Otimização Não Linear

CM106/CMM204/CMI043

Tópico 04 - Controle de Passos

Abel Soares Siqueira - UFPR

2020/s1

Controle de Passos

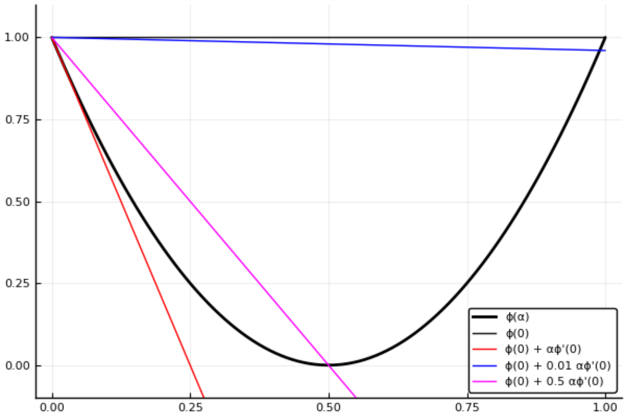
- Vimos que o método do gradiente precisa de controle de passos para funcionar.
- Mesmo Newton pode ser beneficiado por isso.
- Uma maneira é utilizar a busca exata. Quais são outras?
- Aqui o conceito de funcionar é convergência para um ponto crítico.
- **Def.:** Um método é **globalmente convergente** se para qualquer $x_0 \in \mathbb{R}^n$, o método a partir de x_0 gera uma sequência com uma subsequência convergente para um ponto crítico.
- Note que não quer dizer que irá convergir para um minimizador global, mas idealmente para um minimizador local.

Problema: decréscimo

- Uma condição necessária em todos os métodos básicos é que o valor de função diminua.
- Isso é óbvio dado que queremos minimizar a função.
- No entanto, não é suficiente exigir que $f(x_{k+1}) < f(x_k)$.
- Para $f(x) = x^2$, a sequência $x_k = 1 + k^{-1}$ sempre satisfaz $f(x_{k+1}) < f(x_k)$, porém $x_k \to 1$, que não é o minimizador.
- Aqui o problema é que os passos são muito pequenos, mas $x_k = (-1)^k(1 + k^{-1})$ tem o mesmo problema.

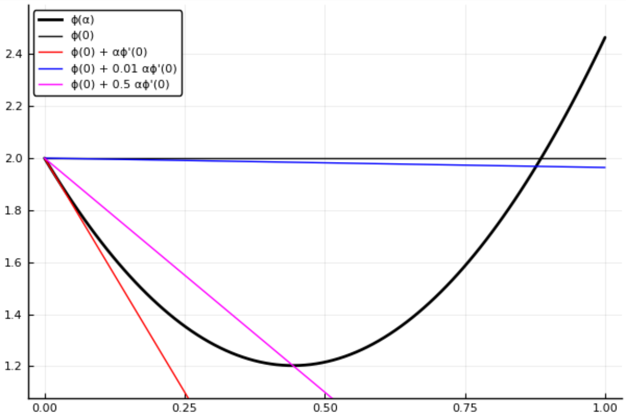
Decréscimo suficiente - condição de Armijo

$$\phi(\alpha) = f(x_k + \alpha d_k) < f(x_k) + \alpha\eta\nabla f(x_k)^T d_k = \phi(0) + \eta\alpha\phi'(0).$$



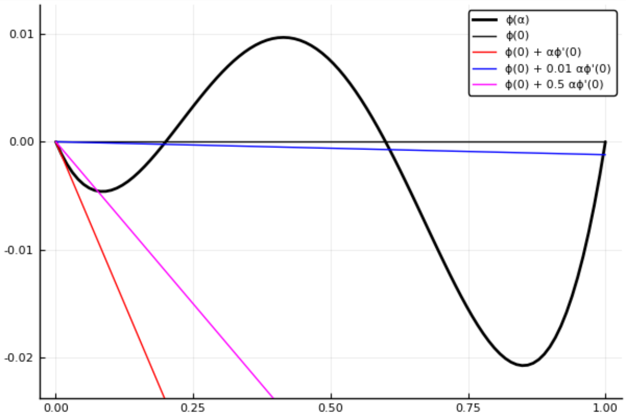
Decréscimo suficiente - condição de Armijo

$$\phi(\alpha) = f(x_k + \alpha d_k) < f(x_k) + \alpha\eta\nabla f(x_k)^T d_k = \phi(0) + \eta\alpha\phi'(0).$$



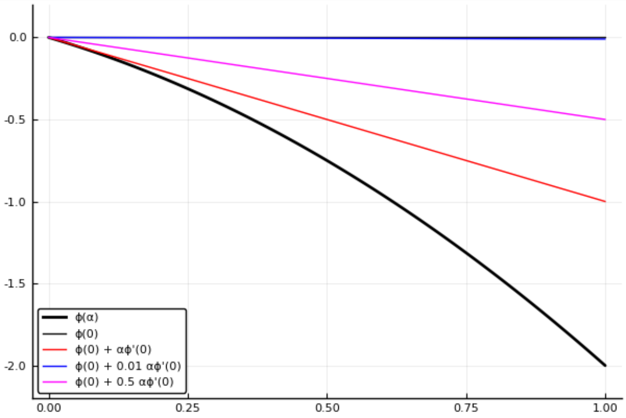
Decréscimo suficiente - condição de Armijo

$$\phi(\alpha) = f(x_k + \alpha d_k) < f(x_k) + \alpha\eta\nabla f(x_k)^T d_k = \phi(0) + \eta\alpha\phi'(0).$$



Decréscimo suficiente - condição de Armijo

$$\phi(\alpha) = f(x_k + \alpha d_k) < f(x_k) + \alpha\eta\nabla f(x_k)^T d_k = \phi(0) + \eta\alpha\phi'(0).$$



Decréscimo suficiente - condição de Armijo

- Para um método com passo $d_k = -\nabla f(x_k)$, quer dizer

$$f(x_k + \alpha d_k) < f(x_k) - \alpha\eta\|\nabla f(x_k)\|^2.$$

- Quanto mais longe de crítico, melhor deve ser o passo.
- De uma maneira geral $d_k^T\nabla f(x_k) < 0$, pois o passo é de descida.
- Pode acontecer, no entanto, de d_k ser muito pequeno ou muito ortogonal, i.e., $d_k^T\nabla f(x_k)$ muito pequeno.
- Isso poderia levar a $\nabla f(x_k)^T d_k \to 0$ antes de $\nabla f(x_k) \to 0$.
- Para escolher α fazemos *backtracking*, começamos com $\alpha_k = 1$, e testamos se a condição vale. Senão, testamos com $\alpha_k = \tau$, depois τ^2, e assim por diante, para $\tau \in (0, 1)$.

Métodos de direção de descida com busca linear

- Dados: $\sigma > 0$ e $\theta, \eta, \tau \in (0, 1)$, $x_0 \in \mathbb{R}^n$.
- A partir de $k = 0$, defina as iterações da seguinte maneira:
- Escolha d_k tal que

$$\|d_k\| \geq \sigma \|\nabla f(x_k)\| \qquad \text{e} \qquad \nabla f(x_k)^T d_k \leq -\theta \|\nabla f(x_k)\| \|d_k\|.$$

- Escolha $\alpha_k = \tau^p$ com p é o menor inteiro não-negativo tal que vale

$$f(x_k + \alpha_k d_k) < f(x_k) + \alpha_k \eta \nabla f(x_k)^T d_k.$$

- Defina $x_{k+1} = x_k + \alpha_k d_k$.

Métodos de direção de descida com busca linear

Teo. (Conv. Global): Seja x_k gerado pelo algoritmo anterior. Então encontramos $\nabla f(x_k) = 0$ para algum k finito, ou então o algoritmo gera uma sequência infinita tal que $\liminf \|\nabla f(x_k)\| = 0$.

Teo.: O método do gradiente com busca linear é globalmente convergente.
Dem.: Como $d_k = -\nabla f(x_k)$, temos $\|d_k\| = \|\nabla f(x_k)\|$ e $d_k^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2$, ou seja valem as condições com $\sigma = \theta = 1$.

