L_0^{(1)}(7) + 11 \cdot L_1^{(1)}(7)$$

$$L_0^{(1)}(7) = \frac{7-8}{4-8} = 0.25$$

$$L_1^{(1)}(7) = \frac{7-4}{8-4} = 0.75$$

$$f(7) \approx P_1(7) = 7 \cdot (0.25) + 11 \cdot (0.75) = 10$$

THE ERROR TERM IN THE LAGRANGIAN INTERPOLATION

$$\varepsilon^{(n)}(x) \equiv f(x) - P_n(x)$$

Applying Rolle's theorem we may find that

$$\varepsilon^{(n)}(x) = \frac{f^{(n+1)}(\xi)\prod_{i=0}^{n}(x - x_i)}{(n+1)!} \leq \left| f_{max}^{(n+1)} \frac{\prod_{i=0}^{n}(x - x_i)}{(n+1)!} \right| \quad \text{for} \quad x_0 \leq \xi \leq x_n$$

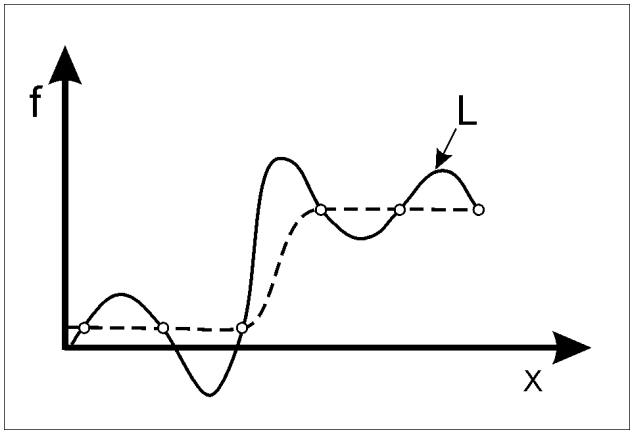
Example

Given $f(x) = ln(x), \quad x \in [1,2], \quad n = 3 \quad \rightarrow \quad f^{IV} = \frac{6}{x^4} \quad \rightarrow \quad f^{IV}_{max} = 6 \quad \rightarrow$

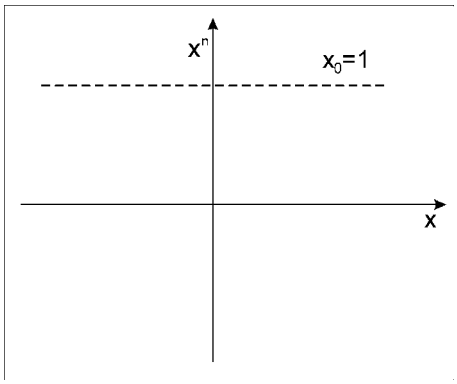
$$\varepsilon^{(3)}(x) \leq \frac{6}{4!}(x - x_0)(x - x_1)(x - x_2)$$

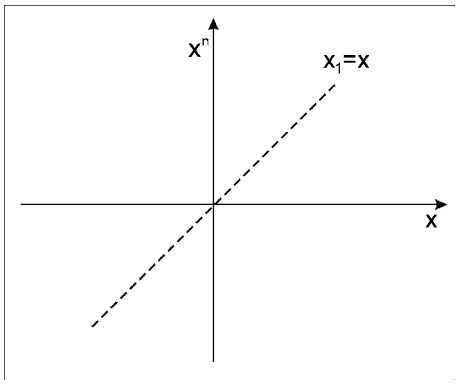
Remarks

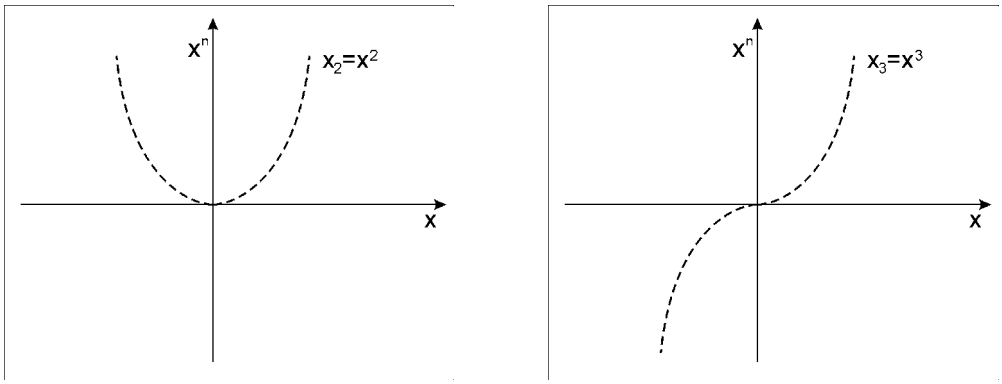
- The interpolating polynomials y of degree n reproduces exactly the polynomial function f of degree n or less
- Lagrangian interpolation of higher orders becomes numerically not stable, therefore it is not suggested to use $n > 3$



The reason is that monomials x^n are not orthogonal to each other and hardly can be distinguished for higher orders, e.g. x^{15} and x^{17}







Only monomials $1, x, x^2, x^3$ are clearly distinct from each other

Problems

Use the Lagrangian polynomial of degrees 1, 2, 3, 4 to evaluate approximately *f(1.25)* and *f(1.15)*. The function *f(x)* is given in a discrete form :

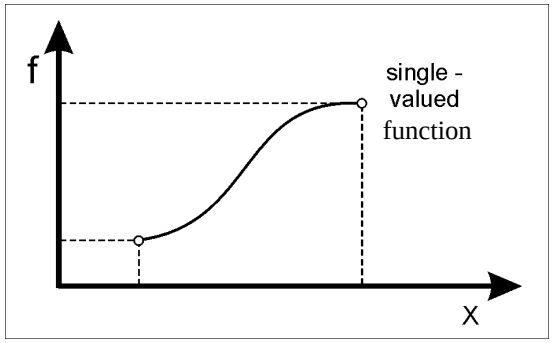
x	*1.0*	*1.1*	*1.2*	*1.3*	*1.4*
f(x)	*0.3679*	*0.3329*	*0.3012*	*0.2725*	*0.2466*

8.4. INVERSE LAGRANGIAN INTERPOLATION

Let

$$f(x)=\sum_{i=0}^{n} a_i\varphi_i(x) \;\rightarrow\; x(f)=\sum_{i=0}^{n} b_i\psi_i(f)$$

In many cases the contrasting question is asked : "find the value of the variable *x* at which the function *f (x)* takes on a particular value" (zero say) .



Sometimes, when *x* is a *single–valued* function of *f(x)* in the interval in question – it is convenient to use the Lagrangian interpolation formula

$$x(f)=\sum_{i=0}^{n} L_i(f)x_i \qquad L_i^{(n)}(f)=\prod_{\substack{j=0\\ j\neq i}} \frac{f-f_j}{f_i-f_j}, \qquad j=0,1,\ldots,n$$

Such type of interpolation is called the inverse interpolation.

Example

Find $x \in [1.0,\ 2.0]$ for which f =0.50. The following values are given :

X	*1.0*	*1.2*	*1.4*	*1.6*	*1.8*	*2.0*
f(x)	*0.000000*	*0.127524*	*0.302823*	*0.517000*	*0.764127*	*1.039725*

$$L_0^{(5)}(0.5) = 0.01122105 \qquad L_3^{(5)}(0.5) = 0.93966347$$
$$L_1^{(5)}(0.5) = -0.04725516 \qquad L_4^{(5)}(0.5) = -0.02204592$$
$$L_2^{(5)}(0.5) = 0.11677892 \qquad L_5^{(5)}(0.5) = 0.00163764$$

Hence

$$x = \sum_{i=0}^{5} x_i L_i^{(5)}(0.5) = 1.58505952$$

8.5. CHEBYCHEV POLYNOMIALS

In order to minimize the error term in the Lagrangian interpolation we may minimize the $\prod_{i=0}^{n}(x - x_i)$ term by an appropriate choice of the interpolation nodes x_i, $i = 0, 1, \ldots, n$. We may ask, therefore, to find

$$\min_{x_i}\left(\max_{-1 \le x \le 1} \left|(x - x_0)(x - x_1)\cdots(x - x_{n-1})\right| \right)$$

The solution is called Chebyshev polynomial of degree n, which is defined as

$$T_n(x) = \cos(n \arccos(x)) \quad \text{for} \quad -1 \le x \le 1$$

The Chebyshev polynomials satisfy a useful recurrence relation

$$T_{n+1}(x) = 2x \cdot T_n(x) - T_{n-1}(x), \qquad n = 1, 2, \ldots$$

where

$T_0(x) = \cos(0) = 1$

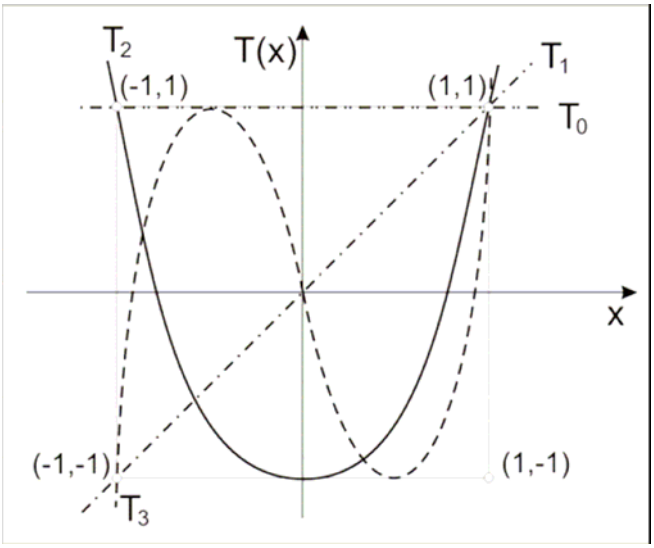
$T_1(x) = \cos(\vartheta) = x$

Chebyshev polynomial has exactly n zeros on $[-1,1]$ being located at

$$x_i = \cos\left(\frac{2i+1}{n}\frac{\pi}{2}\right), \qquad i = 0,1, \ldots, n-1$$

