Universidade Federal de Santa Catarina
Departamento de Engenharia Elétrica

MÉTODOS NUMÉRICOS DE OTIMIZAÇÃO - PARTE I

Profa. Katia Campos de Almeida, Ph.D.
Prof. Roberto Salgado, Ph.D.

Florianópolis - SC
2017.

Apostila feita para a disciplina *Métodos Numéricos Aplicados à Engenharia*, do Curso de Especialização em Sistemas de Energia Elétrica, oferecido pelo Programa de Pós Graduação em Engenharia Elétrica a engenheiros de empresas do setor elétrico.

Sumário

Capítulo 1

Conceitos Preliminares

1.1 Conjuntos Convexos

Definição 1 *Um conjunto C em E^n é dito convexo, se para quaisquer dois pontos $\mathbf{x}_1,\ \mathbf{x}_2 \in C$ e qualquer número real α, com $0 < \alpha < 1$, o ponto $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in C$.*

Esta definição pode ser interpretada da seguinte forma: se dois pontos pertencem a um conjunto convexo, todos os pontos do segmento de reta que une estes dois pontos pertencem também a este conjunto.

As seguintes relações são válidas para conjuntos convexos:

1. Se C é um conjunto convexo e β um número real, o conjunto

$$\beta C = \{\mathbf{x} : \mathbf{x} = \beta\mathbf{c}, \mathbf{c} \in C\}$$

é convexo.

2. Se C e D são convexos, então o conjunto

$$C + D = \{\mathbf{x} : \mathbf{x} = \mathbf{c} + \mathbf{d}, \mathbf{c} \in C, \mathbf{d} \in D\}$$

é convexo.

3. A interseção de vários conjuntos convexos é um conjunto convexo.

Definição 2 *Um conjunto C é um cone se $\mathbf{x} \in C$ implica que $\alpha\mathbf{x} \in C$ para todo $\alpha > 0$.*

Um *cone convexo* é um tipo especial de conjunto convexo.

1.2 Hiperplanos

Sejam $\mathbf{a}$ um vetor coluna não nulo de dimensão n e c um número real. O conjunto

$$H = \{\mathbf{x} \in E^n : \mathbf{a}^t\mathbf{x} = c\}$$

define todos os pontos pertencentes a um hiperplano em E^n.

A representação matemática de um hiperplano é definida pelo vetor que lhe é perpendicular. Para exemplificar esta afirmação, considere o hiperplano definido pela equação

$$\mathbf{a}^t\mathbf{x} = c$$

Se os pontos $\mathbf{x_1}$ e $\mathbf{x_2}$ pertencem ao hiperplano, então

$$\mathbf{a}^t\mathbf{x_1} = c$$

e

$$\mathbf{a}^t\mathbf{x_2} = c$$

A subtração dessas duas equações fornece

$$\mathbf{a}^t\mathbf{x_2} - \mathbf{a}^t\mathbf{x_1} = \mathbf{a}^t(\mathbf{x_2} - \mathbf{x_1}) = c - c = 0$$

Desde que a diferença $(\mathbf{x_2} - \mathbf{x_1})$ define um vetor pertencente ao hiperplano, o produto escalar nulo indica que os vetores $\mathbf{a}$ e $(\mathbf{x_2} - \mathbf{x_1})$ são ortogonais.

1.3 Funções

Uma função de valores reais f definida num subconjunto de E^n é dita contínua em $\mathbf{x}$ se $\mathbf{x}_k \to \mathbf{x}$ implica que $f(\mathbf{x}_k) \to f(\mathbf{x})$.

Um conjunto de funções de valores reais $f_1, f_2, ..., f_m$ em E^n pode ser interpretada como uma única *função vetorial*$\mathbf{f}$, que associa um vetor

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$$

pertencente ao espaço E^m, a todos os vetores $\mathbf{x}$ pertencentes ao espaço E^n. Esta função vetorial é dita contínua se cada uma de suas componentes for contínua.

Definição 3 *Uma função $f : S \to E^1$, onde S é um conjunto convexo em E^n, é denominada* convexa *em S se*

$$f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)$$

para todo $x_1, x_2 \in S$ e todo $\lambda \in (0,1)$.

A função é denominada estritamernte convexa se

$$f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) < \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)$$

Definição 4 *Uma função é côncava se*

$$f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \geq \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)$$

e estritamente côncava se

$$f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) > \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2)$$

.

A interpretação geométrica destas definições é mostrada na figura 1.1. Observe que um ponto $\mathbf{v} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ pertence ao segmento de reta que une $\mathbf{x}_1$ a $\mathbf{x}_2$.

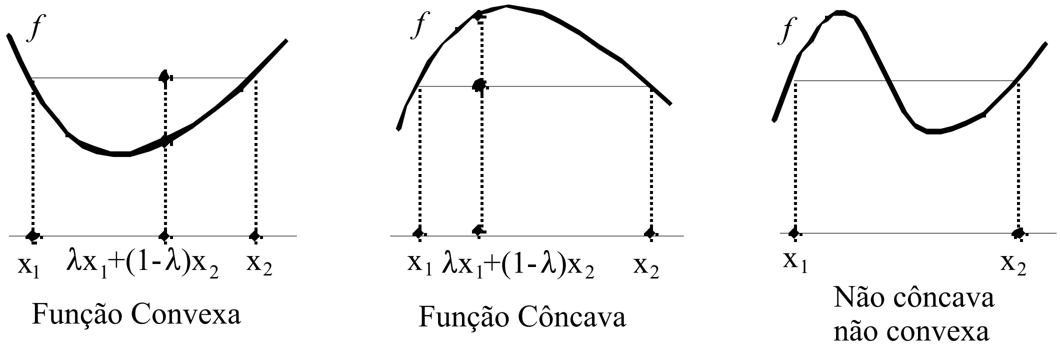

Figura 1.1: Tipos de Funções

Se cada componente de $\mathbf{f} = (f_1, f_2, ..., f_m)$ é contínua num conjunto aberto de E^n, então diz-se que $\mathbf{f} \in C$. Se além disso, cada função f_i tem derivadas parciais de primeira ordem contínuas no mesmo conjunto, diz-se que $\mathbf{f} \in C^1$. Generalizando, se as funções componentes têm derivadas parciais de ordem p contínuas, diz-se que $\mathbf{f} \in C^p$

Seja uma função $f : R^n \to R^1$, duas vezes diferenciável e contínua (isto é, $f \in C^2$). O vetor coluna $(n \times 1)$ de derivadas parciais de f em relação a $\mathbf{x}$, $\mathbf{x} \in R^n$, é uma função de $\mathbf{x}$ denominada vetor gradiente de f e denotada por $\nabla f(\mathbf{x})$ (ou $\nabla_{\mathbf{x}} f(\mathbf{x})$) [1]. Em termos matriciais,

$$\nabla f(\mathbf{x}) = \nabla_x f(\mathbf{x}) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}$$

Definição 5 *Para qualquer função multivariável, a equação $f(\mathbf{x}) = k$ define uma superfície no espaço R^n chamada Curva de Nível ou Curva de Contorno [2].*

Esta curva consiste num conjunto de pontos $\mathbf{x} \in R^n$ onde a função $f : R^n \to R^1$ tem contorno constante. Analiticamente, isto pode ser expresso como,

$$C_n = \{\mathbf{x} \mid f(\mathbf{x}) = k, \quad k = escalar\}$$

Exemplo 1 *: Para $f(x, y) = 8 + (x_1 - 4)^2 + 2(x_2 - 2)^2$ (figura 1.3-a) o gráfico de f é a superfície $z = 8 + (x_1 - 4)^2 + 2(x_2 - 2)^2$. A curva no plano x_1x_2,obtida para $z = 24$ é a elipse $(x_1 - 4)^2 + 2(x_2 - 2)^2 = 16$. Para $z = k$, as curvas de nível de f são as elipses $(x_1 - 4)^2 + 2(x_2 - 2)^2 = k - 8$.*

1.4 Derivada Direcional

Definição 6 *A derivada direcional de uma função de duas vairáveis $f(x, y)$ na direção de um vetor unitário $\mathbf{u} = (cos\theta, sen\theta)$ é definida como*

$$D_u f(x, y) = \lim_{h \to 0} \frac{f(x + hcos\theta, y + hsen\theta) - f(x, y)}{h}$$

se este limite existir.

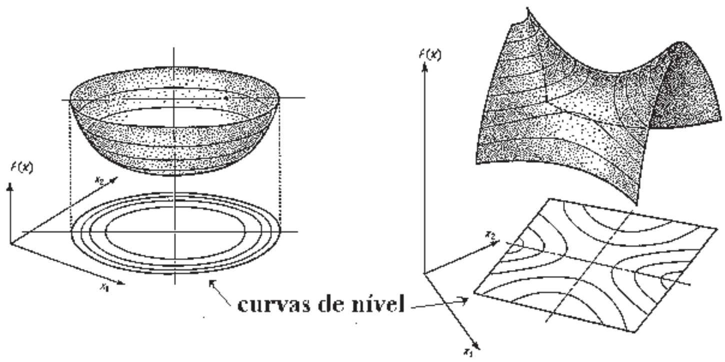

Figura 1.2: Curvas de nível

A derivada direcional fornece a taxa de variação do valor da função f em elaçào à distância no plano xy medida na direção e sentido do vetor $\mathbf{u}$.

A derivada direcional pode ser obtida pelo produto escalar do vetor gradiente da função pelo vetor unitário da direção considerada:

$$D_u f(x, y) = \mathbf{u}^t \nabla f(x, y)$$

Se α é a medida em radianos, do ângulo entre os vetores $\mathbf{u}$ e $\nabla f(x, y)$, então

$$D_u f(x, y) = |\mathbf{u}||\nabla f(x, y)| \cos \alpha$$

A expressão anterior mostra que se $\alpha = 0$ (isto é, se os vetores $\mathbf{u}$ e $\nabla f(x, y)$ são colineares) a função f tem a sua máxima taxa de variação. Portanto, o vetor gradiente indica a direção da máxima taxa de variação de uma função.

A representação do vetor gradiente no plano $f(x, y) = k$ permite verificar que este vetor é perpendicular às curvas de nível da função f.

1.5 Derivadas de Ordem Superior

Derivadas de ordem mais elevada para uma função multivariável são definidas da mesma forma que no caso unidimensional. O número de quantidades associadas a essas derivadas aumenta de um fator n a cada nível de diferenciação. Portanto, a 1^a derivada de uma função $f : R^n \to R^1$ é um vetor n-dimensional; a 2^a derivada da mesma função é definida por n^2 derivadas parciais das n 1^{as} derivadas parciais com relação as n variáveis, ou seja,

$$\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j}) \quad i = 1 \ldots n, \qquad j = 1 \ldots n$$

ou, de uma forma mais geral,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}, \quad para \quad i \neq j \quad e \quad \frac{\partial^2 f}{\partial x_i^2} \quad para \quad i = j$$

Se f é uma função contínua e duas vezes diferenciável, estas n^2 segundas derivadas parciais podem ser representadas por uma matriz quadrada, simética, denominada matriz Hessiana de $f(\mathbf{x})$, denotada por

$$\nabla^2 f(\mathbf{x}) = \nabla^2_{xx} f(\mathbf{x}) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \dots & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \dots & \partial^2 f/\partial x_2 \partial x_n \\ \vdots & \ddots & & \vdots \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \dots & \partial^2 f/\partial x_n^2 \end{bmatrix}$$

Uma vez que

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

pode ser facilmente verificado que a Hessiana é uma matriz simétrica.

Observe que a matriz Hessiana da função escalar $f(\mathbf{x}) : R^n \rightarrow R^1$ é a matriz Jacobiana da função vetorial $\nabla f(\mathbf{x})$.

Se a matriz Hessiana da função f é constante, f é dita *função quadrática*, e neste caso é expressa como [3],

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{G}\mathbf{x} + \mathbf{c}^{\mathbf{T}}\mathbf{x} + \alpha$$

onde, $\mathbf{G}$ é uma matriz constante; $\mathbf{c}$ é um vetor, e α é um escalar.

Note que, o vetor gradiente da função quadrática é

$$\nabla f(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{c}$$

e a sua matriz Hessiana é dada por

$$\nabla^2 f(\mathbf{x}) = \mathbf{G}$$

Para um vetor constituído de funções contínuas e diferenciáveis, as derivadas são definidas diferenciando-se cada função separadamente. Assim, a matriz Jacobiana de um conjunto de funções $\mathbf{f}(\mathbf{x})$, de dimensão $(m \times 1)$, é definida como uma matriz $(m \times n)$, cujo elemento $i - j$ é a derivada de $f_i(\mathbf{x})$ com relação a x_j; ou seja, a i-ésima linha é o vetor gradiente transposto da função $\nabla f_i(\mathbf{x})$. Portanto, se

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{x}) \\ \vdots \\ \mathbf{f}_m(\mathbf{x}) \end{bmatrix}$$

então

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \dots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \dots & \partial f_2/\partial x_n \\ \vdots & & & \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \dots & \partial f_m/\partial x_n \end{bmatrix}$$

Exemplo 2 *: Equações não-lineares da rede elétrica - Matriz Jacobiana do problema de Fluxo de Potência convencional.*

Se a função vetorial é duas vezes diferenciável ($\mathbf{f} \in C^2$), é possível definir m matrizes Hessianas $\nabla^2 f_1(\mathbf{x}), \nabla^2 f_2(\mathbf{x}), \ldots, \nabla^2 f_m(\mathbf{x})$ correspondentes às m funções que compõem $\mathbf{f}$. A segunda derivada de $\mathbf{f}$ é um tensor de 3^aordem, porém seu uso explícito não é freqüente.

Dado um vetor

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix}$$

a função $\lambda^t \mathbf{f}$ tem gradiente igual a $\lambda^t \nabla \mathbf{f}(\mathbf{x})$ e Hessiana

$$\lambda^t \nabla^2 \mathbf{f}(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\mathbf{x})$$

1.6 Expansão em Série de Taylor

Se a função $f(x)$ é univariável, contínua e r vezes diferenciável, então existe um escalar θ $(0 \leq \theta \leq 1.0)$, tal que [3]

$$f(x+h) = f(x) + h\frac{df(x)}{dx} + \frac{1}{2}h^2\frac{d^2f(x)}{dx^2} \cdots \frac{1}{(r-1)!}h^{r-1}\frac{d^{r-1}f(x)}{dx^{r-1}}\bigg|_{x+\theta h} + \frac{1}{r!}h^r\frac{d^rf(x)}{dx^r}\bigg|_{x+\theta h}$$

onde $\frac{d^rf(x)}{dx^r}$ denota a r-ésima derivada de f em relação a x, calculada no ponto x.

Caso multivariável: Sejam $\mathbf{x}$ um ponto, $\mathbf{p}$ um vetor de comprimento unitário e α um escalar. A função $f(\mathbf{x} + \alpha\mathbf{p})$ pode ser considerada como uma função unidimensional de α, e a expansão acima pode ser aplicada diretamente. Na maioria dos casos práticos, apenas a expressão até os três primeiros termos é importante. Assim,

$$f(\mathbf{x} + \alpha\mathbf{p}) = f(\mathbf{x}) + \alpha\nabla f(\mathbf{x})^T\mathbf{p} + \frac{1}{2}\alpha^2\mathbf{p}^T\nabla^2 f(\mathbf{x})\mathbf{p}$$

Observação 1 *: A taxa de variação de f no ponto $\mathbf{x}$ ao longo da direção $\mathbf{p}$ é dada pela quantidade $\nabla f(\mathbf{x})^T\mathbf{p}$, a qual é denominada derivada direcional.*

Observação 2 *: De forma semelhante ao da observação anterior, o escalar $\mathbf{p}^T\nabla^2 f(\mathbf{x})\mathbf{p}$ pode ser interpretado como 2^a derivada de f ao longo de $\mathbf{p}$, e é comumente conhecido como curvatura de f ao longo de $\mathbf{p}$.*

Observação 3 *: A direção $\mathbf{p}$ tal que $\mathbf{p}^T\nabla^2 f(\mathbf{x})\mathbf{p} > 0 (< 0))$ é chamada direção de curvatura positiva (negativa).*

Observação 4 *: A expansão em série de Taylor de uma função geral f em torno de um ponto $\mathbf{x}$ permite determinar aproximações simples da função na vizinhança de $\mathbf{x}$. Assim,*

$$f(\mathbf{x} + \mathbf{p}) \cong f(\mathbf{x}) + \nabla f(\mathbf{x})^T\mathbf{p}$$

define uma função linear do vetor n-dimensional $\mathbf{p}$ (distante $||p||$ de $\mathbf{x}$) e é uma aproximação de f com erro da ordem $||p||^2$.

Observação 5 *: De maneira análoga à da observação prévia,*

$$f(\mathbf{x}+\mathbf{p}) \cong f(\mathbf{x}) + \nabla f(\mathbf{x})^T\mathbf{p} + \frac{1}{2}\mathbf{p}^T\nabla^2 f(\mathbf{x})\mathbf{p}$$

define uma aproximação quadrática de f com erro da ordem $||p||^3$.

Exemplo 3 *: Expansão da função $f(x) = x^2 - 6x + 8$ em série de Taylor.*

1.7 Formulação do Problema de Otimização

O conceito de otimização está implícito no método de resolução de muitos problemas complexos. O uso da filosofia da otimização permite analisar um problema de decisão envolvendo a seleção de valores para um dado número de variáveis inter-relacionadas. Esta análise é realizada concentrando-se a atenção num único objetivo, formulado para quantificar o desempenho e medir a qualidade da decisão. Este objetivo é maximizado ou minimizado, dependendo da formulação, sujeito às restrições que podem limitar a seleção dos valores das variáveis envolvidas. Se um aspecto único de um problema pode ser isolado e caracterizado como um objetivo (por exemplo, lucro ou prejuízo num negócio, velocidade ou distância num problema físico, etc), a otimização pode fornecer uma metodologia de análise consistente.

Deve-se notar entretanto, que freqüentemente é não é fácil representar todas as complexidades das interações das variáveis, restrições e selecionar índices de desempenho apropriados quando se resolve um problema de decisão real. Sendo assim, uma formulação particular deve ser entendida apenas como uma aproximação do problema prático. Observe-se que uma modelagem satisfatória e uma interpretação consistente dos resultados fornecidos são sempre necessárias para se obter conclusões coerentes.

Portanto, problemas de otimização podem ser vistos como formulações matemáticas de problemas de decisão, onde o objetivo a ser alcançado é tratado de forma distinta das interações entre as variáveis e as restrições que limitam a escolha das variáveis envolvidas.

Um problema de otimização estática é em geral constituído de três elementos básicos:

1. um conjunto de *variáveis independentes*, sobre as quais a otimização é executada;
2. um conjunto de condições, chamadas *restrições*, que definem valores aceitáveis das variáveis;
3. um índice a ser otimizado, o qual representa uma medida de desempenho e é chamado convencionalmente de *função objetivo*.

A solução de um problema desse tipo é um conjunto de valores possíveis das variáveis, para os quais a função objetivo assume valor *ótimo*.

Em termos matemáticos, a otimização em geral envolve a *maximização* ou a *minimização* do índice de desempenho, sujeito a um conjunto de restrições. O problema é analiticamente expresso como [1],

$$\begin{array}{lll} \text{M}inimizar & f(\mathbf{x}), & \mathbf{x} \in E^n \;\; (ou \;\; R^n) \\ sujeito\ a & C_i(\mathbf{x}) = 0 & i = 1 \dots m \\ & C_i(\mathbf{x}) \geq 0 & i = (m+1) \dots l \end{array}$$

onde $\mathbf{x}$ é o conjunto de variáveis sobre as quais a otimização é efetuada; E^n (R^n) é o espaço Euclideano n-dimensional; $f(.)$ é o índice de desempenho a ser otimizado; e, $C_i(.)$ é o conjunto de restrições a serem satisfeitas. Este conjunto de restrições define o *conjunto viável* ou *conjunto factível* do problema, denotado S, e expresso como

$$S = \{\mathbf{x} : C_i(\mathbf{x}) = 0 \; para \; i = 1 \ldots m, \quad e \;\; C_i(\mathbf{x}) \geq 0 \; para \; i = (m+1) \ldots l\}$$

Observação 6 *: Deve-se notar que minimizar uma função qualquer é equivalente a maximizar o negativo desta função; isto é,*

$$\begin{array}{lcl} Minimizar \;\; f(\mathbf{x}) & = & Maximizar - f(\mathbf{x}) \\ \mathbf{x} \in S & & \mathbf{x} \in S \end{array}$$

Exemplo 4 *: Minimização das perdas de potência ativa nas linhas de transmissão de um sistema de potência - Neste caso, o índice a ser otimizado é relativo às perdas de potência ativa nas linhas de transmissão, as quais para uma geração de potência ativa pré-especificada, variam com a magnitude da tensão nas barras e com o tap dos transformadores com comutação sob carga. As seguintes restrições devem ser satisfeitas:*

- *equações da rede elétrica (balaços de potência ativa e reativa) satisfeitas;*
- *magnitude das tensões nodais dentro dos limites;*
- *fluxos de potência dentro dos limites;*
- *potências reativas geradas dentro dos limites;*

É importante observar, que a solução de um problema de otimização não necessariamente existe. E se ela existe, não necessariamente tem que ser única. São inúmeros os casos em que mais do que uma solução pode levar ao mesmo valor mínimo do índice a ser otimizado (a minimização das perdas nas linhas de transmissão, por exemplo).

1.8 Classificação dos Problemas de Otimização

A grande maioria dos problemas de otimização pode ser expressa na forma mostrada anteriormente. Mesmo aqueles que não se enquadram no modelo apresentado, podem geralmente ser expressos como uma seqüência de problemas padrões. Entretanto, a existência de uma representação geral não implica em que todas as peculiaridades de cada um devam ser ignoradas. É necessário conhecer a priori as características particulares de cada problema a fim de que a sua solução seja determinada de forma eficiente. Por exemplo, pode ser conveniente omitir testes relativos a situações que não ocorrem, ou evitar a repetição de cálculos de quantidades que não variam etc. Assim, uma classificação é necessária visando aumentar a conveniência na escolha do método de solução.

As principais diferenças em relação aos tipos de problemas estão vinculadas às características matemáticas da função objetivo e das restrições; isto é, linear ou não linear, contínua ou discreta e assim por diante. Outras diferenças são relativas a fatores tais como: dimensão do problema no qual o método de otimização é aplicado, precisão requerida etc. A lista a seguir mostra uma classificação típica dos problemas de otimização, baseada na natureza das funções analíticas envolvidas no problema.

1. Função objetivo linear, restrições lineares;
2. Função objetivo não-linear, restrições lineares;
3. Função objetivo não-linear, restrições não-lineares;
4. Função objetivo linear, restrições não-lineares;

Os problemas de otimização estática em sistemas elétricos operando em regime permanente em geral são enquadrados num destes quatro ítens.

Quanto à técnica de solução de cada tipo de problema, os aspectos listados a seguir devem ser considerados.