Métodos de direção de descida com busca linear

Teo.: O método de Newton com busca linear é globalmente convergente se o conjunto $S = \{x : f(x) \leq f(x_0)\}$ é compacto, $\nabla^2 f(x)$ é uniformemente definida positiva e limitada em S, i.e., existem $0 < \lambda_{\min} \leq \lambda_{\max}$ tais que

$$\lambda_{\min}\|v\|^2 \leq v^T \nabla^2 f(x) v \leq \lambda_{\max}\|v\|^2, \qquad \forall x \in S.$$

Dem.: Vamos escrever $g_k = \nabla f(x_k)$ e $B_k = \nabla^2 f(x_k)$. Como $B_k d_k = -\nabla f(x_k)$, temos

$$-g_k^T d_k = d_k^T B_k d_k \geq \lambda_{\min}\|d_k\|^2$$
$$-g_k^T d_k = g_k^T B_k^{-1} g_k \geq \frac{1}{\lambda_{\max}}\|g_k\|^2.$$

Logo, $-g_k^T d_k = \sqrt{(g_k^T d_k)^2} \geq \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}\|d_k\|\|g_k\|$. Então vale a condição com $\theta = \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}} \in (0,1)$. Também vale $\|d_k\| \geq \frac{1}{\lambda_{\min}}\|g_k\|$. FIM.

Região de confiança

Região de confiança

- Os métodos de busca linear acham d_k e depois controlam o tamanho do passo.
- Métodos de região de confiança trabalham limitando o tamanho do passo durante a obtenção de d_k.
- A ideia original volta na aproximação quadrática em torno do ponto x_k:
$$q_k(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T B_k (x - x_k).$$
- Aqui B_k pode ser a Hessiana para uma aproximação quadrática de Taylor de segunda ordem, mas veremos outras opções viáveis.
- Também definimos, para simplificação, $m_k(d) = q_k(x_k + d)$, ou seja, a aproximação em função do passo.

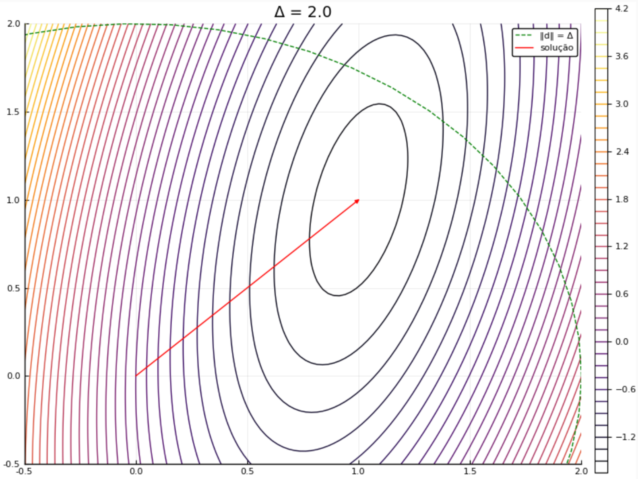
Região de confiança

- Para encontrar Newton antes, definimos d_k como o minimizador de m_k.
- Num método de região de confiança, definimos um raio $\Delta_k > 0$, e buscamos d_k solução de

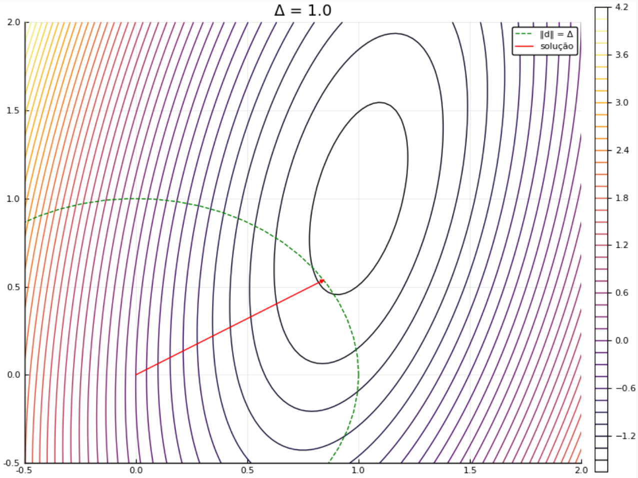
$$\min_d m_k(d) \qquad \text{suj. a} \qquad \|d\| \leq \Delta_k.$$

- Poderíamos a priori fazer $x_{k+1} = x_k + d_k$, mas isso não lida com o problema de decréscimo suficiente.

