CC0323: Otimização Irrestrita

Michael Souza

michael@ufc.br

Problema de Otimização Não-Linear

Dado um conjunto $\Omega \subset \mathbb{R}^n$ e uma função $f : \Omega \to \mathbb{R}$, o problema de otimização não-linear consiste em encontrar um ponto $x^* \in \Omega$ que minimize (ou maximize) a função f. Ou seja, desejamos encontrar

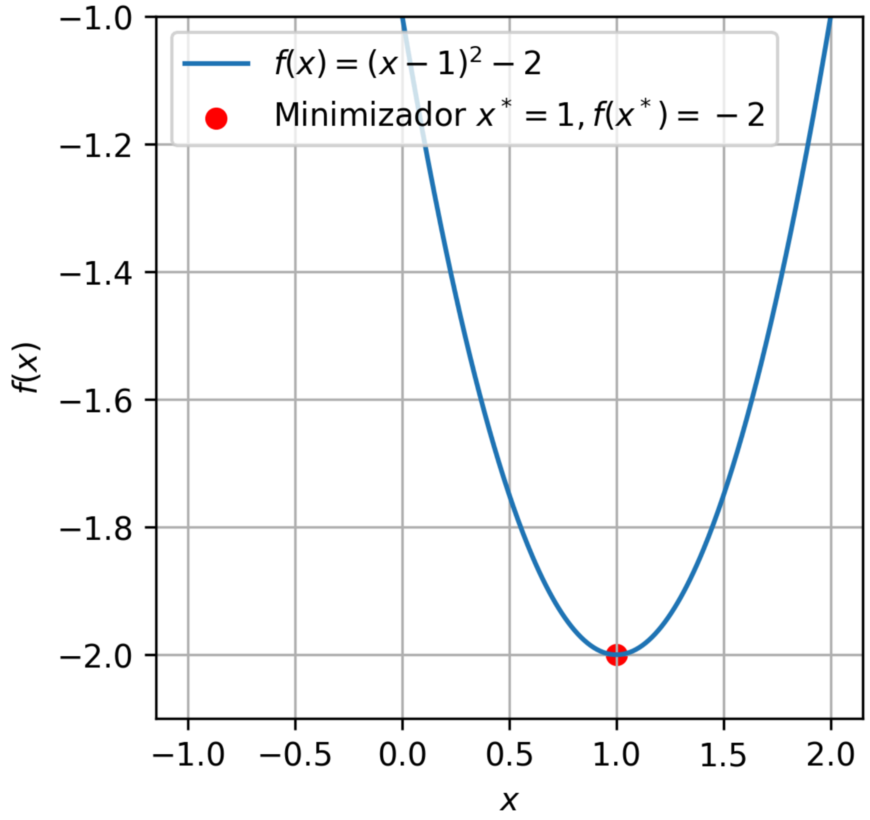
$$x^* = \arg\min_{x \in \Omega} f(x).$$

Mínimo vs Minimizador

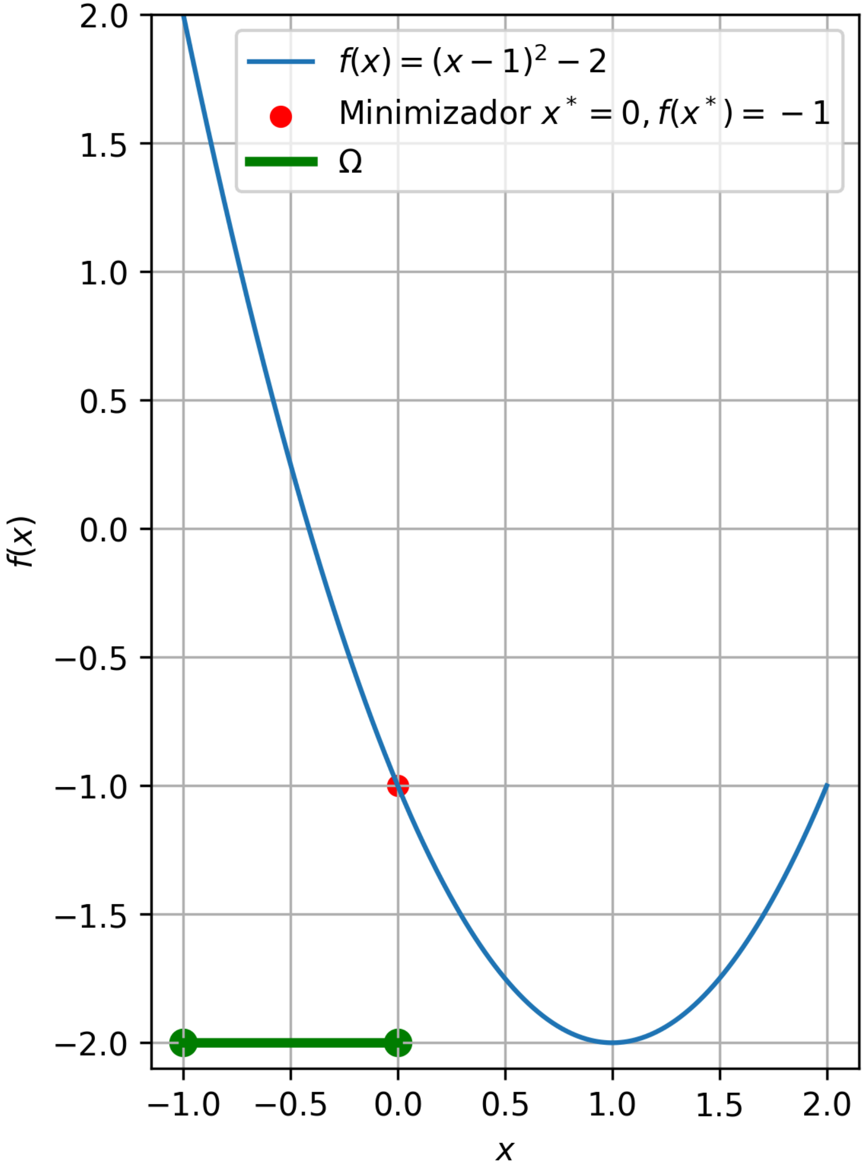
O ponto x^* é chamado de ***minimizador*** (ou ***maximizador***) de f em Ω, enquanto o valor $f(x^*)$ é chamado de ***mínimo*** (ou ***máximo***) de f em Ω.

Exemplo:

1. Considere a função $f(x) = (x-1)^2 - 2$ e o conjunto $\Omega = \mathbb{R}$. O ponto $x^* = 1$ é o minimizador de f em Ω, enquanto o valor $f(x^*) = f(1) = -2$ é o mínimo de f em Ω.



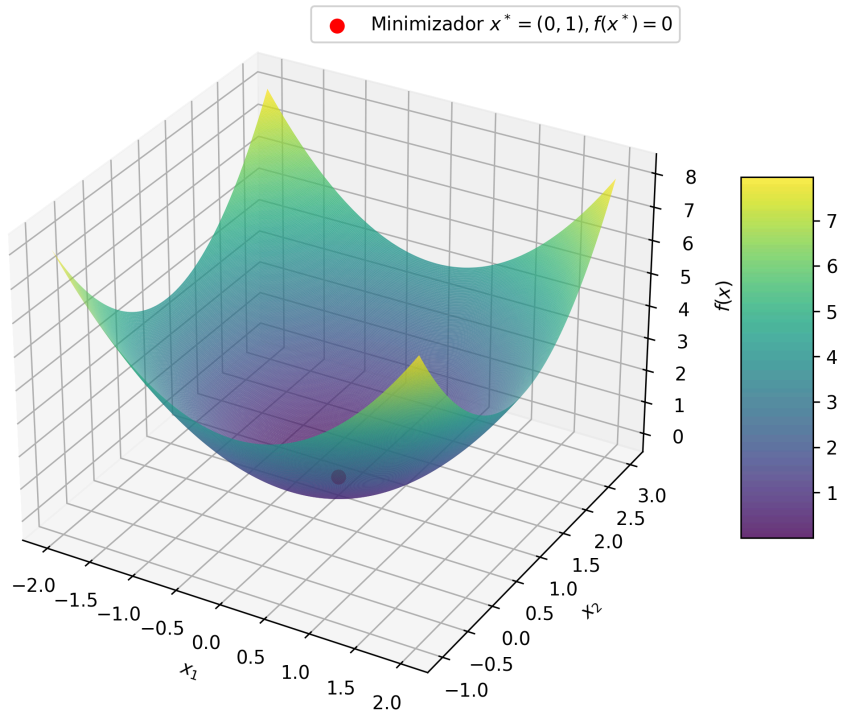
2. Considere a mesma função do exemplo anterior, mas $\Omega[-1,0]$.
Neste caso, $x = 1$ ***não é viável*** (não pertence a Ω) e o mimizador x^* passa a ser 0 e o mínimo será $f(x^*) = f(0) = -1$.