Its maximum value is $|T_n(x)| = 1$

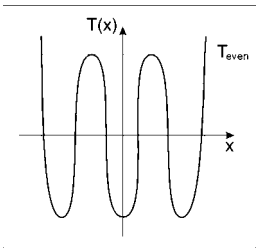
Examples of some Chebychev polynomials of the lowest order



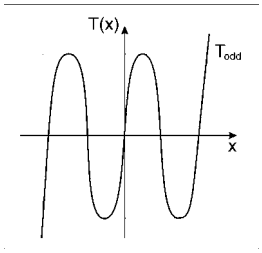
$T_0 = 1$ $T_1 = x$

$T_2 = 2x^2 - 1$ $T_3 = 4x^3 - 3x$

$T_4 = 8x^4 - 8x^2 + 1$ $T_5 = 16x^5 - 20x^3 + 5x$



behavior of even terms



behavior of odd terms

Transformation to the interval $[a,b]$

$$z = \frac{1}{2}\left[(b-a)x + (b+a)\right] \quad \rightarrow \quad x = \frac{2z-(b+a)}{b-a}$$

$$z \in [a,b] \qquad x \in [-1,1]$$

Orthogonality (weighted)

The Chebyshev polynomials form an orthogonal set over $[-1,1]$ with respect to the weight function $\frac{1}{\sqrt{1-x^2}}$, i.e.

$$I_{ij} = \int_{-1}^{1} \frac{T_i(x)T_j(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} 0 & i \neq j \\ \frac{\pi}{2} & i = j \neq 0 \\ \pi & i = j = 0 \end{cases}$$

Example

Find the Chebyshev polynomials in the interval $[1,4]$

$$x = \frac{2z-(4+1)}{4-1} = \frac{2}{3}z - \frac{5}{3}$$

$T_0 = 1$ $T_1 = \frac{2}{3}z - \frac{5}{3}$

$T_2 = 2\left(\frac{2}{3}z - \frac{5}{3}\right)^2 - 1$ $T_3 = 4\left(\frac{2}{3}z - \frac{5}{3}\right)^3 - 3\left(\frac{2}{3}z - \frac{5}{3}\right)$

$T_4 = 8\left(\frac{2}{3}z - \frac{5}{3}\right)^4 - 8\left(\frac{2}{3}z - \frac{5}{3}\right)^2 + 1$ $T_5 = 16\left(\frac{2}{3}z - \frac{5}{3}\right)^5 - 20\left(\frac{2}{3}z - \frac{5}{3}\right)^3 + 5\left(\frac{2}{3}z - \frac{5}{3}\right)$

Remark

Choice of interpolation nodes in zeros of the Chebyshev polynomials increases accuracy of interpolation

8.6. HERMITE INTERPOLATION

Given are :

- function values $f_i \qquad i=0, 1, \ldots, n$
- values of the derivative f_i'

Let us introduce an approximation polynomial

$$P_{2n+1}(x) = \sum_{j=0}^{n} f_j h_j(x) + \sum_{j=0}^{n} f_j' g_j(x)$$

where $h_j(x)$ and $g_j(x)$ are polynomials of degree $2n+1$ satisfying the conditions

$$h_j(x_i) = \delta_{ij}, \qquad i, j = 0,1,...,n \tag{1}$$

$$g_j(x_i) = 0, \qquad i, j = 0,1, \ldots,n \tag{2}$$

Differentiating $P_{2n+1}(x)$ we find

$$P_{2n+1}' = \sum_{j=0}^{n} f_j h_j'(x) + \sum_{j=0}^{n} f_j' g_j'(x) \approx f'(x)$$

If $P_{2n+1}'(x)$ is to interpolate the derivative values at x_i, $i=0, 1, \ldots, n,$ then

$$h_j'(x_i) = 0, \qquad i, j = 0,1, \ldots,m \tag{3}$$

$$g_j'(x_i) = \delta_{ij}, \qquad i, j = 0,1,...,n \tag{4}$$

Let us introduce a polynomial

$$l_j(x) = \prod_{\substack{i=0 \\ i \neq j}}^{n} (x - x_i)^2$$

When normalized it satisfies all conditions (1) and (3) except conditions

$$h_j(x_j) = 1 \quad \text{and} \quad h_j'(x_j) = 0 .$$

However, if we write

$$h_j(x) = \left[a(x - x_j) + b \right] \prod_{\substack{i=0 \\ i \neq j}}^{n} (x - x_i)^2 = \left[a(x - x_j) + b \right] l_j(x)$$

and require satisfaction of the conditions (1) we obtain

$$h_j(x_k)_{j \neq k} = \left[a(x_k - x_j) + b \right] \prod_{\substack{i=0 \\ i \neq j}}^{n} (x_k - x_i)^2 \equiv 0$$

as well as

$$h_j(x_j) = \left[\overbrace{a(x_j - x_j)}^{\equiv 0} + b \right] \prod_{\substack{i=0 \\ i \neq j}}^{n} (x_j - x_i)^2 = 1 \quad \rightarrow \quad b$$

hence we get

$$\boxed{b = \frac{1}{\prod_{\substack{i=0 \\ i\neq j}}^{n}\left(x_j - x_i\right)^2} = \frac{1}{l_j(x_j)}}$$

From (3) we have then

$$h_j'(x_k) = a\prod_{\substack{i=0 \\ i\neq j}}^{n}\left(x_k - x_i\right)^2 + \left[a\left(x_k - x_j\right) + b\right] l_j'(x_k) \equiv 0$$

and

$$h_j'(x_j) = a\prod_{\substack{i=0 \\ i\neq j}}^{n}\left(x_j - x_i\right)^2 + \left[\overbrace{a\left(x_j - x_j\right)}^{\equiv 0} + b\right] l_j'(x_j) = 0 \quad \rightarrow \quad a$$

we get

$$\boxed{a = -\frac{l_j'(x_j)}{l_j^2(x_j)}}$$

Hence

$$h_j(x) = \frac{l_j(x)}{l_j(x_j)}\left[1 - \left(x - x_j\right)\frac{l_j'(x_j)}{l_j(x_j)}\right]$$

Since the following holds

$$\frac{l_j(x)}{l_j(x_j)} \equiv L_j^{(n)2}(x), \qquad \frac{l_j'(x)}{l_j(x_j)} = 2\overbrace{L_j(x_j)}^{\equiv 1} L_j^{(n)'}(x_j) = 2L_j^{(n)'}(x_j)$$

we finally have

$$\boxed{h_j(x) = L_j^{(n)2}(x)\left[1 - 2\left(x - x_j\right)L_j^{(n)'}(x_j)\right]}$$

In a similar way we may derive

$$\boxed{g_j(x) = \left(x - x_j\right)L_j^{(n)2}(x)}$$

So that

$$\boxed{P_{2n+1}(x) = \sum_{j=0}^{n} f_j L_j^{(n)2}(x)\left[1 - 2\left(x - x_j\right)L_j^{(n)'}(x_j)\right] + \sum_{j=0}^{n} f_j'\left(x - x_j\right)L_j^{(n)2}(x)}$$

This is the Hermite interpolation formula.

In case when values f_j' are given in x_j ; $j=0, 1, \ldots, r < n$ the Hermite formula becomes

$$P_{n+r+1}(x) = \sum_{j=0}^{n} h_j(x)\, f_j + \sum_{j=0}^{r} g_j(x)\, f_j'$$

where

$$h_j(x)=\begin{cases}\left\{1-(x-x_j)\left[L_j^{(n)'}(x_j)+L_j^{(r)'}(x_j)\right]\right\}L_j^{(n)}(x)L_j^{(r)}(x), & j=0,1,\ldots,r\\ L_j^{(n)}(x)L_r^{(n)}(x)\dfrac{x-x_r}{x_j-x_r}, & j=r+1,\ldots,n\end{cases}$$

and

$$g_j(x)=(x-x_j)L_j^{(r)}(x)L_j^{(n)}(x), \qquad j=0,1,\ldots,r$$

The error term in the Hermite interpolating formula

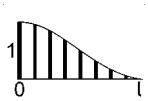
$$E(x)=\frac{\pi_n(x)\pi_r(x)}{(n+r+2)!}\,f^{(n+r+2)}(\xi)\ , \qquad \xi\in[x_0,x_n]$$

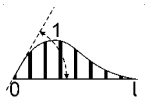
where

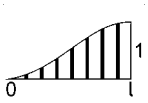
$$\pi_n(x)=\prod_{i=0}^{n}(x-x_i)$$

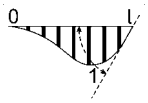
Example

Find Hermite interpolation satisfying the conditions

a) $f'(0)=f'(l)=0,\quad f(0)=1,\quad f(l)=0$ $\qquad f\equiv N_{00}$ 

b) $f'(0)=1,\quad f'(l)=0,\quad f(0)=f(l)=0$ $\qquad f\equiv N_{01}$ 

c) $f'(0)=f'(l)=0,\quad f(0)=0,\quad f(l)=1$ $\qquad f\equiv N_{10}$ 

d) $f'(0)=0,\quad f'(l)=1,\quad f(0)=f(l)=0$ $\qquad f\equiv N_{11}$ 

Ad a)

