

Solução de Equações Não Lineares Univariadas

1.1 – Introdução

Nessa parte do curso, nosso objetivo é construir aproximações numéricas para a solução de *equações algébricas em uma única variável real*.

Diversos problemas reais são modelados por equações não lineares: modelos financeiros, relacionados a juros e/ou disponibilidade de recursos, modelos de engenharia de diversas sub áreas, modelos matemáticos de otimização univariada, só para citar algumas áreas.

Uma equação não linear é definida como

$$f(x) = 0, \quad (1.1)$$

sendo $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função não linear, tipicamente algébrica com polinômios de grau 2 ou superior, ou transcendental.

Uma função é dita ser algébrica se ela pode ser expressa como

$$P_0(x)y^n + P_1(x)y^{n-1} + P_2(x)y^{n-2} + \cdots + P_n(x) = 0, \quad (1.2)$$

sendo $P_n(x)$ um polinômio de grau n em x , definido como

$$P_n(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_nx^n. \quad (1.3)$$

Funções transcendentais são funções que não são algébricas, como funções trigonométricas, exponenciais, logarítmicas ou a combinação delas.

O problema a ser estudado consiste, então, em encontrar todos os valores de x para os quais a equação (1.1) seja satisfeita. Esses valores são chamados *zeros* ou *raízes* dessa equação, e serão referenciados daqui pra frente como x^* . Essas raízes podem ser reais, caso típico de equações formadas por funções transcendentais, ou complexas, mais comumente encontradas para equações formadas por polinômios.

As soluções para o problema de encontrar as raízes de uma equação não linear podem

ajudar a resolver outros problemas relacionados. Por exemplo, o problema de encontrar o ponto de encontro entre duas funções, isto é, encontrar os valores para os quais $f(x) = g(x)$, sendo $f, g : \mathbb{R} \rightarrow \mathbb{R}$. Para solucionar esse problema, pode-se “transformá-lo” no problema de solução de equações não lineares simplesmente considerando que

$$f(x) = g(x) \Rightarrow h(x) = f(x) - g(x) = 0. \quad (1.4)$$

Dessa forma, as raízes de $h(x)$ são os pontos de encontro das funções $f(x)$ e $g(x)$.

Exemplo 1.1

Considere duas funções $f(x) = e^{-x} + 0.5$ e $g(x) = \cos(x)$. A Figura 1.1 mostra os gráficos dessas funções e mostra o ponto de intersecção delas, em que $f(x) = g(x)$.

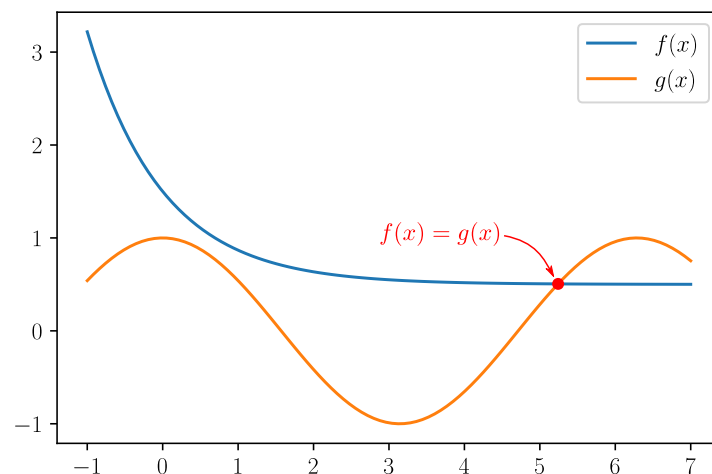


Figura 1.1: Gráficos das funções $f(x) = e^{-x} + 0.5$ (azul) e $g(x) = \cos(x)$ (laranja), destacando o ponto comum entre as duas.

Pode-se definir uma função $h(x) = f(x) - g(x)$, de tal forma que a raiz da equação linear formada por essa função, $h(x) = 0$, será também o ponto exato em que as duas funções se encontram. A Figura 1.2 ajuda a visualizar essa afirmação.