Observação:
O minimizador depende tanto da função f quanto do conjunto viável Ω.

3. Agora, considere a função $f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$. Como f é uma soma de quadrados, temos que $f(x) \geq 0$ para todo $x \in \mathbb{R}^2$.

   Portanto, o minizador de f em $\mathbb{R}^2$ é $x^* = (0, 1)$, pois $f(x^*) = 0$.

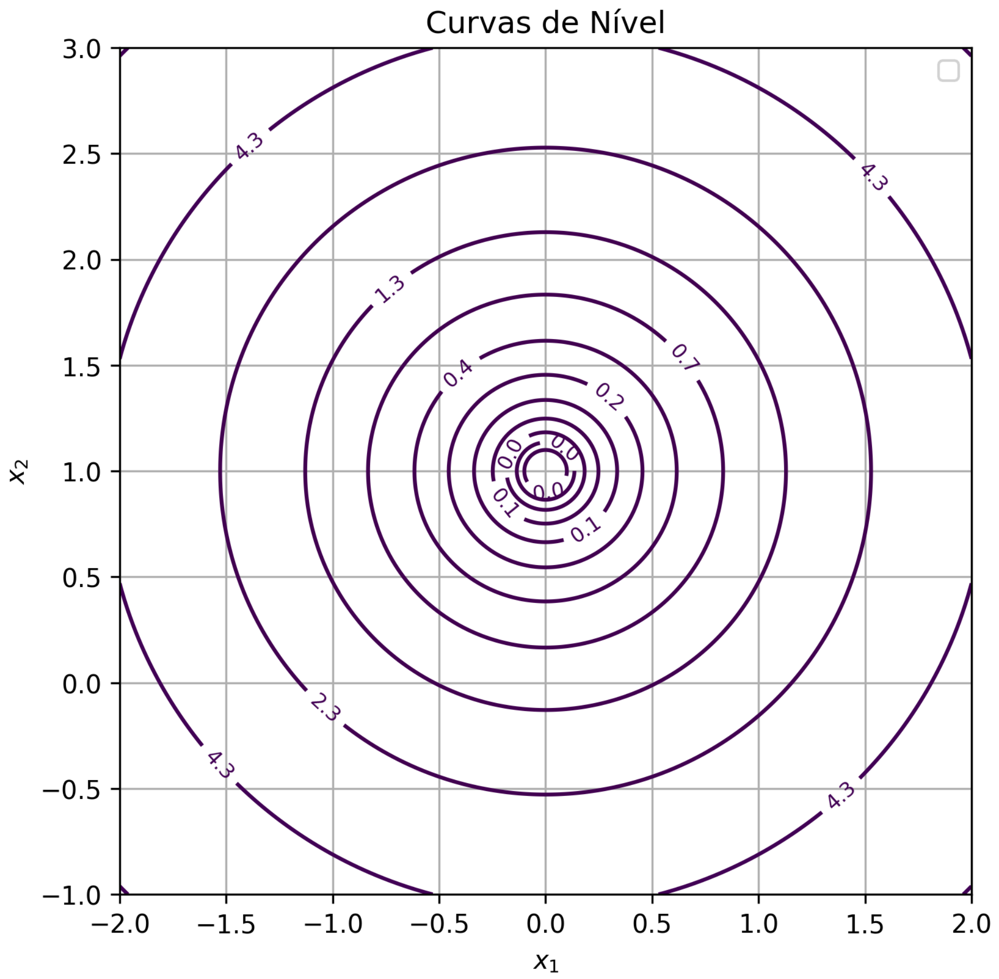


Curva de Nível

A curva de nível de uma função $f : \mathbb{R}^n \to \mathbb{R}$ é o conjunto de pontos $(x_1, x_2, \ldots, x_n)$ onde $f(x) = c$ para algum $c \in \mathbb{R}$.

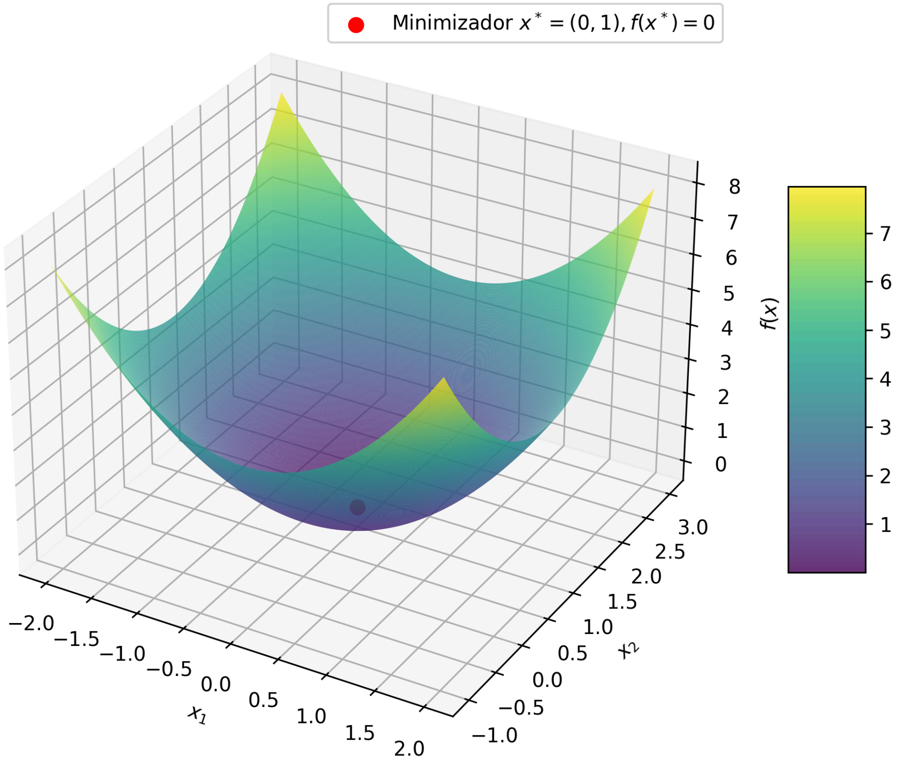
Observações:

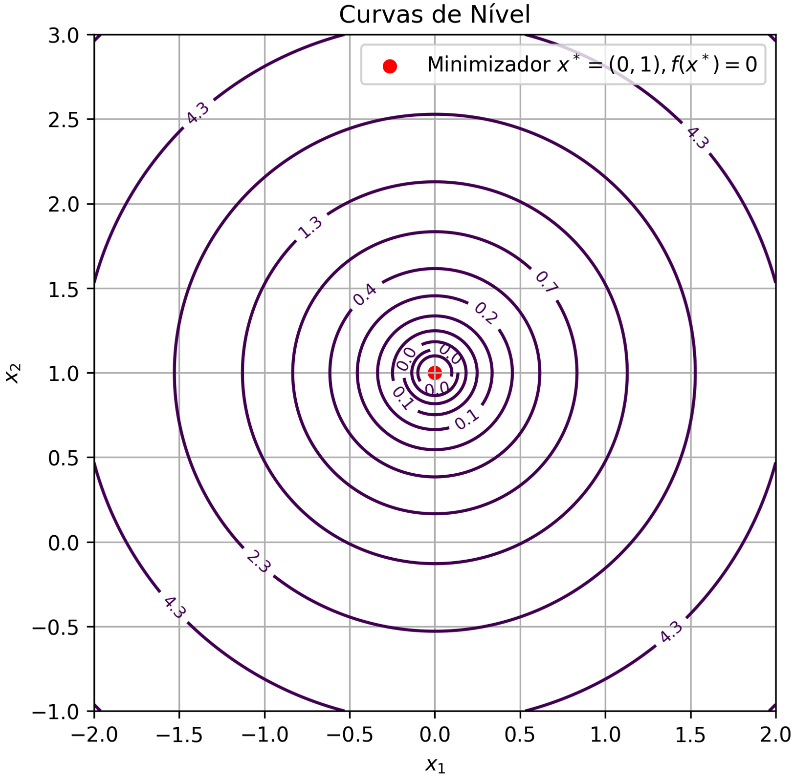
As curvas de nível são úteis para visualizar funções de duas variáveis, cujos gráficos são pontos $(x_1, x_2, f(x_1, x_2)) \in \mathbb{R}^3$.