$n=1, \quad x_0=0, \quad x_1=l$
$f_0=1, \quad f_1=0, \quad f_0'=f_1'=0$

$$L_0^{(1)}(x)=\frac{l-x}{l-0}=\frac{l-x}{l},\qquad L_0^{(1)'}(x)=-\frac{1}{l},\qquad L_1^{(1)}(x)=\frac{x-0}{l-0}=\frac{x}{l},\qquad L_1^{(1)'}(x)=\frac{1}{l}$$

$$N_{00} = \sum_{j=0}^{1} f_j L_j^{(1)^2}(x)\left[1 - 2(x - x_j)L_j^{(1)'}(x_j)\right] + \sum_{j=0}^{1} f_j'(x - x_j)L_j^{(1)^2}(x) =$$

$$= 1 \cdot \left(\frac{l - x}{l}\right)^2 \left[1 - 2(x - 0)\left(-\frac{1}{l}\right)\right] + 0 \cdot (x - 0)\left(\frac{l - x}{l}\right)^2 + 0 \cdot \left(\frac{x}{l}\right)^2 \left[1 - 2(x - l)\frac{1}{l}\right] + 0 \cdot (x - l)\left(\frac{x}{l}\right)^2$$

$$N_{00} = \left(1 - \frac{x}{l}\right)^2 \cdot \left(1 + 2\frac{x}{l}\right)$$

Ad b)

$n=1$, $x_0=0$, $x_1=l$
$f_0 = f_1 = 0$, $f_0' = 1$, $f_1' = 0$

$$N_{01} = 0 \cdot \left(\frac{l - x}{l}\right)^2 \left[1 - 2(x - 0)\left(-\frac{1}{l}\right)\right] + 1 \cdot (x - 0)\left(\frac{l - x}{l}\right)^2 + 0 \cdot \left(\frac{x}{l}\right)^2 \left[1 - 2(x - l)\frac{1}{l}\right] + 0 \cdot (x - l)\left(\frac{x}{l}\right)^2$$

$$N_{01} = x\left(1 - \frac{x}{l}\right)^2$$

Ad c)

$n=1$, $x_0=0$, $x_1=l$
$f_0=0$, $f_1=1$, $f_0' = f_1' = 0$

$$N_{10} = 1 \cdot \left(\frac{x}{l}\right)^2 \left[1 - 2(x - l)\frac{1}{l}\right] = \left(\frac{x}{l}\right)^2 \left(3 - 2\frac{x}{l}\right)$$

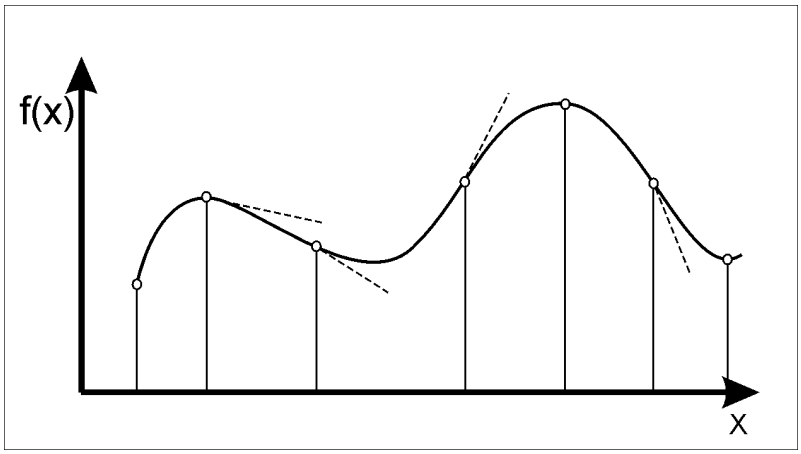
Ad d)

$n=1$, $x_0=0$, $x_1=l$
$f_0 = f_1 = 0$, $f_0' = 0$, $f_1' = 1$

$$N_{11} = (x - l)\left(\frac{x}{l}\right)^2$$

8.7. INTERPOLATION BY SPLINE FUNCTIONS

8.7.1. Introduction



8.7.2. Definition

A function which is a polynomial of degree k in each interval $[x_i, x_{i+1}]$ and which has continuous derivatives up to and including order $k-1$ is called a spline function of degree k.

Example

Given is

$$S(x) = \begin{cases} 1 - 2x & = S_1(x) \quad for \quad x \le -3 \\ 28 + 25x + 9x^2 + x^3 & = S_2(x) \quad for \quad -3 \le x \le -1 \\ 26 + 19x + 3x^2 - x^3 & = S_3(x) \quad for \quad -1 \le x \le 0 \end{cases}$$

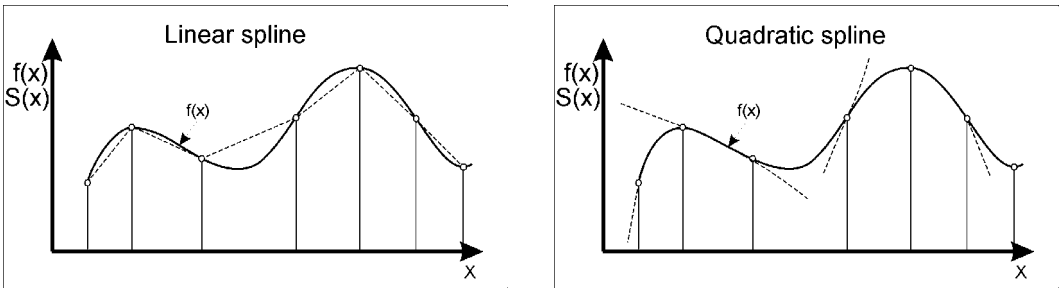
Verify, that $S(x)$ presents a cubic spline. Evaluate:

$$S_1'(x) = -2 \qquad S_1'' = 0$$
$$S_2'(x) = 25 + 18x + 3x^2 \qquad S_2'' = 18 + 6x$$
$$S_3'(x) = 19 + 6x - 3x^2 \qquad S_3'' = 6 - 6x$$

hence

$$\begin{cases} S_1(-3) = S_2(-3) = 7 \\ S_1'(-3) = S_2'(-3) = -2 \\ S_1''(-3) = S_2''(-3) = 0 \end{cases} \qquad \begin{cases} S_2(-1) = S_3(-1) = 11 \\ S_2'(-1) = S_3'(-1) = 10 \\ S_2''(-1) = S_3''(-1) = 12 \end{cases}$$

Examples



THE SPLINE $S(x)$ of degree k on the tubular points $x_0, x_1, ..., x_n$ is represented as :

$$S(x) = p_k(x) + \sum_{i=1}^{n-1} b_i \left(x - x_i\right)_+^k$$

where

$$p_k(x) = \sum_{i=0}^{k} a_i x^i$$

is a polynomial of degree k and $(x - x_i)_+^k$ is a truncated power function defined as :

$$(x - x_i)^k \equiv \begin{cases} (x - x_i)^k & \text{if } x - x_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

8.7.3. Extra conditions

In order to define a_i coefficients in the interval $\left[x_0, x_1\right]$ additional *k-1* conditions should be imposed. We call them "natural" conditions, if all of them are imposed at the point x_0.

Examples

$$S'(x_0) = 0 \qquad \text{- for quadratic spline}$$
$$S'(x_0) = 0,\ S''(x_0) = 0 \qquad \text{- for cubic spline}$$

Evaluation of a_i and b_i coefficients is then simple. Otherwise e.g. for conditions

$$S''(x_0) = S''(x_n) = 0$$

they cannot be obtained without solving full system of linear equations.

Example

Determine the quadratic spline on the tabular points $x_0, x_1, ..., x_n$ and such that $S'(x_0) = 0$.

$$S(x) = p_2(x) + \sum_{i=1}^{n-1} b_i (x - x_i)_+^2$$

$$p_2(x) = a_0 + a_1 x + a_2 x^2$$

For $x \in \left[x_0, x_1\right]$

$$S(x_0) = a_0 + a_1 x_0 + a_2 x_0^2 = f(x_0) \equiv f_0$$
$$S(x_1) = a_0 + a_1 x_1 + a_2 x_1^2 = \ \ldots \ \equiv f_1$$
$$S'(x_0) = \quad a_1 + 2a_2 x_0 = \ \ldots \ \equiv 0$$

Hence we have

$$a_0 = f_0 + a_2 x_0^2, \quad a_1 = -2\, x_0\, a_2, \quad a_2 = \frac{f_1 - f_0}{(x_1 - x_0)^2}$$

For $x \in \left[x_j, x_{j+1}\right]$

$$S(x_{j+1}) \equiv p_2(x_{j+1}) + \sum_{i=1}^{j} b_i (x_{j+1} - x_i)_+^2 = f_{j+1}$$

hence

$$\boxed{b_j = \frac{f_{j+1} - p_2(x_{j+1}) - \sum_{i=1}^{j-1} b_i (x_{j+1} - x_i)^2}{(x_{j+1} - x_j)^2}}$$

So the coefficients b_j are in this case defined explicitly.

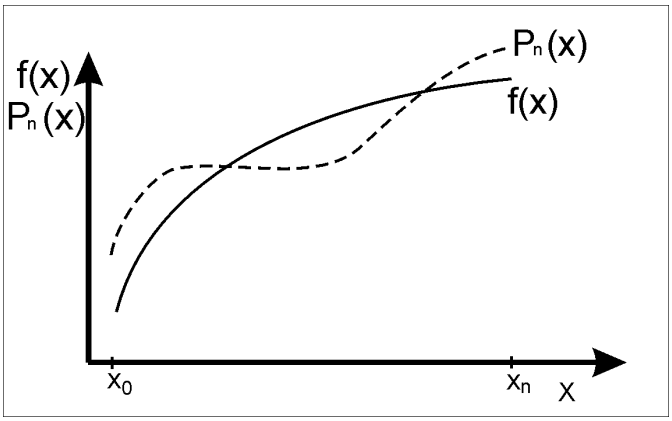
Homework

- Evaluate coefficients b_j for cubic spline, if $S'(x_0) = 0$, $S''(x_0) = 0$.
- Find the spline given in the first example; assume $S(x_0)$, $S(x_1)$, $S(x_2)$, $S'(x_0)$, $S''(x_0)$.

