

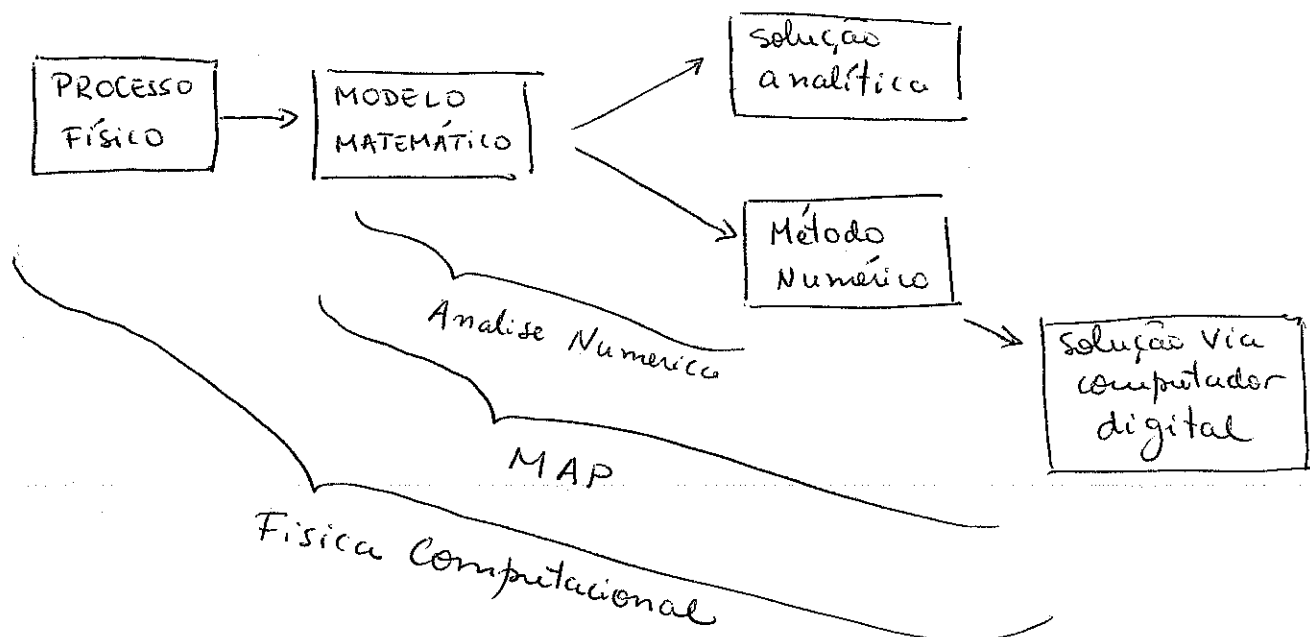
MAP214 - CÁLCULO NUMÉRICO COM APLICAÇÕES EM FÍSICA

Tópicos a serem abordados

- I) Erros, precisão numérica e ponto flutuante
- II) zeros de funções
- III) Matrizes e sistemas lineares, eliminação de Gauss, Gauss-Jacobi, Gauss-Seidel, inversão de Matrizes
- IV) interpolação de funções
- V) aproximação de funções por mínimos quadrados
- VI) Integração Numérica
- VII) Eq. diferenciais ordinárias -

Introdução

Métodos numéricos para solução de problemas matemáticos



Exemplo

PROCESSO FÍSICO : colisão de galáxias

MODELO MATEMÁTICO : EQ. DIFERENCIAIS DA GRAVITAÇÃO

MÉTODO NUMÉRICO : SOLUÇÃO NUMÉRICA DAS EQS. Diferenciais.

SOLUÇÃO VIA COMPUTADOR DIGITAL : PROGRAMAÇÃO DO MÉTODO Numérico.

Método Numérico

— utiliza somente as operações aritméticas

$+$, $-$, \times , \div

— Permite a solução de modelos matemáticos que não possuem solução analítica.

CAP I - ERROS, PRECISÃO NUMÉRICA E PONTO FLUTUANTE

① Onde aparecem os erros

a) ERROS NA CRIAÇÃO de um modelo matemático

Consideremos o caso da emissão de luz de uma lâmpada fluorescente (PLASMA).

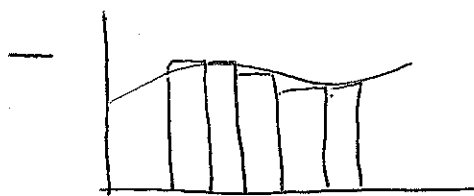
- conhecimento incompleto dos fenômenos
excitações podem gerar moléculas incomuns
- Natureza estatística
colisões entre partículas do gás não são determinísticas
- Incerteza nos parâmetros experimentais
- Eliminação de efeitos de ordem superior

ERRO INERENTE

b) ERROS ao adotarmos um método numérico

- truncamento do cálculo π através da série

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \dots$$



aproximação de áreas por retângulos

c) ERROS AO USARMOS o computador digital

→ nos reais são representados por um sistema de ponto flutuante → TRUNCAMENTO + ARREDONDAMENTO
ex.: Faça $1 \div 3$ e $2 \div 3$ na calculadora "ROUND OFF"

② ERRO ABSOLUTO E ERRO RELATIVO

$$e_x = x - \bar{x}$$

$\bar{x} \rightarrow$ valor verdadeiro

$x \rightarrow$ valor aproximado

$e_x \rightarrow$ erro absoluto

$$\text{erro relativo} = \frac{e_x}{\bar{x}} = \frac{x - \bar{x}}{\bar{x}}$$

Às vezes o erro absoluto é pequeno mas o erro relativo é grande

$$\text{Ex.: } \bar{x} = 0.0005$$

$$x = 0.0006$$

$$e_x = 0.0001 \rightarrow \frac{e_x}{\bar{x}} = 0.2 = 20\%$$

③ Definição de sistema de n.º de ponto flutuante (P.F.)

$$x = \pm .d_1 d_2 d_3 \dots d_n \times \beta^e$$

β - base

$d_1 \neq 0$ o n.º é dito normalizado

n = precisão (n.º de algarismos significativos na base β)

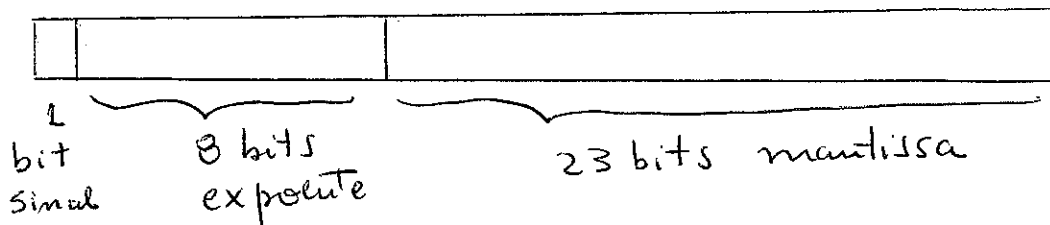
$f = (.d_1 d_2 \dots d_n)$ é a mantissa

$$L \leq e \leq U$$

, e = expoente

onde $0.1 \leq f < 1$

tipicamente com 32 bits



single precision - precisão simples 32 bits (4 bytes)

aprox. 7 dígitos decimais

double precision - dupla precisão 64 bits (8 bytes)

~ 16 dígitos decimais

PRECISION	FORTRAN	C, C++	sinal nº bits	expoente nº bits	valor do expoente	mantissa nº bits	TOTAL bits	nº aprox. decimais
SINGLE	REAL	float	1	8	(-126, +127)	23	32	7
DOUBLE	REAL*8	double	1	11	(-1022 + 1023)	52	64	16

↳ Entrar na Wikipedia e procurar por

IEEE 754 - 1985

ou " floating point arithmetic

Simplex	VALORES	MAIS DISTANTES DO ZERO	$\pm (1 - 2^{-24}) \times 2^{128}$	$\approx \pm 3,40 \times 10^{+38}$
	"	próximos	$\pm 2^{-126}$	$\approx \pm 1,17 \times 10^{-38}$
dupla	"	distantes	$\pm (1 - (\frac{1}{2})^{53}) \times 2^{+1024}$	$\approx \pm 1,79 \times 10^{+308}$
	"	próximos	2^{-1022}	$\approx \pm 2,22 \times 10^{-308}$

Ex: Construa um programa que calcule $n!$

de $n=1$ até o maior valor que possa ser representado

fator = 1

```

faça n=1 até 200
  fator = n*fator
  imprima n, fator
FIM DO LOOP
FIM DO PROGRAMA

```

FAÇA ISSO EM
PRECISÃO SIMPLES (4 bytes)
E DUPLA PRECISÃO (8 bytes)

Ex2: imprimir c/ todos os decimais

$1/3$, $\arcsin(-1)$ em precisão simples e dupla

④ Vamos denotar

\oplus , \ominus , \otimes e \odot operações em P.F.

em geral

$$x \oplus y \neq x + y$$

$$x \otimes y \neq x \cdot y$$

Associatividade em geral não é válida

Ex.: Faça em precisão simples

$$\left(\frac{1}{2} + \frac{1}{3}\right) + \frac{1}{9} \text{ e compare com}$$

$$\frac{1}{2} + \left(\frac{1}{3} + \frac{1}{9}\right)$$

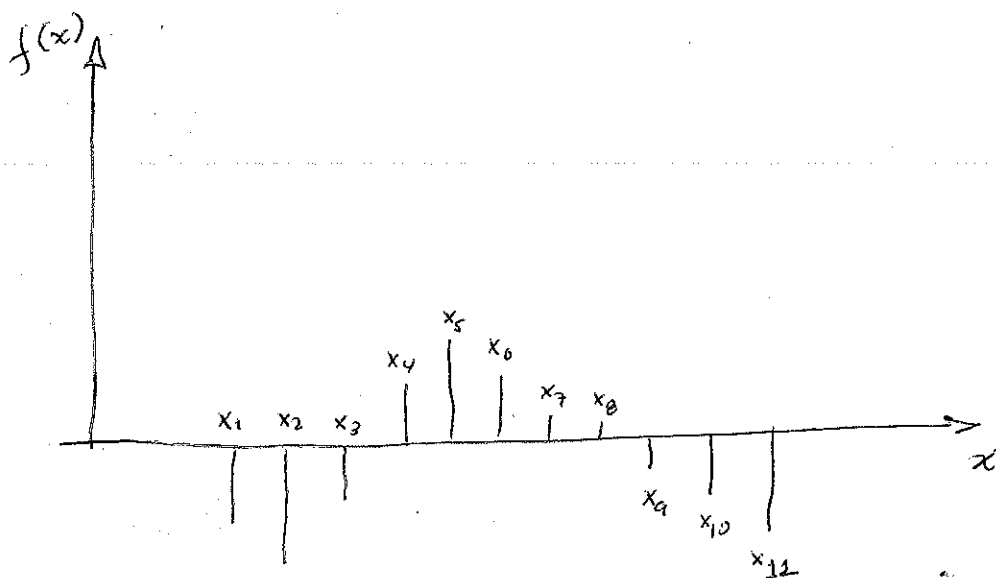
Distributiva não é válida em geral

$$x \otimes (y \oplus z) \neq x \otimes y \oplus x \otimes z$$

CAP II SOLUÇÃO DE EQUAÇÕES NÃO LINEARES $f(x)=0$

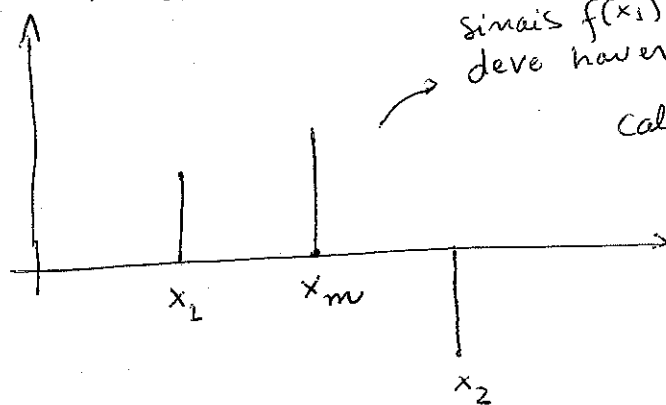
① O problema pode ser dividido em 2 etapas