Exemplo:

Para a função $f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$, uma curva de nível c é o conjunto solução da equação $f(x) = x_1^2 + (x_2 - 1)^2 = c$, ou seja, é o círculo de raio $\sqrt{c}$ centrado em $(0, 1)$.





Problema de Otimização Irrestrita

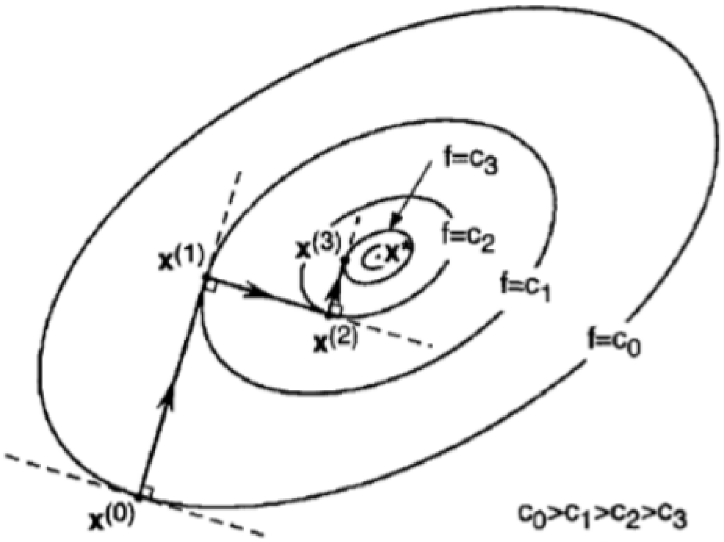
Quando o conjunto Ω é todo o espaço $\mathbb{R}^n$, dizemos que o problema de otimização é ***irrestrito*** (sem restrições).

Ideia Básica

Dado um ponto inicial x_0, a ideia é gerar uma sequência de pontos $\{x_k\}$ tal que $f(x_k) \to f(x^*)$.

Em cada ponto teremos uma direção d_k e um passo α_k que nos levará a um novo ponto x_{k+1}.

$$x_k = x_{k-1} + \alpha_k d_k \quad \text{e} \quad f(x_k) < f(x_{k-1})$$

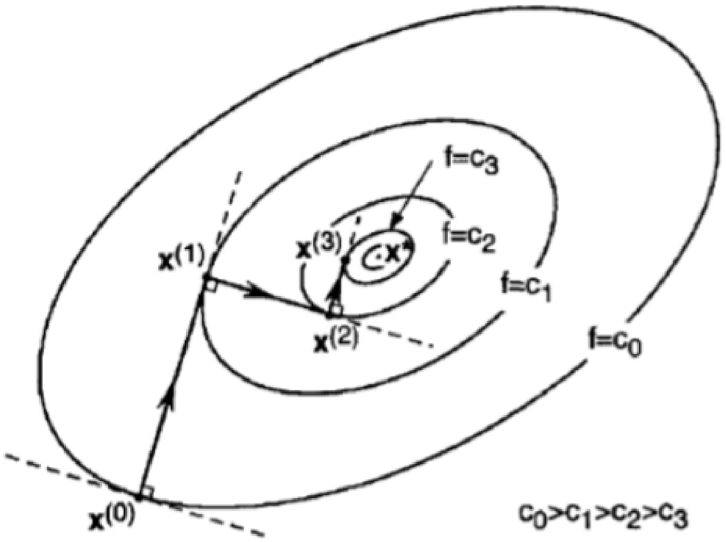


Otimização Unidimensional

Em cada passo do método geral, precisamos resolver um problema de otimização unidimensional. Ou seja, dado um ponto x_k e uma direção d_k, precisamos encontrar um passo (minimizador)

$$\alpha_k = \arg\min_{\alpha} \phi_k(\alpha) = \arg\min_{\alpha} f(x_k + \alpha d_k).$$

Portanto, precisamos de um método para resolver problemas de otimização unidimensional.



Método da Seção Áurea

O método da seção áurea serve para encontrar o mínimo de uma função unimodal em um intervalo $[a_0, b_0] \subset \mathbb{R}$.

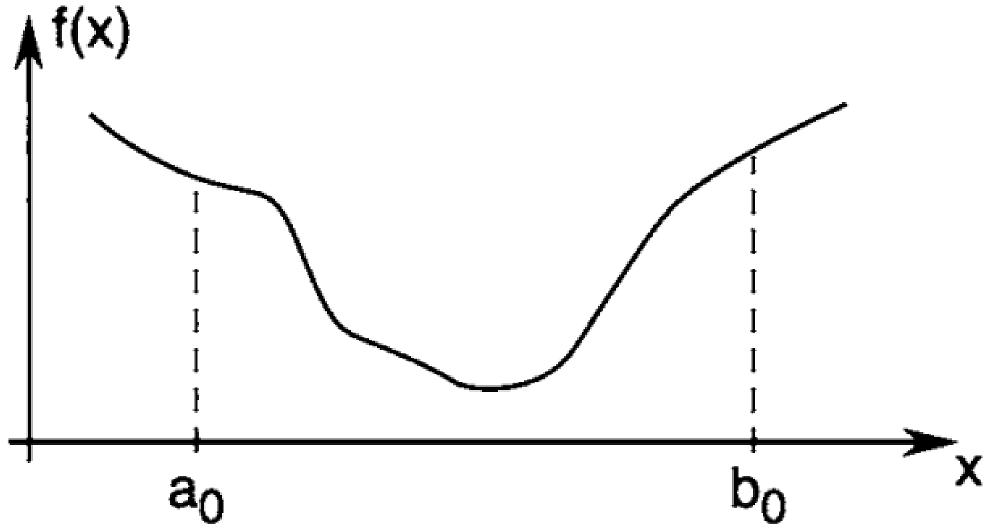


Figure 7.1 Unimodal function.

Algoritmo

1. Escolhemos a_1 e b_1 tais que

$$a_0 < a_1 < b_1 < b_0,$$
$$b_1 - a_1 = \rho(b_0 - a_0),$$

onde $\rho = \frac{3-\sqrt{5}}{2} \approx 0.382$.

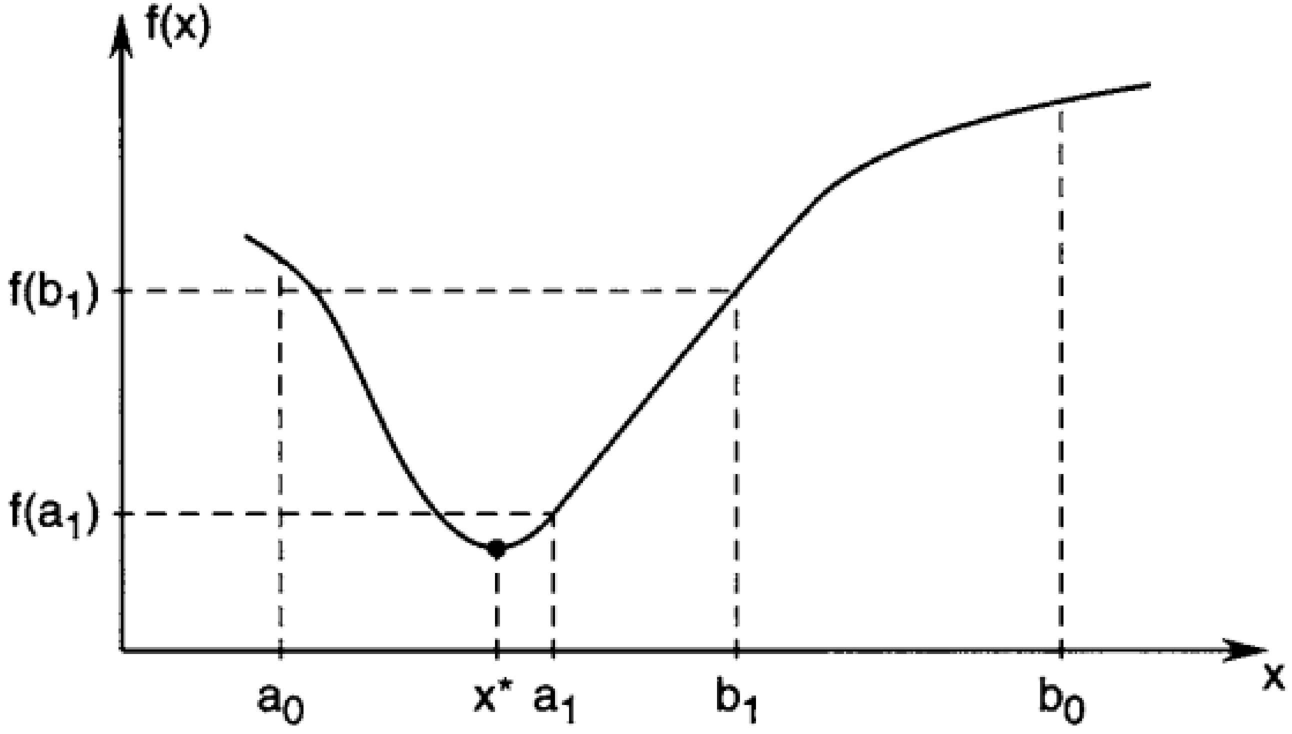


Figure 7.3 The case where $f(a_1) < f(b_1)$; the minimizer $x^* \in [a_0, b_1]$.