8.8. THE BEST APPROXIMATION

Introduction

$$f(\mathbf{x}) \approx P_n(\mathbf{x}) = \mathbf{a}^T \varphi$$
$$\varepsilon \equiv f - P_n \qquad \text{in } \left[x_0, x_n\right]$$



Required

$$\min \|\varepsilon\| = \min_{\mathbf{a}} \|f - P_n\|$$

Approximation is the best with the respect to a chosen norm. Mostly are used :

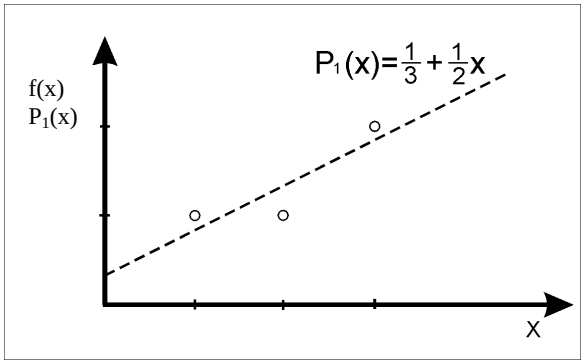
(i) $\|\varepsilon\|_\infty = \max|\varepsilon|$ - MINIMAX approximation (Chebyshev) : $\min_{a} \max_{x_0 \le x \le x_n} |f - P_n|$

$$\text{(ii)}\quad \|\varepsilon\|_2 = \begin{cases} \left(\int_{x_0}^{x_n} \varepsilon^2 dx \right)^{\frac{1}{2}} & - \quad \text{for } f \text{ continuous} \\ \left(\sum_{i=1}^{n} \varepsilon_i^2 \right)^{\frac{1}{2}} & - \quad \text{for } f \text{ discrete} \end{cases} \quad \textit{EUCLIDEAN}$$

Example

Given

n	*0*	*1*	*2*
x	*1*	*2*	*3*
f(x)	*1*	*1*	*2*



Find the best linear approximation using the norm $\|\varepsilon\|_2$

$$P_1(x) = a_0 + a_1 x$$

$$I \equiv \|\varepsilon\|_2^2 = \sum_{i=0}^{2} \left(f_i - p_n(x_i) \right)^2 = (1 - a_0 - a_1)^2 + (1 - a_0 - 2a_1)^2 + (2 - a_0 - 3a_1)^2$$

$$\frac{\partial I}{\partial a_0} = 2\left[-1(1 - a_0 - a_1) - (1 - a_0 - 2a_1) - a_0(2 - a_0 - 3a_1) \right] = 0 \quad \rightarrow \quad 2a_0 + 6a_1 = 4$$

$$\frac{\partial I}{\partial a_1} = 2\left[-1(1 - a_0 - a_1) - 2(1 - a_0 - 2a_1) - 3(2 - a_0 - 3a_1) \right] = 0 \quad \rightarrow \quad 6a_0 + 14a_1 = 9$$

Hence solution is

$$P_1(x) = \frac{1}{3} + \frac{1}{2} x$$

This is called the *Least Squares Approach*

8.9. LEAST SQUARES APPROACH

Let

$$I = \int_{x_0}^{x_n} \left[f - \sum_{i=1}^{n} a_i \varphi_i(x) \right]^2 dx$$

$$\frac{\partial I}{\partial a_k} = -2\int_{x_0}^{x_n} \varphi_k(x)\left[f - \sum_{i=1}^{n} a_i \varphi_i(x) \right] dx = 0, \qquad k = 0,1, \ldots, n$$

Hence

$$\int_{x_0}^{x_n} f(x)\varphi_k(x)dx - \sum_{i=1}^{n} a_i \int_{x_0}^{x_n} \varphi_i(x)\,\varphi_k(x)\,dx = 0$$

Let

$$\Phi_{ik} \equiv \int_{x_0}^{x_n} \varphi_i(x)\varphi_k(x)\,dx, \qquad\qquad F_k = \int_{x_0}^{x_n} f(x)\varphi_k(x)\,dx$$

Then

$$\sum a_i \Phi_{ik} = F_k \qquad \Leftrightarrow \qquad \Phi\mathbf{a} = \mathbf{F} \qquad \rightarrow \qquad \mathbf{a} = \Phi^{-1}\mathbf{F}$$

$$\underset{(n+1)\times 1}{\mathbf{a}} = \{ a_0,\ a_1, \ldots\ , a_n \}, \qquad\qquad \underset{(n+1)\times(n+1)}{\Phi} = \left[\Phi_{ik}\right]$$

$$\underset{(n+1)\times 1}{\mathbf{F}} = \{ F_0,\ F_1, \ldots\ , F_n \}$$

If bases functions are orthogonal matrix Φ is diagonal : $\Phi = \text{diag}\,\Phi_{ii}$ and coefficients **a** are explicitly determined.

8.10. INNER PRODUCT

Given are functions $f(x),\ g(x),\ x \in [a,b]$. Let us define the INNER PRODUCT :

$$(f,g) = \begin{cases} \int_a^b f(x)g(x)\,dx & - \text{ if } f \text{ and } g \text{ are continuous} \\ \sum_{i=0}^{n} f(x_i)g(x_i) & - \text{ if } f \text{ and } g \text{ are discrete} \end{cases}$$

Examples

a) given: $f(x) = x, \quad g(x) = 2x^2 + 1, \qquad [a,b] = [0,1]$

$$(f,g) = \int_0^1 x\left(2x^2 + 1\right)dx = \frac{1}{2}\left(x^4 + x^2\right)\Big|_0^1 = 1$$

b) given:

x	*0.0*	*0.5*	*0.8*	*1.0*
f(x)	*0.0*	*0.5*	*0.8*	*1.0*
g(x)	*1.0*	*1.5*	*2.28*	*3.0*

$$(f,g)=\sum_{i=0}^{3} f(x_i)g(x_i)=0.0\cdot 1.0+0.5\cdot 1.5+0.8\cdot 2.28+1.0\cdot 3.0=5.574$$

8.11. GENERATION OF ORTHOGONAL FUNCTIONS BY THE GRAM - SCHMIDT PROCESS

Consider a linearly independent set of functions $\varphi_j(x)$, $j = 0,1,\ldots,m$ in an interval $[a,b]$. We want to construct functions $q_j(x)$ so that

$$(q_j,q_k)=\left\{\begin{array}{l}\int_a^b q_j(x)q_k(x)\,dx\\ \sum_{i=0}^{n} q_j(x_i)q_k(x_i)\end{array}\right\}=\left\{\begin{array}{ll}0 & \text{if } j\neq k\\ \neq 0 & \text{if } j=k\end{array}\right. \qquad j,k = 0, 1, \ldots, m$$

Let

$$q_0(x)=\varphi_0(x)$$
$$q_1(x)=\varphi_1(x)-\alpha_{01}\,q_0(x) \qquad \text{but} \qquad (q_1,q_0)=0$$

hence

$$\underbrace{(q_1,q_0)}_{=0}=(\varphi_1,q_0)-\alpha_{01}(q_0,q_0) \quad\rightarrow\quad \alpha_{01}=\frac{(\varphi_1,q_0)}{(q_0,q_0)}$$
$$q_2(x)=\varphi_2(x)-\alpha_{02}\,q_0(x)-\alpha_{12}\,q_1(x)$$

hence

$$\underbrace{(q_2,q_0)}_{=0}=(\varphi_2,q_0)-\alpha_{02}(q_0,q_0)-\underbrace{\alpha_{12}(q_1,q_0)}_{=0} \quad\rightarrow\quad \alpha_{02}=\frac{(\varphi_2,q_0)}{(q_0,q_0)}$$

and

$$\underbrace{(q_2,q_1)}_{=0}=(\varphi_2,q_1)-\underbrace{\alpha_{02}(q_0,q_1)}_{=0}-\alpha_{12}(q_1,q_1) \quad\rightarrow\quad \alpha_{12}=\frac{(\varphi_2,q_1)}{(q_1,q_1)}$$

Generally

let

$$\boxed{q_p(x)=\varphi_p(x)-\sum_{i=0}^{p-1}\alpha_{ip}\,q_i(x)}$$

Since we require that

$$\left(q_j, q_k\right) = 0 \qquad \textit{if} \qquad j \neq k$$

we get

$$\left(q_p, q_j\right) = \left(\varphi_p, q_j\right) - \sum_{i=0}^{p-1} \alpha_{ip} \left(q_i, q_j\right) = \left(\varphi_p, q_j\right) - \alpha_{jp} \left(q_j, q_j\right)$$

hence

$$\boxed{\alpha_{jp} = \frac{\left(\varphi_p, q_j\right)}{\left(q_j, q_j\right)}} \qquad \begin{array}{l} p = 1,2,\ldots,n \\ j = 0,1,\ldots,p-1 \end{array}$$

