Capítulo 3. Métodos Numéricos Iterativos

1. Métodos numéricos

Sempre que se pretende resolver um problema cuja solução é um valor numérico, é habitual ter de se considerar, para além de conceitos mais abstratos (que fornecem um modelo consistente para a análise do problema), conceitos de natureza mais prática relacionados com os cálculos a efetuar ou com os números necessários à realização de tais cálculos.

Por exemplo, suponha-se que se pretende determinar o volume V de um paralelepípedo a partir dos comprimentos das três arestas que o definem A, B e C. O modelo abstrato consiste na expressão matemática V = A.B.C, que permite calcular o volume a partir dos comprimentos daquelas arestas. No entanto, para se aplicar aquela expressão, é necessário antes fazer a medição de cada uma das três arestas. Ora, a medição de cada uma das arestas está associada um erro (erros iniciais). Desta forma, o processo de medição fornece apenas valores aproximados dos comprimentos das arestas, sendo eventualmente possível obter-se uma caracterização dos erros iniciais (vindos das medições). Ao efetuar-se, de seguida, o cálculo do volume (V = A.B.C), o resultado será um valor que apenas poderá ser considerado uma aproximação do volume exato do paralelepípedo, o qual terá associado um erro que dependerá dos erros cometidos no processo de medição ocorrido no início.

A situação descrita neste exemplo, de não se conseguir obter um valor numérico exato como solução de um problema, é a mais comum para muitos deles. No entanto, esta situação não é necessariamente má, pois para a grande maioria dos problemas bastará apenas obter um valor numérico suficientemente próximo do valor exato.

De uma forma simples, os métodos numéricos conduzem com eficiência a soluções aproximadas de um modelo matemático (e de um sistema real). Os modelos matemáticos usam os métodos numéricos para a obtenção de soluções numéricas para problemas reais quando, por uma qualquer razão, não se pode ou não se deseja usar métodos analíticos.

2. Métodos analíticos versus métodos numéricos

Um **método analítico** para resolver um modelo matemático é qualquer método baseado na análise matemática rigorosa, cuja aplicação conduz a uma ***solução verdadeira*** (***exata***), também conhecida como ***solução analítica*** do modelo.

Por sua vez, um **método numérico** para resolver um modelo matemático é qualquer método baseado numa análise matemática rigorosa, cuja aplicação, na grande maioria dos casos, simplesmente conduz a uma ***solução aproximada*** (não exata), também conhecida como ***solução numérica***. Em alguns casos (raros) pode-se obter uma solução exata, usando um método numérico.

Por exemplo, as soluções exatas da equação não linear $x^2 - 5x + 3 = 0$ podem ser obtidas usando a conhecida fórmula quadrática (método analítico):

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Esta fórmula dá uma solução analítica

$$x_{1,2} = \frac{5 \pm \sqrt{13}}{2}.$$

Por sua vez, a fórmula de iteração (método numérico)

$$x_{n+1} = \sqrt{5x_n - 3}, \quad n = 0,1,2,\ldots; \quad x_0 = 4.5$$

pode também ser aplicada para aproximar uma das duas soluções da equação quadrática dada. Este método pode somente dar uma solução numérica (aproximada).

Em geral a diferença entre soluções analíticas e soluções numéricas pode ser resumida na seguinte frase: *soluções analíticas são exatas enquanto soluções numéricas são aproximadas*.

3. Necessidade de se usar métodos numéricos

Porque é que alguém usaria métodos numéricos? Os métodos numéricos são necessários?

A partir da distinção apresentada entre métodos analíticos e métodos numéricos, facilmente pode-se ser levado a concluir que é suficiente usar métodos analíticos na resolução de modelos matemáticos. Por outras palavras, que não há necessidade de usar métodos numéricos pois eles conduzem somente a soluções aproximadas. Tal conclusão é enganadora.

Precisamos de usar métodos numéricos pelas seguintes razões:

- ➔ Porque existem situações em que é preferível um método numérico ao método analítico, ainda que este exista; por exemplo, quando a solução para um problema envolve vários cálculos, os quais podem ser muito demorados.
- ➔ Porque a maioria dos problemas reais são, em geral, complexos e envolvem fenómenos não lineares, pelo que é comum os conhecimentos de matemática não serem suficientes para a descoberta de uma solução para um problema real.

➔ Porque quando os dados do problema requerem um tratamento que inclua, por exemplo, diferenciação ou integração, terá de ser feito através de um método numérico.

Uma vez que, em geral, o modelo matemático é demasiado complexo para ser tratado analiticamente, deve-se construir modelos aproximados ou obter soluções aproximadas.

No primeiro caso, implica alterar e simplificar o modelo por forma a torná-lo tratável, e assim obter uma solução exata de um sistema ou modelo aproximado. No entanto, tal solução é "suspeita" pelo facto de ocorrerem simplificações do modelo, pelo que terão que fazer várias experiências para ver se as simplificações são compatíveis com os dados experimentais.

No segundo caso, implica usar métodos numéricos e assim produzir soluções aproximadas para o sistema real. No entanto, tais soluções são apenas aproximações, mas que podem ser melhoradas à custa de esforço computacional.

4. Métodos iterativos

Os métodos iterativos estão associados aos conceitos de **iteração** (ou **aproximação sucessiva**) e de **aproximação local**. Em sentido lato, **iteração** significa a repetição sucessiva de um processo. A **aproximação local** consiste em aproximar uma função por outra que seja de manuseio mais simples. Por exemplo, aproximar uma função não linear por uma função linear num determinado intervalo do domínio das funções. Um método iterativo caracteriza-se por envolver os seguintes elementos: ***aproximação inicial***, que consiste numa primeira aproximação para a solução do problema numérico, e ***teste de paragem***, que é o instrumento por meio do qual o procedimento iterativo é finalizado.

Define-se *sequência* de números, $\{ x_k \}_{k = 1, 2, \ldots}$, como sendo uma transformação do conjunto dos inteiros positivos no conjunto dos reais. O número real associado a *k* é designado por x_k.

A sequência resultante

$$x_k = f\left(x_{k-1}, \ldots\right)$$

chama-se *sequência iterativa* gerada por *f*.

Esta sequência diz-se definida por *iteração* se a função *f* é independente de *k*.

Este processo iterativo gera uma sucessão de aproximações x_k, cada uma com ***erro*** associado,

$$e_k = \alpha - x_k$$

sendo α um ***zero*** da função f; isto é, $f(\alpha) = 0$.

A sequência iterativa diz-se ***convergente*** se

$$\lim_{k \to \infty} x_k = \alpha$$

ou seja,

$$\lim_{k \to \infty} e_k = 0.$$

Um método iterativo é definido por uma equação iterativa, com a qual se constrói aproximações à solução do problema. A implementação da equação iterativa obriga ao conhecimento de uma aproximação inicial e à definição de um conjunto de condições que garantam que a aproximação calculada, numa certa iteração, se encontra suficientemente próxima da solução. Quando estas condições forem verificadas, pode-se ***parar o processo***. Desta forma, antes de se iniciar o processo iterativo, deve-se ter resposta para as seguintes questões:

1. Interessa saber se o método iterativo converge ou não para a solução procurada. Desta forma, devem ser analisadas as condições necessárias e/ou suficientes de convergência do método.
2. Tendo a garantia da convergência do método, deve-se saber qual a razão de convergência: seja $\{x_k\}$ uma sucessão convergente para α; se existirem constantes positivas P e C tais que,

$$\lim_{k\to\infty} \frac{\left|\alpha - x_{k+1}\right|}{\left|\alpha - x_k\right|^P} = C$$

   então diz-se que a sucessão $\{x_k\}$ é convergente para α de ordem **P** com uma constante de convergência assimptótica igual a **C**:

   a) P = 1, convergência *linear/1ª ordem* (C < 1); dígitos ganhos por iteração: constante.

   b) P > 1, convergência *super-linear*; dígitos ganhos por iteração: aumenta.

   c) P = 2, convergência *quadrática*/2ª *ordem*; dígitos ganhos por iteração: duplica.

   Quanto maior for a ordem de convergência de um método iterativo menor será, em princípio, o número de iterações necessárias para atingir uma dada precisão.

   No entanto a rapidez depende também do esforço computacional requerido em cada iteração.
3. A implementação de um método iterativo exige a realização de um número "infinito" de operações para se chegar à solução. No entanto, face aos recursos limitados disponíveis, o processo iterativo tem de ser terminado após um número finito de operações. Esta paragem tem de ser feita com a ajuda de condições que, sendo verificadas, dão melhor garantia de que se está perto da solução. O valor obtido na última iteração é a melhor aproximação calculada. Estas condições definem o que é designado por critério de paragem de um processo iterativo.

5. Resolução de problemas

Os problemas reais que podem ser resolvidos usando métodos numéricos podem ser modelados matematicamente, através de equações não lineares (uma só equação ou um sistema de equações) e através de equações lineares (uma só equação ou um sistema de equações).

As funções associadas às equações matemáticas podem ser obtidas de 3 formas: diretamente (já conhecidas), por interpolação a partir de vários pontos conhecidos, e por aproximação a partir de vários pontos conhecidos.

6. Problemas com equações não lineares

6.1. Forma geral do problema

Uma equação não linear na variável *x* é representada na forma

$$f(x) = 0,$$

em que $f : R \rightarrow R$ é uma função contínua não linear em $x \in R$. A variável *x* diz-se *independente* e a variável $y = f(x)$ é a variável *dependente*. Resolver a equação $f(x) = 0$ consiste em calcular as r**aízes** da equação ou os **zeros** da função f(x). Na representação gráfica da função f(x) no plano XOY, os pontos de interseção da curva f(x) com o eixo dos XX definem as raízes reais de $f(x) = 0$. Pode-se esperar que uma equação não linear tenha raízes reais e/ou complexas.

6.2. Características do problema

Existem dois tipos de de equações não lineares: as **algébricas** e as **transcendentes**. As ***equações algébricas*** envolvem apenas as operações aritméticas básicas, sendo a forma polinomial um caso particular,

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = 0$$

em que os coeficientes a_i ($i = 0, \ldots, n$) são números reais ou complexos. Do grupo das equações algébricas fazem parte as *diofantinas*, que são equações polinomiais com apenas soluções inteiras, e que nem sempre têm solução. Por exemplo, $x^n + y^n = z^n$ não tem solução inteira para $n > 2$. Também as equações polinomiais lineares, quadráticas, cúbicas e quárticas fazem parte deste conjunto de equações, as quais têm fórmulas resolventes, umas mais complicadas do que outras.

As ***equações transcendentes*** envolvem funções trigonométricas, exponenciais, logarítmicas, entre outras, para além das polinomiais. São exemplos de equações transcendentes,

$$f(x) = x - e^{-x} = 0,$$
$$f(x) = x + \ln(x) = 0,$$
$$f(x) = (2x + 1)^2 - 4\cos(\pi x) = 0.$$

Se, para um dado valor p da variável x, pretende-se calcular o correspondente valor de f(p), o ***problema*** diz-se ***direto***. O problema diz-se ***inverso***, se o objetivo é determinar os valores de x que satisfazem a equação $f(x) = 0$. O *problema direto* não oferece qualquer dificuldade, apenas o *problema inverso* requer, na grande maioria dos casos, a utilização de um **método numérico**.

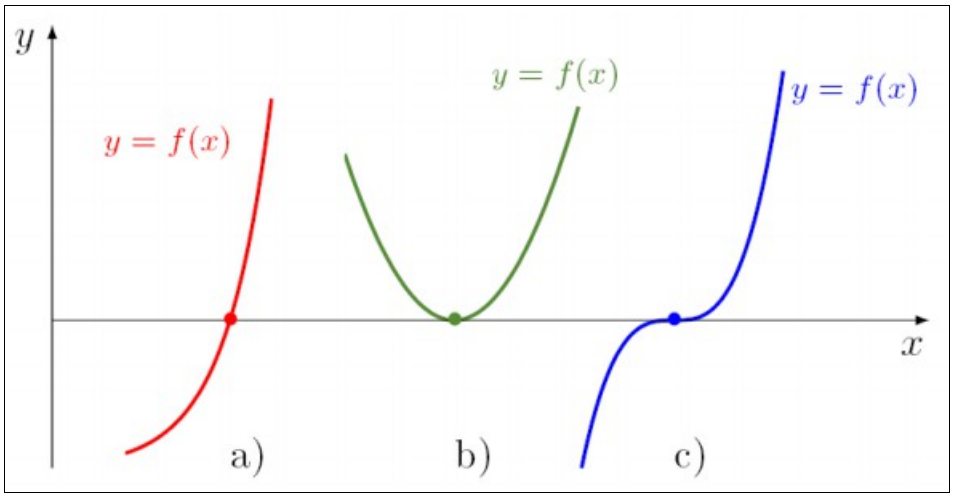
6.3. Zeros (raízes) e multiplicidade

Se $f(\alpha) = 0$ diz-se que α é uma raiz da equação $f(x) = 0$ ou que α é um zero da função f(x). Um zero de uma função pode ser de vários tipos. Por exemplo,

a) simples: $f(\alpha) = 0$

b) duplo: $f(\alpha) = f'(\alpha) = 0$

c) triplo: $f(\alpha) = f'(\alpha) = f''(\alpha) = 0$



Definição:

A multiplicidade de um zero α da função f(x) é o supremo *m* dos valores *k* tais que,

$$\lim_{x \to \alpha} \frac{|f(x)|}{|x-\alpha|^k} = c < \infty \qquad \text{(m é o maior valor de k)}$$

Se m = 1 o zero diz-se simples, se m = 2 o zero diz-se duplo, …

Exemplos:

a) $\alpha = 0$ é um zero simples da função f(x) = sen(x) porque,

$$\lim_{x \to 0} \frac{|\text{sen}(x)|}{|x|} = 1$$

b) $\alpha = 0$ é um zero duplo da função f(x) = 1 – cos(x) porque,

$$\lim_{x \to 0} \frac{|1-\cos(x)|}{|x|^2} = \frac{1}{2}$$

Nota: a multiplicidade de um zero pode não ser um número inteiro, nem sequer finita.

Teorema:

Se α for um zero da função f(x) e se f(x) for *m* vezes diferenciável em α então a multiplicidade de α é m se e só se,

$$f(\alpha) = f'(\alpha) = \ldots = f^{(m-1)}(\alpha) = 0, \quad \text{mas } f^m(\alpha) \neq 0.$$

Exemplos:

a) para f(x) = sen(x), f(0) = 0 mas $f'(0) \neq 0$, portanto m = 1.

b) para f(x) = 1 - cos(x), f(0) = f'(0) = 0 mas $f''(0) \neq 0$, portanto m = 2.

6.4. Utilização de métodos iterativos

A maior parte dos métodos numéricos para a resolução da equação não linear f(x) = 0 pertence à classe dos métodos iterativos.

Os métodos iterativos para resolver o problema f(x) = 0 podem ser classificados em dois grandes grupos: *os métodos de encaixe e os métodos de intervalo aberto*.

Os primeiros caracterizam-se por definir, em cada iteração, um intervalo que contém a raiz e construir, para a iteração seguinte, outro intervalo encaixado neste e que continue a conter a raiz. Os intervalos, como aparecem encaixados uns nos outros, têm amplitudes sucessivamente menores. Como exemplos de métodos de encaixe são o da Bissecção e o da Falsa Posição.

No grupo dos métodos de intervalo aberto não é necessário definir um intervalo que contenha a raiz. O processo iterativo pode ser iniciado com uma única aproximação à raiz, ou mesmo duas. A convergência destes métodos depende dos valores iniciais atribuídos na primeira iteração. Deste grupo de métodos fazem parte o do Ponto Fixo, o de Newton-Raphson e o da Secante.

Independentemente do método utilizado, é possível em muitos casos obter-se um majorante para o erro.

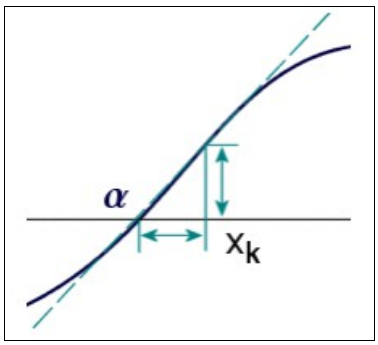
Teorema:

Seja α a raiz exata e x_k um valor aproximado da raiz da equação

$f(x) = 0$ com $\alpha, x_k \in [a, b]$.