2. Calculamos
$f(a_1)$ e $f(b_1)$.

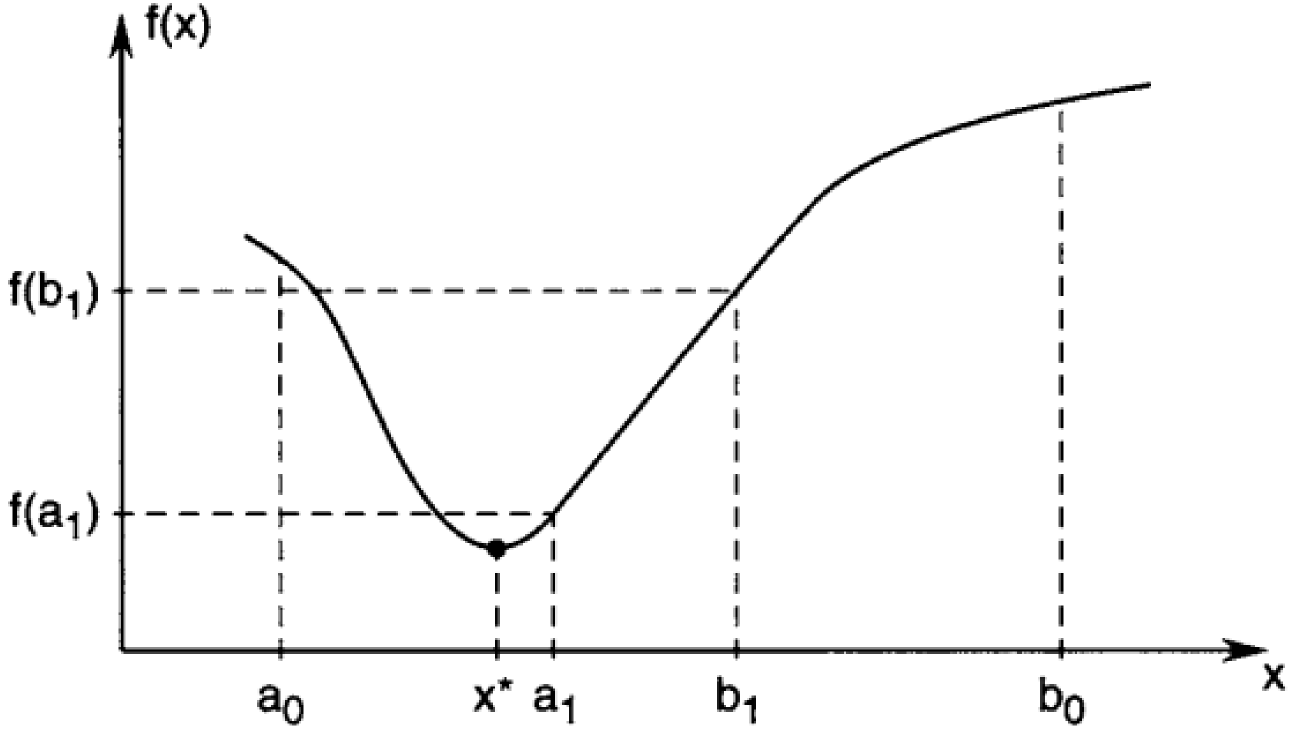


Figure 7.3 The case where $f(a_1) < f(b_1)$; the minimizer $x^* \in [a_0, b_1]$.

3. Se $f(a_1) < f(b_1)$, então o mínimo está no intervalo $[a_0, b_1]$. Caso contrário, o mínimo está no intervalo $[a_1, b_0]$.

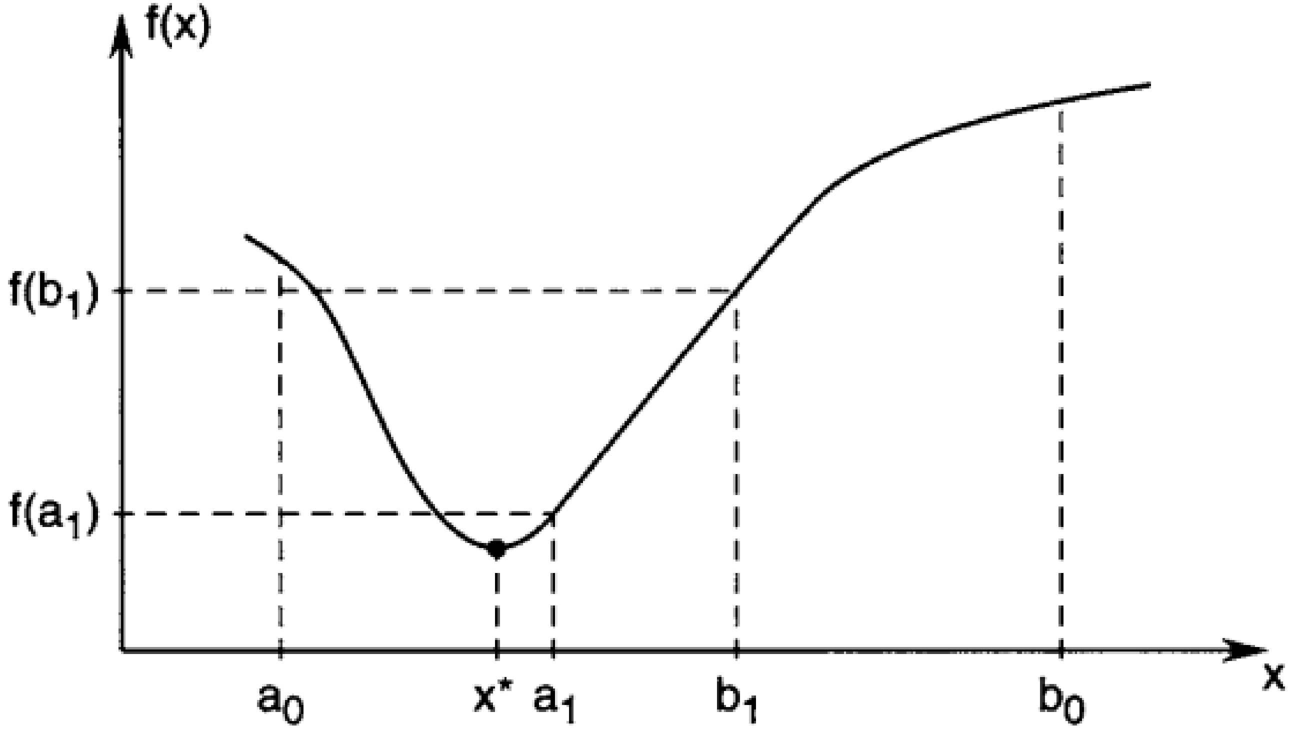


Figure 7.3 The case where $f(a_1) < f(b_1)$; the minimizer $x^* \in [a_0, b_1]$.

4. Agora, repetimos o processo até que o intervalo seja suficientemente pequeno.

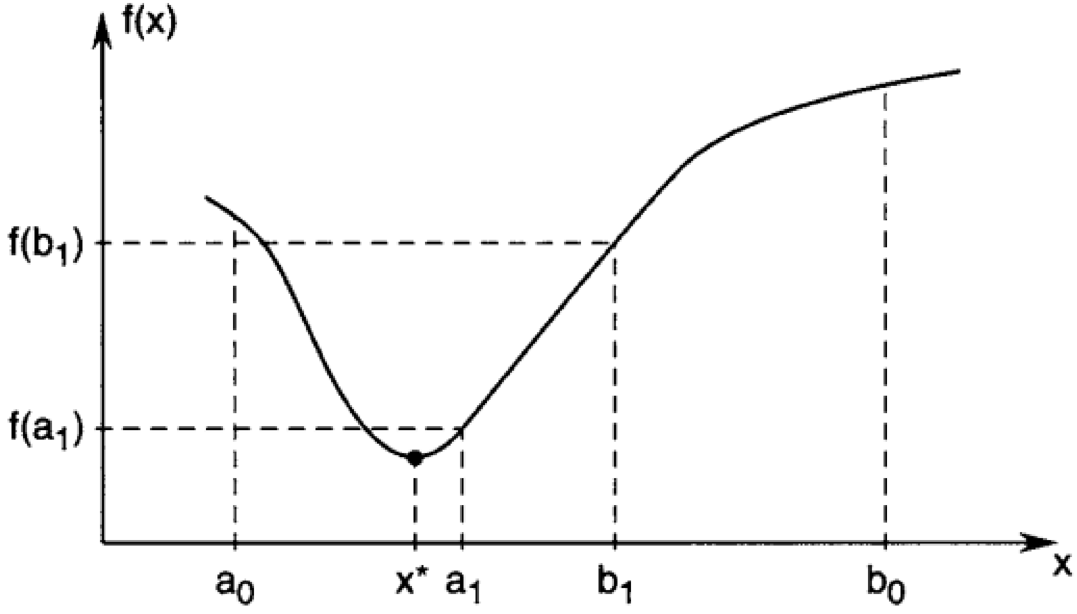


Figure 7.3 The case where $f(a_1) < f(b_1)$; the minimizer $x^* \in [a_0, b_1]$.

