

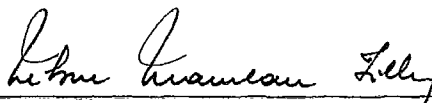
PROBLEMA DE PROGRAMAÇÃO NÃO LINEAR:

UM ALGORITMO DE DIREÇÃO VIÁVEL

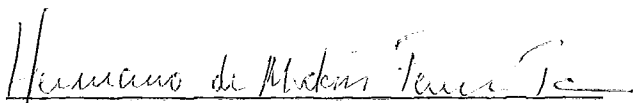
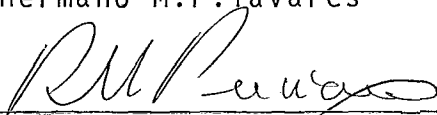
Luiz Carlos da Rocha

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS (D.Sc.).

Aprovada por:



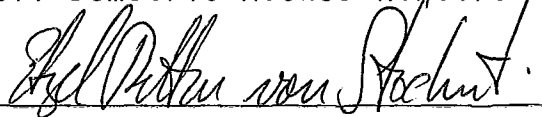
Prof. Nelson Maculan Filho
(Presidente)


Prof. Hermano M.F. Tavares

Prof. Ronaldo C. Marinho Persiano



Prof. Demétrio Alonso Ribeiro



Prof. Etzel Ritter Von Stockert

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 1979

CURRICULUM VITAE

Luis Carlos da Rocha nasceu em 1940 no Rio de Janeiro, graduou-se em bacharel na Escola Nacional de Ciências Estatística em 1964, R.J.

No período 1965-1966, com bolsa de estudo da O.E.A, obteve o grau de Mestre no Centro Interamericano de Enseñaza de Estatística em Santiago-Chile. De 1967 a 1975 trabalhou em diversas empresas privadas e estatais, desempenhando na maioria delas a função de analista de Pesquisa Operacional. Durante o período 1976-1977 como bolsista da Embrapa, Empresa Brasileira de Pesquisa Agropecuária, frequentou os cursos da COPPE, matriculado no Programa de Engenharia de Sistemas para a obtenção do grau de Doutor.

ROCHA, LUIZ CARLOS

Problema de Programação Não Linear. Um algoritmo de
Direção Viável |RIO DE JANEIRO| 1979.

VI, 51p., 29,7cm (COPPE-UFRJ), D.Sc., Engenharia de
Sistemas, 1979)

Tese - Universidade Federal do Rio de Janeiro -Fac.En
genharia

1. Otimização I.COPPE/UFRJ II. Título (série)

A

HAYDÉE

AGRADECIMENTOS

Agradeço a

Gilberto Paez Bogarin

Nelson Maculan Filho

O trabalho recebeu o suporte financeiro da EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária, da qual o autor foi bolsista no período 1976-1977.

SUMÁRIO

O algoritmo desenvolvido neste trabalho é dirigido para resolver problemas de programação matemática não linear com restrições lineares. Pertence a classe de algoritmos de direção viável com máxima derivada direcional. Inicialmente é caracterizado o conjunto de direções viáveis para o problema proposto, e em seguida desenvolvido um processo de obtenção da direção de máxima derivada direcional, através da otimização de um problema quadrático, em que as únicas restrições são as de não-negatividade das variáveis.

SUMMARY

This work describes a method of feasible-direction that maximizes the directional derivative of the objective function at each iteration, when are the variables linearly constrained. The feasible direction is obtained by minimization of a quadratic form with the restriction being the non-negativity of the variables.

INDICE

	<u>Páginas</u>
I. Introdução	1
II. Fundamentos Teóricos	11
III. O Algoritmo	22
IV. Conclusão	32
V. Obras Citadas	34
Anexo A	36
Anexo B, Exemplos	44

I. Introdução

O algoritmo de direção viável que desenvolveremos neste trabalho é proposto para resolver problemas de programação matemática do tipo;

$$\begin{array}{ll}
 \text{Max } f(X) & X = (x_1, x_2, \dots, x_n) \\
 \\
 \text{sujeito a } \sum_j l_{kj} x_j = b_k & k \in I_1 \\
 \sum_j l_{kj} x_j \leq b_k & k \in I_2 \\
 x_j \geq 0 & j \in J
 \end{array} \tag{I}$$

com $f(X)$ suposta diferenciável em todo ponto do conjunto de soluções.

