IA881 – Otimização Linear

Aula: Conceitos básicos de otimização

Ricardo C. L. F. Oliveira & Takaki Ohishi

Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas

1^o Semestre 2017

Tópicos

Introdução

- Otimização Linear $\rightarrow$ caso particular da Otimização Não-Linear.
- Otimização Não-Linear:
  - Condições de otimalidade.
  - Métodos de otimização.
- Tipo de restrições (igualdade, desigualdade ou irrestrito) $\rightarrow$ define classes de problemas e métodos de otimização.
- Esta apresentação:
  - Irrestritos.
  - Igualdades.
  - Desigualdades.
- Má notícia: em otimização não-linear só é possível garantir otimalidade local. Otimalidade global apenas sob condições específicas (convexidade).

Princípios de otimalidade

Considere o seguinte problema de otimização

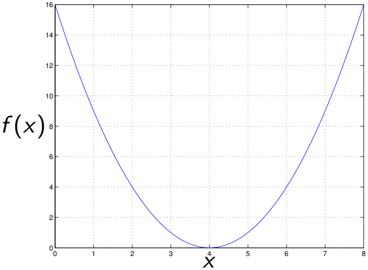
$$\begin{array}{ll} \text{minimizar} & f(x) \\ x \in \mathbb{R} & \end{array}$$

Um ponto $x^\star$ é um mínimo local do problema se a seguinte condição é satisfeita

$$f(x^\star) \leq f(x), \quad \forall x \text{ tal que } |x - x^\star| < \varepsilon$$

Exemplo

- Exemplo 1: Minimizar $f(x) = (x-4)^2$



Condições Necessárias de Otimalidade

$$\frac{df(x^0)}{dx} = 0$$

- Diz-se que x^0 é um ponto de estacionaridade.

Princípios de Otimalidade

- Seja $f : \mathbb{R} \to \mathbb{R}$. Em torno de um dado ponto x^0, pode-se aproximar a função por meio de uma expansão de Taylor dada por:

$$f(x) = f(x^0) + f'(x^0)(x - x^0) + o(\cdot) \tag{1}$$

sendo que $o(\cdot)$ representa os termos de ordem maior do que um. Este termo vai a zero rapidamente à medida que o ponto x se aproxima do ponto x^0. Neste contexto, a equação (1) resulta em

$$f(x) - f(x^0) = f'(x^0)(x - x^0)$$

- Portanto, se x^0 é um ponto de mínimo, então a derivada primeira neste ponto deve ser nula (se não fosse, por exemplo, escolha $x - x^0 = -f'(x^0)$, implicando que $f(x) - f(x^0) < 0$, o que contradiz o fato de x^0 ser mínimo local).

Princípios de Otimalidade

- Se a derivada primeira é nula, então pode-se aproximar a função por meio de:

$$f(x) = f(x^0) + \frac{1}{2} f''(x^0)(x - x^0)^2 + o(\cdot) \quad (2)$$

Quando o ponto x se aproxima do ponto x^0, então a equação (2) resulta em

$$f(x) - f(x^0) = f''(x^0)(x - x^0)^2 \quad (3)$$

Portanto, se x^0 é um ponto de mínimo, então a condição necessária de ótimo requer que a derivada segunda neste ponto deve ser não-negativa. (se x^0 é um ponto de máximo, então a derivada segunda neste ponto deve ser não-positiva).

Condições de Otimalidade

Teorema 1 (condições necessárias para ótimo local)

Se $x^\star$ é um mínimo local, então as seguintes condições devem ser satisfeitas:

- $f'(x^\star) = 0$
- $f''(x^\star) \geq 0$

Agora, se a derivada segunda é positiva, então de (3) tem-se que esta é condição suficiente para mínimo local.

Teorema 2 (condições suficientes para ótimo local)

Se $x^\star$ satisfaz as condições a seguir, então estas condições são suficientes para assegurar um ótimo local:

- $f'(x^\star) = 0$
- $f''(x^\star) > 0$

Princípios de Otimalidade

- Seja $f : \mathbb{R}^n \to \mathbb{R}$. Em torno de um dado ponto x^0, pode-se aproximar a função por meio de uma expansão de Taylor dada por:

$$f(x) = f(x^0) + \nabla f(x^0)^T (x - x^0) + o(\cdot) \tag{4}$$

em que $\nabla f(x^0)$ é o gradiente da função no ponto x^0. Considerando que $o(\cdot)$ vai a zero rapidamente à medida que x se aproxima de x^0, tem-se que a equação (4) resulta em

$$f(x) - f(x^0) = \nabla f(x^0)^T (x - x^0)$$

- Portanto, para que x^0 seja um mínimo local deve-se ter o gradiente igual a zero.