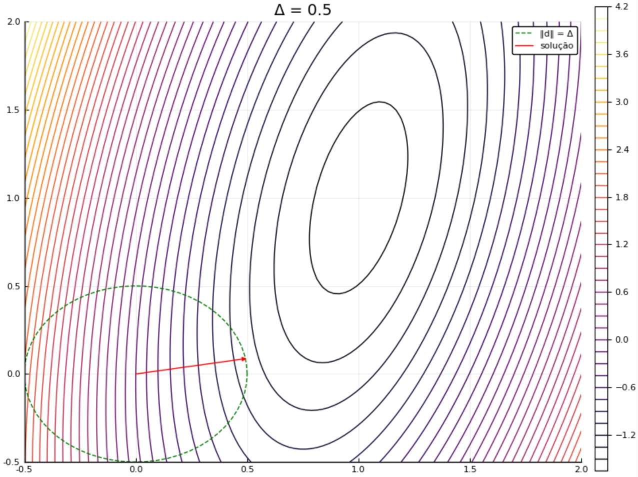
Região de confiança



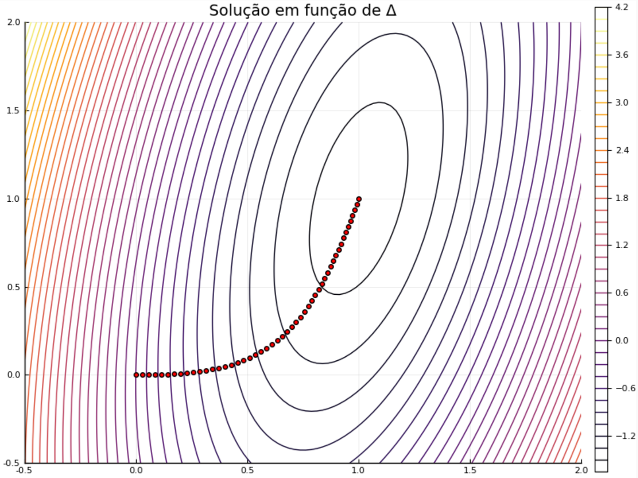
Região de confiança



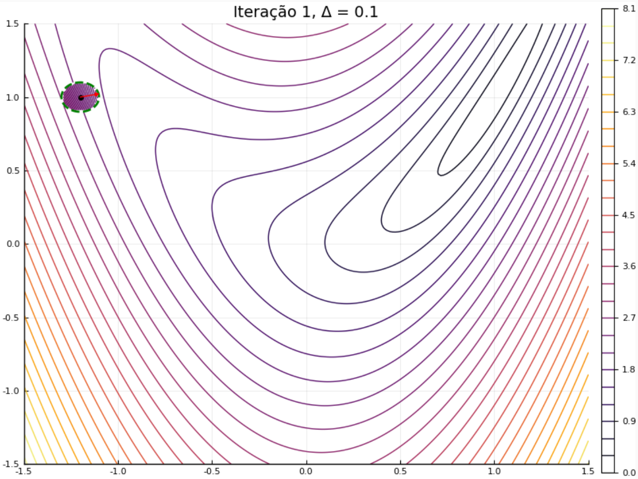
Região de confiança



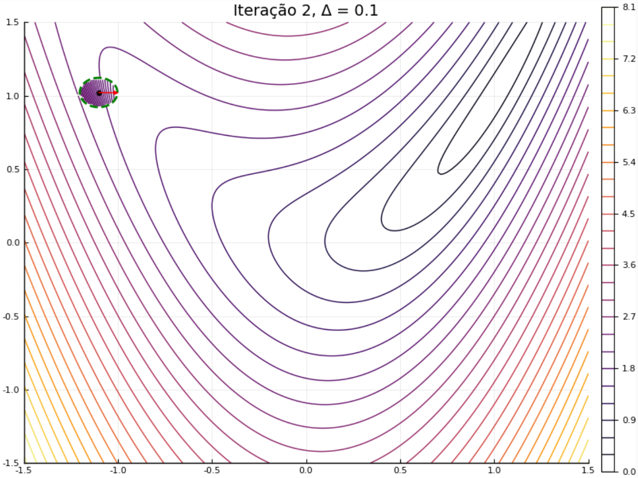
Região de confiança



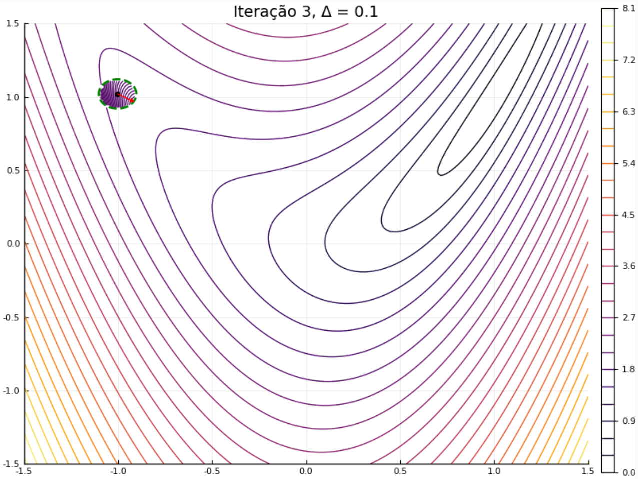
Região de confiança



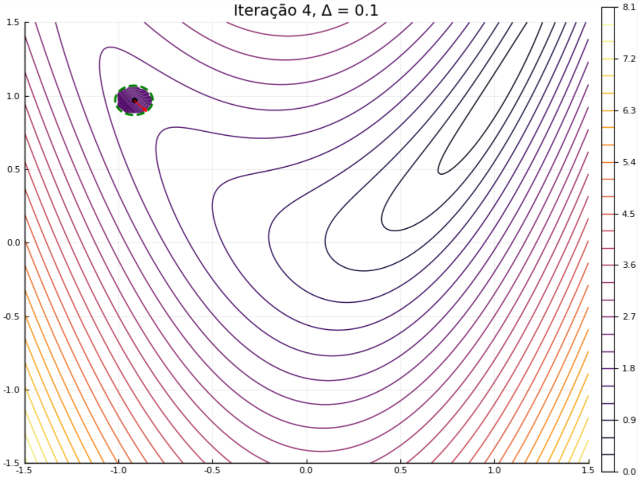
Região de confiança



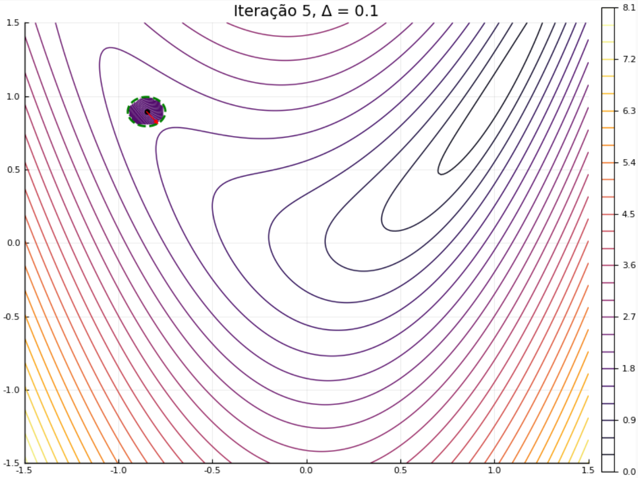
Região de confiança



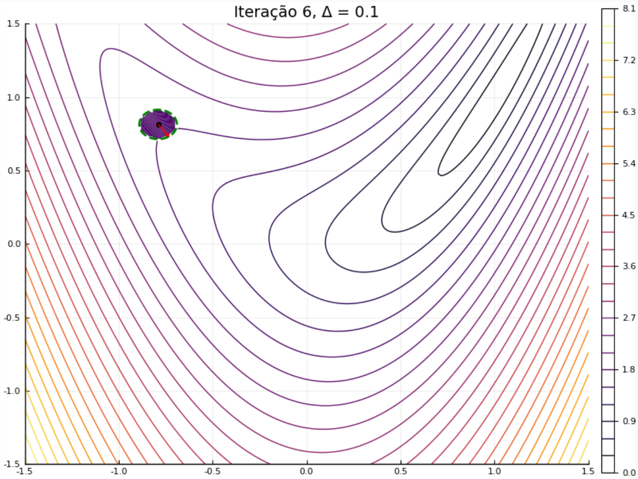
Região de confiança



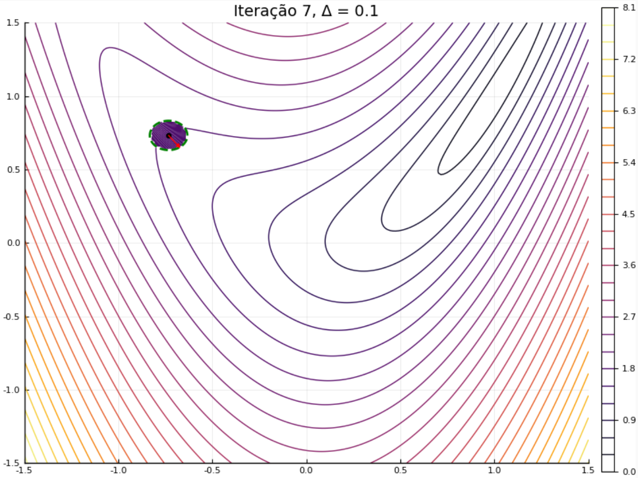
Região de confiança



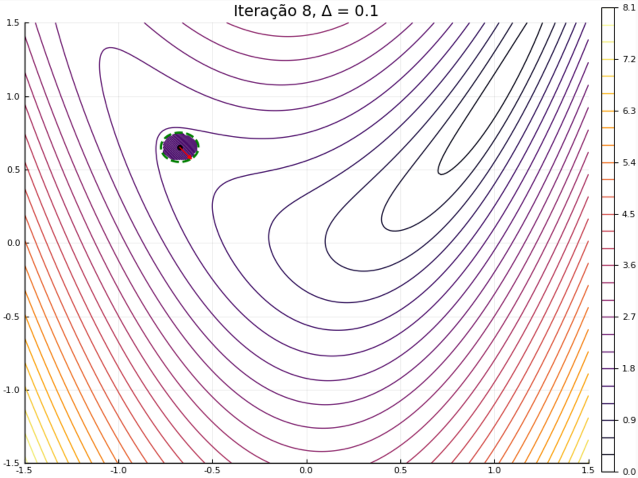
Região de confiança



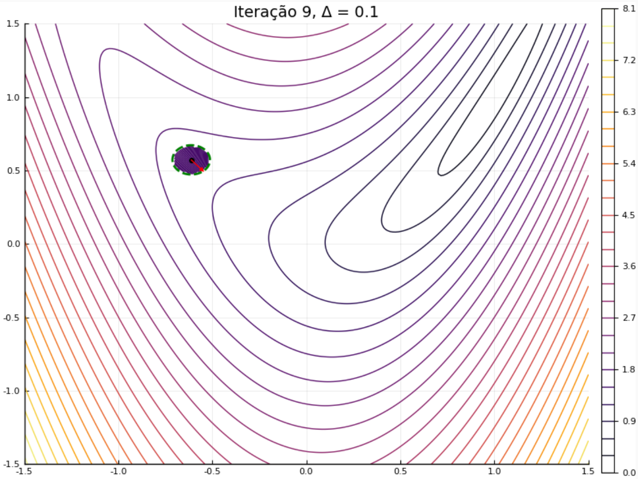
Região de confiança



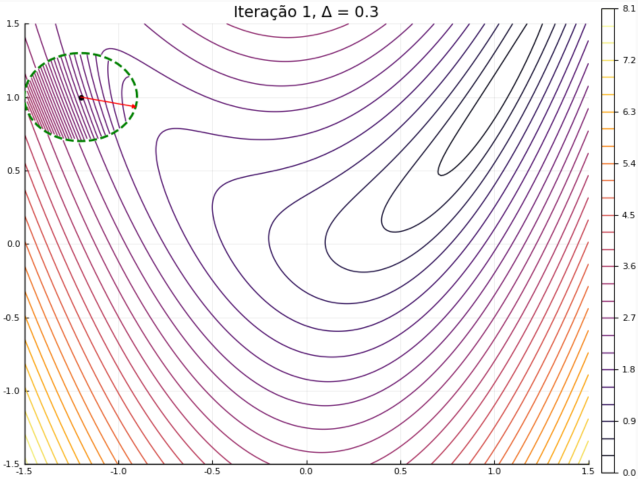
Região de confiança