Os métodos de direção viável surgiram na literatura na forma de livro texto em 1960, com a publicação de *Methods of Feasible Directions* por G. Zoutendijk [12]. Tais métodos procuram a cada passo, determinar uma direção viável que produza, no caso de maximização, um acréscimo da função objetivo $f(X)$. Os diversos métodos existentes se diferenciam na escolha desta direção, e diz-se que o algoritmo é de direção viável ótima, quando esta direção maximiza a derivada direcional de $f(X)$ dentre todas as possíveis direções viáveis. O algoritmo por nós desenvolvido é de tipo de direção viável ótima.

Os primeiros esforços para a obtenção destes tipos de algoritmos foram desenvolvidos por Zoutendijk [11,12] para resolver problemas da forma (I). Vamos considerar de início,

um ponto viável X de (I) , e uma direção de $d \in E^n$. Tratemos de definir condições para que d seja viável, ou seja para que exista $\lambda_M > 0$ tal que todos os pontos $X + \lambda d$, com $\lambda \in (0, \lambda_M]$ sejam pontos viáveis de (I) . Notaremos por $J(X)$ e $I_2(X)$ os conjuntos de índices correspondentes às restrições ativas para X . Isto é,

$$x_j = 0 \leftrightarrow j \in J(X) \subseteq J$$

$$\sum 1_{kj} \cdot x_j = b_k \leftrightarrow K \in I_2(X) \subseteq I_2$$

Para que d seja viável, devemos ter

$$\begin{aligned} \langle L_k, d \rangle &= 0 & K \in I_1 \\ \langle L_k, d \rangle &\leq 0 & K \in I_2(X) \\ d_j &\geq 0 & j \in J(X) \end{aligned} \tag{A}$$

Onde $\langle x, y \rangle$ é a notação usada para representar o produto interno vetorial. Se além da viabilidade de d , quisermos que esta maximize a derivada direcional de $f(X)$, deveremos resolver o seguinte problema de programação matemática.

$$\begin{aligned} \text{Max: } D_r f(X) &= \langle \nabla f(X), d \rangle \\ \langle L_k, d \rangle &= 0 & K \in I_1 \\ \langle L_k, d \rangle &\leq 0 & K \in I_2(X) \\ d_j &\geq 0 & j \in J(X) \\ \langle d, d \rangle &= 1 \end{aligned} \tag{B}$$

0 que representa, pela inclusão de $|d| = 1$, um

problema com restrições não linear e cujo conjunto de soluções é não convexo. Supondo (B) resolvido, seja d_x a solução com $\langle \nabla f(X), d_x \rangle > 0$ e sejam λ_M e λ_x definidos por:

$$\lambda_M = \max \{ \lambda / X + \lambda \cdot d_x \text{ é viável para (I)} \}$$

$$\lambda_x \text{ maximiza } f(X + \lambda \cdot d_x) \text{ com } \lambda \in (0, \lambda_M]$$

se λ_x é finito, então $X + \lambda_x \cdot d_x$ é tal que $f(X + \lambda_x \cdot d_x) > f(X)$.

Assim sendo, a partir de X foi possível obter um outro ponto com acréscimo da função objetivo.

Vamos supor que num determinado passo do algoritmo tenha-se X_* tal que a solução d_* de (B) para este ponto forneça $\langle \nabla f(X_*), d_* \rangle \leq 0$. Neste caso, demonstra-se que X_* é um ponto satisfazendo Kuhn-Tucker, e portanto é um ponto estacionário restrito. O processo é interrompido também se num determinado passo se tem que $\lambda_x = \infty$. Finalmente pode acontecer que o processo gere uma sucessão de pontos viáveis de (I), neste caso se ela for divergente então uma solução ilimitada é evidenciada, caso contrário o ponto limite será necessariamente um ponto estacionário restrito. Como veremos no desenvolvimento do trabalho, esta propriedade de convergência para pontos satisfazendo Kuhn-Tucker é uma característica dos métodos de direção viável ótima.

Voltando ao problema (B), mostraremos que este é equivalente o seguinte problema de programação quadrática.

$$\begin{aligned} & \text{Min } \langle u, u \rangle \\ & \text{sujeito a } \begin{cases} \langle L_k, u \rangle = 0 & K \in I_1 \\ \langle L_k, u \rangle \leq 0 & K \in I_2(X) \\ u_j \geq 0 & j \in J(X) \\ \langle \nabla f(X), u \rangle = 1 \end{cases} \end{aligned} \quad (C)$$

É fácil ver que existe uma relação bi-unívoca entre as soluções viáveis de (B) tendo a desigualdade $\langle \nabla f(X), d \rangle > 0$, e as soluções viáveis de (C). Demonstraremos a seguir a equivalência entre as soluções dos problemas (B) e (C).

