Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Departamento de Física
Laboratório de Teoria da Matéria Condensada

Introdução à teoria de otimização

Tiago de Souza Farias

23 de novembro de 2012

Conceitos de otimização

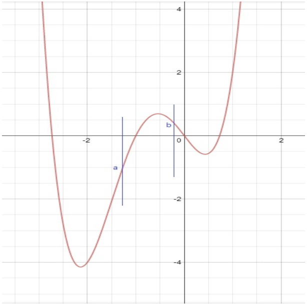
- Otimizar é buscar valores ótimos de uma ou mais funções (função objetivo);
- Em geral, valores ótimos correspondem a máximos, mínimos, melhores caminhos ou maior eficiência;
- Chame-se ponto crítico um possível valor ótimo;
- Uma função terá máximo global se para um ponto crítico c se $f(c) \geq f(x)$ para todo o domínio da função;
- Uma função terá máximo local se $f(c) \geq f(x)$ para todo x próximo de c;

Conceitos de otimização

- Otimização irrestrita é a busca do valor ótimo para todo o domínio de uma função;
- Otimização restrita é a busca do valor ótimo para uma função restrita a certos parâmetros;
- Cada conjunto de problemas a ser otimizado está associado a uma técnica ou método diferente;
- A formulação de técnicas de otimização é baseada em princípios matemáticos como teoremas;

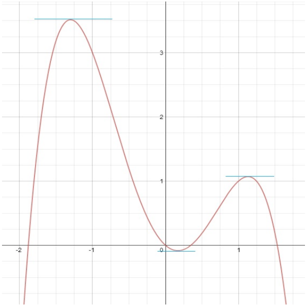
Teorema do valor extremo

- Se uma função f for contínua em um intervalo fechado [a,b], então f possue um valor máximo global e mínimo global entre [a,b];



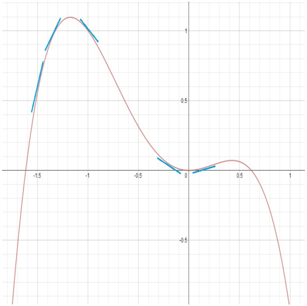
Teorema de Fermat

- Se f tiver um valor extremo em c, e se df(c) existir, então df(c) = 0 ou df(c) não existe;



Teste da derivada primeira

- A derivada primeira avalia o crescimento de uma função;
- Se o crescimento da derivada primeira mudar de sinal em c, então c é um ponto crítico;



Teste da derivada segunda

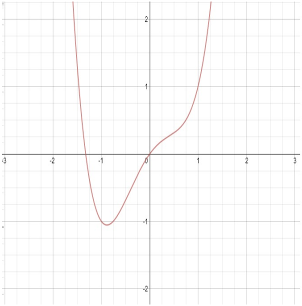
- A derivada segunda avalia concavidades de uma função;
- Se o crescimento da derivada segunda mudar de sinal em c, então c é um ponto crítico;
- Portanto:

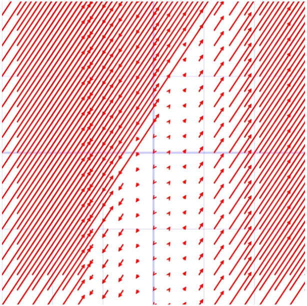
  c será um mínimo local se df(c) $= 0$ e d^2f(c) > 0;

  c será um máximo local se df(c) $= 0$ e d^2(c) < 0;

Gradiente

- Gradiente é um vetor representado por ∇ que indica as direções de crescimento de uma função, define-se: $\nabla f(x_1, x_2, ..., x_n) = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} + ... + \frac{\partial f}{\partial x_n}$





Derivada direcional

- Define-se derivada direcional como a taxa a qual uma função aumenta ou diminui quando f varia de um ponto a em direção a um ponto u;
- Para uma certa função $f(x_1, x_2, ..., x_n)$ e seja o ponto $a(a_1, a_2, ..., a_n)$, a derivada direcional de f na direção de u é: $(\nabla f(a))(u) = \sum_{i=1}^{N} \frac{\partial f(a)}{\partial x_i} u_i$
- Por regras de produto escalar: $(\nabla f(a))(u) = \parallel \nabla f(a) \parallel \ \parallel u \parallel cos\theta$

Matriz Hessiana

- Para uma função de várias variáveis $f(x_1, x_2, ..., x_n)$, c será um ponto crítico se todas as derivadas parciais de f forem nulas;
- O ponto crítico de f é calculado por: $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \ldots = \frac{\partial f}{\partial x_n} = 0$
- Define-se matriz Hessiana a matriz quadrada derivada segunda de f que descreve a curvatura da função;

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Matriz Hessiana

- A relação entre a matriz Hessiana e a matriz Jacobiana é: $H(f(x)) = J(\nabla f(x))$
- Uma matriz simétrica A é definida positiva se seus autovalores são reais e positivos ou se todas as suas submatrizes possuirem determinante positivo;
- Uma matriz simétrica A é definida negativa se seus autovalores são reais e negativos ou se suas submatrizes, não necessariamente todas, possuirem determinante negativo;