Se f(x) for diferenciável em [a, b] e $|f'(x)| \geq m > 0, \forall x \in [a, b]$, então,

$$|\alpha - x_k| \leq \frac{|f(x_k)|}{m}$$



Demonstração:

Pelo teorema do Valor Médio,

$$\frac{f(\alpha) - f(x_k)}{\alpha - x_k} \leq f'(\zeta), \quad \zeta \in (\alpha, x_k)$$

aplicando módulo,

$$|\alpha - x_k| \leq \frac{|f(x_k)|}{|f'(\zeta)|} \quad (f(\alpha) = 0)$$

e portanto,

$$|\alpha - x_k| \leq \frac{|f(x_k)|}{m} \quad \text{(fazendo } m = |f'(\zeta)|.$$

6.5. Localização e separação das raízes

Para se aplicar um método iterativo, na resolução de uma equação não linear, é necessário conhecer uma aproximação inicial. Além disto, para certos métodos, para haver convergência a

aproximação inicial deve estar suficientemente próxima da raiz. Deste trabalho de análise feito *à priori* depende do sucesso na resolução numérica do problema proposto. Desta forma, antes de se aplicar um método iterativo para resolver a equação f(x) = 0, é necessário obter uma aproximação inicial, o que exige a separação das possíveis raízes em intervalos tão pequenos quanto possível.

O método mais prático consiste em analisar a representação gráfica de f(x) ou a combinação dos termos que formam a sua expressão analítica. Se o gráfico de f pode ser esboçado facilmente, então são obtidas geometricamente estimativas para raízes. Se a equação f(x) = 0 pode ser escrita na forma g(x) = h(x), onde g e h são facilmente representadas graficamente, então os pontos α tais que g(α) = h(α) verificam f(α) = 0. O processo consiste no seguinte:

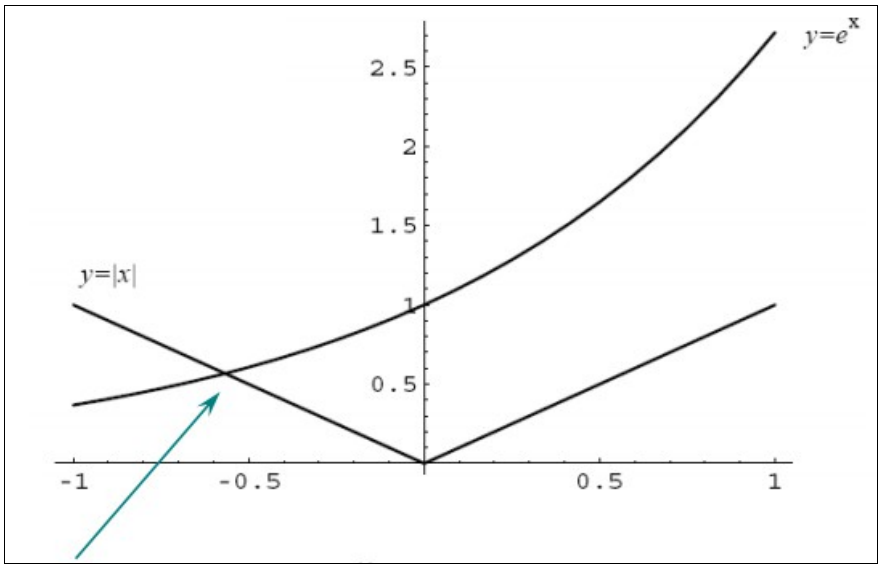
1º escolher um intervalo [a, b] onde se estima estar um zero da função (ou seja, que contenha um ponto de interseção das duas funções);

2º verificar que existe um *ponto de interseção* das duas funções no intervalo (a, b). Confirmar esta observação, com base nos dois resultados seguintes:

1. Se f(x) é uma função real e contínua em [a, b], sendo a e b números reais, tendo f(a) e f(b) sinais contrários (f(a).f(b) < 0), então existe pelo menos uma raiz real entre a e b.
2. Se f'(x) existe, é contínua e mantém o sinal no intervalo [a, b], então existe no máximo uma raiz em [a, b].

Portanto, se estas duas condições se verificarem, então existe uma raiz em [a, b] e é única.

Por exemplo, para f(x) = |x| - e^x,



1º escolher o intervalo (-1, 0) como contendo o ponto de interseção das duas funções;

2º verificar que existe um *ponto de interseção* de **|x|** com **e^x** no intervalo (-1, 0), com base nos dois resultados mencionados:

f(x) $\in$ C((-1, 0))

f(-1) = 0.632 > 0 e f(0) = -1 < 0

f'(x) = -1 – e^x < 0 em todo o intervalo (-1, 0)

Chamam-se *números de Rolle* da equação f(x) = 0, definida em $D \subseteq R$, ao conjunto dos pontos fronteira de D e dos zeros da função derivada de f. Se ordenados por ordem crescente, então entre dois números de Rolle consecutivos existe no máximo uma raiz real da equação.

6.6. Estimativa para o erro de truncatura

Seja $\{ x_k \}_{k = 1,2,...}$, uma sequência de aproximações convergindo para uma raiz real simples α de f(x) = 0, obtidas usando um método iterativo. Deduz-se uma expressão que dá um limite para o erro na aproximação x_k para α.

Pelo teorema do valor médio, $f(x_k) - f(\alpha) = (x_k - \alpha) f'(\eta_k)$,

sendo $\eta_k \in \left[\min\{x_k, \alpha\}, \max\{x_k, \alpha\}\right]$. Então, $\varepsilon_k = \left|\alpha - x_k\right|$ verifica $\varepsilon_k = \dfrac{\left|f(x_k)\right|}{\left|f'(\eta_k)\right|}$.

Se f é contínua em [a, b] contendo α e $f'(\alpha) \neq 0$ então existe $N_r(\alpha) = [\ \alpha - r, \alpha + r\ ] \subseteq [a, b]$ tal que $x \in N_r(\alpha)$. Além disso, existe uma constante $M_1 > 0$ sendo $|f'(x)| \geq M_1$ para $x \in N_r(\alpha)$. Dado que $\{ x_k \}$ converge para α existe k' tal que se $k > k'$, $|x_k - \alpha| < r$, e consequentemente $|\eta_k - \alpha| < r$, isto é, $\eta_k \in N_r(\alpha)$. Donde,

$$\varepsilon_k \leq \frac{\left|f(x_k)\right|}{M_1}.$$

6.7. Critérios de paragem

Note-se que há duas possíveis interpretações computacionais para o problema colocado com a equação f(x) = 0. Uma é calcular um valor x_k muito próximo de α onde $f(\alpha) = 0$. Outra é calcular x_k tal que $|f(x_k)|$ é muito pequeno (muito próximo de zero).

Assim, critério de paragem dos métodos iterativos para calcular uma raiz de f(x) = 0 contém três parâmetros: ε_1, ε_2 e kmax. O objetivo é terminar o processo após o cálculo de x_k quando

$$\left|x_k - x_{k-1}\right| \leq \varepsilon_1 \ \left(\text{ou } \left|x_k - x_{k-1}\right| \leq \varepsilon_1 \left|x_k\right|\right) \text{ ou } \left|f(x_k)\right| \leq \varepsilon_2 \text{ ou } k = kmax.$$

O primeiro parâmetro, $\boldsymbol{\varepsilon_1}$, serve para verificar a proximidade de $\mathbf{x_k}$ em relação a $\boldsymbol{\alpha}$ (um zero da função), o segundo, $\boldsymbol{\varepsilon_2}$, para verificar se $\mathbf{f(x_k)}$ está próximo de 0 ($\boldsymbol{f(x_k) \approx 0}$), e o terceiro, **k**, para controlar o número de iterações (se atingiu o número máximo de iterações predefinido, **kmax**).

6.8. Método da Bissecção

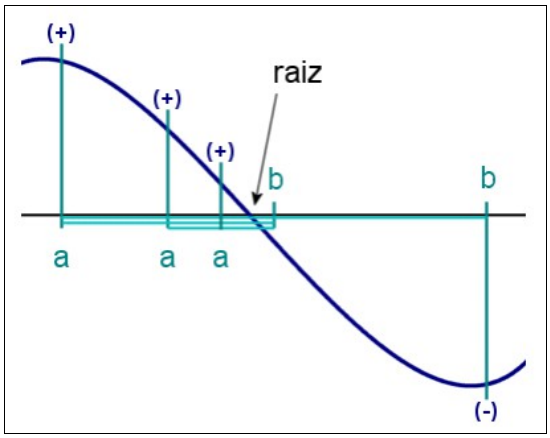
6.8.1. Fórmula geral

Este método é baseado no teorema do valor intermédio e consiste no seguinte: partindo de um intervalo [a, b] que contém a raiz, construir uma sucessão de intervalos, sendo cada um deles a metade do intervalo anterior que contém a raiz.

Seja $[a, b] \subseteq D$ e $f(a).f(b) < 0$. Então (a, b) contém uma raiz real de f(x) = 0.

Seja $I_0 = [a, b]$ e x_0 o ponto médio de I_0. Se $f(a).f(x_0) < 0$ então a raiz está em (a, x_0) senão está em (x_0, b).

Suponha que $f(a).f(x_0) < 0$. Seja $I_1 = [a, x_0]$ e seja x_1 o ponto médio de I_1. Se $f(a).f(x_1) < 0$ então a raiz está em (a, x_1) senão está em (x_1, x_0).



6.8.2. Algoritmo para o método da Bissecção

Objetivo: Calcular uma raiz real de f(x) = 0 em [a, b], $\alpha \in (a, b)$

Parâmetros de entrada: a, b, $\varepsilon_1, \varepsilon_2 \in R^+$, kmax $\in$ N; com f(a).f(b) < 0

```
fa ← f(a)
k ← 0
repita
    m ← (a + b) / 2
    fm ← f(m)
    se fa.fm < 0 então   { α ∈ (a, m) }
        b ← m
    senão   { α ∈ (m, b) }
        a ← m
        fa ← fm
    fim_se
    k ← k + 1
    se (|a - b| < ε1) ou (|fm| < ε2) ou (k > kmax) então
            escolher α ∈ (a, b)
            interromper
    fim_se
fim_repita
{  α ∈ (a, b) e |a - b| < ε1 ou |fm| < ε2 ou k > kmax }
```

Note-se que:

- Só é necessário calcular o valor de f(x) uma vez por iteração.
- Em aritmética de reais é extremamente improvável atingir o valor exato da raiz, pelo que não é necessário testar a igualdade.

Para um dado erro absoluto ε, em cada iteração k, utilizou-se o teste:

$$\frac{\left| b_k - a_k \right|}{2} \leq \varepsilon_1$$

de modo que o erro cometido seja inferior a metade da amplitude do intervalo.

Deste modo, sendo c_k os sucessivos pontos médios,

$$\left| c_1 - \alpha \right| \leq \frac{\left| b-a \right|}{2}; \left| c_2 - \alpha \right| \leq \frac{\left| b-a \right|}{2^2}; \ldots ; \left| c_n - \alpha \right| \leq \frac{\left| b-a \right|}{2^n}$$

é possível estimar o número **n** de iterações necessárias, para garantir uma aproximação da raiz com um erro absoluto máximo de ε_1:

$$\frac{b-a}{2^n} \leq \varepsilon_1$$

ou seja,

$$2^n \geq \frac{b-a}{\varepsilon_1} \Leftrightarrow \ln(2^n) \geq \ln\left((b-a)/\varepsilon_1\right) \Rightarrow n \geq \frac{\ln\left((b-a)/\varepsilon_1\right)}{\ln 2}$$

Por exemplo, aplicando o método da bissecção à equação

$f(x) = |x| - e^x$, com $\varepsilon_1 = 10^{-6}$

e considerando [a, b] = [-1, 0], obtiveram-se os seguintes resultados:

k	a_k	b_k	k	a_k	b_k
1	-1.000000	0.000000	11	-0.567383	-0.566406
2	-1.000000	-0.500000	12	-0.567383	-0.566895
3	-0.750000	-0.500000	13	-0.567383	-0.567139
4	-0.625000	-0.500000	14	-0.567261	-0.567139
5	-0.625000	-0.562500	15	-0.567200	-0.567139
6	-0.593750	-0.562500	16	-0.567169	-0.567139
7	-0.578125	-0.562500	17	-0.567154	-0.567139
8	-0.570313	-0.562500	18	-0.567146	-0.567139
9	-0.570313	-0.566406	19	-0.567146	-0.567142
10	-0.568359	-0.566406	20	-0.567144	-0.567142

Analisando os resultados obtidos, pode-se concluir que:

- a raiz da equação em estudo encontra-se no intervalo -0.567144, -0.567142);
- o ponto médio deste intervalo é -0.567143 que é um valor aproximado da raiz com um erro absoluto que não excede 10^{-6};

- O processo terminou na iteração k = 20, em que

$$\frac{b-a}{2^n} = 0.00000095,$$

ou seja,

$$n \geq \frac{\ln\left((b-a)/\varepsilon_1\right)}{\ln 2} = 19.931569.$$

As vantagens do método da Bissecção são:

- converge sempre (desde que exista raiz no intervalo inicial);
- possibilidade de prever um majorante para o erro cometido ao fim de um certo número de iterações;
- custo computacional de cada iteração muito baixo.

As desvantagens do método da Bissecção são:

- A maior desvantagem reside no facto da sua convergência ser muito lenta (muitas iterações) quando comparada com a dos outros métodos. A ordem de convergência do método da Bissecção é linear (P = 1) e a constante de convergência é igual a 1/2 (C = 1/2).

6.9. Método da Falsa Posição (ou da Corda Falsa)

6.9.1. Fórmula geral

Este método pode ser encarado como um melhoramento do método da Bissecção. Em vez de se determinar o ponto médio, é determinado um ponto $\mathbf{c_k}$ resultante da interseção da secante que passa pelos pontos $(a_k, f(a_k))$ e $(b_k, f(b_k))$ com o eixo dos XX.

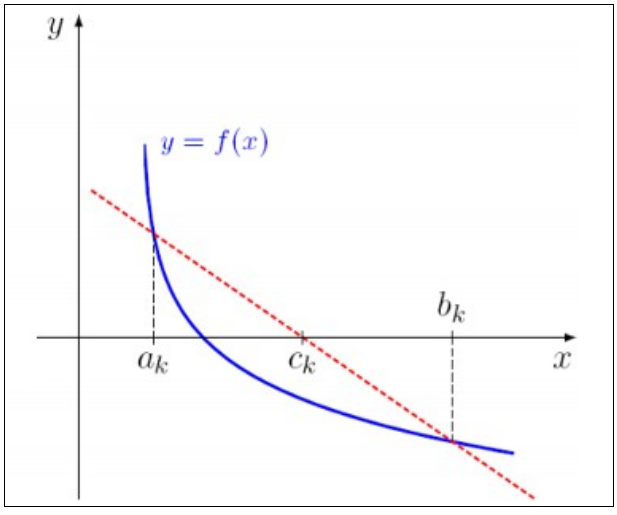
A partir da equação da secante,

$$y - f(b_k) = \frac{f(b_k)-f(a_k)}{b_k-a_k}\left(x-b_k\right)$$

e fazendo y = 0 e substituindo x por c_k obtém-se,

$$c_k = b_k - \frac{f(b_k)}{f(b_k)-f(a_k)}\left(b_k-a_k\right)$$

Note-se que os sucessivos cálculos desta fórmula não provocam efeitos de cancelamento subtrativo pois $f(b_k)$ e $f(a_k)$ têm sinais contrários.



6.9.2. Algoritmo para o método da Falsa Posição

Objetivo: Calcular uma raiz real de f(x) em [a,b], $\alpha \in (a,b)$

Parâmetros de entrada: a, b, ε_1, $\varepsilon_2 \in R^+$, kmax $\in$ N; com f(a).f(b) < 0

fa $\leftarrow$ f(a)

k $\leftarrow$ 0

repita

$$m = b - \frac{f(b)}{f(b) - f(a)}(b - a)$$

fm $\leftarrow$ f(m)

se fa.fm < 0 **então** { $\alpha \in$ (a, m) }

b $\leftarrow$ m

senão { $\alpha \in$ (m, b) }

a $\leftarrow$ m

fa $\leftarrow$ fm

fim_se

k $\leftarrow$ k + 1

se (|a - b| < ε_1) ou |fm| < ε_2) ou (k > kmax) **então**

escolher $\alpha \in$ (a, b)

interromper

fim_se

fim_repita

{ $\alpha \in$ (a, b) e |a - b| < ε_1 ou |fm| < ε_2 ou k > kmax }

6.10. Método do Ponto Fixo

6.10.1. Fórmula geral

Pretende-se determinar a solução α de uma equação não linear da forma,

x = g(x).

Dada uma equação na forma f(x) = 0 é sempre possível fazer,

x = x + f(x), em que g(x) = x + f(x).

