Capítulo 5   Programação Não Linear

Quando o problema de otimização possui função objetivo e/ou funções de restrições não lineares, o problema é dito de **Programação Não Linear.**

Existem três tipos de métodos numéricos para solução de problemas de Programação Não Linear:

- baseados em gradientes;
- busca direta;
- métodos inspirados na natureza.

5.1 Otimização unidimensional: o segmento áureo

Nesta seção se discute a solução numérica do problema de encontrar máximos ou mínimos de uma função escalar de uma só variável, sem restrições. De partida, essa função será considerada não linear, já que o caso linear em uma dimensão não tem máximos ou mínimos finitos.

Um dos métodos numéricos mais utilizados para esse fim é o da busca utilizando o segmento áureo. A razão áurea foi primeiro definida por Euclides (cerca de 300 AC), na construção do pentagrama, e enunciada como: "uma linha reta é dita dividida em extrema e média razão quando a linha toda está para o maior segmento como esse último está para o menor". Sejam L_1 e L_2 o maior e o menor segmento, respectivamente. Fazendo

$$\varphi = \frac{L_1}{L_2} \tag{5.1}$$

tem-se, pelo enunciado de Euclides,

$$\varphi = \frac{L_1+L_2}{L_1} \tag{5.2}$$

resultando a equação de segundo grau

$$\varphi^2 - \varphi - 1 = 0 \tag{5.3}$$

cuja raiz positiva é a chamada “razão áurea”

$$\varphi = \frac{1+\sqrt{5}}{2} = 1{,}618034\ldots \tag{5.4}$$

Esse valor, as vezes conhecido como *divino*, aparece em um grande número de relações encontradas na natureza, nas ciências e nas artes. O leitor é convidado a fazer sua própria pesquisa sobre essa interessante recorrência.

O que se quer é determinar o mínimo de uma função unidimensional por um processo em etapas. Inicia-se definindo um intervalo x_L e x_U dentro do qual o mínimo é procurado. São necessários 2 pontos dentro desse intervalo para detectar a ocorrência de um mínimo, e eles serão escolhidos de acordo com a razão áurea,

$$x_a = x_L + d \quad e \quad x_b = x_U - d, \qquad d = (\varphi - 1)(x_U - x_L) \tag{5.5}$$

