

10. INTRODUÇÃO À OTIMIZAÇÃO

10.1 *Formulação de Problemas*

Os componentes chaves da formulação de um problema de otimização são:

- 1) A função objetivo
- 2) O modelo do processo
- 3) As restrições

A função objetivo representa lucro, custo, energia, produção, distância, etc., em termos das variáveis de decisão do processo ou sistema em análise. O modelo do processo e as restrições descrevem as inter-relações entre estas variáveis.

10.1.1 *Aplicações*

As técnicas de otimização, que buscam a melhor solução para um problema (máximos ou mínimos de grandezas mensuráveis em seus domínios de definição), fazem-se necessárias em muitas áreas da engenharia, tais como:

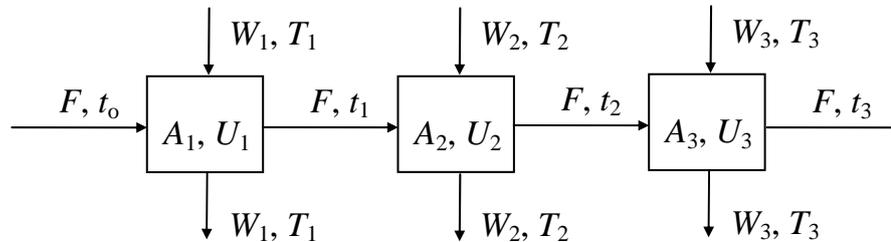
- pesquisa operacional: otimização de sistemas técnico-econômicos, controle de estoques, planejamento de produção, etc.;
- engenharia de processos: dimensionamento e otimização de processos, integração mássica e energética de processos, estimação de parâmetros, reconciliação de dados, análise de flexibilidade, etc.;
- controle de processos: identificação de sistemas, controle ótimo, controle adaptativo, controle preditivo, estimadores de estados, etc.;
- análise numérica: aproximações, regressão, solução de sistemas lineares e não-lineares, etc.

A otimização pode ser efetuada em diferentes níveis dentro de uma empresa, desde uma combinação complexa de plantas e unidades de suprimentos, passando por plantas individuais, combinações de unidades, equipamentos, peças de equipamentos, até entidades menores. Portanto, o escopo de um problema de otimização pode ser toda a empresa, uma planta, um processo, um equipamento, uma peça de um equipamento, ou qualquer sistema intermediário entre estes. Em uma indústria típica existem três níveis que podem ser otimizados: nível gerencial, nível de projeto de processo e especificação de equipamento e nível de operação de plantas.

Existe uma série de atributos de processos que afetam os custos ou lucros, tornando-os atrativos para aplicação de otimização, tais como:

- vendas limitadas pela produção → aumento de capacidade produtiva
- vendas limitadas pelo mercado → aumento de eficiência ou produtividade
- grandes unidades de processamento → redução dos custos de produção
- elevados consumos de matéria-prima ou energia → redução do consumo
- qualidade do produto superior às especificações → operar perto das restrições
- perda de componentes valiosos para os efluentes → ajustes e re-aproveitamentos
- alto custo de mão-de-obra → automação e replanejamento das unidades.

Exemplo 1: No projeto de um sistema de resfriamento representado abaixo:



deseja-se dimensionar os trocadores de calor de modo a resfriar a corrente a temperatura t_0 para t_3 com o menor custo possível. Uma análise econômica levou a seguinte expressão do custo do sistema:

$$\text{custo por unidade: } c_i = a_i \sqrt{A_i} + b_i W_i$$

$$\text{custo total: } c_T = \sum_{i=1}^3 c_i$$

onde o coeficiente a_i é função do tipo de trocador de calor e do fluido refrigerante e b_i é o custo unitário do refrigerante. Uma análise técnica (balanço de energia) gerou as seguintes relações restritivas:

$$A_i = \frac{F C_p}{U_i} \ln \left(\frac{t_{i-1} - T_i}{t_i - T_i} \right) \quad \text{e} \quad W_i = \frac{F C_p}{\lambda_i} (t_{i-1} - t_i)$$

onde C_p é o calor específico da corrente, U_i é o coeficiente global de troca térmica e λ_i é o calor de vaporização do refrigerante. Portanto, dados os tipos de trocadores de calor e fluidos refrigerantes, e conhecidas as temperaturas T_i e as propriedades C_p , λ_i e U_i , o problema de levar a corrente (F, t_0) para a condição (F, t_3) da maneira mais econômica possível, consiste em minimizar c_T no projeto dos trocadores de calor.

10.1.2 Nomenclatura

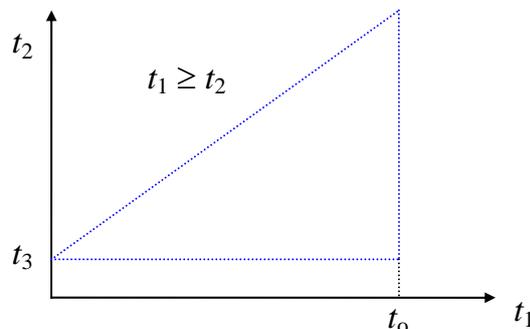
No contexto de otimização os problemas são tratados dentro das seguintes definições:

função objetivo: é a função matemática cujo máximo ou mínimo deseja-se determinar. No exemplo acima a função objetivo é o custo total.

variáveis de decisão: são as variáveis independentes que aparecem na função objetivo. Correspondem, em número, ao excesso de variáveis em relação ao número de equações (restrições de igualdade), isto é, o **grau de liberdade** do sistema. No exemplo acima tem-se 6 equações (3 para A_i e 3 para W_i) e 8 variáveis (t_1 , t_2 , A_i e W_i), portanto 2 variáveis de decisão ou grau de liberdade igual a 2.

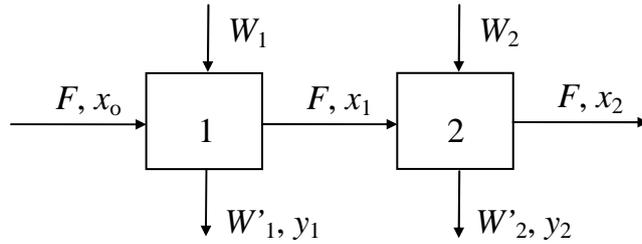
restrições: são os limites impostos ao sistema ou estabelecidos pelas leis naturais que governam o comportamento do sistema, a que estão sujeitas as variáveis de decisão. As restrições podem ser de igualdade (equações) ou de desigualdade (inequações). No exemplo acima tem-se 6 restrições de igualdade e 3 restrições de desigualdade: $t_3 \leq t_2 \leq t_1 \leq t_0$.

região de busca: ou região viável, é a região do espaço definido pelas variáveis de decisão, delimitada pelas restrições, em cujo interior ou em cuja fronteira se localiza o ótimo da função objetivo. No exemplo acima, tomando t_1 e t_2 como variáveis de decisão, a região de busca é aquela delimitada pelas restrições de desigualdade.



No exemplo acima, o problema de otimização foi gerado pela busca de um projeto ótimo do processo. Outra aplicação típica de técnicas de otimização aparece na melhoria das condições de operação de um processo em produção.

Exemplo 2: No processo de extração por solvente puro, ilustrado abaixo, deseja-se encontrar a condição de operação com a maior lucratividade possível.



onde W_i e W'_i são vazões mássicas de solvente, F é a vazão mássica de água, x_i é a massa de soluto por unidade de massa de água e y_i é a massa de soluto por unidade de massa de solvente. Uma análise econômica levou a seguinte expressão do lucro do sistema:

$$\text{lucro: } L = R - C$$

$$\text{receita: } R = P_s (W'_1 y_1 + W'_2 y_2)$$

$$\text{custo: } C = P_x (W_1 + W_2)$$

$$\text{restrição: } R > C$$

onde P_s é o preço do soluto no extrato, P_x é o preço do solvente puro. Uma análise técnica gerou as seguintes relações restritivas:

$$\text{balanços de massa para o soluto: } F x_0 - W'_1 y_1 - F x_1 = 0$$

$$F x_1 - W'_2 y_2 - F x_2 = 0$$

$$\text{balanços de massa para o solvente: } W_1 - W'_1 - s F = 0$$

$$W_2 + s F - W'_2 - s F = 0$$

$$\text{relações de equilíbrio: } y_1 = m x_1$$

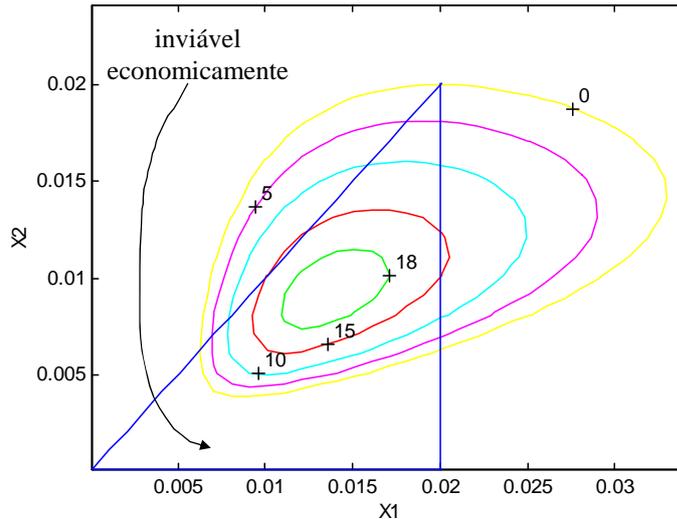
$$y_2 = m x_2$$

onde s é a solubilidade do solvente em água (massa de solvente / massa de água) e m é a constante de equilíbrio entre as fases. Portanto, dados F , x_0 , s , m , P_s e P_x , o problema de extrair o soluto da água da maneira mais lucrativa possível, consiste em maximizar L em função das condições de operação. Este exemplo possui também 6 equações (4 equações de balanço de massa e 2 equações de equilíbrio) e 8 variáveis (W_i , W'_i , y_i , x_1 e x_2), tendo portanto 2 variáveis de decisão, ou dois graus de liberdade. Tomando x_1 e x_2 como variáveis de decisão, a região de busca para o problema de otimização está delimitada pelas seguintes restrições de desigualdade:

$$x_0 > x_1 > x_2 > 0$$

$$L(x_1, x_2) > 0$$

A figura a seguir ilustra a região de busca para $F = 1,0 \times 10^4$ kg-água / h, $x_0 = 0,02$ kg-soluto / kg-água, $s = 7,0 \times 10^{-4}$ kg-solvente / kg-água, $m = 4,0$ kg-água / kg-solvente, $P_s = 0,4$ R\$ / kg-soluto e $P_x = 0,01$ R\$ / kg-solvente.



Outra aplicação muito comum, que utiliza as técnicas de otimização, é o ajuste de modelos matemáticos a dados experimentais (ajuste de curvas), onde a função objetivo utilizada, $S(x)$, é formada pelos quadrados dos desvios do modelo em relação aos dados experimentais (mínimos quadrados):

$$S(x) = \sum_{j=1}^N [y_j(x) - y_j^{\text{exp}}]^2$$

onde $y(x)$ é o modelo proposto (ou tipo de curva) para ser ajustado aos N dados experimentais y^{exp} , e x são os parâmetros de ajuste do modelo. Ou, se for utilizado o método da máxima verossimilhança:

$$S(x) = \sum_{j=1}^N \frac{1}{\sigma_j^2} [y_j(x) - y_j^{\text{exp}}]^2$$

onde σ_j^2 são as variâncias dos dados experimentais.

10.1.3 Procedimento geral

Não existe um único método que pode ser aplicado eficientemente para todos os problemas. O método escolhido para um caso particular depende das características da função objetivo, da natureza das restrições e do número de variáveis do problema. Contudo, os seguintes passos devem ser observados na formulação e solução de um problema de otimização:

- 1) Análise do processo e suas variáveis;
- 2) Determinação do critério para otimização e especificação da função objetivo em termos das variáveis do processo;

- 3) Desenvolvimento do modelo para o processo, relacionando as suas variáveis através de restrições de igualdade e desigualdade. Cálculo do grau de liberdade do sistema;
- 4) Se a formulação do problema é de dimensão elevada, então procurar particionar o problema em partes menores ou simplificar a função objetivo e/ou o modelo;
- 5) Aplicação de técnicas apropriadas de otimização;
- 6) Analisar a solução obtida e a sua sensibilidade frente a variações em parâmetros do modelo e suas considerações (hipóteses).

10.1.4 Dificuldades encontradas

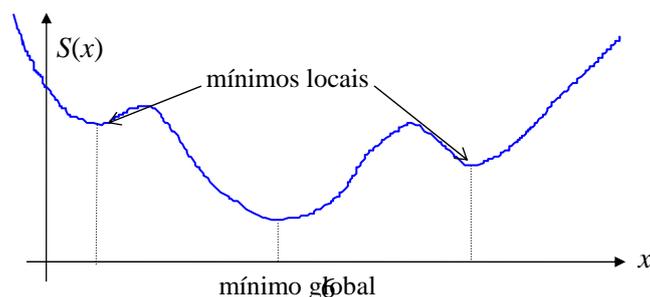
Problemas de otimização que apresentam funções objetivo e/ou restrições complicadas podem apresentar grandes dificuldades para obter a solução pelo uso de algumas técnicas de otimização. Algumas destas complicações estão listadas abaixo:

- função objetivo e/ou restrições podem apresentar descontinuidades;
- função objetivo e/ou restrições não lineares;
- função objetivo e/ou restrições podem estar definidas em termos de complicadas interações entre as variáveis, podendo ocorrer valores não únicos destas variáveis para o valor ótimo da função objetivo;
- função objetivo e/ou restrições podem exibir patamares (pouca sensibilidade à variação das variáveis) para alguns intervalos das variáveis e comportamento exponencial (muita sensibilidade) para outros intervalos;
- função objetivo pode apresentar ótimos locais.

10.2. Conceitos Básicos

10.2.1 Mínimos e máximos

Seja uma função $S: X \subseteq \mathfrak{R}^n \rightarrow \mathfrak{R}$. Diz-se que x^* é um mínimo global (ou absoluto) de S se $S(x^*) \leq S(x) \forall x \in X$, e que x^* é um mínimo local (ou relativo) de S se existe $\varepsilon > 0$, tal que $S(x^*) \leq S(x) \forall x$ tal que $\|x - x^*\| < \varepsilon$. Se as desigualdades forem estritas, isto é, $S(x^*) < S(x)$ tem-se mínimos globais e locais estritos.