- 1- Determinar intervalos (ou regiões) onde exista uma única solução p/ $f(x)=0$
- 2- Sabendo os intervalos, determinar a solução ali contida por meio de um algoritmo



Método pode ter problemas pois pode haver zeros entre x_2 e x_3 por exemplo

a) Método de bissecão [bisection or "half interval"]
ou dicotomia



se sinais $f(x_1)$ e $f(x_2)$ diferentes então
deve haver raiz entre eles.

$$\text{Calcula } x_m = \frac{x_1 + x_2}{2}$$

Calcula $f(x_m)$

se $\text{sinal } f(x_1) = \text{sinal } f(x_m)$

$$\rightarrow x_1 = x_m$$

caso contrário

$$\rightarrow x_2 = x_m$$

Algoritmo

seja uma raiz no intervalo $[a, b]$

$x_1, x_2 \leftarrow b$

calcular
 $f(x_1)$ e $f(x_2)$

$$x_m = \frac{x_1 + x_2}{2}$$

calcular $f(x_m)$

Se $\text{Sinal } f(x_m) = \text{Sinal } f(x_1)$

V

$x_1 \leftarrow x_m$

F

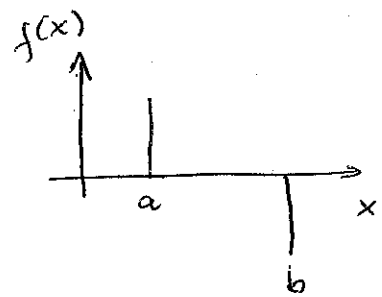
$x_2 \leftarrow x_m$

F

$|x_2 - x_1| < \epsilon$

V

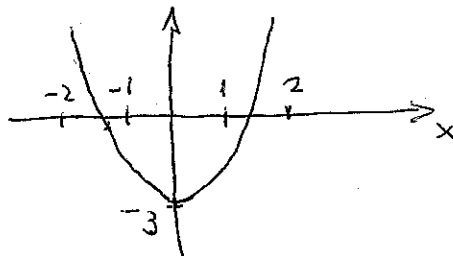
sol $\leftarrow x_1$ ou x_2



Encontre a solução de

Exemplo : $x^2 - 3 = 0$ pelo método de bissecção.

Solução $f(x) = x^2 - 3$



x_1	x_2	x_m	$f(x_1)$	$f(x_m)$	e_n
1	2	1.5	-3	-0.75	1
1.5	2	1.75	-0.75	0.0625	0.5
1.5	1.75	1.625	-0.75	-0.359	0.25
1.625	1.75	1.6875	-0.359	-0.152	0.125
1.6875	1.75	1.71875	-0.152	-0.0459	0.0625
1.71875	1.75	1.734375	-0.0459	$+8.05 \times 10^{-3}$	0.03625
1.71875	1.734375	1.7265625	-0.0459	-0.01898	2^{-6}
1.7265625	1.734375	1.73046875	-0.01898	-5.47×10^{-3}	2^{-7}
1.73046875	1.734375	1.732421875	-5.47×10^{-3}	1.285×10^{-3}	2^{-8}

Exercício Refazer c/ calculadora 4 operações $x^2 - 5 = 0$

$$e_{n+1} = \frac{1}{2} e_n$$

De uma forma geral um método é dito de convergência de ordem m se na vizinhança da solução temos

$$|e_{n+1}| = cte |e_n|^m \quad \text{ou} \quad \lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|^m} = cte$$

No caso da bissecção, $cte = \frac{1}{2}$ e $m=1$ é dito de convergência linear

b) Método de substituições sucessivas

Escrevemos a equação $f(x) = 0$ na forma

$$x = G(x) \Rightarrow x_{n+1} = G(x_n) \quad (1)$$

↳ função iterativa

Ex: $x^2 - 3x + 2 = 0$

Pode ser escrita na forma

$$x = (x^2 + 2)/3$$

$$x_{n+1} = (x_n^2 + 2)/3, \quad n = 0, 1, 2, \dots$$

Por exemplo, tomando $x_0 = 0$ fornece

$$x_1 = 0.666666667$$

$$x_2 = 0.887974394$$

⋮

$$x_9 = 0.99151400 \rightarrow \text{converge p/ o ponto fixo } a = 1$$

Estudo de convergência

$$e_{n+1} = x_{n+1} - a \stackrel{(1)}{=} G(x_n) - G(a) = G(a + e_n) - G(a) \quad (2)$$

Pela expansão em série de Taylor temos que

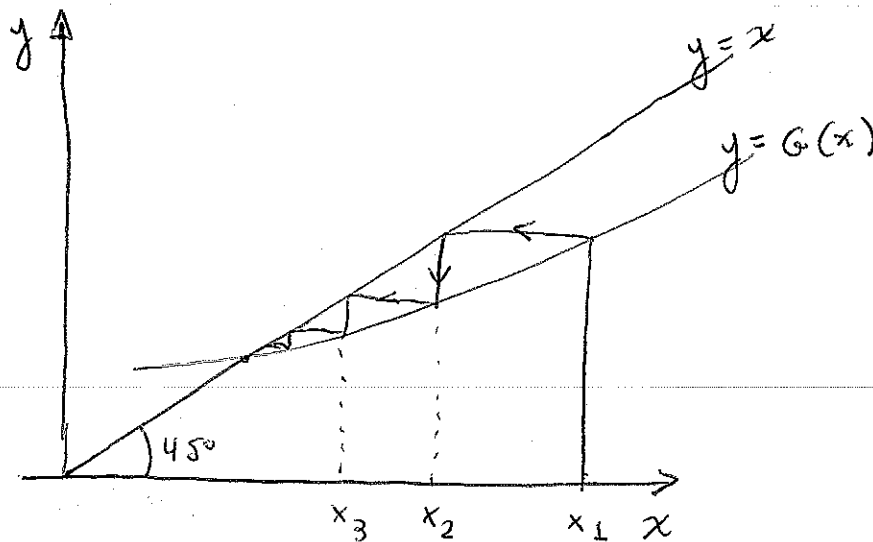
$$G(a + e_n) = G(a) + e_n G'(\xi); \quad a < \xi < a + e_n \quad (3)$$

Substituindo (3) em (2) temos

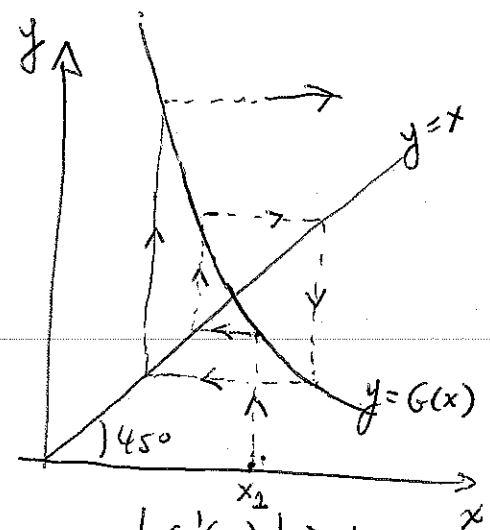
$$e_{n+1} = G'(\xi) e_n$$

Se existe uma constante C tal que $|G'(x)| \leq C < 1$ perto de $x=a$, então $|e_{n+1}| \leq C|e_n| \leq C^2|e_{n-1}| \leq \dots C^{n+1}|e_0|$ e e_{n+1} tende a zero. Como há uma relação linear entre e_n e e_{n+1} , a iteração converge linearmente para a raiz. Se $|G'(x)| < 1$ em algum intervalo contendo a solução e a estimativa inicial x_0 .

Interpretação geométrica



$|G'(x)| < 1$
convergente



$|G'(x)| > 1$
instável

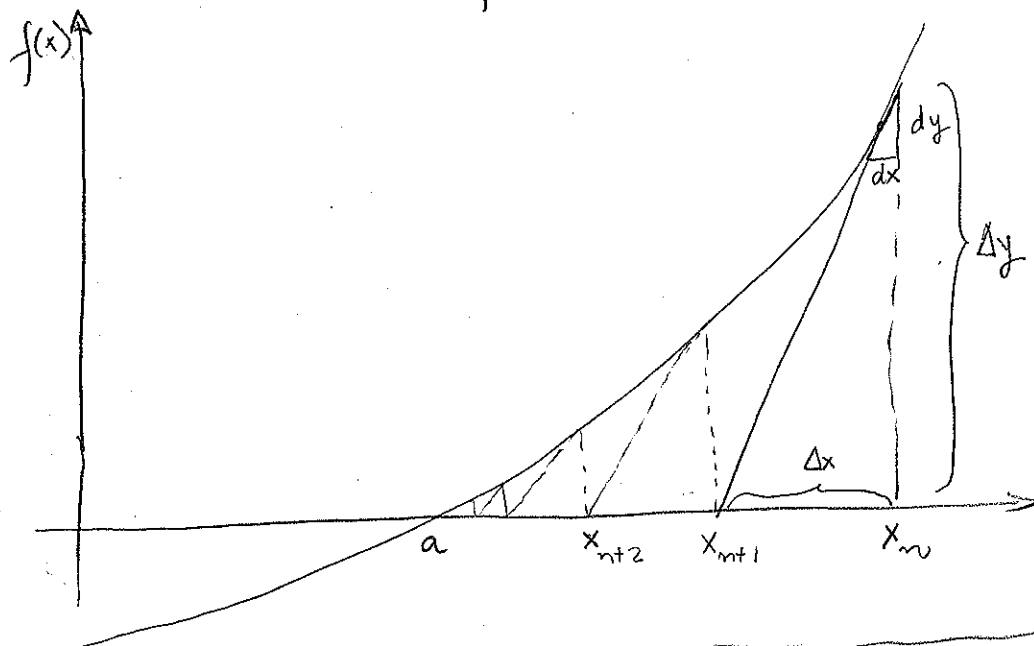
Nota importante

Funções iterativas podem levar a resultados não convergentes. Em alguns casos são caóticas como no clássico exemplo da equação logística

$$x_{n+1} = r x_n (1 - x_n)$$

↳ procurar na internet...