Example

Let $\varphi_i = x^i$, $\quad [a,b] = [0,2]$

Then

$$\left(\varphi_i, \varphi_j\right) = \int_0^2 \varphi_i(x)\,\varphi_j(x)\,dx$$

$$q_0 = 1$$

$$q_1 = x - \alpha_{01} \cdot 1$$

$$\alpha_{01} = \frac{\int_0^2 x \cdot 1\,dx}{\int_0^2 1 \cdot 1\,dx} = \frac{\left.\frac{x^2}{2}\right|_0^2}{\left.x\right|_0^2} = \frac{2}{2} = 1$$

$$q_1 = x - 1$$

$$q_2 = x^2 - \alpha_{02} \cdot 1 - \alpha_{12} \cdot (x-1)$$

$$\alpha_{02} = \frac{\int_0^2 x^2 \cdot 1\,dx}{\int_0^2 1 \cdot 1\,dx} = \frac{\left.\frac{x^3}{3}\right|_0^2}{2} = \frac{4}{3}$$

$$\alpha_{12} = \frac{\int_0^2 x^2 \cdot (x-1)\,dx}{\int_0^2 (x-1) \cdot (x-1)\,dx} = \frac{\left.\left(-\frac{x^3}{3} + \frac{x^4}{4}\right)\right|_0^2}{\left.\left(x - 2\frac{x^2}{2} + \frac{x^3}{3}\right)\right|_0^2} = \frac{\frac{4}{3}}{\frac{2}{3}} = 2$$

$$q_2 = x^2 - \frac{4}{3} \cdot 1 - (+2) \cdot (x-1) = x^2 - 2x + \frac{2}{3}$$

$$q_3 = x^3 - a_{03}\cdot 1 - \alpha_{13}\cdot(x-1) - \alpha_{23}\cdot\left(x^2 - 2x + \frac{2}{3}\right)$$

$$a_{03} = \frac{\int_0^2 x^3\cdot 1\,dx}{\int_0^2 1\cdot 1\,dx} = \frac{\left.\frac{x^4}{4}\right|_0^2}{2} = 2$$

$$\alpha_{13} = \frac{\int_0^2 x^3\cdot(x-1)\,dx}{\int_0^2 (x-1)\cdot(x-1)\,dx} = \frac{\left.\left(-\frac{x^4}{4} + \frac{x^5}{5}\right)\right|_0^2}{\left.\left(x - 2\frac{x^2}{2} + \frac{x^3}{3}\right)\right|_0^2} = \frac{\frac{12}{5}}{\frac{2}{3}} = \frac{18}{5}$$

$$\alpha_{23} = \frac{\int_0^2 x^3\cdot\left(x^2 - 2x + \frac{2}{3}\right)dx}{\int_0^2 \left(x^2 - 2x + \frac{2}{3}\right)^2 dx} = \frac{\left.\left(\frac{x^6}{6} - \frac{2}{5}x^5 + \frac{1}{6}x^4\right)\right|_0^2}{\left.\left(\frac{1}{5}x^5 - x^4 + \frac{16}{9}x^3 - \frac{4}{3}x^2 + \frac{4}{9}x\right)\right|_0^2} = \frac{\frac{8}{15}}{\frac{8}{45}} = 3$$

$$q_3 = x^3 - 2 + \frac{18}{5}(x-1) - 3\left(x^2 - 2x + \frac{2}{3}\right) = x^3 - 3x_2 + \frac{12}{5}x - \frac{2}{5}$$

8.11.1. Orthonormalization

$$\left(\bar{q}_i,\ \bar{q}_j\right) = \delta_{ij} \qquad \rightarrow \qquad \bar{q}_i = \frac{q_i}{\left(q_i, q_i\right)^{1/2}}$$

normalized function

q_i non-normalized function

8.11.2. Weighted orthogonalization

$$\left(q_i, wq_i\right) = \left(wq_i, q_i\right) = \begin{cases} 0 & for \quad i \neq j \\ c_j \neq 0 & for \quad i = j \end{cases}$$

8.11.3. Weighted orthonormalization

$$\left(\bar{q}_i, w\bar{q}_j\right) = \delta_{ij}$$

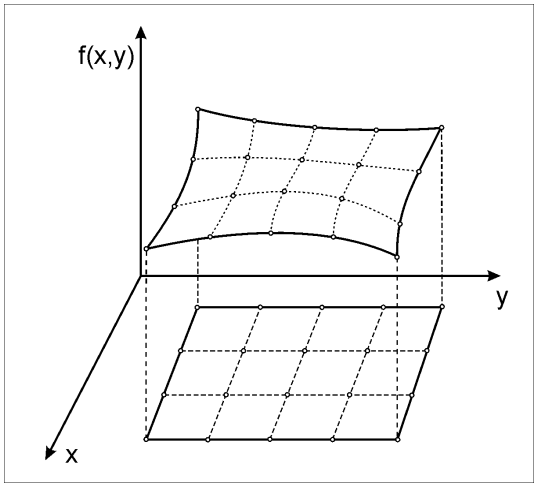
Homework

- Solve the above example for orthonormalized functions q_0, q_1, q_2 calculated as defined above
- Assume $w(x)=x$ and solve the above example again
  (i) for orthogonalization
  (ii) for orthonormalization

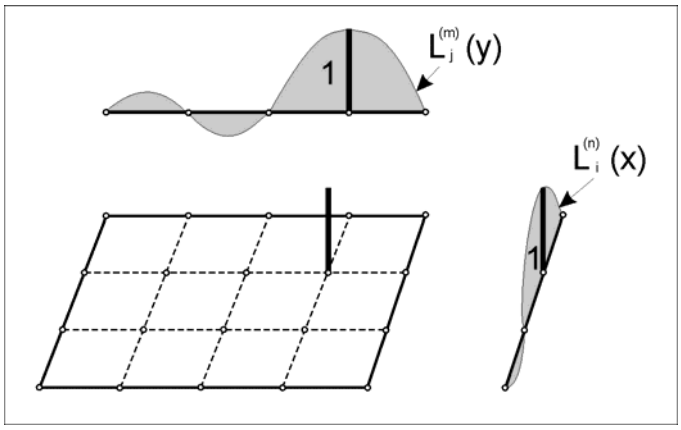
8.12. APPROXIMATION IN A 2D DOMAIN

8.12.1. Lagrangian approximation over rectangular domain

Let a function $f(x,y)$ be given over a rectangular domain.



The basic concept of the Lagrangian approximation holds in $2D$ space



Let

$$L_{ij}^{(n,m)}(x,y)=\begin{cases}1 & \text{if } x=x_j \text{ and } y=y_j\\ 0 & \text{otherwise}\end{cases}$$

$$L_{ij}^{(n,m)}(x,y)=L_i^{(n)}(x)\cdot L_j^{(m)}(y)$$

$$\boxed{f(x,y)=\sum_{i=0}^{n}\sum_{j=0}^{m}a_{ij}L_i^{(n)}(x)\cdot L_j^{(m)}(y)}$$

Example

$$n=3, \quad m=4, \quad i=1, \quad j=3$$

$$L_{13}^{(3,4)}(x,y) = L_1^{(3)}(x) \cdot L_3^{(4)}(y)$$

as shown in Figure above

9. NUMERICAL DIFFERENTATION

Problem

Given is a discrete function $f(x_i)$, $i=0, 1, \ldots, n$. Find the derivative of this function at a point $x=x_j$.

Solution

9.1. BY MEANS OF THE APPROXIMATION AND DIFFERENTATION

Find an approximation of this function

$$f(x) \approx P_n(x) = \mathbf{a}^T \varphi(x)$$

and perform its differentiation. Then

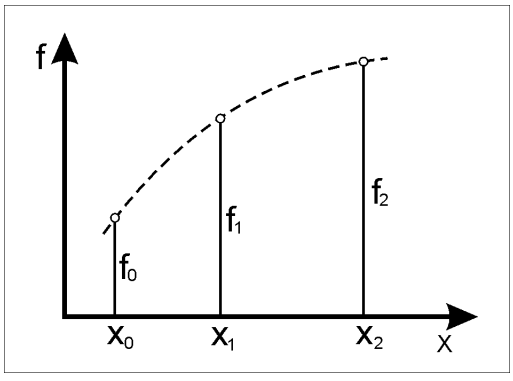
$$f'(x) \approx \frac{dP_n(x)}{dx} = \mathbf{a}^T \varphi'(x)$$

Example

Use the Lagrangian approximation

$$f(x) \approx \sum_{i=0}^{n} a_i L_i^{(n)}(x) \quad \rightarrow \quad f'(x) \approx \sum_{i=0}^{n} a_i L_i'^{(n)}(x)$$

e.g. in case of $\boxed{n=2}$



$$f(x) = f_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + f_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + f_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$

$$f'(x) = f_0 \frac{2x-x_1-x_2}{(x_0-x_1)(x_0-x_2)} + f_1 \frac{2x-x_0-x_2}{(x_1-x_0)(x_1-x_2)} + f_2 \frac{2x-x_0-x_1}{(x_2-x_0)(x_2-x_1)}$$

In case of $x_2 - x_1 = x_1 - x_0 = h$

$$f'(x) = \frac{f_0}{2h^2}(2x - x_1 - x_2) - \frac{f_1}{2h^2}(2x - x_0 - x_2) + \frac{f_2}{2h^2}(2x - x_0 - x_1)$$

hence finite difference formulas

$$f'(x_0) = -\frac{3}{2h} f_0 + \frac{2}{h} f_1 - \frac{1}{2h} f_2$$

$$f'(x_1) = \frac{f_2 - f_0}{2h} \qquad \text{- central difference}$$

$$f'(x_2) = \frac{1}{2h} f_0 - \frac{2}{h} f_1 + \frac{3}{2h} f_2$$

Let the following will be given

x	0.0	½	1.0
$f(x)$	1.0	13/16	0.0

with the true function being $f(x) = x^4 - 2x^3 + 1 \rightarrow f'(x) = 4x^3 - 6x^2$.