Se existe um número β tal que $S(x) \geq \beta \forall x \in X$ e, para um ε suficientemente pequeno, $S(x) < \beta + \varepsilon$ para algum $x \in X$, então β é o ínfimo (ou valor inferior) de $S(x)$. Considerando os pontos $\pm\infty$, então toda função $S(x)$ tem um ínfimo e um supremo (ou valor superior) em X . Nem toda função tem mínimo (máximo), mas se ele existir deve ser finito e é obtido no ínfimo (supremo), isto é,

$$S(x^*) = \min_{x \in X} S(x) = \inf_{x \in X} S(x) \quad [S(x^*) = \max_{x \in X} S(x) = \sup_{x \in X} S(x)]$$

Por exemplo, a função $S(x) = e^x$ não tem máximo em $X = \mathfrak{R}$ e a função $S(x) = e^{-x}$ não tem mínimo em $X = \mathfrak{R}$. Contudo, o supremo de e^x é $+\infty$, e o ínfimo de e^{-x} é zero.

10.2.2 Condições para a otimalidade

Otimização sem restrição

No caso da otimização sem restrições, onde deseja-se encontrar os pontos extremos da função objetivo, tem-se as seguintes condições de otimalidade.

- Condição necessária de primeira ordem:

Para que x^* seja um mínimo (máximo) local da função $S(x)$, diferenciável em x^* , é necessário que:

$$\nabla S(x^*) = 0$$

- Condição necessária de segunda ordem:

Para que x^* seja um mínimo (máximo) local da função $S(x)$, duas vezes diferenciável em x^* , é necessário que:

$$\nabla S(x^*) = 0 \text{ e que}$$

$$H(x^*) \equiv \nabla^2 S(x^*) \text{ seja positiva (negativa) semidefinida}$$

onde $H(x^*)$ é chamada de matriz Hessiana.

Observa-se que estas condições são apenas necessárias porque os termos de primeira e segunda ordem podem estar nulos, deixando ainda dúvida sobre a natureza de x^* .

- Condição suficiente:

Seja $S(x)$ duas vezes diferenciável em x^* tal que:

$$\nabla S(x^*) = 0 \text{ e}$$

$$H(x^*) \text{ seja positiva (negativa) definida}$$

então x^* é um mínimo (máximo) local estrito de S .

Pode-se analisar a condição da matriz Hessiana, $H(x^*)$, pelas seguintes formas:

1) Pela sua contribuição no termo de segunda ordem da expansão em série de Taylor em torno do ponto ótimo.

$$S(x) - S(x^*) = \frac{1}{2}(x - x^*)^T H(x^*)(x - x^*) + \dots$$

2) Pelos sinais dos valores característicos de $H(x^*)$.

Decompondo a matriz Hessiana em seus valores e vetores característicos:

$$H(x^*) = V \Lambda V^1$$

onde V é a matriz dos vetores característicos (nas colunas) e Λ é a matriz dos valores característicos (na diagonal). Definindo $z(x) = V^1(x - x^*)$ e lembrando que sendo a matriz Hessiana simétrica então $V^1 = V^T$ (matriz ortogonal) e $(x - x^*)^T V = z^T$. Desta forma a expansão em série de Taylor pode ser escrita como:

$$S(x) - S(x^*) = \frac{1}{2} z^T \Lambda z + \dots = \frac{1}{2} \sum_{i=1}^n \lambda_i z_i^2 + \dots$$

3) Pelos sinais dos determinantes das primeiras menores principais de $H(x^*)$ (critério de Sylvester).

A *menor* M_{ij} de uma matriz H é definida como a matriz obtida pela remoção da i -ésima linha e da j -ésima coluna de H . Uma *menor principal* de ordem k é uma matriz obtida pela remoção de quaisquer $n - k$ colunas e suas linhas correspondentes de uma matriz de ordem n . A *primeira menor principal* de ordem k de uma matriz H , denotada por $M_k(H)$, é obtida pela remoção das últimas $n - k$ colunas e linhas da matriz H . Observa-se que os determinantes das primeiras menores principais de ordem $1, 2, \dots, n$ da matriz Λ são, respectivamente: $\lambda_1, \lambda_1\lambda_2, \dots, \lambda_1\lambda_2\lambda_3 \dots \lambda_n$.

Na tabela a seguir apresenta-se as relações entre a matriz Hessiana e as três formas de analisar a sua condição.

$H(x^*)$	Taylor	λ	$\Delta_k = \det(M_k)$
positiva semidefinida	$x^T H x \geq 0$	≥ 0	≥ 0
positiva definida	$x^T H x > 0$	> 0	> 0
negativa semidefinida	$x^T H x \leq 0$	≤ 0	$(-1)^k \Delta_k \geq 0$
negativa definida	$x^T H x < 0$	< 0	$(-1)^k \Delta_k > 0$
não definida	$x^T H x \begin{matrix} > 0 \\ < 0 \end{matrix}$	$\begin{matrix} > 0 \\ < 0 \end{matrix}$	\neq dos acima

Desta forma, pode-se afirmar que x^* é:

- ponto de mínimo local se $H(x^*)$ for positiva definida, $S(x) > S(x^*)$
- ponto de máximo local se $H(x^*)$ for negativa definida, $S(x) < S(x^*)$
- ponto de sela se $H(x^*)$ for não definida, ora $S(x) > S(x^*)$ e ora $S(x) < S(x^*)$

Nas situações onde $H(x^*)$ é semidefinida deve-se ainda investigar os termos de ordem superior da expansão em série de Taylor.

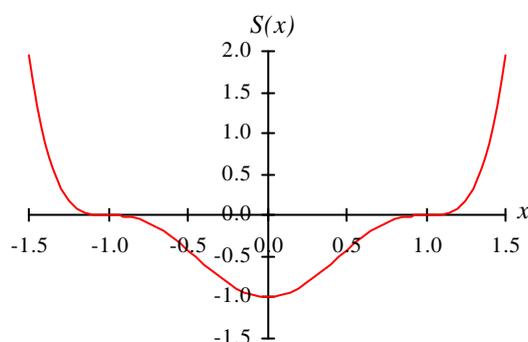
Exemplo 3: $S(x) = (x^2 - 1)^3$

$$\nabla S(x) = 6x(x^2 - 1)^2 \rightarrow \nabla S(x) = 0 \begin{cases} x_1 = 0 \\ x_2 = 1 \\ x_3 = -1 \end{cases}, \text{ satisfazem a condição necessária de primeira}$$

ordem;

$H(x) = (x^2 - 1)(30x^2 - 6) \rightarrow H(x_1) = 6; H(x_2) = 0; H(x_3) = 0$ satisfazem a condição necessária de segunda ordem; contudo somente x_1 satisfaz a condição suficiente. Neste caso (univariável) x_2 e x_3 são pontos de inflexão, como pode ser visto no gráfico abaixo, ou avaliando o valor da derivada terceira de $S(x)$ nestes pontos:

$$\nabla^3 S(x) = 24x(5x^2 - 3) \rightarrow \nabla^3 S(x_2) = 48 \neq 0 \text{ e } \nabla^3 S(x_3) = -48 \neq 0.$$



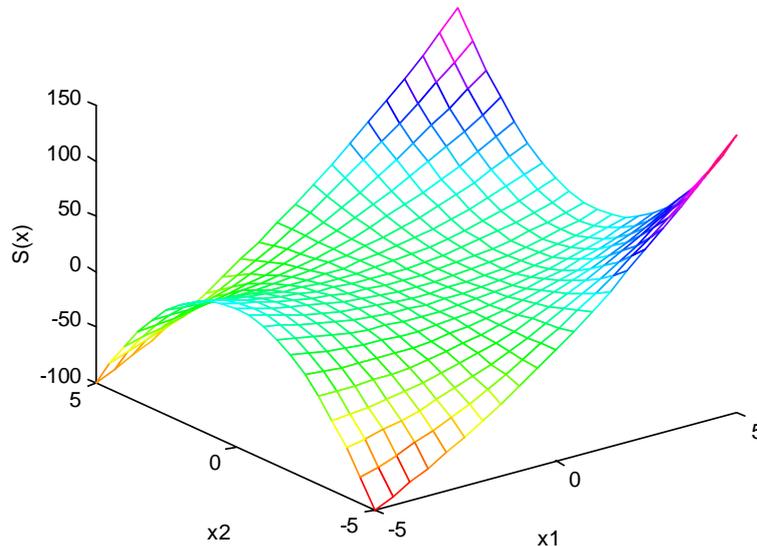
Exemplo 4: $S(x_1, x_2) = x_1^2 + x_1 x_2^2$

$$\nabla S(x) = \begin{bmatrix} 2x_1 + x_2^2 \\ 2x_1 x_2 \end{bmatrix}$$

então, $x_1^* = x_2^* = 0$, e:

$$H(x) = \begin{bmatrix} 2 & 2x_2 \\ 2x_2 & 2x_1 \end{bmatrix} \rightarrow H(x^*) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

isto é, uma matriz positiva semidefinida. O gráfico abaixo ilustra a função $S(x)$, onde observa-se que $x^* = 0$ não é um ponto de mínimo. Fazendo a mesma análise com a mudança de variável $y = x_2^2$, verifica-se que a origem é um ponto sela.



Otimização com restrições

Na otimização com restrição o problema que está sendo resolvido é:

$$\min S(x) \quad \text{ou} \quad \max S(x)$$

sujeito a restrições de igualdade e/ou de desigualdade, que definem a região viável, sendo que qualquer ponto nesta região é uma solução viável. Dependendo do tipo de função objetivo e de suas restrições, os problemas de otimização com restrição são comumente chamados de programação linear, programação quadrática, programação não-linear, programação inteira e programação mista.

Programação linear: função objetivo e restrições lineares

$$\begin{aligned} \min S(x) &= c^T x \\ \text{sujeito a: } & A x \leq b \\ & x \geq 0 \end{aligned}$$

onde A é uma matriz $m \times n$, isto é, m restrições e n variáveis.

Programação quadrática: função objetivo quadrática e restrições lineares

$$\begin{aligned} \min S(x) &= c^T x + \frac{1}{2} x^T Q x \\ \text{sujeito a: } & A x \leq b \\ & x \geq 0 \end{aligned}$$

Programação não-linear: função objetivo e/ou restrições não-lineares

$$\begin{aligned} & \min S(x) \\ & \text{sujeito a: } h_j(x) = 0, \quad j = 1, 2, \dots, m \\ & \quad \quad \quad g_j(x) \leq 0, \quad j = 1, 2, \dots, p \end{aligned}$$

onde $h_j(x)$ são m restrições de igualdade e $g_j(x)$ são p restrições de desigualdade.

Programação inteira: variáveis de decisão pertencem ao campo dos números inteiros.

Programação mista: é uma combinação da programação inteira com as demais. Por exemplo, no problema de projeto de trocadores de calor as variáveis de decisão poderiam ser as temperaturas das correntes (campo real) e os tipos de trocadores existentes (campo inteiro).

Seja o problema de otimização sujeito a restrições de igualdade, $h(x)$, e desigualdade, $g(x)$:

$$\begin{aligned} & \min S(x) \\ & \text{sujeito a: } h_j(x) = 0, \quad j = 1, 2, \dots, m \\ & \quad \quad \quad g_j(x) \leq 0, \quad j = 1, 2, \dots, p \\ & \quad \quad \quad x \in X \subseteq \mathfrak{R}^n \end{aligned}$$

onde $S(x)$, $g(x)$, e $h(x) \in C^2$. O conjunto de todos os pontos viáveis é definido por:

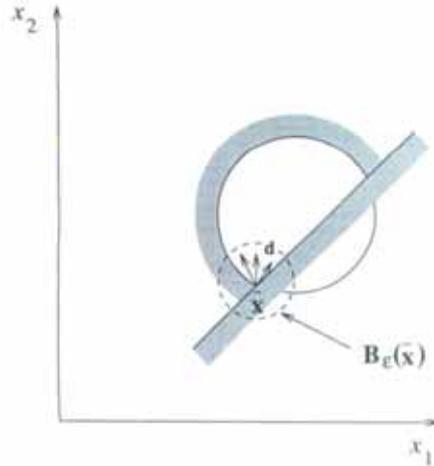
$$K = \{x \in X \subseteq \mathfrak{R}^n / h(x) = 0, g(x) \leq 0\}$$

Uma restrição de desigualdade $g_j(x)$ é chamada de *ativa* em um ponto viável \bar{x} se $g_j(\bar{x}) = 0$, caso contrário ela é uma *restrição inativa*. As restrições ativas restringem a região de viabilidade, enquanto que as inativas não impõem restrição alguma na vizinhança do ponto \bar{x} , definida pela hipersfera de raio ε em torno deste ponto, denotada por $B_\varepsilon(\bar{x})$.

Um vetor d é chamado de *vetor de direção viável* a partir do ponto \bar{x} se existe uma hipersfera de raio ε tal que:

$$(\bar{x} + \lambda d) \in \{B_\varepsilon(\bar{x}) \cap K\} \text{ para todo } 0 \leq \lambda \leq \frac{\varepsilon}{\|d\|}.$$

O conjunto de vetores de direções viáveis a partir de \bar{x} é chamado de *cone de direções viáveis* de K no ponto \bar{x} . A figura abaixo ilustra estas definições.