c) Método de Newton-Raphson (ou Método de Newton)



$$f'(x_n) = \frac{dy}{dx} = \frac{\Delta y}{\Delta x} = \frac{f(x_n)}{x_n - x_{n+1}}$$

$$\Rightarrow \boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}} \quad f'(x_n) \neq 0$$

Podemos escrever na forma $x_{n+1} = G(x_n)$
onde $G(x) = x - \frac{f(x)}{f'(x)}$ ②

De forma geral podemos escrever o erro como

$$e_{n+1} = x_{n+1} - a = G(x_n) - G(a) = G(a + e_n) - G(a) \quad ③$$

Expandindo em série de Taylor

$$G(a + e_n) = G(a) + e_n G'(a) + e_n^2 \frac{G''(\xi)}{2}, \quad a < \xi < a + e_n \quad ④$$

derivando $G(x)$ em relação a x no ponto a

$$G'(a) = 1 - \left[\frac{f'(a)}{f'(a)} - \frac{f''(a)f'(a)}{f'(a)^2} \right] = 0, \quad ⑤$$

e $G''(x)$ próximo a a vale aproximadamente $G''(a) = \frac{f''(a)}{f'(a)} \quad ⑥$

Subst. 4, 5 e 6 em ③ temos próximo a a

$$e_{n+1} \approx \frac{f''(a)}{2f'(a)} e_n^2 \rightarrow \text{método converge quadraticamente na vizinhança de } a$$

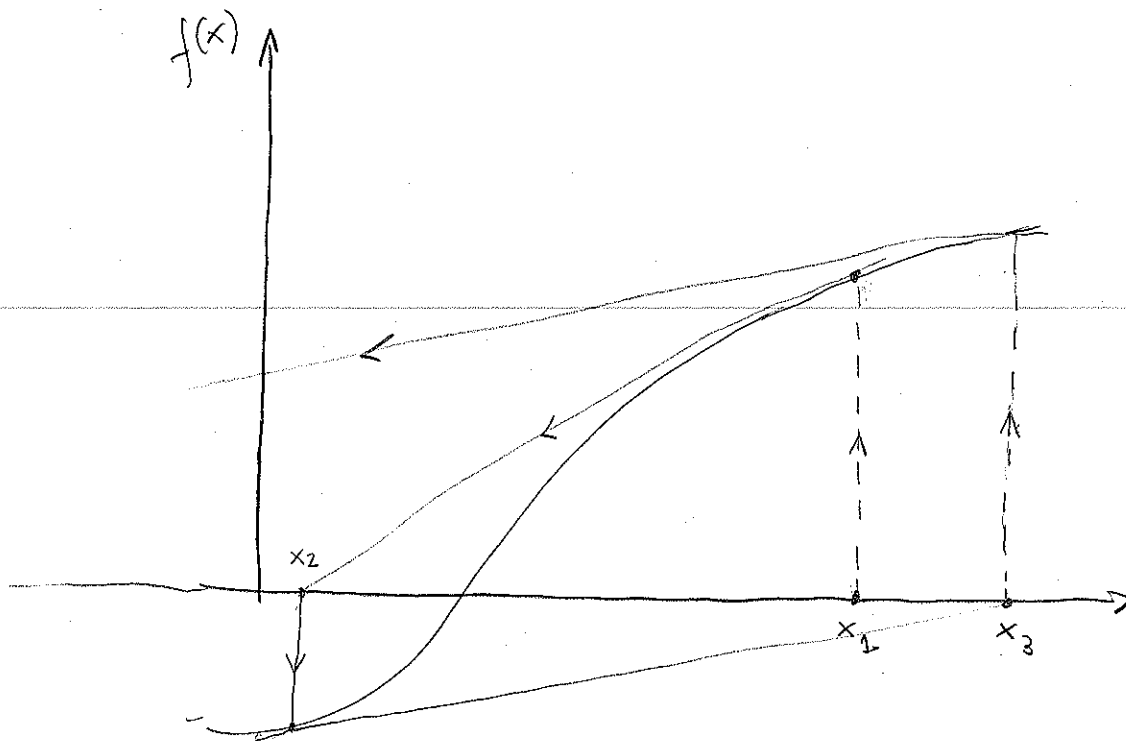
Para convergência devemos ter $e_1 = \underbrace{\frac{f''(x)}{2f'(x)}}_{< 1 \text{ em módulo}} e_0 e_0$

$$\Rightarrow |e_0 f''(x)/f'(x)| < 2$$

↳ condição inicial p/ convergência p/ $\forall x$ perto de a .

Teorema: Se $f(a) = 0$, $f'(a) \neq 0$ e f'' é contínua então há um intervalo aberto N contendo a tal que se $x_1 \in N$ no método de N.R. então $x_n \rightarrow a$ quando $n \rightarrow \infty$.

— Analisando o teorema observemos a seguinte figura



- A escolha de x_1 fez c/ que o método não convergisse ($x_1 \notin N$)
- ^{Na prática} Como determinar x_1 ?
- Faz-se previamente um gráfico de $f(x)$ ou usa-se um método seguro (e.g. bissecção) para encontrar x_1 próximo de a e em seguida aplica-se N.R. p/ chegar à solução final.

Exemplo: encontre a solução de $e^{-x} = x$ usando o método de Newton-Raphson

Solução $f(x) = e^{-x} - x$

$$f'(x) = -e^{-x} - 1, \quad f''(x) = e^{-x}$$

então $|f(x)/f'(x)| < 1$ para todos os valores de x

Desta forma, desde que $e_0 < 2$ devemos ter convergência quadrática.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{e^{-x_n} - x_n}{-e^{-x_n} - 1} = x_n + \frac{e^{-x_n} - x_n}{e^{-x_n} + 1}$$

tomando $x_0 = 0.5$

n	x_n	$e^{-x_n} - x_n$	$-(e^{-x_n} + 1)$
0	0.5	0.106530659	-1.60653066
1	0.566311002	$1.304511685 \times 10^{-3}$	-1.5676155
2	0.567143165	1.96511×10^{-7}	-1.567143362
3	0.56714329	4.2×10^{-10}	

↓ feito c/ CALCULADORA

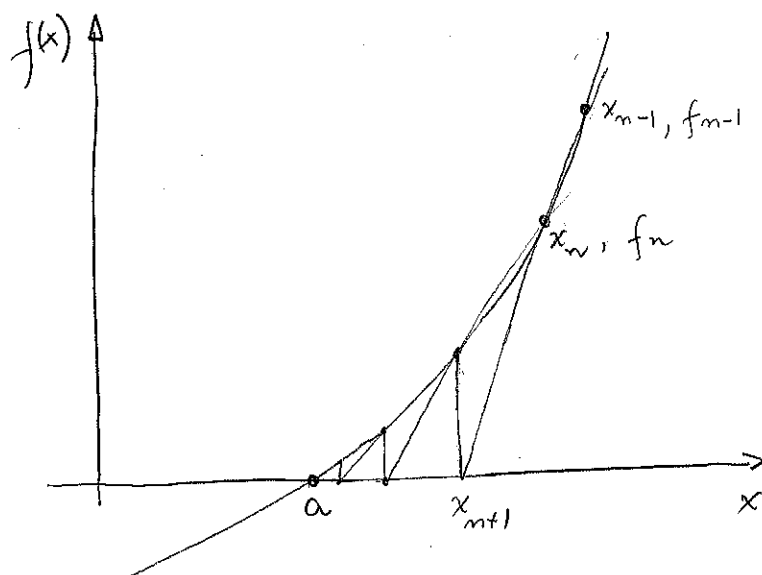
Resultados ligeiramente diferente no computador.

Um critério p/ parar o programa é

$$\left| \frac{x_{n+1} - x_n}{x_n} \right| < \epsilon, \text{ onde } \epsilon \text{ é especificado.}$$

↳ erro relativo

d) Método das secantes



$$f_n \equiv f(x_n)$$

O método das secantes pode ser obtido a partir do N.R. substituindo $f'(x_n)$ por $\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$

$$\Rightarrow x_{n+1} = x_n - \frac{f(x_n)}{[f(x_n) - f(x_{n-1})]/(x_n - x_{n-1})}$$

$$\text{ou } x_{n+1} = x_n - \frac{(x_n - x_{n-1}) f(x_n)}{f(x_n) - f(x_{n-1})}$$

Teorema: Se $f(a) = 0$, $f'(a) \neq 0$, $f''(a) \neq 0$ e f'' contínua então:

1- Existe um intervalo aberto N contendo a tal que se x_0 e x_1 são distintos em N a sequência x_n do método das secantes converge para a qda $n \rightarrow \infty$.

$$2 - \lim_{n \rightarrow \infty} \left| \frac{e_{n+1}}{e_n^m} \right| = cte \neq 0, \text{ onde } m = \frac{1 + \sqrt{5}}{2} \approx 1,618$$

entre linear e quadrático

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})} \quad (1)$$

$$e_{n-1} \equiv x_{n-1} - a$$

$$e_n \equiv x_n - a$$

$$e_{n+1} \equiv x_{n+1} - a$$

$$e_{n+1} + a = e_n + a - \frac{f(x_n)(e_n - e_{n-1})}{f(x_n) - f(x_{n-1})}$$

$$e_{n+1} = \frac{\cancel{e_n f(x_n)} - e_n f(x_{n-1}) - \cancel{f(x_n) e_n} + f(x_n) e_{n-1}}{f(x_n) - f(x_{n-1})}$$

$$= \frac{e_{n-1} f(x_n) - e_n f(x_{n-1})}{f(x_n) - f(x_{n-1})}$$

$$= \frac{e_{n-1} \left\{ \cancel{e_n f'(a)} + \frac{e_n^2}{2!} f''(a) + \frac{e_n^3}{3!} f'''(\xi) \right\} - e_n \left\{ \cancel{e_{n-1} f'(a)} + \frac{e_{n-1}^2}{2!} f''(a) + \frac{e_{n-1}^3}{3!} f'''(\xi) \right\}}{f(x_n) - f(x_{n-1})}$$

$$\approx \frac{e_{n-1} e_n \left[(e_n - e_{n-1}) \frac{f''(a)}{2} \right]}{f(x_n) - f(x_{n-1})} = \frac{e_{n-1} e_n \underbrace{\left(\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \right)}_{\approx f'(a)} \frac{f''(a)}{2}}{f(x_n) - f(x_{n-1})}$$

$$\boxed{e_{n+1} \approx \underbrace{\frac{f''(a)}{2 f'(a)}}_A e_{n-1} e_n} \quad (2)$$