De uma forma mais geral pode-se considerar,

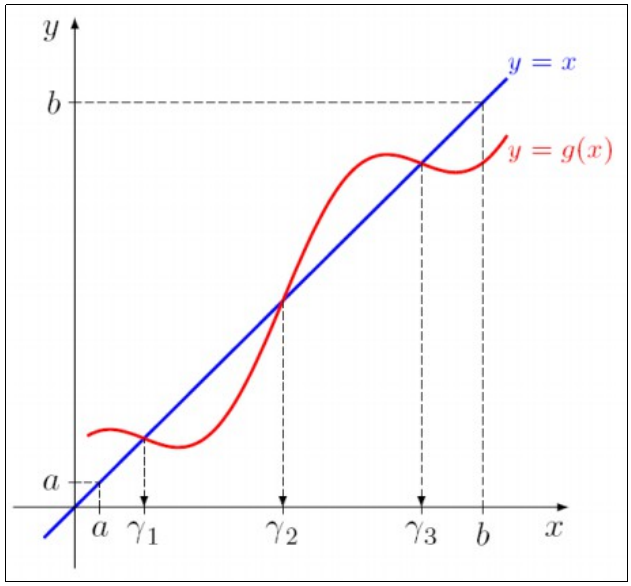
g(x) = x + c(x).f(x)

onde g(x) é uma função contínua, não nula e limitada no intervalo [a,b], contendo a raiz α de f(x).

Definição:

Um **ponto fixo** de uma função g(x) é um número real α tal que $\alpha = g(\alpha)$. Dada uma aproximação inicial $x_0 \in [a, b]$, o método do Ponto Fixo consiste numa sucessão de aproximações $\{ x_k \} \rightarrow \alpha$ tal que,

$$x_{k+1} = g(x_k) , \quad k = 0, 1, 2, \ldots$$

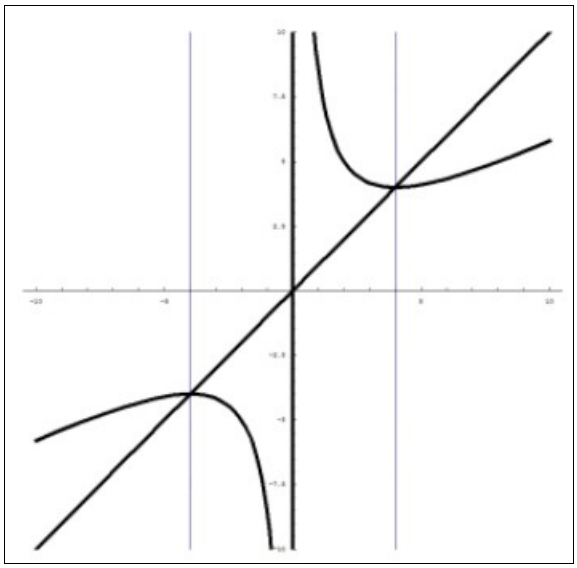


Geometricamente, os ***pontos fixos*** de uma função y = f(x) são os pontos de intersecção de y = g(x) com y = x.

Assim, se $f(x) = 0 \Leftrightarrow x = g(x)$, então determinar a raiz de f(x) = 0 em [a, b] é o mesmo que procurar o ponto fixo de g(x) em [a, b].

Por exemplo, o cálculo de $\sqrt{a}$ consiste na sucessão de aproximações:

$$x_{k+1} = \frac{1}{2}\left(\frac{a}{x_k} + x_k\right)$$



Experimente-se para a = 16, começando com $x_0 = 10$:

0 10.00000000
1 5.80000000
2 4.27931034
3 4.00911529
4 4.00001036
5 4.00000000

O método utilizado tem por base a equação,

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

que é equivalente a $\mathbf{x^2 = a}$ e consiste na *pesquisa de um ponto fixo* da função g(x).

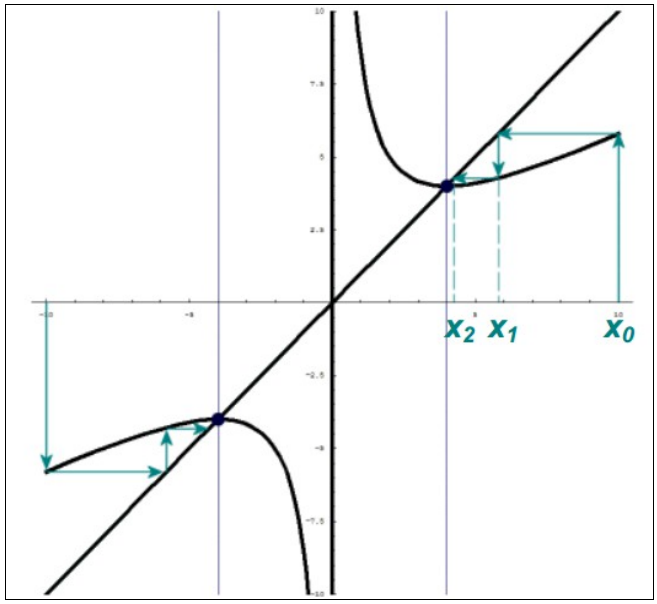
Para a = 16 a função g(x) tem dois pontos fixos, em x = 4 e x = -4.

Ao partir-se de uma estimativa inicial negativa, o método encontra a raiz negativa de 16.

0 -10.00000000
1 -5.80000000
2 -4.27931034
3 -4.00911529
4 -4.00001036
5 -4.00000000

Como funciona?

A partir de uma aproximação inicial x_0, uma sucessão de aproximações da forma $x_{k+1} = g(x_k)$ converge para um ponto fixo da função g(x).



Porque funciona?

Teorema:

Seja g(x) uma função contínua e $\{x_k\}$ uma sucessão de aproximações gerada pelo método do Ponto Fixo, $x_{k+1} = g(x_k)$. Se $\lim_{k\to\infty} x_k = \alpha$ então α é um *ponto fixo* de g(x).

Quando existe ponto fixo ?

Teorema:

Seja $g(x) \in C([a, b])$. Se para todo o $x \in [a, b]$ se verifica-se que $g(x) \in [a, b]$ (isto é, g é uma contração), então g(x) tem pelo menos um ponto fixo em [a, b].

Quando é único o ponto fixo ?

Teorema:

Se g'(x) está definida em [a, b] e existe uma constante positiva $0 < L < 1$, tal que $|g'(x)| \leq L$ para todo o $x \in [a, b]$, então g(x) *tem um único ponto fixo* em [a, b].

6.10.2. Convergência

Quando converge o método do ponto fixo?

Teorema do Ponto Fixo:

Sejam $g(x), g'(x) \in C([a, b])$:

$g(x) \in [a, b]$, para todo o $x \in [a, b]$,

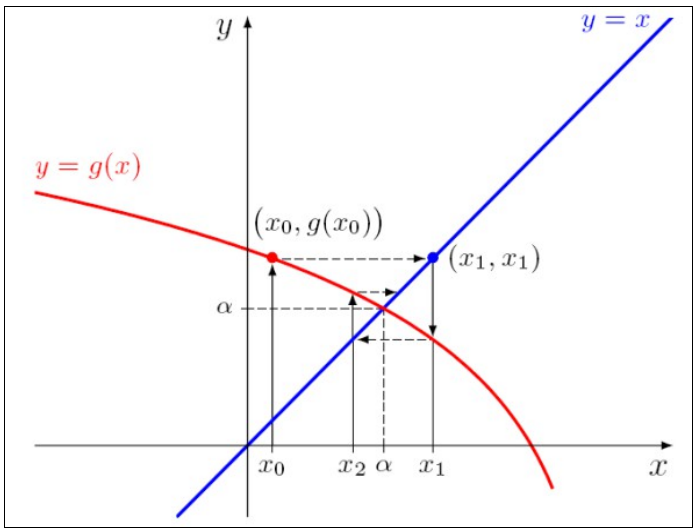
$|g'(x)| < 1$, para todo o $x \in [a, b]$,

$x_0 \in [a, b]$.

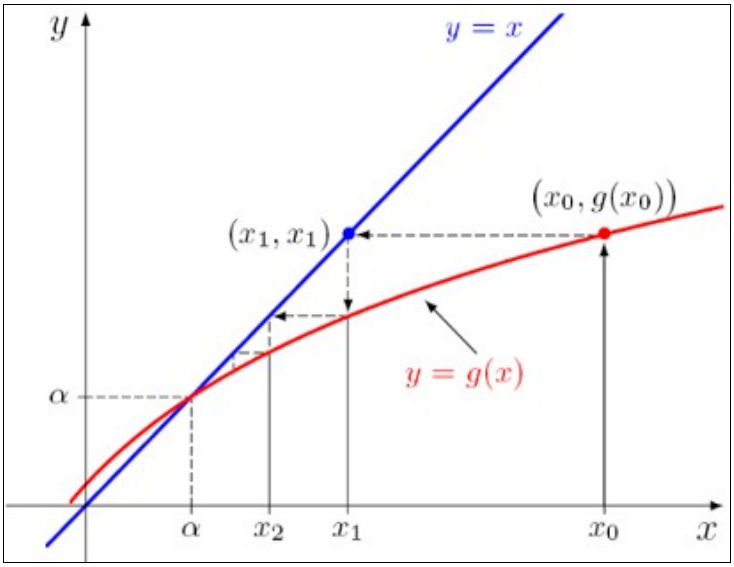
Então a sucessão $\{ x_k \}$ gerada por $x_{k+1} = g(x_k)$, $k = 0, 1, 2, \ldots$, converge para o único ponto fixo $\alpha \in [a, b]$.

Como converge o método do ponto fixo?

Convergência *monótona* quando $0 < g'_0(x) < 1$:



Convergência *oscilante* quando $-1 < g'_0(x) < 0$:



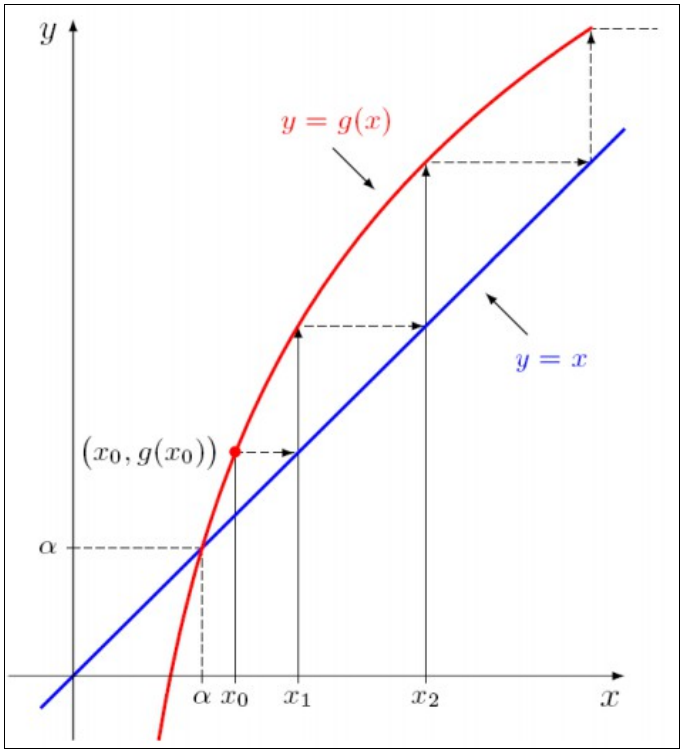
Quando *diverge* o método do ponto fixo ?

Teorema:

Seja $g : D \subseteq R \rightarrow R$. Se $g, g' \in C(D)$, $g(x)$ com um ponto fixo $\alpha \in [a, b] \subset D$, $|g'(x)| > 1$ para todo o $x \in D$, $x_0 \in [a, b]$ (com $x_0 \neq \alpha$). Então a sucessão $\{ x_k \}$ gerada por $x_{k+1} = g(x_k)$, $k = 0, 1, 2, \ldots$, não converge para o ponto fixo $\alpha \in [a, b]$.

***Como diverge* o método do ponto fixo ?**

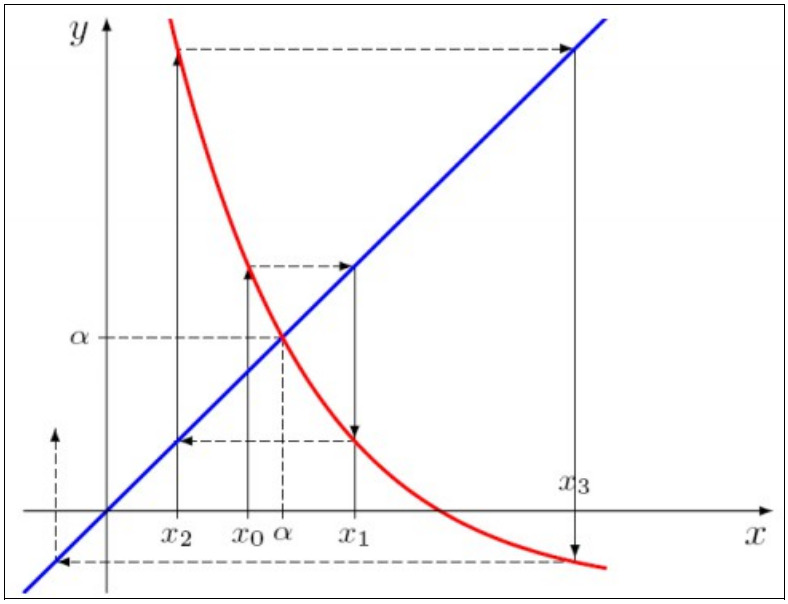
Divergência *monótona* quando g'(x) > 1:



Divergência *oscilante* quando g'(x) < -1 :

Quando converge, qual a ordem de convergência do método do ponto fixo ?

Consideremos que g(x), g'(x) ∈ C([a, b]) e que o método do ponto fixo é *convergente* para α.



1) No caso de g'(α) ≠ 0, e como |g'(α)| < 1, então o método do ponto fixo apresenta ordem de convergência linear sendo |g'(α)| a constante assimptótica de convergência.

2) Para o caso de $g'(\alpha) = 0$ e $g''(\alpha) \neq 0$, o método do ponto fixo apresenta ordem de convergência quadrática sendo $|g''(\alpha)|/2$ a constante assimptótica de convergência.

3) De um modo geral, assumindo que $g(x) \in C^n([a, b])$, se

$g'(\alpha) = g''(\alpha) = \ldots = g^{(n-1)}(\alpha) = 0$, mas $g^{(n)}(\alpha) \neq 0$

prova-se que o método do Ponto Fixo apresenta ordem de convergência n.

6.10.3. Algoritmo do método do Ponto Fixo

Majoração do erro:

$$\left|x_{n+1} - \alpha\right| \leq \frac{1-L}{L} \left|x_{n+1} - x_n\right| \text{ (com } L < 1)$$

Critério de paragem:

$$\frac{1-L}{L} \left|x_{n+1} - x_n\right| \leq \varepsilon_1 \quad \Rightarrow \quad \left|x_{n+1} - \alpha\right| \leq \varepsilon_1$$

Objetivo: Calcular raiz real simples de $f(x) = 0$, α

Parâmetros de entrada: $\mathbf{x_0}$, $\boldsymbol{\varepsilon_1}$, $\boldsymbol{\varepsilon_2}$ e **kmax**; garantia de convergência

$k \leftarrow 0$
repita
  $k \leftarrow k + 1$
  $x_1 \leftarrow g(x_0)$
  se $(|x_1 - x_0| < \varepsilon_1)$ ou $(|f(x_1)| < \varepsilon_2)$ ou $(k = kmax)$ **então**
    $\alpha \leftarrow x_1$ **{ aproximação obtida }**
    interromper
  fim_se
  $x_0 \leftarrow x_1$
fim_repita

6.10.4. Exemplo

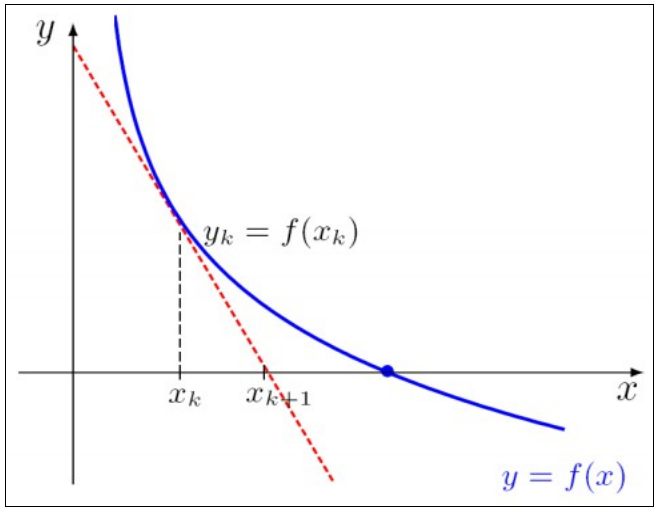
Determinar, com um erro absoluto inferior a 5×10^{-5}, o zero da função $f(x) = 1 + x + e^x$ no intervalo [-2, -1].

k	x_k	$x_{k+1} = g(x_k)$	δ	k	x_k	$x_{k+1} = g(x_k)$	δ
0	-2.00000	-1.13534	$+5.0 \times 10^{-1}$	5	-1.27756	-1.27872	$+6.8 \times 10^{-4}$
1	-1.13534	-1.32131	$+1.1 \times 10^{-1}$	6	-1.27872	-1.27839	$+1.9 \times 10^{-4}$
2	-1.32131	-1.26678	$+3.2 \times 10^{-2}$	7	-1.27839	-1.27848	$+5.2 \times 10^{-5}$
3	-1.26678	-1.28174	$+8.7 \times 10^{-3}$	8	-1.27848	-1.27846	$+1.5 \times 10^{-5}$
4	-1.28174	-1.27756	$+2.4 \times 10^{-3}$				

6.11. Método de Newton-Raphson

6.11.1. Fórmula geral

Em cada iteração x_k, a curva y = f(x) é aproximada pela sua tangente e a interseção desta com o eixo dos XX é a nova aproximação x_{k+1}.