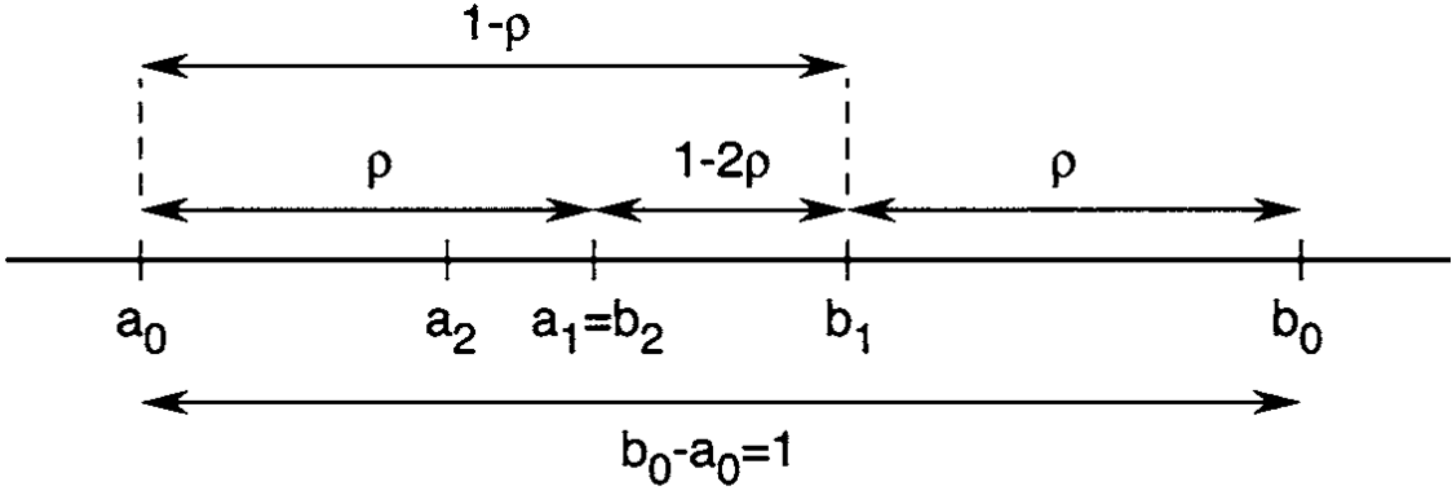
O valor $\rho = 0.382$ decorre de duas ideias simples:

- Simetria, pois não temos razão para preferirmos um dos lados.

$$\frac{b_0 - b_1}{b_0 - a_0} = \frac{a_1 - a_0}{b_0 - a_0} = \rho < \frac{1}{2}$$

- Reuso, queremos reduzir o número de avaliações de f.

$$\frac{b_1 - b_2}{b_1 - a_0} = \rho$$



A cada passo, o intervalo de incerteza é reduzido por um fator

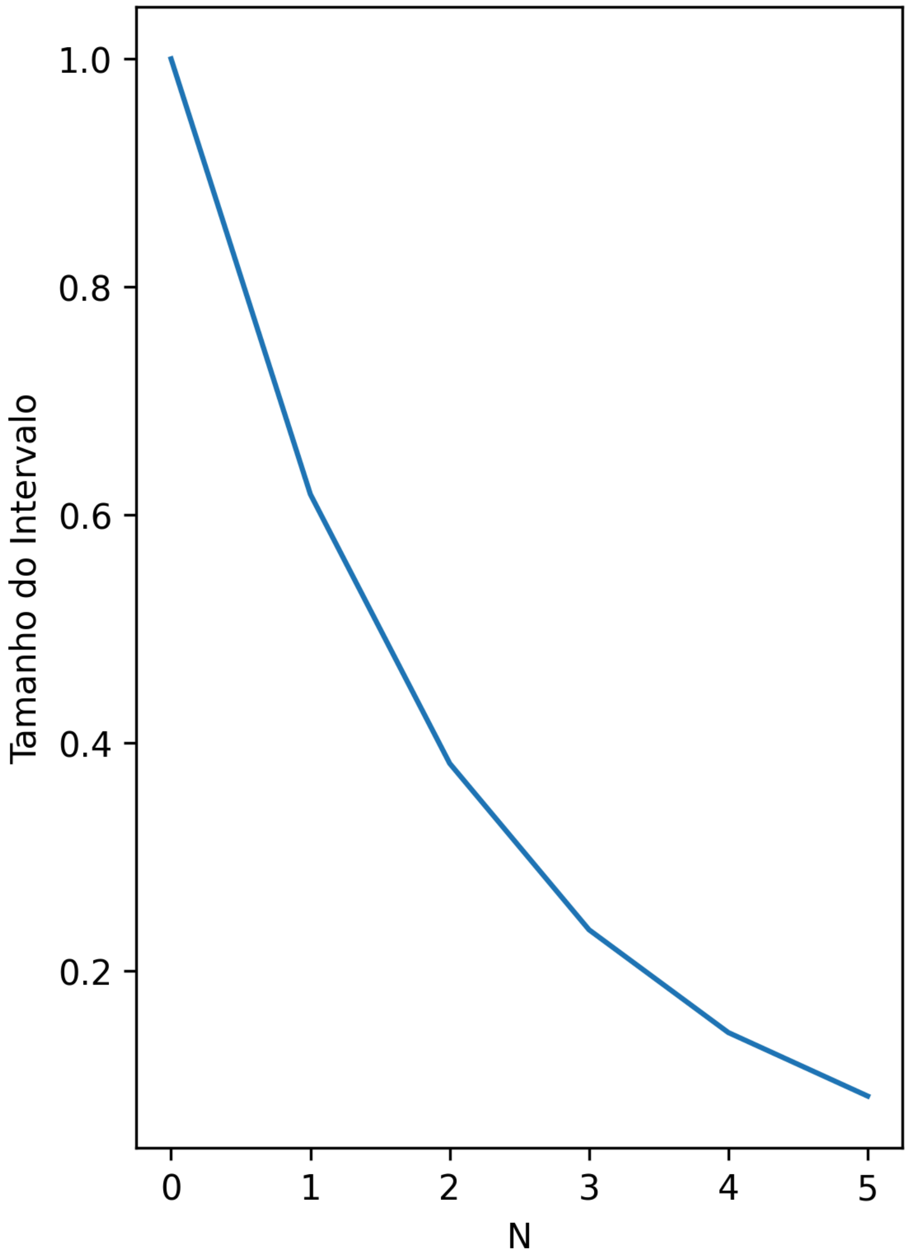
$$(1-\rho) \approx 0.61803$$

Então, após N passos o intervalo original será reduzido por um fator

$$(1-\rho)^N = (0.61803)^N$$

E, como consequência, os erros $\epsilon_k = x_k - x^*$ satisfazem a relação

$$\epsilon_{k+1} = (1-\rho)\epsilon_k.$$



Exemplo

Suponha que queremos utilizar a seção áurea para encontrar o mínimo da função $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ no intervalo $[0, 2]$. Desejamos uma precisão de 10^{-3}, ou seja, $|b_k - a_k| < 10^{-3}$.

Método de Newton

Quando a função f é duas vezes diferenciável, podemos utilizar o método de Newton para encontrar seu mínimo.

A ideia é aproximar a função f por uma função quadrática q e encontrar o mínimo da aproximação.