Para obtermos a ordem de convergência m buscamos uma expressão na forma

$$\text{ou } \left. \begin{aligned} e_n &= B e_{n-1}^m \\ e_{n+1} &= B e_n^m \end{aligned} \right\} e_{n+1} = B [B e_{n-1}^m]^m$$

Substituindo em (2) temos

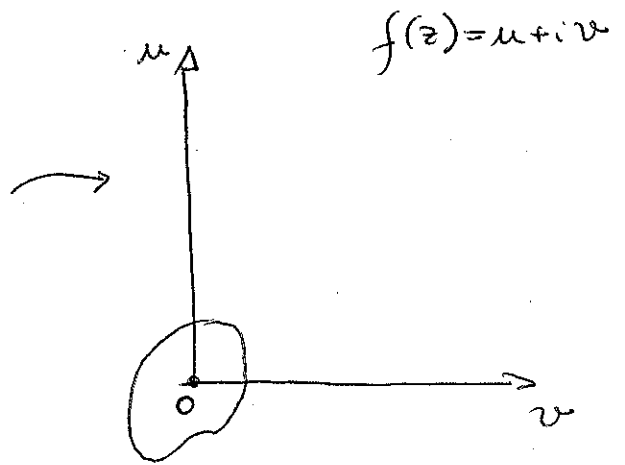
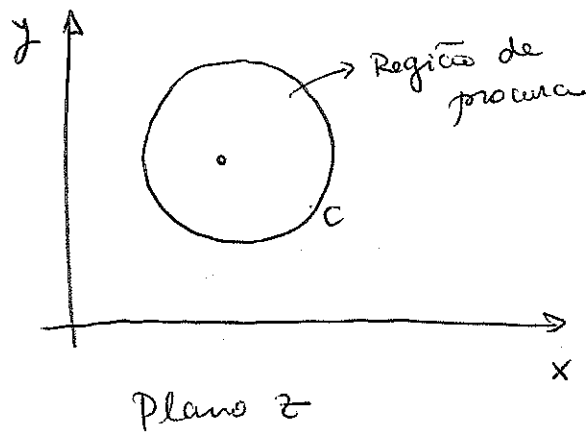
$$B [B e_{n-1}^m]^m = A e_{n-1} B e_{n-1}^m$$

$$e_{n-1}^{m^2 - m - 1} = \frac{A}{B^m} = \text{cte} \Rightarrow m^2 - m - 1 = 0$$

$$\Rightarrow m = \frac{1 + \sqrt{5}}{2} = 1.618...$$

- possui problema semelhante ao N.R. na localização de x_0 e x_1 de modo a convergir p/ a raiz. Tb pode iniciar com bissecção para depois aplicar secantes.
- NÃO precisa do cálculo de derivadas \rightarrow pode ser muito vantajoso qdo $f(x)$ for muito complicada ou desconhecida analiticamente.
 - \hookrightarrow ex. resultado de outro cálculo numérico.
- Do ponto de vista da eficiência, o método do N.R. precisa avaliar 2 funções ($f(x_n)$ e $f'(x_n)$) por iteração, então sua "ordem de convergência efetiva" é $\sqrt{2} \approx 1.414$. Já no método de secantes só é necessária 1 avaliação de função a cada iteração. Sob esse aspecto podemos dizer que o método de secantes é mais eficiente do que o N.R. Isto é evidenciado em funções muito complicadas, onde teremos um gde nº de operações envolvidas.

② Raízes complexas



- Se ao variarmos z sobre C , $f(z)$ circundará a origem ao menos uma vez, podemos dizer que dentro de C existe uma raiz complexa.
- Métodos de N.R. e secantes podem ser aplicados s/ alterações

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)} \quad \text{e} \quad z_{n+1} = z_n - \frac{f(z_n)(z_n - z_{n-1})}{f(z_n) - f(z_{n-1})}$$

↙
ver Carnahan p.171

- São necessários z_0, z_1 próximos à raiz $\left\{ \begin{array}{l} \text{divergência} \\ \text{convergência p/ raiz} \\ \text{circula em} \\ \text{torno de } z \end{array} \right.$
- Se uma raiz for encontrada podemos encontrar outras raízes a partir da função $f_1(z) = \frac{f(z)}{z - a}$
"deflation"
- Uma outra ideia é minimizar $|f(z)|^2 = u^2 + v^2$
ou localizar proximidade do mínimo e \rightarrow veja Carnahan.
aplicar N.R.

VERSO
 \rightarrow

Raízes de polinômios

- Alguns polinômios são patológicos e pequenas mudanças nos coef. podem no pior caso fazer com que raízes reais vão para o plano complexo
→ caso Wilkinson

$$(x-1)(x-2)\dots(x-19)(x-20)$$

raízes

{1, 2, ..., 19, 20
↳ veja ^{livro} Action

- P/ cálculo de raízes pode-se alternativamente usar método de determinação de auto-valores de um polinômio característico
↳ veja Numerical Recipes.

CAP III - SISTEMAS DE EQUAÇÕES LINEARES

① Introdução

- É um dos problemas mais frequentes na área científica

- Forma geral:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

\vdots

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n$$

Isto equivale a

$$Ax = b$$

ou

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

A

matriz $n \times n$

x

↓
vetor
dimensão
 n

b

↓
vetor
dimensão
 n

A e b dados, $x_1 \dots x_n$ a serem determinados.

- O mais conhecido método de resolução é a Regra de Cramer

$x_i =$

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1,i-1} & b_1 & a_{1,i+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2,i-1} & b_2 & a_{2,i+1} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n,i-1} & b_n & a_{n,i+1} & \dots & a_{nn} \end{vmatrix}$$

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

— No entanto se tentarmos resolver um sistema de 20 equações por Cramer teríamos:

21 determinantes de ordem 20

— Cada determinante é decomposto em

$$a_{11} \det_{19} + a_{12} \det_{19} + \dots + a_{1,20} \det_{19}$$

— Até chegarmos em determinantes de ordem 2 temos

$$\underbrace{21 \times 20!}_{(n+1)!} \text{ operações} \approx 10^{20} \text{ operações}$$

- Podemos ter $\begin{cases} \text{matrizes cheias} \\ \text{matrizes esparsas} \end{cases}$

- exemplo de matrizes esparsas

$$\begin{bmatrix} 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 6 & 0 \end{bmatrix};$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix};$$

$$\begin{bmatrix} -2 & 1 & & 0 \\ 1 & -2 & 1 & \\ & 1 & -2 & \ddots \\ 0 & & \ddots & 1 \\ & & & 1 & -2 \end{bmatrix}$$

matriz
de banda

matriz
tridiagonal

- Sistemas esparsos possuem soluções particulares e + rápidas que os cheios
- Estudaremos inicialmente métodos de solução para matrizes cheias.

② Métodos diretos

2.1 Método de Eliminação de Gauss (ou Método de Gauss)

a) O método

→ é o método mais simples p/ solução de um sistema de equações

→ Considere, por exemplo, o seguinte sistema de eqs.:

$$\begin{cases} 2x + y + z = 7 \\ 4x + 4y + 3z = 21 \\ 6x + 7y + 4z = 32 \end{cases}$$

Multiplicando a 1ª eq. por (-2) e somando na 2ª,
" a 1ª eq. por (-3) e somando na 3ª temos

$$\begin{cases} 2x + y + z = 7 \\ 2y + z = 7 \\ 4y + z = 11 \end{cases}$$

Multiplicando a 2ª eq. por (-2) e somando na 3ª temos

$$\begin{cases} 2x + y + z = 7 \\ 2y + z = 7 \\ -z = -3 \end{cases}$$

Da 3ª eq. temos $-z = -3 \Rightarrow \boxed{z = 3}$

Substituindo na 2ª eq. temos $2y + 3 = 7 \Rightarrow \boxed{y = 2}$

Substituindo na 1ª eq. temos $2x + 2 + 3 = 7 \Rightarrow \boxed{x = 1}$

— Na forma matricial estamos resolvendo

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & 4 & 3 \\ 6 & 7 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ 21 \\ 32 \end{bmatrix}$$

e transformamos o problema em

$$\begin{bmatrix} 2 & 1 & 1 \\ & 2 & 1 \\ & & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ -3 \end{bmatrix}$$

matriz
triangular superior

— Poderíamos só ter trabalhado c/ os números s/ escrever as eq. Para tanto é conveniente escrever a chamada matriz aumentada

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 4 & 4 & 3 & 21 \\ 6 & 7 & 4 & 32 \end{array} \right]$$

Como antes multiplicamos a 1ª eq. por 2 e subtraímos da 2ª
1ª eq. " 3 e " da 3ª

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 0 & 2 & 1 & 7 \\ 0 & 4 & 1 & 11 \end{array} \right]$$

Então multiplicamos a 2ª eq por 2 e subtraímos da 3ª

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 0 & 2 & 1 & 7 \\ 0 & 0 & -1 & -3 \end{array} \right]$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ & a_{22} & & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

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

b) N° de operações realizadas

$$\begin{matrix} n-1 \\ \text{linhas} \end{matrix} \left\{ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & | & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & | & b_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & | & b_n \end{bmatrix} \right.$$

i) Eliminação

(n-1) linhas e p/ cada uma delas calculamos n+1 multiplicações
n adições

	multiplicações	adições
eliminação de 1ª coluna	(n-1)(n+1)	(n-1)n
2ª coluna	(n-2)n	(n-2)(n-1)
⋮	⋮	⋮
(n-1)ª coluna	1.3	2.1
Total	$\sum_{i=1}^{n-1} i(i+2)$	$\sum_{i=1}^{n-1} (i+1)i$

P/ multiplicações $\sum_{i=1}^{n-1} i(i+2) = \sum_{i=1}^{n-1} i^2 + \sum_{i=1}^{n-1} 2i$

$$\sum_{i=1}^{n-1} i^2 = \sum_{i=1}^n i^2 - n^2 = \frac{n(n+1)(2n+1)}{6} - n^2 = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}$$

e $\sum_{i=1}^{n-1} i = \sum_{i=1}^n i - n = \frac{n(n+1)}{2} - n = \frac{n^2}{2} - \frac{n}{2}$

$$\Rightarrow \sum_{i=1}^{n-1} i^2 + \sum_{i=1}^{n-1} 2i = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6} + 2 \left\{ \frac{n^2}{2} - \frac{n}{2} \right\} = \frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6}$$