A equação tangente à curva no ponto $(x_k, f(x_k))$ é,

$$y = f(x_k) + f'(x_k)\,(x - x_k)$$

e a interseção desta com o eixo dos XX determina a nova aproximação,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

A partir de uma aproximação inicial x_0 esta fórmula gera uma sucessão $\{ x_k \}$ que, em certos casos, deverá convergir para um zero da função.

Por exemplo, para a função $f(x) = x^2 - a$,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^2 - a}{2x_k} = \frac{1}{2}\left(\frac{a}{x_k} + x_k\right)$$

e para o caso particular de a = 16, com x_0 = 10 (aproximação inicial), a sucessão das aproximações tende para um **zero** de $f(x) = x^2 - 16$.

k	x_k	k	x_k
0	10.00000000	3	4.00911529
1	5.80000000	4	4.00001036
2	4.27931034	5	4.00000000

6.11.2. Newton-Raphson como caso particular do método do Ponto Fixo

Dada uma equação f(x) = 0, pode-se passar para a forma x = g(x) através da relação,

$$g(x) = x + c(x).f(x)$$

onde c(x) é uma função contínua, não nula e limitada no intervalo [a,b], contendo a raiz α de f(x).

Pretende-se definir c(x) de modo a que o método do Ponto Fixo (no caso de convergir) tenha uma ordem de convergência pelo menos quadrática (ordem 2).

Assumindo que f(x) e c(x) são diferenciáveis em [a, b],

$$g'(x) = 1 + c'(x).f(x) + c(x).f'(x)$$

e calculando no ponto α,

$$g'(\alpha) = 1 + c'(\alpha).f(\alpha) + c(\alpha).f'(\alpha).$$

Para que a convergência seja quadrática, devemos ter g'(α) = 0. E como f(α) = 0 então,

$$c(\alpha) = -\frac{1}{f'(\alpha)}$$

Assim, basta escolher,

$$c(x) = -\frac{1}{f'(x)}$$

assumindo que f'(x) ≠ 0 em todo o intervalo [a, b].

Substituindo, temos a nova forma,

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

que corresponde ao método de Newton-Raphson,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, 2, \ldots$$

e que, por esta construção, se convergir é quadrático.

6.11.3. O método de Newton-Raphson a partir da série de Taylor

Suponha-se que f $\in C^2$([a, b]), que o Método de Newton-Raphson é convergente e considere-se o desenvolvimento de Taylor de ordem 1 em torno de x_k :

$$f(x) = f(x_k) + f'(x_k)\left(x - x_k\right) + \frac{f''(\zeta_k)}{2}\left(x - x_k\right)^2, \quad \zeta_k \in (x, x_k)$$

Calculando em x = α,

$$0 = f(\alpha) = f(x_k) + f'(x_k)\left(\alpha - x_k\right) + \frac{f''(\zeta_k)}{2}\left(\alpha - x_k\right)^2, \quad \zeta_k \in (\alpha, x_k)$$

donde,

$$\alpha = \left(x_k - \frac{f(x_k)}{f'(x_k)}\right) - \left(\frac{f''(\zeta_k)}{2f'(x_k)}\left(\alpha - x_k\right)^2\right)$$

e assim obtemos a *nova aproximação* x_{k+1} e o *erro* cometido.

Note-se que assumiu-se que $|\alpha - x_k|$ é pequeno para todo o k, incluindo a aproximação inicial k = 0.

6.11.4. Ordem de convergência do método de Newton-Raphson

Teorema:

A razão de convergência do método de Newton-Raphson é igual a 2 (convergência quadrática).

Prova-se, pela expressão anterior,

$$\alpha - x_{k+1} = -\frac{f''(\zeta_k)}{2f'(x_k)}\left(\alpha - x_k\right)^2$$

que, no caso de o método convergir,

$$\lim_{k\to\infty} \frac{\left|\alpha - x_{k+1}\right|}{\left|\alpha - x_k\right|^2} = \frac{\left|f''(\alpha)\right|}{2\left|f'(\alpha)\right|}$$

e a *convergência é quadrática* com *constante* de convergência assimptótica igual a $\frac{1}{2}\times\frac{\left|f''(\alpha)\right|}{\left|f'(\alpha)\right|}$

Observação:

Se o zero de f não for simples a ordem de convergência do método degrada-se.

Demonstra-se que, no caso dos de multiplicidade 2 a convergência é apenas linear.

6.11.5. Um majorante do erro absoluto

Pela expressão anterior,

$$\alpha - x_{k+1} = -\frac{f''(\zeta_k)}{2f'(x_k)}(\alpha - x_k)^2, \quad \zeta_k \in \left(\alpha, x_k\right)$$

tem-se

$$\left|e_{k+1}\right| = \frac{f''(\zeta_k)}{2f'(x_k)}\left|e_k\right|^2$$

Se identificar-se um majorante da segunda derivada

$$M_2 \geq \left|f''(x)\right|, \quad \forall x \in [a,b]$$

e um minorante da primeira derivada, para todo o intervalo,

$$0 < m_1 \leq \left|f'(x)\right|, \quad \forall x \in [a,b]$$

é simples calcular

$$\left|e_{k+1}\right| \leq \frac{M_2}{2m_1}\left|e_k\right|^2.$$

6.11.6. Uma estimativa do erro absoluto

Assumindo que f $\in$ C([a, b]) e que o Método de Newton-Raphson é convergente, pelo Teorema do Valor Médio (TVM),

$$\frac{f(x_k) - f(\alpha)}{x_l - \alpha} = f'(\zeta_k), \quad \zeta_k \in \left(\alpha, x_k\right)$$

donde, assumindo ainda que

$$f'(x) \neq 0, \quad \forall x \in \left(\alpha, x_k\right)$$

então,

$$x_k - \alpha = \frac{f(x_k)}{f'(\zeta_k)}.$$

Por outro lado, da expressão do próprio método,

$$x_k - x_{k+1} = \frac{f(x_k)}{f'(x_k)}$$

Para k suficientemente grande, $x_{k+1} \approx \alpha$, donde, $\zeta_k \approx x_k$, e portanto,

$$x_k - x_{k+1} \approx x_k - \alpha$$

Assim, pode-se estimar,

$$\left|e_k\right| \approx \left|x_{k+1} - x_k\right|$$

Em termos algorítmicos, é mais cómodo calcular,

$$\left|e_{k-1}\right| \approx \left|x_k - x_{k-1}\right|$$

De facto, para o exemplo anterior

k	x_k	$\|e_k\|$	$\|e_{k-1}\|$	$\|x_k - x_{k-1}\|$
0	10.00000000	6.00000000		
1	5.80000000	1.80000000	6.00000000	4.20000000
2	4.27931034	0.27931034	1.80000000	1.52068966
3	4.00911529	0.00911529	0.27931034	0.27019506
4	4.00001036	0.00001036	0.00911529	0.00910492
5	4.00000000	0.00000000	0.00001036	0.00001036

6.11.7. Critérios de convergência do método de Newton-Raphson

Teorema:

Seja $f \in C^2([a, b])$. Se

(i) $f(a).f(b) < 0$

(ii) $f'(x) \neq 0$ para todo o $x \in [a, b]$

(iii) $f''(x)$ não muda de sinal em $[a, b]$

(iv) $\left|\frac{f(a)}{f'(a)}\right| < b - a$ e $\left|\frac{f(b)}{f'(b)}\right| < b - a$

Então para qualquer $x_0 \in [a, b]$, a sucessão $\{ x_k \}$ gerada pelo método de Newton-Raphson converge para o único zero de f em [a, b].

Observações:

(i) + (ii) garantem a existência de uma só solução em [a, b];

(ii) + (iii) garantem que a função é monótona, convexa ou côncava;

(iv) garante que as tangentes à curva em (a, f(a)) e (b, f(b)) intersetam o eixo dos XX em (a, b).

6.11.8. Algoritmo para o método de Newton-Raphson

Objetivo: Calcular raiz real simples de $f(x) = 0$, α

Parâmetros de entrada: x_0, ε_1 (limite do erro absoluto), ε_2 e kmax; garantia de convergência

```
k ← 0
f0 ← f(x0)
repita
   x1 ← x0 - f0 / f'(x0)
   k ← k + 1
   f0 ← f(x1)
   se (|x1 - x0| < ε1) ou (|f0| < ε2) ou (k = kmax) então      { ver secção 6.11.6 }
      α ← x1   { aproximação obtida }
      interromper
   fim_se
   x0 ← x1
fim_repita
```

6.11.9. Vantagens e desvantagens do método de Newton-Raphson

As vantagens são as seguintes:

- ➔ Quando converge, tem *convergência quadrática*.
- ➔ Necessita apenas de *um ponto*, para estimativa inicial.

As desvantagens são as seguintes:

- ➔ Exige uma boa aproximação inicial; caso contrário pode divergir, ou encontrar outra raiz.
- ➔ Exige o cálculo da derivada em cada iteração, o que pode ser lento ou mesmo impossível.
- ➔ Exige que a derivada (no denominador) nunca se anule. Note-se que, mesmo para valores da derivada próximos de zero, a intersecção da tangente com o eixo dos XX é um ponto muito afastado.

6.11.10. Alguns casos patológicos do método de Newton-Raphson

Para a função $f(x) = x^3 - 2x + 2$, se escolhermos $x_0 = 0$, o método calcula $x_1 = 1$, gerando a sucessão de aproximações: 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, ...

Para a função $f(x) = \sqrt[3]{x}$ o método gera uma sucessão tal que, $x_{k+1} = -2\,x_k$.

Para a função $f(x) = \sqrt{x}$, obtém-se $x_{k+1} = -x_k$ de modo que, para qualquer x_0 , o método gera a sucessão: x_0 , $-x_0$, x_0 , $-x_0$, x_0 , $-x_0$, …

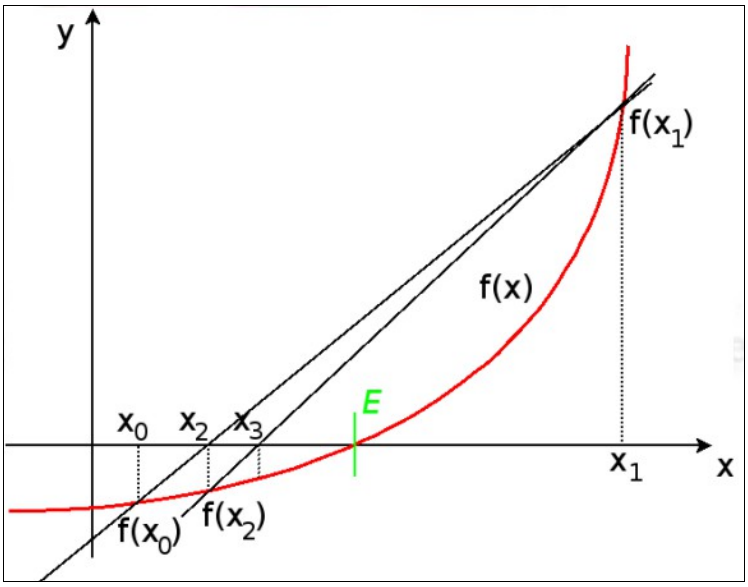
Todo o ponto de inflexão provoca um afastamento da raiz.

6.12. Método da Secante

Uma desvantagem da utilização do método de Newton-Raphson é a necessidade de calcular a derivada da função envolvida. Quando a expressão da derivada é complexa e o cálculo de valores da derivada pouco eficiente, quando comparado com o cálculo de valores da função, o uso do método da Secante poderá ser preferido ao do método de Newton-Raphson.

6.12.1. Forma geral

Para calcular a raiz da equação f(x) = 0 este método baseia-se na aproximação de f(x) por uma reta, na vizinhança da raiz de f(x). O ponto de interseção da reta com o eixo dos XX é considerado como aproximação à raiz de f(x) = 0. Se ainda estiver longe da solução α, o processo é repetido iterativamente.



Para iniciar o processo iterativo são escolhidos dois pontos x_0 e x_1. O intervalo definido por eles não necessita de conter a raiz. O ponto de interseção da reta, que passa pelos dois pontos, com o eixo dos XX obtém-se a partir da equação iterativa, cuja sua forma geral é

$$x_{k+1} = x_k - \frac{\left(x_k - x_{k-1}\right)}{f(x_k) - f(x_{k-1})} f(x_k), \quad k = 1, 2, \ldots.$$

Embora sejam necessários dois pontos para iniciar o processo iterativo, apenas um novo ponto e o correspondente valor da função são calculados em cada iteração. O gráfico anterior ilustra este processo iterativo.

Para evitar o efeito do cancelamento subtrativo quando x_k está muito próximo de x_{k-1}, aconselha-se a implementação da seguinte fórmula iterativa:

$$x_{k+1} = x_k - \frac{\left(x_{k-1} - x_k\right)\dfrac{f(x_k)}{f(x_{k-1})}}{1 - \dfrac{f(x_k)}{f(x_{k-1})}}.$$

que é preferível à fórmula iterativa anterior, desde que $|f(x_k)| < |f(x_{k-1})|$.

Usando a fórmula iterativa anterior, deve-se verificar se $f(x_k) \approx f(x_{k-1})$ que, quando acontecer, o processo deve terminar (pode provocar uma divisão por 0).

6.12.2. Exemplo

Determinar aproximações para a raiz real da função $x^3 - 2x - 5 = 0$, tomando como aproximações iniciais os pontos $x_0 = 3$ e $x_1 = 2$. Usando o método da Secante, foi calculada a seguinte sequência de iterações convergindo para a raiz real daquela função:

$x_2 = 2.058824$	$x_4 = 2.094511$	$x_6 = 2.094552$
$x_3 = 2.096559$	$x_5 = 2.094511$	$x_7 = 2.094552$

6.12.3. Convergência

Para que a sequência gerada por este método convirja para uma raiz real simples de f(x) é, em geral, necessário que as aproximações iniciais, x_0 e x_1, estejam suficientemente próximas da raiz.

Teorema:

Seja $f \in C^2([a,b])$ e α uma raiz simples de $f(x) = 0$ em $[a,b]$. Então existe $r > 0$ tal que a sequência $\{ x_k \}_{k = 2,3,...}$ gerada pelo método da Secante converge sempre que $|x_i - \alpha| < r$ $(i = 0,1)$.

As condições do teorema anterior (relacionado com a convergência do método de Newton-Raphson) são também suficientes para estabelecer convergência para o método da Secante.

Teorema:

Seja $f \in C^2([a,b])$. Se (i), (ii), (iii) e (iv) do teorema sobre convergência do método de Newton-Raphson se verificam, então para x_0, $x_1 \in [a,b]$ a sequência gerada pelo método da Secante converge para α único zero de f em $[a,b]$.

Teorema (Ordem de convergência do método da Secante):

A ordem de convergência deste método é $((1+\sqrt{5}) / 2) = 1.618...$ (convergência super-linear).

6.12.4. Algoritmo do método da Secante

Objetivo: Cálculo de uma raiz real simples de $f(x) = 0$, α

Parâmetros de entrada: x_0, x_1, ε_1 (limite do erro absoluto), ε_2 e kmax; garantia de convergência

$k \leftarrow 0$

$f_0 \leftarrow f(x_0)$

repita

$k \leftarrow k + 1$

$f_1 \leftarrow f(x_1)$

$x_2 \leftarrow x_1 - ((x_0 - x_1) / (f_1 - f_0)) \times f_1$

$x_0 \leftarrow x_1$

$x_1 \leftarrow x_2$

$f_0 \leftarrow f_1$

```
se (|x1 - x0| < ε1) ou (|f1| < ε2) ou (k > kmax) então      { ver secção 6.11.6 }
   α ← x1  { aproximação obtida }
   interromper
  fim_se
 fim_repita
```

7. Problemas com equações lineares

Neste capítulo será abordado o problema de resolução de sistemas de equações lineares, o qual é um dos problemas que na prática ocorre com maior frequência. Os métodos para resolver este tipo de problemas são classificados em duas classes: métodos diretos e métodos iterativos. Neste documento, apenas serão abordados com mais detalhe os métodos iterativos.