Observação 7 *: Problemas do tipo* 1 *são usualmente resolvidos por técnicas de Programação Linear.*

Observação 8 *: A solução dos problemas do tipo* 2 *é freqüentemente determinada levando-se em conta as características da função objetivo. Se esta é não-linear e possui uma forma analiticamente bem definida, então algoritmos específicos para problemas com função objetivo não-linear e restrições lineares são utilizadas. No caso de funções objetivo quadráticas, por exemplo, técnicas de Programação Quadrática podem ser utilizadas. Se por outro lado o índice de desempenho é representado por uma função segmentada linear, técnicas que levam em consideração funções separáveis devem ser aplicadas. Características funcionais tais como convexidade podem simplificar a aplicação da técnica e adicionalmente acelerar a convergência a solução ótima. Os métodos de Programação Não Linear que tratam com restrições lineares são indicados para a solução de problemas desse tipo.*

Observação 9 *: Problemas do tipo* 3 *são certamente os mais difíceis de serem resolvidos. Tipicamente, a menos que as funções envolvidas possuam características que permitam um tratamento em que a sua forma primitiva seja levada em consideração, a solução de tais problemas é encontrada resolvendo-se uma seqüência de problemas simplificados que levam a solução do problema original. Os métodos de Newton e Pontos Interiores têm sido recomendados para a solução desse tipo de problema.*

Observação 10 *:*

A solução de problemas do tipo 4 *é quase tão difícil de ser obtida quanto a dos problemas do tipo* 3*, em virtude da não-linearidade do freqüentemente grande número de restrições envolvidas. Em geral, o Método Simplex Convexo é aplicado para solução de problemas desta espécie.*

Um tipo específico de problema é definido quando algumas (ou todas) variáveis de decisão só devem assumir valores discretos. Este tipo de problema pertence à classe incluída na *Programação Inteira* ou *Inteira-Mista* (dependendo se todas as suas variáveis são discretas ou se possui variáveis *discretas* e variáveis *contínuas*). Dependendo da forma da função objetivo e das restrições estes problemas podem ser classificados como lineares ou não-lineares. A existência de variáveis discretas aumenta muito a dificuldade de resolução de problemas de otimização. Existem técnicas específicas para a resolução desses problemas, a maioria das quais fornece apenas soluções aproximadas. Entre estas técnicas podem ser citadas: Método da Decomposição de Givens, Método de

Branch and Bound, Método da Relaxação Lagrangeana e as chamadas Metaheurísticas (Algoritmos Genéticos, Recozimento Simulado, Busca Tabu, etc.).

Um problema de otimização cuja função objetivo e cujo conjunto viável sejam convexos é denominado um problema convexo.[1] Em geral, problemas convexos são muito mais simples de resolver do que problemas não convexos. Isto porque muitos métodos de resolução se baseiam em aproximações quadráticas da função objetivo e/ou das restrições. Tais aproximações são razoáveis para funções convexas (pelo menos localmente). Além disso, funções convexas possuem um ponto de mínimo único, o que facilita a resolução por métodos numéricos.

1.9 Exercícios

1. Verifique se são convexos ou não convexos os conjuntos viáveis definidos abaixo:

$$
\begin{aligned}
S_1 &= \{x_1, x_2 \mid x_1^2 + x_2^2 \leq 4, x_1 + x_2 \leq 2\} \\
S_1 &= \{x_1, x_2 \mid x_1^3 + x_2 \geq 8, -2 \leq x_1 \leq 3, -20 \leq x_2 \leq 20\} \\
S_1 &= \{x_1, x_2 \mid x_1^2 + x_2^2 = 4, x_1 + x_2 \geq 2\}
\end{aligned}
$$

2. Verifique se são côncavas ou convexas as funções abaixo:

$$
\begin{aligned}
f(x_1, x_2) &= x_1^2 + 2x_1x_2 - 10x_1 + 5x2 \\
f(x_1, x_2) &= x_1 e^{-(x_1 + x_2)}
\end{aligned}
$$

3. Funções separáveis são aquelas que podem ser expressas na forma

$$\psi(\mathbf{x}) = \sum_{i=1}^{n} \psi_i(\mathbf{x})$$

(Por exemplo $f(\mathbf{x}) = x_1^2 + x_2^2$ é separável). Mostre que se os termos em uma função separável são convexos, a própria função é convexa.

4. Diga se as seguintes funções objetivo são separáveis e explique porque:

   - Custo total de geração de potência ativa do conjunto de usinas térmicas de um sistema de energia elétrica;
   - Perdas totais na transmissão de potência ativa em um sistema de energia elétrica.

5. Esboce as curvas de nível da função abaixo e represente seu vetor gradiente.

$$f(x_1, x_2) = x_1^2 + 2x_2^2 - 2x_1 - 12x_2 + 20$$

6. Encontre a direção a partir do ponto $[1 \;\; 3]^T$ para a qual o valor de f não muda se

$$f(x, y) = e^{2y} \text{arc tg}(y/3x)$$

[1]Note que um problema com função objetivo e restrições lineares é um problema convexo.

7. Encontre o gradiente e a hessiana da função

$$f(x_1, x_2) = x_1^2 + 4x_2^2 - 2x_1 - 32x_2 + 64$$

e represente f na forma $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{G}\mathbf{x} + \mathbf{c}^{\mathbf{T}}\mathbf{x} + \alpha$. Mostre, usando a série de Taylor, que um ponto (x_1^0, x_2^0) é um ponto de mínimo se $\nabla f(x_1^0, x_2^0) = 0$ e que $\mathbf{G}$ é definida positiva neste ponto.

8. Usando a série de Taylor, aproxime a função $f(\mathbf{x}) = x_1 e^{x_2} + x_1 x_2$ no ponto $\mathbf{x} = [1 \quad 1]^T$ por uma função quadrática.

9. Encontre o Jacobiano da seguinte função vetorial:

$$\begin{bmatrix} f_1(x_1, x_2, x_3) \\ f_2(x_1, x_2, x_3) \end{bmatrix} = \begin{bmatrix} 3x_1^4 + 4x_2^3 x_3 + 2x_1^2 x_2 + 6x_2 x_3^2 + 10x_1 x_3 + 3x_2 + 4 \\ 4x_1^3 + 4x_2^2 + 2x_3^3 + x_2^2 x_3 + 2x_1 x_2 + 6x_2 + 10 \end{bmatrix}$$

Além disso, encontre a função $\lambda^t \nabla^2 \mathbf{f}(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(\mathbf{x})$, onde $\lambda = [1 \ \ 2]^T$.

10. Quantas variáveis independentes existem no problema de otimização abaixo?

$$\begin{array}{lrl} Min \ \ f(\mathbf{x}) = & \multicolumn{2}{l}{\frac{13}{2}x_1^2 + \frac{5}{2}x_2^2 + \frac{1}{2}x_3^2 - 4x_1x_2 + 3x_1x_3 - 2x_2x_3} \\ & \multicolumn{2}{l}{+y_4^2 + y_4y_5 + y_5^2 + 3y_4 - 2y_5} \\ \text{sujeito a} & y_1 \equiv & 13x_1 - 4x_2 + 3x_3 + y_4 + 3y_5 \geq 0 \\ & y_2 \equiv & -4x_1 + 5x_2 - 2x_3 + y_5 - 1 \geq 0 \\ & y_3 \equiv & 3x_1 - 2x_2 + x_3 - 2y_4 - 2 \geq 0 \\ & y_4 \geq & 0 \\ & y_5 \geq & 0 \end{array}$$

11. Para o problema

$$\begin{array}{lrl} Min \ \ f(\mathbf{x}) = & \multicolumn{2}{l}{\frac{1}{4}x_1^4 + \frac{1}{2}x_1^2 - x_2} \\ \\ \text{sujeito a} & g_1(\mathbf{x}) = & x_1^2 + x_2^2 = 0 \\ & g_2(\mathbf{x}) = & x_1 - x_2 \leq 2 \\ & g_3(\mathbf{x}) = & 3x_1 - 2x_2 + x_3 - 2y_4 - 2 \geq 0 \end{array}$$

Responda às seguintes perguntas:

- O problema é convexo?
- É o ponto $\mathbf{x} = [1 \ \ 1]^T$ um ponto viável?
- É o ponto $\mathbf{x} = [2 \ \ 2]^T$ um ponto interior?
- $f(\mathbf{x})$ possui apenas um ponto de mínimo (é unimodal)?

12. Deseja-se operar um sistema de potência constituído por uma usina térmica e uma usina hidrelétrica de forma a minimizar o custo total de geração da térmica

num intervalo de 4 meses. A usina térmica tem capacidade máxima de geração igual a 1000 MW e a função que representa seu custo de geração é

$$c(pt) = 0.0001Pt^2 + 0.02.Pt \quad \$/h$$

sendo Pt a potência gerada pela usina expressa em MW. A potência gerada pela usina hidrelétrica depende do volume turbinado, sendo esta dependência expressa pela equação

$$Ph = k[hv(\cdot) - hq(\cdot)]q$$

sendo

$$hv(v) = 735.24 + 0.00349v - 1.9743 \times 10^{-7}v^2$$
$$hq = 671.63 + 0.001017(q+s) - 1.7997 \times 10^{-7}(q+s)^2 + 2.5132 \times 10^{-11}(q+s)^3$$

onde Ph é a potência gerada (em MW), q é a vazão turbinado (em hm^3/h), s é a vazão vertida (em hm^3/h), v é o volume do reservatório (hm^3) e $k = 245,22 \times 10^{-4} \frac{MW}{m.hm^3/h}$. A potência gerada pela usina hidrelétrica é no máximo igual a $1272W$ e a vazão turbinada máxima é igual a $5.48hm^3/h$. O balanço de água no reservatório é descrito da seguinte forma: a cada mês j, *entra* no reservatório a água da vazão afluente r_j, (hm^3/mês) e *sai* do reservatório a água turbinada q_j (hm^3/mês) e a água vertida s_j (hm^3/mês). Portanto, o volume do reservatório no mês $j+1$, depende do volume no mês j da seguinte forma

$$v_{j+1} = v_j + horas(r_j - q_j - s_j)$$

onde $horas = 720$ é o número de horas em 1 mês. O volume inicial do reservatório no início do intervalo é de $10000hm^3$ e ao final do período de 4 meses este volume deve maior ou igual a $11000hm^3$. Durante os 4 meses da programação, o volume do reservatório deve estar entre $5733hm^3$ e $22950hm^3$. A vazão afluente e a demanda total a ser atendida em cada hora são dadas na tabela abaixo

Mês	1	2	3	4
$r(hm^3/h)$	11.16	13,14	8.352	19.19
$Pd(MW)$	2000	1800	1200	1600

Formule o problema de otimização a ser resolvido.

13. Classifique o problema abaixo como linear ou não-linear, convexo ou não convexo. Faça uma representação gráfica das curvas de nível da função objetivo e das restrições no plano (x_1, x_2). Encontre a solução de *mínimo irrestrito* da função objetivo (isto é, aquela que não respeita as restrições impostas ao problema) e indique no gráfico a solução de *mínimo restrito* (isto é, aquela que possui menor custo *e* que respeite as restrições impostas). O conjunto viável é convexo?

$$\begin{aligned} Min \;\; f(\mathbf{x}) = \;& x_1^2 + x_2^2 - 4x_1 + 4 \\ \text{sujeito a} \qquad g_1(\mathbf{x}) = \;& x_1 - x_2 + 2 \geq 0 \\ g_2(\mathbf{x}) = \;& x_1^2 + x_2 - 1 \geq 0 \\ g_3(\mathbf{x}) = \;& x_1 \geq 0 \\ g_4(\mathbf{x}) = \;& x_2 \geq 0 \end{aligned}$$

14. É o problema abaixo um problema convexo? Por quê?

$$\begin{array}{lll} Min \;\; f(\mathbf{x}) = & 100x_1 + \frac{200}{x_1x_2} & \\ \text{sujeito a} & g_1(\mathbf{x}) = & 2x_2 + \frac{300}{x_1x_2} \leq 1 \\ & g_2(\mathbf{x}) = & x_1 \geq 0 \\ & g_3(\mathbf{x}) = & x_2 \geq 0 \end{array}$$

15. Plote a função objetivo e as restrições do problema abaixo nos eixos coordenados (x_1, x_2). Indique no gráfico o conjunto de soluções viáveis e a solução ótima do problema. Como se classifica este problema? Ele é convexo?

$$\begin{array}{lll} Min \;\; f(\mathbf{x}) = & 2x_1 - x_2 & \\ \text{sujeito a} & g_1(\mathbf{x}) = & 5x_1 + 7x_2 \leq 45 \\ & g_2(\mathbf{x}) = & -2x_1 + x_2 \leq 1 \\ & g_3(\mathbf{x}) = & 2x - 5x_2 \leq 5 \\ & g_4(\mathbf{x}) = & x_1 \geq 0, \;\; x_1 \; inteiro \\ & g_5(\mathbf{x}) = & x_2 \geq 0 \end{array}$$

Capítulo 2

Otimização Irrestrita

São analisados aqui os problemas de otimização cujas soluções não precisam respeitar restrição alguma. Embora sejam problemas pouco encontrados na prática, uma vez que há sempre limitações nos sistemas físicos, seu estudo é importante pois fornece uma base para as técnicas de resolução de problemas mais complexos. O capítulo, em primeiro lugar, mostra as condições matemáticas que as soluções dos problemas irrestritos devem satisfazer e, em segundo lugar, apresenta alguns dos algoritmos de resolução mais utilizados.

2.1 Introdução

Neste capítulo serão estudados problemas com a forma

$$\begin{array}{ll} Min & f(\mathbf{x}) \\ \text{sujeito a} & \\ & \mathbf{x} \in \Omega \end{array} \tag{2.1}$$

onde f é uma função com valores reais e Ω (conjunto viável) é um subconjunto de E^n.

O objetivo é caracterizar em termos matemáticos, a solução de 2.1. Feito isso, discutiremos métodos de busca unidimensional e de otimização irrestrita.

A primeira pergunta a ser feita a respeito da solução do problema representado pela equação 2.1, é se essa existe. Na realidade, algumas condições devem ser satisfeitas para que o valor ótimo do índice de desempenho sujeito às condições indicadas exista. Em termos matemáticos essas condições são:

- f deve ser uma função contínua;
- Ω deve ser um conjunto e *limitado* e *compacto* (isto é, *fechado*).

Um conjunto é *fechado* se todo ponto arbitrariamente próximo ao conjunto pertence ao conjunto.

Exemplo 5 *O conjunto* $\{\mathbf{x} \mid |\mathbf{x}| \leq 1\}$ *é fechado, enquanto que o conjunto* $\{\mathbf{x} \mid |\mathbf{x}| < 1\}$ *não é fechado.*

Supondo que a solução ótima existe, derivamos abaixo as condições que caracterizam tal solução.

2.2 Condições de Otimalidade de Primeira Ordem

Existem dois tipos de solução para o problema 2.1: pontos de mínimo local e pontos de mínimo global.

Definição 7 *Um ponto $\mathbf{x}^*$ é dito de* mínimo local *ou de* mínimo relativo *de f na região definida pelo conjunto Ω, se este ponto pertence à região; isto é, $\mathbf{x}^* \in \Omega$, e se para todo ponto arbitrário $\mathbf{x} \in \Omega$, situado à uma distância $\xi > 0$ de $\mathbf{x}^*$, a condição*

$$f(\mathbf{x}) \geq f(\mathbf{x}^*)$$

é satisfeita.

A função $f(\mathbf{x})$ tem um mínimo local forte *ou tem um* mínimo relativo forte *em $\mathbf{x}^*$, se $\mathbf{x}^* \in \Omega$ e se existe uma vizinhança de $\mathbf{x}$ também em Ω tal que $f(\mathbf{x}) > f(\mathbf{x}^*)$ em todos os pontos $\mathbf{x}$ desta vizinhança, ou seja,*

$$\exists \;\; \xi > 0 \mid f(\mathbf{x}) > f(\mathbf{x}^*), \forall \mathbf{x} \in X$$

com $||\mathbf{x} - \mathbf{x}^*|| < \xi$.

Definição 8 *Um ponto $\mathbf{x}^*$ é dito de* mínimo global *de f em Ω se para todo $\mathbf{x} \in \Omega$, a condição*

$$f(\mathbf{x}) \geq f(\mathbf{x}^*)$$

é satisfeita. Se $f(\mathbf{x}) > f(\mathbf{x}^*)$ para todo $\mathbf{x} \in \Omega$, então $\mathbf{x}^{ast}$ é dito um mínimo global forte *de f em Ω*

Devido às limitações da maioria dos métodos numéricos, o mínimo global de um problema poderá não ser obtido nos casos onde f não é convexa.

Para derivar as condições que devem ser satisfeitas para que um ponto $\mathbf{x}$ seja um mínimo local $\mathbf{x}^*$, a idéia básica consiste em analisar os movimentos à partir do ponto $\mathbf{x}^*$, numa direção arbitrária.

Definição 9 *: Dado um ponto $\mathbf{x} \in \Omega$, um vetor $\mathbf{d}$ é uma direção viável em $\mathbf{x}$ se existe um $\overline{\alpha}$ tal que $\mathbf{x} + \alpha\mathbf{d} \in \Omega$ para todo α, com $0 \leq \alpha \leq \overline{\alpha}$.*

Teorema 1 *: Seja Ω um subconjunto de E^n e $f \in C^1$ uma função em Ω. Se $\mathbf{x}^*$ é um ponto de mínimo relativo de f em Ω, então para qualquer direção $\mathbf{d} \in E^n$ viável em $\mathbf{x}^*$, tem-se $\nabla f(\mathbf{x}^*)^T\mathbf{d} \geq 0$.*

Deve ser observado que, se $\mathbf{x}^*$ é interior a região definida pelo conjunto Ω, existem direções viáveis a partir de $\mathbf{x}^*$, tal que para todas as direções

$$\nabla f(\mathbf{x}^*)^T\mathbf{d} \geq 0$$

o que implica que $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Corolário 1 *: Seja Ω um subconjunto de E^n e $f \in C^1$ uma função em Ω. Se x^* é um ponto de mínimo relativo de f em Ω e x^* é interior a Ω, então $\nabla f(x^*) = 0$.*

A condição acima pode ser explicada considerando-se a expansão da função f em série de Taylor, no ponto $\mathbf{x}^*$, na direção $\mathbf{d}$, até o termo de primeira ordem; isto é,

$$f(\mathbf{x}^* + \mathbf{d}) \approx f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T\mathbf{d}$$

O vetor $\nabla f(\mathbf{x})$ fornece a direção de maior variação da função f. Se $\nabla f(\mathbf{x}^*) = \mathbf{0}$, a taxa de variação da função em qualquer direção $\mathbf{p}$ é nula, portanto $\nabla f(\mathbf{x}^*)^T\mathbf{p} = 0$.

Portanto, a *condição necessária* para que a função $f(\mathbf{x})$ tenha um mínimo em $\mathbf{x}^*$ é que $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Os *pontos estacionários* de $f(\mathbf{x})$ são aqueles para os quais $\nabla f(\mathbf{x}) = \mathbf{0}$. Nesses pontos, o movimento em qualquer direção resulta numa taxa de variação nula da função f. Note que esta condição também é suficiente para que $f(\mathbf{x})$ tenha um ponto estacionário em $\mathbf{x}^*$. Esses pontos podem ser mínimos locais, máximos ou pontos de sela, conforme ilustrado na figura 2.1. A distinção entre esses casos requer o estabelecimento das condições de otimalidade de 2^a ordem.

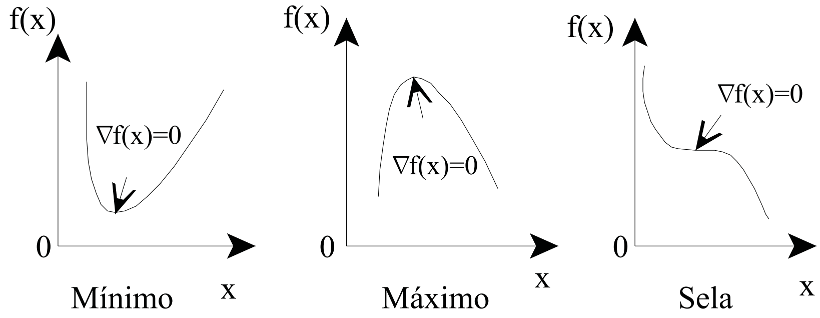

Figura 2.1: Pontos estacionários

Exemplo 6 *: Determine os pontos estacionários do problema*

$$\begin{array}{ll} Min & f(\mathbf{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_2 \\ sujeito\ a & \\ & \mathbf{x} \in E^n \end{array}$$

Exemplo 7 *: O problema abaixo tem mínimo global em $x_1 = \frac{1}{2}$ e $x_2 = 0$. Verifique as condições de otimalidade neste ponto.*

$$\begin{array}{ll} Min & f(\mathbf{x}) = x_1^2 - x_1 + x_2 + x_1x_2 \\ sujeito\ a & \\ & x_1 \geq 0, \quad x_2 \geq 0 \end{array}$$

2.3 Condições de Otimalidade de Segunda Ordem

Condições adicionais para ponto de ótimo podem ser derivadas considerando-se aproximações de ordem mais elevada de f. As condições de segunda ordem, definidas em termos da matriz Hessiana da f, são de extrema importância em problemas não-lineares genéricos.

Suponha que o vetor gradiente da função f é nulo no ponto $\mathbf{x}^*$; isto é $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Adicionalmente, considere a expansão da função f em série de Taylor, no ponto $\mathbf{x}^*$, na direção $\mathbf{d}$, até o termo de segunda ordem; ou seja,

$$f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\nabla^2 f(\mathbf{x}^*)\mathbf{d}$$

A condição adicional para que $\mathbf{x}^*$ seja um mínimo de f é que a matriz de segundas derivadas de f com relação a $\mathbf{x}$, denotada $\nabla^2 f(\mathbf{x}^*)$, seja positiva semidefinida; isto é,

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0$$

para qualquer $\mathbf{d}$ viável.

Teorema 2 : *Seja Ω um subconjunto de E^n e $f \in C^2$ uma função em Ω. Se $\mathbf{x}^*$ é um ponto de mínimo relativo de f em Ω, então para qualquer direção $\mathbf{d} \in E^n$ viável em $\mathbf{x}^*$,*

1. $\nabla f(\mathbf{x}^*)^T\mathbf{d} \geq 0$
2. Se $\nabla f(\mathbf{x}^*)^T\mathbf{d} = 0$, então $\mathbf{d}^t\nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0$

Exemplo 8 : *Verificar as condições de segunda ordem para a solução do problema*

$$\begin{array}{ll} Min & f(\mathbf{x}) = x_1^2 - x_1 + x_2 + x_1x_2 \\ \textit{sujeito a} & \\ & x_1 \geq 0, \quad x_2 \geq 0 \end{array}$$

Um caso de interesse especial é aquele onde $\mathbf{x}^*$ é interior a Ω, como por exemplo, o caso de problemas completamente irrestritos. Neste caso, a condição descrita a seguir é observada.

Teorema 3 (**Condições Necessárias de** 2^a **Ordem**): *Seja $\mathbf{x}^*$ um ponto interior a Ω. Suponha que $\mathbf{x}^*$ é um mínimo local de $f \in C^2$ em Ω. Então:*

1. $\nabla f(\mathbf{x}^*)^T\mathbf{d} = 0$;
2. *para todo* $\mathbf{d}$,
$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0$$

Observe que a segunda condição é equivalente a dizer que $\nabla^2 f(\mathbf{x}^*)$ é uma matriz positiva semidefinida.

Exemplo 9 : *Verificar as condições de segunda ordem para a solução do problema*

$$\begin{array}{ll} Min \;\; f(\mathbf{x}) = & x_1^3 - x_1^2x_2 + 2x_2^2 \\ \textit{sujeito a} & \\ & x_1 \geq 1, \quad x_2 \geq 1 \end{array}$$

Tornando mais restritiva a condição imposta acima, obtem-se o conjunto de condições que implicam que $\mathbf{x}^*$ é um forte mínimo local. Estas condições se aplicam a problemas irrestritos onde $\mathbf{x}^*$ é um ponto interior. O caso onde a solução de 2.1 pertence à *fronteira* (ou *borda*) do conjunto viável é tratado juntamente com os problemas de otimização restrita.

Teorema 4 (**Condições Suficientes de** 2^a **Ordem**): *Seja $f \in C^2$ uma função definida numa região onde o ponto $\mathbf{x}^*$ é um ponto interior. Suponha que*

1. $\nabla f(\mathbf{x}^*)^T \mathbf{d} = 0$*;*

2. $\nabla^2 f(\mathbf{x}^*)$ *é positiva definida.*

Então $\mathbf{x}^*$ *é um mínimo local de* f*.*

Exemplo 10 *: Determinar a solução do problema*

$$Min \quad f(\mathbf{x}) = \quad x_1^3 - x_1^2 x_2 + 2x_2^2 + x_1^2$$

e verificar as condições de otimalidade.