P/ adições De forma análoga $\sum_{i=1}^{n-1} (i+1)i = \frac{n^3}{3} - \frac{n}{3}$

ii) Back substitution

1º passo (linha n)	1 multiplicação	nenhuma adição
2º passo (linha n-1)	2 multiplicações	uma adição
⋮	⋮	⋮
nésimo passo (linha 1)	n multiplicações	n-1 adições
TOTAL	$\frac{n(n+1)}{2}$	$\frac{(n-1)n}{2}$ adições

Gaussian Elimination = forward elimination + back^{ward} substitution

$$\text{TOTAL}_{\text{multiplicações}} = \frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6} + \left\{ \frac{n(n+1)}{2} \right\} = \frac{n^3}{3} + n^2 - \frac{n}{3}$$

$$\text{TOTAL}_{\text{adições}} = \frac{n^3}{3} - \frac{n}{3} + \left\{ \frac{(n-1)n}{2} \right\} = \frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6}$$

→ n gde temos ^{entre adições e multiplicações} um total de

$$\approx \frac{2n^3}{3} \quad \text{operações de ponto flutuante}$$

Voltando ao nosso problema, um ^{sistema}matriz $20 \times 20 \rightarrow \frac{2^3 \times 10^3}{3} \approx 3 \times 10^3$ flop
com um PC com 1000 Mflops será resolvido

$$\text{em } t = \frac{3 \times 10^3 \text{ flop}}{1000 \text{ Mflop/s}} \approx 10^{-6} \text{ s}!$$

↓ só tempo de CPU
sem contar perda de
eficiência no acesso
à memória.

c) O pivotamento e sua necessidade
Seja o sistema

$$\begin{bmatrix} 10 & -7 & 0 \\ -3 & 2.099 & 6 \\ 5 & -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3.901 \\ 6 \end{bmatrix}$$

não solução real e $\vec{x} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$

— Vamos considerar que nosso P.F. tenha apenas
5 algarismos significativos

- Multiplicando 1ª eq por 0.3 e somando na 2ª
- Multiplicando 1ª eq por 0.5 e somando na 3ª

$$\begin{bmatrix} 10 & -7 & 0 & 7 \\ 0 & -0.001 & 6 & 6.001 \\ 0 & 2.5 & 5 & 2.5 \end{bmatrix}$$

- Multiplicando a 2ª eq por $\frac{-2.5}{-0.001} = 2500$ e somando na 3ª

$$\begin{bmatrix} 10 & -7 & 0 & 7 \\ & -0.001 & 6 & 6.001 \\ & 0 & 15005 & 15004 \end{bmatrix}$$

$$\begin{array}{r} 6.001 \times 2500 = 15002.5 \\ + 2.5 \\ \hline 15004. \end{array}$$

↓
só 5 algarismos
significativos

→ Na substituição teremos

$$x_3 = \frac{15004}{15005} = 0.99993$$

$$x_2 = \frac{6.001 - 6 \times 0.99993}{-0.001} = \frac{6.001 - 5.99958}{-0.001} = -1.5$$

$$x_1 = \frac{7 + 7 \times (-1.5)}{10} = \frac{7 - 10.5}{10} = -0.35$$

(erro muito grande)

O pivô (-0.001) é muito pequeno em relação aos outros coeficientes o que resulta num enorme multiplicador (2500) que faz aparecerem erros em P.F. Estes erros por sua vez são ampliados devido à divisão pelo pequeno pivô.

No método de Gauss utilizaremos o pivotamento parcial que consiste em trocar linhas de forma que tenhamos o maior valor ^{absoluto} do pivô. Isto garantirá multiplicadores ≤ 1 em módulo.

No exemplo

$$\begin{bmatrix} 10 & -7 & 0 & | & 7 \\ 0 & -0.001 & 6 & | & 6.001 \\ & 2.5 & 5 & | & 2.5 \end{bmatrix} \xrightarrow{\text{PIVOTAMENTO PARCIAL}} \begin{bmatrix} 10 & -7 & 0 & | & 7 \\ & 2.5 & 5 & | & 2.5 \\ & -0.001 & 6 & | & 6.001 \end{bmatrix}$$

$$\text{Multiplicador: } \frac{-0.001}{-2.5} = 0.0004$$

2ª eq

$$0.0004 \times [0 \quad 2.5 \quad 5 \quad | \quad 2.5] \rightarrow$$

$$+ \begin{array}{ccc|c} 0. & 0.0004 & 0.002 & 0.001 \\ 0 & -0.001 & 6 & 6.001 \\ \hline 0 & 0 & 6.002 & 6.002 \end{array}$$

$$\begin{bmatrix} 10 & -7 & 0 & | & 7 \\ & 2.5 & 5 & | & 2.5 \\ & 0 & 6.002 & | & 6.002 \end{bmatrix}$$

$$\rightarrow x_3 = 1$$

$$x_2 = -1$$

$$x_1 = 0 //$$

d)

Obtenção da Matriz Inversa por Gauss"almost
NEVER INVERT A MATRIX"

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & & \\ 4 & 4 & 3 & & 1 & \\ 6 & 7 & 4 & & & 1 \end{array} \right]$$

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

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & -2 & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array} \right]$$

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array} \right]$$

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array} \right]$$

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

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

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

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

e) Soluções com diferentes vetores do lado direito

$$A\vec{x}^{(1)} = \vec{b}^{(1)} ; A\vec{x}^{(2)} = \vec{b}^{(2)} ; \dots$$

Ex:

$$\left[\underbrace{\quad}_{A} \quad \underbrace{\quad}_{\vec{b}^{(1)}} \right] ; \left[\quad \underbrace{\quad}_{\vec{b}^{(2)}} \right]$$

Em vez de fazer a mesma eliminação 2 vezes podemos "guardar" a sequência de operações aplicadas na triangulação da matriz A p/ depois aplicar em $\vec{b}^{(1)} ; \vec{b}^{(2)}$.

Ex: Seja a matriz

$$A = \begin{bmatrix} 2 & 6 & -2 \\ 1 & 3 & -4 \\ 3 & 6 & 9 \end{bmatrix}$$

$$\vec{p} = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}$$

vetor de pivotamento
que contém o n.º da
linha que foi pivotada

Pivotamento da 3ª linha

$$\begin{bmatrix} 3 & 6 & 9 \\ 1 & 3 & -4 \\ 2 & 6 & -2 \end{bmatrix} ; \begin{bmatrix} 3 \\ \quad \\ \quad \end{bmatrix}$$

1ª linha $(x - 1/3)$ somar 2ª linha
1ª linha $(x - 2/3)$ " 3ª linha

$$\begin{bmatrix} 3 & 6 & 9 \\ -1/3 & 1 & -7 \\ -2/3 & 2 & -8 \end{bmatrix} ; \begin{bmatrix} 3 \\ \quad \\ \quad \end{bmatrix}$$

Pivotamento da 3ª linha

$$\begin{bmatrix} 3 & 6 & 9 \\ -1/3 & 2 & -8 \\ -2/3 & 1 & -7 \end{bmatrix}; \quad \vec{p} = \begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$$

2ª linha $\times -1/2$ e somar na 3ª

$$\begin{bmatrix} 3 & 6 & 9 \\ -1/3 & 2 & -8 \\ -2/3 & -1/2 & -3 \end{bmatrix}; \quad \vec{p} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

P/aplicação, vamos agora fornecer o vetor $b = \begin{bmatrix} 4 \\ -7 \\ 39 \end{bmatrix}$

No 1º passo devemos trocar linha 1 com linha $p(1)=3 \rightarrow \begin{bmatrix} 39 \\ -7 \\ 4 \end{bmatrix}$

multiplicar 1ª linha $\times (-1/3)$ e somar na 2ª
1ª linha $\times (-2/3)$ e somar na 3ª $\rightarrow \begin{bmatrix} 39 \\ -20 \\ -22 \end{bmatrix}$

Trocar linha 2 com linha $p(2)=3 \rightarrow \begin{bmatrix} 39 \\ -22 \\ 20 \end{bmatrix}$

multiplicar 1ª linha $\times -1/2$ e somar na 3ª $\rightarrow \begin{bmatrix} 39 \\ -22 \\ -9 \end{bmatrix}$

$$\begin{bmatrix} 3 & 6 & 9 \\ 0 & 2 & -8 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 39 \\ -22 \\ -9 \end{bmatrix}$$

$$x_3 = -9 / (-3) = 3$$

$$x_2 = [-22 - (8) \cdot 3] / 2 = 1$$

$$x_1 = [39 - 6 \cdot 1 - 9 \cdot 3] / 3 = 2$$

Rotina para eliminação

de $k=1$ até $n-1$ faça
 determine o índice do pivot l
 troque as linhas k e l , colunas k até n
 $p(k) = l$

loop $\left[\begin{array}{l} \text{determine os multiplicadores} \\ \text{guarde os multiplicadores onde os elementos eliminados irão} \\ \text{adicione múltiplos da linha } k \text{ às linhas} \end{array} \right.$ remanescentes

— fim da etapa

f) Determinante de uma matriz A

Pelas propriedades do determinante, o determinante não se altera se somarmos um múltiplo de uma linha à outra. Assim a eliminação de Gauss até atingir a matriz triangular superior não afeta o valor do determinante

$$\det \underbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}}_A = \det \underbrace{\begin{bmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ \vdots & 0 & & \vdots \\ 0 & 0 & 0 & a'_{nn} \end{bmatrix}}_U$$

Mas $\det U = a'_{11} a'_{22} \dots a'_{nn}$

$$\Rightarrow \det A = \det U = a'_{11} a'_{22} \dots a'_{nn}$$

produto dos elementos da diagonal principal da matriz triangular superior.

(*) Cada operação de pivotamento troca o sinal do determinante!

Ex: Calcule o determinante da matriz abaixo usando o método de Gauss

Solução

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 4 & 3 \\ 6 & 7 & 4 \end{bmatrix}$$

Solução usando a eliminação de Gauss p/ frente termos

$$U = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ & & -1 \end{bmatrix} \rightarrow \det A = \det U = 2 \cdot 2 \cdot (-1) = \boxed{-4}$$

2.2 Método de Gauss-Jordan

É uma variante do método de Gauss, só que são eliminados todos os elementos acima e abaixo do pivô. O pivotamento é total com trocas de linhas e colunas de forma que o pivô seja o maior em valor absoluto. Tem se mostrado vantajoso quando aplicado em cálculos paralelos.