É importante frisar que a raiz x^* é o ponto no eixo das abscissas em que as duas funções se encontram e isso não significa que o valor das funções nesse ponto será igual. De fato, como o gráfico mostra

$$f(x^*) = g(x^*) \neq h(x^*), \quad (1.5)$$

apesar de

$$f(x^*) = g(x^*) \text{ e } h(x^*) = 0 \quad (1.6)$$

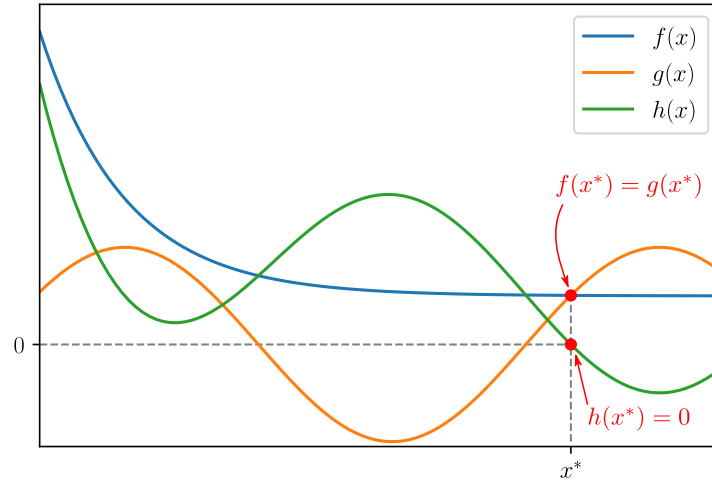


Figura 1.2: Gráficos das funções $f(x)$ (azul), $g(x)$ (laranja), mostradas anteriormente, e $h(x) = f(x) - g(x)$ (verde), destacando o ponto comum entre as duas primeiras e a raiz da equação não linear $h(x) = 0$.

Um segundo tipo de problema que pode utilizar as abordagens para solução de equações não lineares é o de encontrar o máximo ou mínimo local de uma função contínua, com base no Teorema de Fermat, tipicamente visto em disciplina de Cálculo I, que diz que *se uma função f tem um máximo ou mínimo local no ponto c , então $f'(c) = 0$* , sendo $f'(\cdot)$ a primeira derivada da função $f(\cdot)$. Assim, para encontrar o máximo ou mínimo de uma dada função, basta considerar a sua derivada (caso exista) como a função que constitui a equação não linear,

$$f'(x) = 0, \quad (1.7)$$

e os valores de x para os quais a equação (1.7) for satisfeita serão os máximos ou mínimos locais dessa função.

Exemplo 1.2

Seja a função $f(x) = \sin(x) - \cos(3x)$, e sua primeira derivada, $f'(x) = 3 \sin(3x) + \cos(x)$, cujos gráficos são mostrados na Figura 1.3.

Como é possível ver nos gráficos, os valores das raízes da equação $f'(x) = 0$ são exatamente os pontos de máximo e mínimo locais para a função $f(x)$, em concordância com o Teorema de Fermat, mencionado logo acima.

Importante também frisar, como no exemplo anterior, que as raízes x_1^* e x_2^* são os pontos no eixo das abscissas em que ocorrem os máximo e mínimo locais da função no intervalo considerado, mas que isso não significa que os valores da função e de sua derivada nesse ponto serão iguais. De fato, como o gráfico mostra

$$f(x_1^*) \neq f'(x_1^*) \quad \text{e} \quad f(x_2^*) \neq f'(x_2^*). \quad (1.8)$$

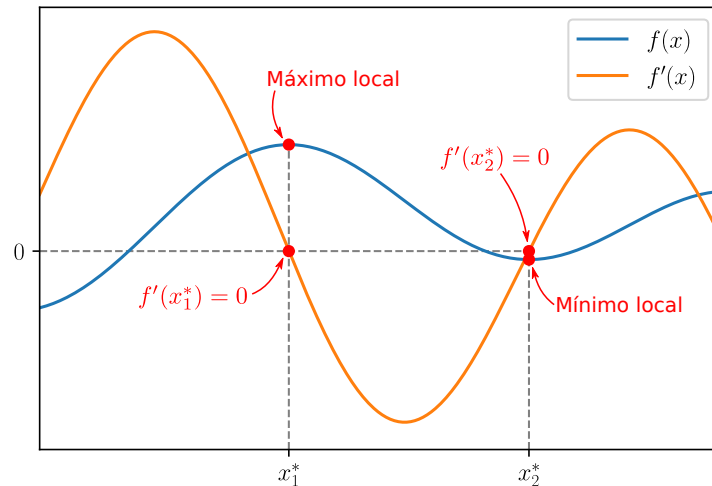


Figura 1.3: Gráficos das funções $f(x) = \sin(x) - \cos(3x)$ (azul) e de sua primeira derivada $f'(x)$ (laranja), destacando os pontos de máximo e mínimo locais de $f(x)$ e as raízes da equação $f'(x) = 0$.