Se $d \neq 0$, então \bar{x} deve satisfazer as seguintes condições:

$$d^T \nabla h(\bar{x}) = 0$$

$$d^T \nabla g(\bar{x}) \leq 0 \text{ para as } g(\bar{x}) \text{ ativas, pois } g(x) \approx g(\bar{x}) + \lambda \nabla^T g(\bar{x})d \leq 0$$

e se $d^T \nabla S(\bar{x}) < 0$, então d é uma direção viável e promissora, isto é,

$$S(\bar{x} + \lambda d) < S(\bar{x}) \text{ para todo } 0 \leq \lambda \leq \frac{\varepsilon}{\|d\|}, \text{ pois } S(x) - S(\bar{x}) \approx \lambda \nabla^T S(\bar{x})d < 0.$$

Se $\bar{x} = x^*$ é um ponto de mínimo local do problema, então para um λ suficientemente pequeno, tem-se:

$$S(x^*) \leq S(x^* + \lambda d).$$

A idéia chave para desenvolver as condições necessárias e suficientes para um problema de otimização com restrições é transformá-lo em um problema de otimização sem restrições e aplicar as condições para este caso. Uma forma de fazer esta transformação é através da introdução de uma função auxiliar, chamada de função de Lagrange, $L(x, \lambda, \mu)$, definida como:

$$L(x, \lambda, \mu) = S(x) + \lambda^T h(x) + \mu^T g(x) \quad , \quad \mu \geq 0$$

onde λ e μ são os multiplicadores de Lagrange associados com as restrições de igualdade e desigualdade, respectivamente (μ são também conhecidos como multiplicadores de Kuhn-Tucker). Deste modo, o problema transformado torna-se:

$$\max_{\lambda, \mu \geq 0} \min_x L(x, \lambda, \mu)$$

onde os multiplicadores λ associados com as restrições de igualdade, $h(x) = 0$, assumem sinais positivos quando $h(x) \geq 0$ e negativos quando $h(x) \leq 0$. No ponto ótimo tem-se:

$$L(x^*, \lambda^*, \mu^*) = S(x^*)$$

Cada multiplicador de Lagrange indica o quão sensível é a função objetivo em relação à restrição associada. Por exemplo, se as restrições de igualdade são perturbadas por um vetor b , isto é, $h_b(x) = b$, então:

$$\nabla_b S(x^*) = -\lambda^*.$$

Note que neste caso a função de Lagrange tem a forma:

$$L(x, \lambda) = S(x) + \lambda^T [h_b(x) - b]$$

e sua sensibilidade em relação ao parâmetro b é dada por:

$$\nabla_b L = \nabla_b^T x [\nabla_x S(x) + \nabla_x^T h_b(x) \lambda] + \nabla_b^T \lambda [h_b(x) - b] - \lambda$$

Como os termos entre colchetes da expressão acima devem ser nulos no ponto ótimo (condição necessária de primeira ordem), então:

$$\nabla_b L(x^*, \lambda^*) = -\lambda^* \quad \text{e} \quad \nabla_b S(x^*) = -\lambda^*$$

pois $\nabla_b L(x^*, \lambda^*) = \nabla_b S(x^*) + \nabla_b^T \lambda [h_b(x^*) - b] + \nabla_b^T h_b(x^*) \lambda^* - \lambda^*$,

$$h_b(x) = b \quad \text{e} \quad \nabla_b h_b(x) = I$$

Portanto, o valor de $S(x)$ aumenta ou diminui a partir de $S(x^*)$ com um aumento ou diminuição em b , dependendo do sinal de λ^* . Por isso, os multiplicadores de Lagrange são também conhecidos como “*shadow prices*” ou “custos marginais” das restrições, porque a mudança no valor ótimo da função objetivo por unidade de acréscimo no lado direito da restrição de igualdade é dado por λ^* .

Exemplo 5: Seja o seguinte problema de otimização com restrição

$$\min S(x) = (x_1 - 5)^2 + (x_2 - 5)^2$$

$$\text{sujeito a: } h(x) = x_1 - x_2 = 0$$

Introduzindo uma perturbação na restrição de igualdade do tipo: $x_1 - x_2 = b$, então a função de Lagrange toma a forma:

$$L(x, \lambda) = (x_1 - 5)^2 + (x_2 - 5)^2 + \lambda (x_1 - x_2 - b)$$

cujos gradientes nulos com relação a x_1 , x_2 e λ leva ao seguinte sistema de equações:

$$2(x_1 - 5) + \lambda = 0$$

$$2(x_2 - 5) - \lambda = 0$$

$$x_1 - x_2 - b = 0$$

resultando na solução ótima: $x_1^* = 5 + b/2$, $x_2^* = 5 - b/2$ e $\lambda^* = -b$.

Deste modo $S(x^*) = b^2/2$ e $\nabla_b S(x^*) = b = -\lambda^*$.

Para entender a origem da função de Lagrange, o ótimo do exemplo acima deve satisfazer as seguintes condições:

$$\delta S = \frac{\partial S}{\partial x_1} \delta x_1 + \frac{\partial S}{\partial x_2} \delta x_2 = 0$$

$$\delta h = \frac{\partial h}{\partial x_1} \delta x_1 + \frac{\partial h}{\partial x_2} \delta x_2 = 0$$

Se $S(x)$ fosse uma função sem restrição, então as suas duas derivadas parciais seriam nulas no ponto ótimo e $\delta S(x^*)$ seria nulo para quaisquer valores das variações δx_1 e δx_2 . Entretanto, como as variáveis x_1 e x_2 estão restritas (δx_1 e δx_2 não são independentes), as duas derivadas parciais de $S(x)$ não precisam ser igualadas a zero. Contudo, $S(x)$ deve ser um ponto estacionário no ponto ótimo e portanto $\delta S(x^*) = 0$. A segunda condição, $\delta h(x^*) = 0$, existe porque $h(x) = 0$. Para se obter uma solução (δx_1 e δx_2) não-trivial do sistema de equações acima, a matriz dos coeficientes do sistema:

$$\begin{bmatrix} \frac{\partial S}{\partial x_1} & \frac{\partial S}{\partial x_2} \\ \frac{\partial h}{\partial x_1} & \frac{\partial h}{\partial x_2} \end{bmatrix}$$

deve ter determinante nulo, ou seja, as linhas da matriz são linearmente dependentes:

$$\frac{\partial S}{\partial x_1} + \lambda \frac{\partial h}{\partial x_1} = 0 \quad \text{e} \quad \frac{\partial S}{\partial x_2} + \lambda \frac{\partial h}{\partial x_2} = 0$$

Então, definindo uma função auxiliar: $L(x, \lambda) = S(x) + \lambda^T h(x)$

as condições acima são satisfeitas se: $\nabla_x L(x, \lambda) = 0$. Para que a restrição de igualdade, $h(x) = 0$, seja também satisfeita é necessário que $\nabla_\lambda L(x, \lambda) = 0$. Portanto, no ponto ótimo é necessário que $\nabla L(x^*, \lambda^*) = 0$.

A existência dos multiplicadores de Lagrange depende da forma das restrições, e estará garantida se e somente se os gradientes das restrições de desigualdade ativas, $\nabla g_j(x^*)$, e das restrições de igualdade, $\nabla h(x^*)$, forem linearmente independentes. Por exemplo, no caso de um problema somente com restrições de igualdade, a condição necessária de primeira ordem para $L(x, \lambda)$ fica:

$$\nabla_x S(x) + [\nabla_x h(x)]^T \lambda = 0$$

cuja solução para λ existirá somente se a matriz $\nabla_x h(x)$ possuir posto completo, m , isto é, estar composta por m vetores linearmente independentes.

- Condição necessária de primeira ordem de Karush-Kuhn-Tucker (KKT):

Para que x^* seja um ótimo local do problema com restrições, com $S(x)$, $g(x)$, e $h(x)$ diferenciáveis em x^* , é necessário que:

os gradientes das restrições de desigualdade ativas, $\nabla g_j(x^*)$, e das restrições de igualdade, $\nabla h(x^*)$, sejam linearmente independentes (*qualificação de segunda ordem* das restrições), e que as seguintes condições sejam satisfeitas:

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla S(x^*) + (\lambda^*)^T \nabla h(x^*) + (\mu^*)^T \nabla g(x^*) = 0$$

$$h(x^*) = 0$$

$$g(x^*) \leq 0$$

$$\mu_j^* g_j(x^*) = 0, j = 1, 2, \dots, p \text{ (condições de complementaridade)}$$

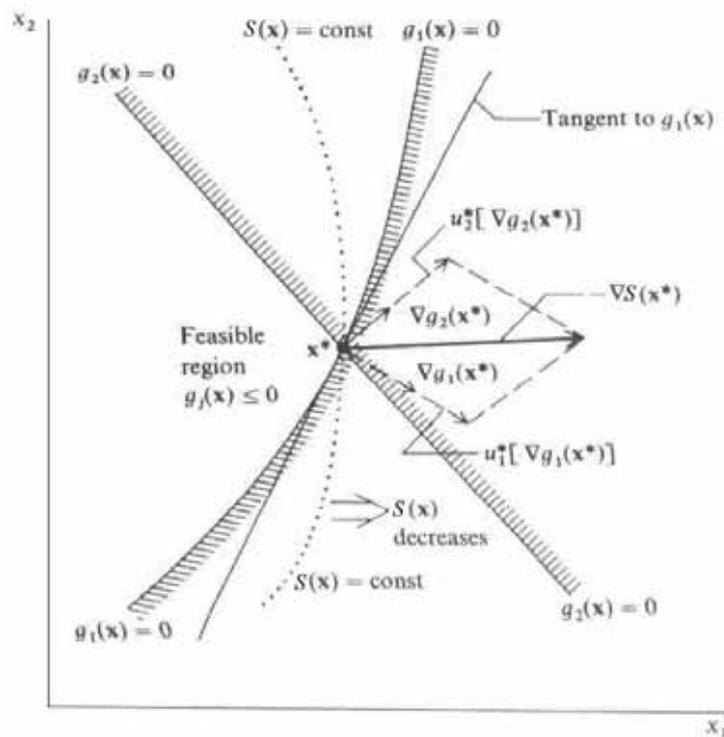
$$\mu^* \geq 0$$

A condição de independência linear pode ser relaxada por outras qualificações de primeira e segunda ordens das restrições (Floudas, 1995, pg. 59 e 64).

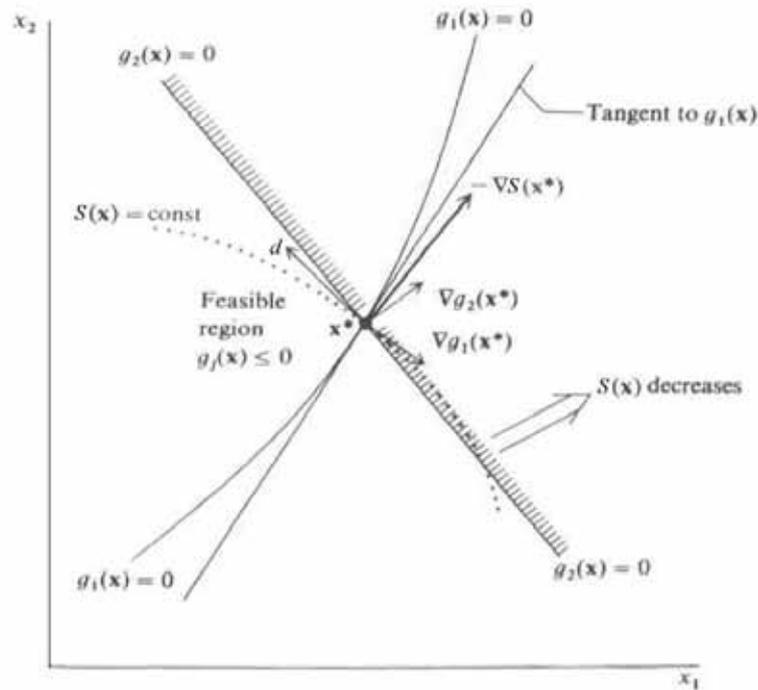
A condição do gradiente nulo, $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, implica em:

$$(\lambda^*)^T \nabla h(x^*) + (\mu^*)^T \nabla g(x^*) = -\nabla S(x^*)$$

que interpretada graficamente, figura abaixo, mostra que o vetor $-\nabla S(x^*)$ pertence ao cone das direções viáveis, formado pelos gradientes das restrições de igualdade e desigualdade ativas (uma vez que $\mu_j^* = 0$ para as restrições inativas).



Supondo que $-\nabla S(x^*)$ caia fora do cone das direções viáveis, então haveria uma direção d tal que $d^T \nabla S(x^*) < 0$, $d^T \nabla g(x^*) \leq 0$ e $d^T \nabla h(x^*) = 0$, isto é, existiria um ponto melhor que x^* , como ilustra a figura abaixo.



• Condição necessária de segunda ordem de KKT:

Para que x^* seja um mínimo local do problema com restrições, com $S(x)$, $g(x)$, e $h(x)$ duas vezes diferenciáveis em x^* , é necessário que:

a condição de primeira ordem de KKT seja satisfeita e, que a matriz Hessiana da função de Lagrange, $\nabla_x^2 L(x^*, \lambda^*, \mu^*)$, seja positiva semidefinida para todo vetor não nulo d tal que:

$$d^T \nabla h_i(x^*) = 0, \quad i = 1, 2, \dots, m$$

$$d^T \nabla g_j(x^*) = 0 \text{ para as } g_j(x^*) \text{ ativas}$$

isto é, $d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d \geq 0$.

• Condição suficiente de KKT:

Para que x^* seja um mínimo local do problema com restrições, com $S(x)$, $g(x)$, e $h(x)$ duas vezes diferenciáveis em x^* , é suficiente que:

a condição de primeira ordem de KKT seja satisfeita e, que a matriz Hessiana da função de Lagrange, $\nabla_x^2 L(x^*, \lambda^*, \mu^*)$, seja positiva definida para todo vetor não nulo d tal que:

$$d^T \nabla h_i(x^*) = 0, \quad i = 1, 2, \dots, m$$

$$d^T \nabla g_j(x^*) = 0 \text{ para as } g_j(x^*) \text{ ativas } \{g_j(x^*) = 0 \text{ e } \mu_j^* > 0\}$$

$$d^T \nabla g_j(x^*) \leq 0 \text{ para as } g_j(x^*) \text{ inativas } \{g_j(x^*) < 0 \text{ e } \mu_j^* = 0\}$$

isto é, $d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d > 0$.

A positividade da matriz Hessiana com restrição, isto é:

$$d^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) d > 0 \quad \forall d \in \{d / d^T \nabla h_i(x^*) = 0, d^T \nabla g_j(x^*) = 0, d \neq 0\}$$

é garantida se todas as raízes do polinômio característico

$$p(\lambda) = \begin{vmatrix} \lambda I - \nabla_x^2 L & M \\ M^T & 0 \end{vmatrix} = 0$$

forem positivas, onde M é a matriz formada pelos gradientes de $h(x^*)$ e $g(x^*)$ ativas, isto é, a matriz tal que $d^T M = 0$, com $m+p^a < n$ e com posto completo (p^a é o número de restrições g ativas). O mesmo critério se aplica para semipositividade, negatividade e seminegatividade, com os respectivos sinais das raízes.