Observação 11 *: Qualquer ponto estacionário de uma função* $f(\mathbf{x})$ *pode ser determinado diferenciando-se a função* $f(\mathbf{x})$ *e igualando a zero a expressão resultante. Este ponto estacionário, pode ser de mínimo, máximo ou de sela da função. A diferenciação entre essas condições é estabelecida a partir da matriz de* 2^a *derivada.*

Observação 12 *: Não há método que forneça diretamente o mínimo ou máximo global de uma função qualquer. O procedimento referido na observação anterior é apropriado apenas para problemas de pequena dimensão com uma função objetivo simples e bem definida em termos analíticos. Em geral, mesmo para pequenos problemas, a equação resultante da* 1^a *derivada é não linear e somente pode ser resolvida por métodos iterativos.*

Observação 13 *Em muitas situações práticas, apenas os pontos estacionários na vizinhança de um determinado ponto são calculados. O mínimo é determinado das soluções disponíveis, comparando-se os valores correspondentes da função objetivo. Tal procedimento não garante o alcance do mínimo global, o que constitui uma das principais limitações desse procedimento.*

As propriedades apresentadas anteriormente são relativas à mínimos locais e não fornecem informações sobre as condições para mínimos globais, exceto se os mínimos global e local forem coincidentes. Nos casos práticos, uma alternativa para a determinação do mínimo global é a seleção entre os mínimos locais daquele que resulta no menor valor da função objetivo.

2.4 Otimização de Funções Convexas

Seja uma função $f(\mathbf{x})$ duas vezes diferenciável (isto é, f é da classe C^2). Então a condição

$$f(\mathbf{x} + \mathbf{h}) \geq \mathbf{f}(\mathbf{x}) + \nabla f(\mathbf{x})^T \mathbf{h}$$

é satisfeita para toda direção definida pelo vetor $\mathbf{h}$, se e somente se a função f é convexa.

Com base nas condições de necessidade e suficiência e nas definições de funções quadráticas e convexas apresentadas anteriormente, segue que

1. A função f é convexa e da classe C^2 dentro de um conjunto convexo Ω que contém um ponto interior se e somente se $\nabla^2 f(\mathbf{x})$ é positiva semidefinida para qualquer ponto $\mathbf{x} \in \Omega$.

2. Se $f(\mathbf{x})$ é quadrática e expressa analiticamente por

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} - \mathbf{b}^T\mathbf{x} + \alpha$$

onde $\mathbf{G}$ é uma matriz simétrica e independente de $\mathbf{x}$, é possível afirmar o seguinte [4]:

   - a função $f(\mathbf{x})$ tem um forte mínimo local, se e somente se $\mathbf{G}$ é positiva definida;
   - caso a condição anterior seja satisfeita, o mínimo local é dado por

$$\mathbf{x}^* = \mathbf{G}^{-1}\mathbf{b}.$$

Uma proriedade muito importante, a qual pode ser demonstrada para funções convexas, é estabelecida pelo teorema apresentado a seguir.

Teorema 5 *: Se $f(\mathbf{x})$ é convexa e possui um forte mínimo local, então este é único, e também é mínimo global.*

Verdadeiras funções convexas ocorrem raramente na prática, porém esse tipo de função é importante teoricamente porque um grande número de funções utilizadas em problemas reais são convexas na vizinhança do mínimo local.

Uma situação incomum é aquela na qual se deseja maximizar uma função convexa dentro de um conjunto convexo. Nestes casos um resultado importante é estabelecido pelo teorema mostrado a seguir.

Teorema 6 *: Seja f uma função convexa definida num conjunto convexo compacto Ω. Se f tem um máximo em Ω, então este máximo é atingido num ponto extremo de Ω.*

2.5 Convergência de Algoritmos de Descida.

Diferentes algoritmos para a resolução de problemas não-lineares de minimização são apresentados nesta seção. Embora esses algoritmos sejam distintos, eles têm em comum o fato de serem *algoritmos iterativos de descida*. O adjetivo iterativo significa que o algoritmo gera uma série de pontos, cada um calculado a partir do ponto anterior. O fato de serem algoritmos de descida significa que o valor de uma função avaliada no ponto mais recentemente calculado decresce com relação ao valor anterior.

Um algoritmo é iniciado especificando-se um ponto de partida. Se o algoritmo converge para uma solução para um ponto inicial arbitrário, ele é dito de *convergência global*. Caso contrário, ele é dito de *convergência local*.

Dois itens são muito importantes na hora de se escolher um algoritmo de otimização: a influência do ponto inicial na sua convegência e a velocidade com que a seqüência de pontos gerada pelo algoritmo converge para a solução ótima.

Com o objetivo de fornecer uma ideia preliminar sobre o significado da *Ordem de Convergência* dos algoritmos de otimização, considere uma seqüência de números reais $\{r^k\}_{k=0}^{\infty}$, convergindo para o limite r^*. A ordem de convergência desta seqüência, denotada p, é o maior número não negativo tal que

$$\lim_{k\to\infty} \frac{\left|r^{k+1} - r^*\right|}{\left|r^k - r^*\right|^p} = \beta < \infty$$

Se $p = 1$ diz-se que o algoritmo tem convergência *linear* se a razão de convergência $\beta < 1$. Se $p > 1$, o algoritmo tem convergência superlinear. No caso especial definido para $p = 2$ o algoritmo é dito de convergência *quadrática*. Um exemplo deste último tipo de ordem de convergência é aquela obtida através do método de Newton-Raphson.

2.6 Métodos Básicos de Descida

Existe uma grande variedade de algoritmos de resolução de problemas irrestritos. São analisados aqui apenas os mais conhecidos. Existe uma estrutura básica em todos os algoritmos de descida que serão discutidos: À partir de um ponto inicial, determina-se de acordo com uma regra bem estabelecida, uma direção de movimento e o prossegue-se nessa direção até o mínimo da função objetivo. No novo ponto, uma nova direção é determinada e o processo é repetido. A diferença principal entre os vários algoritmos encontrados na literatura está na regra para se determinar a direção de busca da solução ótima. Uma vez que a seleção é feita, todos os algorimos usam subrotinas que calculam o ponto de mínimo no espaço unidimensional correspondente à esta direção. A Figura 2.2 ilustra o procedimento.

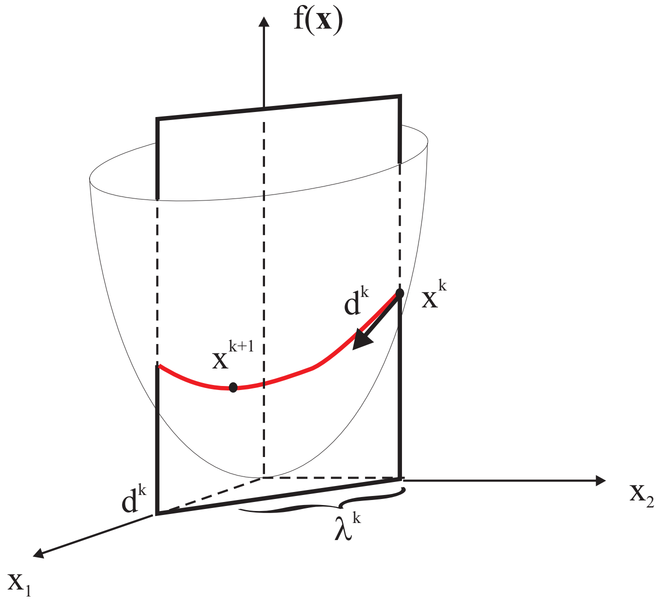

Figura 2.2: Algoritmos de Descida

O procedimento para se determinar um ponto de mínimo numa dada direção é chamado de *Busca Unidimensional*. As técnicas de busca unidimensional proporcionam de certa forma uma base para os algoritmos de programação não-linear, uma vez que problemas de maiores dimensões são muitas vezes resolvidos executando-se uma seqüência de buscas unidimensionais.

2.7 Métodos de Busca Unidimensional

Para uma dada direção $\mathbf{d}$, seja $f(\lambda) = f(\mathbf{x} + \lambda\mathbf{d})$ uma função unimodal[1] e $\lambda \in R$. Algumas medodologias para a minimização de $f(\lambda)$ são:

- Regra de Armijo
- Método de Newton

A regra de Armijo foi desenhada para garantir que haverá um decréscimo da função objetivo na direção especificada, ou seja, não encontra o mínimo de $f(\lambda)$. Essa regra se baseia no valor da função f e de seu gradiente.

2.7.1 Regra de Armijo

Tendo sido fornecida uma direção de descida $\mathbf{d}^k$, esta regra é aplicada para se obter um tamanho de passo na direção especificada λ^k que garanta a diminuição da função objetivo. O algoritmo pode ser resumido nos seguintes passos:

1. Passo inicial:
   - Escolha um escalar $\beta \in (0,1)$ e outro escalar $\sigma \in (0,1)$.
   - Faça $k = 1$.
2. Passo Principal:
   - Faça $\lambda^k = \beta^{m_k}$, sendo m_k o primeiro inteiro não negativo, m para o qual

$$f(\mathbf{x}) - f(\mathbf{x} + \beta^m \mathbf{d}^k) \geq -\sigma.\beta^m \nabla f(\mathbf{x}^k)^T \mathbf{d}^k$$

Os tamanhos de passo $1, \beta, \beta^2, ...$ são testados até que a desigualdade anterior seja satisfeita para $m = m_k$, como mostra a figura 2.3.

2.7.2 Método de Newton

O método de Newton é aplicável apenas na minimização de funções duas vezes diferenciáveis. Esse método é baseado na aproximação quadrática da função f num dado ponto λ^k. Esta aproximação é dada por

$$q(\lambda) = f(\lambda^k) + f'(\lambda^k)(\lambda - \lambda^k) + \frac{1}{2} f''(\lambda^k)(\lambda - \lambda^k)^2$$

O ponto λ^{k+1} é tomado onde a derivada de q é igual a zero. Isto resulta na equação

$$f'(\lambda^k) + f''(\lambda^k)(\lambda - \lambda^k) = 0$$

De forma que

$$\lambda^{k+1} = \lambda^k - \frac{f'(\lambda^k)}{f''(\lambda^k)}$$

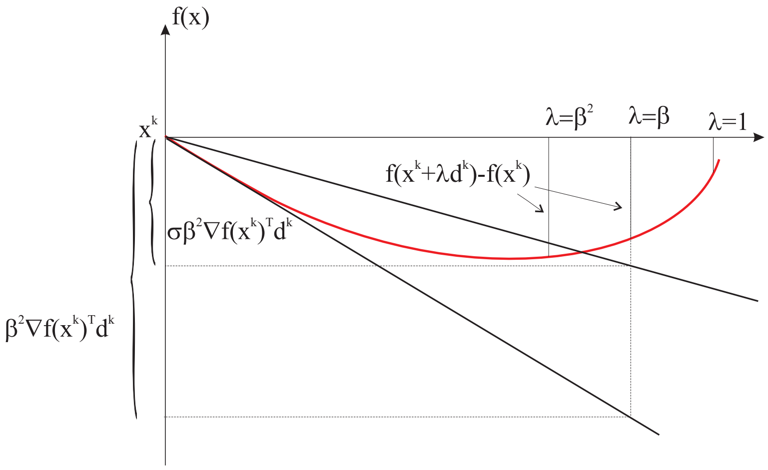

Figura 2.3: Regra de Armijo

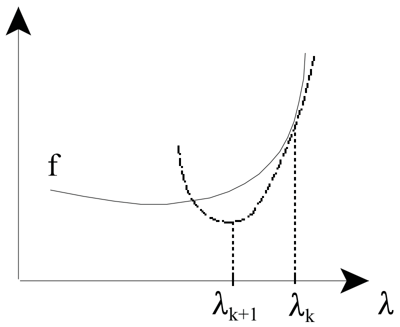

Figura 2.4: Busca unidimensional via método de Newton

O processo é finalizado quando $\left|\lambda^{k+1} - \lambda^k\right| < \varepsilon$ ou quando $\left|f'(\lambda^k)\right| < \varepsilon$.

A Figura 2.4 ilustra a busca unidimensional via método de Newton.

Exemplo 11 *Considere novamente o exemplo* **??**. *Dê um passo de busca unidimensional pelo método de Newton para minimizar $f(\mathbf{x})$ a partir de $\mathbf{x}^0$ na direção $\mathbf{d}$.*

2.8 Minimização de uma Função Multivariável

São descritos aqui dois dos métodos mais usados na otimização irrestrita. Há, na realidade, um grande número de métodos numéricos que podem ser aplicados para resolver problemas de otimização sem restrições. Os métodos descritos a seguir foram escolhidos por serem basicos e terem ampla utilização.

2.8.1 Método do Gradiente

Esse é provavelmente o método mais conhecido para a resolução do problema representado pela equação 2.1. Embora existam métodos mais avançados, muitos deles são

[1]A função unimodal possui um único ponto de mínimo.

modificações desse método. O método do gradiente constitui portanto uma referência padrão usada para analizar outras técnicas e por esta razão é discutido aqui.

Sejam f uma função contínua e diferenciável em E^n e o seu correspondente vetor de primeiras derivadas, expresso como

$$\mathbf{g}^k = \frac{\nabla f(\mathbf{x}^k)}{|\nabla f(\mathbf{x}^k)|}$$

O método da máxima descida é definido pelo algoritmo iterativo:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \mathbf{g}^k \tag{2.2}$$

onde α^k é um escalar não negativo que minimiza a função $f(\mathbf{x}^k - \alpha^k \mathbf{g}^k)$.

A análise da equação 2.2 revela que a cada iteração o ponto de mínimo da função f, denotado $(\mathbf{x}^{k+1})$, é determinado à partir da solução corrente $(\mathbf{x}^k)$, ao longo da direção do negativo do vetor gradiente $(-\mathbf{g}^k)$.

Observação 14 *O método do gradiente possui* convergência global.
Considere agora um problema quadrático

$$Minimizar \quad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} - \mathbf{b}^T\mathbf{x}$$

onde $\mathbf{G}$ é uma matriz simétrica definida positiva.

O vetor gradiente de $f(\mathbf{x})$ é dado por

$$\mathbf{g}(\mathbf{x}) = \mathbf{G}\mathbf{x} - \mathbf{b}$$

e as condições de otimalidade estabelecem que o ponto de mínimo de f, denotado $\mathbf{x}^*$, satisfaz a relação

$$\mathbf{G}\mathbf{x}^* = \mathbf{b}$$

Assim, na iteração k,

$$\mathbf{g}^k = \mathbf{G}\mathbf{x}^k - \mathbf{b}$$

tal que, para a aplicação do algoritmo do gradiente representado pela expressão 2.2, α^k pode ser determinado explicitamente, diferenciando-se a função $f(\mathbf{x}^k - \alpha\mathbf{g}^k)$ com relação a α e igualando-se o resultado a zero.

Portanto, para

$$f(\mathbf{x}^k - \alpha^k\mathbf{g}^k) = \frac{1}{2}(\mathbf{x}^k - \alpha^k\mathbf{g}^k)^T\mathbf{G}(\mathbf{x}^k - \alpha^k\mathbf{g}^k) - (\mathbf{x}^k - \alpha^k\mathbf{g}^k)^T\mathbf{b}$$

tem-se

$$-\left(\mathbf{g}^k\right)^T\mathbf{G}(\mathbf{x}^k - \alpha^k\mathbf{g}^k) + \left(\mathbf{g}^k\right)^T\mathbf{b} = \mathbf{0}$$

ou ainda

$$-\left(\mathbf{g}^k\right)^T(\mathbf{G}\mathbf{x}^k - \mathbf{b}) + \alpha^k\left(\mathbf{g}^k\right)^T\mathbf{G}\mathbf{g}^k = \mathbf{0} \tag{2.3}$$

Da Equação 2.3, lembrando que $\mathbf{g}_k = \mathbf{G}\mathbf{x}_k - \mathbf{b}$, obtém-se

$$\alpha_k = \frac{\left(\mathbf{g}^k\right)^T\mathbf{g}^k}{\left(\mathbf{g}^k\right)^T\mathbf{G}\mathbf{g}^k}$$

Portanto, o método da máxima descida toma a forma

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{\left(\mathbf{g}^k\right)^T \mathbf{g}^k}{\left(\mathbf{g}^k\right)^T \mathbf{G}\mathbf{g}^k}\mathbf{g}^k$$

com $\mathbf{g}^k = \mathbf{G}\mathbf{x}^k - \mathbf{b}$.

A Figura 2.5 mostra as curvas de nível da função f com uma seqüência de pontos gerados durante o processo iterativo. Os contornos de f são elipsóides com eixos nas direções dos n autovetores (ortogonais) de $\mathbf{Q}$. O eixo correspondendo ao i-ésimo autovetor tem comprimento porporcional a $\frac{1}{\lambda_i}$.

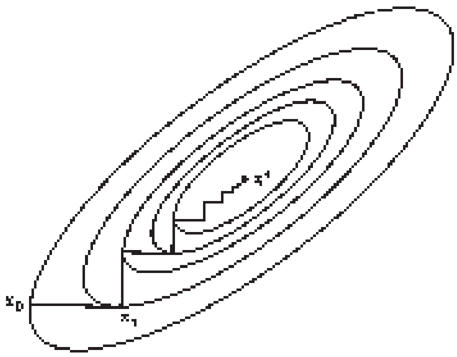

Figura 2.5: Método do Gradiente

Seja a função

$$\hat{E}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{G}(\mathbf{x} - \mathbf{x}^*)$$

Portanto,

$$\hat{E}(\mathbf{x}) = \frac{1}{2}\left[\mathbf{x}^t\mathbf{G}(\mathbf{x} - \mathbf{x}^*) - \mathbf{x}^{*t}\mathbf{G}(\mathbf{x} - \mathbf{x}^*)\right] = \mathbf{x}^t\mathbf{G}\mathbf{x} + \mathbf{x}^{*t}\mathbf{G}\mathbf{x}^*$$

ou ainda

$$\hat{E}(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{2}\mathbf{x}^{*t}\mathbf{G}\mathbf{x}^*$$

Ou seja, que $\hat{E}(\cdot)$ difere de $f(\cdot)$ por apenas uma constante. Sendo assim, minimizar f é equivalente a minimizar $\hat{E}$. O teorema apresentado a seguir é útil para compreender o comportamento do método do gradiente.

Teorema 7 *: Para todo $\mathbf{x}_0 \in E^n$ o método da máxima descida converge para o ponto de mínimo de f. Em cada passo a seguinte desigualdade é satisfeita*

$$\hat{E}(\mathbf{x}^k) \leq \left(\frac{A - a}{A + a}\right)^2 \hat{E}(\mathbf{x}^k)$$

onde a e A são, respectivamente, o menor e maior autovalores de $\mathbf{G}$.

Este teorema indica que a taxa de convergência do método de máxima descida é diminuída quando os contornos de f se tornam mais deformados (mais elípticos). Se os contornos são circunferências, a convergência ocorre em uma iteração.

O resultado anterior pode ser generalizado para o caso de uma função não quadrática.

Teorema 8 *Seja a função f definida em E^n, com derivadas de segunda ordem contínuas e com um mínimo relativo em $\mathbf{x}^*$. Suponha também que $\nabla^2 f(\mathbf{x})$ tem o menor autovalor $\alpha > 0$ e o maior autovalor $A > 0$. Se $\{\mathbf{x}^k\}$ é a seqüência gerada pelo método da máxima descida que converge para $\mathbf{x}^*$, então a seqüência $\{f(\mathbf{x}^k)\}$ converge para $f(\mathbf{x}^*)$ linearmente com uma taxa de convergência não maior do que $[(A-a)/(A+a)]^2$.*

As afirmações dos últimos dois teoremas indicam que o método do gradiente não possui boas características de convergência nos casos em que a matriz Hessiana da função f é *mal condicionada*. Isto significa que o método é dependente da escolha das variáveis do problema. Uma escolha que leve a uma melhor estrutura para os autovalores de $\nabla^2 f(\mathbf{x})$ pode levar a uma convergência muito mais rápida. Isto leva à introdução de fatores para escalonar as variáveis. Tais mecanismos são conhecidos na literatura como *mecanismos de condicionamento* e são muitas vezes imprescindíveis para o bom desempenho de métodos derivados do método do gradiente.

O método do gradiente usualmente tem boa performance no início do processo iterativo. Entretanto, à medida em que se aproxima do ponto estacionário, a convergência do algoritmo se torna lenta devido aos pequenos passos ortogonais que gera. Esse fenômeno é conhecido como *zig-zag*.

O algoritmo mostrado a seguir resume os passos para a aplicação do método da máxima descida.

1. Passo Inicial: Seja $\varepsilon > 0$ uma tolerância. Escolha um ponto inicial $\mathbf{x}^0$;

2. Passo Principal:

   - Se $\left|\nabla f(\mathbf{x}^k)\right| < \varepsilon$, o processo é encerrado;
   - Caso contrário, calcule o vetor gradiente

   $$\mathbf{g}^k = \frac{\nabla f(\mathbf{x}^k)}{|\nabla f(\mathbf{x}^k)|}$$

   e determine α^k como a solução ótima do problema

   $$Minimizar \quad f(\mathbf{x}^k - \alpha \mathbf{g}^k)$$

   sujeito a restrição $\alpha \geq 0$;
   - Faça $\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \mathbf{g}^k$, $k = k+1$ e repita o passo principal.

Exemplo 12 *Considere novamente o exemplo* **??**. *Dê um passo do método do gradiente a partir de $\mathbf{x}^0$ para minimizar $f(\mathbf{x})$.*

2.8.2 Método de Newton

A idéia básica do método de Newton é que a função f a ser minimizada pode ser aproximada localmente por uma função quadrática. Essa função quadrática aproximada é minimizada de forma exata. Então, na vizinhança de $\mathbf{x}^k$ podemos aproximar f por uma série de Taylor até o termo de segunda ordem; isto é,

$$f(\mathbf{x}) \cong f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T \left(\mathbf{x} - \mathbf{x}^k\right) + \frac{1}{2}\left(\mathbf{x} - \mathbf{x}^k\right)^T \nabla^2 f(\mathbf{x}^k)\left(\mathbf{x} - \mathbf{x}^k\right)$$

A minimização da função do lado direito da última expressão fornece

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \left[\nabla^2 f(\mathbf{x}^k)\right]^{-1} \nabla f(\mathbf{x}^k)$$

que representa a forma recursiva utilizada quando se aplica o método de Newton.

Tendo em vista as condições de segunda ordem para a existência do ponto de mínimo, podemos supor que $\nabla^2 f(\mathbf{x}^k)$ é definida positiva próximo de $\mathbf{x}^*$, e portanto o método é bem definido na vizinhança da solução.

Teorema 9 *Seja $f \in C^3$ em E^n e suponha que, num ponto de mínimo local $\mathbf{x}^*$, a matriz Hessiana $\nabla^2 f(\mathbf{x}^k)$ é definida positiva. Então, se iniciado num ponto suficientemente próximo de $\mathbf{x}^*$, os pontos gerados pelo método de Newton convergem para $\mathbf{x}^*$. A ordem de convergência é pelo menos igual a 2.*

Embora o método de Newton tenha boa convergência próximo da solução, são necessárias modificações para que possa ser usado em pontos distantes da solução ótima. A primeira modificação introduzida é expressa pela expressão abaixo:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \left[\nabla^2 f(\mathbf{x}^k)\right]^{-1} \nabla f(\mathbf{x}^k)$$

onde α^k é selecionado para minimizar f.

Na modificação descrita acima é introduzido um passo de busca unidimensional no algoritmo. Próximo da solução ótima, $\alpha^k \simeq 1$, enquanto que, para pontos distantes α^k pode assumir valores mais próximos de zero.

Outras modificações podem ser feitas no método de Newton a partir de diferentes aproximações para a matriz $\nabla^2 f(\mathbf{x}^k)$ [4].

O algoritmo básico do método de Newton é descrito abaixo. Note que, neste algoritmo, α^k é calculado através de uma busca unidimensional. Nem sempre este cálculo é necessário. Se $\mathbf{x}^k$ está suficientemente próximo de $\mathbf{x}^*$ α^k pode ser feito igual a 1 em todas as iterações. O algoritmo consiste nos passos sumarizados a seguir.