Princípios de Otimalidade

- Se o ponto $x^\star$ é um ponto estacionário, então a função $f(x)$ pode ser aproximada por:

$$f(x) = f(x^\star) + \frac{1}{2}(x - x^\star)^T F(x^\star)(x - x^0)$$

sendo que F é a matriz Hessiana. Esta última equação pode ser reescrita na forma

$$f(x) - f(x^\star) = \frac{1}{2}(x - x^\star)^T F(x^\star)(x - x^0) \tag{5}$$

Assim, para que o ponto $x^\star$ seja um mínimo local, a matriz $F(x^\star)$ deve ser semi-definida positiva.

Condições de Otimalidade

Teorema 3 (condições necessárias para ótimo local)

Se $x^\star$ é um mínimo local, então as seguintes condições devem ser satisfeitas:

- $\nabla f(x^\star) = 0$
- $F(x^\star) \geq 0$

Agora, se a Hessiana é positiva, então de (5) tem-se que esta é condição suficiente para mínimo local.

Teorema 4 (condições suficientes para ótimo local)

Se $x^\star$ satisfaz as condições a seguir, então estas condições são suficientes para assegurar um ótimo local:

- $\nabla f(x^\star) = 0$
- $F(x^\star) > 0$

Positividade de matrizes (simétricas)

- Uma matriz simétrica P é definida positiva se e somente se $x^T Px > 0$, $\forall\ x \neq 0$, $x \in \mathbb{R}^n$.

- Uma matriz simétrica P é semi-definida positiva se e somente se $x^T Px \geq 0$, $\forall x \in \mathbb{R}^n$.

- Uma matriz simétrica P é (semi)definida negativa se $-P$ é (semi)definida positiva.

- Note que uma condição necessária para que uma matriz seja definida positiva é que todos os elementos da diagonal sejam positivos.

Positividade de matrizes (simétricas)

Teorema 5

Uma matriz simétrica $P \in \mathbb{R}^{n \times n}$ é (semi)definida positiva se e somente se qualquer uma das condições for verificada.

- *Todos os autovalores são positivos (não negativos);*
- *Todos os menores principais líderes de P são positivos (todos os menores principais de P são não negativos);*
- *Existe uma matriz $N \in \mathbb{R}^{n \times n}$ não singular (uma matriz singular $N \in \mathbb{R}^{n \times n}$ ou uma matriz $N \in \mathbb{R}^{m \times n}$ com $m < n$) tal que $P = N^T N$.*

Observação: Para $P = P^T$ definida positiva, autovalores, traço, determinante e menores principais são positivos

Notação: $P > 0$ (≥ 0) indica que P é (semi)definida positiva.
$P \geq Q \Rightarrow P - Q \geq 0$.

Menores Principais

- Seja a matriz $M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$

- Menores Principais (determinantes de todas as submatrizes de M cujas diagonais coincidem ou fazem parte da diagonal de M):

$$m_{11} \quad , \quad m_{22} \quad , \quad m_{33}$$

$$\det\left(\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}\right), \ \det\left(\begin{bmatrix} m_{11} & m_{13} \\ m_{31} & m_{33} \end{bmatrix}\right), \ \det\left(\begin{bmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{bmatrix}\right), \ \det(M)$$

- Menores Principais Líderes (determinantes das submatrizes de M obtidas ao eliminarem-se as últimas k colunas e k linhas, $k = 2, 1, 0$):

$$m_{11} \quad , \quad \det\left(\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}\right) \quad , \quad \det(M)$$

- Observação: todos os menores principais líderes positivos $\implies$ menores principais positivos. No entanto, todos os menores principais líderes não negativos não implica que todos os menores são não negativos. Por exemplo,