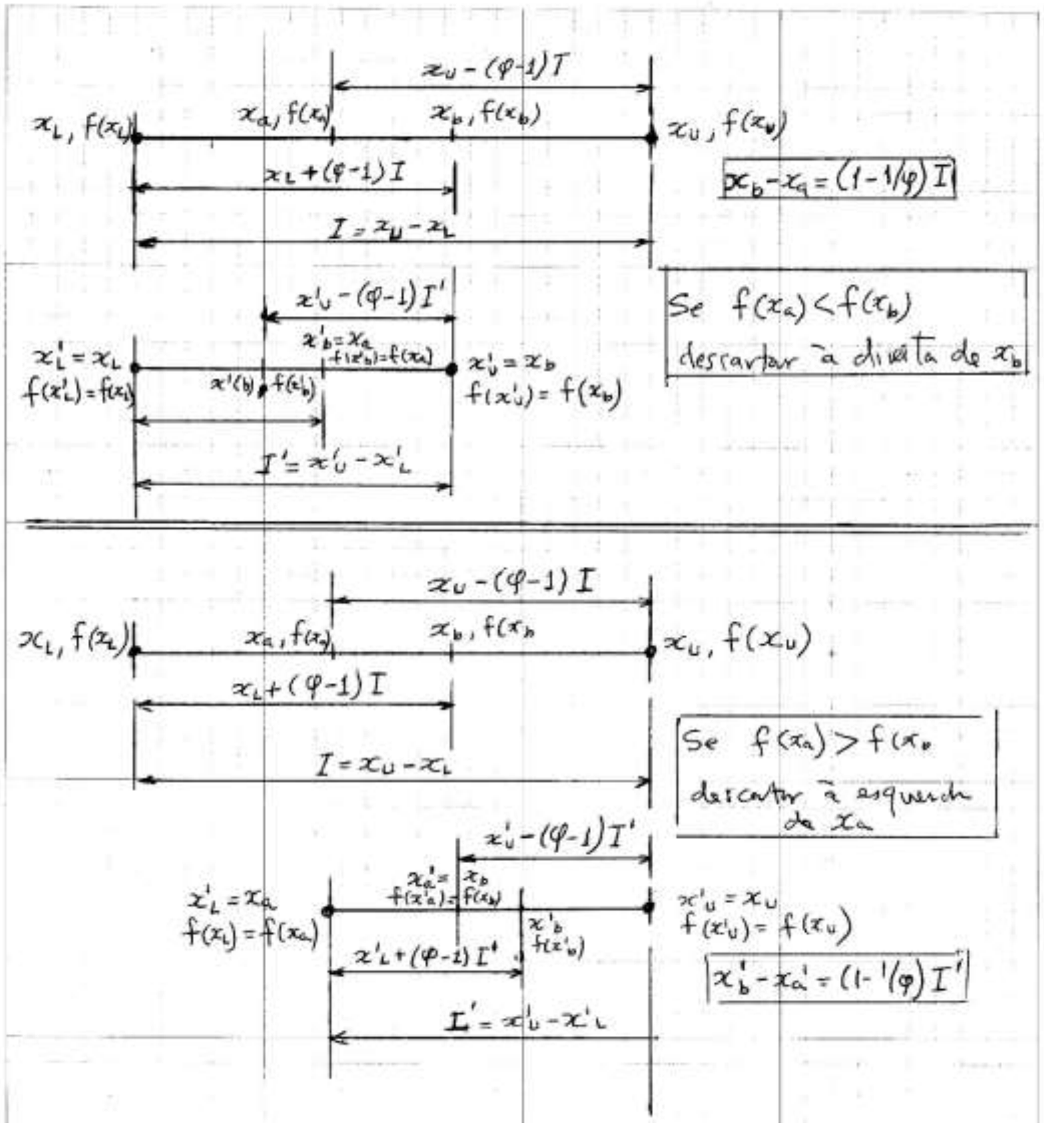
Calcula-se o valor da função nesses 2 pontos interiores. Dois resultados podem ocorrer.

1. Se $f(x_a) < f(x_b)$, então $f(x_a)$é o mínimo desse intervalo, e o domínio à esquerda de x_a, de x_L a x_b, pode ser eliminado da busca porque não contém o mínimo. Neste caso, x_b transforma-se no novo x_L para a próxima etapa.

2. Se $f(x_b) < f(x_a)$, então $f(x_b)$é o mínimo desse intervalo, e o domínio à direita de x_a, de x_a a x_U, pode ser eliminado da busca porque não contém o mínimo. Neste caso, x_a transforma-se no novo x_U para a próxima etapa.

Como os valores de x_a e x_b foram escolhidos usando a razão áurea, não é necessário calcular todos os valores da função na próxima iteração. Por exemplo, se ocorrer a hipótese 1 acima, o antigo x_a passa a ser o novo x_b, e, assim, o novo $f(x_b)$não precisa ser calculado, pois é o mesmo que o antigo $f(x_a)$. Para completar o algoritmo, determina-se o novo x_a, aplicando a Eq. (6.8), baseando-se nos novos valores de x_L e x_U . No caso 2, o procedimento é análogo. Demonstra-se que, em cada iteração, o intervalo é reduzido por um percentual de cerca de 61,8%. Assim, por exemplo, depois de 10 iterações, o intervalo diminuiu de cerca de $0{,}618^{10}$, 0,8% de seu comprimento original.

Se for desejado encontrar o máximo de uma função $f(x)$, em vez do mínimo, basta determinar o mínimo dessa função com sinal trocado, $F(x) = -f(x)$.


xL, f(xL)
xa, f(xa)
xb, f(xb)
xu, f(xu)
xu - (φ-1) I
xL + (φ-1) I
I = xu - xL
xb - xa = (1 - 1/φ) I
Se f(xa) < f(xb)
descartar à direita de xb
x'u - (φ-1) I'
x'L = xL
f(x'L) = f(xL)
x'b = xa
f(x'b) = f(xa)
x'u = xb
f(x'u) = f(xb)
I' = x'u - x'L
Se f(xa) > f(xb)
descartar à esquerda de xa
x'L = xa
f(xL) = f(xa)
x'a = xb
f(x'a) = f(xb)
x'u = xu
f(x'u) = f(xu)
x'b
f(x'b)
x'L + (φ-1) I'
x'b - x'a = (1 - 1/φ) I'

Exemplo

Determinar o mínimo da função

$$f(x) = \frac{x^2}{10} - 2\ \text{sen}\ x$$

no intervalo $x_L = 0$ a $x_U = 4$.