Using the above derived formula we get

x	0.0	⅓	½	1.0
f'_{exact}	0.0	−14/27	-1.0	-2.0
f'_{approx}	¼	−7/12	-1.0	−9/4

$$f'(x) \approx \frac{1}{2\cdot\left(\frac{1}{2}\right)^2}\left(2x - \frac{1}{2} - 1\right) - \frac{\frac{13}{16}}{\left(\frac{1}{2}\right)^2}\left(2x - 0 - \frac{1}{2}\right) + \frac{0}{2\cdot\left(\frac{1}{2}\right)^2}\left(2x - 0 - \frac{1}{2}\right) = -\frac{5}{2}x + \frac{1}{4}$$

and the following values for the first derivative

9.2. GENERATION OF NUMERICAL DERIVATIVES BY UNDETERMINED COEFFICIENTS METHOD

As may be seen above a numerical derivative of a function $f(x)$ is a linear combination of values of this function in some chosen points, e.g.

$$f'(x_i) = \sum_j \alpha_{j(i)} f_j$$

where $\alpha_{j(i)}$ are coefficients to be determined. We may find them :

(i) expanding the function in the Taylor series about $x = x_i$ end equalizing both sides of the above equation, or

(ii) imposing requirement that formula should be exact for monomials x^k, $k = 0, 1, \ldots, n$ up to the highest order n

Example

Required

$$f'(x_i) = \alpha_{i-1} f_{i-1} + \alpha_i f_i + \alpha_{i+1} f_{i+1}$$

Find

$$\alpha_{i-1},\ \alpha_i,\ \alpha_{i+1} \quad \text{if} \quad x_{i-1} = x_i - h \quad \text{and} \quad x_{i+1} = x_i + h\,.$$

Let

$$\begin{aligned}
f_{i-1} = f(x_i - h) &= f_i - h f_i' + \frac{1}{2}h^2 f_i'' - \frac{1}{6}h^3 f_i''' + \ldots && \big|\alpha_{i-1} \\
f_i &= f_i && \big|\alpha_i \\
f_{i+1} = f(x_i + h) &= f_i + h f_i' + \frac{1}{2}h^2 f_i'' + \frac{1}{6}h^3 f_i''' + \ldots && \big|\alpha_{i+1}
\end{aligned}$$

$$\begin{aligned}
0 f_i + 1 f_i' + 0 f_i'' + \ldots = {} & f_i \overbrace{(\alpha_{i-1} + \alpha_i + \alpha_{i+1})}^{=0} + f_i' \overbrace{(-h\alpha_{i-1} + h\alpha_{i+1})}^{=1} + \\
& + f_i'' \frac{1}{2}h^2 \overbrace{(\alpha_{i-1} + \alpha_{i+1})}^{=0} + f_i''' \frac{1}{6}h^3 \overbrace{(-\alpha_{i-1} + \alpha_{i+1})}^{=R} + \cdots
\end{aligned}$$

hence

$$\begin{aligned}
&\alpha_{i-1} + \alpha_i + \alpha_{i+1} = 0 \\
&-h\cdot\alpha_{i-1} + h\cdot\alpha_{i+1} = 1 \\
&\alpha_{i-1} + \alpha_{i+1} = 0
\end{aligned} \quad \rightarrow \quad \begin{cases} \alpha_{i-1} = -\dfrac{1}{2h} \\ \alpha_i = 0 \\ \alpha_{i+1} = \dfrac{1}{2h} \end{cases}$$

Finally

$$f'(x_i) = \frac{f_{i+1} - f_{i-1}}{2h} + \frac{h^2}{6} f_i''' + \cdots \equiv \frac{f_{i+1} - f_{i-1}}{2h} + O\left(h^2\right)$$

This is the central finite difference formula of the second order of accuracy

Homework

- Derive the formula $f'(x_i) = \alpha_{i-3} f_{i-3} + \alpha_{i-2} f_{i-2} + \alpha_{i-1} f_{i-1} + \alpha_i f_i$

- Derive the same formula using Lagrangian approximation. Find the derivative *f'(2.5)* if

x	*0.0*	*1.0*	*2.0*	*3.0*
f(x)	*0.0*	*1.0*	*8.0*	*27.0*

Derive the same formulas by using approach *(ii)*, i.e. substituting for *f(x)* subsequently x^j, $j = 0, 1, \ldots, k$ in order to obtain simultaneous equations for α_t, $i = 0,1,\ldots,k$

$$jx^{j-1} = \sum_{i=0}^{k} \alpha_t x_i^j \quad \rightarrow \quad \alpha_t = \ldots$$

10. NUMERICAL INTEGRATION

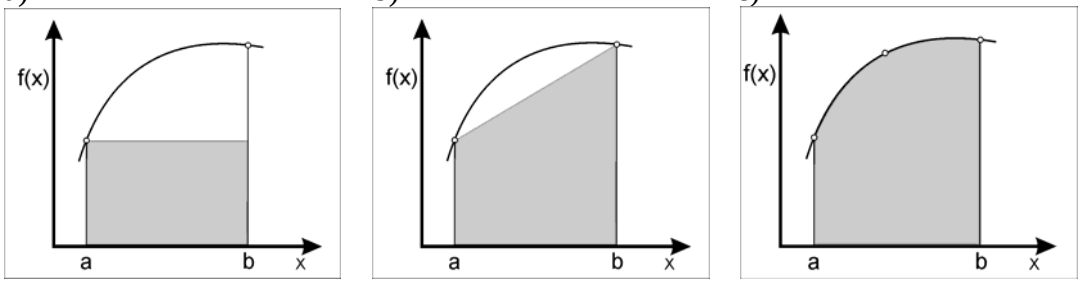
10.1. INTRODUCTION

Like in a case of numerical differentiation an approximation may be used in order to replace the integrated function *f(x)*. Thus

$$I_{exact} = \int_a^b f(x)\,dx \approx \int_a^b P_n(x)\,dx = \int_a^b \mathbf{a}^T \varphi(x)\,dx \equiv I_{approx}$$

Some simplest cases are shown below

a) b) c)



a) approximation by the zero order polynomial

$$\int_a^b f(x)\,dx \approx \int_a^b f_0 dx = f_0 \int_a^b dx = hf_0$$ - rectangular rule

b) approximation by the first order polynomial

$$\int_a^b f(x)\,dx \approx \int_a^b \left(f_0 \frac{1-x}{h} + f_1 \frac{x}{h} \right) dx = \frac{h}{2}\left(f_0 + f_1 \right)$$ - trapezoidal rule

c) approximation by the second order polynomial

$$\int_a^b f(x)\,dx \approx \int_a^b \left[f_0 L_0^{(2)}(x) + f_1 L_1^{(2)}(x) + f_2 L_2^{(2)}(x) \right] dx = \frac{h}{3}\left(f_0 + 4 f_1 + f_2 \right)$$ - Simpson rule

where

$$h = b - a$$

Example

Evaluate

$$I = \int_0^{\frac{1}{2}} \sqrt{1+x}\,dx = \tfrac{2}{3}\left(1+x\right)^{\frac{3}{2}}\Big|_0^{\frac{1}{2}} = \tfrac{2}{3}\left[\left(\tfrac{2}{3}\right)^{\frac{3}{2}} - 1 \right] = 0.5580782$$

Rectangular rule

$$I \approx \left(1+0\right)^{\frac{1}{2}} \cdot \left(\tfrac{1}{2} - 0\right) = 0.5$$

Trapezoidal rule

$$I \approx \tfrac{1}{2}\left(\tfrac{1}{2}-0\right)\left[\left(1+0\right)^{\frac{1}{2}}+\left(1+\tfrac{1}{2}\right)^{\frac{1}{2}}\right]=0.5561862$$

Simpson rule

$$I \approx \tfrac{1}{3}\left(\tfrac{1}{4}-0\right)\left[\left(1+0\right)^{\frac{1}{2}}+4\left(1+\tfrac{1}{4}\right)^{\frac{1}{2}}+\left(1+\tfrac{1}{2}\right)^{\frac{1}{2}}\right]=0.5580734$$

General approach

If $x_{i+1}-x_i=x_i-x_{i-1}=h=\text{const}$ → Newton - Cotes Formulas
Otherwise we consider Gauss Formulas

10.2. NEWTON – COTES FORMULAS

Let the equally spaced tabular points be denoted by x_i, $i=0, 1, \ldots, n,$ with $a=x_0+ph$, $b=x_0+qh$, $p\geq 0$, $q\leq n$. Using an interpolating formula based on these $n+1$ points we have:

$$I=\int_a^b f(x)\,dx \approx \int_a^b \sum_{j=0}^{n} L_j(x)\,f_j\,dx \equiv I_{n+1}$$

$$I_{n+1}=\sum_{j=0}^{n}\int_a^b L_j(x)\,f_j\,dx$$

Introducing an independent variable $s \rightarrow x=x_0+sh$, we get

$$I_{n+1}=\sum_{j=0}^{n}\int_p^q \prod_{\substack{k=0\\k\neq j}}^{n}\frac{\left(x_0+sh\right)-\left(x_0+kh\right)}{\left(x_0+jh\right)-\left(x_0+kh\right)}\,f_j\,h\,dx = h\sum_{j=0}^{n}\int_p^q \prod_{\substack{k=0\\k\neq j}}^{n}\frac{\left(s-k\right)}{\left(j-k\right)}\,f_j\,ds$$

After reorganization we get

$$I_{n+1}=h\sum_{j=0}^{n} f_j \prod_{\substack{k=0\\k\neq j}}^{n}\frac{1}{\left(j-k\right)}\int_p^q \prod_{k=0}^{n}\frac{\left(s-k\right)}{\left(s-j\right)}\,ds$$