1. Passo Inicial: Seja $\varepsilon > 0$ uma tolerância. Escolha um ponto inicial $\mathbf{x}^0$;
2. Passo Principal:
   - Se $\left|\nabla f(\mathbf{x}^k)\right| < \varepsilon$, o processo é encerrado;
   - caso contrário, calcule

     $$\mathbf{d}^k = \left[\nabla^2 f(\mathbf{x}^k)\right]^{-1} .\nabla f(\mathbf{x}^k)$$

     e minimize a função $f(\mathbf{x}^k - \alpha \mathbf{d}^k)$ sujeito a $\alpha \geq 0$ com relação a α. A solução é α^k;
   - Faça $\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \mathbf{d}^k$, $k = k + 1$ e repita o passo principal.

Exemplo 13 *Considere novamente o exemplo* **??**. *Dê um passo do método do Newton a partir de $\mathbf{x}^0$ para minimizar $f(\mathbf{x})$.*

Tanto o método do gradiente como o método de Newton dão origem a algoritmos de otimização largamente utilizados em problemas práticos. O método do gradiente forma a base para os *métodos de direções conjugadas*, os quais têm sido usados recentemente na resolução de sistemas lineares de grande porte. Por outro lado, o método de Newon é a base da maioria dos algoritmos de otimização de problemas não lineares com restrições. Infelizmente, a propriedade de convergência local se mantém também para estas aplicações, de forma que muitas vezes são necessárias modificações para aumentar o raio de convergência deste método.

2.9 Exercícios

1. Usando as condições necessárias de primeira ordem, encontre o ponto de mínimo da função

$$f(x,y,z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9$$

   - Verifique que o ponto é um ponto de mínimo relativo através das condições suficientes de 2^a ordem.
   - O ponto obtido é um ponto de mínimo global? Mostre porque.

2. Mostre que a função objetivo

$$f(x) = \mathbf{x}^T \begin{bmatrix} -3.03 & 1.345 \\ 1.345 & -6.96 \end{bmatrix} \mathbf{x} + [7.31 \quad 26.6]\mathbf{x} + 55.84$$

   onde $\mathbf{x} = [x_1 \quad x_2]^T$ é convexa. Mostre que $-f(x)$ é côncava.

3. Suponha que desejamos encontrar as raízes reais da equação

$$f_1(x) = 3000 - 100x^2 - 4x^5 - 6x^6 = 0$$

   Como este problema pode ser convertido em um problema de otimização?

4. Encontre os pontos que satisfazem as condições para pontos estacionários da função

$$f(\mathbf{x}) = \frac{x_1 + x_2}{3 + (x_1)^2 + (x_2)^2 + x_1 x_2}$$

   Estabeleça a natureza destes pontos através das condições suficientes de otimalidade.

5. Considere o problema de despacho econômico de usinas térmicas

$$\begin{array}{ll} Min & C = \sum_{i=1}^{3} a_i Pg_i^2 + b_i Pg_i + c_i \\ \text{s.a} & \\ & \sum_{i=1}^{3} Pg_i = Pd \\ & 20 \leq Pg_1 \leq 100 \\ & 20 \leq Pg_2 \leq 80 \\ & 10 \leq Pg_3 \leq 200 \end{array}$$

   onde Pg_i são expressos em MW

Gerador	a_i	b_i	c_i
1	0.0020	6.7	180
2	0.0083	6.43	743
3	0.0022	6.75	360

Uma das formas de resolvê-lo constitui em transformá-lo em um problema irrestrito com a seguinte forma

$$\begin{array}{ll} \underset{Pg_i}{Min} & C = \sum_{i=1}^{3}\left(a_i Pg_i^2 + b_i Pg_i + c_i\right) + \\ & w_1\left(\sum_{i=1}^{3} Pg_i - Pd\right)^2 + w_2\left(Pg_1 - \frac{100-20}{2}\right)^2 + \\ & w_3\left(Pg_2 - \frac{80-20}{2}\right)^2 + w_4\left(Pg_3 - \frac{200-10}{2}\right)^2 \end{array}$$

Verifique se a resolução do problema acima com os pesos $w_1 = 5, w_2 = w_3 = w_4 = 1$ fornece uma solução viável para o problema original. Caso não forneça, aumente os valores dos pesos até que se obtenha uma solução viável.

6. Implemente uma subrotina em SCILAB ou em MATLAB para resolver, usando o método da razão áurea, o seguinte problema de busca unidimensional:

$$\begin{array}{ll} Min & \lambda^2 + 2\lambda \\ \text{s.a} & -3 \leq \lambda \leq 5 \end{array}$$

Considere que o intervalo de incerteza final deve ter no máximo comprimento igual a 0.2 [Bazaraa].

7. Os dados de linha e de barra para o sistema de potência da Figura 2.6 são:

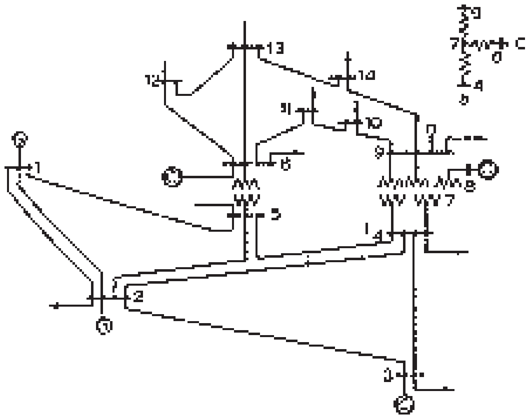

Figura 2.6: Diagrama unifilar - exercício 7

linha	de	para	reatância (pu)
1	1	2	0.0597
2	1	5	0.22304
3	2	3	0.19797
4	2	4	0.17632
5	2	5	0.17388
6	3	4	0.17103
7	4	5	0.04211
8	4	7	0.20912
9	4	9	0.55618
10	5	6	0.25202
11	6	11	0.19890
12	6	12	0.25581
13	6	13	0.13027
14	7	8	0.17615
15	7	9	0.11001
16	9	10	0.08450
17	9	14	0.27038
18	10	11	0.19207
19	12	13	0.19988
20	13	14	0.34802

barra	ger. (MW)	carga (MW)
1	0	0
2	40	21.7
3	0	94.2
4	0	47.8
5	0	7.6
6	0	11.2
7	0	0
8	0	0
9	0	29.5
10	0	9.0
11	0	3.5
12	0	6.1
13	0	13.5
14	0	14.9

Base do Sistema: $100MVA$

(a) Modele o problema do fluxo de carga linearizado para este sistema tomando a barra 1 como referência.

(b) Transforme este problema num problema de otimização irrestrita.

(c) Implemente uma rotina em SCILAB ou em MATLAB para a resolução do problema de otimização pelo método do gradiente.

2.10 Trabalhos Dirigidos

1. Considere o sistema indicado na Figura 2.7, cujos dados em *pu* na base $100MVA$ e $138kV$ são mostrados nas tabelas abaixo. Usando o SCILAB ou o MATLAB

Tabela 2.1: Dados de barra

Barra	P_g	reat. trafo	P_d
1	–	0.05	0.5
2	–	0.025	0.5
3	–	0.025	0.25
4	–	0.05	0.6
5	–	–	1.34
6	–	–	1.73
7	–	–	1.4
8	–	–	1.56
9	–	–	–
10	1.8	–	–
11	1.8	–	–
12	2.0	–	–

Tabela 2.2: Dados de linha

Linha	De	Para	x
1	1	2	0.07
2	1	6	0.01
3	1	5	0.03
4	1	4	0.065
5	4	5	0.02
6	3	4	0.063
7	3	8	0.015
8	3	7	0.023
9	2	3	0.081
10	2	7	0.030
11	6	7	0.010
12	5	6	0.013
13	5	8	0.021
14	7	8	0.031

   (a) Obtenha a matriz **B**.
   (b) Retirando linha e a coluna referentes à barra 9, obtenha os fatores **LU** e **LDU** do sistema.
   (c) Por substituição direta e inversa, obtenha o vetor dos ângulos das barras, δ.
   (d) Obtenha os fluxos nas linhas e nos transformadores do sistema. Além disso, calcule também a potência fornecida pelo gerador da barra 9.

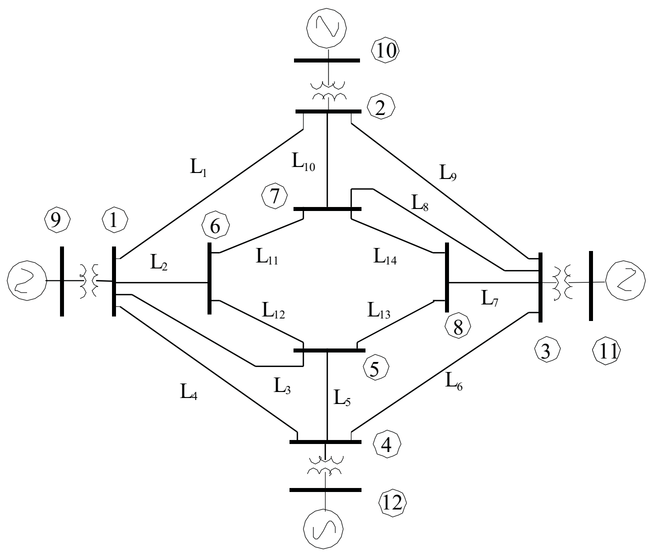

Figura 2.7: Sistema teste 1

(e) Considere agora o sistema linear formado pelas equações de balanço potência das barras 1, 3, 5, 7, 10 e 12. Supondo $\delta_9 = 0$, obtenha as soluções genérica e particular desse sistema para

$$r = (-0.1646, -0.1409, -0.1935, -0.0794, -0.1629)$$

Compare com a solução obtida no item (c).

(f) Forme um novo sistema linear, introduzindo no sistema do item anterior as equações que representam os fluxos nas linhas $1, 2, 5, 6, 9, 11, 13$ e nos quatro transformadores. Supondo $\delta_9 = 0$, resolva esse sistema e analise o resultado.

Para obter a solução genérica do sistema subdeterminado, o espaço nulo da matriz **B** pode ser obtido em SCILAB ou MATLAB com o comando "$null(\mathbf{B})$". Duas rotinas podem ser implementadas em SCILAB ou MATLAB para desenvolver o trabalho:

(a) Rotina com os dados de entrada. Para:

- nb = número de barras
- nl = número de linhas
- de = vetor com as barras origem das linhas/trafos
- $para$ = vetor com as barras destino das linhas/trafos
- Pg = vetor com as potências geradas nas barras
- Pd = vetor com as cargas das barras

A rotina possui as seguintes declarações:

$nb = 12;$
$nl = 18;$
$de = [1\ 1\ 1\ 1\ 4\ 3\ 3\ 3\ 2\ 2\ 6\ 5\ 5\ 7\ 9\ 10\ 11\ 12\ ]';$
$para = [2\ 6\ 5\ 4\ 5\ 4\ 8\ 7\ 3\ 7\ 7\ 6\ 8\ 8\ 1\ 2\ 3\ 4\ ]';$
$x =$[0.07 0.01 0.03 0.065 0.02 0.063 0.015 0.023 0.081
0.03 0.01 0.013 0.021 0.0311 0.05 0.025 0.025 0.05]';
$Pg = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.0\ 1.8\ 1.8\ 2.0\ ]';$
$Pd = [0.5\ 0.5\ 0.25\ 0.6\ 1.34\ 1.73\ 1.40\ 1.56\ 0\ 0\ 0\ 0\ ]';$

(b) Rotina para a montagem da matriz **B** do fluxo de carga CC:

$$
\begin{array}{l}
B = zeros(nb, nb); \\
for\ linha = 1 : nl \\
\quad k1 = de(linha); \\
\quad k2 = para(linha); \\
\quad aux = 1/x(linha); \\
\quad B(k1, k1) = B(k1, k1) + aux; \\
\quad B(k1, k2) = B(k1, k2) - aux; \\
\quad B(k2, k1) = B(k2, k1) - aux; \\
\quad B(k2, k2) = B(k2, k2) + aux; \\
end
\end{array}
$$

2. Considere o sistema de 4 barras da Figura 2.8. As equações de balanço de potência ativa e reativa desse sistema são fornecidas pelas expressões:

$$P_{g_i} - P_{d_i} = P_i,\ i = 2, 3, 4$$

$$Q_{g_i} - Q_{d_i} = Q_i,\ i = 3, 4$$

onde P_i, Q_i são as potências ativas e reativas injetadas nas barras expressas por

$$P_i = \sum_{k=1}^{4} V_i V_k [G_{ik} cos(\delta_{ik}) + B_{ik} sen(\delta_{ik})]$$

$$Q_i = \sum_{k=1}^{4} V_i V_k [G_{ik} sen(\delta_{ik}) - B_{ik} cos(\delta_{ik})]$$

sendo P_{g_i}, Q_{g_i} as potências ativas e reativas geradas e P_{d_i}, Q_{d_i} as cargas ativas e reativas fornecidas na Tabela 2.3, G_{ik}, B_{ik} elementos das matrizes **G** e **B** abaixo e $\delta_{ik} = \delta_i - \delta_k$ a diferença angular entre as barras.

(a) Calcule a matriz jacobiana do sistema formado pelas equações de injeção de potência ativa e reativa para os valores de tensão e ângulo indicados na Tabela 2.3. Considere as tensões nas barras 1 e 2 e o ângulo da barra 1 constante (isto é, não há derivadas em relação a essas variáveis).

(b) Através da série de Taylor, obtenha uma aproximação linear para o sistema formado pelas injeções de potência para os valores de tensão e ângulo indicados na Tabela 2.3. Compare esses valores com aqueles obtidos pelas expressões não lineares e comente os resultados.

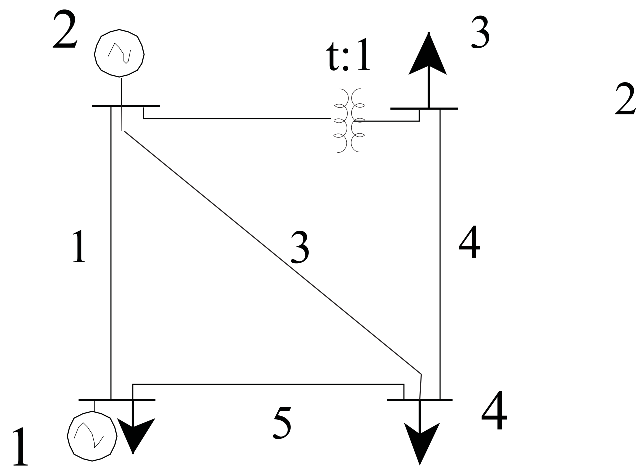

Figura 2.8: Sistema teste 2

Tabela 2.3: Dados - sistema de 4 barras

Barra	δ(rad.)	V(pu)	P_g(pu)	Q_g(pu)	P_d(pu)	Q_d(pu)
1	0	1.03	0.6371	0.5944	0.6	0.3
2	0.1069	1.02	1.4	0.6692	0	0
3	-0.1256	0.9137	-	-	0.9	0.4
4	-0.0921	0.9352	-	-	0.7	0.2

$$\mathbf{G} = \begin{bmatrix} 1.1424 & -0.5541 & 0 & -0.5883 \\ -0.5541 & 0.9880 & 0 & -0.4339 \\ 0 & 0 & 0.5765 & -0.5765 \\ -0.5883 & -0.4339 & -0.5765 & 1.5988 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} -4.8917 & 2.3249 & 0 & 2.5819 \\ 2.3249 & -7.8093 & 3.5000 & 1.8275 \\ 0 & 3.5000 & -4.6418 & 1.3085 \\ 2.5819 & 1.8275 & 1.3085 & -5.6998 \end{bmatrix}$$

3. Considere a função f_4 do exercício 12 do capítulo 2 da apostila:

   (a) Usando o SCILAB ou MATLAB, plote essa função com as curvas de nível definidas pelo ponto $\mathbf{x}^0$ e pelo ponto $\mathbf{x}^1 = (0.5, 0.5)$.

   (b) Plote os vetores gradiente de f_4 nos pontos $\mathbf{x}^0$ e $\mathbf{x}^1$ nas curvas de nível obtidas anteriormente. Comente o resultado.

   (c) Calcule a hessiana de f_4, seus autovalores e autovetores correspondentes.

   (d) Forme a matriz ortogonal $\mathbf{Q}$ com os autovalores da hessiana e, para $\mathbf{x}' = \mathbf{Q}\mathbf{x}$, expresse f_4 em função de $\mathbf{x}'$.

   (e) Plote a função modificada e suas curvas de nível.

   (f) Obtenha os autovalores e autovetores da hessiana da função modificada. Plote os novos autovetores no mesmo gráfico das curvas de nível da função modificada. Verifique, por observação, quais autovetores definem direções de aumento ou diminuição de f_4.

4. Resolva o exercício 19 do capítulo 2 da apostila.

5. Faça uma rotina computacional para resolver o exercício 5 do capítulo 3 da apostila através do método de Newton. Considere o passo da busca unidimentional, α, constante e igual a 1.

6. Faça uma rotina computacional para resolver o exercício 7 do capítulo 3 da apostila.

Capítulo 3

Otimização com Restrições

Este capítulo visa o estudo de problemas de otimização com restrições de igualdade e de desigualdade. Primeiramente, a caracterização de problemas restritos e de suas correspondentes soluções é analisada. Pretende-se com isto, mostrar como as restrições podem influenciar a localização da solução ótima, a ponto de em certos casos inviabilizar a solução do problema de otimização. A seguir, as condições de otimalidade de primeira e segunda ordem para problemas com restrições de igualdade e/ou de desigualdade são estabelecidas. Estas condições fornecem subsídios para se determinar se uma especificada solução constitui um ponto estacionário, e em caso afirmativo, qual a natureza (máximo, mínimo, ponto de sela etc) desta solução.

3.1 Conceitos Básicos

A forma geral de um problema de otimização com restrições é

$$\begin{array}{ll} Min & f(\mathbf{x}) \\ \text{s.a} & g_i(\mathbf{x}) = 0, \;\; i = 1 \ldots m \\ & h_k(\mathbf{x}) \geqslant 0, \;\; k = 1 \ldots l \end{array} \tag{3.1}$$

onde, $\mathbf{x}$ é um vetor $n \times 1$, das variáveis de otimização; $g_i(\mathbf{x}) = 0$ representa a i-ésima equação do conjunto de m restrições de igualdade; e $h_k(\mathbf{x}) \geqslant 0$ representa a k-ésima inequação do conjunto de l restrições de desigualdade.

As equações $g_i(\mathbf{x}) = 0$ $(i = 1, \ldots, m)$ e $h_k(\mathbf{x}) \geqslant 0$ $(k = 1, \ldots, l)$ definem o que é denominada *região das soluções viáveis,* ou seja, das soluções que satisfazem a ambas restrições de igualdade e de desigualdade simultaneamente.

A solução ótima do problema de otimização com restrições, denotada $\mathbf{x}^*$, é uma solução viável (isto é, $g_i(\mathbf{x}^*) = 0$ para $i = 1, \ldots, m$ e $h_k(\mathbf{x}^*) \geqslant 0$ para $k = 1, \ldots, l$), correspondente ao valor mínimo da função objetivo, denotado $f(\mathbf{x}^*)$.

O problema representado na Equação (3.1) pode também ser expresso na forma matricial mostrada a seguir.

$$\begin{array}{rl} Min & f(\mathbf{x}) \\ s.a & \mathbf{g}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) \geqslant \mathbf{0} \end{array} \tag{3.2}$$

onde $\mathbf{g}(\cdot)$ e $\mathbf{h}(\cdot)$ são vetores de ordens $m \times 1$ e $l \times 1$, cujas componentes são as funções que representam as restrições de igualdade e desigualdade, respectivamente.

Visando ilustrar como as restrições afetam a solução de um problema de otimização, consideremos os casos mostrados a seguir.

Exemplo 14

$$\begin{array}{ll} Min & f(\mathbf{x}) = (x_1 - 1,5)^2 + (x_2 - 1,5)^2 \\ s.a & h_1(\mathbf{x}) = x_1 + x_2 - 2,0 \leqslant 0 \\ & h_2(\mathbf{x}) = -x_1 \leqslant 0 \\ & h_3(\mathbf{x}) = -x_2 \leqslant 0 \end{array}$$

A análise das equações envolvidas neste problema de otimização revela que:

- *as curvas de nível (ou de contorno) da função objetivo são círculos concêntricos no ponto* $(1,5\ ;\ 1,5)$*;*
- *a restrição* $h_1(\mathbf{x}) = x_1 + x_2 - 2,0 \leqslant 0$ *corresponde à parte inferior da região delimitada pela reta que passa pelos pontos* $(0,0; 2,0)$ *e* $(2,0; 0,0)$*, incluindo esta reta;*
- *as restrições* $h_2(\mathbf{x}) = -x_1 \leqslant 0$ *e* $h_3(\mathbf{x}) = -x_2 \leqslant 0$ *são restrições de não-negatividade nas variáveis de otimização;*

Se as restrições fossem ignoradas, o valor mínimo da função objetivo seria no ponto $(1,5; 1,5)$*, para o qual a função objetivo vale* $0,0$ *e a restrição* $h_1(\mathbf{x})$ *é violada. Se as restrições são consideradas, a solução ótima é* $(1,0; 1,0)$*, com o valor* $0,5$ *para a função objetivo neste ponto. Nota-se que, neste caso a localização da solução ótima é na fronteira da região das soluções viáveis, sendo diretamente influenciada pelas restrições.*

As curvas de nível da função objetivo e a região definida pelas restrições são representadas na Figura 3.1

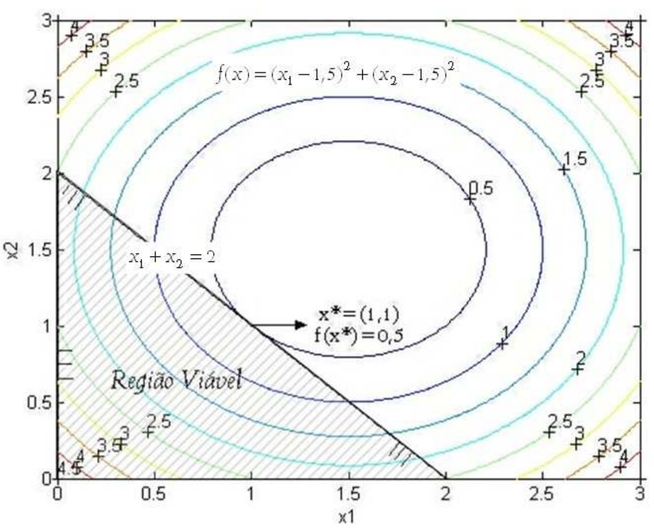

Figura 3.1: Representação do problema do exemplo 14

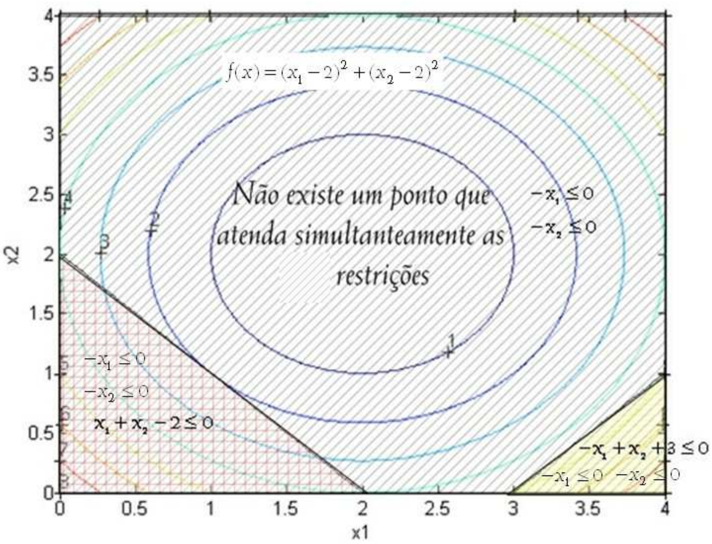

Figura 3.2: Representação do problema do exemplo 16

Exemplo 15

$$
\begin{array}{ll}
Min & f(\mathbf{x}) = (x_1 - 0,5)^2 + (x_2 - 0,5)^2 \\
s.a & h_1(\mathbf{x}) = x_1 + x_2 - 2,0 \leqslant 0 \\
 & h_2(\mathbf{x}) = -x_1 \leqslant 0 \\
 & h_3(\mathbf{x}) = -x_2 \leqslant 0
\end{array}
$$

O conjunto de restrições deste problema é o mesmo que o do problema anterior. As curvas de contorno da função objetivo também possuem a mesma forma, estando porém centradas no ponto $(0,5; 0,5)$.