1ª iteração

$$d = 0{,}61803(4 - 0) = 2{,}4721$$
$$x_1 = 0 + 2{,}4721 = 2{,}4721 \qquad x_2 = 4 - 2{,}4721 = 1{,}5279$$
$$f(x_2) = -1{,}7647 \qquad f(x_1) = -0{,}63, \quad \therefore \quad f(x_2) < f(x_1)$$

O mínimo atual é $f(x_2) = -1{,}7647$.

2aª iteração

$$d = 0{,}61803(2{,}4721 - 0) = 1{,}5279$$
$$x_1 = 1{,}5279 \quad x_2 = 2{,}4721 - 1{,}5279 = 0{,}9443$$
$$f(x_2) = -1{,}5310, \quad \therefore \quad f(x_2) > f(x_1)$$

O mínimo atual ainda é $f(x_1) = -1{,}7647$

Após 8 iterações, a localização do mínimo é aproximada por x = 1,4427, e o valor mínimo estimado da função é -1,7755.

A busca do mínimo utilizando a razão áurea pode ser uma rotina que se repete dentro de um algoritmo maior, como será visto a seguir.

5.2 Otimização multidimensional

Num problema multidimensional, ou seja, com várias variáveis de projeto, sem restrições, formalmente:

Achar $\mathbf{x}^*$ para minimizar $f(\mathbf{x})$,

em cada passo do processo k, conhecido o valor do vetor $\mathbf{x}^{(k)}$projeta-se um novo valor $\mathbf{x}^{(k+1)}$ por

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x} \tag{5.6}$$

onde

$$\Delta\mathbf{x} = \alpha_k \mathbf{d}^{(k)} \tag{5.7}$$

em que $\mathbf{d}^{(k)}$ é a direção desejada e α_k é o tamanho do passo nessa direção. Para isso, utilizar o algoritmo que se segue.

Algoritmo

1. Estimar um projeto inicial razoável $\mathbf{x}^{(0)}$, k =0,
2. computar a direção $\mathbf{d}^{(k)}$ no ponto $\mathbf{x}^{(k)}$,
3. checar a convergência; se atingida parar, senão
4. calcular um tamanho de passo positivo $\alpha_k$na direção $\mathbf{d}^{(k)}$, e $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x}$,
5. $k = k + 1$ e vá para o passo 2.

Para minimização da função objetivo, em cada passo seu valor deve diminuir

$$f\left(\mathbf{x}^{(k+1)}\right) < f\left(\mathbf{x}^{(k)}\right) \tag{5.8}$$

$$f(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) < f\left(\mathbf{x}^{(k)}\right) \tag{5.9}$$

Aproximando o lado esquerdo em série de Taylor em torno do ponto $\mathbf{x}^{(k)}$

$$f\left(\mathbf{x}^{(k)}\right) + \alpha_k(\mathbf{c}^{(k)} \cdot \mathbf{d}^{(k)}) < f\left(\mathbf{x}^{(k)}\right) \tag{5.10}$$

onde $\mathbf{c}^{(k)}$ é o gradiente da função objetivo nesse ponto, e o produto escalar indicado pelo ponto entre os vetores entre parêntesis deve ser negativo,

$$(\mathbf{c}^{(k)} \cdot \mathbf{d}^{(k)}) < 0 \tag{5.11}$$

uma vez que α_k foi definido como sempre positivo. Isso implica em que o ângulo entre esses dois vetores, o gradiente e direção desejável, deve estar entre 90º e 270º .

Qualquer vetor $\mathbf{d}^{(k)}$satisfazendo essa desigualdade, a **condição de descida** da função objetivo, é uma **direção desejável**. Métodos deste tipo são métodos de descida "morro abaixo", em que se quer atingir o fundo do vale da função objetivo a partir de um ponto mais alto.