Para resolver esses problemas, é possível utilizar três abordagens:

1. **soluções analíticas diretas** – para o caso em que pode-se encontrar uma expressão analítica para calcular as raízes da equação diretamente. Os leitores com certeza já tiveram contato com esse tipo de abordagem, quando aprenderam como achar as raízes de uma equação de segundo grau, definida como

$$f(x) = ax^2 + bx + c = 0, \quad (1.9)$$

usando a famosa *fórmula de Bhaskara*,

$$x^* = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (1.10)$$

Usando a equação (1.10), pode-se encontrar diretamente todas as raízes da equação (1.9). No entanto, poucos são os casos em que a solução de uma equação não linear pode ser expressa por uma fórmula analítica direta. Sendo sincero, isso é razoavelmente raro... Por esses motivos, outras abordagens são necessárias para a obtenção desses valores.

2. **soluções gráficas** – para o caso em que se tem uma função bem definida na equação não linear, uma possível abordagem é gerar o gráfico dessa função e, por inspeção visual, identificar os pontos nos quais o gráfico toca o eixo das abscissas. Esses pontos serão as raízes ou zeros da equação. O grande problema desse método é sua imprecisão, uma vez que uma busca visual pode acarretar numa aproximação grosseira do valor, a depender de vários fatores como resolução do gráfico, precisão da representação, ou até mesmo a disposição de elementos como grades e valores nos eixos. O exemplo 1.3 ilustra esse tipo de abordagem e as suas dificuldades.

Exemplo 1.3

Considere o gráfico a seguir. É possível verificar que, visualmente, é impossível determinar com precisão a raiz da equação não linear formada pela função do gráfico, mesmo com a representação de valores nos eixos e com elementos como a grade do gráfico. No máximo, é possível estimar que essa raiz se encontra dentro do intervalo entre $[5, 6]$, talvez, $[5, 5.5]$, mas, precisamente, não é possível dar uma estimativa dessa raiz.

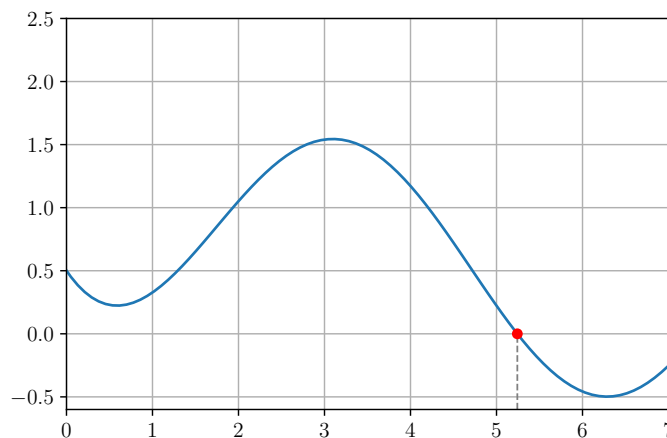


Figura 1.4: Exemplo das dificuldades do uso de inspeção visual para determinação de raízes de equações não lineares.

3. **métodos iterativos** – na impossibilidade de ter um resultado direto, via definição analítica de uma expressão para as raízes da equação, ou de um valor preciso obtido por inspeção visual do gráfico da função que descreve essa equação, a melhor opção para resolver o problema de encontrar a raiz de uma equação não linear é usar métodos iterativos. Nessa abordagem, algoritmos realizam iterações para se aproximar do valor da raiz, até encontrar o valor exato ou uma aproximação cujo erro seja aceitável.

Métodos Iterativos

Os métodos iterativos podem ser divididos em dois grupos: *intervalares* e *abertos*.

Nos métodos intervalares define-se um intervalo que contenha a raiz e, a cada iteração, esse intervalo vai diminuindo até que os limites dele fiquem suficientemente próximos da raiz procurada. Os métodos da *bisseção* e da *falsa posição* são dois dos mais famosos representantes desse grupo.

Nesse ponto, é interessante responder uma pergunta: *como garantir que a raiz da equação está de fato dentro do intervalo desejado?*. A resposta à essa pergunta é dada pelo Teorema do Valor Intermediário (visto com certeza na disciplina de Cálculo II!)