7.1. O problema da resolução de um sistema de equações lineares

Pretende-se calcular a solução de um sistema de equações lineares, cuja forma geral é,

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \qquad\qquad \ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n \end{cases}$$

onde

$x_1, x_2, \ldots, x_n$ são as incógnitas do sistema,

a_{ij} (i, j = 1,2, ..., n) são os coeficientes do sistema,

$b_1, b_2, \ldots, b_n$ são os termos independentes (ou segundos membros) do sistema.

Problema:

Pretende-se determinar valores para $\mathbf{x_1, x_2, \ldots, x_n}$ de modo que as **n** equações do sistema em cima sejam satisfeitas simultaneamente.

O sistema pode também escrever-se na sua f**orma matricial**

$$\mathbf{A\,x = b}$$

onde,

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_n \end{bmatrix}$$

sendo $A = (a_{ij})$ a matriz dos coeficientes, $b = (b_i)$ o vetor dos termos independentes e $x = (x_i)$ o vetor das incógnitas.

Definição:

Diz-se que um sistema de equações lineares é **determinado** se tem uma única solução.

Teorema:

Um sistema de equações lineares (escrito na sua forma matricial) é **determinado** se e só se verificar qualquer uma das duas condições (equivalentes):

1) A^{-1} (inversa de A) existir (A é invertível)
2) $\det A \neq 0$

Nota:

Ao longo deste texto é assumido que os sistemas de equações considerados têm solução única.

7.2. Utilização de métodos iterativos

Teoricamente, os métodos diretos permitem calcular a solução exata com um número finito de operações aritméticas básicas. No entanto, na prática isto não é bem assim, pois, devido à acumulação de erros de arredondamento, ao cancelamento subtrativo, etc., estes métodos permitem apenas calcular uma solução aproximada. Exemplos de métodos diretos são a Regra de Cramer, Eliminação de Gauss, Decomposição LU e Método de Choleski.

Teoricamente, nos métodos iterativos a solução é definida como um limite de uma sucessão (infinita) de vetores. Na prática, calcula-se apenas um número finito de vetores da sucessão, isto é, calcula-se um um número finito de iterações. Exemplos de métodos iterativos são os métodos de Jacobi e de Gauss Seidel. Estes métodos são apropriados para sistemas de grande dimensão, cujas matrizes dos coeficientes são dispersas.

Seja

$$\mathbf{A\,x = b} \qquad (1)$$

um sistema de n equações em n incógnitas.

A matriz A pode ser escrita na forma

$$\mathbf{A = M - N}, \qquad (2)$$

sendo M e N matrizes de ordem n e M invertível.

Substituindo (2) na expressão (1), obtém-se

$$(M - N)\,x = b$$

ou

$$M\,x = N\,x + b$$

donde

$$x = M^{-1}\,(N\,x + b) \qquad (3)$$

Assim, a solução de (3) é ponto fixo de (1) e reciprocamente.

Note-se que, $\forall\ x, y \in R^n$

$$\|x - y\| = \left\|M^{-1}\,(Nx + b) - M^{-1}\,(Ny + b)\right\| = \left\|M^{-1}\,N\,(x - y)\right\| \leq \left\|M^{-1}\,N\right\| \|x - y\|$$

se a norma de matriz for compatível com a norma de vetor.

Assim, tem-se o teorema que se segue.

Teorema 1:

Se $\left\|M^{-1} N\right\| < 1$ então a sequência definida pela iteração

$$x^{k+1} = M^{-1}\left(N x^{(k)} + b\right), \quad (k = 0, 1, \ldots) \tag{4}$$

converge para o ponto fixo de (3) qualquer que seja $x^{(0)} \in R^n$.

Teorema 2:

A razão de convergência do método iterativo definido por (4) é igual a 1. A constante de convergência é menor ou igual a $\left\|M^{-1} N\right\|$.

Com escolhas especiais para M e N serão definidos os métodos de Jacobi e de Gauss Seidel.

7.3. Método de Jacobi

7.3.1. Fórmula geral

Considerando a matriz dos coeficientes de (1), $A = (a_{ij})$, definam-se $D = (d_{ij})$ uma matriz diagonal, $L = (l_{ij})$ uma matriz estritamente triangular inferior e $U = (u_{ij})$ uma matriz estritamente triangular superior, tais que

$$d_{ij} = \begin{cases} a_{ij}, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases}, \quad l_{ij} = \begin{cases} a_{ij}, & \text{se } i > j \\ 0, & \text{se } i \leq j \end{cases}, \quad u_{ij} = \begin{cases} a_{ij}, & \text{se } i < j \\ 0, & \text{se } i \geq j \end{cases} \tag{5}$$

Então, A = D + (L + U)

A escolha M = D e N = L + U resulta no método de Jacobi.

Do teorema 1 conclui-se que, se

$$\left\|D^{-1}(L + U)\right\| < 1 \tag{6}$$

a sequência definida pela iteração

$$x^{(k+1)} = D^{-1}\left(-(L + U) x^{(k)} + b\right), \quad (k = 0,1,\ldots) \tag{7}$$

converge para a solução de (1), qualquer que seja $x^{(0)} \in R^n$.

Naturalmente que (6) e (7) pressupõem que D é invertível. Mas, sendo A invertível é sempre possível por troca de linhas transformá-la numa matriz cujos elementos da diagonal são não nulos.

Um caso importante e frequente é o dos sistemas dispersos cuja matriz dos coeficientes é ***estritamente diagonal dominante***. A convergência do método de Jacobi é então garantida.

Teorema 3:

Se a matriz dos coeficientes do sistema *Ax = b* de **n** equações em **n** incógnitas é ***estritamente diagonal dominante***, então o método de Jacobi converge. Isto é, se

$$\left|a_{ii}\right| \geq \sum_{j \neq i} \left|a_{ij}\right|$$

As componentes da iteração $x^{(k+1)}$ de (7) são dadas por

$$x^{(k+1)} = -\sum_{\substack{j=1 \\ j\neq i}}^{n} a'_{ij}\, x_j^{(k)} + b'_i, \quad (i=0,1,2,\ldots,n) \tag{8}$$

onde,

$$a'_{ij} = \frac{a_{ij}}{a_{ii}} \text{ e } b'_i = \frac{b_i}{a_{ii}}.$$

7.3.2. Algoritmo para o método de Jacobi

Objetivo: resolução de Ax = b supondo satisfeitas as condições de convergência (x é a solução)

Parâmetros de entrada: $x^{(0)}$, ε e kmax

para i **de** 1 **até** n **fazer**

$x_i \leftarrow x_i^{(0)}$

fim_para

$k \leftarrow 0$

repita

para i **de** 1 **até** n **fazer**

$y_i \leftarrow x_i$

fim_para

para i **de** 1 **até** n **fazer**

$$x_i \leftarrow b'_i - \sum_{\substack{j=1 \\ j\neq i}}^{n} a'_{ij}\, y_j$$

fim_para

$k \leftarrow k + 1$

se $\left((\|x-y\|_\infty < \varepsilon \|x\|_\infty) \text{ ou } (k = kmax)\right)$ **então**

interromper

fim_se

fim_repita

7.4. Método de Gauss Seidel

7.4.1. Fórmula geral

Considere-se novamente as matrizes D, L e U, definidas em (5). Tem-se A = (D + L) + U. A escolha M = D + L e N = U dá o método de Gauss Seidel.

Sendo

$(D + L)\, x^{(k+1)} = -U\, x^{(k)} + b, \quad (k = 0, 1, \ldots)$

obtém-se

$$x^{(k+1)} = D^{-1} (-L\, x^{(k+1)} - U\, x^{(k)} + b), \quad (k = 0, 1, \ldots) \tag{9}$$

Se

$$\left\| (D + L)^{-1} U \right\| < 1 \tag{10}$$

então a sequência definida por (9) converge para a solução de (1) qualquer que seja $x^{(0)} \in R^n$.

Se A for ***estritamente diagonal dominante*** é garantida a convergência para o método de Gauss Seidel qualquer que seja $x^{(0)} \in R^n$.

As componentes de $x^{(k+1)}$ de (9) são dadas por

$$x_i^{(k+1)} = -\sum_{j=1}^{i-1} a'_{ij}\, x_j^{(k+1)} - \sum_{j=i+1}^{n} a'_{ij}\, x_j^{(k)} + b'_i, \quad (i = 0,1,2,\ldots,n) \tag{11}$$

Para interpretar a diferença entre (11) e (8) note-se que no método de Gauss Seidel, no cálculo da componente ***i*** da iteração ***k+1*** são usadas as primeiras ***i-1*** componentes já "atualizadas".

7.4.2. Algoritmo para o método de Gauss Seidel

Objetivo: Resolução de Ax = b supondo satisfeitas as condições de convergência (x é a solução)

Parâmetros de entrada: $x^{(0)}$, ε e kmax

para i **de** 1 **até** n **fazer**

$x_i \leftarrow x_i^{(0)}$

fim_para

$k \leftarrow 0$

repita

para i **de** 1 **até** n **fazer**

$y_i \leftarrow x_i$

fim_para

para i **de** 1 **até** n **fazer**

$$x_i \leftarrow b'_i - \sum_{j=1}^{i-1} (a'_{ij}\, x_j) - \sum_{j=i+1}^{n} (a'_{ij}\, y_j)$$

fim_para

$k \leftarrow k + 1$

se $\left((\|x - y\|_\infty < \varepsilon \|x\|_\infty) \text{ ou } (k = kmax) \right)$ **então**

interromper

fim_se

fim_repita

7.5. Exemplo

Considere-se o seguinte sistema de equações lineares

$$\begin{cases} 7x_1 - 3x_3 + x_5 = 1 \\ 2x_1 + 8x_2 = 1 \\ x_3 = 1 \\ 3x_1 + 5x_4 = 1 \\ x_2 + 4x_5 = 1 \\ 2x_4 + 6x_6 = 1 \end{cases}$$

A matriz dos coeficientes é estritamente diagonal dominante e assim haverá convergência para a solução do sistema usando o método de Jacobi ou o de Gauss Seidel.

Tomando $x^{(0)} = [1/7 \;\; 1/8 \;\; 1 \;\; 1/5 \;\; 1/4 \;\; 1/6]^T$ foram obtidos os seguintes resultados:

Iter.	x_i	Jacobi	Gauss Seidel	Iter.	x_i	Jacobi	Gauss Seidel
1	1	0.607143	0.607143	4	1	0.606186	---
	2	0.089286	-0.026786		2	-0.027582	---
	3	1.000000	1.000000		3	1.000000	---
	4	0.285714	0.564286		4	0.563712	---
	5	0.281250	0.243304		5	0.243104	---
	6	0.233333	0.354762		6	0.355400	---
2	1	0.611607	0.606186	...		...	...
	2	-0.026786	-0.026547				
	3	1.000000	1.000000				
	4	0.564286	0.563712				
	5	0.272322	0.243363				
	6	0.354762	0.354571				
3	1	0.610332	0.606195	10	1	0.606195	---
	2	-0.027902	-0.026549		2	-0.026549	---
	3	1.000000	1.000000		3	1.000000	---
	4	0.566964	0.563717		4	0.563717	---
	5	0.243304	0.243363		5	0.243363	---
	6	0.354762	0.354572		6	0.354572	---

7.6. Eficiência

Se a matriz A em Ax = b (1) tiver *p* elementos não nulos, então cada iteração dos métodos de Jacobi e Gauss Seidel requer *p-n* adições/subtrações e *p-n* multiplicações/divisões. Donde, *k* iterações requerem *k(p-n)* adições/subtrações (e multiplicações/divisões). Adicionalmente, no cálculo dos (a')´s e dos (b')´s tem-se *p+n* divisões.

Definindo

$$\alpha = p/n^2 \Rightarrow p = \alpha.n^2$$

o número 100α dá a percentagem de elementos não nulos de A, portanto uma medida da dispersão dos elementos de A. Donde $k(p-n) = k\alpha n^2 - kn \approx k\alpha n^2$ se n for grande.

No sistema associado ao exemplo dado (7.5) verifica-se que n = 6 e p = 12, logo α = 12/36 = 1/3.

8. Interpolação polinomial

8.1. Introdução

Seja f uma função real definida em [a, b] $\subset$ R, sendo conhecidos os seus valores nos pontos x_0, x_1, …, $x_n \in$ [a, b]. Suponha-se que se pretende calcular o valor não tabulado f(x), sendo x $\in$ [a, b]. Por exemplo, dada a tabela de valores da função $\log_{10}$ seguinte

x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
$\log_{10}$(x)	0.32222	0.34242	0.36173	0.38021	0.39794	0.41497	0.43136	0.44716	0.46240

considere-se os seguintes problemas:

➔ calcular $log_{10}(2.45)$;

➔ determinar ***x*** tal que $log_{10}(x) = 0.4$.

Qualquer um destes problemas pode ser resolvido por *interpolação*. Em linhas gerais, este processo consiste em obter uma aproximação para o valor que se pretende conhecer "representando" a função f por uma função "simples", a *função interpoladora*, que assume os mesmos valores que f para certos valores do argumento em [a, b].

Um caso particular de interpolação com grande importância devido ao grande número de aplicações é a interpolação polinomial. Os polinómios interpoladores constituem meios de aproximação de funções muito usados. Além disso, fórmulas desenvolvidas para interpolação polinomial estão na base do desenvolvimento de muitos métodos numéricos para o cálculo de integrais e resolução de equações diferenciais.

8.2. Polinómio interpolador

8.2.1. Definição

Seja f $\in$ C([a,b]) e $x_i \in$ [a, b] (i = 0, 1, …, n). Um polinómio *p* que assume os mesmos valores que f nos pontos x_0, x_1, …, x_n, isto é, que satisfaz

$$p(x_i) = f(x_i), \quad (i = 0, 1, \ldots, n)$$

chama-se ***polinómio interpolador*** *de f nos pontos* $x_0, x_1, \ldots, x_n$.

Exemplo:

Considere-se a tabela de $\log_{10}$ anterior. Para se obter estimativas para $\log_{10}(2.45)$, vai-se "representar" $\log_{10}$ por diferentes polinómios interpoladores.

Começar por calcular o polinómio p_3 de grau menor ou igual a 3, interpolador de $\log_{10}$ nos pontos 2.3, 2.4, 2.5 e 2.6. De acordo com a definição anterior ter-se-á

$$p_3(2.3) = 0.36173,\ \ p_3(2.4) = 0.38021,\ \ p_3(2.5) = 0.39794,\ \ p_3(2.6) = 0.41497.$$

Isto é, se $p_3(x) = a_0 + a_1\, x + a_2\, x^2 + a_3\, x^3$ então

$$\begin{cases} a_0 + 2.3\,a_1 + 5.29\,a_2 + 12.167\,a_3 = 0.36173 \\ a_0 + 2.4\,a_1 + 5.76\,a_2 + 13.824\,a_3 = 0.38021 \\ a_0 + 2.5\,a_1 + 6.25\,a_2 + 15.625\,a_3 = 0.39794 \\ a_0 + 2.6\,a_1 + 6.76\,a_2 + 17.576\,a_3 = 0.41497 \end{cases}$$

Sendo o sistema possível e determinado tal polinómio existe e é único. Assim,

$$p_3(x) = -0.33540 + 0.50502\, x - 0.09750\, x^2 + 0.00833\, x^3\ ,$$

é o polinómio de grau menor ou igual a 3 interpolando $\log_{10}$ nos pontos 2.3, 2.4, 2.5 e 2.6.

Tem-se então $\log_{10}(2.45) \approx p_3(2.45) = 0.38916$. Sendo $\log_{10}(2.45) = 0.38916608...$ o erro na aproximação calculada não excede 0.7×10^{-5}.