Uma maneira de resolver o problema de valor característico acima é usando a decomposição QR (ou decomposição ortogonal-triangular, ou decomposição de Householder) da matriz $M (= Q R)$ para obter a matriz de projeção, Z^T , do vetor d no sub-espaço nulo de M , isto é, $d^T M = 0$, ou então a decomposição em valores singulares da matriz $M^T (= U S V^H)$. A matriz Z é formada pelas últimas w colunas de Q , ou ainda pelas últimas w colunas de V , onde w é a dimensão do espaço nulo (número de valores singulares nulos, ou número de linhas nulas da matriz R):

$$Z_{ij} = Q_{ij} = V_{ij} \quad i = 1, 2, \dots, n, \quad j = m+p^a-w+1, \dots, m+p^a$$

onde $Q^H M = \begin{bmatrix} R \\ 0 \end{bmatrix}$ (Q é uma matriz unitária e $Q^{-1} = Q^H$ é a transposta conjugada)

Uma vez encontrada a matriz Z , obtém-se os valores característicos da matriz Hessiana projetada neste sub-espaço: $Z^T \nabla_x^2 L Z$.

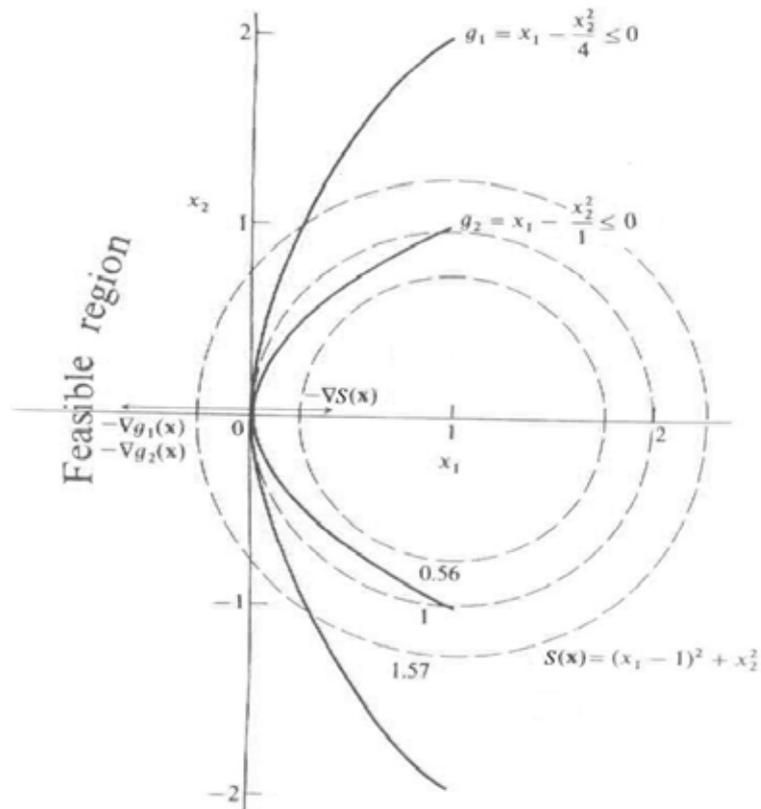
Exemplo 6: Verificar as condições necessárias e suficientes para o seguinte problema (Edgar & Himmelblau, 1988, pg. 314):

$$\min S(x) = (x_1 - 1)^2 + x_2^2$$

$$\text{sujeito a: } g_1(x) = x_1 - x_2^2 / 4 \leq 0$$

Exemplo 7: Verificar as condições necessárias e suficientes para o problema com a mesma função objetivo do exemplo 2.4, mas usando a seguinte restrição:

$$g_2(x) = x_1 - x_2^2 \leq 0$$



Exemplo 8: Verificar se o ponto $x^* = [1 \ 4,9]^T$ é um mínimo local para o seguinte problema (Edgar & Himmelblau, 1988 , pg. 316):

$$\min S(x) = 4 x_1 - x_2^2 - 12$$

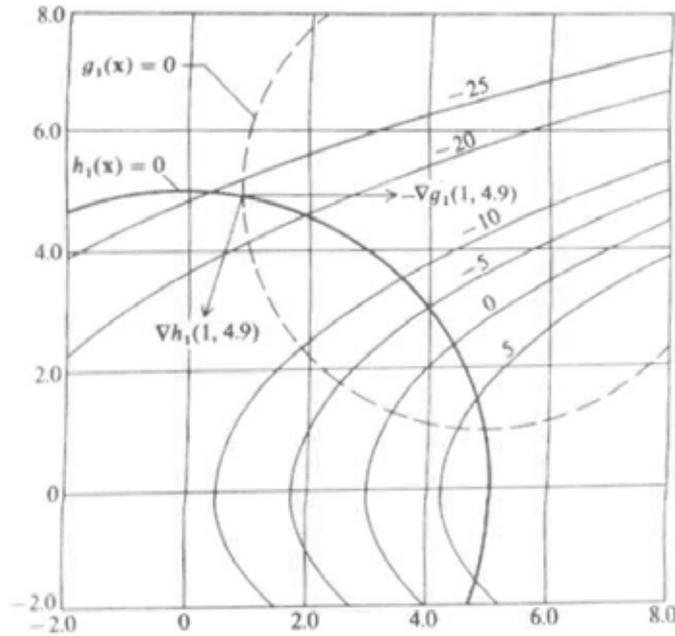
$$\text{sujeito a: } h_1(x) = 25 - x_1^2 - x_2^2 = 0$$

$$g_1(x) = x_1^2 - 10 x_1 + x_2^2 - 10 x_2 + 34 \leq 0$$

$$g_2(x) = -(x_1 - 3)^2 - (x_2 - 1)^2 \leq 0$$

$$g_3(x) = -x_1 \leq 0$$

$$g_4(x) = -x_2 \leq 0$$



10.2.3 Convexidade

Um subconjunto K de um espaço vetorial X é dito convexo se $\forall x_1, x_2 \in K$ e $0 \leq \alpha \leq 1$:

$$\alpha x_1 + (1 - \alpha) x_2 \in K$$

e apresenta as seguintes propriedades:

- $\beta K = \{x \in K / x = \beta y, y \in K\}$ é convexo para $\forall \beta \in \mathfrak{R}$;
- $K + L$ e $K \cap L$ são convexos para \forall subconjunto convexo L de X .

Seja K um convexo não vazio do \mathfrak{R}^n . A função $S: K \rightarrow \mathfrak{R}$ é dita convexa se $\forall x_1, x_2 \in K$ e $0 \leq \alpha \leq 1$:

$$S[\alpha x_1 + (1 - \alpha) x_2] \leq \alpha S(x_1) + (1 - \alpha) S(x_2)$$

A função $S(x)$ é estritamente convexa se a desigualdade for estrita. Uma função $T(x)$ é côncava se a função $S(x) = -T(x)$ for convexa.

Sejam $S(x), S_i(x), i=1,2,\dots,n$, funções convexas em um convexo não vazio K do \mathfrak{R}^n , então as seguintes propriedades são apresentadas:

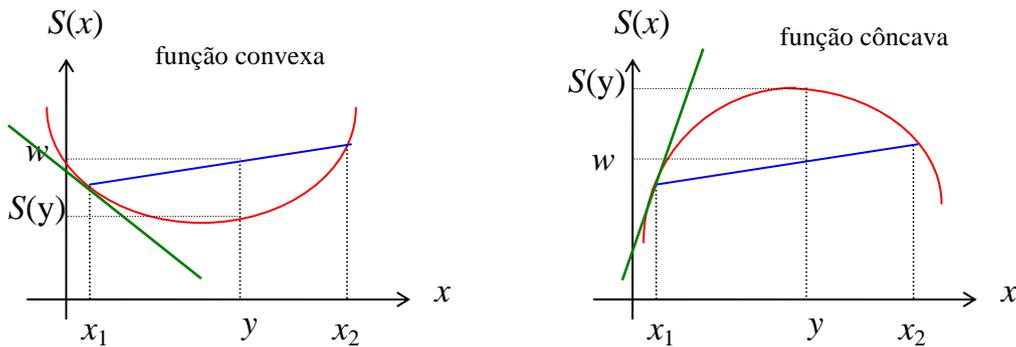
- $S(x)$ é contínua em qualquer ponto do interior de K ;
- as seções $K_\varepsilon = \{x \in K / S(x) \leq \varepsilon\}$ são conjuntos convexas;
- $S(x) = \sum_{i=1}^n S_i(x)$ é uma função convexa. Se no mínimo uma $S_i(x)$ é estritamente convexa, então $S(x)$ é estritamente convexa;

d) $\beta S(x)$ é convexa para $\beta > 0 \in \mathfrak{R}$.

e) se todas $S_i(x) < \infty \forall x \in K$, então $S(x) = \max\{S_1(x), S_2(x), \dots, S_n(x)\}$ é convexa.

Quando $S(x)$ é convexa, as condições de otimalidade simplificam-se, porque as condições de segunda ordem são equivalentes à convexidade local da função. Além disso, um mínimo local será também global e se a função for estritamente convexa o mínimo global é único.

Para ilustrar, define-se $y = \alpha x_1 + (1 - \alpha) x_2$ e $w = \alpha S(x_1) + (1 - \alpha) S(x_2)$



Seja X um conjunto aberto não vazio do \mathfrak{R}^n e $S(x)$ uma função diferenciável em $x^0 \in X$. Se $S(x)$ é convexa em x^0 , então

$$S(x) - S(x^0) \geq \nabla^T S(x^0)(x - x^0)$$

Para uma função $S(x)$ diferenciável em X ,

$$S(x) \text{ é convexa} \Leftrightarrow S(x^2) - S(x^1) \geq \nabla^T S(x^1)(x^2 - x^1) \quad \forall x_1, x_2 \in X.$$

Seja X um conjunto aberto não vazio do \mathfrak{R}^n e $S(x)$ uma função duas vezes diferenciável em $x^0 \in X$. Se $S(x)$ é convexa em x^0 , então $\nabla^2 S(x^0)$ é positiva semidefinida. Para uma função $S(x)$ duas vezes diferenciável em X ,

$$S(x) \text{ é convexa} \Leftrightarrow \nabla^2 S(x) \text{ é positiva semidefinida} \quad \forall x \in X.$$

Seja K um conjunto convexo não vazio do \mathfrak{R}^n , $x^0 \in K$ e d um vetor não nulo tal que $(x^0 + \alpha d) \in K$ para um $\alpha > 0$ suficientemente pequeno. Então, a *derivada direcional* de $S(x)$ no ponto x^0 , ao longo da direção α , denotada por $S'(x^0, d)$, é definida pelo seguinte limite (incluindo $\pm\infty$):

$$S'(x^0, d) = \lim_{\alpha \rightarrow 0^+} \frac{S(x^0 + \alpha d) - S(x^0)}{\alpha} \approx \lim_{\alpha \rightarrow 0^+} (\nabla^T S(x^0) d + \alpha d^T \nabla^2 S(x^0) d)$$

Portanto, a derivada direcional no ponto x^0 é dada por $S'(x^0, d) = \nabla^T S(x^0) d$.

Seja K um conjunto convexo não vazio do \mathfrak{R}^n e $S(x)$ uma função convexa. Então, o sub-gradiente de $S(x)$ no ponto $x^0 \in K$, denotado por d , é definido como:

$$S(x) \geq S(x^0) + d^T (x - x^0) \quad , \quad \forall x \in K$$

Isto é, a aproximação linear com o uso do vetor d sempre resulta em uma sub-estimativa de $S(x)$.

Generalização de funções convexas

Para generalizar a definição de funções convexas é necessário definir a continuidade de funções escalares multivariáveis.

Sejam $X \subseteq \mathfrak{R}^n$, $x^0 \in X$, e $S(x): X \rightarrow \mathfrak{R}$.

$S(x)$ é contínua em x^0 se:

para cada $\varepsilon > 0$ existir $\delta(\varepsilon)$ tal que $\|x - x^0\| < \delta \Rightarrow |S(x) - S(x^0)| < \varepsilon$, ou

para cada seqüência x^1, x^2, \dots, x^m em X convergindo para x^0 :

$$\lim_{m \rightarrow \infty} S(x^m) = S\left(\lim_{m \rightarrow \infty} x^m\right) = S(x^0).$$

$S(x)$ é contínua em X se ela for contínua para todo $x^0 \in X$.

$S(x)$ é semicontínua inferior em x^0 se:

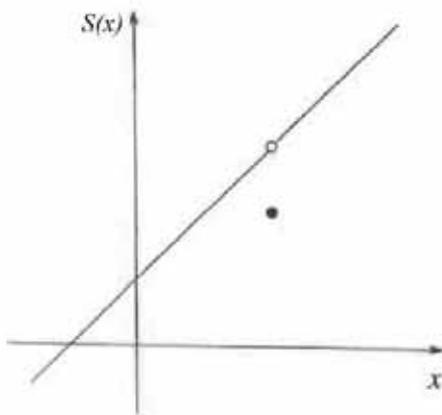
para cada $\varepsilon > 0$ existir $\delta(\varepsilon)$ tal que $\|x - x^0\| < \delta \Rightarrow -\varepsilon < S(x) - S(x^0)$, ou

para cada seqüência x^1, x^2, \dots, x^m em X convergindo para x^0 :

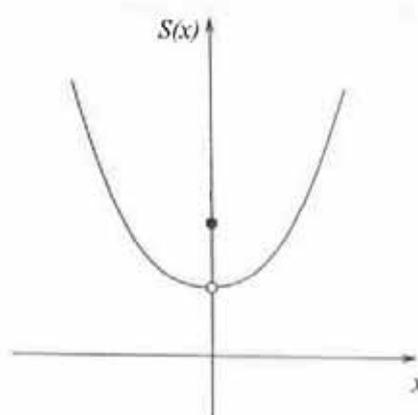
$$\liminf_{m \rightarrow \infty} S(x^m) = S\left(\lim_{m \rightarrow \infty} x^m\right) = S(x^0),$$

onde $\liminf_{m \rightarrow \infty} S(x^m) = \lim_{m \rightarrow \infty} \min\{S(x^1), S(x^2), \dots, S(x^m)\}$.