Menores Principais – Exemplos

$$M = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{bmatrix} \quad \begin{array}{l} \Rightarrow \ m_{11} = 0 \ , \ \det \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0 \ , \ \det(M) = 0 \\ \text{mas} \ \Rightarrow \ m_{22} = 0, \ m_{33} = 2, \ \det \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} = -1 \ ; \ \det \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} = 0 \end{array}$$

- Exemplo 2:

$M = \begin{bmatrix} 5 & -1 \\ -1 & 8 \end{bmatrix}$ é definida positiva, pois $m_{11} = 5 > 0$; $\det(M) = 39 > 0$; os autovalores de M são $\lambda_1 = 4.697$ e $\lambda_2 = 8.303$ (traço$(M) = 13$) e

$$M = N'N \quad ; \quad N = \begin{bmatrix} 2.2273 & -0.1981 \\ -0.1981 & 2.8215 \end{bmatrix} \quad ; \quad \det(N) = 6.2450 \neq 0$$

Métodos de otimização

- Na seção anterior foram apresentadas as condições que uma solução ótima deve atender.
- Nesta seção são apresentados alguns métodos para a obtenção de uma solução que atenda às condições de otimalidade.
- No caso de otimização irrestrita há muitos métodos, e aqui são apresentados o método direto; o método de gradiente e o método de Newton.
- Os métodos de gradiente e de Newton são processos iterativos, que partem de um ponto inicial, e iterativamente são calculadas novas soluções melhores que as anteriores, até chegar à otimalidade.

Processo iterativo

- O processo iterativo pode ser representado por:

$$x^{k+1} = x^k + \alpha_k d^k$$

com x^0 dado.

- Esta equação descreve a dinâmica dos métodos, nos quais uma nova solução, x^{k+1}, é obtida a partir da última solução, x^k. Esta transição depende de uma direção de caminhada, d^k, e também do tamanho passo dado nesta direção, α_k.
- Em geral, a principal diferença entre os métodos é a forma de cálculo da direção de caminhada.
- Neste contexto, dois conceitos são importantes: as direções de melhorias e a busca unidimensional. Esta última refere-se ao cálculo do tamanho do passo.

Métodos diretos

- Os métodos diretos resolvem diretamente as condições de otimalidade. Isto é, determinam as soluções que atendem estas condições.

- Basicamente, estes métodos determinam os pontos de estacionaridade:

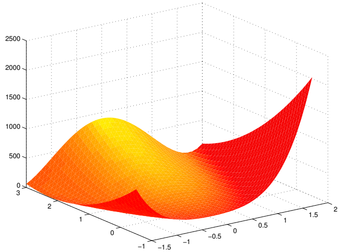
$$\nabla f(x) = 0$$

e, para cada ponto de estacionaridade é analisado o sinal da matriz hessiana para determinar a natureza de cada um destes pontos.

Exemplo

- Exemplo: Função de Rosenbrock

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



com gradiente dado por

$$\nabla f(x) = \begin{bmatrix} -400(x_2 - x_1)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

Exemplo: Função de Rosenbrock

Ponto de estacionaridade: $\nabla f(x) = 0 \Rightarrow x^\star(1,1)$. A Hessiana é dada por

$$F(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

No ponto estacionário temos

$$F(x^\star) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \Rightarrow m_{11} = 802 > 0, \quad \det(F(x^\star)) = 400 > 0$$

Portanto, como $F(x^\star) > 0$, então é um ponto de mínimo.

Direção de melhoria

- Seja o problema:

$$\begin{array}{ll} \text{maximizar} & f(x) \\ x \in \mathbb{R} & \end{array}$$

- Os métodos iterativos partem de uma solução inicial e definem uma direção de caminhada.

- Quando uma dada direção de caminhada proporciona uma nova solução com melhor valor na função objetivo, diz-se que esta é uma direção de melhoria.

- Em particular, em um problema de minimização, uma direção de melhoria é uma direção que minimiza a função objetivo. Em um problema de maximização seria uma direção de maximização da função objetivo.