Problema:

Dado um conjunto de pontos,

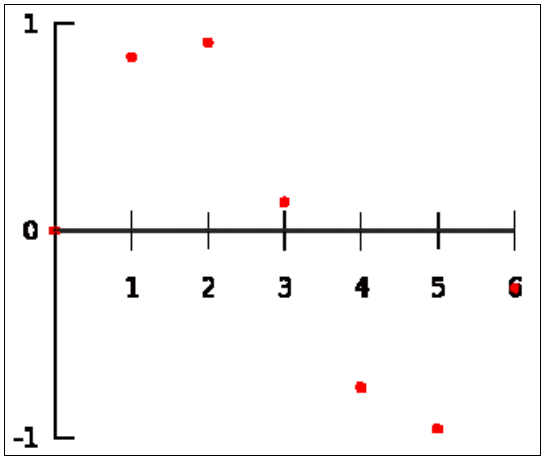
$$(x_0, y_0),\ (x_1, y_1),\ \ldots,\ (x_n, y_n)$$

com $x_i \neq x_j$ (para $i \neq j$) e $i, j = 0, 1, \ldots, n$,

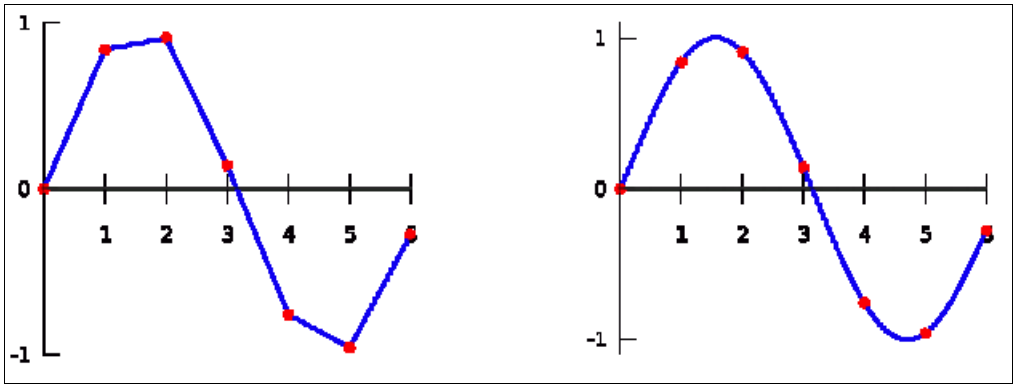
determinar uma **função interpoladora f** tal que

$$f(x_i) = y_i,\ i = 0, 1, \ldots, n$$

Por exemplo, dado o conjunto de pontos,
duas possíveis soluções seriam



Terminologia associada a esta problemática:



➔ Os valores $x_0, x_1, \ldots, x_n$ chamam-se **nós de interpolação** e os respetivos $y_0, y_1, \ldots, y_n$ são os **valores nodais**.

➔ O conjunto $\{ (x_i, y_i), i = 0, 1, \ldots, n \}$ chama-se **suporte de interpolação**.

➔ $\{ f(x_i) = y_i , \ i = 0, 1, \ldots, n \}$ é a **função de interpolação** nesse suporte.

Existem vários **tipos de funções de interpolação**, tais como:

➔ ***Interpolação polinomial***

$$f(x) = a_n x^n + \ldots + a_1 x + a_0$$

➔ ***Interpolação trigonométrica***

$$f(x) = a_{-M} e^{-iMx} + \ldots + a0 + \ldots + a_M e^{iMx}$$

onde,

M é um inteiro igual a **n/2** se **n** é *par* e **(n-1)/2** se **n** é *ímpar*,

i é a unidade imaginária

➔ ***Interpolação racional***

$$f(x) = \frac{a_k x^k + \ldots + a_1 x + a_0}{a_{k+1} x^n + \ldots + a_{k+n} x + a_{k+n+1}}$$

8.2.2. Polinómios

Um **polinómio de grau n** é uma função da forma,

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \quad (a_n \neq 0, \text{ para } n > 0)$$

onde $a_0, a_1, \ldots, a_n$ são os **coeficientes** reais do polinómio.

O problema da determinação de zeros de um polinómio p pode ser visto como o de calcular as raízes da equação p(x) = 0.

Teorema (Fundamental da Álgebra):

Seja *p* um polinómio de grau $n \geq 1$ definido pela expressão anterior.

Então, existe $\alpha \in R$ tal que $p_n(\alpha) = 0$.

Se α é um zero real de $p_n(x)$ então $p_n(x) = (x - \alpha)\, q_{n-1}(x)$.

8.2.3. Cálculo de valores de um polinómio

Como calcular o valor de um polinómio num dado ponto ?

Seja *p* um polinómio de grau *n* de coeficientes reais definido por

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \quad (a_n \neq 0, \text{ para } n > 0).$$

Pretende-se calcular $p_n(y)$, $y \in R$. Ao usar-se

$$p_n(y) = a_n y^n + a_{n-1} y^{n-1} + \ldots + a_1 y + a_0 ,$$

serão efetuadas *n* adições/subtrações e *2n-1* multiplicações/divisões.

Mas $p_n(x)$ pode ser escrita na forma

$$p_n(x) = ((\ldots(a_n x + a_{n-1}) x + \ldots + a_2) x + a_1) x + a_0 ,$$

que representa a chamada *forma encaixada* do polinómio e é a base do método de Horner para o cálculo de valores do polinómio, que requer *n* adições/subtrações e *n* multiplicações/divisões.

Por exemplo,

$p_3(x) = a_3 x^3 + a_2 x^2 + a_1 x + a_0$ { n + (n-1) + ... + 2 + 1 = n (n+1) / 2 = 6 multiplicações }

$p_3(x) = ((a_3 x + a_2) x + a_1) x + a_0$ { n = 3 multiplicações }

Algoritmo:

{ Objetivo: cálculo do valor de $p_n(x) = ((\ldots(a_n x + a_{n-1}) x + \ldots + a_2) x + a_1) x + a_0$ }

{ parâmetros de entrada: $a_0, a_1, \ldots, a_n , z \in R$ }

{ parâmetros de saída: polinómio = $((\ldots(a_n z + a_{n-1}) z + \ldots + a_2) z + a_1) z + a_0$ }

polinómio $\leftarrow a_n$

para k **desde** (n-1) **até** 0 **fazer**

polinómio $\leftarrow$ polinómio * z + a_k

fim_para

Complexidade:

n multiplicações e **n** adições

Algoritmo (método de Horner):

{ Objetivo: calcular $p_n(z)$, *valor de um polinómio* de grau **n** *no ponto* **z** }

{ parâmetros de entrada: $a_0, a_1, \ldots, a_n, z \in R$ }

{ parâmetros de saída: $c_0 = p_n(z)$ }

$c_n \leftarrow a_n$

para k **desde** (n-1) **até** 0 **fazer**

$c_k \leftarrow a_k + z\, c_{k+1}$

fim_para

{ $c_0 = p_n(z)$ }

Exemplo (método de Horner):

Calcular $p_5(x) = x^5 - 6\,x^4 + 8\,x^3 + 8\,x^2 + 4\,x - 40$, para $x = 3$

$c_5 = a_5 = 1$

$c_4 = a_4 + 3\,c_5 = -6 + 3 = -3$

$c_3 = a_3 + 3\,c_4 = 8 - 9 = -1$

$c_2 = a_2 + 3\,c_3 = 8 - 3 = 5$

$c_1 = a_1 + 3\,c_2 = 4 + 15 = 19$

$c_0 = a_0 + 3\,c_1 = -40 + 57 = 17$

Logo, $p_5(3) = c_0 = 17$.

Seja

$p_n(x) = a_n\,x^n + a_{n-1}\,x^{n-1} + \ldots + a_1\,x + a_0$

um polinómio de grau **n** e **z** um número real. Então,

$p_n(x) = (x - z)\,q_{n-1}(x) + r$

onde **q** é um polinómio de grau **n-1** e **r** uma constante ($r = 0$ se e só se z é um zero de p).

Seja

$q_{n-1}(x) = b_{n-1}\,x^{n-1} + b_{n-2}\,x^{n-2} + \ldots + b_1\,x + b_0$.

Então, a expressão $p_n(x) = (x - z)\,q_{n-1}(x) + r$ pode ser escrita da seguinte forma:

$a_n\,x^n + a_{n-1}\,x^{n-1} + \ldots + a_1\,x + a_0 = (x - z)\,(b_{n-1}\,x^{n-1} + b_{n-2}\,x^{n-2} + \ldots + b_1\,x + b_0) + r$

donde, e igualando os coeficientes de potências de x do mesmo grau, obtém-se

$b_{n-1} = a_n$,

$b_k = a_{k+1} + z\,b_{k+1}$ $(k = n-2, n-3, \ldots, 0)$,

$r = a_0 + z\,b_0$.

Algoritmo (fatorização de um polinómio):

{ Objetivo: fatorizar $p_n(x) = (x - z)\, q_{n-1}(x) + r$, onde $z \in R$ e $p_n(x) = a_n x^n + ... + a_1 x + a_0$ }

$b_{n-1} \leftarrow a_n$

para k **desde** (n – 2) **até** 0 **fazer**

$b_k \leftarrow z\, b_{k+1} + a_{k+1}$

fim_para

$r \leftarrow z\, b_0 + a_0$

Exemplo (método da fatorização):

Sendo $p_5(3) = x^5 - 6\, x^4 + 8\, x^3 + 8\, x^2 + 4\, x - 40$, fatorizar $p_5(x) = (x - z)\, q_{n-1}(x) + r$, para z = 3

$b_4 = a_5 = 1$

$b_3 = 3\, b_4 + a_4 = 3 - 6 = -3$

$b_2 = 3\, b_3 + a_3 = -9 + 8 = -1$

$b_1 = 3\, b_2 + a_2 = -3 + 8 = 5$

$b_0 = 3\, b_1 + a_1 = 15 + 4 = 19$

$r = 3\, b_0 + a_0 = 57 - 40 = 17$

Logo, $q_4(x) = x^4 - 3\, x^3 - x^2 + 5\, x + 19$.

E assim, $p_5(x) = (x - 3)\, (x^4 - 3\, x^3 - x^2 + 5\, x + 19) + 17$.

8.3. Interpolação polinomial de Lagrange

Os *polinómios* são excelentes candidatos a *funções interpoladoras*, porque:

- ➔ O cálculo dos valores é realizável em ordem linear ao número de multiplicações e adições.
- ➔ As operações de derivação e primitivação são simples e podem ser facilmente programáveis.
- ➔ Aproximam tanto quando se queira qualquer função contínua num intervalo finito (Teorema de Weierstrass).

Sempre que as funções de interpolação consideradas são polinómios então está-se perante **Interpolação Polinomial**.

Problema:

Dado um **suporte de interpolação** com **n+1** pontos,

$\{ (x_i, y_i),\ i = 0, 1, ..., n \}$

encontrar um **polinómio de grau menor ou igual a n** tal que,

$y_i = p_n(x_i),\ i = 0, 1, ..., n$

Questões:

- ➔ ***Existe*** sempre um polinómio que satisfaz as condições acima ?
- ➔ Caso exista, é ***único*** ?

Teorema (da existência e unicidade):

Seja P_n o conjunto dos polinómios de *grau menor ou igual a* n.

Dados $n+1$ *pontos suporte distintos* $(x_i, f(x_i))$, $i = 0, 1, \ldots, n$,

existe um e um só polinómio $p_n \in P_n$ tal que,

$$p_n(x_i) = f(x_i), \quad i = 0, 1, \ldots, n$$

Observações:

➔ O teorema anterior mostra-nos que o polinómio interpolador *existe e é único* (podem ser deduzidas *várias fórmulas* para ele, mas todas representam **o mesmo polinómio interpolador**).

8.3.1. Fórmula de Lagrange

Definição:

Os polinómios de grau **n** dados por,

$$L_k(x) = \prod_{\substack{i=0 \\ i \neq k}}^{n} \frac{x - x_i}{x_k - x_i}, \quad k = 0, 1, \ldots, n$$

são designados por **polinómios de Lagrange** *associados aos nós* $x_0, x_1, \ldots, x_n$.

Teorema:

O ***polinómio interpolador*** p_n de grau menor ou igual a **n** que interpola os valores nodais $y_0, y_1, \ldots, y_n$ nos nós distintos $x_0, x_1, \ldots, x_n$ é dado por,

$$p_n(x) = \sum_{k=0}^{n} L_k(x)\, y_k \ .$$

Exemplo:

Construir o polinómio interpolador de grau menor ou igual a 3 que interpola os seguintes valores:

x_i	0	1	3	4
y_i	1	-1	1	2

Os ***polinómios de Lagrange*** associados aos nós ($x_0 = 0$, $x_1 = 1$, $x_2 = 3$, $x_3 = 4$) obtêm-se diretamente da ***definição*** anterior,

$$L_0(x) = \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} = -\frac{1}{12}(x-1)(x-3)(x-4)$$

$$L_1(x) = \frac{(x - x_0)(x - x_2)(x - x_3)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)} = \frac{1}{6}x(x-3)(x-4)$$

$$L_2(x) = \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} = -\frac{1}{6}x(x-1)(x-4)$$

$$L_3(x) = \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} = \frac{1}{12}x(x-1)(x-3)$$

Assim sendo, nas condições do teorema, o **polinómio interpolador** é dado por:

$$p_3(x) = \sum_{k=0}^{3} L_k(x)\, y_k = L_0(x)y_0 + L_1(x)y_1 + L_2(x)y_2 + L_3(x)y_3 = L_0(x) - L_1(x) + L_2(x) + 2.L_3(x) =$$

$$= -\frac{1}{12}(x-1)(x-3)(x-4) - \frac{1}{6}x(x-3)(x-4) - \frac{1}{6}x(x-1)(x-4) + \frac{2}{12}x(x-1)(x-3)$$

Algoritmo (fórmula de Lagrange):

{ Objetivo: cálculo de $p_n(z)$ sendo p_n interpolador de f nos pontos distintos $x_0, x_1, \ldots, x_n$ }

```
q ← 0
para i desde 0 até n fazer
   p ← 1
   para j desde 0 até n fazer
      se j ≠ i então
         p ← p.(z - x_j) / (x_i - x_j)
      fim_se
   fim_para
   q ← q + p.y_i
fim_para
```

Observação:

A fórmula de Lagrange ***pode não ser a representação mais conveniente*** do polinómio interpolador, fundamentalmente por duas razões:

1. É possível obter este polinómio com ***menos operações aritméticas*** do que as requeridas por aquela fórmula (o cálculo de um valor do polinómio interpolador requer ***n(n+2)*** adições/subtrações e ***n(n+1)*** multiplicações/divisões);
2. Os polinómios de Lagrange ***estão associados a um conjunto de nós*** e uma mudança de posição, ou do número destes, altera completamente estes polinómios.

8.3.2. Fórmula de Newton

Definição:

A fórmula de Newton para polinómios de grau **n** é dada por,

$$p_n(x) = a_0 + a_1(x - c_1) + a_2(x - c_1)(x - c_2) + \ldots + a_n(x - c_1)(x - c_2)\ldots(x - c_n)$$

onde os parâmetros c_i, i = 1, 2, ... , n são chamados **centros do polinómio**.

Construção da Fórmula de Newton:

Considerando os **nós** $x_0, x_1, \ldots, x_{n-1}$ como *centros do polinómio*, temos:

$$p_n(x) = a_0 + a_1 (x - x_0) + a_2 (x - x_0) (x - x_1) + \ldots + a_n (x - x_0) (x - x_1) \ldots (x - x_{n-1})$$

Os coeficientes $a_0, a_1, \ldots, a_n$ vão ser determinados de modo que p_n seja o *polinómio interpolador* nos nós $x_0, x_1, \ldots, x_n$ dos valores nodais $y_0, y_1, \ldots, y_n$:

$$p_n(x_0) = y_0 \; ; \; p_n(x_1) = y_1 \; ; \ldots ; \; p_n(x_n) = y_n$$

ou, se os valores nodais y_i forem ***valores nodais de uma <u>função</u> f*** tem-se,

$$p_n(x_i) = f(x_i), \quad i = 0, 1, \ldots, n$$

Assim, a partir de,

$$p_n(x) = a_0 + a_1 (x - x_0) + a_2 (x - x_0) (x - x_1) + \ldots + a_n (x - x_0) (x - x_1) \ldots (x - x_{n-1})$$

e fazendo sucessivamente $x = x_0$, $x = x_1$, ..., $x = x_n$ obtém-se os **coeficientes**:

$$a_0 = f(x_0)$$

$$a_1 = \frac{f(x_1) - a_0}{x_1 - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

$$a_2 = \frac{f(x_2) - a_0 - a_1(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

. . .

$$a_n = \frac{f(x_n) - a_0 - a_1(x_n - x_0) - a_2(x_n - x_0)(x_n - x_1) - \ldots - a_{n-1}(x_n - x_0)\ldots(x_n - x_{n-2})}{(x_n - x_0)(x_n - x_1)\ldots(x_n - x_{n-1})} = \ldots$$

Observação:

Cada coeficiente a_k , $k = 0, 1, \ldots, n$:

➔ pode ser calculado a partir dos a_i , $i = 0, 1, \ldots, k-1$, *já determinados*.