Se as restrições fossem ignoradas, o valor mínimo da função objetivo seria no ponto $(0,5; 0,5)$*, para o qual a função objetivo vale* $0,0$. *Se as restrições são consideradas, a solução ótima permanece a mesma. Neste caso a localização da solução ótima é no interior da região das soluções viáveis, não sendo portanto afetada pelas restrições.*

Exemplo 16

$$
\begin{array}{ll}
Min & f(\mathbf{x}) = (x_1 - 2,0)^2 + (x_2 - 2,0)^2 \\
s.a & h_1(\mathbf{x}) = x_1 + x_2 - 2,0 \leqslant 0 \\
 & h_2(\mathbf{x}) = -\,x_1 + x_2 + 3,0 \leqslant 0 \\
 & h_3(\mathbf{x}) = -x_1 \leqslant 0 \\
 & h_4(\mathbf{x}) = -x_2 \leqslant 0
\end{array}
$$

Neste caso, o conjunto de restrições deste problema é o mesmo que o do problema anterior, acrescido da restrição $C_2(\mathbf{x}) = -\,x_1 + x_2 + 3,0 \leqslant 0$. *Esta restrição delimita a região abaixo da reta* $-x_1 + x_2 + 3,0 = 0$. *As curvas de contorno da função objetivo também possuem a mesma forma, estando porém centradas no ponto* $(2,0; 2,0)$. *A análise da região das soluções viáveis mostra que esta não é convexa. A solução deste problema de otimização é inviável.*

Esta situação é ilustrada na Figura 3.2

3.2 Restrições de Igualdade

Visando simplificar a derivação das condições de otimalidade para o problema de otimização restrito, considere inicialmente a minimização de uma função $f(\mathbf{x})$, tal que $\mathbf{x} \in R^n$ satisfaça m restrições $(m < n)$ de igualdade representadas pelas funções

$$g_i(\mathbf{x}) = 0, \quad i = 1, 2 \dots m$$

O conjunto de equações $g_i(\mathbf{x}) = 0, \quad i = 1, 2 \dots m$ define a região das soluções viáveis, a qual pertence a um subespaço de dimensão $(n - m)$.

Denotando $\mathbf{g}(\mathbf{x})$ a matriz cujos componentes são as m restrições $g_i(\mathbf{x}) = 0$; isto é,

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{bmatrix} \tag{3.3}$$

o problema de otimização com restrições de igualdade é expresso como

$$\begin{aligned} Min \quad & f(\mathbf{x}) \\ s.a \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0} \end{aligned} \tag{3.4}$$

A maioria dos algoritmos de solução do problema de otimização com restrições de igualdade são baseados na busca de uma solução que satisfaça determinadas condições de otimalidade. Nos problemas de otimização irrestritos, estas condições requerem que o vetor gradiente da função objetivo seja nulo e que a matriz Hessiana seja positiva semidefinida na solução ótima.

Conforme mencionado anteriormente, uma solução $\mathbf{x} \in R^n$ é dita viável com relação ao conjunto de restrições de igualdade do problema de otimização, se cada equação do conjunto $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ é satisfeita no ponto considerado. Caso contrário, a solução $\mathbf{x}$ é denominada inviável com relação ao conjunto de restrições de igualdade.

Uma solução $\mathbf{x}^*$ é denominada *mínimo local* do problema de otimização com restrições de igualdade se:

- $\mathbf{x}^*$ é viável com relação a todas as restrições de igualdade;
- existe uma vizinhança $V(\mathbf{x}^*)$, tal que

  $$f(\mathbf{x}^*) \leqslant f(\mathbf{x}) \ \text{ para todo } \mathbf{x} \in V(\mathbf{x}^*)$$

  Se $\mathbf{x}^*$ é viável e $f(\mathbf{x}^*) < f(\mathbf{x})$, a solução é classificada como *mínimo local forte*. Por outro lado, se $(f(\mathbf{x}^*) \leqslant f(\mathbf{x}))$ a solução é dita ser um *mínimo local fraco*.

Com o objetivo de verificar se a condição de mínimo local se aplica à uma solução viável, é necessário caracterizar a vizinhança $V(\mathbf{x}^*)$. Em geral, a viabilidade de uma solução em relação a um conjunto de equações não-lineares é mantida apenas por movimento ao longo de um caminho não-linear, denominado *arco viável*. Por exemplo, a restrição não-linear $x_1^2 + x_2^2 = 1$ representa um arco viável definido por um círculo de raio unitário no R^2, centrado na origem.

Se um arco viável existe com relação à uma solução viável $\mathbf{x}$, então a vizinhança desta solução ao longo deste arco contém pontos viáveis. A verificação da otimalidade de uma determinada solução requer o estabelecimento de uma condição indicando a existência do arco viável. Para isto, considere a expansão da equação $\mathbf{g}(\mathbf{x}) = 0$ em série de Taylor, no ponto $\mathbf{x}^*$ e ao longo da direção $\mathbf{p}$; isto é,

$$\mathbf{g}(\mathbf{x}^* + \mathbf{p}) = \mathbf{g}(\mathbf{x}^*) + \left.\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}^*} \mathbf{p} + \text{termos de ordem superior} \tag{3.5}$$

onde $\left.\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}^*}$ é a matriz de ordem $m \times n$, cujas linhas são os vetores gradiente das restrições de igualdade calculados no ponto onde é feita a expansão. Esta matriz é expressa analiticamente por

$$\left.\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}^*} = \mathbf{A}(\mathbf{x}^*) = \begin{bmatrix} \nabla g_1(\mathbf{x}^*) \\ \nabla g_2(\mathbf{x}^*) \\ \nabla g_3(\mathbf{x}^*) \\ \vdots \\ \nabla g_m(\mathbf{x}^*) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1(\mathbf{x}^*) \\ \mathbf{a}_2(\mathbf{x}^*) \\ \mathbf{a}_3(\mathbf{x}^*) \\ \vdots \\ \mathbf{a}_m(\mathbf{x}^*) \end{bmatrix}$$

onde $\mathbf{a}_i(\mathbf{x}^*) = \nabla g_i(\mathbf{x}^*)$ é um vetor linha de ordem n.

Desprezando os termos de ordem superior da expansão em série de Taylor, os pontos na vizinhança de $\mathbf{x}^*$ ao longo de qualquer direção de movimento $\xi\mathbf{p}$ serão viáveis apenas se a condição

$$\mathbf{A}(\mathbf{x}^*)\mathbf{p} = \mathbf{0} \tag{3.6}$$

for satisfeita.

Cada componente da Equação Matricial (3.6) ($\mathbf{a}_i(\mathbf{x}^*)\mathbf{p} = 0$) representa um hiperplano tangente à superfície definida pelo conjunto de restrições de igualdade no ponto $\mathbf{x}^*$. Portanto, a condição estabelecida pela Equação (3.6) pode ser interpretada geometricamente como o requisito de que o vetor gradiente das restrições seja normal ao plano tangente ao arco viável, ou em outras palavras à superfície definida pelas restrições de igualdade no ponto $\mathbf{x}^*$.

Se a condição estabelecida pela Equação (3.6) não é satisfeita, um movimento ao longo de qualquer caminho tangente a $\mathbf{p}$ pode resultar na violação de uma restrição. Assim, deve ser observado que se $\mathbf{g}(\mathbf{x})$ é acentuadamente não-linear, a condição imposta pela equação (3.6) é insuficiente para assegurar que $\mathbf{p}$ é tangente ao arco viável. O requisito que $\mathbf{g}(\mathbf{x})$ deve satisfazer de forma a assegurar a existência dos arcos viáveis é denominado *qualificação das restrições*.

Para uma solução viável qualquer $\mathbf{x}$, os requisitos de qualificação das restrições $\mathbf{g}(\mathbf{x})$ são satisfeitos se cada vetor não nulo satisfazendo a equação (3.6) é tangente a um arco duas vezes diferenciável originado em $\mathbf{x}$.

Quando a restrição é linear, a qualificação da mesma é imediata. Caso contrário, a equação (3.6) não representa rigorosamente a não linearidade da restrição, e portanto a análise da mesma pode conduzir a conclusões errôneas sobre a viabilidade da solução.

Um *ponto regular* é aquele para o qual as restrições $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ são simultaneamente satisfeitas e os vetores gradiente das restrições (linhas da matriz $\mathbf{A}(\mathbf{x})$) são linearmente independentes. Nesses pontos, a equação (3.6) caracteriza completamente os hiperplanos tangentes aos arcos viáveis. Esta equação requer que o vetor $\mathbf{p}$ pertença ao espaço nulo formado pelas linhas da matriz $\mathbf{A}(\mathbf{x})$ (vetores gradiente das restrições no ponto considerado).

Se $\mathbf{Z}(\mathbf{x})$ é uma matriz de ordem $n \times (n-m)$ cujas colunas são uma base do espaço nulo das linhas da matriz $\mathbf{A}(\mathbf{x})$, então $\mathbf{A}(\mathbf{x})\mathbf{Z}(\mathbf{x}) = \mathbf{0}$ e

$$\mathbf{A}(\mathbf{x})\mathbf{p} = \mathbf{0} \text{ se e somente se } \mathbf{p} = \mathbf{Z}(\mathbf{x})\mathbf{p}_z$$

onde $\mathbf{p}_z$ é um vetor arbitrário.

Resumindo, se $\mathbf{x}^*$ é um ponto de mínimo local de $f(\mathbf{x})$ no problema expresso pela Equação (3.4), então $\mathbf{x}^*$ é uma solução viável; isto é,

$$\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

e para qualquer outra solução viável, obtida à partir de $\mathbf{x}^*$ na direção $\mathbf{p}$; isto é, satisfazendo a condição $\mathbf{A}(\mathbf{x})\mathbf{p} = \mathbf{0}$, o valor da função objetivo não é menor do que aquele correspondente a $\mathbf{x}^*$. Isto é expresso analiticamente pela inequação

$$f(\mathbf{x}^* + \mathbf{p}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^t\mathbf{p} \quad \geq \quad f(\mathbf{x}^*)$$

o que implica em

$$\nabla f(\mathbf{x}^*)^t\mathbf{p} \geqslant 0, \quad \text{para todo } \mathbf{p}$$

onde $\nabla f(\mathbf{x}^*) = \left.\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}^*}$ é o vetor gradiente da função objetivo no ponto $\mathbf{x}^*$.

Desde que $\mathbf{A}(\mathbf{x})\mathbf{p} = \mathbf{0}$ e pela condição de mínimo local, o produto escalar $\nabla f(\mathbf{x}^*)^t\mathbf{p}$ é nulo na melhor das hipóteses. Então, o vetor gradiente da função objetivo pode ser expresso como uma combinação linear das linhas da matriz $\mathbf{A}(\mathbf{x})$; isto é,

$$\nabla f(\mathbf{x}^*) = \mathbf{A}(\mathbf{x}^*)^t\lambda$$

onde λ é um vetor arbitrário de ordem $m \times 1$. Desta forma, a condição $\nabla f(\mathbf{x}^*)^t\mathbf{p} \geqslant 0$ é satisfeita, pois

$$\nabla f(\mathbf{x}^*)^t\mathbf{p} = \left[\mathbf{A}(\mathbf{x}^*)^t\lambda\right]^t\mathbf{p} = \lambda^t\mathbf{A}(\mathbf{x}^*)\mathbf{p} = 0$$

Portanto, a condição necessária para que uma solução viável $\mathbf{x}^*$ seja um ponto de mínimo local do problema de otimização com restrições de igualdade representado pela equação (3.4), é que

$$\nabla f(\mathbf{x}^*) = \mathbf{A}(\mathbf{x}^*)^t\lambda$$

onde, λ é um vetor formado pelos escalares λ_i, denominados *multiplicadores de Lagrange.* Esta condição também pode ser escrita como

$$\mathbf{Z}(\mathbf{x}^*)^t\nabla f(\mathbf{x}^*) = \mathbf{0}$$

pois $\mathbf{A}(\mathbf{x}^*)\mathbf{Z}(\mathbf{x}^*) = \mathbf{0}$, conforme definido anteriormente.

Exercício 1 *: Analise as soluções viáveis da região definida pelas funções*

$$\begin{aligned} c_1(\mathbf{x}) = &\ x_1 + x_2 + 3x_3 - 2 = 0 \\ c_2(\mathbf{x}) = &\ 5x_1 + 2x_2 + x_3 - 5 = 0 \end{aligned}$$

3.2.1 Função Lagrangeana

Quando apenas restrições de igualdade são consideradas no problema de otimização, é conveniente analisar o mesmo sob o ponto de vista de um problema irrestrito, o que é possível através da função Lagrangeana.

A função Lagrangeana correspondente ao problema expresso pela equação (3.4),

$$\begin{aligned} Min \quad & f(\mathbf{x}) \\ s.a \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0} \end{aligned}$$

é dada por

$$\pounds(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda^t \mathbf{g}(\mathbf{x})$$

onde λ é o vetor $(m \times 1)$, dos multiplicadores de Lagrange.

Observe que, na solução ótima $(\mathbf{x}^*, \lambda^*)$ o valor crítico de $\pounds(\mathbf{x}^*, \lambda^*)$ é apenas $f(\mathbf{x}^*)$ pois $g_i(\mathbf{x}^*) = 0$ para $i = 1, \ldots, m$.

Os pontos estacionários desta função irrestrita são determinados aplicando-se as condições de otimalidade de 1^a ordem; isto é, derivando-se a função Lagrangeana $\pounds(\mathbf{x}, \lambda)$ e igualando-se o resultado a zero. Isto fornece

$$\frac{\partial \pounds(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = \nabla_x \pounds(\mathbf{x}, \lambda) = \mathbf{0} = \nabla_x f(\mathbf{x}) - \nabla_x \mathbf{g}(\mathbf{x})^t \lambda$$

e

$$\frac{\partial \pounds(\mathbf{x}, \lambda)}{\partial \lambda} = \nabla_\lambda \pounds(\mathbf{x}, \lambda) = \mathbf{0} = -\mathbf{g}(\mathbf{x})$$

onde $\nabla_x \mathbf{g}(x) = \mathbf{A}(\mathbf{x})$ é uma matriz $(m \times n)$, cuja $i - \acute{e}sima$ linha é o vetor gradiente da função $g_i(\mathbf{x})$, conforme definido anteriormente.

A primeira equação pode ser escrita como

$$\nabla_x f(\mathbf{x}) - \mathbf{A}(\mathbf{x})^t \lambda = \mathbf{0}$$

o que resulta em

$$\nabla_x f(\mathbf{x}) = \mathbf{A}(\mathbf{x})^t \lambda$$

isto é, em um ponto estacionário o vetor gradiente é expresso como uma combinação linear das colunas de $\nabla_x \mathbf{g}(\mathbf{x}) = \mathbf{A}(\mathbf{x})$.

Exemplo 17 *Determinar os pontos estacionários da função* $f(\mathbf{x}) = 3x_1^2 + 4x_2^2$ *sujeito a restrição* $g(\mathbf{x}) = 3,5x_1 + 4x_2 - 14 = 0$.

Os vetores gradiente da função objetivo e das restrições são, respectivamente

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 6x_1 \\ 8x_2 \end{bmatrix}$$

$$\nabla g(\mathbf{x}) = \begin{bmatrix} 3,5 \\ 4,0 \end{bmatrix}$$

De acordo com a condição de otimalidade

$$\begin{bmatrix} 6x_1 \\ 8x_2 \end{bmatrix} = \begin{bmatrix} 3,5 \\ 4,0 \end{bmatrix} \lambda$$

tal que, $3,5x_1 + 4x_2 - 14 = 0$.

A solução deste sistema linear fornece $\lambda = -3,46$, $x_1 = 2,02$ *e* $x_2 = 1,73$.

A ilustração geométrica deste problema é mostrada na Figura 3.3

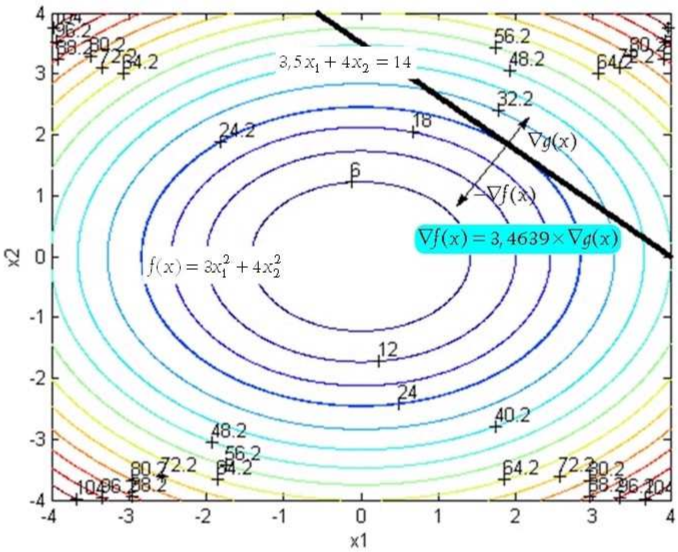

Figura 3.3: Representação do problema do exemplo 17

Exemplo 18 *Determinar os pontos estacionários da função* $f(\mathbf{x}) = 2x_1 + 3x_2$ *sujeito a restrição* $g(\mathbf{x}) = x_1^2 + 1,5x_2^2 - 6 = 0$.

3.2.2 Multiplicadores de Lagrange

Visando facilitar a interpretação geométrica dos multiplicadores de Lagrange, suponha que o conjunto de m restrições de igualdade seja expresso por um sistema de equações lineares representadas pela equação ($\mathbf{Ax}^* = \mathbf{b}$, onde $\mathbf{A}$ possui m linhas linearmente independentes), e que na solução ótima $\mathbf{x}^*$ com $\mathbf{x} \in R^n$, o valor da função objetivo seja $f(\mathbf{x}^*)$.

A condição de otimalidade de primeira ordem estabelece que o gradiente da função objetivo deve ser expresso como uma combinação linear dos gradientes das restrições; ou seja,

$$\nabla_x f(\mathbf{x}^*) = \mathbf{A}^t \lambda$$

onde λ e o vetor $(m \times 1)$ dos multiplicadores de Lagrange.

Desde que *posto* $(\mathbf{A}) = m$, é possível particionar a matriz $\mathbf{A}$ em duas submatrizes $\mathbf{A}_1$, de ordem $(m \times m)$ e $\mathbf{A}_2$, de ordem $m \times (n - m)$, sendo $\mathbf{A}_1$ não singular, tal que

$$[\mathbf{A}_1 | \mathbf{A}_2] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \mathbf{b} \tag{3.7}$$

onde o vetor

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$$

é particionado de forma semelhante a da matriz $\mathbf{A}$.

Desde que $\mathbf{A}_1$ é não singular,

$$\mathbf{x}_1 = \mathbf{A}_1^{-1}(\mathbf{b} - \mathbf{A}_2\mathbf{x}_2)$$

tal que a função objetivo é expressa como

$$f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2) = f[\mathbf{A}_1^{-1}(\mathbf{b} - \mathbf{A}_2\mathbf{x}_2), \mathbf{x}_2]$$

e o vetor gradiente de $f(x)$ particionado nesta mesma base fornece

$$\nabla_x f(\mathbf{x}) = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$$

onde $\mathbf{g}_1 = \nabla_{x_1} f(\mathbf{x})$ e $\mathbf{g}_2 = \nabla_{x_2} f(\mathbf{x})$.

A equação $\mathbf{A}^t\lambda = \nabla_x f(\mathbf{x})$ pode ser re-escrita como,

$$\begin{bmatrix} \mathbf{A}_1^t \\ \mathbf{A}_2^t \end{bmatrix} \lambda = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$$

tal que

$$\lambda = \mathbf{A}_1^{-t}\mathbf{g}_1$$

e

$$\mathbf{g}_2 = \mathbf{A}_2^t \mathbf{N}_1^{-t} \mathbf{g}_1$$

A derivada da função objetivo $f[\mathbf{A}_1^{-1}(\mathbf{b} - \mathbf{A_2 x_2}), \mathbf{x_2}]$ com relação a $\mathbf{b}$ é

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{b}} = \mathbf{A}_1^{-t}\mathbf{g}_1 = \lambda$$

ou seja, os multiplicadores de Lagrange fornecem na solução ótima informações sobre a sensibilidade instantânea da função objetivo com relação a pertubações nas restrições.

Exemplo 19 *: Minimizar a função $f(x_1, x_2)$ sujeito a restrição $g(x_1, x_2) = 0$.*

Aplicação a Sistemas Elétricos de Potência 1 *Considere dois geradores suprindo potência a uma carga de 3 MW. Os geradores possuem curvas de custo de geração de potência ativa dadas por $C_1(P_1) = 1,10P_1^2$ e $C_1(P_2) = 0,88P_2^2$. Determinar a potência de saída de cada gerador na solução de mínimo custo de geração. As perdas de potência ativa nas linhas de transmissão são desprezíveis.*

3.3 Condições de Suficiência

As condições de otimalidade necessárias para que uma solução viável $\mathbf{x}^*$ seja um ponto de mínimo do problema de otimização com restrições de igualdade são:

1. o gradiente da função objetivo deve ser expresso como uma combinação linear dos vetores gradiente das restrições no ponto $\mathbf{x}^*$, isto é

$$\nabla f(\mathbf{x}^*) = \mathbf{A}(\mathbf{x}^*)^t \lambda^*$$

2. para quaisquer vetores $\mathbf{p}$ e λ^* satisfazendo respectivamente as condições $\mathbf{A}(\mathbf{x}^*)^t\mathbf{p} = \mathbf{0}$ e $\nabla f(\mathbf{x}^*) = \mathbf{A}(\mathbf{x}^*)^t\lambda^*$, então $\mathbf{p}^t\nabla^2\mathcal{L}(\mathbf{x}^*, \lambda^*)\mathbf{p} \geqslant \mathbf{0}$.

A segunda condição implica no requisito de que a função Lagrangeana $\pounds(\mathbf{x}, \lambda^*)$ seja convexa em $\mathbf{x}$ e λ. Em outras palavras, se as funções $f(\cdot)$ e $g_i(\cdot)$ são duas vezes diferenciáveis, a matriz Hessiana $\nabla^2 \pounds(\mathbf{x}^*, \lambda^*)$ deve ser positiva semi-definida.

Se $\nabla^2 \pounds(\mathbf{x}^*, \lambda^*)$ é positiva definida, estas condições são suficientes para assegurar que $(\mathbf{x}^*, \lambda^*)$ é um ponto de mínimo de $\pounds(\mathbf{x}, \lambda)$ e de $f(\mathbf{x})$, pois $\mathbf{p}^t \nabla^2 \pounds(\mathbf{x}^*, \lambda^*)\mathbf{p} > 0$ para qualquer $\mathbf{p}$ e $\mathbf{p}^t \mathbf{a}_i(\mathbf{x}^*) = 0$. Desde que $\mathbf{p} = \mathbf{Z}\mathbf{p}_z$ ($\mathbf{Z}$ é a matriz de espaço nulo das linhas de $\mathbf{A}(\mathbf{x}^*)$), então a segunda condição requer que a matriz

$$\mathbf{Z}(\mathbf{x}^*)^t \nabla^2 \pounds(\mathbf{x}^*, \lambda^*)\mathbf{Z}(\mathbf{x}^*)$$

denominada *Hessiana reduzida da função Lagrangeana*, seja positiva definida.

Exemplo 20 : *Verificar as condições de otimalidade dos seguintes problemas:*

- $$\begin{aligned} Min \quad & f(\mathbf{x}) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8x_1 \\ s.a \quad & x_1 + x_2 = 4,0 \end{aligned}$$

- $$\begin{aligned} Min \quad & f(\mathbf{x}) = (x_1 - 1,0)^2 + (x_2 - 1,0)^2 \\ s.a \quad & x_1 + x_2 - 4,0 = 0,0 \\ & x_1 - x_2 - 2,0 = 0,0 \end{aligned}$$

- $$\begin{aligned} Min \quad & f(\mathbf{x}) = 2x_1 + 3x_2 \\ s.a \quad & x_1^2 + 1,5x_2^2 - 6,0 = 0 \end{aligned}$$

3.4 Restrições de Desigualdade

Suponha que o problema de otimização a ser resolvido é representado por

$$\begin{aligned} Min \quad & f(\mathbf{x}) \\ s.a \quad & \mathbf{h}(\mathbf{x}) \geq \mathbf{0} \end{aligned}$$

onde as inequações $\mathbf{h}(\mathbf{x}) \geqslant \mathbf{0}$ representam as restrições de desigualdade.