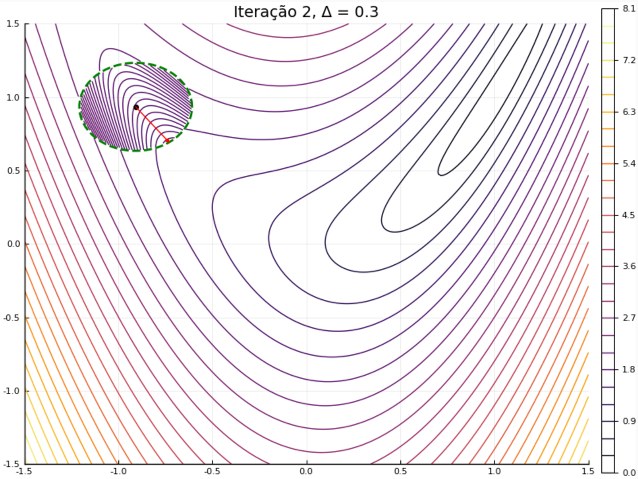
Região de confiança



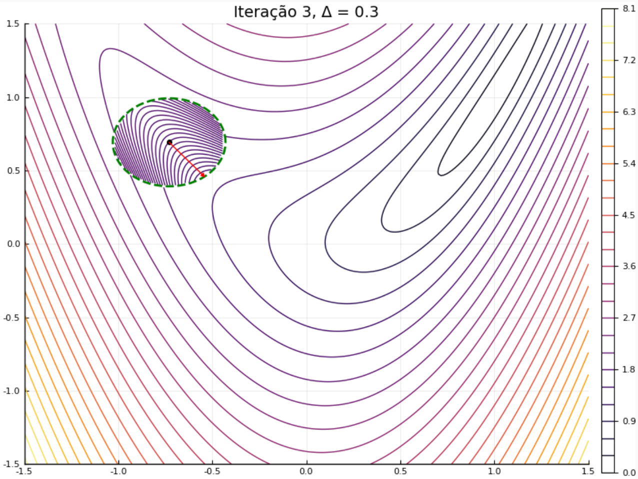
Região de confiança



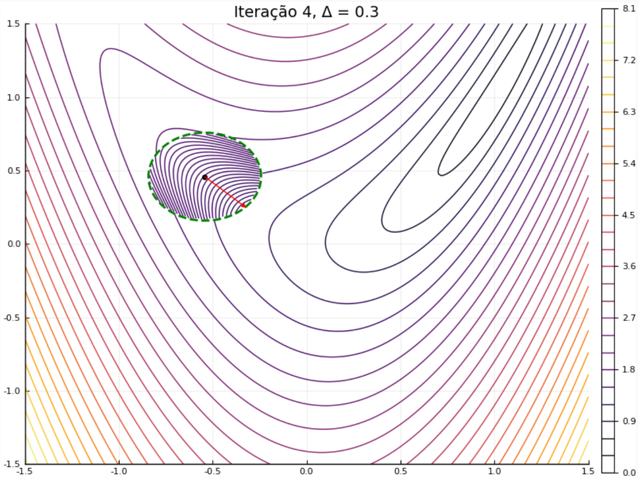
Região de confiança



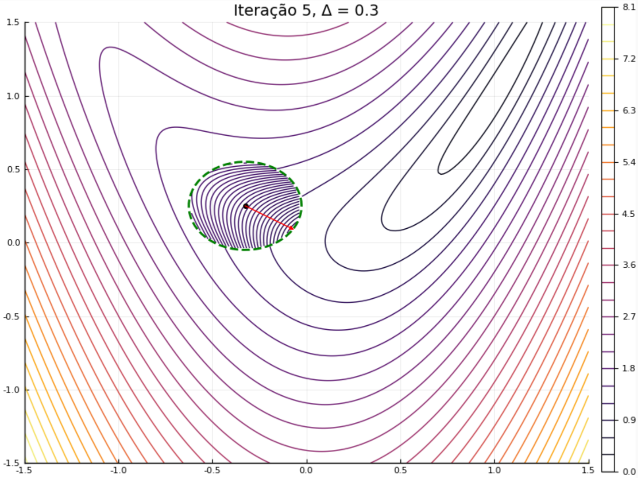
Região de confiança



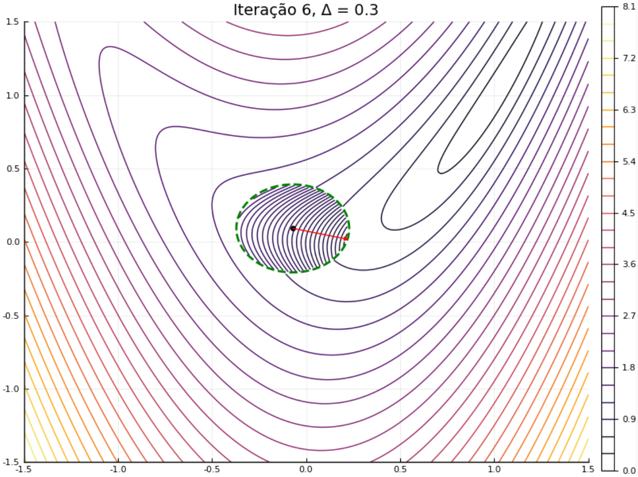
Região de confiança



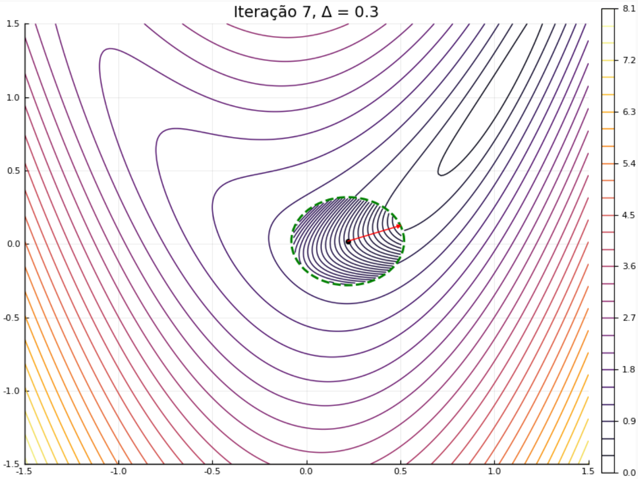
Região de confiança



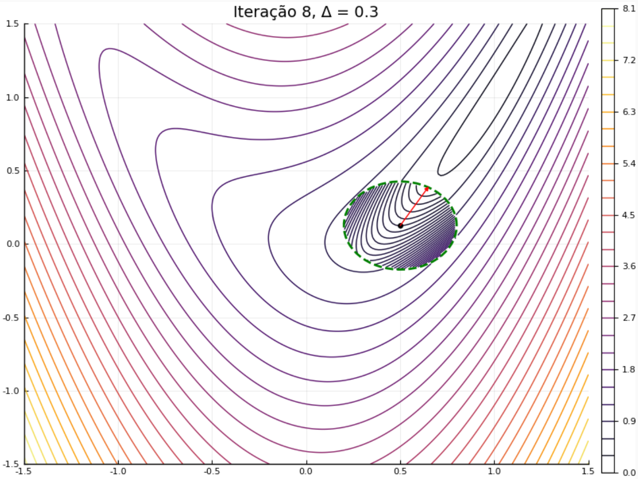
Região de confiança



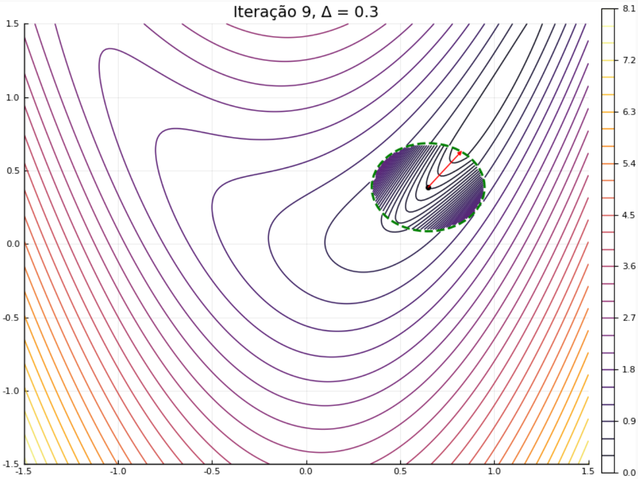
Região de confiança



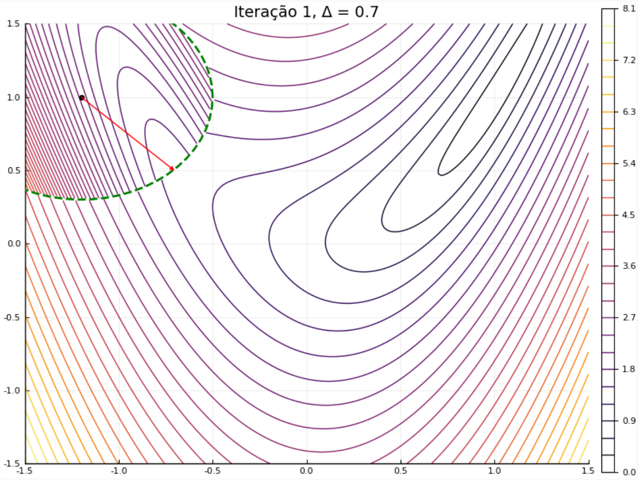
Região de confiança



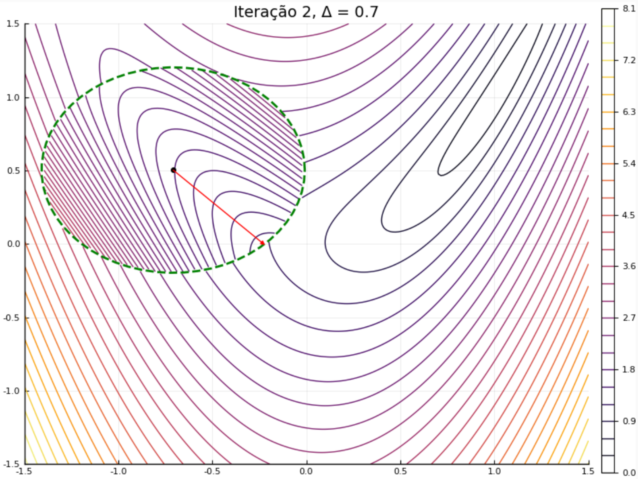
Região de confiança



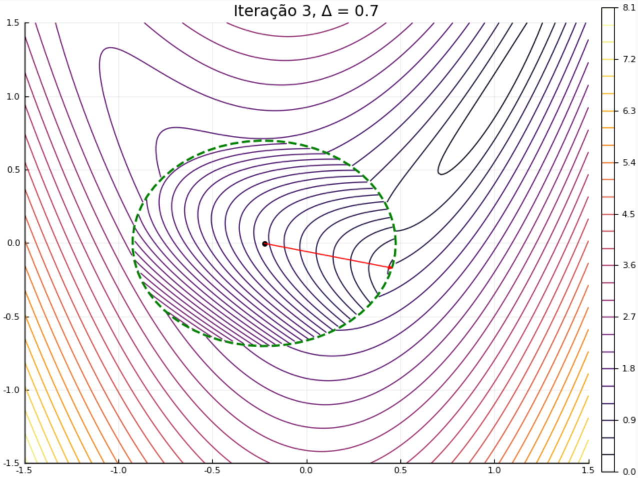
Região de confiança



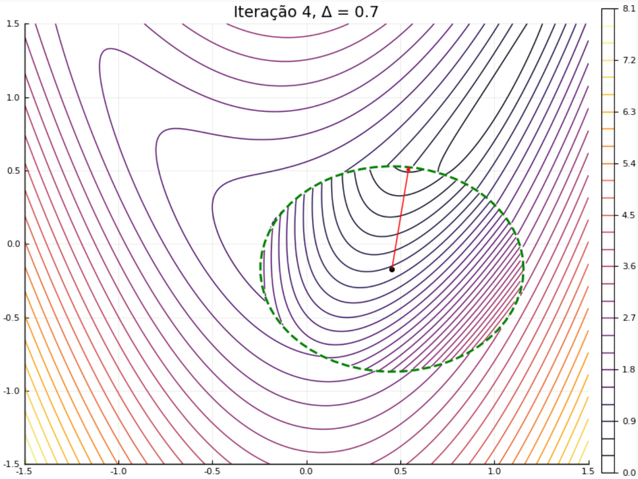
Região de confiança