Hence the *j-th* coefficient in the integrating formulas

$$\boxed{I_{n+1}=\sum_{j=0}^{n}\alpha_j\,f_j}$$

is given by

$$\alpha_j = \frac{h}{\prod\limits_{\substack{k=0 \\ k\neq j}}^{n}(j-k)} \int_p^q \prod_{k=0}^{n} \frac{(s-k)}{(s-j)} ds$$

since

$$\frac{1}{\prod\limits_{\substack{k=0 \\ k\neq j}}^{n}(j-k)} = \frac{(-1)^{n-j}}{j!(n-j)!} = \frac{(-1)^{n-j}}{n!}\binom{n}{j}$$

then finally we have

$$\boxed{\alpha_j = h\frac{(-1)^{n-j}}{n!}\binom{n}{j}\int_p^q \prod_{k=0}^{n}\frac{s-k}{s-j}ds} \qquad j=0, 1, \ldots, n$$

The following relation holds

$$I = \int_a^b f(x)\,dx = \sum_{j=0}^{n} \alpha_j f_j + E$$

where $E = I - I_{n+1}$ is the error term which may be evaluated for the Newton – Cotes formulas as follows

$$E = \begin{cases} \dfrac{2h^{n+2}}{(n+1)!} f^{(n+1)}(\xi) \displaystyle\int_0^{m+\frac{1}{2}} \prod_{k=\frac{-(2r-1)}{2}}^{\frac{(2r-1)}{2}} (s-k)\,ds & \text{- for } n \text{ odd } \xi \in (a,b) \\ \dfrac{2h^{n+3}}{(n+2)!} f^{(n+2)}(\eta) \displaystyle\int_0^{m} \prod_{k=0}^{r} \left(s^2-k^2\right) ds & \text{- for } n \text{ even} \end{cases}$$

The results of the above formulas may be presented in the tabular form given below.

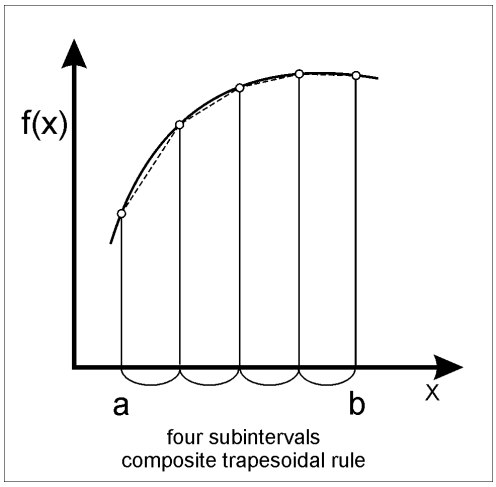
Coefficients α_j / h for NEWTON – COTES formulas (closed)

n	*j=0*	*j=1*	*j=2*	*j=3*	*Error term*	*Formula name*
0	*1*				$-h^3 f^{(2)}(\xi)$	*Rectangular*
1	$\frac{1}{2}$	$\frac{1}{2}$			$-\frac{1}{12}h^3 f^{(2)}(\xi)$	*Trapezoidal*
2	$\frac{1}{3}$	$\frac{4}{3}$	$\frac{1}{3}$		$-\frac{1}{90}h^5 f^{(4)}(\xi)$	*Simpson*
3	$\frac{3}{8}$	$\frac{9}{8}$	$\frac{9}{8}$	$\frac{3}{8}$	$-\frac{3}{80}h^5 f^{(4)}(\xi)$	

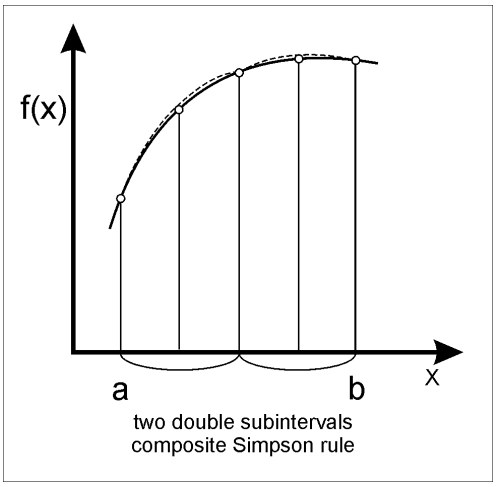
Conclusion

The Simpson formula displays the best accuracy up to third order integrating formulas.

10.2.1. Composite rules



four subintervals
composite trapesoidal rule



two double subintervals
composite Simpson rule

Idea

We subdivide the interval $[a,b]$ into a certain number of equal subintervals and apply in each of them the same appropriate rule (rectangular, trapezoidal, …). We get this way :

Composite rectangular rule

$$\int_a^b f(x)dx \approx h\sum_{i=0}^{n-1} f_i + (b-a)\frac{h}{2} f^{(1)}(\eta), \qquad a<\eta<b$$

Composite trapezoidal rule

$$\int_a^b f(x)dx \approx h\left[\frac{1}{2}(f_0+f_n)+\sum_{i=1}^{n-1} f_i\right] - \frac{(b-a)h^2}{12} f^{(2)}(\eta), \qquad a<\eta<b$$

Composite Simpson rule

$$\int_a^b f(x)dx \approx \frac{h}{3}\left[f_0 + f_{2n} + 2\sum_{i=1}^{n-1} f_{2i} + 4\sum_{i=0}^{n-1} f_{2i+1}\right] - \frac{(b-a)h^4}{180} f^{(4)}(\eta), \quad a<\eta<b$$

Multiple integrals

$$I = \int_a^b \int_{c(x)}^{d(x)} f(x,y)\,dydx$$

Idea : Apply twice 1D rule, e.g. for Simpson integration

$$I = \frac{\Delta x}{3}\left[g(a)+g(b)+4\sum_{\substack{j=1\\ j\ odd}}^{n-1} g(a+j\Delta x)+2\sum_{\substack{j=2\\ j\ even}}^{n-2} g(a+j\Delta x)\right]$$

where :

$$g(a) = \int_{c(a)}^{d(a)} f(a, y)dy$$

$$g\left(a + j\Delta x\right) = \int_{c(a+j\Delta x)}^{d(a+j\Delta x)} f\left(a + j\Delta x, y\right)dy$$

$$g(b) = \int_{c(b)}^{d(b)} f(b, y)dy$$

10.3. GAUSSIAN QUADRATURES

Find

$$I = \int_a^b F(z)dz$$

Let

$$x = \frac{2z - a - b}{b - a} \qquad \Leftrightarrow \qquad z = \frac{b-a}{2}\, x \; + \; \frac{a+b}{2}$$

$$dx = \frac{2}{b-a}\, dz \qquad \rightarrow \qquad dz = \underbrace{\frac{b-a}{2}}_{=J} dx$$

$$I = \int_a^b F(z)\, dz = \int_{-1}^{1} F\left(z(x)\right) \underbrace{\frac{dz}{dx}}_{=J} dx = \frac{b-a}{2} \int_{-1}^{1} F\left(\frac{b-a}{2}x + \frac{b+a}{2}\right) dx \equiv \frac{b-a}{2} \int_{-1}^{1} f\left(x\right)\, dx$$

Numerical quadrature

$$\int_{-1}^{1} f\left(x\right)\, dx = \sum_{i=1}^{n} w_i\, f\left(x_i\right), \qquad i = 1, 2, \ldots, n$$

n – number of sampling points

For arbitrary interval [*a*, *b*]

$$\int_a^b F\left(z\right)\, dz = \frac{b-a}{2} \int_{-1}^{1} F\left(z(x)\right) dx = \frac{b-a}{2} \int_{-1}^{1} f\left(x\right) dx$$

In the Gaussian integrating formulas we require *2n-1* order of accuracy, i.e. they should be exact for monomials x^k, $k = 0, 1, 2, \ldots, 2n-1$

$$\sum_{i=1}^{n} w_i\, x_i^k = \int_{-1}^{1} x^k\, dx = \frac{1}{k+1}\left[1 - \left(-1\right)^{k+1}\right]$$

Example

$$n = 2 \quad \rightarrow \quad 2n - 1 = 2 \cdot 2 - 1 = 3, \quad k = 0, 1, 2, 3$$

$$\begin{aligned} w_1 + w_2 &= 2 \\ w_1 x_1 + w_2 x_2 &= 0 \\ w_1 x_1^2 + w_2 x_2^2 &= \frac{2}{3} \\ w_1 x_1^3 + w_2 x_2^3 &= 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} w_1 &= 1 \qquad x_1 = -\frac{1}{\sqrt{3}} = -0.5773502 \\ w_2 &= 1 \qquad x_2 = +\frac{1}{\sqrt{3}} = +0.5773502 \end{aligned}$$

Example

Find

$$I = \int_0^4 \sqrt{1+z}\, dz = \frac{2}{3}\left(5\sqrt{5} - 1\right) = 6.78689$$

$$x = \frac{2z - 0 - 4}{4 - 0} = \frac{1}{2} z - 1 \quad \rightarrow \quad z = \frac{4-0}{2} x + \frac{0+4}{2} = 2\left(x + 1\right)$$

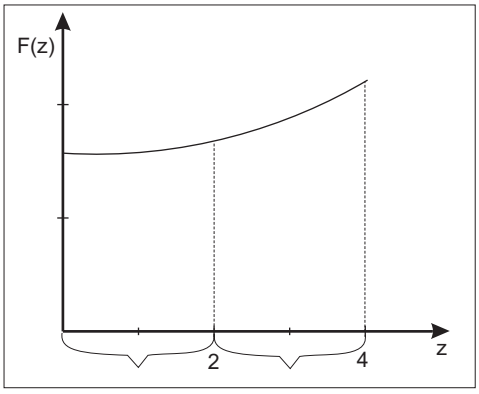
$$I = \frac{4-0}{2} \int_{-1}^{1} \sqrt{2x+3}\, dx \quad \rightarrow \quad f\left(x\right) = \sqrt{2x+3}$$

$$\tilde{I} = 1 \cdot 2\sqrt{2\left(-\frac{1}{\sqrt{3}}\right) + 3} \; + \; 1 \cdot 2\sqrt{2\left(\frac{1}{\sqrt{3}}\right) + 3} = 6.79345$$