A determinação do comprimento do passo é uma busca unidimensional, ou em linha

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha\, \mathbf{d}^{(k)} \tag{5.12}$$

$$f\left(\mathbf{x}^{(k+1)}\right) = f(\mathbf{x}^{(k)} + \alpha\, \mathbf{d}^{(k)}) = \bar{f}(\alpha) \tag{5.13}$$

$\bar{f}(\alpha)$ é uma função apenas de α. Quando $\alpha = 0 \rightarrow \bar{f}(0) = f\left(\mathbf{x}^{(k)}\right)$. Logo, para que se diminua o valor da função objetivo, como desejado, $\bar{f}(\alpha)$ deve sempre decrescer. Trata-se de um problema de minimização unidimensional sem restrições. A condição necessária para um mínimo é

$$\frac{d\bar{f}}{d\alpha} = 0 \tag{5.14}$$

e a condição suficiente é

$$\frac{d^2\bar{f}}{d\alpha^2} > 0 \tag{5.15}$$

Simplificando a notação suprimindo a barra sobre a função, e aplicando a regra da cadeia,

$$\frac{df(\mathbf{x}^{(k+1)})}{d\alpha} = \frac{\partial f^T\left(\mathbf{x}^{(k+1)}\right)}{\partial \mathbf{x}} \frac{d\left(\mathbf{x}^{(k+1)}\right)}{d\alpha} = \nabla f\left(\mathbf{x}^{(k+1)}\right) \cdot \mathbf{d}^{(k)} = \mathbf{c}^{(k+1)} \cdot \mathbf{d}^{(k)} = 0 \tag{5.16}$$

Ou seja, que a direção desejada seja ortogonal ao vetor gradiente em cada passo.

5.3 Método do Gradiente: da maior declividade

Algoritmo

1. Estimar um projeto inicial $\mathbf{x}^{(0)}$, fazer $k = 0$, adotar um parâmetro de convergência pequeno $\varepsilon > 0$,
2. calcular o gradiente da função objetivo no ponto atual $\mathbf{x}^{(k)}$, $\mathbf{c}^{(k)} = \nabla f\left(\mathbf{x}^{(k)}\right)$,
3. calcular o comprimento do vetor gradiente, $\left\|\mathbf{c}^{(k)}\right\|$; se $\left\|\mathbf{c}^{(k)}\right\| < \varepsilon$, pare, pois $\mathbf{x}^{(k)} = \mathbf{x}^*$, senão, continue,
4. fazer a direção de busca neste passo $\mathbf{d}^{(k)} = -\mathbf{c}^{(k)}$,
5. calcular o tamanho do passo neste ponto, minimizando $f(\alpha) = f\left(\mathbf{x}^{(k)}\right) + \alpha\, \mathbf{d}^{(k)}$, pelo segmento áureo, por exemplo,
6. atualizar o projeto $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha\, \mathbf{d}^{(k)}$, fazer $k = k + 1$ e voltar ao passo 2.

Exemplo 1

Minimizar

$$f(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1x_2$$

projeto inicial dado $\mathbf{x}^{(0)} = [1 \quad 0]^T$, $k = 0$ e $\varepsilon = 10^{-4}$,

$\mathbf{c}^{(0)} = [2x_1 - 2x_2 \quad 2x_2 - 2x_1]^T = [2 \quad -2]^T$

$\left\|\mathbf{c}^{(0)}\right\| = 2\sqrt{2} > \varepsilon$, continue,

$\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} = [-2 \quad 2]^T$

minimizar $f(\alpha) = f\left(\mathbf{x}^{(0)}\right) + \alpha\, \mathbf{d}^{(0)}$, onde $\mathbf{x}^{(0)} + \alpha\, \mathbf{d}^{(0)} = [1 - 2\alpha \quad 2\alpha]^T$

$$f(\alpha) = f\left(\mathbf{x}^{(0)}\right) + \alpha\, \mathbf{d}^{(0)} = (1 - 2\alpha)^2 + (2\alpha)^2 - 2(1 - 2\alpha)(2\alpha)$$

$$f(\alpha) = 16\alpha^2 - 8\alpha + 1$$

$$\frac{df}{d\alpha} = 0 \;\therefore\; \alpha_0 = 1/4$$

$$\frac{d^2f}{d\alpha^2} = 32 > 0$$

$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_0\,\mathbf{d}^{(0)} = [0.5 \quad 0.5]^T$

$\mathbf{c}^{(1)} = [0 \quad 0]^T$, $\left\|\mathbf{c}^{(1)}\right\| < \varepsilon$, pare.