2.3 Decomposição LU

Suponha que possamos escrever a matriz A como o produto de duas matrizes

$$A = L \cdot U$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & l_{n,n-1} & \dots & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ & u_{22} & \dots & u_{2n} \\ & 0 & \ddots & \\ & & & u_{nn} \end{bmatrix}}_U$$

triangular inferior triangular superior

$$Ax = (L \cdot U)x = L(\underbrace{Ux}_y) = b$$

1º Resolvemos $Ly = b$ e depois $Ux = y$

$Ly = b$ pode ser resolvido por substituição pora frente

$$y_1 = b_1$$

$$y_i = b_i - \sum_{j=1}^{i-1} l_{ij} y_j, \quad i = 2, 3, \dots, n$$

$Ux = y$ pode ser resolvido por substituição p/ trás

$$x_n = \frac{y_n}{u_{nn}}$$

$$x_i = \frac{1}{u_{ii}} \left[y_i - \sum_{j=i+1}^n u_{ij} x_j \right] \quad i = n-1, n-2, \dots, 1$$

— Algoritmos p/ realização da decomposição LU são mais elaborados quando incluem pivotamento e podem ser encontrados por exemplo no Numerical Recipes.

Veremos um caso mais simples de matriz tridiagonal

Suponha que A seja uma matriz na forma

$$A = \begin{bmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & a_n & b_n \end{bmatrix}$$

a decomposição LU então tem uma forma simples

$$L = \begin{bmatrix} 1 & & & & \\ l_2 & 1 & & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & \ddots & l_n & 1 \end{bmatrix}; \quad U = \begin{bmatrix} u_1 & v_1 & & & \\ & u_2 & v_2 & & \\ & & \ddots & \ddots & \\ & & & v_{n-1} & \\ & & & & u_n \end{bmatrix}$$

Exercício se A é tridiagonal mostre que A pode ser decomposta e

LU sendo

$$u_1 = b_1, \quad v_i = c_i$$

$$l_j u_{j-1} = a_j$$

$$l_j v_{j-1} + u_j = b_j, \quad j = 2, 3, \dots, n$$

Determinados L e U procede-se como no caso geral, i.e. p/ encontrar solução $Ax = d \rightarrow LUx = d \Rightarrow Ly = d; Ux = y$ resolvidos por forward e backward substitution. $\mathcal{O}(n)$ operações.

$$\begin{cases} u_1 = b_1 \\ l_j = a_j / u_{j-1} \\ u_j = b_j - l_j c_{j-1}, \quad j = 2, 3, \dots, n \end{cases}$$

Decomposição
LU

$$\begin{cases} z_1 = d_1 \\ z_i = d_i - l_i z_{i-1}, \quad i = 2, \dots, n \end{cases}$$

"Forward
Substitution"

$$\begin{cases} x_n = \frac{z_n}{u_n} \\ x_k = z_k - c_k \frac{x_{k+1}}{u_k}, \quad k = n-1, \dots, 1 \end{cases}$$

"Backward
Substitution"

$$\begin{bmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & a_n & b_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

Comparação do Nº de operações

Soluções de
sistemas lineares

Inversão
de Matriz

m
lados
direitos

método	multiplicações (ou adições)
gauss	$\frac{1}{3} n^3$
gauss-Jordan	$\frac{1}{2} n^3$
LU	$\frac{1}{3} n^3$
gauss	n^3
gauss-Jordan	n^3
LU	n^3
gauss	$\frac{1}{3} n^3 + \frac{1}{2} mn^2$
gauss-Jordan	$\frac{1}{2} n^3 + mn^2$
LU	$\frac{1}{3} n^3 + \frac{1}{2} mn^2$

2.4 Refinamento

Seja o sistema $A\vec{x} = \vec{b}$ (1)

Resolvendo por gauss obtemos a solução $\vec{x}^{(0)}$

Vamos chamar de erro a diferença entre o valor verdadeiro e o valor obtido $\rightarrow \vec{e}^{(0)} = \vec{x} - \vec{x}^{(0)}$

Substituindo no sistema (1) temos

$$A(\vec{x}^{(0)} + \vec{e}^{(0)}) = \vec{b}$$

$$\text{ou } A\vec{e}^{(0)} = \underbrace{\vec{b} - A\vec{x}^{(0)}}_{\equiv \vec{r} \equiv (\text{resíduos})} \quad (2)$$

Resolvendo o sistema (2) determinamos $\vec{e}^{(0)}$ então $\vec{x} = \vec{x}^{(0)} + \vec{e}^{(0)}$ deveria ser a solução correta mas ainda pode conter erros e podemos chamá-la de $\vec{x}^{(1)} \equiv \vec{x}^{(0)} + \vec{e}^{(0)}$

Calcula-se a seguir $\vec{e}^{(1)} \equiv \vec{x} - \vec{x}^{(1)}$ resolvendo

$$A\vec{e}^{(1)} = \vec{b} - A\vec{x}^{(1)} \quad \text{obtendo-se } \vec{x}^{(2)} = \vec{x}^{(1)} + \vec{e}^{(1)} \quad \text{e}$$

assim por diante.

2.5 Sistemas mal-condicionados ("ill conditioned")

Exemplo:
$$\begin{cases} x + y = 1 \\ 99x + 100y = 99.5 \end{cases}$$

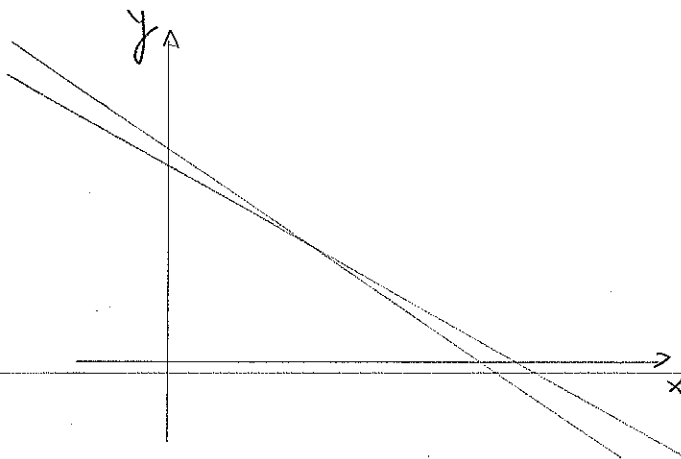
cuja solução única e exata é $x = 0.5$, $y = 0.5$

Agora considere o sistema

$$\begin{cases} x + y = 1 \\ 99.4x + 99.9y = 99.2 \end{cases}$$

cuja solução única e exata é $x = 1.4$
 $y = -0.4$

Desenhe as retas e veja porque isto acontece



→ retas correspondentes às eq. são quase paralelas
≡ quase linearmente dependentes.

Pensando nos vetores linha (ou vetores coluna)

$\vec{a}_1, \vec{a}_2, \dots ; j = 1 \dots n$, se eles forem quase linearmente dependentes então o sistema será mal-condicionado.

Uma maneira de "medir" o condicionamento de uma matriz é calcular seu determinante. Mas o valor do determinante depende da norma de cada um dos vetores. Assim normalizamos os vetores para depois ^{em módulo} calcular o determinante. Isto produzirá um valor entre zero e 1. Se o módulo do determinante for próximo de zero então o sistema é mal-condicionado.

ex.:
$$\begin{bmatrix} 1 & 1 \\ 99 & 100 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 99.5 \end{bmatrix}$$

normalização
dos
vetores
linha

$$\rightarrow \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{99}{140.716} & \frac{100}{140.716} \end{bmatrix}$$

Cálculo do módulo do determinante

$$\left| \frac{1}{\sqrt{2}} \left(\frac{100}{140.716} \right) - \frac{1}{\sqrt{2}} \left(\frac{99}{140.716} \right) \right| = \frac{1}{\sqrt{2}} 7.1 \times 10^{-3} \approx 5 \times 10^{-3}$$

→ próximo de zero
→ vetores quase paralelos

Há outras medidas do condicionamento de uma matriz assim como há fórmulas que relacionam o erro cometido no método de Gauss com essas medidas e com o n° de algarismos significativos. Porém isso vale além dessas notas. Veja por exemplo:

- Blum (1972)
- Numerical Recipes (1992) - "SVD"
"Singular Value Decomposition"

③ Métodos Iterativos (ou métodos indiretos)

— O método de eliminação de Gauss se torna ineficiente quando o n° de equações atinge $n \geq 100$, pois o n° de operações de P.F. é $O(n^3)$. Mais detalhes Blum (1972) p.131.

— Métodos iterativos: arbitra-se um vetor inicial $\vec{x}^{(0)}$ e calcula-se um vetor $\vec{x}^{(1)}$ como função de $\vec{x}^{(0)}$ e assim sucessivamente,

$$\vec{x}^{k+1} = g(\vec{x}^k) \text{ onde } k \text{ é a } k\text{-ésima iteração}$$

até obter uma precisão desejada que seja uma diferença muito pequena entre \vec{x}^{k+1} e \vec{x}^k

3.1 Método de Jacobi

Dado o sistema linear $Ax = b$ na forma

$$\begin{array}{ccccccc} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n} & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n} & = & b_2 \\ \vdots & & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn} & = & b_n \end{array}$$

Podemos reescrever o sistema na forma

$$x_1 = \frac{1}{a_{11}} (b_1 - a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n)$$

$$x_2 = \frac{1}{a_{22}} (b_2 - a_{21}x_1 - a_{23}x_3 - \dots - a_{2n}x_n)$$

\vdots

$$x_n = \frac{1}{a_{nn}} (b_n - a_{n1}x_1 - a_{n2}x_2 - \dots - a_{n,n-1}x_{n-1})$$