Uma solução é considerada *viável* se o conjunto de restrições $\mathbf{h}(\mathbf{x}) \geqslant \mathbf{0}$ é simultaneamente satisfeito. Caso alguma restrição não seja satisfeita, a solução é dita *inviável*. Portanto as equações do conjunto $\mathbf{h}(\mathbf{x}) \geqslant \mathbf{0}$ definem a *região das soluções viáveis.*

Se $\mathbf{x}^*$ é um ponto viável e mínimo local de $f(\mathbf{x})$, então,

1. $\mathbf{x}^*$ pertence à região das soluções viáveis definida pelo conjunto de restrições $\mathbf{h}(\mathbf{x}) \geqslant \mathbf{0}$;

2. não existe nenhum vetor $\mathbf{p}$, tal que

   - $\mathbf{x} = \mathbf{x}^* + \mathbf{p} \in S$; e
   - $\nabla f(\mathbf{x}^*)^t \mathbf{p} < \mathbf{0}$.

O significado destas condições é que qualquer movimento a partir de $\mathbf{x}^*$ não tem componente negativa $\nabla f(\mathbf{x})^t\mathbf{p}$, e portanto o valor da função objetivo não pode ser reduzido.

Com relação ao conjunto de inequações do problema de otimização, a restrição de desigualdade $h_i(\mathbf{x}) \geq 0$ é denominada *restrição ativa* se ela estiver no limite no ponto $\mathbf{x}^*$, isto é $h_i(\mathbf{x}^*) = 0$. Caso contrário, se $h_i(\mathbf{x}) > 0$, a inequação é chamada *restrição inativa*.

Quanto à localização da solução ótima, as duas seguintes possibilidades são consideradas:

1. $\mathbf{x}^*$ não está no limite da região das soluções viáveis. Isto implica em que não há nenhuma restrição de desigualdade ativa, ou seja

$$\mathbf{h}(\mathbf{x}^*) > \mathbf{0} \quad (h_i(\mathbf{x}^*) > 0, \quad i = 1, 2 \ldots m)$$

   Então existe uma vizinhança de $\mathbf{x}^* \in S$, e para que $\mathbf{x}^*$ seja um ponto estacionário é necessário que

$$\nabla f(\mathbf{x}^*) = 0$$

2. $\mathbf{x}^*$ é tal que existe uma ou mais restrições ativas (por exemplo, $h_i(\mathbf{x}^*) = 0$ para algum $i \in I$). O conjunto I, chamado *conjunto de restrições ativas*, é em geral desconhecido, exceto no caso onde existem apenas restrições de igualdade.

3.4.1 Condições de Karush-Kuhn-Tucker

Seja o ponto $\mathbf{x}^*$ para o qual,

$$h_i(\mathbf{x}^*) = 0, \quad i \in I$$

$$h_i(\mathbf{x}^*) > 0, \ i \notin I$$

Se os vetores $\mathbf{a}_i(\mathbf{x}^*) = \nabla h_i(\mathbf{x}^*)$, $i \in I$ são linearmente independentes e se $\mathbf{x}^*$ é um ponto de *mínimo local forte* de $f(\mathbf{x}^*)$ sujeito as restrições $h_i(\mathbf{x}) \geqslant 0$, então:

1. $\nabla f(\mathbf{x}^*) = \mathbf{A}(\mathbf{x}^*)^t\lambda^*$, onde as linhas da matriz $\mathbf{A}$ são os vetores gradiente das restrições ativas;

2. $\lambda_i^* = 0$ para as restrições inativas $(h_i(\mathbf{x}^*) > 0, \ i \notin I)$;

3. $\lambda_i^* h_i(\mathbf{x}^*) = 0$ (condição de complementaridade), com $h_i(\mathbf{x}^*) \geq 0$;

4. $\lambda^* > \mathbf{0}$ para as restrições ativas $(h_i(\mathbf{x}^*) = 0, \ i \in I)$.

Essas condições, as quais estabelecem a suficiência para que a solução de um problema de otimização com restrições de desigualdade seja ótima, são chamadas *Condições de Karush-Kuhn-Tucker* para o ponto de mínimo.

3.4.2 Interpretação da Condição $\lambda > 0$

Seja $\mathbf{p}$ uma direção de movimento a partir da solução $\mathbf{x}^*$, tal que todas as restrições do conjunto I permanecem ativas, com exceção da $k - ésima$ restrição, a qual permanece viável mas se torna inativa. Então no ponto $\mathbf{x}^* + \mathbf{p}$,

$$\mathbf{p}^t\mathbf{a}_i(\mathbf{x}^*) = 0, \quad i \in I, \quad i \neq k$$

$$h_k(\mathbf{x}^* + \mathbf{p}) \geq h_k(\mathbf{x}^*) = 0$$

tal que

$$\mathbf{p}^t\mathbf{a}_k(\mathbf{x}^*) \geq 0$$

A expansão em série de Taylor de $f(x)$ na direção $\mathbf{p}$ e em torno de $\mathbf{x}^*$ até o termo de 1^a ordem fornece

$$\begin{aligned} f(\mathbf{x}^* + \mathbf{p}) &= f(\mathbf{x}^*) + \mathbf{p}^t\nabla f(\mathbf{x}^*) \\ &= f(\mathbf{x}^*) + \mathbf{p}^t\mathbf{A}(\mathbf{x}^*)^t\lambda^* \\ &= f(\mathbf{x}^*) + \mathbf{p}^t\lambda_k^*\mathbf{a}_k(\mathbf{x}^*) \end{aligned}$$

pois $\mathbf{p}^t\mathbf{a}_i(\mathbf{x}^*) = 0$ para $i \in I$, de forma que a expressão resultante é

$$f(\mathbf{x}^* + \mathbf{p}) = f(\mathbf{x}^*) + \mathbf{p}^t\lambda_k^*\mathbf{a}_k(\mathbf{x}^*)$$

sendo que $\mathbf{p}^t\mathbf{a}_k(\mathbf{x}^*) \geq 0$ e então, se $\lambda_k^* > 0$ o 2^o termo sempre resultará em acréscimo no valor da função objetivo.

As condições de Kuhm-Tucker são suficientes em primeira ordem para assegurar que $\mathbf{x}^*$ é ponto de mínimo, pois é possível provar que não existe vetor $\mathbf{p}$ satisfazendo simultaneamente as inequações $\mathbf{p}^t\nabla f(\mathbf{x}^*) < 0$ e $\mathbf{p}^t\mathbf{a}_i(\mathbf{x}^*) \geq 0$.

A condição $\lambda_k^* > 0$ mostrada acima, pode também ser explicada a partir do fato de que

$$\frac{\partial f}{\partial b_k} = \lambda_k \quad (quando \quad h_k(\mathbf{x}) = b_k)$$

Um aumento em b_k moverá o correspondente limite para o interior da região viável.

Se $\lambda_k^* > 0$, desde que

$$\frac{\partial f}{\partial b_k} = \lambda_k$$

a única forma de reduzir o valor da função objetivo com relação a restrição $h_k(\mathbf{x}) \geq 0$ seria diminuindo b_k, tal que $\partial b_k < 0$. Entretanto, isto poderia resultar numa solução inviável. Então $\lambda_k^* \leq 0$ indica que, se a restrição $h_k(\mathbf{x}) \geq b_k$ é ativa, não é possível reduzir o valor da função objetivo via manipulação da restrição h_k. Daí, a condição

$$\frac{\partial f(\mathbf{x}^*)}{\partial b_k} = \lambda_k^*$$

com

- $\lambda_k^* > 0$, para as restrições ativas;
- $\lambda_k^* = 0$, para as restrições não ativas.

A particularização do estabelecimento das condições de Karush-Kuhn-Tucker para o caso do problema de otimização com restrições de desigualdade lineares é mostrada a seguir.

3.5 Otimização com Restrições Lineares

Seja o seguinte problema de otimização, no qual as restrições são representadas por um conjunto de desigualdades.

$$\begin{array}{ll} \textit{Minimizar} & f(\mathbf{x}) \\ \textit{sujeito a} & \mathbf{Ax} \geq \mathbf{b} \end{array} \tag{3.8}$$

onde, $f(\mathbf{x})$ é a função objetivo a ser otimizada, $\mathbf{A}$ é a matriz de coeficiente das restrições formada por vetores linha $\mathbf{a}_i^t$, $\mathbf{b}$ é o vetor do lado direito com componentes b_i e $\mathbf{x}$ é o vetor das variáveis de otimização. A determinação de uma solução viável para este problema requer uma distinção preliminar entre as inequações que representam as restrições. Para um dado ponto $\mathbf{x}$ é possível distinguir três situações na qual podem estar as mesmas:

- restrições satisfeitas, mas não no limite, isto é, $\mathbf{a}_i^t\mathbf{x} > b_i$, chamadas restrições inativas;
- restrições satisfeitas e no limite, isto é, $\mathbf{a}_i^t\mathbf{x} = b_i$, chamadas restrições ativas;
- não satisfeitas, isto é, $\mathbf{a}_i^t\mathbf{x} < b_i$, chamadas restrições violadas.

Em termos gerais duas situações podem ser previstas com relação a solução inicial no problema de otimização exposto na equação (3.8):

- solução inicial viável, isto é, $\mathbf{x}^0$ tal que $\mathbf{Ax}^0 \geq \mathbf{b}$, ou seja, $\mathbf{x}^0$ para o qual todas as restrições são satisfeitas;
- solução inicial inviável, isto é, $\mathbf{x}^0$ para o qual existem restrições i tal que $\mathbf{a}_i^t\mathbf{x}^0 < b_i$

Naturalmente, é mais simples resolver o problema representado pela equação (3.8) a partir de uma solução inicial viável. Entretanto, nem sempre é possível se dispor de tal solução.

Nas seções subseqüentes mostra-se os movimentos à partir de uma solução inicial viável com o objetivo de se estabelecer as condições de otimalidade do problema expresso pela equação (3.8).

3.5.1 Movimentos a Partir de um Ponto Inicial Viável

Seja $\mathbf{x}^k$ uma solução viável. Se a j-ésima restrição é inativa porém satisfeita neste ponto, ou seja $\mathbf{a}_j^t\mathbf{x}^k > b_j$, então os movimentos a partir de $\mathbf{x}^k$ são possíveis em qualquer direção sem violar a restrição considerada. Isto é, para qualquer vetor $\mathbf{p}$, $\mathbf{x}^k + \xi\mathbf{p}$ será viável para uma escolha adequada de $\parallel \xi \parallel$.

Uma restrição ativa, por outro lado, restringe os movimentos a partir de um ponto viável. Se a i-ésima restrição é tal que $\mathbf{a}_i^t\mathbf{x}^k = b_i$, existem dois tipos de movimento que manterão a restrição viável. Se $\mathbf{p}$ satisfaz

$$\mathbf{a}_i^t\mathbf{p} = \mathbf{0} \tag{3.9}$$

a direção $\mathbf{p}$ corresponde a um movimento ao longo da restrição i, e esta permanece ativa para todos os pontos $\mathbf{x}^k + \alpha\mathbf{p}$ para qualquer α. Se $\mathbf{p}$ é tal que $\mathbf{a}_i^t\mathbf{p} > \mathbf{0}$, então

$$\mathbf{a}_i^t(\mathbf{x}^k + \alpha\mathbf{p}) = b_i + \alpha\mathbf{a}_i^t\mathbf{p} > b_i \quad se \quad \alpha > 0 \tag{3.10}$$

isto é, a i-ésima restrição se torna inativa no ponto $\mathbf{x}^k + \alpha\mathbf{p}$. O vetor $\mathbf{p}$ representa, portanto, um movimento com a direção apontando para o interior da região viável.

3.5.2 Análise da Condição de Otimalidade

Para determinar se o ponto $\mathbf{x}^*$ é a solução ótima do problema de otimização com restrições de desigualdade lineares, deve-se inicialmente identificar as restrições ativas em $\mathbf{x}^*$. Para isto, seja $\hat{\mathbf{A}}$ uma matriz de ordem $(t \times n)$ cujas linhas são os coeficientes das restrições ativas em $\mathbf{x}^*$. A condição necessária para que $\mathbf{x}^*$ seja a solução ótima é

$$\nabla f(\mathbf{x}^*) = \hat{\mathbf{A}}^t\lambda \tag{3.11}$$

onde, $\nabla f(\mathbf{x}^*)$ é o gradiente da função Lagrangeana calculado no ponto $\mathbf{x}^*$, e λ é o vetor dos multiplicadores de Lagrange das restrições.

Esta condição assegura que $f(\cdot)$ é estacionária para todos os movimentos de $\mathbf{x}^*$ ao longo das restrições ativas. Entretanto, desde que os movimentos à partir de $\mathbf{x}^*$ direcionados para o interior da região viável correspondem também a soluções viáveis, o ponto $\mathbf{x}^*$ não será ótimo se houver alguma direção $\mathbf{p}$ de movimento descendente, isto é, ao longo do qual $f(\cdot)$ decresça. Para evitar esta possibilidade, é necessário assegurar que, para qualquer vetor $\mathbf{p}$ satisfazendo $\hat{\mathbf{A}}\mathbf{p} \geq 0$, $\nabla^t f(\mathbf{x}^*)\mathbf{p} \geq 0$. Desde que, pela equação (3.11)

$$\nabla^t f(\mathbf{x}^*)\mathbf{p} = \lambda_1\hat{\mathbf{a}}_1^t\mathbf{p} + \lambda_2\hat{\mathbf{a}}_2^T\mathbf{p} + \cdots + \lambda_t\hat{\mathbf{a}}_t^t\mathbf{p} \tag{3.12}$$

onde, t é o número de restrições ativas; a condição desejada é representada por

$$\nabla^t f(\mathbf{x}^*)\mathbf{p} = \lambda_1\hat{a}_1^t\mathbf{p} + \lambda_2\hat{a}_2^T\mathbf{p} + \cdots + \lambda_t\hat{a}_t^T\mathbf{p} \geq 0 \tag{3.13}$$

onde, $\hat{a}_i^t\mathbf{p} \geq 0, i = 1, 2, \cdots, t$

Esta condição é satisfeita somente se $\lambda_i \geq 0, i = 1, 2, \cdots, t$, isto é, $\mathbf{x}^*$ não será a solução ótima se existirem multiplicadores de Lagrange negativos, pois isto indica que ainda existe alguma direção ao longo da qual é possível reduzir o valor da função objetivo.

Para se estabelecer o conjunto total de condições de otimalidade para o problema definido pela equação (3.8), deve-se fazer a imposição adicional de que as t linhas da matriz $\hat{\mathbf{A}}$ são linearmente independentes e que $\mathbf{Z}$ é uma matriz cujas colunas formam uma base para o conjunto de vetores ortogonais as linhas de $\hat{\mathbf{A}}$. Neste caso, cada vetor $\mathbf{p}$ tal que $\hat{\mathbf{A}}\mathbf{p} = 0$ pode portanto ser escrito como uma combinação linear das colunas de $\mathbf{Z}$, isto é, $\mathbf{p} = \mathbf{Z}\mathbf{p}_z$ onde, $\mathbf{p}_z$ é um vetor arbitrário. A expansão da função $f(\mathbf{x})$ no ponto $\mathbf{x}^*$ ao longo da direção $\mathbf{p}$ (para a qual $\hat{\mathbf{A}}\mathbf{p} = 0$ e $\mathbf{p} = \mathbf{Z}\mathbf{p}_z$) é dada por

$$f(\mathbf{x}^* + \xi\mathbf{Z}\mathbf{p}_z) = f(\mathbf{x}^*) + \xi\mathbf{p}_z^t\mathbf{Z}^t\nabla f(\mathbf{x}^*) + \frac{1}{2}\xi^2\mathbf{p}_z^t\mathbf{Z}^t\mathbf{G}(\mathbf{x}^*)\mathbf{Z}\mathbf{p}_z \tag{3.14}$$

onde, $\mathbf{G}(\mathbf{x}^*)$ é a matriz de segunda derivada de $f(\cdot)$ calculada em $\mathbf{x}^*$. A análise desta equação mostra que uma condição adicional para que $\mathbf{x}^*$ seja a solução ótima é que a matriz $\mathbf{Z}^t\mathbf{G}(\mathbf{x}^*)\mathbf{Z}$ seja positiva definida, pois neste caso $\frac{1}{2}\xi^2\mathbf{p}_z^T\mathbf{Z}^t\mathbf{G}(\mathbf{x}^*)\mathbf{Z}\mathbf{p}_z \geq 0$ para qualquer $\mathbf{p}$, isto é, não há direção ao longo da qual o valor da função objetivo possa ser reduzido. Portanto, as condições de otimalidade para o problema de otimização expresso pela equação (3.8) podem ser sumarizadas como:

1. $\mathbf{a}_i^t\mathbf{x}^* > b_i$, para as restrições inativas e $\mathbf{a}_i^t\mathbf{x}^* = b_i$, para as restrições ativas;
2. $\nabla f(\mathbf{x}^*) = \hat{\mathbf{A}}^t\lambda$, onde, $\hat{\mathbf{A}}$ corresponde as restrições ativas;
3. $\lambda > 0$ para as restrições ativas;
4. $\mathbf{Z}^t\mathbf{G}(\mathbf{x}^*)\mathbf{Z}$ positiva definida.

Essas condições são chamadas *Condições de Karush-Kuhn-Tucker* para o problema de otimização com restrições lineares expresso em (3.8). É comum ainda, atribuir multiplicadores de Lagrange nulos as restrições inativas, tal que a condição 3 poderia ser expressa de forma generalizada por:

- $\lambda_i > 0$ para restrições ativas
- $\lambda_i = 0$ para restrições inativas

3.6 Exercícios

Exercício 2 *: Determinar os pontos estacionários e analisar as condições de otimalidade dos seguintes problemas:*

1.

$$\begin{array}{ll} Min & f(\mathbf{x}) = 2x_1 + 3x_2 - x_1^3 - 2x_2^2 \\ s.a & x_1 + 3x_2 - 6,0 \leqslant 0 \\ & 5x_1 + 2x_2 - 10,0 \leqslant 0 \\ & x_1, x_2 \geqslant 0 \end{array}$$

2.

$$\begin{array}{ll} Min & f(\mathbf{x}) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8x_1 \\ s.a & x_1 + x_2 - 4,0 \leqslant 0 \end{array}$$

3.

$$\begin{array}{ll} Min & f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 - 2x_2 + 6,0 \\ s.a & x_1 + x_2 - 4,0 \geqslant 0 \end{array}$$

4.

$$\begin{array}{ll} Min & f(\mathbf{x}) = 2x_1^2 + 9x_2^2 - 6x_1x_2 - 18x_1 + 9x_2 \\ s.a & x_1 + 2x_2 - 10,0 \leqslant 0 \\ & 4x_1 - 3x_2 - 20,0 \leqslant 0 \\ & x_1, x_2 \geqslant 0 \end{array}$$

5.

$$\begin{array}{ll} Min & f(\mathbf{x}) = 9x_1^2 + 13x_2^2 - 18x_1x_2 - 4,00 \\ s.a & x_1^2 + x_2^2 + 2x_1 - 16,0 \geqslant 0 \end{array}$$

6.

$$\begin{array}{ll} Min & f(\mathbf{x}) = x_1^3 - 16x_1 + 2x_2 - 3x_2^2 \\ s.a & x_1 + x_2 - 3,0 \leqslant 0 \end{array}$$

7.

$$\begin{array}{ll} Min & f(\mathbf{x}) = 3x_1^2 - 2x_1x_2 + 5x_2^2 + 8x_2 \\ s.a & x_1^2 - x_2^2 + 8x_2 - 16,0 \leqslant 0 \end{array}$$

Exercício 3 *: Encontre o ponto estacionário* $\mathbf{x}^*$ *da função*

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

sujeito a

$$c_1(\mathbf{x}) = x_1 + x_2 + 3x_3 - 2 = 0$$

$$c_2(\mathbf{x}) = 5x_1 + 2x_2 + x_3 - 5 = 0$$

e determine a sua natureza. Verifique se o vetor gradiente de $f(\mathbf{x})$ *pode ser escrito como uma combinação linear dos vetores gradientes das restrições* $c_1(\mathbf{x})$ *e* $c_2(\mathbf{x})$.

Exercício 4 *: Mostre que* $\mathbf{x} = (1,0; 1,0; 1,0)$ *é um ponto estacionário da função*

$$f(\mathbf{x}) = x_1^2 + 2x_2^2 + 4x_3^2 + 5x_1x_2$$

sujeito as restrições

$$c_1(\mathbf{x}) = x_1 + x_2^2 + 3x_2x_3 - 5 = 0$$

$$c_2(\mathbf{x}) = x_1^2 + 5x_1x_2 + 3x_3^2 - 9 = 0$$

e determine a sua natureza.

Exercício 5 *: Verifique que a função*

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

sujeito a

$$c_1(\mathbf{x}) = x_1^2 - 4x_2 = 0$$

$$c_2(\mathbf{x}) = x_1 - x_2 - 1 = 0$$

tem um mínimo em $\mathbf{x}^* = (2,0; 1,0; 0,0)$*, porém o gradiente da função objetivo não pode ser expresso como uma combinação linear dos vetores gradientes das restrições.*

Exercício 6 *: Estabeleça as condições de Kuhn-Tucker para o problema*

$$Minimizar \quad f(\mathbf{x}) = -x_1^3 + x_2^2 - 2x_1x_3^2$$

sujeito a

$$c_1(\mathbf{x}) = 2x_1 + x_2^2 + x_3 - 5 = 0$$

$$c_2(\mathbf{x}) = 5x_1^2 - x_2^2 - x_3 \geq 2$$

$$c_3(\mathbf{x}) = x_1 \geq 0$$

$$c_4(\mathbf{x}) = x_2 \geq 0$$

$$c_5(\mathbf{x}) = x_3 \geq 0$$

e verifique se elas são satisfeitas no ponto $\mathbf{x} = (1,0; 0,0; 3,0)$.

Exercício 7 : *Estabeleça as condições de Kuhn-Tucker para o problema*

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

sujeito a

$$c_1(\mathbf{x}) = -x_1 + x_2 - x_3 \geq -10,0$$
$$c_2(\mathbf{x}) = x_1 + x_2 + 4x_3 \geq 20,0$$

e obtenha a solução deste problema através de qualquer método.

Exercício 8 : *Para o problema*

$$Minimizar \quad f(\mathbf{x}) = (x_1 - 1)^2 + (x_2 - 1)^2$$

sujeito a

$$c_1(\mathbf{x}) = x_1 + x_2 \leq 0,50$$
$$x_1 \geq 0,0$$

1. *plotar os contornos de* $f(\mathbf{x}) = 0,0$, $f(\mathbf{x}) = 1,0$ *e* $f(\mathbf{x}) = 4,0$ *e as restrições do problema de otimização;*
2. *identificar o mínimo irrestrito na Figura;*
3. *identificar o mínimo restrito na Figura;*
4. *calcular os multiplicadores de Lagrange.*

Exercício 9 : *Seja o problema*

$$Minimizar \quad f(\mathbf{x}) = x_1 + x_2$$

sujeito a

$$2 - x_1^2 - x_2 \leq 0,0$$
$$4 - x_1 - 3x_2 \leq 0,0$$
$$-30 + x_1 + x_2^2 \leq 0,0$$

1. *verificar se as condições de otimalidade de primeira ordem são satisfeitas no ponto* $(1,0; 1,0)$*;*
2. *calcular os multiplicadores de Lagrange neste ponto.*

Aplicação a Sistemas Elétricos de Potência 2 *Determine a potência de saída de três geradores, que resulta no mínimo custo de geração de potência ativa, quando uma carga de 800 MW é suprida. Considere as perdas de transmissão desprezíveis. As curvas de custo de geração são quadráticas, dadas por*

$$C_1(P_1) = 300 + 7,3P_1 + 0,001P_1^2$$

$$C_2(P_2) = 150 + 7,8P_2 + 0,002P_2^2$$

$$C_3(P_3) = 75 + 7,5P_3 + 0,005P_3^2$$

não havendo restrição na geração de potência.