Matriz Hessiana

- Para um ponto crítico c:

  [•]Se a matriz Hessiana aplicada no ponto c for simétrica e definida negativa, então c é um máximo local;

  [•]Se a matriz Hessiana aplicada no ponto c for simétrica e definida positiva, então c é um mínimo local;

  [•]Se a matriz Hessiana aplicada no ponto c for indefinida, então c é um ponto de cela;

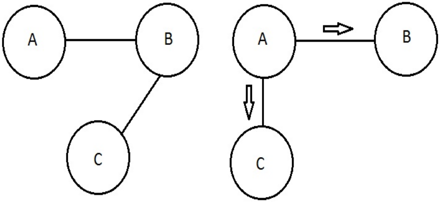
Teorema do envelope

- Os valores ótimos de uma função constituem uma outra função chamada função parâmetro. Se os valores ótimos forem aplicados na função objetivo, a função objetivo se tornará indiretamente a função parâmetro.

Grafos

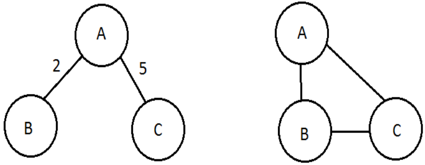
- Define-se grafo como uma representação gráfica de vértices ligados por arestas;
- Propriedades:

    [•] Grafos direcionados e não direcionados;



Grafos

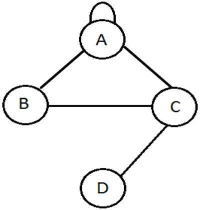
[•] Com peso e sem peso;



[•] Duas arestas serão adjacentes se possuirem um vértice em comum;

Grafos

- Sendo o grafo G, seus vértices v, um grafo pode ser representado como uma matriz A, cuja ordem é o número de vértices e a_{ij} terá o peso dos vértices se v_i e v_j forem adjacentes e 0 caso contrário;



$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Otimização exata

- Otimização exata usa diretamente teoremas de otimização para buscar valores ótimos exatas através de $Ax = B$;
- Conceitualizado por Fermat e Lagrange;
- Técnicas limitadas a encontrar máximos e mínimos e restritas a funções simples;
- Exemplo: técnica do gradiente;

Otimização iterativa

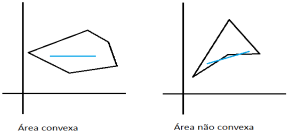
- Conceitualizado por Newton e Gauss;
- Aplicável em uma grande variedade de funções;
- Em alguns casos o custo computacional pode se tornar excessivamente alto;
- Exemplo: Método de Newton-Raphson;

Otimização geométrica

- Ultiliza recursos do cálculo, geometria e análise de imagens;
- Muito eficiente, mas bastante limitado;
- Exemplo: Método de granulação;

Otimização convexa

- Limitada a funções convexas e restritas;
- Função convexas são funções que possuem áreas convexas;
- Se um segmento de reta qualquer ligando dois pontos A e B estiver inteiramente contido em uma área delimitada pelo domínio de uma função, então esta área é convexa.

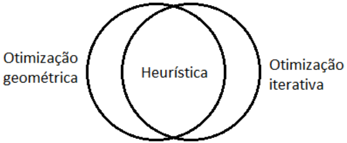


- Exemplo: Programação linear;

Otimização combinatória

- Conceitualizado por George B. Dantzin em 1947;
- Otimização combinatória consiste em procurar pontos críticos através dos vértices de uma área qualquer (espaço de pesquisa);
- Alto custo computacional, direcionado a funções de muitas variáveis;
- Exemplo: Método simplex, busca tabu, Simulated Annealing;

Heurística

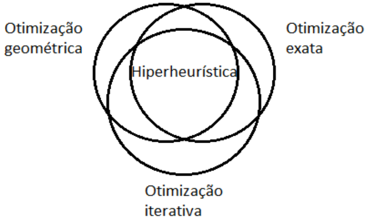


- Simula fenômenos da natureza para obter o valor ótimo;
- Em geral, aumento na complexidade programacional implica em aumento de eficiência e redução de custo computacional;
- Baseado em paradigmas como:

  [•] Paradigma do guloso: a cada iteração o algoritmo escolhe o valor mais "apetitoso";

  [•] Paradigma do míope: a cada iteração o algoritmo escolhe o valor com melhor visualização;
- Ultiliza amplamente grafos;
- Exemplo: Sistemas fórmicos;

Hiperheurística



- Os algorítmos são capazes de aprender e se adaptar;
- Formado a partir de conceitos de inteligência artificial e biologia;
- Muito complexo, porém muito eficiente;
- Exemplo: Otimização memética, redes neurais e algoritmos genéticos;

Bibliografia

STEWART, James. Antonio Carlos Moretti. Cálculo, volume 1. São Paulo: Cengage Learning, 2012.

STEWART, James. Antonio Carlos Moretti. Cálculo, volume 2. São Paulo: Cengage Learning, 2012.

BOLDRINI, José Luiz. Álgebra linear. São Paulo: Harba, 1980.

CHONG, Edwin K. P. An introduction to optimization. USA: Wiley, 2001.