Se d_x é a solução de (B) com $\langle \nabla f(X), d_x \rangle = \sigma_x > 0$ então $u_x = \frac{d_x}{\sigma_x}$ é a solução para (C), pois supondo que u_x não minimiza $\langle u, u \rangle$, então existirá u^* viável tal que, $\langle u^*, u^* \rangle < \langle u_x, u_x \rangle$ ou seja $|u^*| < |u_x|$. Fazendo $d^* = \frac{u^*}{|u^*|}$ então d^* é viável para (B) e teremos,

$$\begin{aligned} \langle \nabla f(X), d^* \rangle &= \frac{1}{|u^*|} \langle \nabla f(X), u^* \rangle \\ &= \frac{1}{|u^*|} > \frac{1}{|u_x|} = \langle \nabla f(X), d_x \rangle \end{aligned}$$

Logo d_x não é solução de (B), o que é absurdo. Por outro lado, se u_x é solução de (C) então $d_x = \frac{u_x}{|u_x|}$ é a solução de (B), pois caso contrário existiria d^* viável tal que:

$$\langle \nabla f(X), d^* \rangle = \sigma^* > \langle \nabla f(X), d_x \rangle = \sigma_x > 0$$

então $u^* = \frac{d^*}{\sigma^*}$ é viável para (C) com

$$|u^*| = \frac{1}{(\sigma^*)^2} < |u_x| = \frac{1}{(\sigma_x)^2}$$

o que contradiz a hipótese de ser u_x solução de (C).

Assim, nosso problema (B) foi transformado em (C)

que é um problema de programação quadrática convexa, com restrições lineares.

Tal problema, abordado pelos métodos clássicos como por exemplo o método de Wolfe, deverá resolver a cada iteração um simplex de pelo menos $n + m_1 + m_2 + m_j$ restrições, onde m_1 , m_2 e m_j representam respectivamente o número de restrições do tipo I_1 , $I_2(X)$ e $J(X)$ do problema (C), o que significa um esforço computacional muito grande, mesmo para computadores potentes.

Zontendijk [12] sugere uma abordagem alternativa, procurando simplificar o trabalho computacional. Substitui no problema (B) a restrição $\langle d, d \rangle = 1$ pela $\langle d, d \rangle \leq 1$, com isso o conjunto de soluções passa ser convexo e a solução ótima não se modifica, já que estamos maximizando $\langle \nabla f(X), d \rangle$. Incorporando as restrições de não-negatividade, o problema (B) pode ser colocado na forma;

$$\begin{aligned} \text{Max } D_r f(X) &= \langle \nabla f(X), d \rangle \\ \text{sujeito a } \quad M_X d &\leq 0 \\ \langle d, d \rangle &\leq 1 \end{aligned} \quad (D)$$

Aplicando as condições de Kuhn-Tucker e utilizando-se ainda do método simplex adaptado ao problema resultante, Zontendijk desenvolveu uma metodologia no qual as matrizes básicas são de ordem $m_1 + m_2 + m_j$ a cada interação.

Lemke [2], descreve um método ligeiramente diferente para resolver (C), mas neste caso as matrizes são de ordem n a cada passo, o que representa uma vantagem, pois depen

dendo da sucessão de pontos viáveis o processo de Zontendijk trabalhará com matrizes consideravelmente menores.

Estamos vendo portanto, que a busca da direção ótima gera um problema de otimização a cada passo, que por sua vez é o responsável pelo grande esforço computacional. Principalmente por isso, surgiram os métodos de direção viável não ótima, que basicamente procuram determinar uma direção viável que garanta tão somente a relação $\langle \nabla f(X), d \rangle > 0$, pois a partir desta, sabemos que um acréscimo pode ser obtido para a função $f(X)$. Tais métodos, embora sejam mais simples computacionalmente, podem criar problemas de convergência do tipo Zig-Zag, que representa uma convergência para um ponto limite não satisfazendo Kuhn - Tucker. Dentro desta classe, aquele que mais se destaca pela simplicidade computacional é o Gradient Projection Method desenvolvido por Rosen [7].

Retornemos as condições (A) de viabilidade para d , com as seguintes modificações:

$$\langle L_k, d \rangle = 0 \quad K \in \{I_1 + I_2(X)\}$$

$$\langle e_j, d \rangle = d_j = 0 \quad j \in J(X)$$

ou matricialmente, $M_x \cdot d = 0$

Onde M_x representa a matriz formada pelas linhas L_k com $K \in \{I_1 + I_2(X)\}$ e pelas linhas e_j com $j \in J(X)$. A projeção \bar{d}_x de $\nabla f(X)$ no espaço complementar e ortogonal ao gerado pelas linhas de M_x , satisfaz $M_x \bar{d}_x = 0$ e ademais apresenta a interessante propriedade $\langle \nabla f(X), \bar{d}_x \rangle \geq 0$.