$S(x)$ é semicontínua inferior em X se ela for semicontínua inferior para todo $x^0 \in X$.



função semicontínua inferior



função semicontínua superior

$S(x)$ é semicontínua superior em x^0 se:

para cada $\varepsilon > 0$ existir $\delta(\varepsilon)$ tal que $\|x - x^0\| < \delta \Rightarrow S(x) - S(x^0) < \varepsilon$, ou para cada seqüência x^1, x^2, \dots, x^m em X convergindo para x^0 :

$$\limsup_{m \rightarrow \infty} S(x^m) = S\left(\lim_{m \rightarrow \infty} x^m\right) = S(x^0),$$

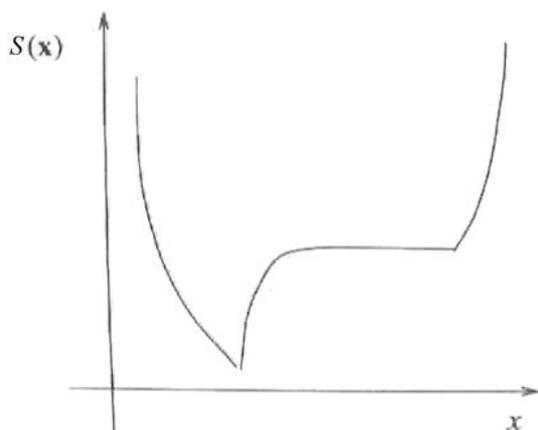
onde $\limsup_{m \rightarrow \infty} S(x^m) = \lim_{m \rightarrow \infty} \max\{S(x^1), S(x^2), \dots, S(x^m)\}$.

$S(x)$ é semicontínua superior em X se ela for semicontínua superior para todo $x^0 \in X$.

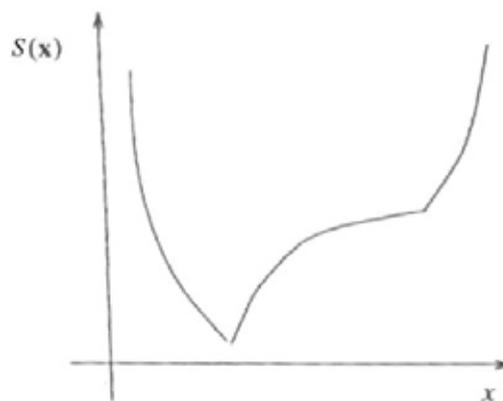
Seja K um convexo não vazio do \mathfrak{R}^n . A função $S: K \rightarrow \mathfrak{R}$ é dita *quasi-convexa* se $\forall x_1, x_2 \in K$ e $0 \leq \alpha \leq 1$:

$$S[\alpha x_1 + (1 - \alpha) x_2] \leq \max\{S(x_1), S(x_2)\}$$

A função $S(x)$ é estritamente quasi-convexa se a desigualdade for estrita quando $S(x_1) \neq S(x_2)$. Uma função $T(x)$ é (estritamente) *quasi-côncava* se a função $S(x) = -T(x)$ for (estritamente) quasi-convexa. Um mínimo local de uma função estritamente quasi-convexa será também global.



Função quasi-convexa



Função estritamente quasi-convexa

Definindo as seções $K_\varepsilon = \{x \in K / S(x) \leq \varepsilon\}$, então:

$$S(x) \text{ é quasi-convexa} \Leftrightarrow K_\varepsilon \text{ é convexo } \forall \varepsilon \in \mathfrak{R}.$$

Seja $S(x)$ uma função semicontínua inferior no convexo $K \subseteq \mathfrak{R}^n$. Se $S(x)$ é estritamente quasi-convexa em K , então $S(x)$ é quasi-convexa, mas o converso não é verdadeiro.

A soma de funções quasi-convexas não é necessariamente uma função quasi-convexa, propriedade válida para funções convexas. Por outro lado, o recíproco de uma função quasi-convexa, $1/S(x)$, é uma função quasi-côncava, mas o recíproco de uma função convexa não é

necessariamente uma função côncava (já o recíproco de uma função côncava positiva é uma função convexa). Por exemplo, $S(x) = e^x$ é convexa e $1/S(x)$ é convexa também.

Seja $S(x)$ uma função diferenciável em um conjunto convexo aberto não vazio $K \subseteq \mathfrak{R}^n$, então:

$$S(x) \text{ é quasi-convexa} \Leftrightarrow S(x^2) \leq S(x^1) \rightarrow \nabla^T S(x^1)(x^2 - x^1) \leq 0 \quad \forall x_1, x_2 \in K.$$

Seja $S(x)$ uma função quasi-côncava duas vezes diferenciável em um conjunto convexo aberto não vazio $K \subseteq \mathfrak{R}^n$, então uma direção, d , ortogonal ao $\nabla S(x)$ exibe as seguintes propriedades:

- a) se $x^0 \in K$, $d \in \mathfrak{R}^n$ e $d^T \nabla S(x^0) = 0$, então $d^T \nabla^2 S(x^0) d \leq 0$;
- b) a matriz Hessiana de $S(x)$ tem no máximo um valor característico positivo $\forall x \in K$. Isto é, a generalização da concavidade de funções (para quasi-côncavidade) é equivalente à existência de no máximo um valor característico positivo da Hessiana.

Seja $S(x)$ uma função diferenciável em um conjunto aberto não vazio $X \subseteq \mathfrak{R}^n$.

$$S(x) \text{ é pseudo-convexa se } S(x^2) < S(x^1) \rightarrow \nabla^T S(x^1)(x^2 - x^1) < 0 \quad \forall x_1, x_2 \in X.$$

Uma função $T(x)$ é *pseudo-côncava* se a função $S(x) = -T(x)$ for pseudo-convexa. O recíproco de uma função pseudo-côncava, $1/S(x)$, é uma função pseudo-côncava. As seguintes relações são válidas:

- a) uma função convexa diferenciável é pseudo-convexa;
- b) uma função convexa é estritamente quasi-convexa;
- c) uma função convexa é quasi-convexa;
- d) uma função pseudo-convexa é estritamente quasi-convexa;
- e) uma função estritamente quasi-convexa e semicontínua inferior é quasi-convexa.

10.2.4 Formas funcionais

O método empregado para solucionar um problema de otimização depende, fundamentalmente, da forma da função objetivo e de suas restrições. Com relação a função objetivo tem-se:

função linear:
$$S(x) = a + c^T x = a + \sum_{i=1}^n c_i x_i$$

como neste caso $\nabla S(x) = c \neq 0 \quad \forall x$, a função $S(x)$ não apresenta pontos de máximo ou mínimo. Para este tipo de função objetivo, só faz sentido a existência de um problema de otimização com restrição, cuja solução, se existir, estará localizada em algum ponto da fronteira da região de busca.

função quadrática: $S(x) = a + c^T x + \frac{1}{2} x^T A x = a + \sum_{i=1}^n c_i x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$

como $x^T A x$ é um escalar $\Rightarrow x^T A x = [x^T A x]^T = x^T A^T x$,

tem-se que: $x^T A x = \frac{1}{2} (x^T A x + x^T A^T x) = x^T \left[\frac{1}{2} (A + A^T) \right] x$,

então definindo $Q = \frac{1}{2} (A + A^T)$, que é uma matriz simétrica: $Q = Q^T$,

resulta em: $x^T A x = x^T Q x$, isto é:

$$S(x) = a + c^T x + \frac{1}{2} x^T Q x = a + \sum_{i=1}^n c_i x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j$$

e portanto:

$$\nabla S(x) = \left[\frac{\partial S}{\partial x_k} \right] = \left[c_k + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n q_{ij} \left(\frac{\partial x_i}{\partial x_k} x_j + x_i \frac{\partial x_j}{\partial x_k} \right) \right],$$

mas como $\frac{\partial x_i}{\partial x_k} = \delta_{ik} = \begin{cases} 0 & , i \neq k \\ 1 & , i = k \end{cases}$, tem-se: $\nabla S(x) = \left[c_k + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (q_{kj} x_j + q_{ik} x_i) \right]$

e sabendo que a matriz Q é simétrica:

$$\nabla S(x) = \left[c_k + \sum_{i=1}^n q_{ik} x_i \right] = c + Qx$$

e $H(x) = \nabla^2 S(x) = Q$

com isto, a condição necessária de primeira ordem, $\nabla S(x^*) = 0$, para este tipo de função resulta em:

$$Q x^* = -c \quad \text{ou de forma similar} \quad (x^*)^T Q = -c^T$$

e

$$S(x^*) = a + c^T x^* + \frac{1}{2} (x^*)^T Q x^* = a + \frac{1}{2} c^T x^*$$

Subtraindo $S(x^*)$ da função $S(x)$ tem-se:

$$S(x) = S(x^*) + c^T (x - x^*) + \frac{1}{2} [x^T Q x - (x^*)^T Q x^*]$$

como $c^T = -(x^*)^T Q$,

$$S(x) = S(x^*) + \frac{1}{2} \left[x^T Q x - (x^*)^T Q x^* - (x^*)^T Q (x - x^*) - (x^*)^T Q (x - x^*) \right]$$

$$S(x) = S(x^*) + \frac{1}{2} \left[x^T Q x - (x^*)^T Q x - (x^*)^T Q (x - x^*) \right]$$

$$S(x) = S(x^*) + \frac{1}{2} \left[(x - x^*)^T Q x - (x^*)^T Q (x - x^*) \right]$$

$$S(x) = S(x^*) + \frac{1}{2} (x - x^*)^T Q (x - x^*)$$

função não linear: $S(x): \mathfrak{R}^n \rightarrow \mathfrak{R}$

a expansão em série de Taylor de $S(x)$ em torno de um ponto x_0 é dada por:

$$S(x) = S(x_0) + \nabla^T S(x_0) (x - x_0) + \frac{1}{2} (x - x_0)^T \nabla^2 S(x_0) (x - x_0) + \dots$$

No caso particular de $x_0 = x^*$, tem-se $\nabla S(x^*) = 0$ e:

$$S(x) \approx S(x^*) + \frac{1}{2} (x - x^*)^T H(x^*) (x - x^*)$$

Se $S(x)$ for duas vezes diferenciável em x^* , então $H(x^*) = \left[\frac{\partial^2 S}{\partial x_i \partial x_j} \right]_{x^*}$ é simétrica.

Portanto, tanto no caso de função quadrática, onde $H(x^*) = Q$, quanto no caso geral de função não-linear, a matriz Hessiana fornece as características do(s) ponto(s) ótimo(s) de $S(x)$.

Algumas observações adicionais:

- 1) no caso de $S(x)$ ser uma função quadrática, o ponto ótimo é global;
- 2) se $H(x^*)$ é positiva definida, então $S(x)$ é estritamente convexa na vizinhança de x^* (ou $\forall x$ no caso da função quadrática);
- 3) se $H(x^*)$ é positiva semidefinida, então $S(x)$ é convexa na vizinhança de x^* (ou $\forall x$ no caso da função quadrática)..
- 4) se $H(x^*)$ é negativa definida, então $S(x)$ é estritamente côncava na vizinhança de x^* (ou $\forall x$ no caso da função quadrática);
- 5) se $H(x^*)$ é negativa semidefinida, então $S(x)$ é côncava na vizinhança de x^* (ou $\forall x$ no caso da função quadrática).

função mista linear e inteira: $S(x, y) = c^T x + d^T y$, $x \in X \subseteq \mathfrak{R}^n$ e $y \in Y$

onde y é um vetor de variáveis inteiras de dimensão q , $Y \subseteq \mathfrak{T}^q$, ou um vetor de variáveis binárias de dimensão q , $Y = \{0, 1\}^q$. Como a função é linear, só faz sentido a existência de um problema de otimização com restrição.

função mista não linear e inteira: $S(x, y): X \times Y \rightarrow \mathfrak{R}$

Problemas de otimização mista não linear e inteira apresentam grandes desafios e dificuldades associados com o caráter combinatorial do domínio discreto (inteiro) junto com as comumente encontradas não convexidades e não linearidades do domínio contínuo.

função unimodal: é a função que apresenta apenas um extremo.

função multimodal: é a função que apresenta mais de um extremo.

NOTA: toda função côncava ou convexa é unimodal, mas nem toda função unimodal é necessariamente côncava ou convexa. Veja exemplo $S(x) = (x^2 - 1)^3$.

10.2.5 Álgebra vetorial

Sejam x e $y \in X \subseteq \mathfrak{R}^n$, A e $B \in \mathfrak{R}^{n \times n}$, $S(x): X \rightarrow \mathfrak{R}$, $g(x)$ e $h(x): X \rightarrow X$. Então as seguintes regras de diferenciação se aplicam:

$$\nabla[g(x)^T h(x)] = \nabla^T g(x) h(x) + \nabla^T h(x) g(x)$$

$$\nabla[S(x) g(x)] = g(x) \nabla^T S(x) + S(x) \nabla g(x)$$

$$\nabla(A x) = A$$

$$\nabla(x^T A) = A$$

$$\nabla(x^T A x) = (A + A^T) x$$

$$\nabla S[g(x)] = \nabla^T g(x) \nabla_g S(g)$$

$$\nabla S\{g[h(x)]\} = \nabla^T h(x) \nabla_h^T g(h) \nabla_g S(g)$$

$$\nabla g[h(x)] = \nabla_h g(h) \nabla h(x)$$

$$\frac{d(x^T y)}{dt} = \dot{x}^T y + x^T \dot{y}$$

$$\frac{d(Ax)}{dt} = \dot{A}x + A\dot{x}$$

$$\frac{d(AB)}{dt} = \dot{A}B + A\dot{B}$$

$$\frac{d(x^T A x)}{dt} = \dot{x}^T A x + x^T \dot{A} x + x^T A \dot{x}$$

Exercícios de fixação

1. Teste as condições necessárias e suficientes do problema abaixo.

$$\min S(x) = x_1 x_2$$

$$\text{sujeito a: } g_1(x) = x_1^2 + x_2^2 - 25 \leq 0$$

2. Determine se as funções abaixo são côncavas, convexas, estritas ou não, ou nenhum destes casos ?

a) $S(x) = 2x_1 + 3x_2 + 6$

b) $S(x) = 2x_1^2 - 3x_1x_2 + 2x_2^2$

c) $S(x) = x_1^3 + x_2^2 - 3x_1 + 8x_2 + 2$

3. Verifique se as restrições abaixo, definindo uma região fechada, formam uma região convexa.

$$g_1(x) = -x_1^2 + x_2 \geq 1 \quad \text{e} \quad g_2(x) = x_2 - x_1 \leq 2$$

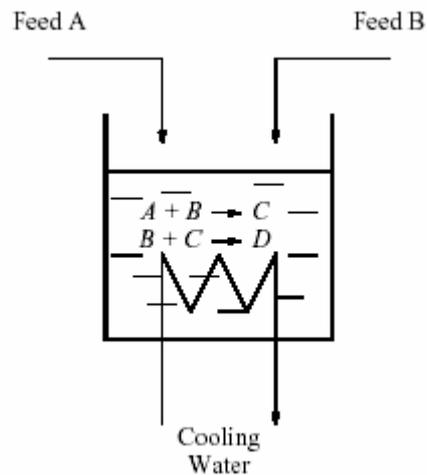
NOTA: se todas as restrições colocadas na forma $g_i(x) \leq 0$ são funções convexas (ou na forma $g_i(x) \geq 0$ são funções côncavas), então elas formam uma região convexa. Funções lineares são tanto convexas quanto côncavas.