➔ *depende* exclusivamente *dos nós* $x_0, x_1, \ldots, x_n$ e dos respetivos *valores nodais* $y_0, y_1, \ldots, y_n$

$$a_k = f[x_0, x_1, \ldots, x_k]$$

em que

$$f[x_0, x_1, \ldots, x_k]$$

é a *diferença dividida de ordem* k ($k \geq 1$) entre os $k+1$ nós $x_0, x_1, \ldots, x_k$.

Definição:

Para designar a ***diferença dividida de ordem k*** (***k ≥ 1***) entre os ***k+1*** nós $x_0, x_1, \ldots, x_k$, são utilizadas indistintamente *duas notações*:

$$D^k f(x_i) \equiv f\left[x_i, x_{i+1}, \ldots, x_{i+k}\right]$$

onde

$$D^k f(x_i) = \frac{D^{k-1} f(x_{i+1}) - D^{k-1} f(x_i)}{x_{i+k} - x_i}$$

ou

$$f\left[x_i, x_{i+1}, \ldots, x_{i+k}\right] = \frac{f\left[x_{i+1}, \ldots, x_{i+k}\right] - f\left[x_i, \ldots, x_{i+k-1}\right]}{x_{i+k} - x_i}$$

Teorema:

Os *coeficientes* a_k , ***k = 0, 1, ..., n*** do polinómio p_n de grau menor ou igual a **n**, na *forma de Newton* que interpola os valores $f(x_0), f(x_1), \ldots, f(x_k)$ nos *nós distintos* $x_0, x_1, \ldots, x_k$ são dados *indutivamente* pela expressão:

$$a_k = f\left[x_0, x_1, \ldots, x_k\right] = \frac{f\left[x_1, \ldots, x_k\right] - f\left[x_0, \ldots, x_{k-1}\right]}{x_k - x_0}$$

Assim, o **Polinómio Interpolador com Diferenças Divididas** tem a forma:

$$p_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) +$$
$$+ \ldots + f[x_0, x_1, \ldots, x_n](x - x_0)(x - x_1)\ldots(x - x_{n-1})$$

Uma tabela de diferenças divididas de uma função f pode ser escrita da forma que segue (denotando-se por $f_{i,i+j}$ a diferença $f[x_i, \ldots, x_{i+j}]$).

x	D^0 / f[]	D^1 / f [,]	D^2 / f [, ,]	D^3 / f [, , ,]	...
x_0	$f(x_0)$				
		$f_{0,1}$			
x_1	$f(x_1)$		$f_{0,2}$		
		$f_{1,2}$		$f_{0,3}$	
x_2	$f(x_2)$		$f_{1,3}$		...
		$f_{2,3}$		...	
x_3	$f(x_3)$		...		...
		...		$f_{n-3,n}$	
...	...		$f_{n-2,n}$		
		$f_{n-1,n}$			
x_n	$f(x_n)$				

Algoritmo (Diferenças divididas):

{ Objetivo: construir uma tabela de diferenças divididas de f por diagonais ascendentes sucessivas }

$f_0 \leftarrow f(x_0)$

para i **desde** 1 **até** n **fazer**

$f_i \leftarrow f(x_i)$

para j **desde** (i-1) **até** 0 **fazer**

$f_{j,i} \leftarrow (f_{j,i-1} - f_{j+1,i}) \ / \ (x_j - x_i)$

fim_para

fim_para

Exemplo:

Determinar o *polinómio interpolador*, na *forma de Newton*, que interpola os seguintes pontos:

x_i	0	1	3	4
y_i	1	-1	1	2

A tabela de diferenças dividida para este caso é a seguinte:

x	**D^0 / f[]**	**D^1 / f [,]**	**D^2 / f [, ,]**	**D^3 / f [, , ,]**
$x_0 = 0$	1			
		$f_{0,1} = -2$		
$x_1 = 1$	-1		$f_{0,2} = 1$	
		$f_{1,2} = 1$		$f_{0,3} = -1/4$
$x_2 = 3$	1		$f_{1,3} = 0$	
		$f_{2,3} = 1$		
$x_3 = 4$	2			

$$f_{0,1} = f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{-1 - 1}{1 - 0} = \frac{-2}{1} = -2$$

$$f_{1,2} = f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{1 - (-1)}{3 - 1} = \frac{2}{2} = 1$$

$$f_{2,3} = f[x_2, x_3] = \frac{f(x_3) - f(x_2)}{x_3 - x_2} = \frac{2 - 1}{4 - 3} = \frac{1}{1} = 1$$

$$f_{0,2} = f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{f_{1,2} - f_{0,1}}{x_2 - x_0} = \frac{1 - (-2)}{3 - 0} = \frac{3}{3} = 1$$

$$f_{1,3} = f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1} = \frac{f_{2,3} - f_{1,2}}{x_3 - x_1} = \frac{1 - 1}{4 - (-1)} = \frac{0}{5} = 0$$

$$f_{0,3} = f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0} = \frac{f_{1,3} - f_{0,2}}{x_3 - x_0} = \frac{0 - 1}{4 - 0} = -\frac{1}{4}$$

Assim calculados os coeficientes do polinómio interpolador na forma de Newton,

$a_0 = f[x_0] = 1$

$a_1 = f[x_0, x_1] = -2$

$a_2 = f[x_0, x_1, x_2] = 1$

$a_3 = f[x_0, x_1, x_2, x_3] = -1/4$

$$p_3(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) \Leftrightarrow$$

$$p_3(x) = 1 + (-2)(x - 0) + 1(x - 0)(x - 1) + \left(-\frac{1}{4}\right)(x - 0)(x - 1)(x - 3) \Leftrightarrow$$

$$p_3(x) = 1 - 2x + x(x - 1) - \frac{1}{4}x(x - 1)(x - 3) .$$

Observações:

➔ A *ordem* pela qual os nós são tomados é arbitrária.

➔ Se é necessário *acrescentar mais algum nó* aos anteriores, basta colocá-lo no fundo da tabela e calcular mais uma linha de valores (as diferenças divididas já obtidas não seriam afetadas).

➔ Se os valores nodais forem os de uma *função*, é possível estabelecer uma *ligação* importante entre as *diferenças divididas* de ordem ***k*** e a *derivada* da mesma ordem dessa função.

Teorema:

Sejam $f \in C^n([a,b])$ e $x_0, x_1, \ldots, x_n$ *nós distintos* no intervalo [a,b].

Então existe um $\xi \in (a,b)$ tal que,

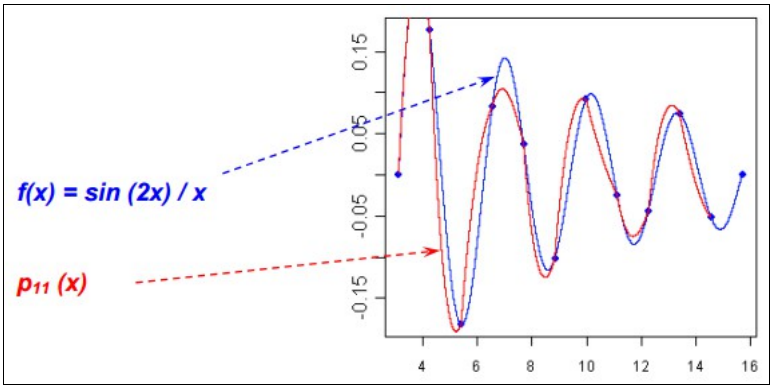
$$f[x_0, x_1, \ldots, x_n] = \frac{1}{n!} f^{(n)}(\zeta)$$

Deste modo, se os valores nodais forem os valores nodais de uma função, este teorema estabelece uma *relação importante* entre as ***diferenças divididas*** de ordem ***n*** e a ***derivada*** da mesma ordem dessa função.

8.3.3. Erros de Interpolação Polinomial

Que **erro** se comete quando se *interpola uma função* por um ***polinómio*** de grau menor ou igual a *n* utilizando o valor da função em ***n+1*** nós distintos?

Por exemplo,



Teorema:

Sejam $f \in C^{n+1}([a, b])$ e p_n o polinómio de grau menor ou igual a **n** que interpola f nos nós distintos $x_0, x_1, \ldots, x_n$, contidos em [a,b]. Então para qualquer $z \in [a,b]$ *existe* um valor $\xi \in (a, b)$, dependente de $x_0, x_1, \ldots, x_n, z$ e de ***f*** tal que

$$e_n(z) \equiv f(z) - p_n(z) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}(z - x_0)(z - x_1)\ldots(z - x_n) .$$

Estimativa do Erro de Interpolação

Como em,

$$e_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}\omega(x), \quad \text{com } \omega(x) = (x - x_0)(x - x_1)\ldots(x - x_n)$$

o valor de ξ é desconhecido, temos de calcular um ***limite superior*** para estimativa do valor do erro.

Para o caso particular da função a interpolar, procuramos um ***majorante*** em $[x_0, x_n]$,

$$\left| f^{(n+1)}(x) \right| \leq M_{n+1}$$

e considerando ***h*** o *espaçamento máximo* entre dois nós consecutivos,

$$\left| \omega(x) \right| \leq \frac{n!}{4} h^{n+1}, \quad x \in [x_0, x_n]$$

temos,

$$\left| e_n(x) \right| \leq \left| \omega(x) \right| \frac{M_{n+1}}{(n+1)!}$$

ou,

$$\left| e_n(x) \right| \leq \frac{1}{4(n+1)} M_{n+1} h^{n+1}.$$

Comportamento do Erro de Interpolação

Analisando

$$e_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}\omega(x)$$

verifica-se que o erro de interpolação depende de:

➔ o *número de nós* considerado,

➔ o comportamento da *derivada de ordem* ***n+1*** da função,

➔ o comportamento do **polinómio** ω de grau ***n+1***.

9. Aproximação polinomial

9.1. Introdução

Em linhas gerais pode dizer-se que aproximar uma função é representá-la por uma outra mais "simples". Há necessidade de aproximar uma função quando a forma com ela é definida dificulta ou impossibilita a resolução de problemas matemáticos envolvendo essa função. São os casos, por exemplo, de funções conhecidas por uma tabela de alguns dos seus valores, funções definidas como soluções de equações, ou definidas explicitamente por expressões envolvendo funções transcendentes. Pode-se estar interessado em calcular o valor do integral da função num dado domínio e não conhecer a primitiva da função, ou calcular um (ou mais) zeros da função não existindo uma fórmula que o permita fazer explicitamente, etc.

Uma forma de resolver estes problemas é substituir a função dada por outra mais "simples" (por exemplo, um polinómio) que num subconjunto relevante do seu domínio não se "afasta" muito da função dada.

As classes de funções mais importantes são as dos polinómios (incluindo polinómios segmentados), das funções racionais (quocientes de polinómios) e das funções de Fourier ({sen(nx), cos(nx)}, n = 0, 1, …). Neste texto será considerada apenas a aproximação de funções por polinómios.

Podem ser colocadas três questões. Para que tipo de funções existe uma aproximação polinomial adequada? Como caraterizar uma "boa" aproximação? Como construí-la?

9.2. Conceitos e resultados básicos

O teorema de *Weierstrass* estabelece que para uma certa classe de funções, as contínuas num intervalo fechado de R, existe um polinómio que aproxima a função tão bem quando se queira.

Teorema (da aproximação de Weierstrass):

Seja $[a, b] \subset R$ e $f \in C([a,b])$. Então, qualquer que seja $\varepsilon > 0$ existe $n = n(\varepsilon)$ tal que

$$|f(x) - p_n(x)| < \varepsilon,$$

para todo o $x \in [a, b]$.

É possível construir, de uma maneira eficiente, aproximações polinomiais úteis sob o ponto de vista prático, e estimar o erro de aproximação. No entanto, têm de ser impostas condicionais adicionais sobre a regularidade da função a aproximar.

9.2.1. Métricas, normas e seminormas

Qualquer estudo sobre aproximação de funções pressupõe a existência de uma maneira de medir a "distância" entre duas funções.

Definição:

Seja F um conjunto. Uma aplicação d : F x F $\rightarrow$ R tal que

$\forall$ f, g $\in$ F, d(f,g) = 0 se e só se f = g,

$\forall$ f, g $\in$ F, d(f,g) = d(g,f),

$\forall$ f, g, h $\in$ F, d(f,h) $\leq$ d(f,g) + d(g,h),

é uma ***métrica***. O conjunto F munido duma métrica é um ***espaço métrico***.

Desta definição resulta que

$\forall$ f, g $\in$ F, d(f, g) $\geq$ 0.

Exemplo:

Considere-se o conjunto C([a,b]) das funções reais contínuas num intervalo fechado [a,b] $\subset$ R. A aplicação d : C([a,b]) x C([a,b]) $\rightarrow$ R definida por

$\forall$ f, g $\in$ C([a,b]), $d(f,g) = \max_{x\in[a,b]} |f(x) - g(x)|$,

é uma *métrica*.

Considere-se a função e^x definida em [0,1] e o polinómio p(x) = 1 + (e - 1)x, interpolador de e^x nos pontos 0 e 1. Tem-se

$d(e^x,p) = \max_{x\in[0,1]} |e^x - (1 + (e - 1)x)| \approx 0.21$

Definição:

Seja F um espaço linear. Uma aplicação $\|.\| : F \rightarrow R$ tal que,

$\forall$ f $\in$ F, $\|f\| = 0$ se e só se f = 0,

$\forall$ f $\in$ F, $\forall$ $\alpha \in R$, $\|\alpha f\| = |\alpha| \|f\|$,

$\forall$ f, g $\in$ F, $\|f + g\| \leq \|f\| + \|g\|$,

é chamada de ***norma***. O espaço linear F onde está definida uma norma diz-se ***espaço normado***.

Definição:

Se a aplicação $\|.\| : F \rightarrow R$ satisfaz as duas últimas condições da definição anterior e

$\forall$ f $\in$ F, $\|f\| = 0$ se f = 0,

então f é uma ***seminorma***.

Note-se que se $\|.\|$ é uma *seminorma* em F pode ter-se, para algum f $\neq$ 0, $\|f\| = 0$.

Se $\|.\| : F \rightarrow R$ é uma norma (ou *seminorma*) tem-se

$\forall f \in F, \quad \|f\| \geq 0.$

Seja C[a,b] o espaço linear das funções reais contínuas no intervalo fechado [a,b] $\subset$ R.

Exemplo:

A aplicação $\|.\|_\infty : C([a,b]) \to R$ definida por

$$\forall f \in C([a,b]), \quad \|f\|_\infty = \max_{x \in [a,b]} |f(x)|,$$

é uma norma chamada *norma uniforme* ou *norma de Chebyshev*.

Exemplo:

A aplicação $\|.\|_2 : C([a,b]) \to R$ definida por

$$\forall f \in C([a,b]), \quad \|f\|_2 = \left\{ \int_a^b w(x) |f(x)|^2 \, dx \right\}^{1/2}, \; w \in C([a,b]) \text{ e } w(x) > 0 \text{ para } x \in [a,b],$$

é uma norma chamada *norma dos quadrados ponderados*.

Exemplo:

Dados n + 1 pontos distintos $x_0, x_1, \ldots, x_n \in [a, b]$, a aplicação $\|.\| : C([a,b]) \to R$ definida por

$$\forall f \in C([a,b]), \quad \|f\| = \max_{0 \le i \le n} |f(x_i)|,$$

é uma *seminorma*.

Exemplo:

Dados n + 1 pontos distintos $x_0, x_1, \ldots, x_n \in [a, b]$, a aplicação $\|.\| : C([a,b]) \to R$ definida por

$$\forall f \in C([a,b]), \quad \|f\| = \left\{ \sum_{i=0}^{n} w(x_i) |f(x_i)|^2 \, dx \right\}^{1/2}, \; w \in C([a,b]) \text{ e } w(x_i) > 0 \; (i = 0, 1, \ldots, n),$$

é uma semin*orma*.

Toda a norma induz uma métrica. Isto é, se F é um espaço normado onde está definida uma norma $\|.\|$, então F é um espaço métrico com a métrica definida por

$$\forall f, g \in F, \quad d(f,g) = \|f - g\|.$$

9.2.2. Melhor aproximação polinomial

Definição:

Seja F um espaço linear de funções onde está definida uma norma $\|.\|$. Se $g \in F$ é uma aproximação para $f \in F$ a

$$\|f - g\|$$

chama-se ***erro de aproximação g*** com respeito à norma $\|.\|$.