Teorema 1.1. Teorema do Valor Intermediário: *Seja a função contínua $f : [a, b] \rightarrow \mathbb{R}$, tal que $f(a) < f(b)$. Para qualquer $k \in (f(a), f(b))$, existe um $x^* \in (a, b)$ tal que $f(x^*) = k$.*

Ora, desse teorema, tomando $k = 0$, é possível afirmar que existe $x^* \in (a, b)$ tal que $f(x^*) = 0$, se $f(a) < f(b)$. Mas, como $k \in (f(a), f(b))$, e, por conseguinte, $0 \in (f(a), f(b))$ então, $f(a) < 0 < f(b)$ e, nesse caso, $f(a) \cdot f(b) < 0$.

Em outras palavras, se $f(x)$ é uma função contínua em um dado intervalo $[a, b]$ no qual ela troca de sinal, então ela tem pelo menos uma raiz neste intervalo. Somente haverá troca de sinal quando a relação $f(a) \cdot f(b) < 0$ for válida, isto é, quando a função “cruzar” o eixo das abscissas.

A Figura 1.5 ilustra essa discussão, mostrando a existência de uma raiz na função quando existe troca de sinal dentro do intervalo. Note que as regiões destacadas na curva ilustram, justamente, que a troca de sinal ocorre na passagem por 0.

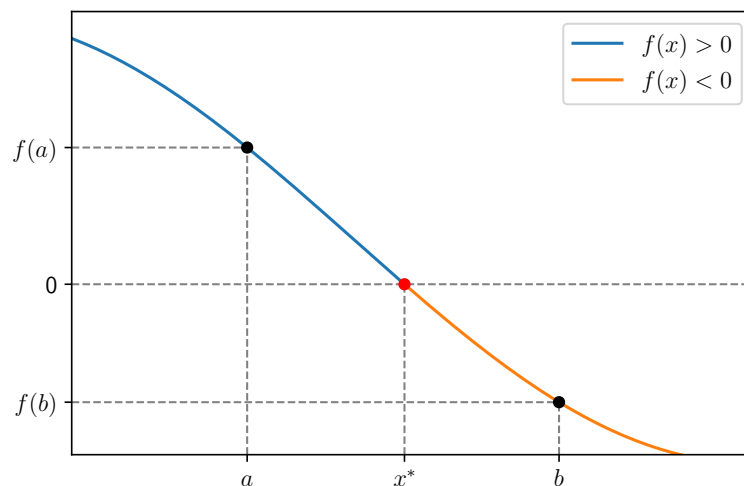


Figura 1.5: Ilustração do Teorema do Valor Intermediário, para o caso em que $k = 0$, mostrando a existência de uma raiz na função quando existe troca de sinal dentro do intervalo.

Exemplo 1.4

Mostre que existe pelo menos uma solução da equação $f(x) = 2 - e^x = 0$ no intervalo $[0, 1]$.

De acordo com o Teorema 1.1, como $f(0) = 1 > 0$ e $f(1) = -0.72 < 0$ e, portanto, $f(0) \cdot f(1) = 1 \cdot (-0.72) = -0.72 < 0$, então existe ao menos uma solução para a equação $f(x) = 2 - e^x = 0$ no intervalo $[0, 1]$.

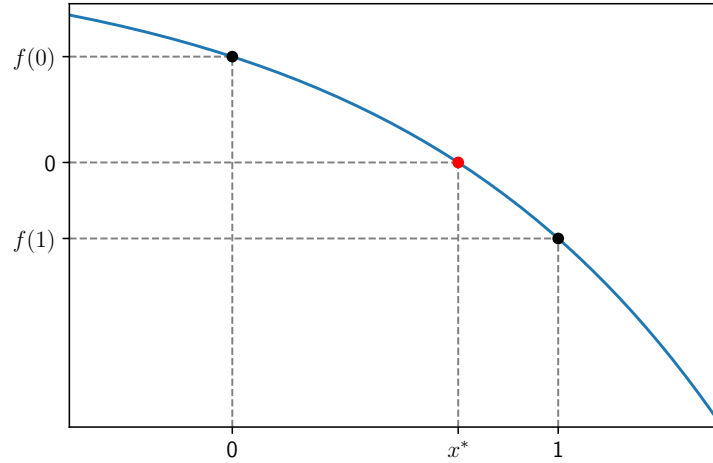


Figura 1.6: Gráfico da função $f(x) = 2 - e^x$, mostrando a existência de uma raiz no intervalo $[0, 1]$.