Aplicação a Sistemas Elétricos de Potência 3 *Verifique as condições de otimalidade do problema da aplicação 2, considerando que as perdas de potência ativa na transmissão são dadas por* $P_l = 0,00003P_1^2 + 0,00005P_2^2 + 0,00007P_3^2$.

Capítulo 4

Fundamentos de Programação Linear

Este capítulo apresenta a classe de problemas no qual a função objetivo e as restrições são funções lineares da variável de otimização, os quais são denominados problemas de Programação Linear. Apesar da maioria dos problemas de interesse prático da análise de sistemas de potência não possuir esta característica, a discussão deste tipo de problema é importante por várias razões. Em geral, é possível simplificar a obtenção da solução de um problema de otimização não-linear, através da linearização sequencial do mesmo e da aplicação de algoritmos relativamente mais simples, envolvendo apenas sistemas lineares. Além disso, algoritmos de Programação Linear podem freqüentemente ser utililizados como base para o desenvolvimento de algoritmos mais complexos de Programação Não-Linear. Finalmente, os conceitos relacionados à Programação Linear fornecem subsídios para o entendimento dos fundamentos gerais da programação matemática.

4.1 Características Gerais

A aplicação de algoritmos de Programação Linear para a obtenção da solução de problemas de otimização com restrições requer que tanto a função objetivo quanto as restrições sejam representadas por funções lineares.

A forma padrão do problema de otimização é [1, 3],

$$\begin{array}{ll} \text{Minimizar} & \mathbf{f}(\mathbf{x}) = \mathbf{c}^t\mathbf{x} \\ \text{sujeito a} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array} \tag{4.1}$$

onde, $\mathbf{c}$ é o vetor $(n \times 1)$ dos coeficientes da função objetivo; $\mathbf{A}$ é a matriz $(m \times n)$ dos coeficientes das restrições; $\mathbf{x}$ é o vetor $(n \times 1)$ das variáveis de otimização e $\mathbf{b}$ é o vetor $(m \times 1)$ do lado direito das restrições.

A equação (4.1) representa um problema com restrições de igualdade e de não negatividade nas variáveis de otimização. Apesar de que muitos problemas de otimização não possuem esta forma, mostra-se no texto a seguir como converter qualquer problema de otimização com função objetivo e restrições de igualdade e de desigualdade lineares para esta forma.

Devido ao fato de ser expresso apenas em função das variáveis de otimização, o problema mostrado na equação (4.1) é dito estar na *forma compacta*. Esta formulação pode ser obtida utilizando-se como índice de desempenho a ser otimizado funções lineares simples, funções multisegmentadas, ou funções linearizadas. Se as restrições são expressas por funções não lineaares, a forma linear das restrições pode ser obtida através da expansão das equações não-lineares em série de Taylor, em torno de um determinado ponto, até o termo de primeira ordem.

Um grande número de metodologias baseadas em Programação Linear têm sido propostas na literatura para a solução de problemas restritos da forma expressa na equação (4.1). Uma das técnicas mais freqüentemente utilizadas para esta finalidade é o algoritmo *Simplex*, proposto por Dantzig na década de 40. Uma variante deste método largamente utilizada em problemas lineares de grande porte é o algoritmo *Simplex Revisado*, pela sua habilidade de manipular um número elevado de restrições sem aumento demasiado no dimensionamento das matrizes envolvidas na solução do problema, e pela sua eficiência no tratamento de restrições com limites bilaterais.

As principais vantagens da utilização das técnicas de Programação Linear na solução de problemas de otimização com restrições são:

- confiabilidade;
- rapidez na determinação de soluções viáveis;
- rapidez na detecção de inviabilidade da solução;
- facilidade de implementação;

A principal desvantagem associada às metodologias baseadas em Programação Linear é a imprecisão resultante da linearização de determinadas funções objetivo, principalmente aquelas relacionadas aos problemas de despacho de potência reativa. Em particular, tanto as perdas de potência ativa nas linhas de transmissão como as injeções de potência reativa são funções com grau acentuado de não-linearidade, o que dificulta a aplicação desses algoritmos.

Exemplo 21 *Minimize a função* $f(\mathbf{x}) = -4x_1 - x_2 + 50$*, sujeito às restrições*

$$\begin{aligned} x_1 - x_2 &\quad \leq 2 \\ x_1 + 2x_2 &\quad \leq 8 \\ x_1 \geq 0 &\quad x_2 \geq 0 \end{aligned}$$

Observe que os contornos da função objetivo e as restrições são todos representados por linhas retas. No espaço n-dimensional esses elementos seriam hipersuperfícies.

A solução ótima e única é o ponto $x_1 = 4,0$ *e* $x_2 = 2,0$*, para o qual a função objetivo vale* $32,0$*. As curvas de nível da função objetivo e a região definida pelas restrições são representadas na Figura 4.1*

Exemplo 22 *Minimize a função* $f(\mathbf{x}) = -x_1 + x_2 + 30$*, sujeito às mesmas restrições do exemplo anterior.*

Neste caso, os coeficientes da função objetivo são iguais aos da primeira restrição em magnitude, porém com sinais opostos. O problema admite infinitas soluções.

A Figura 4.2 ilustra geometricamente este problema.

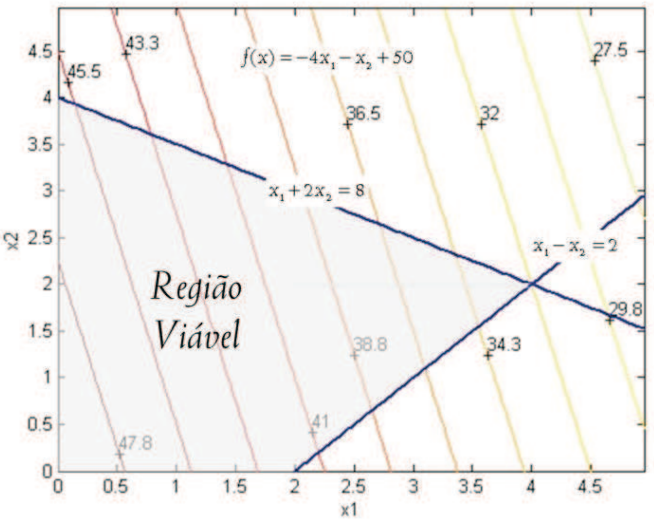

Figura 4.1: Representação do problema do exemplo 21

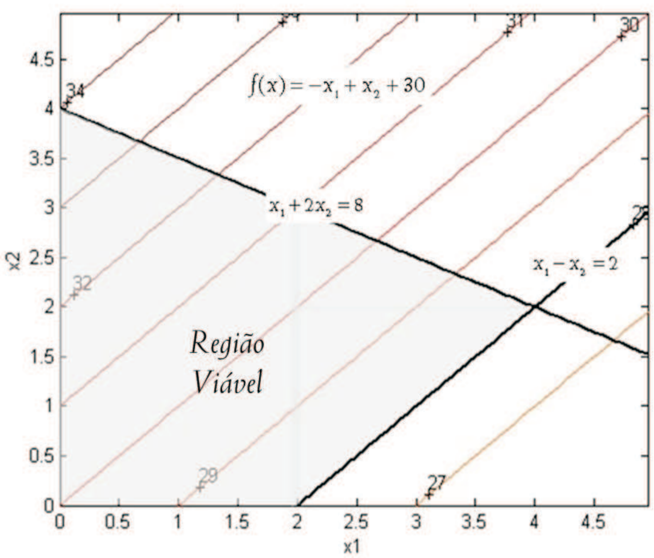

Figura 4.2: Representação do problema do exemplo 22

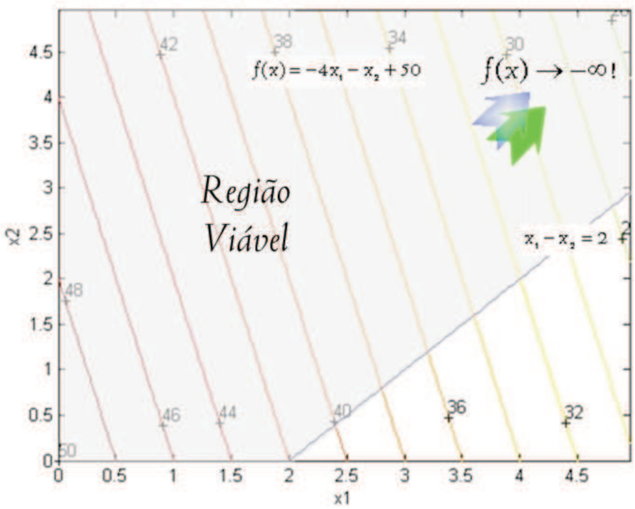

Figura 4.3: Representação do problema do exemplo 23

Exemplo 23 *Minimize a função* $f(\mathbf{x}) = -4x_1 - x_2 + 50$*, sujeito às restrições*

$$\begin{array}{ll} x_1 - x_2 & \leq 2 \\ x_1 \geq 0 & x_2 \geq 0 \end{array}$$

Neste caso, a região das soluções viáveis é um conjunto aberto, o que resulta numa solução com x_1 *e* x_2 *tendendo a infinito. Esta situação é ilustrada na Figura 4.3.*

Exemplo 24 *Minimize a função* $f(\mathbf{x}) = -4x_1 - x_2 + 50$*, sujeito às restrições*

$$\begin{array}{ll} x_1 - x_2 & \leq -5 \\ x_1 + 2x_2 & \leq 8 \\ x_1 \geq 0 & x_2 \geq 0 \end{array}$$

Neste caso não existe solução viável, conforme mostra a Figura 4.4.

4.2 Conceitos Básicos

Para converter um problema de Programação Linear à forma padrão, todas as restrições de desigualdade devem ser transformadas em restrições de igualdade. Além disso, a restrição de não negatividade nas variáveis de otimização deve ser convenientemente manipulada, de forma que as variáveis envolvidas possam assumir também eventualmente valores negativos. O estabelecimento de um procedimento lógico para a determinação da solução do problema de Programação Linear padrão requer o uso das definições mostradas a seguir.

Definição 10 : *Variável irrestrita - é aquela que pode assumir qualquer valor finito, positivo ou negativo. Para ser considerada num problema de Programação Linear na*

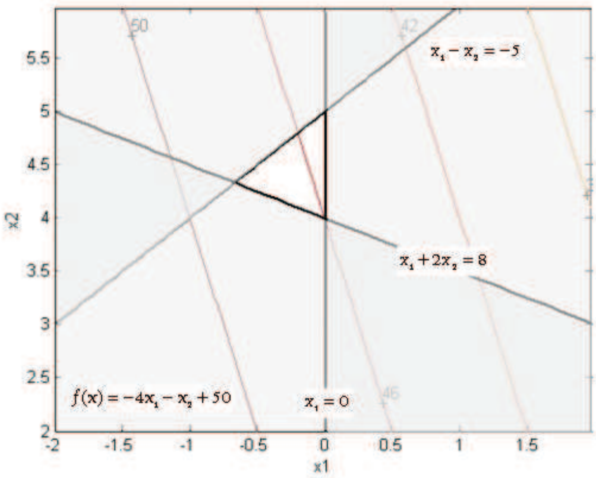

Figura 4.4: Representação do problema do exemplo 24

forma padrão, este tipo de variável deve ser expresso como a diferença entre duas variáveis não-negativas; ou seja, se x_i é variável irrestrita, ela pode ser escrita como

$$x_i = x_i' - x_i'', \quad com \quad x_i' \geq 0 \;\; e \;\; x_i'' \geq 0$$

.

A substituição de uma variável irrestrita no problema representado na equação (4.1), aumenta o número de variáveis de otimização, porém elimina a limitação imposta pela restrição de não negatividade.

Definição 11 : *Variável de folga - é aquela utilizada para se transformar uma restrição de desigualdade em restrição de igualdade. Isto é a inequação do tipo $\leq$ (menor ou igual)*

$$\sum_j a_{ij}x_j \leq b_j$$

pode ser transformada na equação

$$\sum_j a_{ij}x_j + x_{n+1} = b_j$$

com $x_{n+1} \geq 0$ e custo correspondente nulo ($c_{n+1} = 0$).

De maneira análoga, a restrição é do tipo $\geq$ (maior ou igual) pode ser transformada em restrição de igualdade; isto é, a inequação

$$\sum_j a_{ij}x_j \geq b_j$$

pode ser re-escrita como

$$\sum_j a_{ij}x_j - x_{n+1} = b_j$$

com $x_{n+1} \geq 0$ e custo correspondente nulo ($c_{n+1} = 0$).

Exemplo 25 *Converta o seguinte problema a forma padrão do problema de Programação Linear:*

$$\begin{array}{ll} \textit{Maximizar} & f(\mathbf{x}) = x_1 + 4y \\ \textit{sujeito a} & x_1 + 2y \leq 5 \\ & 2x_1 + y = 4 \\ & x_1 - y \geq 3 \\ & x_1 \geq 0, \quad y \ \textit{irrestrito em sinal} \end{array}$$

A variável irrestrita é expressa como a diferença entre duas variáveis não negativas; ou seja,

$$y = x_2 - x_3, \quad \textit{com } x_2, \ x_3 \geq 0$$

As restrições de desigualdade são convertidas em restrições de igualdade com o auxílio de variáveis de folga não negativas; isto é,

$$x_1 + 2y \leq 5 \Rightarrow x_1 + 2y + x_4 = 5$$

$$x_1 - y \geq 3 \Rightarrow x_1 - y - x_5 = 3$$

A inclusão das novas variáveis no problema de otimização original resulta

$$\begin{array}{ll} \textit{Maximizar} & f(\mathbf{x}) = x_1 + 4(x_2 - x_3) \\ \textit{sujeito a} & x_1 + 2(x_2 - x_3) + x_4 = 5 \\ & 2x_1 + (x_2 - x_3) = 4 \\ & x_1 - (x_2 - x_3) - x_5 = 3 \\ & x_1, \ x_2, \ x_3, \ x_4, \ x_5 \geq 0 \end{array}$$

ou, na forma matricial,

$$\textit{Maximizar} \qquad \begin{bmatrix} 1 & 4 & -4 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

$$\textit{sujeito a} \qquad \begin{bmatrix} 1 & 2 & -2 & 1 & 0 \\ 2 & 1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}$$

$$x_1, \quad x_2, \quad x_3, \quad x_4, \quad x_5 \geq 0$$

Definição 12 *Variável artificial - é aquela adicionada à restrição de igualdade, com custo bastante elevado para assegurar que tal variável não fará parte da solução ótima.*

Definição 13 *Solução viável - é aquela que satisfaz simultaneamente as restrições de igualdade e não negatividade.*

Definição 14 *Base - é um conjunto de restrições linearmente independentes, referidas como restrições básicas, que são satisfeitas como igualdades.*

Definição 15 *Solução básica viável - é aquela obtida selecionando-se algumas variáveis (denominadas não básicas e denotadas por x_n) como nulas e resolvendo o sistema linear remanescente para as variáveis básicas (que são maiores ou iguais a zero e denotadas por x_b).*

Esta definição permite particionar os vetores e as matrizes envolvidas no problema de otimização linear, tal que o mesmo pode ser re-escrito como

$$
\begin{array}{ll}
\text{Minimizar} & \mathbf{c}_b^t\mathbf{x}_b + \mathbf{c}_n^t\mathbf{x}_n \\[2ex]
\text{sujeito a} & \begin{bmatrix} \mathbf{A}_n & \mathbf{A}_b \end{bmatrix} \begin{bmatrix} \mathbf{x}_n \\ \mathbf{x}_b \end{bmatrix} = \mathbf{b} \\[2ex]
& \begin{bmatrix} \mathbf{x}_n \\ \mathbf{x}_b \end{bmatrix} \geq \mathbf{0}
\end{array}
$$

onde os vetores $\mathbf{x}_b$, $\mathbf{c}_b$, $\mathbf{A}_b$ e $\mathbf{x}_n$, $\mathbf{c}_n$, $\mathbf{A}_n$ correspondentes às variáveis básicas e não básicas, respectivamente.

Em geral, a solução básica inicial viável é obtida atribuindo-se o valor zero a todas as variáveis do problema original, tal que a matriz base é a matriz identidade (correspondente as variáveis de folga etc).

Definição 16 *Um sistema de equações lineares* $\mathbf{Ax} = \mathbf{b}$*, com* m *equações e* n *incógnitas e posto* $(\mathbf{A}) = m$*, é dito estar na forma canônica se em cada equação existe uma variável (com coeficiente unitário) que não aparece nas outras equações.*

Um sistema linear generalizado, na forma canônica, pode ser expresso como

$$
\begin{bmatrix}
a_{11} & a_{12} & \dots & a_{1m} & 1 & 0 & \dots & 0 \\
a_{21} & a_{22} & \dots & a_{2m} & 0 & 1 & \dots & 0 \\
\vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\
a_{m1} & a_{m2} & \dots & a_{mm} & 0 & 0 & \dots & 1
\end{bmatrix}
\begin{bmatrix} x_1 \\ \vdots \\ x_m \\ x_{m+1} \\ \vdots \\ x_n \end{bmatrix}
=
\begin{bmatrix} b_1 \\ \vdots \\ b_m \\ b_{m+1} \\ \vdots \\ b_n \end{bmatrix}
$$

Note que neste exemplo as variáveis $x_{m+1} \dots x_n$ foram selecionadas para converter o sistema linear a forma canônica, apesar de que outro conjunto de variáveis poderia ser selecionado para esta finalidade. O método de eliminação de Gauss-Jordan pode ser utilizado para converter um dado sistema linear para a forma canônica.

Exemplo 26 *Determine as soluções básicas do problema mostrado a seguir e identifique as mesmas num gráfico representando a região das soluções viáveis.*

$$
\begin{array}{ll}
\textit{Maximizar} & 4x_1 + 5x_2 \\
\textit{sujeito a} & -x_1 + x_2 \leq 4 \\
& x_1 + x_2 \leq 6 \\
& x_1, \quad x_2 \geq 0
\end{array}
$$

A forma padrão deste problema é obtida introduzindo-se variáveis de folga para converter as inequações em restrições de igualdade e transformando o problema de maximização em minimização; isto é,

$$\begin{array}{ll} \textit{Minimizar} & -4x_1 - 5x_2 \\ \textit{sujeito a} & -x_1 + x_2 + x_3 = 4 \\ & x_1 + x_2 + x_4 = 6 \\ & x_1, \quad x_2, \quad x_3, \quad x_4 \geq 0 \end{array}$$

O sistema de equações lineares que representa as restrições de igualdade do problema na forma padrão é dado por

$$\begin{bmatrix} -1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

sendo naturalmente expresso na forma canônica, como pode ser observado.

É conveniente dispor do sistema linear na forma canônica pela facilidade de se obter diretamente as soluções básicas do problema de otimização linear.

Por exemplo, se x_1 e x_2 forem selecionadas como variáveis não básicas ($x_1 = 0$ e $x_2 = 0$), as variáveis de folga x_3 e x_4 são básicas, cujos valores são 4 e 6, respectivamente. Se x_1 e x_2 forem selecionadas como variáveis básicas, então x_3 e x_4 são nulas e os valores das variáveis básicas são obtidos através da aplicação do método de eliminação de Gauss-Jordan no sistema linear que representa as restrições do problema de otimização na forma padrão; isto é,

$$\begin{bmatrix} -1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

e, desde que $x_3 = x_4 = 0$, então $x_1 = 1$ e $x_2 = 5$.

Com quatro variáveis e duas restrições, este problema possui 6 soluções básicas, as quais são mostradas na tabela 4.1

Solução	x_1	x_2	x_3	x_4	f	Viabilidade	Básicas	Não Básicas
1	0	0	4	6	0	Viável	x_3, x_4	x_1, x_2
2	0	4	0	2	-20	Viável	x_2, x_4	x_1, x_3
3	0	6	-2	0	-	Inviável	x_2, x_3	x_1, x_4
4	-4	0	0	10	-	Inviável	x_1, x_4	x_2, x_3
5	6	0	10	0	-24	Viável	x_1, x_3	x_2, x_4
6	1	5	0	0	-29	Viável	x_1, x_2	x_3, x_4

Tabela 4.1: Soluções básicas do exemplo 26

4.3 O Método Simplex

O método Simplex consiste em determinar uma solução inicial básica viável e então modificar a base, intercambiando uma de cada vez, as variáveis básicas e não básicas. Para fazer isto, deve ser observado que em qualquer estágio do processo,

$$\begin{bmatrix} \mathbf{A}_n & \mathbf{A}_b \end{bmatrix} \begin{bmatrix} \mathbf{x}_n \\ \mathbf{x}_b \end{bmatrix} = \mathbf{b}$$

ou seja,

$$\mathbf{A}_n \mathbf{x}_n + \mathbf{A}_b \mathbf{x}_b = \mathbf{b}$$

e, usando o mesmo particionamento de variáveis para a função objetivo,

$$f = \mathbf{c}_n^t \mathbf{x}_n + \mathbf{c}_b^t \mathbf{x}_b$$

onde $\mathbf{c}_n$ e $\mathbf{c}_b$ são os vetores de coeficientes da função objetivo relativos às variáveis não-básicas e básicas, respectivamente.

Desde que as linhas da submatriz $\mathbf{A}_b$ são linearmente independentes, esta matriz é não-singular, e portanto,

$$\mathbf{x}_b = \mathbf{A}_b^{-1}(\mathbf{b} - \mathbf{A}_n \mathbf{x}_n)$$

tal que a função objetivo pode ser escrita como

$$f = \mathbf{c}_n^t \mathbf{x}_n + \mathbf{c}_b^t \mathbf{A}_b^{-1}(\mathbf{b} - \mathbf{A}_n \mathbf{x}_n)$$

ou, em forma compacta,

$$f = \mathbf{c}_b^t \mathbf{A}_b^{-1} \mathbf{b} + \mathbf{c}_n^{t'} \mathbf{x}_n \tag{4.2}$$

onde $\mathbf{c}_n^{t'} = \mathbf{c}_n^t - \mathbf{c}_b^t \mathbf{A}_b^{-1} \mathbf{A}_n$.

A cada iteração do processo, o conjunto das variáveis básicas $\mathbf{x}_b \geq 0$ é modificado para intercambiar uma variável básica com uma variável não-básica. Este procedimento termina apenas quando for assegurado que o intercâmbio de variáveis não resulta em decréscimo no valor da função objetivo. A variável não-básica a ser introduzida na base é escolhida observando-se os coeficientes $\mathbf{c}_n^{'}$ na equação (4.2). A análise desta equação revela que enquanto existir algum componente de $\mathbf{c}_n^{'}$ negativo durante o processo computacional, é possível reduzir o valor da função objetivo através da conversão da variável não básica correspondente ao $c_{ni}^{'}$ mais negativo (para assegurar a máxima melhoria no valor da função objetivo) em variável básica. O processo de intercâmbio de variáveis é efetuado portanto, até que todos os coeficientes $c_{ni}^{'}$ sejam positivos ou nulos.

A variável a deixar a base é selecionada com base na equação

$$\mathbf{x}_b = \mathbf{A}_b^{-1}(\mathbf{b} - \mathbf{A}_n \mathbf{x}_n) \tag{4.3}$$

Supondo que a variável a ser incluída na base é denotada por x_{ni}, a equação (4.3) pode ser re-escrita como

$$\mathbf{x}_b = \mathbf{A}_b^{-1}\mathbf{b} - \mathbf{A}_b^{-1} \begin{bmatrix} \dots & \mathbf{a}_{ni1} & \dots \\ \dots & \mathbf{a}_{ni2} & \dots \\ \vdots & \vdots & \vdots \\ \dots & \mathbf{a}_{nin} & \dots \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ x_{ni} \\ \vdots \\ 0 \end{bmatrix}$$

o que fornece

$$\mathbf{x}_b = \mathbf{A}_b^{-1}\mathbf{b} - \mathbf{A}_b^{-1}\mathbf{A}_{ni}x_{ni}$$

onde $\mathbf{A}_{ni}$ é a coluna da matriz $\mathbf{A}_n$ correspondente à variável x_{ni}.