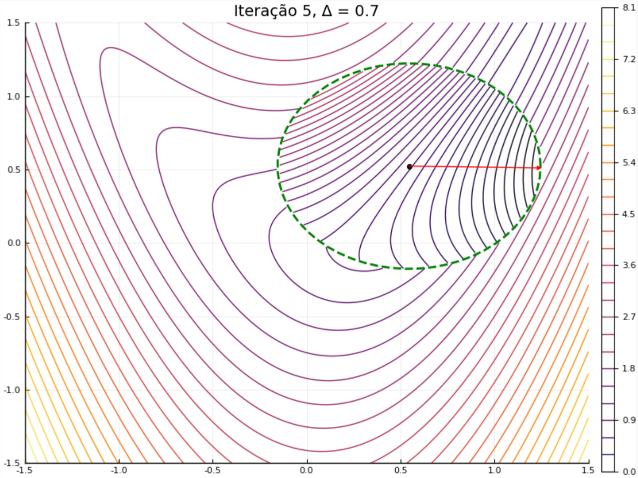
Região de confiança



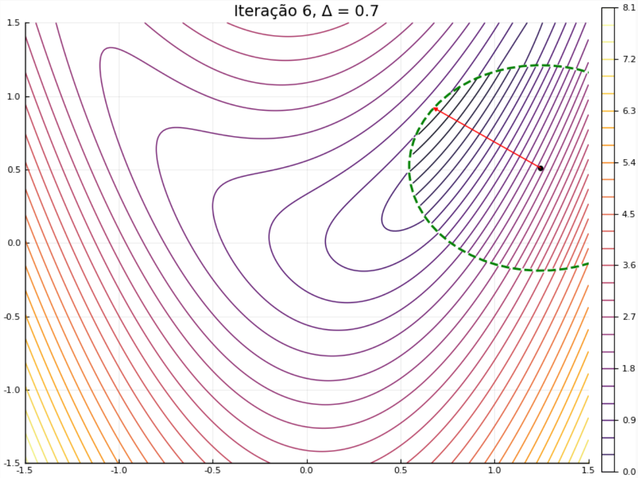
Região de confiança



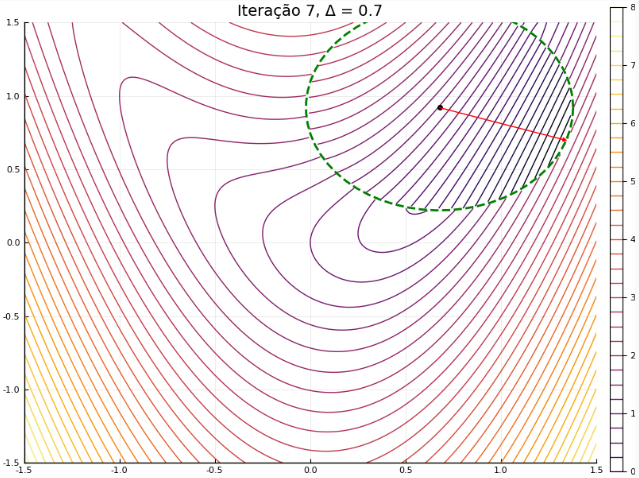
Região de confiança



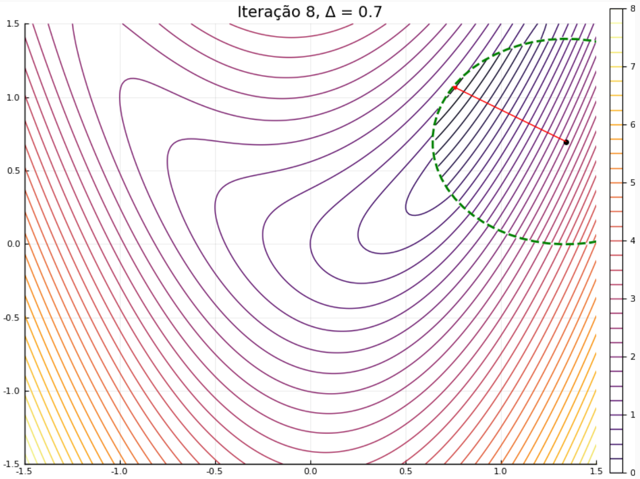
Região de confiança



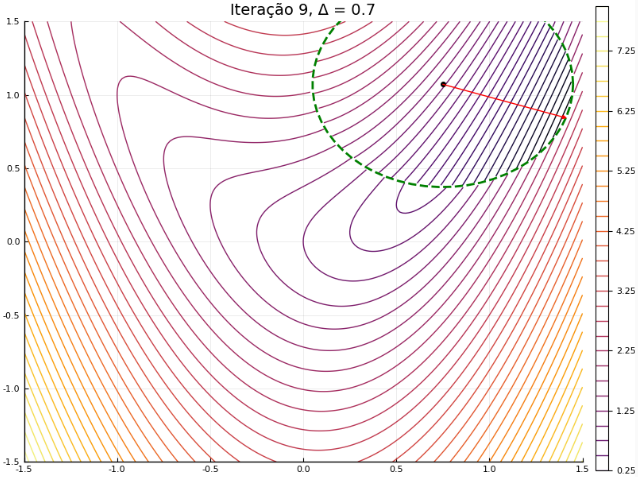
Região de confiança



Região de confiança



Região de confiança



Região de confiança

- A decisão da qualidade do passo é feita considerando as seguintes quantidades:

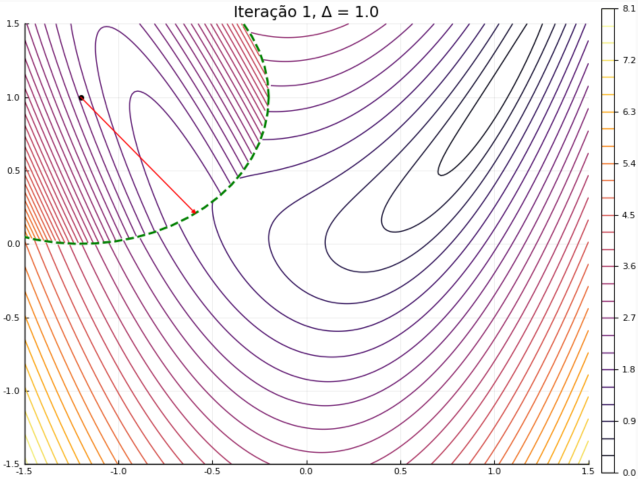
$$\text{Redução real} = \text{Ared} = f(x_k) - f(x_k + d_k)$$
$$\text{Redução prevista} = \text{Pred} = m(0) - m(d_k)$$

- Sempre Pred > 0 se o modelo for resolvido corretamente e x_k não é solução.
- $m(0) = f(x_k)$, então se $m(d_k) \approx f(x_k + d_k)$, a solução é boa.