Direção de melhoria

- Supondo que um método calculou uma nova solução por:

$$x^{k+1} = x^k + \alpha_k d^k \tag{6}$$

e ainda que o novo ponto apresenta um valor na função objetivo maior que o ponto anterior. Então, voltando à expansão de Taylor de primeira ordem, tem-se:

$$f(x^{k+1}) - f(x^k) = \nabla f(x^k)^T (x^{k+1} - x^k)$$

Assim, supondo a melhoria, tem-se

$$f(x^{k+1}) - f(x^k) > 0 \Rightarrow \nabla f(x^k)^T (x^{k+1} - x^k) > 0 \tag{7}$$

Por outro lado, de (6) tem-se que

$$x^{k+1} - x^k = \alpha_k d^k$$

que substituindo em (7) fornece

$$\nabla f(x^k)^T (x^{k+1} - x^k) > 0 \Rightarrow \nabla f(x^k)^T d^k > 0$$

Desta última equação conclui-se que para uma direção ser de melhoria em um problema de otimização, esta deve ter um ângulo menor que 90^o com o vetor gradiente. Além disso, a direção de máximo crescimento é:

$$d^k = \nabla f(x^k)$$

Busca Unidimensional

- Uma vez calculada uma direção de melhoria, deve-se agora determinar o tamanho do passo nesta direção.

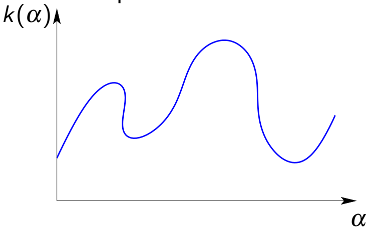
- O processo de cálculo desse passo é conhecido como busca unidimensional.

Função unidimensional

Seja uma função $f : \mathbb{R}^n \to \mathbb{R}$. Quando a partir de um dado ponto desloca-se numa dada direção, então a função f neste corte é função somente do tamanho do passo α:

$$x^{k+1} = x^k + \alpha_k d^k, \qquad f(x^{k+1}) = f(x^k + \alpha d^k) = k(\alpha)$$

O gráfico a seguir mostra a função unidimensional (monovariável), e a busca unidimensional visa determinar o ponto de máximo desta função.



Método de Busca Através da Função Unidimensional

- Há vários métodos de busca unidimensional. Aqui serão apresentados dois métodos: O método de busca através da função unidimensional e o método da falsa posição.

Método de Busca Através da Função Unidimensional

- Este método calcula explicitamente a função unidimensional e após otimiza esta função. Exemplo, seja o problema:

$$\begin{array}{ll} \text{maximizar} & f(x) = -x_1^2 - 2x_2^2 + 3x_1 + x_2 \\ x \in \mathbb{R} & \end{array}$$

O ponto de partida é o ponto $x^0 = (0,0)^T$. A direção de caminhada (dada pelo gradiente) é dada por

$$d^1 = \nabla f(x^0) \begin{bmatrix} -2x_1 + 3 \\ -2x_2 + 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Método de Busca Através da Função Unidimensional

A caminhada nesta direção a partir do ponto x^0 é dada por:

$$x^1 = x^0 + \alpha_1 d^1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3\alpha \\ \alpha \end{bmatrix} \tag{8}$$

Substituindo (8) na função objetivo, tem-se

$$f(x) = -(3\alpha)^2 - 2\alpha^2 + 3(3\alpha) + \alpha \Rightarrow k(\alpha) = -11\alpha^2 + 10\alpha$$

Agora deve-se determinar o tamanho do passo para maximizar a $k(\alpha)$

$$\frac{\partial k(\alpha)}{\partial \alpha} = -22\alpha + 10 = 0 \Rightarrow \alpha = \frac{10}{22}$$

Assim, o novo ponto é dado por

$$x^1 = x^0 + \alpha_1 d^1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 30/22 \\ 10/22 \end{bmatrix}$$

Método do gradiente

- O método do gradiente é o método iterativo mais simples de otimização não-linear.

- Este método parte de uma solução inicial e realiza buscas unidimensionais sempre na direção do vetor gradiente até atingir um ponto de estacionaridade.

- O método do gradiente pode ser resumido nos seguintes passos:

Forneça: x^0 e $\varepsilon > 0$;

$k = 0$
enquanto $||\nabla f(x^k)|| > \varepsilon$ **faça**
  $d^{k+1} = \nabla f(x^k)$
  Fazer uma busca em d^{k+1}
  $x^{k+1} = x^k + \alpha_{k+1} d^{k+1}$
  $k = k + 1$
fim enquanto

Método de Newton

- O método de Newton é muito mais eficiente do que o método do gradiente, e também menos dependente das condições da função objetivo.