Definição:

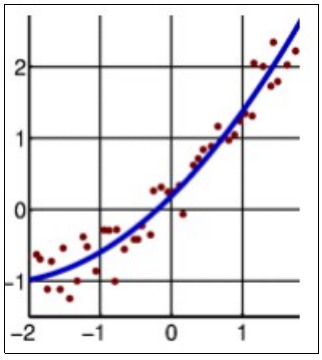
Seja $F \subseteq C([a,b])$ um espaço normado e P_n o conjunto dos polinómios de coeficientes reais de grau menor ou igual a *n*. Então, $p_n \in P_n$ é uma *melhor aproximação polinomial* de grau *n* para uma função $f \in F$ em relação à norma $\|.\|$ de F, se

$$\|f - p_n\| = \inf_{q_n \in P_n} \|f - q_n\|.$$

Por outras palavras, relativamente a uma dada norma, a melhor aproximação polinomial de grau *n* para uma função que minimiza o erro.

Se a norma for de Chebyshev ou dos quadrados ponderados, então existe uma melhor aproximação polinomial e é única. Uma melhor aproximação em relação à norma de Chebyshev é chamada *aproximação minimax*. Uma melhor aproximação em relação à norma dos quadrados ponderados é chamada *aproximação dos mínimos quadrados*.

9.3. Aproximação dos mínimos quadrados para dados discretos

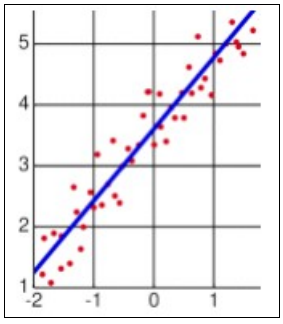


Em muitos casos reais, os ***valores nodais*** que se tem são obtidos experimentalmente, vindo portanto **afetados de erros**. Desta forma, em vez de se tentar construir uma função interpoladora, faz mais sentido procurar a ***função que melhor aproxima*** esses valores.

Seja $\{ (x_i, y_i) \}$, $i = 1, 2, \ldots, m$ um conjunto de pares de números reais onde,

$$y_i \approx f(x_i), \quad i = 1, 2, \ldots, m$$

A partir destes valores, pretende-se construir uma função que, *de alguma forma*, seja *a melhor aproximação* da função ***f(x)***.



Tome-se como ***exemplo o caso linear***, isto é, quando a *função aproximante* pretendida for uma reta $y = a\,x + b$. Para calcular os parâmetros **a** e **b**, podem ser estabelecidos ***diferentes critérios***, tais como:

➔ Minimizar o ***erro máximo***, $\max_{i=1,\ldots,m} \left| y_i - (a\,x_i + b) \right|$

➔ Minimizar a ***soma dos erros***, $\sum_{i=1}^{m} \left| y_i - (a\,x_i + b) \right|$

➔ Minimizar o ***erro quadrático***, $\sum_{i=1}^{m} \left(y_i - (a\,x_i + b) \right)^2$

9.3.1. Funções aproximantes e desvios

No caso geral, o problema consiste em determinar a **função que melhor aproxima** um dado conjunto de pontos $\{ (x_i, y_i) \}$, $i = 1, 2, \ldots, m$.

A **classe das funções aproximantes** é caracterizada por um conjunto de ***parâmetros*** $c_1, \ldots, c_n$. ***Cada função*** da classe é especificada pelos ***valores*** desses parâmetros,

$$f(x) = F(x; c_1, \ldots, c_n)$$

Por exemplo, se pretender-se aproximar os pontos por

➔ uma ***reta***, são dois os parâmetros (c_1 e c_2) e $f(x) = F(x; c_1, c_2) = c_1 + c_2\, x$

➔ uma ***parábola***, são três os parâmetros (c_1, c_2 e c_3) e $f(x) = F(x; c_1, c_2, c_3) = c_1 + c_2\, x + c_3\, x^2$

Para cada classe definem-se os ***desvios***, em relação aos valores y_i dos dados,

$$d_i = y_i - F(x_i; c_1, c_2, \ldots, c_n), \quad i = 1, 2, \ldots, m$$

Em função dos desvios, é necessário decidir qual o ***critério*** a estabelecer. Cada critério define um ***problema de minimização***.

➔ Problema de ***minimax*** (minimização do desvio máximo),

$$\text{minimizar} \quad \max_{i=1,\ldots,m} \left| y_i - F(x_i; c_1, c_2, \ldots, c_n) \right|$$

➔ Problema de minimização (da soma) dos ***desvios absolutos***,

$$\text{minimizar} \quad \sum_{i=1}^{m} \left| y_i - F(x_i; c_1, c_2, \ldots, c_n) \right|$$

➔ Problema de minimização do ***erro quadrático total***,

$$\text{minimizar} \quad \sum_{i=1}^{m} \left(y_i - F\left(x_i; c_1, c_2, \ldots, c_n\right) \right)^2$$

O método de resolução do problema de *minimização do erro quadrático total* chama-se **método dos mínimos quadrados** e a função que o minimiza chama-se **aproximação dos mínimos quadrados.**

9.3.2. Método dos Mínimos Quadrados

Considere-se uma classe de funções,

$$F(x; c_1, c_2, \ldots, c_n) = c_1\, f_1(x) + c_2\, f_2(x) + \ldots + c_n\, f_n(x)$$

onde $f_1(x)$, $f_2(x)$, ..., $f_n(x)$ são funções dadas.

A *aproximação dos mínimos quadrados* consiste na determinação dos **parâmetros $c_1, c_2, \ldots, c_n$** que *minimizam a soma dos quadrados dos desvios*,

$$E(c_1, \ldots, c_n) = \sum_{i=1}^{m} \left(y_i - c_1 f_1(x_i) - c_2 f_2(x_i) - \ldots - c_n f_n(x_i) \right)^2 = \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} c_j f_j(x_i) \right)^2 .$$

Tratando-se de um problema de minimização em R^n, para que $E(c_1, c_2, ..., c_n)$ seja mínimo é necessário que,

$$\nabla E(c_1,\ldots,c_n) = 0 \Leftrightarrow \frac{\partial E}{\partial c_j} = 0, \quad j = 1,\ldots,n$$

Donde se obtém um *sistema de **n** equações a **n** incógnitas*,

$$\begin{cases} c_1 \sum_{i=1}^{m} f_1(x_i)f_1(x_i) + c_2 \sum_{i=1}^{m} f_1(x_i)f_2(x_i) + \ldots + c_n \sum_{i=1}^{m} f_1(x_i)f_n(x_i) = \sum_{i=1}^{m} y_i f_1(x_i) \\ c_1 \sum_{i=1}^{m} f_2(x_i)f_1(x_i) + c_2 \sum_{i=1}^{m} f_2(x_i)f_2(x_i) + \ldots + c_n \sum_{i=1}^{m} f_2(x_i)f_n(x_i) = \sum_{i=1}^{m} y_i f_2(x_i) \\ \qquad\qquad\qquad\qquad\qquad\qquad \ldots \\ c_1 \sum_{i=1}^{m} f_n(x_i)f_1(x_i) + c_2 \sum_{i=1}^{m} f_n(x_i)f_2(x_i) + \ldots + c_n \sum_{i=1}^{m} f_n(x_i)f_n(x_i) = \sum_{i=1}^{m} y_i f_n(x_i) \end{cases}$$

Em certos casos, este sistema tem solução única e permite determinar univocamente os parâmetros $c_1, c_2, ..., c_n$ que caracterizam a ***melhor função aproximante***.

9.3.3. Reta dos Mínimos Quadrados (Reta de Regressão)

No ***caso linear***, o *problema da minimização do erro quadrático*, pretende determinar os valores de **a** e de **b** em,

$F(x; a, b) = a + b\,x$

que minimizam

$$E(a,b) = \sum_{i=1}^{m} \left(y_i - a - b\,x_i\right)^2$$

Para que $E(a, b)$ seja **mínimo** é necessário (e prova-se que também suficiente) que,

$$\nabla E(a,b) = \begin{cases} \frac{\partial E}{\partial a} = 0 \\ \frac{\partial E}{\partial b} = 0 \end{cases}$$

ou seja que,

$$\begin{cases} a\,m + b \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i \\ a \sum_{i=1}^{m} x_i + b \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i\,y_i \end{cases}$$

Assim tem-se um sistema linear com *duas equações* (***equações normais***) e as *duas incógnitas* **a** e **b** que caracterizam a reta pretendida (***reta de regressão***).

Os coeficientes de **a** e de **b** e os termos independentes, obtém-se facilmente pela construção de uma tabela,

x_i	y_i	x_i^2	$x_i\ y_i$
x_1	y_1	x_1^2	$x_1\ y_1$
x_2	y_2	x_2^2	$x_2\ y_2$
...	...	...	...
x_n	y_n	x_n^2	$x_n\ y_n$
$\sum x_i$	$\sum y_i$	$\sum x_i^2$	$\sum x_i\ y_i$

Exemplo:

Para se determinar a ***reta de regressão*** que aproxima os pontos,

x_i	1	2	4	5	7	8	10
y_i	1	2	4	4	5	6	7

constrói-se a ***tabela***

	x_i	y_i	x_i^2	$x_i\ y_i$
	1	1	1	1
	2	2	4	4
	4	4	16	16
	5	4	25	20
	7	5	49	35
	8	6	64	48
	10	7	100	70
Σ	37	29	259	194

Donde se obtém o ***sistema***

$$\begin{cases} a \times 7 + b \times 37 = 29 \\ a \times 37 + b \times 259 = 194 \end{cases}$$

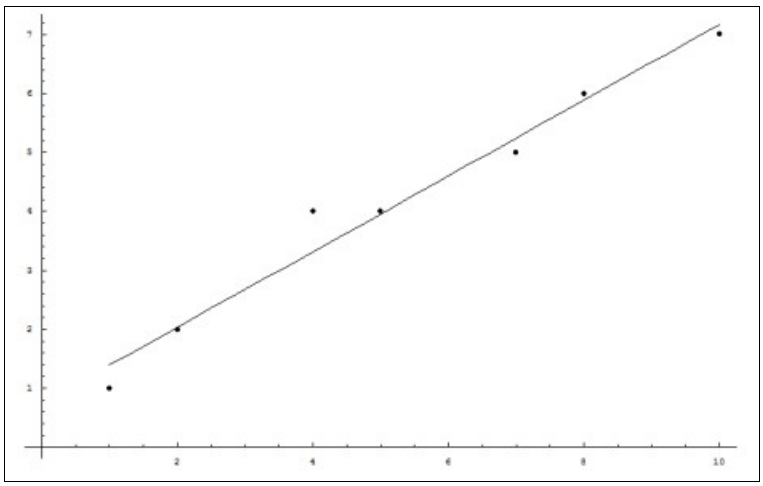
cuja ***solução*** é

a = 0.75

b = 0.6418918918919

o que permite determinar a reta de regressão,

y = 0.75 + 0.6418918918919 x



Em algumas aplicações, são os valores de ***{ x_i }, i = 1, 2, ..., m*** que estão *afetados de erros*, sendo os ***{ y_i }*** considerados exatos. Nesse caso é necessário efetuar uma ***aproximação inversa***.

Assim, dado $\{ x_i, y_i \}, i = 1, 2, ..., m$ um conjunto de pares de números reais onde,

$$x_i \approx g(y_i), i = 1, 2, ..., m$$

podemos calcular uma *aproximação dos mínimos quadrados* para **g(y)**.

Exemplo: no *exemplo anterior*, basta trocar os papéis dos x e y dados,

x_i	1	2	4	5	7	8	10
y_i	1	2	4	4	5	6	7

construindo neste caso a ***tabela***

	x_i	y_i	y_i^2	$y_i\, x_i$
	1	1	1	1
	2	2	4	4
	4	4	16	16
	5	4	16	20
	7	5	25	35
	8	6	36	48
	10	7	49	70
Σ	37	29	**147**	194

donde se obtém o ***sistema***

$$\begin{cases} a \times 7 + b \times 29 = 37 \\ a \times 29 + b \times 147 = 194 \end{cases}$$

cuja ***solução*** é

a = -0.994680851064

b = 1.5159574468085

o que permite determinar a ***reta de regressão inversa***,

x = -0.994680851064 + 1.5159574468085 y

9.3.4. Parábola dos Mínimos Quadrados

Para aproximar o conjunto de pontos por uma ***parábola***, pretende-se determinar os valores de **a**, **b** e **c** em,

$$F(x; a, b, c) = a + b\, x + c\, x^2$$

por forma a *minimizar o erro quadrático total*,

$$E(a,b,c) = \sum_{i=1}^{m} \left(y_i - a - b\, x_i - c\, x_i^2 \right)^2$$

Para que ocorra o ***mínimo*** é necessário (e prova-se que também é suficiente) que,

$$\nabla E(a,b,c) = 0$$

ou seja,

$$\begin{cases} a\,m + b\sum_{i=1}^{m} x_i + c\sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} y_i \\ a\sum_{i=1}^{m} x_i + b\sum_{i=1}^{m} x_i^2 + c\sum_{i=1}^{m} x_i^3 = \sum_{i=1}^{m} x_i\,y_i \\ a\sum_{i=1}^{m} x_i^2 + b\sum_{i=1}^{m} x_i^3 + c\sum_{i=1}^{m} x_i^4 = \sum_{i=1}^{m} x_i^2\,y_i \end{cases}$$

Os coeficientes de **a** e de **b** e os termos independentes, também se obtém pela construção de uma ***tabela***.

Exemplo:

Para o ***exemplo anterior***,

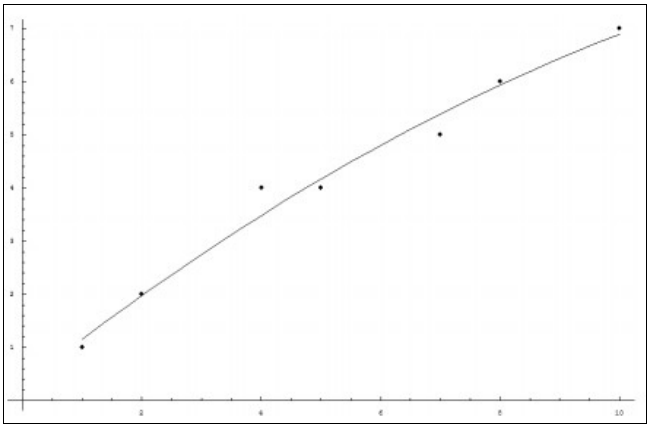
x_i	1	2	4	5	7	8	10
y_i	1	2	4	4	5	6	7

construindo a ***tabela***

	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i\,y_i$	$x_i^2\,y_i$
	1	1	1	1	1	1	1
	2	2	4	8	16	4	8
	4	4	16	64	256	16	64
	5	4	25	125	625	20	100
	7	5	49	343	2401	35	245
	8	6	64	512	4096	48	384
	10	7	100	1000	10000	70	700
Σ	37	29	259	2053	17395	194	1502

donde se obtém o ***sistema***

$$\begin{cases} a \times 7 + b \times 37 + c \times 259 = 29 \\ a \times 37 + b \times 259 + c \times 2053 = 194 \\ a \times 259 + b \times 2053 + c \times 17395 = 1502 \end{cases}$$



cuja solução é

a = 0.28869047619

b = 0.890625

c = -0.02306547619

o que permite determinar a ***parábola***, que se aproxima dos pontos,

$y = 0.28869047619 + 0.890625\,x - 0.02306547619\,x^2$

9.3.5. Algoritmo

O processo pode ser formalizado da forma que se segue. Seja P_n o espaço dos polinómios de grau menor ou igual a n com coeficientes reais e $f_1, f_2, \ldots, f_n \in P_n$. A aproximação polinomial dos mínimos quadrados de grau n para os pares de valores $(x_1, y_1), \ldots, (x_m, y_m)$, é o polinómio

$$p = \sum_{i=0}^{n} c_i f_i(x),$$

que minimiza

$$\|f - p\|_2 = \sqrt{\sum_{i=0}^{m} |y_i - p(x_i)|^2}.$$

Algoritmo (equações normais):

{ Objetivo: Construção das equações normais para dados discretos }

{ Parâmetros de entrada: (x_i, y_i) i = 1,...,m; f_k, k = 1, ..., n }

{ Parâmetros de saída:

para i **de** 1 **até** (n+1) **fazer**

para j **de** i **até** (n+1) **fazer**

$a_{ij} \leftarrow 0$

para k **de** 1 **até** m **fazer**

$a_{ij} \leftarrow a_{ij} + f_{i-1}(x_k)\, f_{j-1}(x_k)$

fim_para

$a_{ji} \leftarrow a_{ij}$

fim_para

$b_i \leftarrow 0$

para k **de** 1 **até** m **fazer**

$b_i \leftarrow b_i + y_k\, f_{i-1}(x_k)$

fim_para

fim_para