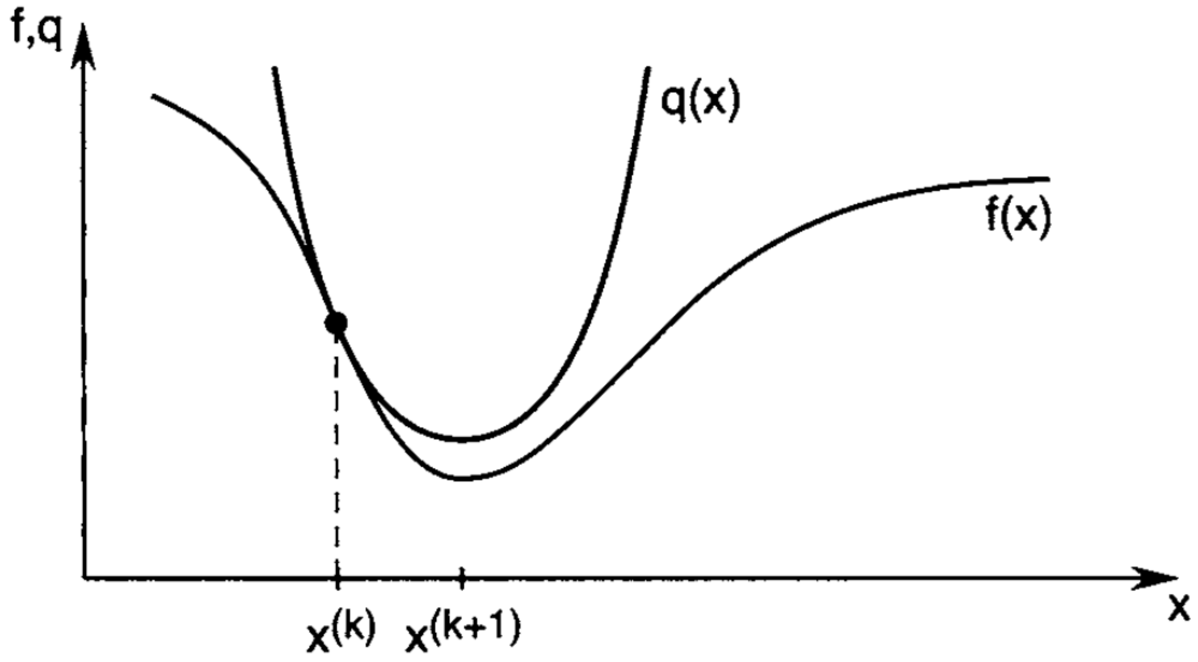


Figure 7.6 Newton's algorithm with $f''(x) > 0$.

A aproximação quadrática q é dada pelo polinômio de Taylor de segunda ordem de f em torno de x_k.

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2.$$

Cujo mínimo (vértice) é dado por

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

Taxa de Convergência do Método de Newton

Se f é duas vezes diferenciável, x^* é um mínimo de f, $f''(x^*) > 0$ e as demais derivadas são limitadas, então o método de Newton converge para x^* com taxa quadrática, ou seja,

$$\epsilon_{k+1} \leq C\epsilon_k^2,$$

onde $\epsilon_k = \left|x_k - x^*\right|$ e C é uma constante positiva.

Prova

Uma vez que $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$, então

$$x_{k+1} - x^* = x_k - x^* - \frac{f'(x_k)}{f''(x_k)} \longrightarrow \epsilon_{k+1} = \epsilon_k - \frac{f'(x_k)}{f''(x_k)}.$$

Agora, tomando a expansão em série de Taylor de f' e f'' em torno de x^*, temos

$$
\begin{aligned}
f'(x_k) &= f'(x^*) + f''(x^*)(x_k - x^*) + \frac{1}{2} f'''(x^*)(x_k - x^*)^2 + \dots \\
&= f'(x^*) + f''(x^*)\epsilon_k + \frac{1}{2} f'''(x^*)\epsilon_k^2 + \dots
\end{aligned}
$$

Uma vez que x^* é um mínimo, temos que $f'(x^*) = 0$ e $f''(x^*) > 0$. Portanto, podemos escrever

$$
f'(x_k) = f''(x^*)\epsilon_k + \frac{1}{2} f'''(x^*)\epsilon_k^2 + \dots .
$$

Por outro lado, podemos escrever a expansão em série de Taylor de f'' em torno de x^* como

$$
\begin{aligned}
f''(x_k) &= f''(x^*) + f'''(x^*)(x_k - x^*) + \frac{1}{2} f''''(x^*)(x_k - x^*)^2 + \dots \\
&= f''(x^*) + f'''(x^*)\epsilon_k + \frac{1}{2} f''''(x^*)\epsilon_k^2 + \dots
\end{aligned}
$$

E, colocando $f''(x^*)$ em evidência, temos

$$
f''(x_k) = f''(x^*) \left(1 + \frac{f'''(x^*)}{f''(x^*)} \epsilon_k + \dots \right).
$$

Combinando as duas equações, temos

$$\begin{aligned}\epsilon_{k+1} &= \epsilon_k - \frac{\left(f''(x^*)\epsilon_k + \frac{1}{2} f'''(x^*)\epsilon_k^2 + \dots\right)}{f''(x^*)} \times \left(1 + \frac{f'''(x^*)}{f''(x^*)}\epsilon_k + \dots\right)^{-1} \\ &= \epsilon_k - \left(\epsilon_k + \frac{1}{2}\frac{f'''(x^*)}{f''(x^*)}\epsilon_k^2 + \dots\right) \times \left(1 + \frac{f'''(x^*)}{f''(x^*)}\epsilon_k + \dots\right)^{-1}\end{aligned}$$

Como $\epsilon_k \to 0$, devemos nos ater apenas aos termos de maior ordem. Mais especificamente, vamos mostrar que o termo ϵ_k do lado direito da equação acima é cancelado e, como consequência, o termo ϵ_k^2 domina a convergência.

Agora, para $0 < x < 1$, temos $(1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots$, e, uma vez que as derivadas de f em x^* são limitadas, então podemos substituir a inversa, obtendo

$$\begin{aligned}
\epsilon_{k+1} &= \epsilon_k - \left(\epsilon_k + \frac{1}{2}\frac{f'''(x^*)}{f''(x^*)}\epsilon_k^2 + \ldots\right) \times \left(1 - \frac{f'''(x^*)}{f''(x^*)}\epsilon_k + \ldots\right). \\
&= \epsilon_k - \epsilon_k + \frac{1}{2}\frac{f'''(x^*)}{f''(x^*)}\epsilon_k^2 + O(\epsilon_k^3) \\
&= C\epsilon_k^2 + O(\epsilon_k^3) \\
&\approx C\epsilon_k^2.
\end{aligned}$$

Observação

Isto significa que cada passo do método de Newton **dobra o número de dígitos corretos do minimizador.***

Limitações do método de Newton

- O método de Newton é sensível à escolha do ponto inicial.