— Escolhemos arbitrariamente um vetor inicial $x_i^{(0)}$ e substituímos no lado direito das equações acima obtendo um novo vetor $x_i^{(1)}$ e assim por diante

$$x_1^{k+1} = \frac{1}{a_{11}} (b_1 - a_{12}x_2^k - \dots - a_{1n}x_n^k)$$

$$x_2^{k+1} = \frac{1}{a_{22}} (b_2 - a_{21}x_1^k - \dots - a_{2n}x_n^k)$$

\vdots

$$x_n^{k+1} = \frac{1}{a_{nn}} (b_n - a_{n1}x_1^k - \dots - a_{n,n-1}x_{n-1}^k)$$

ou de forma mais compacta

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

Exemplo: Considere o seguinte sistema de equações

$$\begin{aligned} 4x_1 + 2x_2 + x_3 &= 11 \\ -x_1 + 2x_2 &= 3 \\ 2x_1 + x_2 + 4x_3 &= 16 \end{aligned} \quad (1) \quad \text{cuja solução é } \vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Reescrevendo as equações como

$$x_1 = \frac{11}{4} - \frac{1}{2}x_2 - \frac{x_3}{4} \quad (2)$$

$$x_2 = \frac{3}{2} + \frac{1}{2}x_1$$

$$x_3 = 4 - \frac{1}{2}x_1 - \frac{1}{4}x_2,$$

o que sugere o método iterativo

$$x_1^{k+1} = \frac{11}{4} - \frac{1}{2}x_2^k - \frac{1}{4}x_3^k$$

$$x_2^{k+1} = \frac{3}{2} + \frac{1}{2}x_1^k \quad (3)$$

$$x_3^{k+1} = 4 - \frac{1}{2}x_1^k - \frac{1}{4}x_2^k$$

Começando com um vetor arbitrário $\vec{x}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ temos

$$x_1^{(1)} = \frac{11}{4} - \frac{1}{2} \cdot 1 - \frac{1}{4} \cdot 1 = 2$$

$$x_2^{(1)} = \frac{3}{2} + \frac{1}{2} \cdot 1 = 2$$

$$x_3^{(1)} = 4 - \frac{1}{2} \cdot 1 - \frac{1}{4} \cdot 1 = \frac{13}{4}$$

Substituindo $\vec{x}^{(1)}$ do lado direito do sistema (3) obtemos

$$x_1^{(2)} = \frac{11}{4} - \frac{1}{2}(2) - \frac{1}{4}\left(\frac{13}{4}\right) = \frac{44 - 16 - 13}{16} = \frac{15}{16}$$

$$x_2^{(2)} = \frac{3}{2} + \frac{1}{2}(2) = \frac{5}{2}$$

$$x_3^{(2)} = 4 - \frac{1}{2}(2) - \frac{1}{4}(2) = \frac{5}{2}$$

Podemos colocar os resultados em forma de tabela:

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\max\{ x_i^k - x_i^{k-1} , i=1 \dots n\}$
0	1	1	1	—
1	2	2	$13/4$	$9/4$
2	$15/16$	$5/2$	$5/2$	$3/4$
3	$7/8$	$63/32$	$93/32$	$17/32$
4	$133/128$	$31/16$	$393/128$	$21/128$
5	$519/512$	$517/256$	$767/256$	$21/256$

dentro de 1%
da solução exata

Convergência do Método de Jacobi

Vamos escrever a matriz A como

$$A = L + D + U, \text{ onde}$$

$L \rightarrow$ "lower triangular matrix" (sem diagonal)

$U \rightarrow$ "upper triangular matrix" (sem diagonal)

$D \rightarrow$ diagonal

$$A\vec{x} = (L + D + U)\vec{x} = \vec{b}$$

$$D\vec{x} = -(L + U)\vec{x} + \vec{b}$$

$$\vec{x} = D^{-1} [-(L + U)\vec{x} + \vec{b}]$$

$$\vec{x} = J\vec{x} + \vec{c}, \text{ onde } J \equiv -D^{-1}(L + U)$$

$$\vec{c} \equiv D^{-1}\vec{b}$$

aplicando o método iterativo teremos

$$\vec{x}^{(k+1)} = J\vec{x}^{(k)} + \vec{c} \quad \text{onde } J = \begin{bmatrix} 0 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} & \dots & \frac{a_{1n}}{a_{11}} \\ \frac{a_{21}}{a_{22}} & 0 & \frac{a_{23}}{a_{22}} & \dots & \frac{a_{2n}}{a_{22}} \\ \vdots & & 0 & & \vdots \\ \frac{a_{n1}}{a_{nn}} & \frac{a_{n2}}{a_{nn}} & \dots & \frac{a_{n,n-1}}{a_{nn}} & 0 \end{bmatrix}$$

Partindo de $\vec{x}^{(0)}$ e fazendo sucessivamente a iteração temos

$$\vec{x}^{(k)} = J^k \vec{x}^{(0)} + \left[\mathbf{1} + J + J^2 + \dots + J^{k-1} \right] \vec{c}$$

Para que convirja requer que

$$\lim_{k \rightarrow \infty} J^k = [0] \quad (I)$$

O que implica que $\lim_{k \rightarrow \infty} [\mathbf{1} + J + J^2 + \dots + J^k] = (\mathbf{1} - J)^{-1}$

Assim quando (I) é satisfeita, $\vec{x} = \lim_{k \rightarrow \infty} \vec{x}^{(k)}$ existe e

$$\vec{x} = 0 + (\mathbf{1} - J)^{-1} \vec{c}, \text{ i.e. } (\mathbf{1} - J)\vec{x} = \vec{c} \text{ ou } \vec{x} = J\vec{x} + \vec{c}.$$

Mas a condição (I) é válida se e somente se todos os autovalores da matriz T forem em módulo < 1 .

Seja $\rho_s = \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}$ ou λ_i são os autovalores da matriz T .
↳ raio espectral "spectral radius"

Então para atingir precisão p após k iterações devemos ter

$$\rho_s^k \approx 10^{-p} \rightarrow \boxed{k \approx -\frac{p \ln 10}{\ln \rho_s}}$$

Assim se ρ_s estiver próximo de 1 a convergência será muito lenta. Existem métodos de aceleração:

Ver: - Quinney

- Numerical Recipes seção 19.5

- Determinar os autovalores da matriz T requerirá outro algoritmo em geral. Na prática muitas vezes é mais fácil testar numericamente a convergência.

- Uma condição⁺ simples porém apenas suficiente que assegura a convergência do método é que o sistema possua diagonal principal estritamente dominante ou seja

$$|a_{ii}| > \sum_j |a_{ij}|, \quad i \neq j$$

que é chamado de critério das linhas.

Exercício Mostre que o sistema

$$\begin{bmatrix} 4 & 2 & 1 \\ -1 & 2 & 0 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 11 \\ 3 \\ 16 \end{bmatrix}$$

satisfaz o critério das linhas.

- Note que um sistema que não satisfaz o critério das linhas pode convergir.
- Alterando a ordem das linhas ou colunas pode tornar o sistema convergente em divergente e vice-versa.

3.2 Método de Gauss-Seidel

— É muito semelhante ao método de Jacobi, só que para calcular x_i^{k+1} aproveita-se os valores já calculados $x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}$

— As equações tomam a seguinte forma

$$x_1^{k+1} = \frac{1}{a_{11}} (b_1 - a_{12}x_2^k - \dots - a_{1n}x_n^k)$$

$$x_2^{k+1} = \frac{1}{a_{22}} (b_2 - a_{21}x_1^{k+1} - a_{23}x_3^k - \dots - a_{2n}x_n^k)$$

$$x_3^{k+1} = \frac{1}{a_{33}} (b_3 - a_{31}x_1^{k+1} - a_{32}x_2^{k+1} - a_{34}x_4^k - \dots - a_{3n}x_n^k)$$

⋮

$$x_n^{k+1} = \frac{1}{a_{nn}} (b_n - a_{n1}x_1^{k+1} - \dots - \dots - a_{n,n-1}x_{n-1}^{k+1})$$

— Em forma mais compacta

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

— Num programa de computador pode-se escrever

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

Exemplo: $x_1 = \frac{11}{4} - \frac{1}{2}x_2 - \frac{1}{4}x_3$

$$x_2 = \frac{3}{2} + \frac{1}{2}x_1$$

$$x_3 = 4 - \frac{1}{2}x_1 - \frac{1}{4}x_2$$

Vamos
escolher
novamente

$$\vec{x}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$x_1^{(1)} = \frac{11}{4} - \frac{1}{2}(1) - \frac{1}{4}(1) = 2$$

$$x_2^{(1)} = \frac{3}{2} + \frac{1}{2}(2) = \frac{5}{2}$$

$$\rightarrow \vec{x}^{(1)} = \begin{bmatrix} 2 \\ 5/2 \\ 19/8 \end{bmatrix}$$

$$x_3^{(1)} = 4 - \frac{1}{2}(2) - \frac{1}{4}\left(\frac{5}{2}\right) = \frac{19}{8}$$

e sucessivamente obtemos

$$\vec{x}^{(2)} = \begin{bmatrix} 29/32 \\ 125/64 \\ 783/256 \end{bmatrix}$$

$$\vec{x}^{(3)} = \begin{bmatrix} 1033/1024 \\ 4095/2048 \\ 24541/8192 \end{bmatrix} \approx \begin{bmatrix} 1.0087 \\ 1.9995 \\ 2.9957 \end{bmatrix}$$

Note que neste exemplo a taxa de convergência é muito maior que no método de Jacobi.

Convergência

$$A \vec{x} = \vec{b}, \quad A = \mathbb{L} + D + U$$

$$(\mathbb{L} + D + U) \vec{x} = \vec{b}$$

$$D \vec{x} = -(\mathbb{L} + U) \vec{x} + \vec{b}$$

$$\vec{x} = D^{-1} [-(\mathbb{L} + U) \vec{x} + \vec{b}]$$

$$\vec{x}^{k+1} = -D^{-1} \mathbb{L} \vec{x}^{k+1} - D^{-1} U \vec{x}^k + D^{-1} \vec{b}$$

$$(\mathbb{I} + D^{-1} \mathbb{L}) \vec{x}^{k+1} = -D^{-1} U \vec{x}^k + D^{-1} \vec{b}$$

$$\vec{x}^{k+1} = \underbrace{(\mathbb{I} + D^{-1} \mathbb{L})^{-1} (-D^{-1} U)}_{\equiv G} \vec{x}^k + \underbrace{(\mathbb{I} + D^{-1} \mathbb{L})^{-1} D^{-1} \vec{b}}_{\equiv \vec{C}}$$

ou

$$\vec{x}^{k+1} = G \vec{x}^k + \vec{C}$$

— Esta última expressão está na mesma forma que a estudada no método de Jacobi, portanto uma condição necessária e suficiente para a convergência é que os autovalores de G devem ser todos em módulo menores que 1.

Ou seja, $\det [G - \lambda \mathbb{I}] = 0$, $|\lambda| < 1$, todo i .

$$\det [(\mathbb{I} + D^{-1} \mathbb{L})^{-1} [-D^{-1} U - \lambda (\mathbb{I} + D^{-1} \mathbb{L})]]$$

$$\Rightarrow \det [-D^{-1} U - \lambda (\mathbb{I} + D^{-1} \mathbb{L})] = 0 //$$

— A convergência (ou não) tanto no método de Jacobi como no de Gauss-Seidel independem do vetor inicial escolhido

— o critério das Linhas também pode ser aplicado ao método de Gauss-Seidel, também como condição apenas suficiente.

→ Para o método de Gauss-Seidel existe um outro critério menos restritivo que o critério das Linhas, chamado de critério de Sassenfeld.