Error

$$\frac{I - \tilde{I}}{I} = \frac{6.78689 - 6.79346}{6.78689} = -0.0009669 \approx -0.01\%$$

Remark



One may use the composite formula

$$\int_0^4 \ldots\, dz = \int_0^2 \ldots\, dz + \int_2^4 \ldots\, dz$$

Composite Gaussian – Legendre integration

$$I = \sum_{i=0}^{m-1} I_i = \frac{h}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{n} \alpha_j \, f\left[\frac{h\,x_j}{2} + \frac{1}{2}\left(z_i + z_{i+1}\right)\right]$$

$$z_0 = a, \quad z_1 = \frac{a+b}{2}, \quad z_2 = b$$

10.3.1. Derivation of the general Gauss-Legendre quadratives

In order to find the general formula of Gaussian integration let us consider the Hermite interpolations formula

$$f(x) = \sum_{i=0}^{n} h_i(x)\, f(x_i) + \sum_{i=0}^{n} g_i(x)\, f'(x_1) + E$$

where

$$h_i(x) = \left[1 - 2\left(x - x_i\right)\right] L'_i(x_i)\, L_i^2(x) \qquad i = 0, 1, \ldots, n$$

$$g_i(x) = \left(x - x_i\right) L_i^2(x) \qquad i = 0, 1, \ldots, n$$

$$E = \frac{\pi_{n+1}^2(x)\; f_{(\xi)}^{(2n+2)}}{(2n+2)!}$$

Integrating and interchanging the integration and summation signs we obtain

$$\int_{-1}^{1} f(x)\,dx = \sum_{i=0}^{n} \int_{-1}^{1} h_i(x)\, f(x_i)\,dx \; + \; \sum_{i=0}^{n} \int_{-1}^{1} g_i(x)\, f'(x_i)\,dx \; + E_I$$

Now, we choose an integrating formula of the form

$$\int_{-1}^{1} f(x)\,dx = \sum_{i=0}^{n} \alpha_i \, f(x_i) \; + \; E_I\,, \quad E_I = \int_{-1}^{1} E\,dx = \frac{f^{(n+2)}}{(2n+2)!} \int_{-1}^{1} \pi_{n+1}^2(x)\,dx$$

assuming that

$$\int_{-1}^{1} g_i(x)\; f'(x_i)\,dx = 0 \quad \rightarrow \quad \int_{-1}^{1} g_i(x)\,dx = \int_{-1}^{1} \left(x - x_i\right) L_i^2(x)\,dx = 0$$

$$i = 0, 1, \ldots, n$$

because $f'(x_i)$ may assume arbitrary value. Since the following holds

$$L_i(x) = \frac{\pi_{n+1}(x)}{\left(x - x_i\right)\pi'_{n+1}(x_i)}$$

the requirement is

$$\int_{-1}^{1} \frac{\pi_{n+1}(x)\; L_i(x)\,dx}{\pi'_{n+1}(x_i)} = 0$$

This may be interpreted so that $\pi_{n+1}(x)$ should be orthogonal to all polynomials of degree *n* or less on the interval [-1, 1].

The function $\pi_{n+1}(x)$ which satisfies this requirement is the appropriate polynomial defined by:

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$\vdots$$
$$P_i(x) = \frac{1}{i}\left[(2i-1)\,x\,P_{i-1}(x) - (i-1)\,P_{i-2}(x)\right] \qquad i = 2, 3, \ldots$$

namely the Legendre polynomials multiplied by the constant

$$\frac{2^{n+1}\left[(n+1)!\right]^2}{\left[2(n+1)\right]!}$$

The zeros of these polynomials are the required abscissas x_i for our integrating formula. Knowing the number of points we want to use in the integrating formula we consider the zeros of the appropriate Legendre polynomial.

The coefficient α_i, $i = 0, 1, \ldots, n$ are found now by integrating

$$\boxed{\alpha_i = \int_{-1}^{1} h_i(x)\,dx = L'_i(x_i)\int_{-1}^{1} L_i^{\,2}\,dx}$$

These values are tabulated

Degree of Legendre polynomials	Zeros of Legendre polynomials x_i	weights α_i
1	0	2
2	$\pm\frac{1}{\sqrt{3}}$	1,1
3	0 $\pm\sqrt{\frac{3}{5}}$	$\frac{8}{9}$ $\frac{5}{9}, \frac{5}{9}$
4	± 0.3399810436 ± 0.8611363116	0.6521451549 0.3478548451

10.3.2. Composite Gaussian – Legendre integration

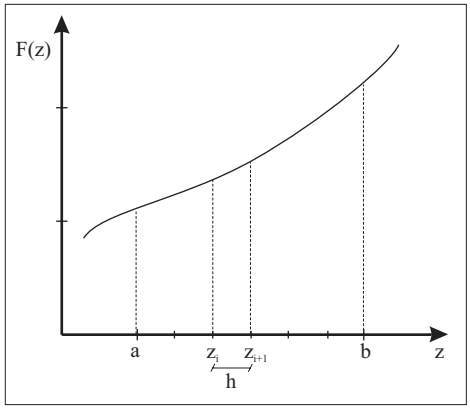
$$I = \sum_{i=0}^{m-1} I_i = \frac{h}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{n} \alpha_i f\left[\frac{h\, y_j}{2} + \frac{1}{2}\left(x_i + x_{i+1}\right) \right]$$

where y_i, $j = 0, 1, ..., n$ are the zeros of the $(n+1)$th degree of Legendre polynomials.

10.3.3. Summary of the Gaussian integration

$$\int_a^b F(z)\, dz = \frac{b-a}{2} \int_{-1}^{1} F\left(\frac{b-a}{2} x + \frac{a+b}{2} \right) dx = \frac{b-a}{2} \sum_{i=1}^{n} \alpha_i F\left(\frac{b-a}{2} x + \frac{a+b}{2} \right)$$

where α_i and x_i are taken from the table. This formula is exact for polynomials up to the $2n-1$ order. Usually the composite formula is being used.



$$I = \sum_{i=0}^{m-1} I_i = \frac{h}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{n} \alpha_j F\left[\frac{h\, z_i}{2} + \frac{1}{2}\left(z_i + z_{i+1}\right) \right]$$

$$a = z_0 < z_1 < ... < z_m = b$$

$$z_{i+1} - z_i = h \qquad i = 0, 1, ..., m-1$$

10.3.4. Special topics

Numerical integration can not be directly applied in case of singularity or infinite intervals. The following measures may be applied then

1) *The use of the weight functions and special quadratures*

$$\int_a^b f(x)\, dx = \int_a^b w(x)\gamma(x)\, dx = \sum_{i=0}^{n} \alpha_i \gamma(x_i) + E_i$$

Such approach is useful in case of singularities, for selected types of singularity.
As a result we come to Gauss – Chebyshev, Gauss – Laguerre, Gauss – Jacobi etc. formulas

2) *General way of dealing with singularities*

Example

$\int_0^1 \frac{x}{\sqrt{1-x}}\,dx$ - Approach: remove singularity by means of integration by parts first and apply numerical integration then

$$\int_0^1 \frac{x}{\sqrt{1-x}}\,dx = -x\cdot 2\sqrt{1-x}\Big|_0^1 + \int_0^1 2\sqrt{1-x}\,dx = 2\int_0^1 \sqrt{1-x}\,dx$$

3) *Integrals with infinite limits*

Approach: eliminate infinite limits by appropriate change of variable

Example

$$\int_0^\infty x^2 \mathrm{e}^{-x^2}\,dx = \int_0^1 x^2 \mathrm{e}^{-x^2}\,dx + \int_1^\infty x^2 \mathrm{e}^{-x^2}\,dx$$

Let $x^2 = y^{-1} \quad \rightarrow \quad dx = -\frac{1}{2y^{3/2}}\,dy$

$$\int_0^\infty x^2 \mathrm{e}^{-x^2}\,dx = \int_0^1 x^2 \mathrm{e}^{-x^2}\,dx + \frac{1}{2}\int_0^1 \frac{\mathrm{e}^{-\frac{1}{y}}}{y^{5/2}}\,dy = I_1 + I_2$$

Remarks

(i) integration by parts required

$$\int_0^1 \frac{\mathrm{e}^{-\frac{1}{y}}}{y^{5/2}}\,dy = \frac{1}{e} + \frac{1}{2}\int_0^1 \frac{\mathrm{e}^{-\frac{1}{y}}}{y^{3/2}}\,dy \quad \rightarrow \quad etc$$

(ii) singularity due to $e^{-\frac{1}{y}}$ remains

(iii) $$\frac{1}{2}\int_0^1 \frac{\mathrm{e}^{-\frac{1}{y}}}{y^{3/2}}\,dy = \frac{1}{2}\int_0^1 \frac{y^{-2}\mathrm{e}^{-\frac{1}{y}}}{y^{-2}y^{3/2}}\,dy = \frac{1}{2}\int_0^1 \frac{d\left(\mathrm{e}^{-\frac{1}{y}}\right)}{y^{-1/2}} = \frac{1}{4e} + \frac{3}{8}\int_0^1 y^{\frac{1}{2}} e^{-\frac{1}{y}}\,dy$$

and

$$\lim_{y\to 0_+} y^{\frac{1}{2}} e^{-\frac{1}{y}} = 0$$