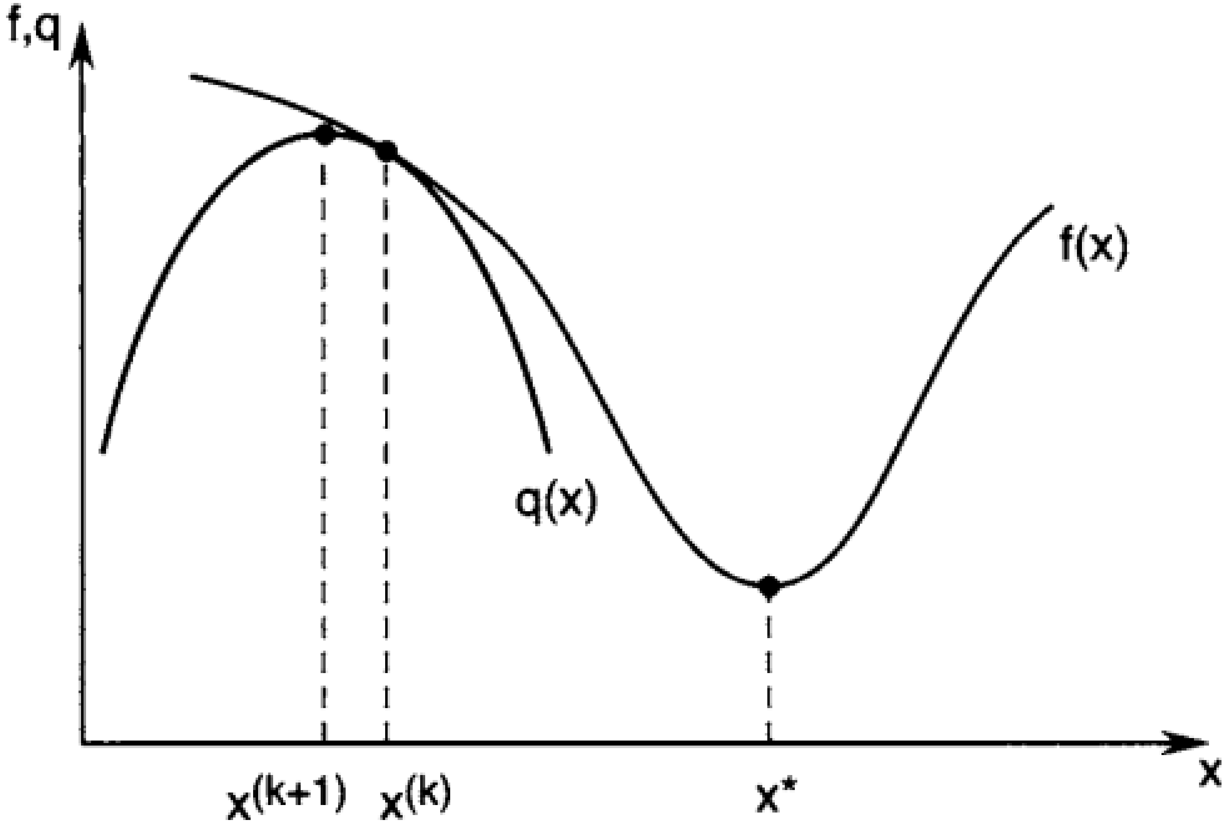


Figure 7.7 Newton's algorithm with $f''(x) < 0$.

Limitações do método de Newton

- O método de Newton é sensível à escolha do ponto inicial.
- O método de Newton pode ciclar ou divergir.

Na imagem ao lado, $g(x) = f'(x)$.

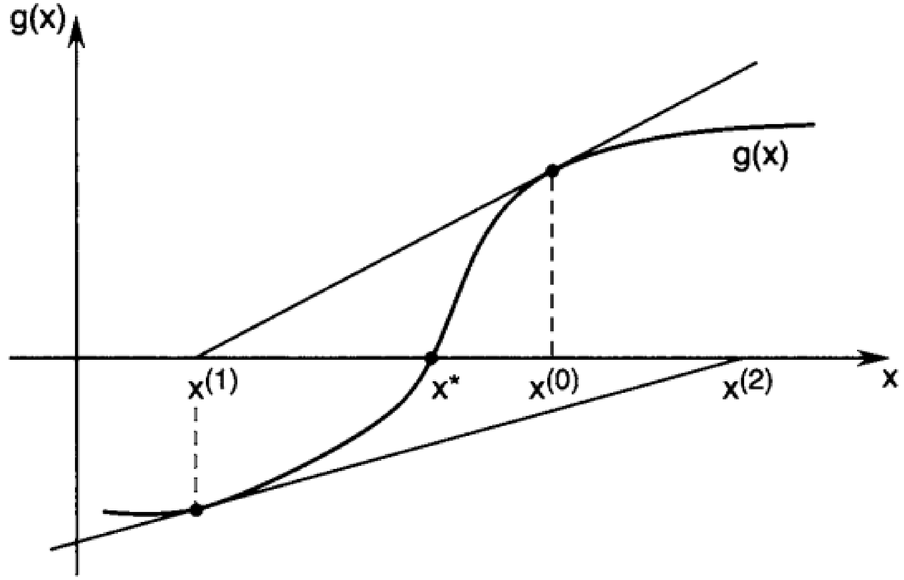


Figure 7.9 Example where Newton's method of tangents fails to converge to the root x^* of $g(x) = 0$.

Método da Secante

O método da secante é uma aproximação do método de Newton para funções unidimensionais.

A ideia é aproximar

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

Assim, a iteração de Newton se torna

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

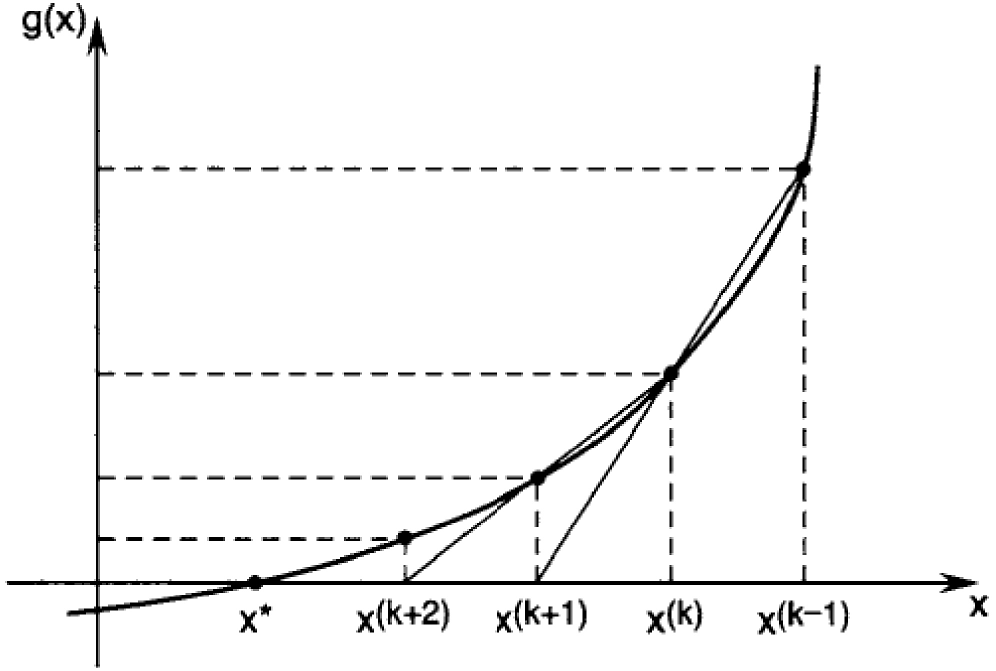


Figure 7.10 Secant method for root finding.

Newton vs Secante

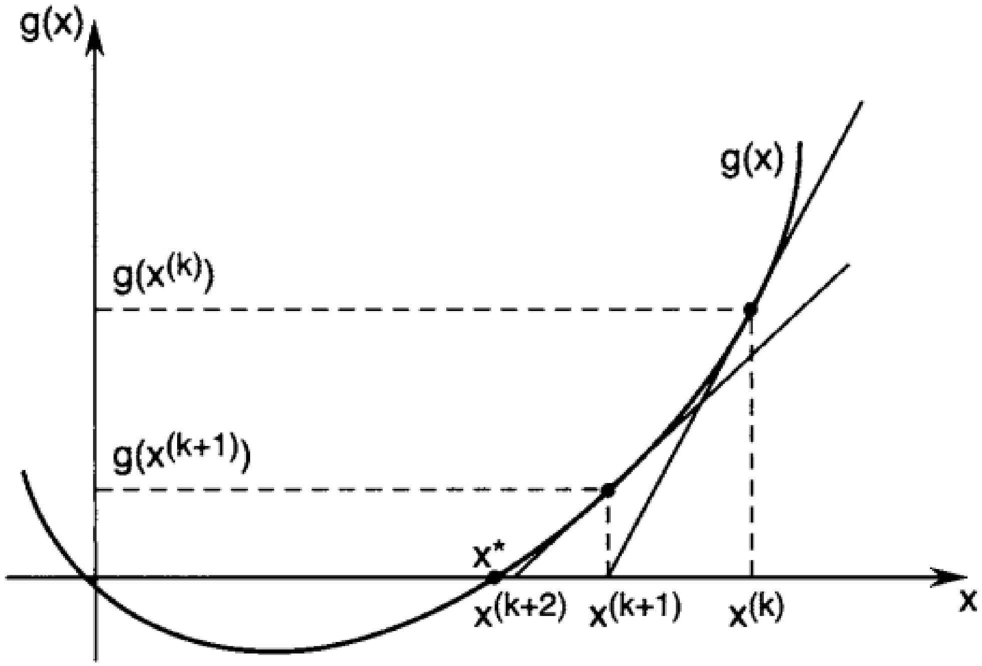


Figure 7.8 Newton's method of tangents.