Tomando $\mathbf{p} = \mathbf{A}_b^{-1}\mathbf{b}$ e $\mathbf{q} = \mathbf{A}_b^{-1}\mathbf{A}_{ni}$ pode-se escrever

$$\begin{array}{rcll} x_{b1} & = & p_1 - & q_1 x_{ni} \\ x_{b2} & = & p_2 - & q_2 x_{ni} \\ \vdots & & \vdots & \vdots \\ x_{bj} & = & p_j - & q_j x_{ni} \\ x_{bk} & = & p_k - & q_k x_{ni} \end{array}$$

Com relação a esta última equação, denotando $\theta_k = p_k/q_k$, as seguintes observações podem ser feitas:

- Ao se tornar uma variável básica, x_{ni} deve ser positivo. Por outro lado, supondo x_{bj} a variável que será excluída da base, p_j também é positivo, o que implica em que a variável a deixar a base deve ser selecionada entre aquelas com correspondente q_k positivo. Observe que x_{bj} se tornará nulo (variável não-básica);

- Visando assegurar que todas as variáveis remanescentes na base serão positivas, a variável selecionada para deixar a base deve ser aquela correspondente ao mínimo θ_k.

Após o intercâmbio das variáveis básica e não básica selecionadas conforme descrito previamente, o vetor $\mathbf{b}$ (que fornece a solução básica em qualquer instante do processo) e as submatrizes $\mathbf{A}_b$ e $\mathbf{A}_n$ devem ser atualizadas. Esta atualização é realizada com base na equação (4.3), a qual pode ser vista como uma operação que converte o sistema linear à forma canônica correspondente às variáveis básicas selecionadas. Portanto,

$$\mathbf{b}^{atual} = \mathbf{A}_b^{-1^{atual}}\mathbf{b} \tag{4.4}$$

e

$$\mathbf{A}_n^{atual} = \mathbf{A}_b^{-1^{atual}}\mathbf{A}_n \tag{4.5}$$

onde, $\mathbf{b}$ e $\mathbf{A}_n$ correspondem aos conjuntos de variáveis básicas e não básicas antes do intercâmbio.

4.3.1 Algoritmo

O procedimento descrito na seção anterior pode ser sumarizado no seguinte algoritmo:

1. Determine uma solução básica inicial viável;

2. Calcule $\mathbf{c}_n^{t'} = \mathbf{c}_n^t - \mathbf{c}_b^t\mathbf{A}_b^{-1}\mathbf{A}_n$;

3. Determine o c'_{nj} mais negativo: a variável x_{nj} é aquela que deve ser incluída na base. Caso todos os c'_{nj} sejam positivos ou nulos o processo é encerrado;

4. Calcule $\mathbf{p} = \mathbf{A}_b^{-1}\mathbf{b}$ e $\mathbf{q} = \mathbf{A}_b^{-1}\mathbf{A}_{ni}$;

5. Calcule $\theta_k = p_k/q_k$, para todos os q_k positivos;

6. Selecione o mínimo θ_j: a variável x_{bj} é aquela que deve ser excluída da base;
7. Determine a nova matriz inversa $\mathbf{A}_b^{-1}$;
8. Atualize $\mathbf{b}$ e $\mathbf{A}_n$ usando as expressões (4.4) e (4.5);
9. Retorne ao segundo passo.

Exemplo 27 *Resolver o problema do exemplo 26 através do método Simplex.*

Conforme mostrado anteriormente, a forma padrão deste problema é

$$\begin{array}{ll} \textit{Minimizar} & -4x_1 - 5x_2 - 0x_3 - 0x_4 \\ \textit{sujeito a} & \begin{bmatrix} -1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} \end{array}$$

$$x_1, \quad x_2, \quad x_3, \quad x_4 \geq 0$$

A aplicação do algoritmo descrito na subseção 4.3.1 requer os seguintes cálculos:

1. Determine uma solução básica inicial viável;

   Visando tirar proveito da forma canônica do sistema linear que representa as restrições de igualdade, as variáveis não básicas da solução inicial são x_1 e x_2 enquanto que as variáveis básicas são x_3 e x_4. Desta forma, o vetor das variáveis de otimização é expresso como

$$\mathbf{x}^0 = \begin{bmatrix} \mathbf{x}_n \\ \cdots \\ \mathbf{x}_b \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 4 \\ 6 \end{bmatrix}$$

   o vetor dos coeficientes da função objetivo é dado por

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_n \\ \cdots \\ \mathbf{c}_b \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} -4 \\ -5 \\ \cdots \\ 0 \\ 0 \end{bmatrix}$$

   e as submatrizes correspondentes às variáveis não básicas e básicas são, respectivamente,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_n & | & \mathbf{A}_b \end{bmatrix} = \left[\begin{array}{cc|cc} -1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right]$$

2. Calcule $\mathbf{c}_n^{t'} = \mathbf{c}_n^t - \mathbf{c}_b^t \mathbf{A}_b^{-1} \mathbf{A}_n$; isto é,

$$\mathbf{c}_n^{t'} = \begin{bmatrix} -4 & -5 \end{bmatrix} - \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{c}_n^{t'} = \begin{bmatrix} -4 & -5 \end{bmatrix}$$

3. Determine o c'_{nj} mais negativo: a variável x_{nj} é aquela que deve ser incluída na base.

   O c'_{nj} mais negativo é aquele correspondente à variável x_2 (de valor -5), e portanto esta variável deve ser incluída na base;

4. Calcule $\mathbf{p} = \mathbf{A}_b^{-1}\mathbf{b}$ e $\mathbf{q} = \mathbf{A}_b^{-1}\mathbf{A}_{ni}$; ou seja,

$$\mathbf{p} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

$$\mathbf{q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

5. Calcule $\theta_k = p_k/q_k$, para todos os q_k positivos;

   Os θ_k valem $\dfrac{4}{1} = 4$ (correspondente à variável x_3) e $\dfrac{6}{1} = 6$ (correspondente à variável x_4)

6. Selecione o mínimo θ_j;

   A variável x_3 é aquela que deve ser excluída da base. O novo particionamento das variáveis fornece

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_n \\ \ldots \\ \mathbf{x}_b \end{bmatrix} = \begin{bmatrix} x_1 \\ x_3 \\ \ldots \\ x_2 \\ x_4 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_n & | & \mathbf{A}_b \end{bmatrix} = \left[\begin{array}{rr|rr} -1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{array}\right]$$

7. Determine a nova matriz inversa $\mathbf{A}_b^{-1}$;

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

8. Atualize $\mathbf{b}$ e $\mathbf{A}_n$ usando as expressões (4.4) e (4.5);

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad e \quad \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$$

9. Retorne ao segundo passo.

Com o objetivo de simplificar a aplicação do método Simplex, assim como de facilitar a compreensão dos passos do algoritmo mostrado na seção 4.3.1, uma tabela (*tableau*) é construída com o sistema linear na forma canônica e atualizada a cada iteração do processo de busca da solução ótima. Para o exemplo em considerado, as tabelas relativas a cada iteração são mostradas a seguir.

Básica	x_1	x_2	x_3	x_4	**b**	Razão
x_3	-1	1	1	0	4	$\frac{4}{1} = 4$
x_4	1	1	0	1	6	$\frac{6}{1} = 6$
Custo	-4	-5	0	0	$f = 0$	
Básica	x_1	x_2	x_3	x_4	**b**	Razão
x_2	-1	1	1	0	4	
x_4	2	0	-1	1	2	
Custo	-9	0	5	0	$f = 20$	
Básica	x_1	x_2	x_3	x_4	**b**	Razão
x_2	0	1	$\frac{1}{2}$	$\frac{1}{2}$	5	
x_1	1	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	
Custo	0	0	$\frac{1}{2}$	$\frac{9}{2}$	$f = 29$	

Tabela 4.2: Resumo das iterações - exemplo 27

4.3.2 Simplex na Forma Tabular

Para facilitar a resolução de problemas à mão, pode-se empregar o *tableau simplex*. Todas as operações necessárias para levar a estimativa inicial da solução do problema, passando por vários vértices do politopo viável, até a solução ótima, pode ser feita facilmente usando o tableau. Existem vários fomatos de tableau simplex. O usado neste curso será explicado através do exemplo abaixo.

Deseja-se resolver o seguinte PL:

$$\begin{array}{ll} \text{min} & f = -4x_1 - 3x_2 \\ \text{sujeito a} & 2x_1 + 3x_2 \leq 6 \\ & -3x_1 + 2x_2 \leq 3 \\ & 2x_2 \leq 5 \\ & 2x_1 + x_2 \leq 4 \end{array} \tag{4.6}$$

Para inicial o processo de resolução, coloca-se o problema na forma padrão e com a função objetivo representada como uma restrição:

$$\begin{array}{lllllllll} \text{min} & f & & & & & & & \\ \text{sujeito a} & -4x_1 & -3x_2 & & & & & = f \\ & 2x_1 & +3x_2 & +x_3 & & & & = 6 \\ & -3x_1 & +2x_2 & & +x_4 & & & = 3 \\ & & 2x_2 & & & +x_5 & & = 5 \\ & 2x_1 & +x_2 & & & & +x_6 & = 4 \\ & x_1, ..., x_6 \geq 0 & & & & & & \end{array} \tag{4.7}$$

A estimativa de solução inicial do problema é: $x_1 = x_2 = 0$ (variáveis não básicas) $x_3 = 6,\ x_4 = 3,\ x_5 = 5,\ x_6 = 4$ (variáveis básicas). Neste ponto $f = 0$.

Uma vez que as restrições de não negatividade não participam explicitamente do processo de solução, os coeficientes da função objetivo e das equações são organizados numa tabela (um tableau):

x_1	x_2	x_3	x_4	x_5	x_6	f	
-4	-3	0	0	0	0	0	L_0
2	3	1	0	0	0	6	L_1
-3	2	0	1	0	0	3	L_2
0	2	0	0	1	0	5	L_3
2	1	0	0	0	1	4	L_4

A tabela acima é o "tableau simplex" inicial. As linhas do tableau são numeradas $L_0, ..., L_4$. As linhas $L_1 - L_4$ formam a matriz de coeficientes, $\mathbf{A}$, cujos elementos são a_{ij}. Na coluna à esquerda do tableau tem-se os elementos do vetor $\mathbf{b}$.

Considere L_0. Uma vez que os coeficientes de x_1 e x_2 são negativos, f pode ser reduzida ao se colocar uma dessas variáveis na base. Para progredir o mais rápido possível escolhe-se a variável com coeficiente mais negativo para entrar na base, ou seja, x_1.

A coluna de x_1 é denominada *coluna do pivot*.

Para determinar o incremento máximo em x_1 faz-se o seguinte teste: Primeiramente verifica-se quais coeficientes de $\mathbf{A}$ na coluna do pivot são positivos. Calculam-se as razões b_j/a_{j1} para todo a_{j1} positivo. A variável básica associada à linha j com menor razão b_j/a_{j1} sairá da base quando x_1 entrar na base. A linha j é denominada *linha do pivot*.

Usando o tableu anterior:

x_1	x_2	x_3	x_4	x_5	x_6	f	Razões
-4	-3	0	0	0	0	0	
2	3	1	0	0	0	6	$6/2 = 3$
-3	2	0	1	0	0	3	$-$
0	2	0	0	1	0	5	$-$
2	1	0	0	0	1	4	$4/2 = 2 \leftarrow$ Mínimo

O teste das razões indica que a variável x_6 deverá sair da base. Note que não foram computadas as raões para L_2 e L_3, pois $a_{2,1}$ e $a_{3,1}$ são negativos.

Com x_1 entrando na base e x_6 saindo da base, deve-se construir um novo tableau com a seguinte configuração:

x_1	x_2	x_3	x_4	x_5	x_6	f
0	?	0	0	0	?	?
0	?	1	0	0	?	?
0	?	0	1	0	?	?
0	?	0	0	1	?	?
1	?	0	0	0	?	?

onde ? indica que o novo coeficiente deve ser calculado.

Para criar este novo tableau, são feitas operações com as linhas, tal como na eliminação Gaussiana. Por exemplo, a nova linha 4 será gerada multiplicando-se L_4 por $1/2$; e a nova linha 0 será gerada multiplicando-se L_4 por 2 e adicionando o resultado a L_0. Fazendo operações similares obtém-se o tableu abaixo.

	x_1	x_2	x_3	x_4	x_5	x_6	f
$2 \times L_4 + L_0$:	0	-1	0	0	0	2	+8
$(-1) \times L_4 + L_1$:	0	2	1	0	0	-1	2
$(3/2) \times L_4 + L_2$:	0	7/2	0	1	0	3/2	9
$0 \times L_4 + L_3$:	0	2	0	0	1	0	5
$(1/2) \times L_4$:	1	1/2	0	0	0	1/2	2

Note que no lado esquerdo deste tableu são indicadas as operações feitas para obter cada linha. Observe também que na última coluna aparece o novo valor da função objetivo com o sinal trocado.

Nesta primeira iteração o *pivot* (o "2" localizado na interseção da linha e coluna do pivot) é usado para gerar todas as outras linhas.

A base associada a este novo tableu é composta por x_1, x_3, x_4, x_5. A nova estimativa de solução é $(x_1, x_2, x_3, x_4, x_5, x_6) = (2, 0, 2, 9, 5, 0)$ e o valor da função objetivo é $f = -8$.

No novo tableu em L_0 o coeficiente de x_2 é negativo. Portanto, deve-se inserir x_2 na base. Para escolher a variável a sair da base faz-se o novo teste das razões:

x_1	x_2	x_3	x_4	x_5	x_6	f	f
0	-1	0	0	0	2	+8	
0	2	1	0	0	-1	2	$1/2 = 1 \leftarrow$ Mínimo
0	7/2	0	1	0	3/2	9	$9/(7/2) = 18/7$
0	2	0	0	1	0	5	$5/2$
1	1/2	0	0	0	1/2	2	$2/(1/2) = 4$

Tem-se que L_1 é a linha do pivot. Assim, a variável a sair da base é x_3.

Usando $a_{12} = 2$ como pivot, faz-se novo pivoteamento para obter o tableau abaixo.

	x_1	x_2	x_3	x_4	x_5	x_6	f
$(1/2) \times L_1 + L_0$:	0	1	1/2	0	0	3/2	+9
$(1/2) \times L_1$:	0	0	1/2	0	0	-1/2	1
$(-7/4) \times L_1 + L_2$:	0	0	-7/4	1	0	13/4	11/2
$(-1) \times L_1 + L_3$:	0	0	-1	0	1	1	3
$(-1/4) \times L_1 + L4$:	1	0	-1/4	0	0	3/4	3/2

A nova solução é $(x_1, x_2, x_3, x_4, x_5, x_6) = (3/2, 1, 0, 11/2, 3, 0)$ e $f = -9$. Como os coeficientes de x_3 e x_6 em L_0 são positivos, esta é a solução ótima.

Aplicação a Sistemas Elétricos de Potência 4 *Considere um sistema de transmissão no qual três barras (1, 2 e 3) estão com a magnitude da tensão abaixo do limite inferior. Essas magnitudes devem ser modificadas de tal forma que os seus incrementos satisfaçam as seguintes restrições*

$$\begin{bmatrix} \Delta V_1 \\ \Delta V_2 \\ \Delta V_3 \end{bmatrix} \geq \begin{bmatrix} 0,04 \\ 0,05 \\ 0,06 \end{bmatrix}$$

Esses incrementos devem ser obtidos por efeito da instalação de capacitores shunt nas baras 4 e 5. A relação entre os incrementos de magnitude da tensão das barras 1, 2 e

3 e de potência reativa das barras 4 e 5 é expressa por

$$\begin{bmatrix} \Delta V_1 \\ \Delta V_2 \\ \Delta V_3 \end{bmatrix} = \begin{bmatrix} 0,01 & 0,03 \\ 0,08 & 0,05 \\ 0,02 & 0,01 \end{bmatrix} \begin{bmatrix} \Delta Q_4 \\ \Delta Q_5 \end{bmatrix}$$

Determine a mínima quantidade de compensação reativa que deve ser instalada nas barras 4 e 5.

4.4 Exercícios

Exercício 10 *Converta os seguintes problemas a forma padrão.*

- $$\begin{array}{ll} \textit{Minimizar} & 8x_1 - 3x_2 + 15x_3 \\ \textit{sujeito a} & 5x_1 - 1,8x_2 - 3,6x_3 \geq 2 \\ & 3x_1 + 6x_2 + 8,2x_3 \geq 5 \\ & 1,5x_1 - 4x_2 + 7,5x_3 \geq -4,5 \\ & -x_2 + 5x_3 \geq 1,5 \\ & x_1, \quad x_2 \geq 0, \quad x_3 \ \textit{irrestrito em sinal} \end{array}$$

- $$\begin{array}{ll} \textit{Maximizar} & 2x_1 + 5x_2 - 4,5x_3 + 1,5x_4 \\ \textit{sujeito a} & 5x_1 + 3x_2 + 1,5x_3 \leq 8 \\ & 1,8x_1 - 6x_2 + 4x_3 + x_4 \geq 3 \\ & -3,6x_1 + 8,2x_2 + 7,5x_3 + 5x_4 = 15 \\ & x_i \geq 0, \quad i = 1,2,3,4 \end{array}$$

Exercício 11 *Determine todas as soluções básicas dos seguintes problemas de Programação Linear. Identifique as soluções viáveis e mostre (se possível) essas soluções num gráfico.*

- $$\begin{array}{ll} \textit{Maximizar} & x_1 + 2x_2 \\ \textit{sujeito a} & 3x_1 + 4x_2 \leq 12 \\ & x_1 + 3x_2 \geq 3 \\ & x_1 \geq 0, \quad x_2 \ \textit{irrestrito em sinal} \end{array}$$

- $$\begin{array}{ll} \textit{Minimizar} & 5x_1 + 4x_2 - x_3 \\ \textit{sujeito a} & x_1 + 2x_2 - x_3 \geq 1 \\ & 2x_1 + x_2 + x_3 \geq 4 \\ & x_1, \quad x_3 \geq 0, \quad x_2 \ \textit{irrestrito em sinal} \end{array}$$

- $$\begin{array}{ll} \textit{Minimizar} & 9x_1 + 2x_2 + 3x_3 \\ \textit{sujeito a} & -2x_1 - x_2 + 3x_3 \leq -5 \\ & x_1 - 2x_2 + 2x_3 \geq -2 \\ & x_1, \quad x_2, \quad x_3 \geq 0 \end{array}$$

Exercício 12 *Resolva os seguintes problemas usando o método Simplex.*

- $$\begin{array}{ll} \textit{Maximizar} & x_1 + 0,5x_2 \\ \textit{sujeito a} & 6x_1 + 5x_2 \leq 30 \\ & 3x_1 + x_2 \leq 12 \\ & x_1 + 3x_2 \leq 12 \\ & x_1, \quad x_2 \geq 0 \end{array}$$

- $$\begin{array}{ll} \textit{Minimizar} & 2x_1 - x_2 \\ \textit{sujeito a} & -x_1 + 2x_2 \leq 10 \\ & 3x_1 + 2x_2 \leq 18 \\ & x_1, \quad x_2 \geq 0 \end{array}$$

- $$\begin{array}{ll} \textit{Minimizar} & 8x_1 - 3x_2 + 15x_3 \\ \textit{sujeito a} & 5x_1 - 1,8x_2 - 3,6x_3 \geq 2 \\ & 3x_1 + 6x_2 + 8,2x_3 \geq 5 \\ & 1,5x_1 - 4x_2 + 7,5x_3 \geq -4,5 \\ & -x_2 + 5x_3 \geq 1,5 \\ & x_1, \quad x_2 \geq 0, \quad x_3 \;\; \textit{irrestrito em sinal} \end{array}$$

- $$\begin{array}{ll} \textit{Maximizar} & 2x_1 + 5x_2 - 4,5x_3 + 1,5x_4 \\ \textit{sujeito a} & 5x_1 + 3x_2 + 1,5x_3 \leq 8 \\ & 1,8x_1 - 6x_2 + 4x_3 + x_4 \geq 3 \\ & -3,6x_1 + 8,2x_2 + 7,5x_3 + 5x_4 = 15 \\ & x_i \geq 0, \quad i = 1,2,3,4 \end{array}$$

Aplicação a Sistemas Elétricos de Potência 5 *Um sistema possui duas unidades térmicas com curvas de custo de geração de potência ativa e limites de capacidade dados por*

$$C_1(P_1) = 10P_1 + 8 \times 10^{-3} P_1^2 \;\; \$/h \qquad 100 \leq P_1 \leq 600 \; Mw$$

$$C_2(P_2) = 8P_2 + 9 \times 10^{-3} P_2^2 \;\; \$/h \qquad 400 \leq P_2 \leq 1000 \; Mw$$

A perda total de potência ativa nas linhas de transmissão é expressa como uma função do tipo

$$P_L = 1,5 \times 10^{-4} P_1^2 + 2 \times 10^{-5} P_1 P_2 + 3,0 \times 10^{-5} P_2^2$$

com P_1 e P_2 expressos em Mw.

- *linearize as curvas de custo de geração e a função das perdas de transmissão no ponto $P_1 = 270MW$ e $P_2 = 405MW$;*
- *utilizando o modelo linearizado do item anterior, determine através do método Simplex a potência de saída de cada unidade geradora que resulta no mínimo custo de geração, necessária para suprir uma demanda de 680MW;*
- *verifique se o balanço de potência (com a expressão não-linear da perda) é satisfeito;*

Linha	Z_{ser}	Barra	P_g^{min}	P_g^{Max}	P_d
	(pu)		(pu)	(pu)	pu
1-2	j 1,0	1	5,0	10,0	2,0
1-2	j 1,0	2	-	-	5,0
1-2	j 1,0	3	-	-	5,0
1-2	j 1,0	4	0,0	10,0	-
1-2	j 1,0				

Tabela 4.3: Dados do sistema de 4 barras - aplicação 6

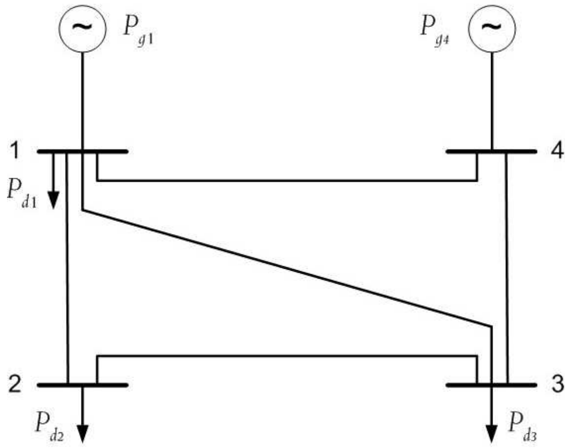

Figura 4.5: Diagrama do sistema da aplicação 6

Aplicação a Sistemas Elétricos de Potência 6 *Considere o sistema da Figura 4.5, cujos dados são mostrados na tabela 4.3. Utilizando as equações do fluxo de potência linearizado para modelar a rede elétrica, determine a geração de mínimo custo que satisfaz a carga, sem sobrecarregar as linhas 1-4 e 2-3.*

Adote o ângulo da barra 4 como referência. Os limites de fluxo de potência ativa nas linhas 1-4 e 2-3 são ambos 2,0 pu.

Referências Bibliográficas

[1] P.E. Gill, W. Murray, and M. Wright. *Practical Optimisation.* Academic Press, 1981

[2] G. Hadley. *Nonlinear and Dynamic Programming.* Addison-Wesley, 1970.

[3] D.M. Greig. *Optimisation.* Longman, 1980.

[4] David G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley Publishing Company, 1984.

[5] C.L. Lawson and R.J. Hanson *Solving Least Squares Problems.* Prentice Hall, 1974.

[6] David M. Himmelblau. *Aplied Nonlinear Programming.* Mc-Graw Hill, 1972.

[7] M. S. Bazaraa e C.M. Shetty. *Nonlinear Programming, Theory and Algorithms.* John-Wiley & Sons, 1984.