- A desvantagem deste método é a necessidade do cálculo da matriz hessiana e do cálculo de sua inversa.

- Suponha um ponto x^k e uma aproximação de segunda ordem do tipo:

$$f(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T F(x^k)(x - x^k)$$

Método de Newton

- Supondo a aproximação quadrática, então o ponto de estacionaridade desta aproximação é dado por:

$$\frac{\partial f(x)}{\partial x} = \nabla f(x^k) + F(x^k)(x - x^k) = 0 \Rightarrow x^{k+1} = x^k - F(x^k)^{-1} \nabla f(x^k)$$

- Uma iteração do método de Newton determina o ponto de estacionaridade da função quadrática aproximada.

- Então, se a função é quadrática, o método de Newton determina a solução ótima em uma única iteração.

Restrições de igualdade

- Considere o problema de otimização

$$\begin{array}{ll} \text{minimizar} & f(x) = (x-4)^2 \\ s.a & x = 6 \end{array}$$

Seja a função Lagrangeana

$$L(x,\lambda) = (x-4)^2 + \lambda(x-6)$$

sendo λ o multiplicador de Lagrange (variável dual).

- Condições de Otimalidade: dada pela estacionaridade da função Lagrangeana.

$$\frac{\partial f(x)}{\partial x} = 0 \Rightarrow 2x - 8 + \lambda = 0 \Rightarrow \lambda = -4.$$

$$\frac{\partial f(x)}{\partial \lambda} = 0 \Rightarrow x - 6 = 0 \Rightarrow x = 6.$$

Restrições de desigualdade

- Considere o problema de otimização com restrição de desigualdade

$$\begin{array}{ll} \text{minimizar} & f(x) = (x-4)^2 \\ s.a. & x \leq 3 \end{array}$$

Resolvendo problema de modo similar ao caso com restrições de igualdade, tem-se a seguinte solução:

$$x^\star = 3, \qquad \lambda^\star = 2$$

A restrição de desigualdade dualizada foi satisfeita na igualdade.

Restrições de desigualdade

- Considere agora o problema de otimização

$$\begin{array}{ll} \text{minimizar} & f(x) = (x-4)^2 \\ s.a. & x \leq 6 \end{array}$$

Resolvendo o problema de modo similar ao caso com restrições de igualdade, tem-se a seguinte solução:

$$x^\star = 6, \qquad \lambda^\star = -4$$

Note que os valores acima não são a solução do problema. O indicativo que esta não é a solução do problema está no sinal negativo do multiplicador de Lagrange λ. Ou seja, para problemas com restrições de desigualdade, os multiplicadores de Lagrange devem apresentar valores não negativos na solução ótima. Caso contrário, a restrição com o multiplicador negativo deve ser retirado da função Lagrangeana. Assim, a função Lagrangeana fica:

$$L(x) = (x-4)^2$$

Cuja condição de estacionaridade é:

$$\frac{\partial L(x)}{\partial x} = 0 \Rightarrow 2x - 8 = 0 \Rightarrow x^\star = 4$$

A solução ótima coincide com a solução irrestrita.

Otimização com Restrições de Desigualdade

- Seja o problema de otimização

$$\begin{array}{ll} \text{minimizar} & f(x) \\ g(x) \leq 0 & \end{array}$$

- Condições de Otimalidade: Condições de Karush-Kuhn-Tucker (KKT)

$$\begin{aligned} \nabla_x L(x^\star, \lambda^\star) &= 0 \\ \nabla_\lambda L(x^\star, \lambda^\star) &= 0 \\ \lambda^\star &\geq 0 \\ \lambda^\star g(x^\star) &= 0 \end{aligned}$$

Estratégias de solução

- Uma estratégia para resolução de problemas com restrições de desigualdade:

  (1) Resolve-se o problema de otimização com um sub-conjunto das restrições de desigualdade, estas tratadas como restrições de igualdade.

  (2) Se todos os multiplicadores de Lagrange associados a restrições $\leq$ forem negativos e se todos os multiplicados de Lagrange associados a restrições $\geq$ forem positivos, então a solução obtida em (1) é a solução ótima do problema. Caso contrário, as restrições de desigualdade com os respectivos multiplicadores de Lagrange com sinais "errados" devem ser relaxados (não serão considerados restrições de igualdade em (1)), e novamente repete-se o processo.