4. Mostre que a função $S(x) = -x_1x_2$ com $x_1 \geq 0$ e $x_2 \geq 0$ é quasi-convexa. Ela é também convexa ? Por quê ?

5. Mostre que a função $S(x) = x_1^2 + 5x_1x_2 + 3x_2^2$ com $x_1 > 0$ e $x_2 > 0$ é pseudo-convexa. Ela é também convexa ? Por quê ?

10.3 Otimização Dinâmica

Exemplo 9: Para ilustrar o problema de otimização dinâmica, algumas vezes chamado de otimização em *dimensão infinita*, pois envolve variáveis de decisão que são funções, considere o sistema abaixo de um reator semi-batelada, com duas reações exotérmicas:



onde C é o produto de interesse e D é um sub-produto. Iniciando com o reator vazio, as alimentações de A e B, bem com a taxa de água de resfriamento, estão livres para variar ao longo da operação. Para um dado reator, com dimensões fixas, o objetivo operacional pode ser o de determinar o tempo de duração da operação e as taxas de alimentação de reagentes e água, de modo a maximizar a concentração final de C. O projeto do equipamento também pode ser parte do objetivo do problema. Geralmente, quando as variáveis de decisão envolvem somente variáveis operacionais, o problema é denominado de *controle ótimo*.

A formulação matemática do processo acima pode ser genericamente representada por:

$$F(x(t), \dot{x}(t), y(t), u(t), v) = 0$$

onde $x(t)$, $\dot{x}(t)$ são as variáveis diferenciais e suas derivadas em relação a variável independente, t (geralmente o tempo), $y(t)$ são as variáveis algébricas, $u(t)$ são as variáveis de controle e v são os parâmetros invariantes no tempo, a serem determinados pela otimização. Para o exemplo, $u(t)$ representa as taxas de alimentação dos reagentes e da água de refrigeração e v pode ser o volume do reator. As condições iniciais para este problema podem ser descritas genericamente por:

$$I(x(0), \dot{x}(0), y(0), u(0), v) = 0$$

que juntamente com a forma funcional de $u(t)$ e os valores de v determinam completamente a resposta transiente do sistema.

10.3.1 Formulação da função objetivo e restrições

A formulação da função objetivo para um problema de otimização dinâmica pode estar associada a determinação dos seguintes valores:

- horizonte de tempo da operação, t_f (tempo final);
- parâmetros invariantes no tempo, v ;
- variação temporal das variáveis de controle, $u(t)$, para $t \in [0, t_f]$.

de modo a minimizar (ou maximizar) um funcional $\Phi(u(t), v, t_f)$. A função objetivo é um *funcional* pois depende da escolha da função $u(t)$. No problema do reator semi-batelada Φ seria a concentração final do produto C e t_f seria a duração da batelada, resultando no problema de otimização abaixo:

$$\begin{aligned} \max_{t_f, v, u(t), t \in [0, t_f]} \quad & \Phi = x_3(t_f) \\ \text{sujeito a:} \quad & F(x(t), \dot{x}(t), y(t), u(t), v) = 0 \\ & I(x(0), \dot{x}(0), y(0), u(0), v) = 0 \end{aligned}$$

onde x_3 é a concentração de C.

Se a função objetivo a ser minimizada for a integral de uma função sobre todo o intervalo $[0, t_f]$:

$$\Phi(u(t), v, t_f) = \int_0^{t_f} \varphi(x(t), \dot{x}(t), y(t), u(t), v) dt$$

pode-se adicionar ao sistema de equações algébrico-diferenciais do problema, a seguinte equação:

$$\dot{\Phi} = \varphi(x(t), \dot{x}(t), y(t), u(t), v)$$

com a condição inicial $\Phi(0) = 0$, para ser integrada simultaneamente. Para o caso particular de $\dot{\Phi} = 1$, tem-se o problema de minimização do tempo de operação, t_f .

Uma formulação mais geral para a função objetivo envolve as duas formas acima, ou seja:

$$\Phi(u(t), v, t_f) = \psi(x(t_f), \dot{x}(t_f), y(t_f), u(t_f), v) + \int_0^{t_f} \varphi(x(t), \dot{x}(t), y(t), u(t), v) dt$$

Dentre as mais variadas formas de restrições adicionais em um problema de otimização dinâmica, tem-se:

a) Limites nas variáveis de decisão

$$t_f^{min} \leq t_f \leq t_f^{max}$$

$$u^{\min} \leq u(t) \leq u^{\max}, \quad \forall t \in [0, t_f]$$

$$v^{\min} \leq v \leq v^{\max}$$

b) Restrições terminais

de igualdade: $w(t_f) = w^*$

de desigualdade: $w^{\min} \leq w(t_f) \leq w^{\max}$

onde w representa alguma variável do sistema (x ou y), ou uma relação entre elas. Para o exemplo do reator, a quantidade final de material no reator pode ser fixada ou a temperatura final pode estar limitada em uma dada faixa.

c) Restrições interiores

Aparecem quando algumas variáveis devem estar limitadas em alguns pontos no interior do intervalo de operação, ou seja:

$$w^{\min}(t_I) \leq w(t_I) \leq w^{\max}(t_I)$$

onde $t_I \in [0, t_f)$. Como por exemplo, limitar a curva de aquecimento do reator dentro de uma faixa determinada.

d) Restrições de trajetória

São aquelas restrições que devem ser satisfeitas durante todo o intervalo de operação, ou seja:

$$w^{\min} \leq w(t) \leq w^{\max}, \quad \forall t \in [0, t_f]$$

Por exemplo, no reator semi-batelada pode-se impor que a temperatura da reação não deva ultrapassar um determinado valor para evitar a degradação do produto.

10.3.2 Princípio do Mínimo de Pontryagin

Antes de enunciar o princípio do mínimo de Pontryagin, é necessário introduzir os conceitos do cálculo variacional, que trata da seleção de uma função desconhecida que aparece no integrando de uma integral que deve ser minimizada ou maximizada em função desta seleção.

Considere a seguinte integral:

$$S(x) = \int_{t_1}^{t_2} \varphi[t, x(t), \dot{x}(t)] dt$$

onde $\dot{x}(t)$ é a derivada de $x(t)$ em relação a t . O problema do cálculo variacional consiste na seleção de $x(t)$ de modo a minimizar (ou maximizar) o funcional $S(x)$. Supondo que $x^*(t)$ é a função que minimiza $S(x)$ e $x(t)$ é uma outra função com uma diferença infinitesimal de $x^*(t)$ em cada ponto dentro do intervalo (t_1, t_2) , define-se:

$$\delta x = x(t) - x^*(t)$$

como o operador variação, ou seja, a variação de uma função representa uma mudança infinitesimal arbitrária na função a um dado valor de t . Note que a variação difere da diferenciação, pois esta última corresponde a uma medida da mudança de uma função resultante de uma mudança específica (não arbitrária) na variável independente. A equação acima pode ser escrita como:

$$\delta x = x(t) - x^*(t) = \varepsilon \phi(t)$$

onde $\phi(t)$ é uma função arbitrária contínua e diferenciável e ε é um parâmetro variável que tende a zero.

Como o operador variação acarreta uma mudança infinitesimal em uma função para um valor fixo de t , tem-se:

$$\delta t = 0$$

ou seja, a variável independente não participa do processo de variação. Outra propriedade importante do operador variação é a comutatividade com os operadores de diferenciação e integração:

$$\frac{d}{dt} \delta x = \frac{d}{dt} [\varepsilon \phi(t)] = \varepsilon \frac{d\phi}{dt} \quad \text{e} \quad \delta \frac{dx}{dt} = \frac{dx}{dt} - \frac{dx^*}{dt} = \frac{d}{dt} (x - x^*) = \varepsilon \frac{d\phi}{dt}$$

$$\delta \int_{t_1}^{t_2} x(t) dt = \int_{t_1}^{t_2} x(t) dt - \int_{t_1}^{t_2} x^*(t) dt = \int_{t_1}^{t_2} [x(t) - x^*(t)] dt = \int_{t_1}^{t_2} \delta x(t) dt$$

Como a condição necessária de primeira ordem para o mínimo local de uma função, $S(x)$, pode ser interpretada como a existência de um ponto estacionário da função objetivo, então dentro de uma região infinitesimal em torno deste ponto tem-se que:

$$\delta S = 0$$

Usando a propriedade comutativa chega-se a:

$$\delta S = \int_{t_1}^{t_2} \delta \varphi dt$$

onde $\delta \varphi = \varphi[x, \dot{x}(t), t] - \varphi[x^*, \dot{x}^*(t), t] = \varphi[x^* + \varepsilon \phi, \dot{x}^*(t) + \varepsilon \dot{\phi}(t), t] - \varphi[x^*, \dot{x}^*(t), t]$.

Expandindo φ em série de Taylor tem-se:

$$\varphi[x^* + \varepsilon \phi(t), \dot{x}^*(t) + \varepsilon \dot{\phi}(t), t] = \varphi[x^*, \dot{x}^*(t), t] + \nabla_x^T \varphi[x^*, \dot{x}^*(t), t] \varepsilon \phi(t) + \nabla_{\dot{x}}^T \varphi[x^*, \dot{x}^*(t), t] \varepsilon \dot{\phi}(t) + O[\varepsilon^2 \phi^2(t), \varepsilon^2 \dot{\phi}^2(t)]$$

que eliminando os termos de ordem $O(\varepsilon^2)$, resulta em:

$$\delta \varphi = \varepsilon \{ \nabla_x^T \varphi[x^*, \dot{x}^*(t), t] \phi(t) + \nabla_{\dot{x}}^T \varphi[x^*, \dot{x}^*(t), t] \dot{\phi}(t) \}$$

Para um ponto estacionário tem-se:

$$\delta S = \varepsilon \int_{t_1}^{t_2} \left(\nabla_x^T \varphi \phi + \nabla_{\dot{x}}^T \varphi \dot{\phi} \right) dt = 0$$

sobre qualquer função arbitrária $\phi(t)$. O segundo termo no integrando pode ser integrado por partes ($u = \nabla_x \varphi$ e $v = \phi$), resultando em:

$$\int_{t_1}^{t_2} \nabla_{\dot{x}}^T \varphi \dot{\phi} dt = [\nabla_{\dot{x}}^T \varphi \phi]_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt} [\nabla_{\dot{x}}^T \varphi] \phi dt$$

Se a função $x(t)$ é fixada nos contornos, t_1 e t_2 , então $\phi(t)$ deve se anular nestes pontos, pois não pode existir variação de $x(t)$ nos contornos. Deste modo o primeiro termo da integração por partes é nulo. Para o caso mais geral, onde algum contorno pode estar livre, as seguintes condições de *complementaridade* devem ser satisfeitas:

$$\nabla_{\dot{x}}^T \varphi \phi = 0 \quad \text{para } t = t_1 \text{ e } t = t_2$$

ou seja, se δS é para ser zero para todas as variações admissíveis (restritas somente pela continuidade e diferenciabilidade) então o primeiro termo da integração por partes deve ser sempre nulo. No caso de existir uma contribuição terminal na função objetivo:

$$S(x) = \Psi[x(t_2), t_2] + \int_{t_1}^{t_2} \varphi[t, x(t), \dot{x}(t)] dt$$

a condição de complementaridade terminal seria:

$$[\nabla_x^T \Psi + \nabla_x^T \varphi] \phi = 0 \quad \text{para } t = t_2$$

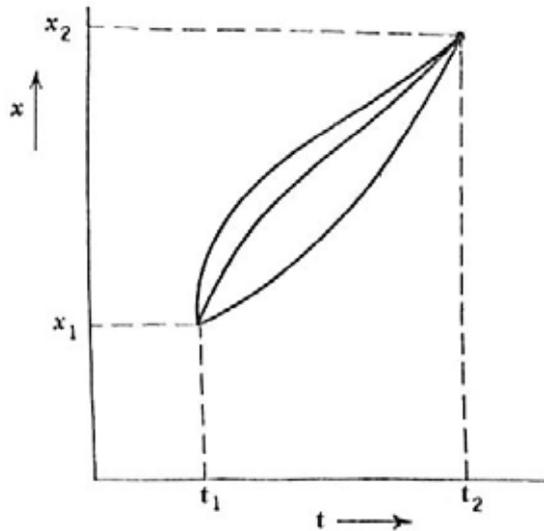
Substituindo a integração por partes em $\delta S = 0$, tem-se então:

$$\delta S = \varepsilon \int_{t_1}^{t_2} \left(\nabla_x^T \varphi - \frac{d}{dt} [\nabla_{\dot{x}}^T \varphi] \right) \phi dt = 0$$

Como a equação acima deve ser satisfeita para qualquer função arbitrária $\phi(t)$, contínua, diferenciável e que satisfaça as condições de contorno, então o termo entre parêntesis deve ser nulo, resultando na equação de *Euler-Lagrange*:

$$\nabla_x \varphi - \frac{d}{dt} [\nabla_{\dot{x}} \varphi] = 0$$

Exemplo 10: Para ilustrar, considere o problema de determinar a curva de menor comprimento conectando dois pontos no espaço, conforme a figura abaixo.