Resposta: $\mathbf{x}^* = [0.5 \quad 0.5]^T$, e o valor mínimo da função objetivo é zero nesse ponto.

Exemplo 2

Minimizar

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2$$

projeto inicial dado $\mathbf{x}^{(0)} = [1 \quad 1 \quad 2]^T$, $k = 0$ e $\varepsilon = 10^{-4}$,

$\mathbf{c}^{(0)} = [2x_1 \quad 2x_2 \quad 2x_3]^T = [2 \quad 2 \quad 4]^T$

$\left\|\mathbf{c}^{(0)}\right\| > \varepsilon$, continue,

$\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} = [-2 \quad -2 \quad -4]^T$

minimizar

$f(\alpha) = f\left(\mathbf{x}^{(0)}\right) + \alpha\,\mathbf{d}^{(0)}$, onde $\mathbf{x}^{(0)} + \alpha\,\mathbf{d}^{(0)} = [1 - 2\alpha \quad 1 - 2\alpha \quad 2 - 4\alpha]^T$

$$f(\alpha) = f\left(\mathbf{x}^{(0)}\right) + \alpha\, \mathbf{d}^{(0)} = (1 - 2\alpha)^2 + (1 - 2\alpha)^2 + (2 - 4\alpha)^2$$

$$f(\alpha) = 24\alpha^2 - 24\alpha + 6$$

$$\frac{df}{d\alpha} = 0 \ \therefore\ \alpha_0 = 1/2$$

$$\frac{d^2 f}{d\alpha^2} > 0$$

$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_0\, \mathbf{d}^{(0)} = [0 \quad 0 \quad 0]^T$

$\mathbf{c}^{(1)} = [0 \quad 0 \quad 0]^T$, $\left\|\mathbf{c}^{(1)}\right\| < \varepsilon$, pare.

Resposta: $\mathbf{x}^* = [0 \quad 0 \quad 0]^T$, e o valor mínimo da função objetivo é zero nesse ponto.

Nesses dois exemplos foi possível determinar o mínimo da função $f(\alpha)$ pelas condições necessária e suficiente. Em casos mais complexos, pode ser necessário o uso de um método numérico de busca em linha, como o segmento áureo ou busca de Armijo.

5.4 Gradientes Conjugados

Esta é uma modificação simples e efetiva do Método da Maior Declividade. É possível provar que no método original as direções em 2 passos consecutivos são ortogonais entre si, o que diminui a velocidade de convergência. Nos Gradientes Conjugados, essas direções não mais são ortogonais. Pouco cálculo adicional é necessário, mas a convergência é muito mais rápida.

O algoritmo agora é:

1. Estimar um projeto inicial $\mathbf{x}^{(0)}$, fazer $k = 0$, adotar um parâmetro de convergência pequeno $\varepsilon > 0$,
   Calcular $\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} = -\nabla f(\mathbf{x}^0)$

Verifique o critério de parada. Se $\|\mathbf{c}^{(0)}\| < \varepsilon$, pare. Senão vá para o passo 5.

2. calcular o gradiente da função objetivo no ponto atual $\mathbf{x}^{(k)}$, $\mathbf{c}^{(k)} = \nabla f\left(\mathbf{x}^{(k)}\right)$,
3. calcular o comprimento do vetor gradiente, $\|\mathbf{c}^{(k)}\|$; se $\|\mathbf{c}^{(k)}\| < \varepsilon$, pare, senão, continue,
4. calcular a nova direção de busca neste passo como

   $\mathbf{d}^{(k)} = -\mathbf{c}^{(k)} + \beta_k \mathbf{d}^{(k-1)}$, onde $\beta_k = \left(\frac{\|\mathbf{c}^{(k)}\|}{\|\mathbf{c}^{(k-1)}\|}\right)^2$
5. calcular o tamanho do passo neste ponto, minimizando $f(\alpha) = f\left(\mathbf{x}^{(k)}\right) + \alpha\ \mathbf{d}^{(k)}$, pelo segmento áureo, ou outro método,
6. atualizar o projeto $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha\ \mathbf{d}^{(k)}$, fazer $k = k + 1$ e voltar ao passo 2.