- Se Ared for pequeno em relação a Pred (ou negativo), o modelo é ruim na região, então diminuímos Δ_k para ter mais confiança. Nessa situação, não atualizamos x, i.e., definimos $x_{k+1} = x_k$.
- Se Ared for muito bom em relação a Pred, então vamos atualizar $x_{k+1} = x_k + d_k$, mas também podemos aproveitar para aumentar Δ_k.

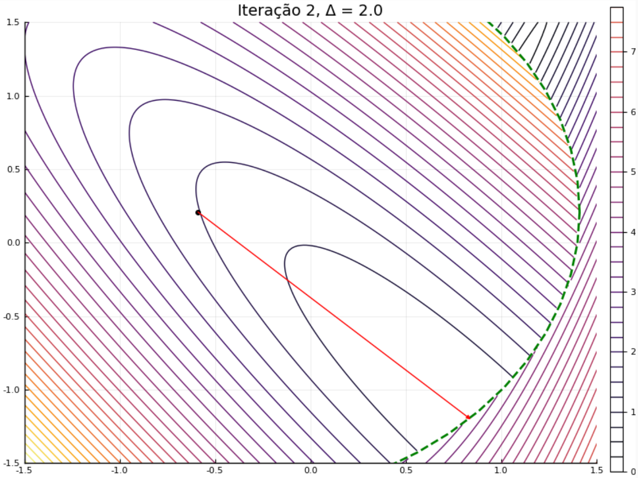
Região de confiança

- Razão: $\rho = \frac{\text{Ared}}{\text{Pred}}$
- Dados $0 < \eta_1 \leq \eta_2$, $0 < \sigma_1 < 1 < \sigma_2$.
- Se $\rho \leq \eta_1$, então $x_{k+1} = x_k$, e $\Delta_{k+1} = \sigma_1 \Delta_k$.
- Se $\eta_1 < \rho \leq \eta_2$, então $x_{k+1} = x_k + d_k$ e $\Delta_{k+1} = \Delta_k$.
- Se $\eta_2 < \rho$, então $x_{k+1} = x_k + d_k$ e $\Delta_{k+1} = \sigma_2 \Delta_k$.