O comprimento de uma curva conectando os pontos (t_1, x_1) e (t_2, x_2) é dado por:

$$S(x) = \int_{s(t_1)}^{s(t_2)} ds = \int_{t_1}^{t_2} \sqrt{1 + \dot{x}^2} dt$$

onde $\varphi[t, x(t), \dot{x}(t)] = \sqrt{1 + \dot{x}^2}$ e $ds = \sqrt{dt^2 + dx^2}$ que é o comprimento de um segmento infinitesimal da curva.

Aplicando a equação de Euler-Lagrange para o problema acima, tem-se:

$$\nabla_x \varphi = \frac{\partial \varphi}{\partial x} = 0$$

$$\nabla_{\dot{x}} \varphi = \frac{\partial \varphi}{\partial \dot{x}} = \frac{\dot{x}}{\sqrt{1 + \dot{x}^2}}$$

$$\nabla_x \varphi - \frac{d}{dt} [\nabla_{\dot{x}} \varphi] = 0 - \frac{d}{dt} \left(\frac{\partial \varphi}{\partial \dot{x}} \right) = - \frac{d}{dt} \left(\frac{\dot{x}}{\sqrt{1 + \dot{x}^2}} \right) = 0$$

ou seja: $\frac{\dot{x}}{\sqrt{1 + \dot{x}^2}} = \text{constante}$, o que implica que $\dot{x}^*(t) = \text{constante}$. Portanto, como não

poderia deixar de ser, o caminho mais curto entre dois pontos é uma reta: $x^*(t) = a t + b$.

Naturalmente, para assegurar que a função $x(t)$ minimize o funcional $S(x)$, a condição suficiente de segunda ordem deve ser verificada, que no caso de variações significa que a segunda variação de S deve ser positiva para todas variações possíveis de x , isto é:

$$\delta^2 S > 0, \quad \forall \delta x$$

que é equivalente a $(\delta x)^T \nabla_x^2 S(x^*) \delta x > 0$, ou ainda pela condição de segunda ordem de Legendre:

$$\nabla_{\dot{x}}^2 \varphi(x^*, \dot{x}^*, t) > 0$$

Aplicando esta condição para o exemplo acima tem-se:

$$\nabla_{\dot{x}}^2 \varphi = \frac{1}{(1 + \dot{x}^2)^{3/2}} > 0$$

logo, a solução ótima $x^*(t)$ minimiza o funcional $S(x)$.

Retornando ao problema de otimização de sistema dinâmicos, seja então o problema:

$$\min_{t_f, x(t), u(t), t \in [0, t_f]} \Phi[x(t), u(t), t_f] = \psi[x(t_f), t_f] + \int_0^{t_f} \varphi[x(t), u(t), t] dt$$

$$\text{sujeito a: } \dot{x}(t) = f[x(t), u(t), t] \quad , \quad x(0) = x_0$$

$$h[(t_f), t_f] = 0$$

Introduzindo os multiplicadores de Lagrange associado às restrições chega-se a:

$$L[x(t), u(t), \lambda(t), \eta, t_f] = \psi[x(t_f), t_f] + \eta^T h[x(t_f), t_f] + \int_0^{t_f} \{ \varphi[x(t), u(t), t] + \lambda^T(t) (f[x(t), u(t), t] - \dot{x}(t)) \} dt$$

que pode ser dividida em duas funções:

$$\Theta[x(t_f), \eta, t_f] = \psi[x(t_f), t_f] + \eta^T h[x(t_f), t_f] \text{ e}$$

$$\Lambda[x(t), u(t), \lambda(t), t] = \varphi[x(t), u(t), t] + \lambda^T(t) (f[x(t), u(t), t] - \dot{x}(t))$$

resultando no funcional:

$$L[x(t), u(t), \lambda(t), \eta, t_f] = \Theta[x(t_f), \eta, t_f] + \int_0^{t_f} \Lambda[x(t), u(t), \lambda(t), t] dt$$

que ao ser minimizado resulta nas equações de Euler-Lagrange:

$$\nabla_x \Lambda - \frac{d}{dt} [\nabla_{\dot{x}} \Lambda] = 0 = \nabla_x \varphi + \nabla_x^T f \lambda + \dot{\lambda}$$

$$\nabla_{\omega} \Lambda - \frac{d}{dt} [\nabla_{\dot{\omega}} \Lambda] = 0 \Rightarrow \nabla_{\dot{\omega}} \Lambda = cte$$

onde $\dot{\omega}(t) = u(t)$, e pelas condições de complementaridade para $\dot{\omega}(t_f)$:

$$[\nabla_{\omega} \Theta + \nabla_{\dot{\omega}} \Lambda]^T \delta \omega_f = 0 \Rightarrow \nabla_{\omega} \Theta + \nabla_{\dot{\omega}} \Lambda = 0 \Rightarrow \nabla_{\dot{\omega}} \Lambda = 0 \quad \text{para } t = t_f$$

pois $\omega(t_f)$ é livre ($\delta \omega_f \neq 0$) e Θ não depende de ω , chega-se a $\nabla_{\dot{\omega}} \Lambda = 0 \quad \forall t$, ou seja:

$$\nabla_{\dot{\omega}} \Lambda = 0 = \nabla_{\dot{\omega}} \varphi + \nabla_{\dot{\omega}}^T f \lambda = \nabla_u \varphi + \nabla_u^T f \lambda$$

Definindo a seguinte *função de Hamilton* (ou *Hamiltoniano*):

$$H[x(t), u(t), \lambda(t), t] = \varphi[x(t), u(t), t] + \lambda^T(t) f[x(t), u(t), t]$$

então, as condições necessárias de primeira ordem (Euler-Lagrange) descritas acima são equivalentes a:

$$\begin{aligned}\dot{x} &= \nabla_{\lambda} H = f \\ \dot{\lambda} &= -\nabla_x H = -\nabla_x \varphi - \nabla_x^T f \lambda \\ \nabla_u H &= \nabla_u \varphi + \nabla_u^T f \lambda = 0\end{aligned}$$

que devem ser resolvidas juntamente com as condições iniciais e finais e as condições de complementaridade:

$$\begin{aligned}x(0) &= x_0 \\ h[x(t_f), t_f] &= 0 \\ [\nabla_{x(t_f)} \Theta - \lambda(t_f)]^T \delta x(t_f) &= 0 \\ \{H[x(t_f), u(t_f), \lambda(t_f), t_f] + \nabla_{t_f} \Theta\} \delta t_f &= 0\end{aligned}$$

A condição de segunda ordem pode ser verificada pela positividade de $\nabla_u^2 H$. Este conjunto de condições são conhecidos como *Princípio do Mínimo de Pontryagin* (1962), e os multiplicadores de Lagrange, $\lambda(t)$, são conhecidos como *variáveis adjuntas*. Para as partes do caminho $u(t)$ sobre as restrições u^{\min} ou u^{\max} , vale a seguinte condição de otimalidade, no lugar de $\nabla_u H = 0$:

$$H[x^*(t), u^*(t), \lambda^*(t), t] \leq H[x^*(t), u(t), \lambda^*(t), t]$$

para todos as possíveis funções $u(t)$.

Para problemas com restrições de desigualdade do tipo:

$$g[x(t), u(t), t] \leq 0, \quad \forall t \in [0, t_f]$$

onde $g \in \mathfrak{R}^p$ é duas vezes continuamente diferenciável e o subconjunto de restrições ativas, g^a , não deve ser maior que o número de variáveis de controle, u , e deve apresentar uma matriz Jacobiana, $\nabla_u g^a$, de posto completo $\forall t \in [0, t_f]$. Exemplos de restrições de desigualdade que satisfazem estas condições são:

$$\begin{aligned}u^{\min} &\leq u(t) \leq u^{\max}, \quad \forall t \in [0, t_f] \\ u^{\min}[x(t), t] &\leq u(t) \leq u^{\max}[x(t), t], \quad \forall t \in [0, t_f] \\ |u(t)| &\leq u^m, \quad \forall t \in [0, t_f] \\ \|u(t)\|^2 &\leq r^2, \quad \forall t \in [0, t_f], \quad r \in \mathfrak{R}\end{aligned}$$

Introduzindo funções de folga $\dot{z}(t)$, isto é:

$$g_i[x(t), u(t), t] + \dot{z}_i^2(t) = 0, \quad i = 1, 2, \dots, p$$

tem-se o integrando da função de Lagrange na forma:

$$\Lambda[x(t),u(t),\lambda(t),t] = \varphi[x(t),u(t),t] + \lambda^T(t) (f[x(t),u(t),t] - \dot{x}(t)) + \mu^T(t) (g[x(t),u(t),t] + \dot{z}^2(t))$$

onde o vetor $\dot{z}^2(t) = [\dots \dot{z}_i^2(t) \dots]^T$. Neste caso as equações de Euler-Lagrange são dadas por:

$$\nabla_x \Lambda - \frac{d}{dt} [\nabla_{\dot{x}} \Lambda] = 0 = \nabla_x \varphi + \nabla_x^T f \lambda + \nabla_x^T g \mu + \dot{\lambda}$$

$$\nabla_{\omega} \Lambda - \frac{d}{dt} [\nabla_{\dot{\omega}} \Lambda] = 0 \Rightarrow \nabla_{\dot{\omega}} \Lambda = cte$$

$$\nabla_z \Lambda - \frac{d}{dt} [\nabla_{\dot{z}} \Lambda] = 0 \Rightarrow \nabla_{\dot{z}} \Lambda = cte$$

De modo análogo às condições de complementaridade para $\omega(t)$, que resultam em $\nabla_{\dot{\omega}} \Lambda = 0 \forall t$, para a função $z(t)$ tem-se:

$$[\nabla_z \Theta + \nabla_z \Lambda]^T \delta z_f = 0 \Rightarrow \nabla_z \Theta + \nabla_z \Lambda = 0 \Rightarrow \nabla_z \Lambda = 0 \quad \text{para } t = t_f$$

logo, $\nabla_z \Lambda = 0 \forall t$ e, pela definição de Λ , $\mu_i \dot{z}_i = 0$, ou ainda $\mu_i g_i = 0, i = 1, 2, \dots, p$.

Definindo o Hamiltoniano para este caso:

$$H[x(t),u(t),\lambda(t),t] = \varphi[x(t),u(t),t] + \lambda^T(t) f[x(t),u(t),t] + \mu^T(t) g[x(t),u(t),t]$$

então o Princípio do Mínimo de Pontryagin (ou condições necessárias de otimalidade para sistemas dinâmicos) é dado por:

$$\dot{x} = \nabla_{\lambda} H = f$$

$$\dot{\lambda} = -\nabla_x H = -\nabla_x \varphi - \nabla_x^T f \lambda - \nabla_x^T g \mu$$

$$\nabla_u H = \nabla_u \varphi + \nabla_u^T f \lambda + \nabla_u^T g \mu = 0$$

$$\mu \geq 0$$

$$\nabla_u^2 H \geq 0 \quad \forall \{ \delta u \mid \nabla_u^T g^a \delta u = 0 \}$$

juntamente com as condições iniciais, finais e de complementaridade:

$$x(0) = x_0$$

$$h[x(t_f), t_f] = 0$$

$$[\nabla_{x(t_f)} \Theta - \lambda(t_f)]^T \delta x(t_f) = 0$$

$$\{ H[x(t_f), u(t_f), \lambda(t_f), t_f] + \nabla_{t_f} \Theta \} \delta t_f = 0$$

$$\mu_i g_i = 0, i = 1, 2, \dots, p.$$

10.3.3 Controle ótimo

Analisando inicialmente o problema de controle ótimo para sistemas lineares:

$$\dot{x}(t) = A(t) x(t) + B(t) u(t) \quad , \quad x(0) = x_0$$

com a seguinte função objetivo quadrática:

$$S[x(t), u(t), t_f] = \frac{1}{2} \|x(t_f)\|_R^2 + \frac{1}{2} \int_0^{t_f} [\|x(t)\|_{Q(t)}^2 + \|u(t)\|_{W(t)}^2] dt$$

onde $\|x\|_R^2 = x^T R x$ e $R \geq 0$, $Q(t) \geq 0$ e $W(t) \geq 0$ são matrizes simétricas de pesos. Definindo o Hamiltoniano para este problema:

$$H[x(t), u(t), \lambda(t), t] = \frac{1}{2} \|x(t)\|_{Q(t)}^2 + \frac{1}{2} \|u(t)\|_{W(t)}^2 + \lambda^T [A x + B u]$$

então, pelo princípio do mínimo de Pontryagin, tem-se:

$$\begin{aligned} \dot{x} &= \nabla_x H = A x + B u \\ \dot{\lambda} &= -\nabla_x H = -Q x - A^T \lambda \\ \nabla_u H &= W u + B^T \lambda = 0 \\ x(0) &= x_0 \\ \lambda(t_f) &= R x(t_f) \end{aligned}$$

que resolvendo para $u(t)$, chega-se a:

$$\begin{aligned} u &= -W^{-1} B^T \lambda \\ \dot{x} &= A x - B W^{-1} B^T \lambda \end{aligned}$$

gerando um problema de TPBVP (*two-point boundary value problem*) para x e λ , cuja solução minimiza a função objetivo, pois $\nabla_u^2 H = W > 0$, e é um mínimo global devido a convexidade do problema. Se a solução $\lambda(t)$ do problema de TPBVP puder ser expressa como uma função da solução $x(t)$ em termos de uma transformação linear $P(t) \in \mathfrak{R}^{n \times n}$, isto é, $x(t)$ ser uma base de soluções para $\lambda(t)$:

$$\lambda(t) = P(t) x(t)$$

então a solução do TPBVP pode ser facilmente obtida. Para verificar tal possibilidade, primeiro substitui-se esta dependência nas equações para \dot{x} e $\dot{\lambda}$, e subtrai-se uma da outra resultando em:

$$(\dot{P} + PA + A^T P - P B W^{-1} B^T P + Q) x(t) = 0$$

que deve ser satisfeita $\forall t$, logo como $x(t) \neq 0$ então:

$$\dot{P} = -PA - A^T P + P B W^{-1} B^T P - Q$$

que é uma equação diferencial matricial não linear, conhecida como equação de *Riccati* dinâmica, cuja condição de contorno é obtida de $\lambda(t_f) = R x(t_f)$, ou seja:

$$P(t_f) = R$$

Fazendo a mudança de variável $\tau = t_f - t$, tem-se $d\tau = -dt$ e:

$$\dot{P} = PA + A^T P - P B W^{-1} B^T P + Q$$

$$P(0) = R$$

Portanto, o problema de TPBVP foi transformado em um problema de valor inicial de dimensão $n(n+1)/2$, pois $P(t)$ é uma matriz simétrica. Com a solução da equação de Riccati dinâmica, a lei de controle ótimo quadrático é dada por:

$$u^* = -K(t) x^*(t)$$

$$K(t) = W(t)^{-1} B(t)^T P(t)$$

que é independente da condição inicial x_0 . O sistema dinâmico em malha fechada pode ser escrito como:

$$\dot{x}^* = (A - B W^{-1} B^T P) x^* = A_K x^*, \quad x(0) = x_0$$

onde $A_K(t) = A - B W^{-1} B^T P$ é a matriz característica da dinâmica do sistema regulado. Observa-se também que o valor ótimo da função objetivo é dado por:

$$S[x^*, u^*, t_f] = \frac{1}{2} x_0^T P(0) x_0$$

onde $P(0)$ é obtido em $t = 0$ ($\tau = t_f$). Este resultado pode ser verificado pela equação:

$$\frac{d}{dt}(x^T P x) = \dot{x}^T P x + x^T \dot{P} x + x^T P \dot{x}$$

que ao substituir as expressões para \dot{P} e \dot{x}^* , resulta em:

$$\frac{d}{dt}(x^T P x) = -x^T Q x - x^T P B W^{-1} B^T P x = -x^T Q x - u^T W u$$

Integrando a equação acima em $[0, t_f]$ tem-se:

$$x(0)^T P(0) x(0) = x(t_f)^T P(t_f) x(t_f) + \int_0^{t_f} (x^T Q x + u^T W u) dt = S[x(t), u(t), t_f]$$

pois $P(t_f) = R$.