Na implementação de métodos intervalares para a solução de equações não lineares, é interessante contemplar um teste para verificar a existência de raiz no intervalo, baseado no teorema anterior, para evitar que a equação seja avaliada em um intervalo sem raiz, o que levaria o método utilizado a nunca convergir ou retornar um valor errado da raiz. No entanto, a depender dos valores trabalhados, a produto $f(a) \cdot f(b)$ pode ocasionar erros de *overflow* ou *underflow*, conforme discutido no capítulo anterior.

Uma forma alternativa para essa avaliação é utilizar a função *senal*, $\text{sgn}(x)$, definida como

$$\text{sgn}(x) = \begin{cases} -1, & \text{se } x < 0, \\ 0, & \text{se } x = 0, \\ 1, & \text{se } x > 0, \end{cases} \quad (1.11)$$

para realizar o teste

$$\text{sgn}(f(a)) \cdot \text{sgn}(f(b)) = -1, \quad (1.12)$$

que, caso seja verdadeiro, indica a presença de raiz no intervalo testado, obtendo o mesmo resultado de $f(a) \cdot f(b) < 0$, mas sem risco de incorrer em erros de *overflow* ou *underflow*, uma vez que, independente do valor de x , a função só retorna os valores mostrados na equação (1.11).

Esse teste pode ser facilmente executado utilizando a função `sign` da biblioteca Numpy, cuja documentação é mostrada no [link](#) e que implementa a função descrita pela equação (1.11). O exemplo a seguir ilustra o uso dessa função.

Exemplo 1.5

Use a função `sign` da biblioteca Numpy para o caso trabalhado no exemplo 1.4.

Antes de tudo, é preciso carregar a biblioteca Numpy,

```
In [ ]: import numpy as np
```

e, sem seguida, criar a função a ser avaliada,

```
In [ ]: f = lambda x: 2 - np.exp(x)
```

e definir os limites do intervalo

```
In [ ]: a,b = 0,1
```

para, então, poder fazer o teste

```
In [ ]: np.sign(f(a))*np.sign(f(b))
```

```
Out [ ]: -1.0
```

indicando, como no exemplo 1.4, que no intervalo existe ao menos uma raiz para a equação avaliada.

Diferentemente dos métodos intervalares, nos métodos abertos não existe a necessidade de definir um intervalo que contenha a raiz para então realizar sua busca. Esses métodos partem de um ou dois valor inicial de x (que não definem um intervalo de busca) e, então, realizam a verificação se esse valor escolhido é, de fato, a raiz x^* . Caso não seja, o valor atual de x é atualizado por uma expressão, e o novo valor é novamente verificado. Esse processo iterativo se mantém até que o valor de x na iteração atual seja próximo o suficiente de x^* , de acordo com algum critério adotado. Os métodos da *Iteração de Ponto Fixo*, de *Newton-Raphson* e da *Secante* são alguns dos métodos abertos mais comumente utilizados.

Crítérios de parada para métodos iterativos

Como visto, métodos iterativos iniciam a partir de um intervalo ou de um valor inicial, e seguem realizando iterações em busca do valor exato da raiz, ou, como acontece na maioria dos casos, da melhor aproximação. No entanto, é preciso definir: *qual a melhor aproximação?*

Para responder isso, é preciso estabelecer uma *tolerância*, ou seja, um limite de erro aceitável, abaixo do qual se considere a aproximação como satisfatória. Assim, três critérios

baseados na tolerância podem ser adotados como critério de paradas para métodos iterativos

$$|x_n - x_{n-1}| < \text{atol}, \quad (1.13a)$$

$$\frac{|x_n - x_{n-1}|}{|x_n|} < \text{rtol}, \quad (1.13b)$$

$$|f(x_n)| < \text{ftol}, \quad (1.13c)$$

sendo `atol` o erro absoluto máximo tolerado, ou *tolerância absoluta*, `rtol`, o erro relativo máximo tolerado, ou *tolerância relativa* e `ftol` o erro absoluto máximo tolerado do valor da função na aproximação. Esses três valores são especificados de acordo com a necessidade de cada problema.

Além desse, muitas implementações também consideram o número de iterações um critério extra para a parada do método, muitas vezes coexistindo com o anterior. Nesses casos em que os dois critérios são considerados, o algoritmo para sua execução sempre que um dos dois critérios for satisfeito.