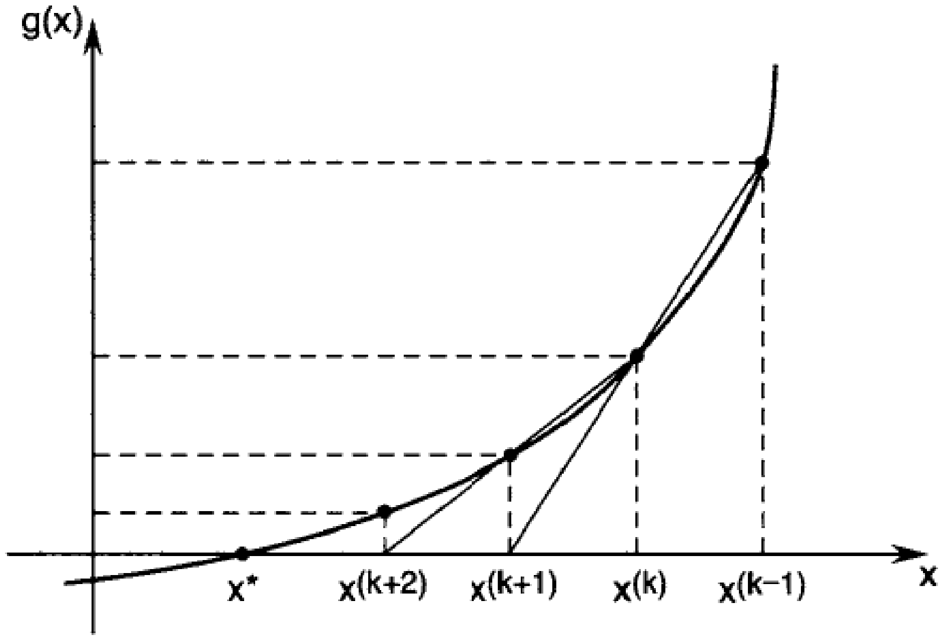
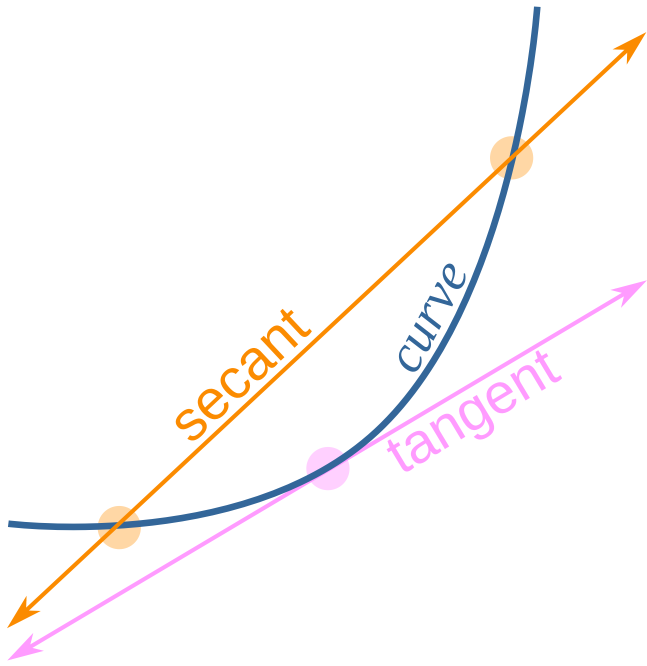


Figure 7.10 Secant method for root finding.

Newton vs Secante

O método de Newton e da Secante são ajustes quadráticos.

- No método de Newton, a aproximação quadrática ajusta f, f' e f'' em x_k.
- No método da Secante, a aproximação quadrática ajusta f em x_k e f' em x_k e x_{k-1}.



S. D. Conte and C. de Boor, Elementary Numerical Analysis: An Algorithmic Approach, 3rd ed. New York: McGraw-Hill Book Co., 1980.

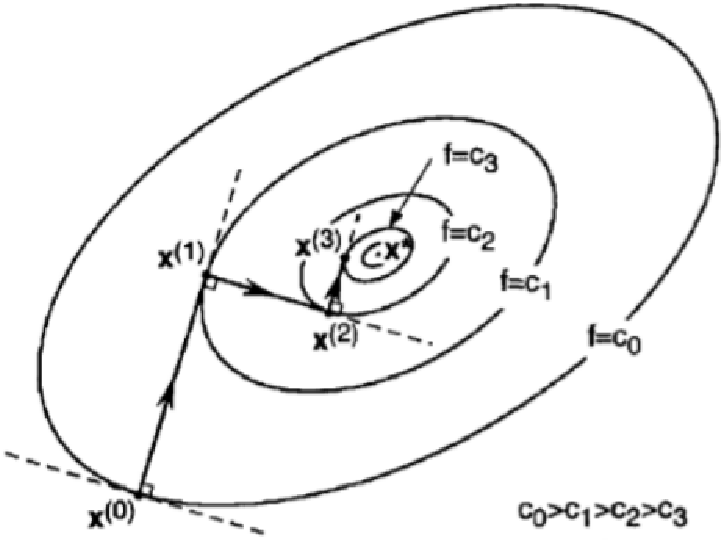
Otimização Inexata

No algoritmo geral, a otimização unidimensional é utilizada para encontrar o passo α_k que minimiza a função f ao longo da direção d_k.

$$\alpha_k = \arg\min_{\alpha} \phi_k(\alpha),$$

onde $\phi_k(\alpha) = f(x_k + \alpha d_k)$.

Na prática, a solução exata de cada problema unidimensional é custosa e desnecessária, portanto, podemos utilizar métodos aproximados.



S. D. Conte and C. de Boor, Elementary Numerical Analysis: An Algorithmic Approach, 3rd ed. New York: McGraw-Hill Book Co., 1980.

Regra de Armijo

A regra de Armijo é um critério de parada, onde

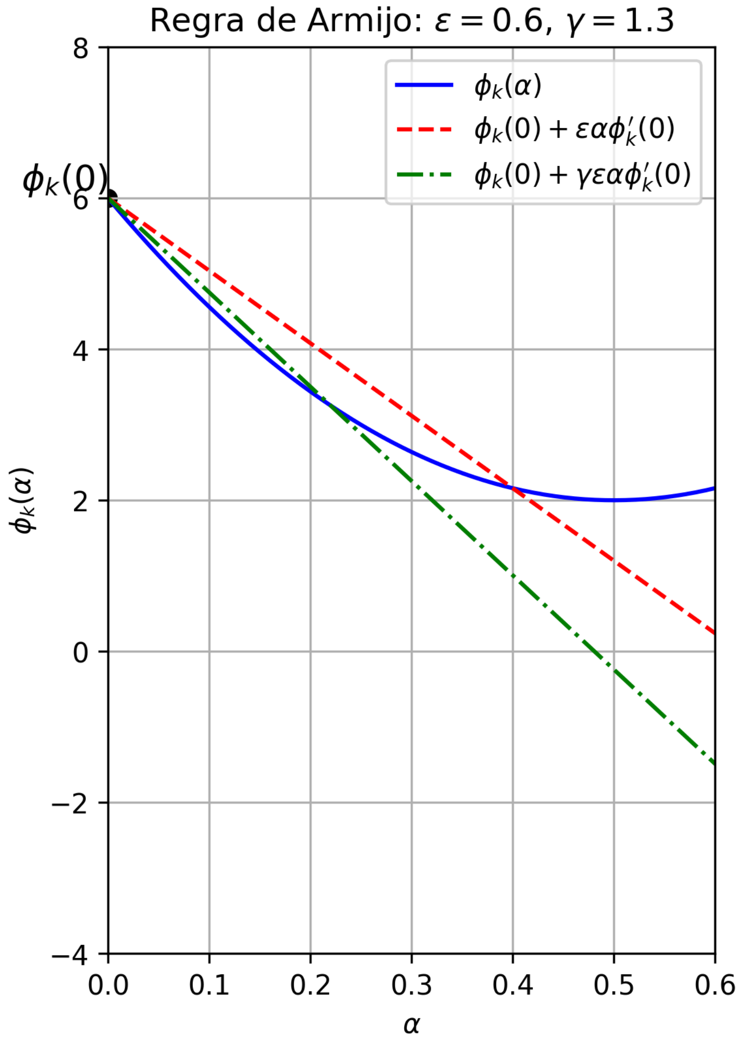
- um passo relativamente menor não reduz significativamente o valor da função.

$$\phi_k(\alpha_k) \leq \phi_k(0) + \epsilon \alpha_k \phi_k'(0),$$

- um passo relativamente maior não reduz significativamente o valor da função.

$$\phi_k(\gamma \alpha_k) \geq \phi_k(0) + \gamma \epsilon \alpha_k \phi_k'(0),$$

onde $\epsilon \in (0, 1)$ e $\gamma > 1$ são constantes.



S. D. Conte and C. de Boor, Elementary Numerical Analysis: An Algorithmic Approach, 3rd ed. New York: McGraw-Hill Book Co., 1980.

Perguntas?

S. D. Conte and C. de Boor, Elementary Numerical Analysis: An Algorithmic Approach, 3rd ed. New York: McGraw-Hill Book Co., 1980.