Fazendo a integração em $[t, t_f]$ tem-se:

$$x(t)^T P(t) x(t) = x(t_f)^T R x(t_f) + \int_t^{t_f} (x^T Q x + u^T W u) dt \geq 0$$

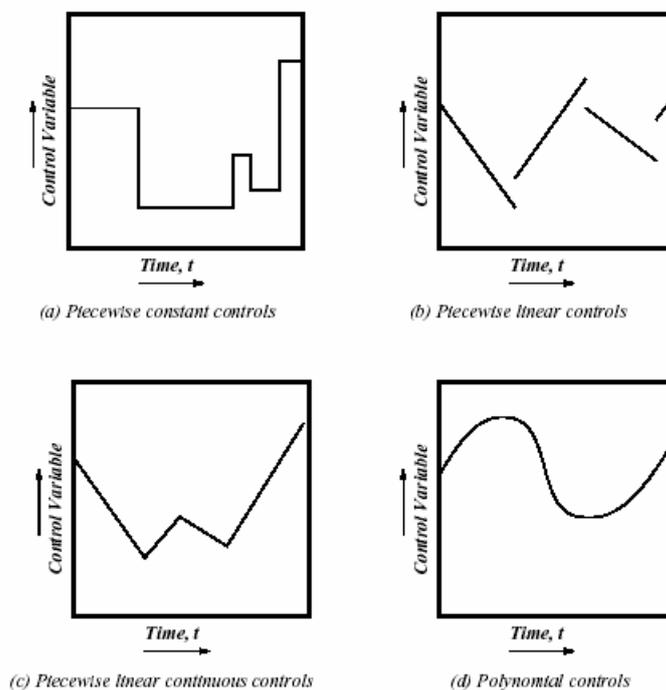
de onde conclui-se que $P(t)$ é positiva semidefinida.

Para o caso particular de um sistema invariante no tempo, com $[A \ B]$ controlável, $t_f = \infty$ e matrizes pesos constantes, a lei de controle ótimo pode ser obtida através da solução da equação de Riccati estacionária:

$$PA + A^T P - P B W^{-1} B^T P + Q = 0$$

que pode ser eficientemente resolvida via decomposição em valores característicos do gradiente do Hamiltoniano associado. Neste caso a matriz de ganhos, K , da lei de controle é invariante no tempo.

Na maioria dos casos de controle ótimo não linear, é mais eficiente tratar o problema como uma programação não linear do que utilizar o princípio do mínimo de Pontryagin, para obter o controle ótimo, $u(t)$. Nestes casos, a escolha da forma das variáveis de controle toma um aspecto mais de engenharia do que matemático, pois dependerá da forma com que as ações de controle serão implementadas na prática. Alguns tipos de variações mais comuns para as variáveis de controle estão ilustrados abaixo.



Sendo as duas primeiras mais freqüentemente utilizadas em controle ótimo de processos.

Para qualquer um dos casos acima, o algoritmo abaixo mostra uma maneira de solucionar o problema de otimização dinâmica via NLP, conhecida como método da tentativa-e-erro ou *single-shooting*.

algoritmo

- 1) Escolher um perfil inicial para as variáveis de controle, $u^0(t)$, $k = 0$.
- 2) Integrar o sistema de equações algébrico diferenciais.
- 3) Calcular a função objetivo e restrições.
- 4) Corrigir as variáveis de controle pelo uso de métodos de NLP.
- 5) Verificar a convergência de $u^k(t)$. Caso não esteja satisfeita $k \leftarrow k + 1$ (ir para 2).
- 6) FIM.

Outra forma de resolver o problema é através da discretização das equações diferenciais (ou método de *multiple-shooting*) por técnicas de elementos finitos, diferenças finitas ou aproximação polinomial, gerando um problema algébrico de programação não linear de elevada dimensão, que também pode ser resolvido por qualquer método de NLP com restrição.

Lista de Exercícios

1) A figura abaixo representa um extrator onde entram uma solução com uma vazão $F = 10^4$ kg H₂O / h e uma concentração $x_0 = 0,02$ kg A / kg H₂O, e um solvente B puro a uma vazão W kg B / h a ser determinada. A corrente refinada sai com uma vazão $F = 10^4$ kg H₂O / h e uma concentração x kg A / kg H₂O e ainda um pouco de solvente B dissolvido em água. No extrato, a vazão é de W kg B / h e a concentração y kg A / kg B. Calcular a vazão ótima W de B.

dados: preço do soluto no extrato: 0,4 \$ / kg A

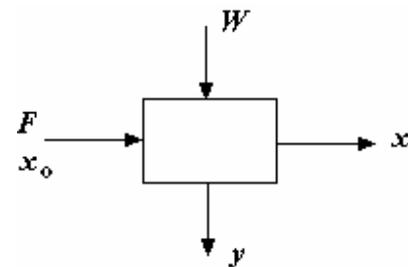
preço do solvente B: 0,01 \$ / kg B

solubilidade de B em água: 0,0007 kg B / kg H₂O

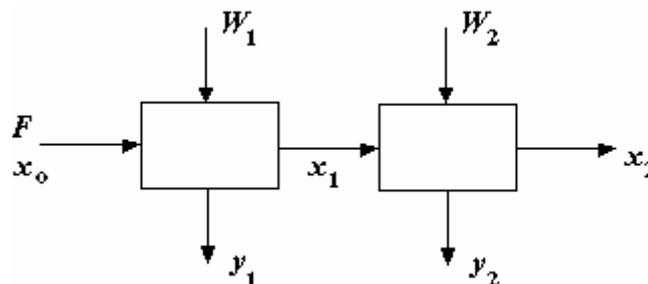
relação de equilíbrio: $y = k x$ ($k = 4$)

o despejo não tem valor comercial.

critério de desempenho: Lucro = \$ soluto extraído - \$ solvente utilizado.



2) Calcular as vazões ótimas de solvente para o exercício anterior quando dois extratores em série são utilizados no lugar de apenas um extrator, conforme figura abaixo. Considere a mesma relação de equilíbrio e a mesma solubilidade de B em água para os dois extratores.



3) A figura abaixo representa um trocador de calor em contra-corrente, onde uma corrente de processo com vazão $W_1 = 1000 \text{ kg/h}$ e temperatura $T_1 = 200 \text{ }^\circ\text{C}$ é resfriada a $T_2 = 100 \text{ }^\circ\text{C}$ por uma corrente fria a $t_2 = 60 \text{ }^\circ\text{C}$. Calcular a vazão ótima W_2 do refrigerante (kg/h) bem como a área de troca térmica ótima A (m^2). Considerar $c_p = 1 \text{ kcal/kg }^\circ\text{C}$ para as duas correntes e $U = 500 \text{ kcal/h m}^2 \text{ }^\circ\text{C}$.

dados: investimento do trocador de calor: $3200 (A/50)^{0,48} \$$

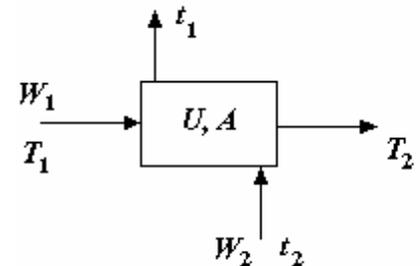
custo do refrigerante: $5 \times 10^{-6} \$ / \text{kg}$

custo de manutenção: 2 % do investimento / ano

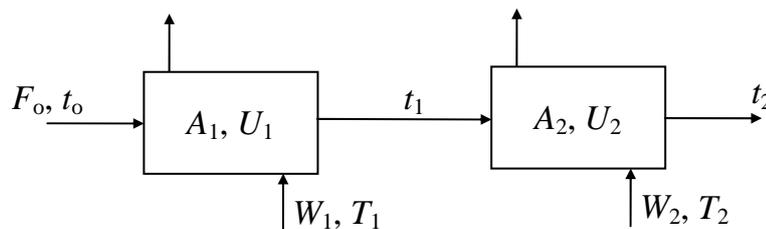
tempo de operação: 8640 h / ano

critério de desempenho: custo anual = 50 % custo

operacional + 10 % ao ano sobre o investimento



4) A figura abaixo representa um sistema de dois trocadores de calor em série, onde uma corrente de processo com vazão $F_o = 1000 \text{ kg/h}$ e temperatura $t_o = 20 \text{ }^\circ\text{C}$ deve ser aquecida a $t_2 = 100 \text{ }^\circ\text{C}$ por correntes de vapor saturado a $T_1 = 100 \text{ }^\circ\text{C}$ e $T_2 = 120 \text{ }^\circ\text{C}$, respectivamente. Calcular as vazões ótimas W_1 e W_2 das correntes de vapor (kg/h) bem como as áreas de troca térmica ótimas A_1 e A_2 (m^2). Considerar $c_p = 1 \text{ kcal/kg }^\circ\text{C}$, $U_1 = 300 \text{ kcal/h m}^2 \text{ }^\circ\text{C}$, $U_2 = 400 \text{ kcal/h m}^2 \text{ }^\circ\text{C}$, $\lambda_1 = 500 \text{ kcal/kg}$ e $\lambda_2 = 600 \text{ kcal/kg}$.



dados: investimento dos trocadores de calor: $160 (A_1)^{0,5} \$$ e $320 (A_2)^{0,5} \$$

custo do vapor: (a $100 \text{ }^\circ\text{C}$) $5 \times 10^{-6} \$ / \text{kg}$ e (a $120 \text{ }^\circ\text{C}$) $1 \times 10^{-5} \$ / \text{kg}$

custo de manutenção: 2 % do investimento / ano

tempo de operação: 8640 h / ano

critério de desempenho:

custo anual = 50 % custo operacional + 10 % ao ano sobre o investimento

5) Um reator CSTR converte um reagente A em um produto B através de uma reação catalítica endotérmica. O catalisador pode ser continuamente injetado no reator e a temperatura do reator pode ser controlada pela injeção de vapor através de uma serpentina. Alguma conversão pode ser obtida somente com o aquecimento, sem a adição de catalisador. Por outro lado, pode-se também obter alguma conversão somente com a adição de catalisador, sem o aquecimento. O produto B pode ser vendido por R\$ 50 / mol de B. O custo do catalisador (R\$ / mol da alimentação) é dado pela expressão:

$$C_c = 2 + 10 X_c + 20 X_c^2$$

onde X_c é a fração molar de catalisador. O custo do vapor (R\$ / mol da alimentação) é dado pela expressão:

$$C_s = 1 + 0,003 S + 2 \times 10^{-6} S^2$$

onde S é a relação de kg de vapor / mol de alimentação. A conversão em termos das quantidades de catalisador e vapor é dada por:

$$X_B = 0,1 + 0,3 X_c + 10^{-4} S - 10^{-4} X_c S$$

onde X_B é a relação de mol de B / mol de alimentação. Deseja-se saber qual o lucro máximo deste processo produtivo por mol de alimentação.

6) Encontre o mínimo global da função

$$f(x) = x_1^2 + x_2^2 - 3x_1 + 15x_2$$

$$\text{sujeito a: } (x_1 + x_2)^2 - 4(x_1 - x_2) = 0$$

Bibliografia Básica

- The Variational Method in Engineering – R. S. Schechter – McGraw-Hill, 1967.
- Optimization: Theory and Practice – G. S. G. Beveridge & R. S. Schechter – McGraw-Hill, 1970.
- Applied Nonlinear Programming – D. M. Himmelblau – McGraw-Hill, 1972.
- Nonlinear Programming – M. S. Bazaraa & C. M. Shetty – John Wiley & Sons, 1979.
- Mathematical Programming. Theory and Algorithms – M. Minoux – John Wiley & Sons, 1986.
- Programação Não-Linear. Introdução à Teoria e aos Métodos – P. Mahey – Editora Campus, 1987.
- Optimization of Chemical Processes – T. F. Edgar & D. M. Himmelblau – McGraw-Hill, 1988.
- Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications – C.A. Floudas – Oxford Press, 1995.
- Optimierung – M. Papageorgiou – Oldenbourg Ed., 1996.
- Systematic Methods of Chemical Process Design – L.T. Biegler, I.E. Grossmann, A.W. Westerberg – Prentice Hall, Inc., 1997.
- Numerical Optimization – J. Nocedal & S. J. Wright – Springer, 1999.
- Numerical Optimization – J. F. Bonnans, J. C. Gilbert, C. Lemaréchal & C. A. Sagastizábal – Springer, 2003.