O critério de Sassenfeld afirma que se definirmos β_i 's de forma que

$$\beta_1 = \frac{|a_{12}| + |a_{13}| + \dots + |a_{1n}|}{|a_{11}|},$$

$$\beta_i = \frac{\sum_{j=1}^{i-1} \beta_j |a_{ij}| + \sum_{j=i+1}^n |a_{ij}|}{|a_{ii}|},$$

então o método de Gauss-Seidel converge se $\beta_i < 1$
p/ todo $i=1, \dots, n$

Exercício: Seja a matriz $A\vec{x} = \vec{b}$ com

$$A = \begin{pmatrix} -1 & -5 & 4 \\ -6 & -2 & 8 \\ 3 & -1 & 1 \end{pmatrix} \quad \text{Verifique se a matriz } A \text{ satisfaz o critério das Linhas ou Sassenfeld por alguma permutação de linhas.}$$

Solução

$$\begin{pmatrix} 3 & -1 & 1 \\ -1 & 5 & 4 \\ -6 & -2 & 8 \end{pmatrix} \rightarrow \text{não satisfaz o critério das Linhas} \\ \rightarrow \text{não sabemos se os métodos de Jacobi ou Gauss-Seidel convergem por esse critério.}$$

$$\beta_1 = \frac{|-1| + 1}{3} = \frac{2}{3}$$

$$\beta_2 = \frac{\frac{2}{3}|-1| + 4}{5} = \frac{14}{15}$$

$$\beta_3 = \frac{\frac{2}{3}|-6| + \frac{14}{15}|-2|}{8} = \frac{4 + \frac{28}{15}}{8} = \frac{88}{120} = \frac{11}{15}$$

$\beta_1, \beta_2, \beta_3 < 1 \rightarrow$ satisfaz Sassenfeld \rightarrow Gauss-Seidel converge

③ Sistemas de Equações não-lineares

Seja $F(\vec{x}) = 0$ $\vec{x} = \{x_1, x_2, \dots, x_n\}$

$$F(x) = \{f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x})\}$$

3.1 — Uma possível solução pode ser obtida como no método de Newton-Raphson.

— Vamos tomar \vec{x}_0 como valor inicial

— A série de Taylor para vetores é

$$F(x) = F(\vec{x}_0) + J(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \dots + \dots$$

onde J é a matriz Jacobiana

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

— A próxima aproximação \vec{x}_1 para a solução de $F(\vec{x}) = 0$ é dada por $F(\vec{x}_0) + J^{(0)}(\vec{x}_1 - \vec{x}_0) = 0$, $J_0 \equiv J(\vec{x}_0)$ ou

$J_0 \Delta \vec{x}_0 = -F(\vec{x}_0)$ que é um sistema linear que pode ser resolvido por eliminação de Gauss e

$$\vec{x}_1 = \vec{x}_0 + \Delta \vec{x}_0$$

— E assim para \vec{x}_{k+1} temos

$$J_k \Delta \vec{x}_k = -F(\vec{x}_k) \quad \text{e} \quad \vec{x}_{k+1} = \vec{x}_k + \Delta \vec{x}_k$$

Exemplo

$$\begin{cases} x_1^2 + x_2^2 + x_3^2 - 1 = 0 \\ 2x_1^2 + x_2^2 - 4x_3 = 0 \\ 3x_1^2 - 4x_2 + x_3^2 = 0 \end{cases}$$

cujas soluções é $\vec{x} = \begin{bmatrix} 0.7852 \\ 0.4966 \\ 0.3699 \end{bmatrix}$

$$J = \begin{bmatrix} 2x_1 & 2x_2 & 2x_3 \\ 4x_1 & 2x_2 & -4 \\ 6x_1 & -4 & 2x_3 \end{bmatrix}$$

$$\vec{x}_0 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}, \quad J_0 = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -4 \\ 3 & -4 & 1 \end{bmatrix}, \quad -F(\vec{x}_0) = \begin{bmatrix} 0.25 \\ 1.25 \\ 1.00 \end{bmatrix}$$

$$J_0(\Delta\vec{x}_0) = -F(\vec{x}_0), \quad \text{onde } \Delta\vec{x}_0 = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{bmatrix}$$

$$\begin{cases} \Delta x_1 + \Delta x_2 + \Delta x_3 = 0.25 \\ 2\Delta x_1 + \Delta x_2 - 4\Delta x_3 = 1.25 \\ 3\Delta x_1 - 4\Delta x_2 + \Delta x_3 = 1.00 \end{cases}$$

Resolvendo por Gauss temos

$$\Delta x_1 = 0.375, \quad \Delta x_2 = 0, \quad \Delta x_3 = -0.125$$

$$\vec{x}_1 = \vec{x}_0 + \Delta\vec{x}_0 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0.375 \\ 0 \\ -0.125 \end{bmatrix} = \begin{bmatrix} 0.875 \\ 0.500 \\ 0.375 \end{bmatrix}$$

e assim sucessivamente.

3.2 MÉTODO DE BROYDEN

O método de Newton para resolver o sistema de equações não lineares $F(\vec{x})=0$ usa o Jacobiano a cada iteração. Mas o cálculo do Jacobiano pode ser muito complicado e caro ($O(n^3)$).

O método de Broyden é uma generalização do método de secantes para múltiplas equações. No método de secantes a derivada é estimada como $f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$.

O Jacobiano poderia então ser estimado usando

$J_k (\vec{x}_k - \vec{x}_{k-1}) \approx F(\vec{x}_k) - F(\vec{x}_{k-1})$.⁽¹⁾ Mas essa equação não determina unicamente J_k e são necessárias restrições.

Broyden sugere usar uma estimativa do Jacobiano J_{k-1} e estimar J_k como solução da eq. (1) que produza a mínima modificação de J_{k-1} . (mínima no sentido da norma de Frobenius $\|J_k - J_{k-1}\|_F$, $\|A\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N |a_{ij}|^2}$)

Assim ele obteve a expressão

$$J_k = J_{k-1} + \frac{\Delta F_k - J_{k-1} \Delta \vec{x}_k}{\|\Delta \vec{x}_k\|^2} \Delta \vec{x}_k^T \quad (2) \quad \begin{array}{l} \text{"good Broyden"} \\ \text{method"} \end{array}$$

$\Delta F_k \equiv F(x_{k+1}) - F(x_k)$

e então prossegue-se a iteração com método de Newton

$$J_k \Delta \vec{x}_k = -F(\vec{x}_k) \quad \text{ou} \quad \vec{x}_{k+1} = \vec{x}_k - J^{-1} F(\vec{x}_k)$$

Broyden também sugeriu usar a fórmula de Sherman-MORRISON[†] para atualizar diretamente o Jacobiano invertido

$$J_k^{-1} = J_{k-1}^{-1} + \frac{\Delta \vec{x}_k - J_{k-1}^{-1} \Delta F_k}{\Delta \vec{x}_k^T J_{k-1}^{-1} \Delta F_k} (\Delta \vec{x}_k^T J_{k-1}^{-1})$$

Alternativamente pode-se usar uma atualização diferente para J_k^{-1} que minimize $\|J_k^{-1} - J_{k-1}^{-1}\|_F$ obtendo

$$J_k^{-1} = J_{k-1}^{-1} + \frac{\Delta x_k - J_{k-1}^{-1} \Delta F_k}{\Delta F_k^T \Delta F_k} \Delta F_k^T$$

Vamos voltar ao nosso exemplo

$$f_1(\vec{x}) = x_1^2 + x_2^2 + x_3^2 - 1 = 0$$

$$f_2(\vec{x}) = 2x_1^2 + x_2^2 - 4x_3 = 0$$

$$f_3(\vec{x}) = 3x_1^2 - 4x_2 + x_3^2 = 0$$

uma solução é $\vec{x} = \begin{bmatrix} 0.7852 \\ 0.4966 \\ 0.3699 \end{bmatrix}$

Vamos iniciar com um chute inicial $\vec{x}_0 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$

Os elementos da matriz Jacobiana são dados por $J_{i,j} = \frac{\partial f_i}{\partial x_j}$ e podem ser estimados na 1ª iteração como

$$J_{i,j} = \frac{f_i(x_j + h_j) - f_i(x_j)}{(x_j + h_j) - x_j} \quad \text{onde } h_j \text{ é um pequeno valor}$$

Usando $h_j = 0.001$, $j = 1, \dots, n$ obtemos a 1ª estimativa do Jacobiano

$$J_0 = \begin{bmatrix} 1.001 & 1.001 & 1.001 \\ 2.002 & 1.001 & -4 \\ 3.003 & -4 & 1.001 \end{bmatrix}$$

e seguimos usando o método de Newton

$$J_0 \Delta \vec{x}_0 = -F(\vec{x}_0), \quad F(\vec{x}_0) = \begin{bmatrix} 0.25 \\ 1.25 \\ 1 \end{bmatrix}$$

Resolvendo por Eliminação de Gauss obtemos

$$\Delta \vec{x}_0 = \begin{bmatrix} 0.37469 \\ 0.00002 \\ -0.12496 \end{bmatrix} \Rightarrow \vec{x}_1 = \vec{x}_0 + \Delta \vec{x}_0 = \begin{bmatrix} 0.87469 \\ 0.50002 \\ 0.37504 \end{bmatrix}$$

Agora podemos estimar J_1 usando a expressão (2)

$$J_1 = J_0 + \frac{F(\vec{x}_1) - F(\vec{x}_0) - J_0 \Delta \vec{x}_0}{\|\Delta \vec{x}_0\|^2}$$

$$= \begin{bmatrix} 1.37509 & 1.00102 & -0.87624 \\ 2.67457 & 1.00104 & -4.22431 \\ 4.04965 & -3.9993 & -0.65193 \end{bmatrix}$$

e prosseguimos com

$$J_1 \Delta \vec{x}_1 = -F(\vec{x}_1)$$

$$\text{obtendo } \Delta \vec{x}_1 = \begin{bmatrix} -0.1935 \\ -0.00233 \\ -0.00349 \end{bmatrix} \Rightarrow \vec{x}_2 = \vec{x}_1 + \Delta \vec{x}_1 = \begin{bmatrix} 0.76534 \\ 0.49770 \\ 0.37154 \end{bmatrix}$$

e usando novamente a expressão (2) obtemos J_2 e assim sucessivamente.

Referências: Broyden, C.G. "A class of Methods for solving Nonlinear Simultaneous Equations", Mathematics of Computation (American Mathematical Society) 19(92), 577-593 (1965)

Wikipedia

BURDEN FAIRES cap.10.

BROYDEN, C.G., "On the discovery of the "good Broyden" method, Math. Program. Ser. B 87: 209-213 (2000).