Proposta: aplicar o método a

Minimizar $\quad f(\mathbf{x}) = x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_2x_3$

Estimativa inicial: $\mathbf{x}^T = [2 \quad 4 \quad 10]^T$

Ponto de ótimo: $\mathbf{x}^{*T} = [1.0E-07 \quad -1.7E-07 \quad 1.04E-09]^T$

Valor da função objetivo no ponto de ótimo: $f(\mathbf{x}^*) = -4.0E-14$

5.5 Busca unidimensional

5.5 Busca unidimensional pelo segmento áureo

Na busca do mínimo da função $f(\alpha)$ a partir de $\alpha = 0$, se procede em 2 fases, como a seguir.

Passo 1. Para um número pequeno δ, calculam-se vários valores $f\left(\alpha_q\right) = \sum_{j=0}^{q} \delta(\varphi)^j; \quad q = 0{,}1{,}2, \ldots$

Se

$f(\alpha_{q-1}) < f(\alpha_{q-2})$ e $(\alpha_{q-1}) < f(\alpha_q)$

o mínimo estará entre α_{q-2} e α_q . Os limites superior e inferior da fase seguinte serão

$\alpha_U = \alpha_q = \sum_{j=0}^{q} \delta(\varphi)^j$ e $\alpha_L = \alpha_{q-2} = \sum_{j=0}^{q-2} \delta(\varphi)^j$

Passo 2. Calcula-se $\alpha_b = \alpha_L + (1-\varphi)(\alpha_U - \alpha_L)$ e $f(\alpha_b)$, sendo que $\alpha_a = \alpha_L + 0.5(1-1/\varphi)(\alpha_U - \alpha_L) = \alpha_{q-1}$ tal que $f(\alpha_a)$ já é conhecido.

Passo 3.

Passo 3.1. Se $f(\alpha_a) < f(\alpha_b)$, o mínimo está entre α_L e α_b, e os novos limites inferior e superior são $\alpha'_L = \alpha_L$ e $\alpha'_U = \alpha_b$, e $\alpha'_b = \alpha_a$ Calcular $f(\alpha'_a)$, com $\alpha'_a = \alpha'_L + (1-1/\varphi)(\alpha'_U - \alpha'_L)$, e ir para o Passo 4.

Passo 3.2. Se $f(\alpha_a) > f(\alpha_b)$, o mínimo está entre α_a e α_U, e os novos limites inferior e superior são $\alpha'_L = \alpha_a$ e $\alpha'_U = \alpha_U$, e $\alpha'_a = \alpha_b$ Calcular $f(\alpha'_b)$, com $\alpha'_b = \alpha'_L + (1-\varphi)(\alpha'_U - \alpha'_L)$, e ir para o Passo 4.

Passo 3.3. Se $f(\alpha_a) = f(\alpha_b)$, $\alpha_L = \alpha_a$ e $\alpha_U = \alpha_b$, e voltar para o Passo 2.

Passo 4. Se o intervalo $(\alpha'_U - \alpha'_L)$ é menor que um certo critério de parada, o mínimo é estimado como $(\alpha'_U - \alpha'_L)/2$. Caso contrário, fazer $\alpha_L = \alpha'_L$, $\alpha_U = \alpha'_U$, $\alpha_a = \alpha'_a$ e voltar para o Passo 3.

5.5.2 Busca unidimensional de Armijo

Algoritmos de busca em linha exatos, como o Golden Rule, podem ser custosos em tempo computacional. Alternativamente se usam métodos inexatos.