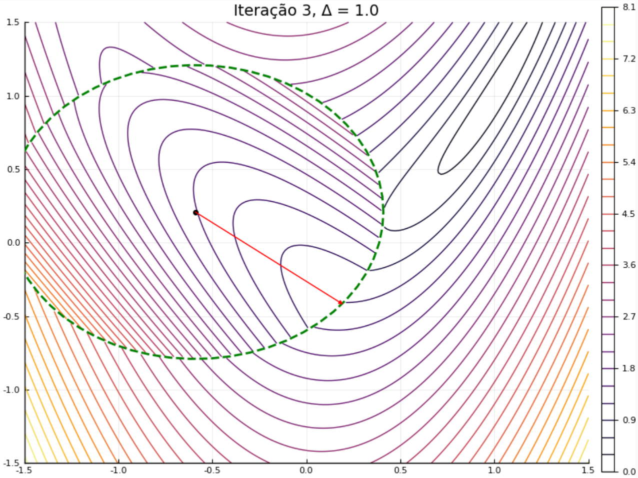
Região de confiança



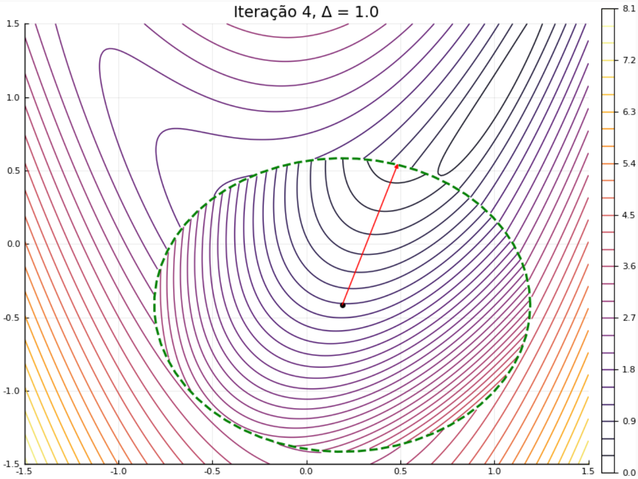
Região de confiança



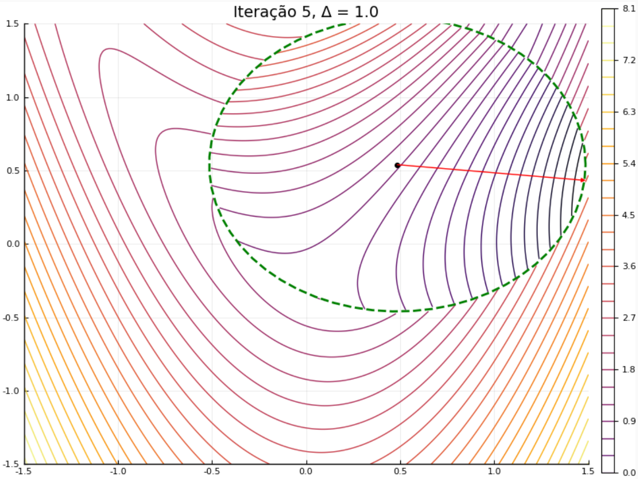
Região de confiança



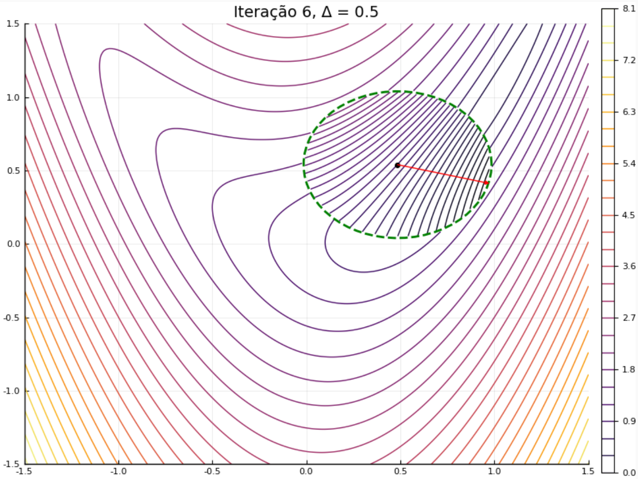
Região de confiança



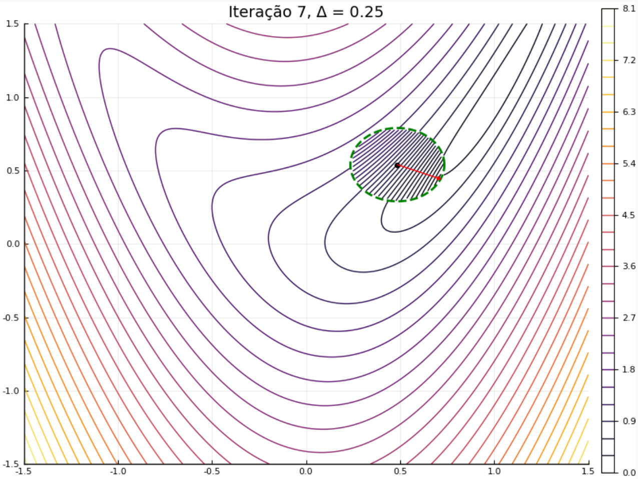
Região de confiança



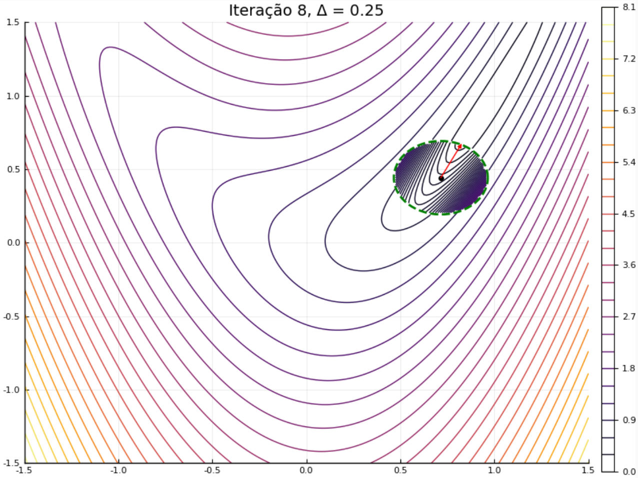
Região de confiança



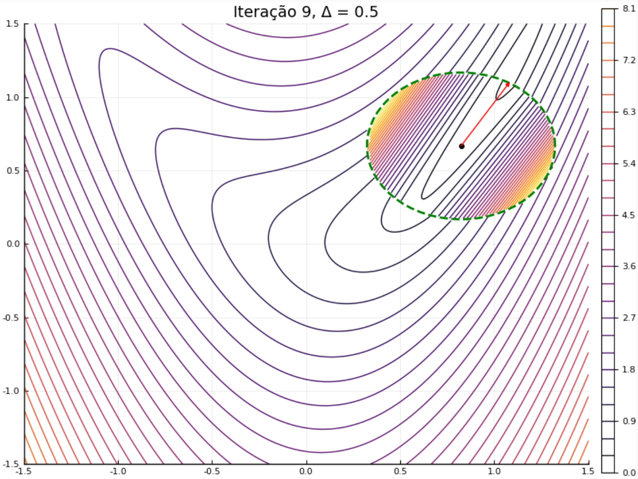
Região de confiança



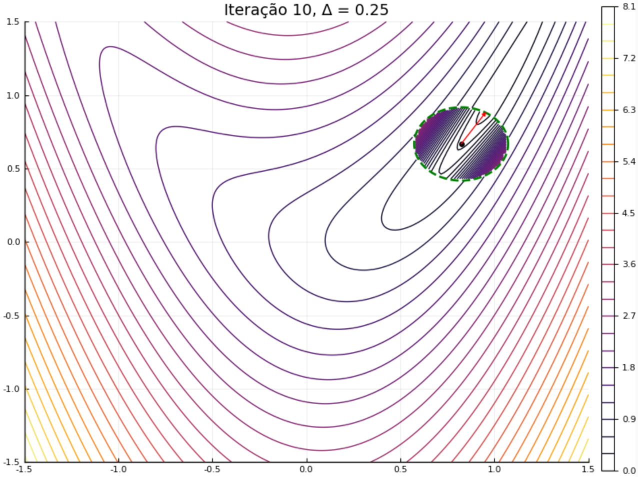
Região de confiança



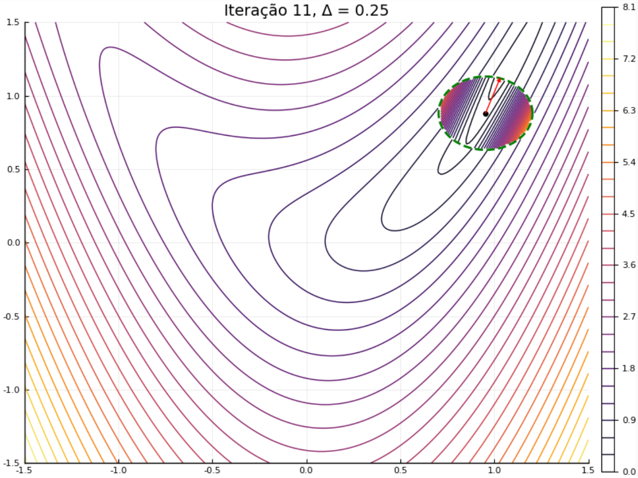
Região de confiança



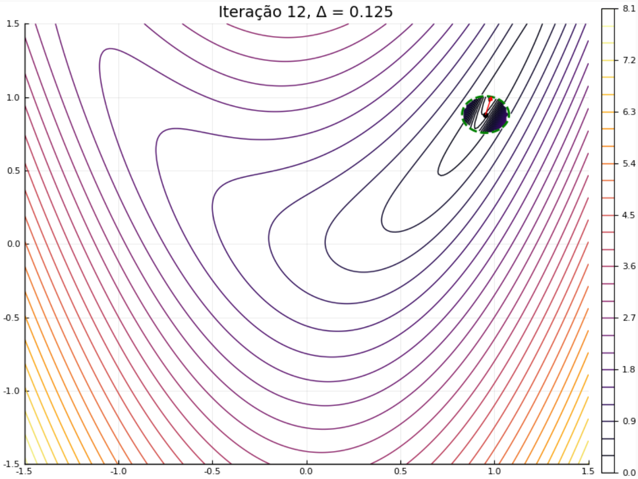
Região de confiança



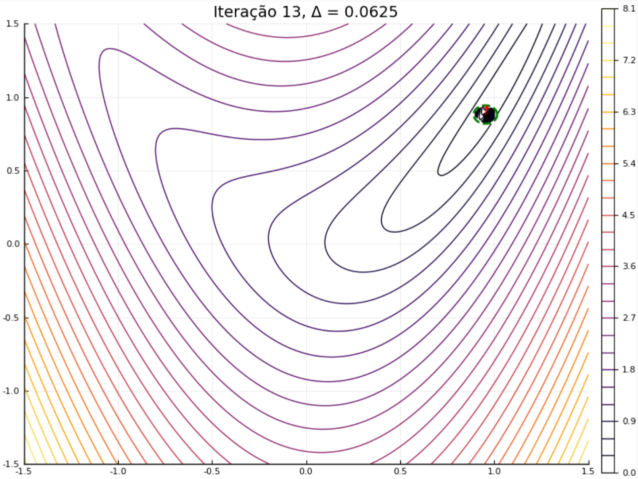
Região de confiança



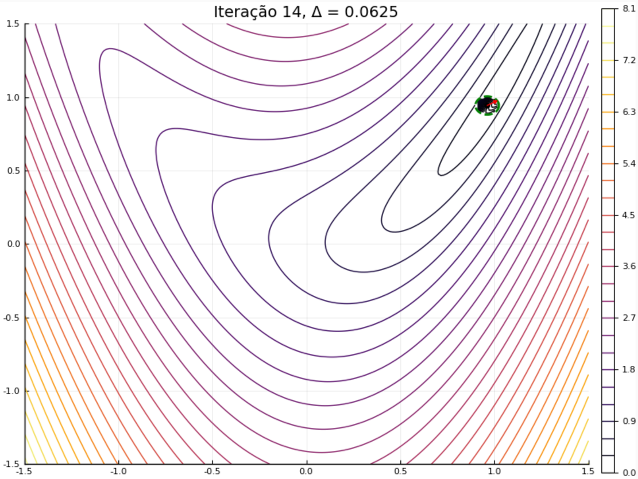
Região de confiança



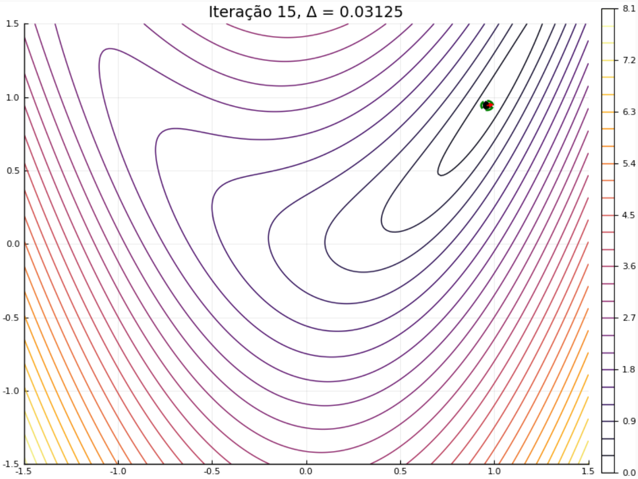
Região de confiança



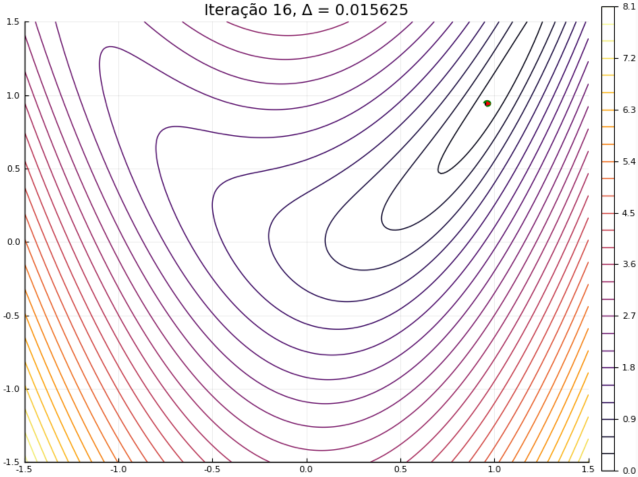
Região de confiança



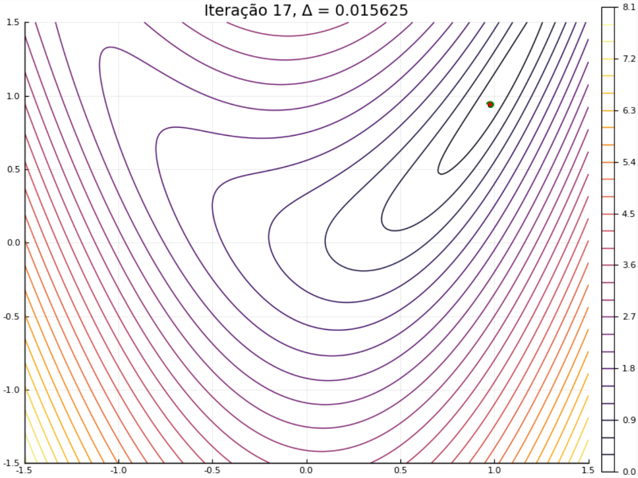
Região de confiança



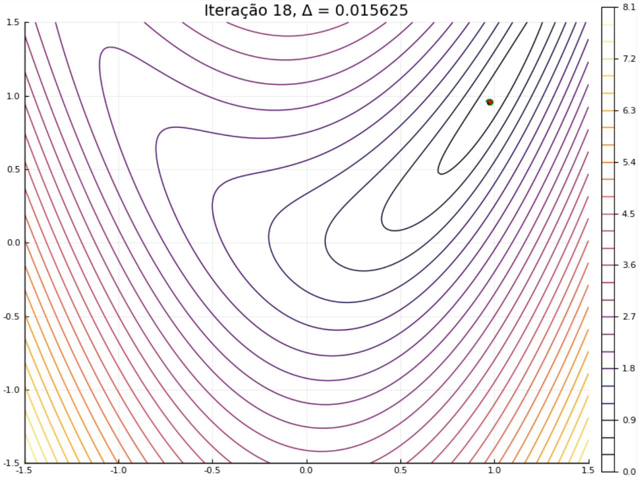
Região de confiança



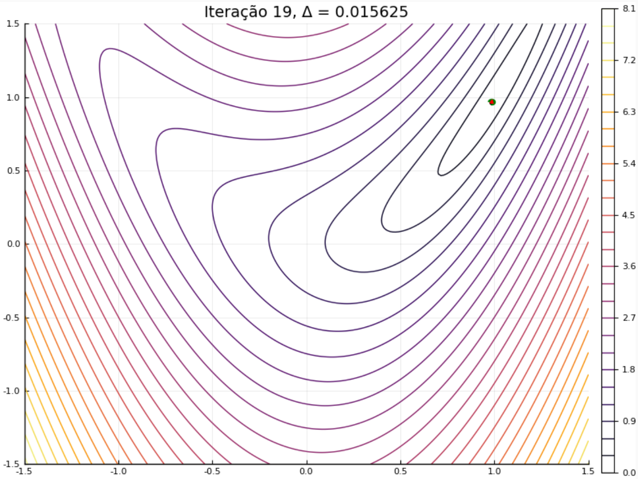
Região de confiança



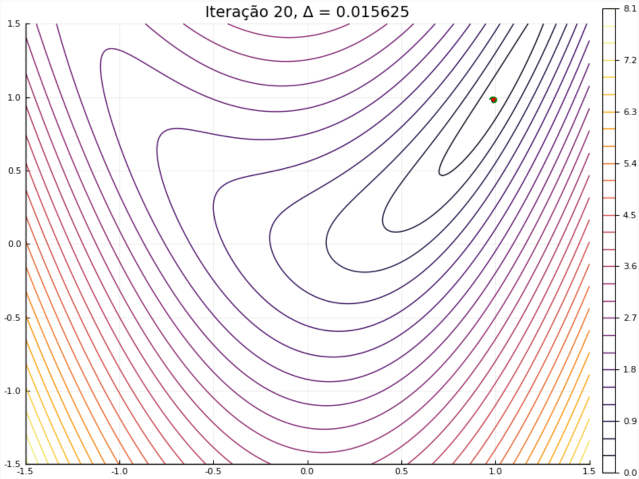
Região de confiança



Região de confiança



Região de confiança



Região de confiança

- $B_k = \nabla^2 f(x_k)$ é uma escolha natural para aproximação quadrática.
- $B_k = I$ é o equivalente ao método do gradiente, e outras escolhas são comuns também.
- Como resolver $\min_d m_k(d)$ sujeito a $\|d\| \leq \Delta_k$?
- A solução exata envolve a resolução de diversos sistemas lineares (método de Sorensen).
- Uma estratégia para quadraticas convexas é o método dogleg.
- Uma estratégia mais geral é uma variante do método de gradientes conjugados por Steihaug e Toint.

Sumário

- Para convergência global, precisamos controlar os passos.
- Busca linear consiste essencialmente de resover $B_k d_k = -\nabla f(x_k)$ e encontrar α_k para decréscimo suficiente usando a condição de Armijo

$$f(x_k + \alpha_k d_k) < f(x_k) + \alpha_k \eta \nabla f(x_k)^T d_k.$$

- Região de confiança consiste essencialmente de resolver $\min_d m_k(d)$ sujeito a $\|d\| \leq \Delta_k$ e controlar o aceite de d_k e o tamanho de Δ_k usando a razão

$$\frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)}.$$

- Encontrar d_k na região de confiança é um problema difícil.

FIM