Um dos mais populares é a busca unidimensional de Armijo. Usa-se uma função linear de α:

$q(\alpha) = f(\alpha) + \alpha[\rho f'(0)]$, onde $0 < \rho < 1$ (em geral $\rho = 0.2$) (5.17)

A seguir, adota-se um valor inicial arbitrário de α , tal que

$$f(\alpha) \leq q(\alpha) \tag{5.18}$$

e ele é repetidamente incrementado de η (geralmente $\eta = 2$) até que a Eq. (5.18) seja violada. O maior valor de α assim encontrado é selecionado como o tamanho do passo. Se o valor inicial já viola a Eq. (5.18), então ele será repetidamente dividido por η até respeitar essa inequação.

5.6 Método de Newton

No Método da Maior Declividade, se usam apenas derivadas de primeira ordem, contidas no vetor gradiente. Se se dispuser das derivadas de segunda ordem, contidas na Matriz Hessiana, uma busca mais eficiente de direção e melhor convergência é de se esperar.

Parte-se da expansão em série de Taylor truncada

$$f(\mathbf{x} + \Delta\mathbf{x}) = f(\mathbf{x}) + \mathbf{c}^T \Delta\mathbf{x} + \frac{1}{2!} \Delta\mathbf{x}^T \mathbf{H}(\mathbf{x})\, \Delta\mathbf{x} \tag{5.19}$$

uma função quadrática em $\Delta\mathbf{x}$, com um mínimo global, se a Hessiana é, no mínimo, positivo semi definida. A condição para esse mínimo é

$$\frac{\partial f}{\partial(\Delta\mathbf{x})} = \mathbf{0} = \mathbf{c} + \mathbf{H}\, \Delta\mathbf{x} \tag{5.20}$$

levando a

$$\Delta\mathbf{x} = -\mathbf{H}^{-1}\mathbf{c} \tag{5.21}$$

que é apenas formal, já que na prática se resolve o sistema $\mathbf{H}\, \Delta\mathbf{x} = -\mathbf{c}$, e o projeto é atualizado com

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \Delta\mathbf{x} \tag{5.22}$$

A seguir, o processo é repetido até atingir o mínimo da função objetivo. Em cada passo, a matriz Hessiana tem que ser recalculada, um esforço computacional considerável. Em compensação o método de Newton converge quadraticamente dentro de certo raio em torno da solução.

Uma alternativa menos custosa computacionalmente são os chamados métodos quasi Newton.

5.7 Problemas com Restrições

A ideia é transformar um problema com restrições numa sequência de problemas sem restrições. Para isso, se constrói uma função composta das funções objetivo e de restrições, e certos parâmetros de penalidade a serem aplicados quando se violam essas restrições. Quanto maior essas violações, maiores as penalizações. Quando a função composta é definida para um conjunto de parâmetros, é minimizada como um problema sem restrições. Os parâmetros são novamente justados e a função composta redefinida e novamente minimizada. O processo continua até que não se tenha ganho significativo na estimativa do ponto de ótimo.

Um dos mais utilizados desses métodos é o Lagrangiano Aumentado, um exemplo dos métodos de multiplicadores.

O problema original é

Minimizar $f(\mathbf{x})$ sujeita a

Restrições de igualdade $\quad g_i(\mathbf{x}) = 0; \quad i = 1, l$

Restrições de desigualdade $\quad g_i(\mathbf{x}) \leq 0; \quad i = l + 1, m$

O Lagrangiano é definido como

$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x})$

em que $\mathbf{u}$ é o vetor de multiplicadores de Lagrange, um multiplicador u_i para cada equação de restrição $g_i(\mathbf{x})$.

A seguir, é construído o Lagrangiano aumentado

$$\Phi(\mathbf{x}, \mathbf{u}, \mathbf{r}) = f(\mathbf{x}) + P(\mathbf{g}(\mathbf{x}), \mathbf{u}, \mathbf{r}) \qquad (5.23)$$

onde $P(\mathbf{g}(\mathbf{x}), \mathbf{u}, \mathbf{r})$ é uma função de penalidade e $\mathbf{r}$ é um vetor contendo os parâmetros de penalidade relacionados a cada equação de restrição. A seguir, o Lagrangiano aumentado é minimizado. Se os critérios de convergência não forem satisfeitos, os vetores de multiplicadores de Lagrange e de penalidades são atualizados para um próximo passo de otimização. É nessa atualização que estão os problemas do método e não são abordados neste curso.