A direção \bar{d}_x é obtida por:

$$\bar{d}_x = P_x \cdot \nabla f(X)$$

onde P_X é a matriz de projeção dada por:

$$P_X = I - Q_X (Q_X^T \cdot Q_X)^{-1} Q_X^T$$

com $Q_X = (q_X^1, q_X^2, \dots, q_X^n)$ onde q_X^k são as colunas linearmente independentes da matriz M_X^T . A simplicidade do método iterativo é devido principalmente ao fato de que a matriz M_X sofre pequenas variações, já que, ou permanece inalterada ou entra e/ou sai um vetor, entre iterações sucessivas. Tal fato facilita o uso de fórmulas de recorrência para a obtenção das matrizes de projeção sucessivas. O processo iterativo fica interrompido quando para algum X viável se tenha $\bar{d}_X = 0$. Neste caso, $\nabla f(X)$ é combinação linear das linhas de M_X , ou das colunas de Q_X ,

$$\nabla f(X) = Q_X \cdot \alpha_X$$

$$\text{com } \alpha_X = (Q_X^T Q_X)^{-1} Q_X^T \nabla f(X)$$

Se $\alpha_{X_k} \geq 0 \quad k \notin I_1$, então são cumpridas as condições de Kuhn-Tucker, e X é um ponto estacionário restrito. Caso contrário, isto é se $\exists \alpha_{X_k} < 0 \quad k \notin I_1$ é possível, relaxando a restrição k -ésima, obter a projeção de $\nabla f(X)$ no sub-espço complementar e ortogonal ao espaço gerado pelos vetores anteriores com exceção de L_k . Esta nova projeção \bar{d}_X^k atenderá as condições de viabilidade com acréscimo da função $f(X)$. Infelizmente, este algoritmo no modo simples como foi apresentado por Rosen, não tem a sua convergência garantida, e pode de acordo com Luemberg [4] teoricamente apresentar o fenômeno de Zig-Zag, embora na prática o método venha sendo largamente utilizado e jamais apresentou este inconveniente. Além disso, nenhum exemplo, nem mesmo teórico, de

Zig-Zag foi apresentado até hoje. Uma versão mais complicada do algoritmo, foi proposta por Polakj E. [6] que garante a convergência se $f(X)$ é concava e se algumas condições de regularidade são atendidas. Entretanto, mesmo nestes casos especiais de convergência, o algoritmo perde a sua mais importante característica que é a simplicidade, deixando de ser competitivo com outros métodos da classe não ótimo como por exemplo o Convex Simplex Method cuja convergência é garantida para problemas do tipo (I). Este algoritmo é devido a Wolfe [8] e aprimorado por Zangwill [10]. Historicamente e didaticamente, o referido algoritmo pode ser apresentado como um caso particular do Reduced Gradiente Method, que é um algoritmo não convergente no caso geral do problema (I) e que apresentaremos a seguir.

Transformando as desigualdades em igualdades, através das variáveis de folga, nosso problema original pode tomar a seguinte forma matricial:

$$\begin{aligned} \text{Max } f(X) \quad X &= (x_1, x_2, \dots, x_n) \\ \text{sujeito a } L \cdot X &= b \\ x &\geq 0 \end{aligned} \quad (\text{II})$$

Onde L é matriz de ordem $m \times n$.

Dado um ponto X viável para (II), consideremos a partição $X = (Y, Z)$ com $Y > 0$. Y é dito vetor de variáveis básicas ou dependentes, e Z o vetor de variáveis não básicas ou independentes.

A condição, $Y > 0$ poderá ser assegurada sempre que a hipótese de não degeneração puder ser garantida. Então (II),

pode ser escrito:

$$\text{Max } f(Y, Z)$$

$$\text{sujeito } \bar{M} \cdot Y + N \cdot Z = b$$

$$Y \geq 0$$

$$Z \geq 0, \text{ com } L = |M, N|$$

$$\text{Como, } Y = M^{-1} b - M^{-1} \cdot N \cdot Z$$

então, $f(X) = f(M^{-1} b - M^{-1} \cdot N \cdot Z, Z) = g(Z)$ e $\nabla g(Z)$ pode ser calculado. Assim, nosso problema ficou reduzido a encontrar uma direção r viável para Z e que mantenha a viabilidade de $Y \geq 0$ com acréscimo da função $g(Z)$. Dado um acréscimo r para Z , temos que $\Delta Y = -M^{-1} \cdot N \cdot r$. Portanto o procedimento para a determinação de r será:

$$r_i = \begin{cases} \nabla g_i(Z) & \text{se } \nabla g_i(Z) \text{ ou } Z_i > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Em seguida, devemos determinar os parâmetros $\alpha_1, \alpha_2, \alpha_3$ definidos por:

$$\alpha_1 = \max \{ \alpha / Y + \alpha \Delta Y \geq 0 \}$$

$$\alpha_2 = \max \{ \alpha / Z + \alpha r \geq 0 \}$$

$$\text{e } \alpha_3 \text{ maximiza } f(X + \alpha \Delta X) \text{ com } 0 < \alpha < \min \{ \alpha_1, \alpha_2 \}$$

Então o novo ponto será $X + \alpha_3 \Delta X$, se $\alpha_3 < \alpha_1$, recalcula-se ΔZ com a mesma partição anterior, caso contrário uma das variáveis básicas será anulada e uma nova partição deverá ser organizada. Para garantir a convergência do método, uma simples modi

ficação na escolha da direção r torna-se necessária, o que dará origem a um algoritmo denominado Convex Simplex Method. A principal diferença entre ele e o Reduced Gradient é que enquanto neste a direção r pode modificar todas as variáveis independentes, no CSM somente uma variável independente é modificada a cada passo.

Dado um ponto viável $X = (Y, Z)$ já devidamente particionado, a direção r será definida pela variável independente Z_i da seguinte forma:

$$r_{i_1} = \max_i \{r_i\}$$

$$r_{i_2} \cdot Z_{i_2} = \min_i \{r_i \cdot Z_i\}$$

Se $r_{i_1} = r_{i_2} \cdot Z_{i_2} = 0$, o ponto X solução. Por outro lado se:

$$r_{i_1} \geq |r_{i_2} \cdot Z_{i_2}| : \text{aumentar } Z_{i_1}$$

$$r_{i_1} \leq |r_{i_2} \cdot Z_{i_2}| : \text{diminuir } Z_{i_2}$$

Tal procedimento torna convergente, ver Zangwill [10], o algoritmo proposto. É importante resaltar que, a hipótese de não degeneração é muito importante, já que na ocorrência desta o algoritmo CSM pode não convergir.

II. Fundamentos Teóricos

O problema de programação não linear com restrições lineares pode ser colocado na seguinte forma geral:

$$\begin{array}{ll}
 \text{Max } f(X) & X = (x_1, x_2, \dots, x_n) \\
 \text{sujeito } \bar{a} & \begin{array}{ll}
 \langle L_k, X \rangle = b_k & k \in I \\
 \langle L_k, X \rangle \leq b_k & k \in J
 \end{array}
 \end{array} \quad (1)$$

Com I e J representando respectivamente os conjuntos de índices para as restrições de igualdade e desigualdades. Onde para todo k, L_k é vetor de coeficientes pertencente a E^n e b_k é número real.

O caso particular de restrição de não-negatividade do tipo $x_i \geq 0$ estará considerada nas restrições do tipo J, fazendo $L_k = -e_i$ onde e_i representa o i-ésimo vetor da base Euclidiana de E^n .

Assumiremos a hipótese que a função $f: E^n \rightarrow R$, possui derivadas de primeira ordem contínua em todo ponto do conjunto de soluções.

Notaremos,

χ : conjunto das soluções. $X \in \chi \leftrightarrow X$ satisfaz (1)

Γ : subespaço de E^n gerado pelos vetores $\{L_k/k \in I\}$

Ω : subespaço de E^n ortogonal e complementar a Γ .

$\omega \in \Omega \leftrightarrow \langle L_k, \omega \rangle = 0$ para todo $k \in I$.

$A(X)$: conjunto de índices associados a $X \in \chi$ e corresponden-
te às desigualdades ativas para X , isto é:

$k \in A(X) \leftrightarrow k \in J$ e $\langle L_k, X \rangle = b_k$

Para evitar complicações desnecessárias, vamos supor que a dimensão do subespaço vetorial Ω dado por

$$d(\Omega) = n - d(\Gamma)$$

é diferente de zero, $d(\Omega) > 0$.

Notemos que se $I = \emptyset$ então $\Omega = E^n$.

Definição 1

Dado $X \in \chi$, uma direção $\beta \in E^n$ é viável para X , se existir u_* real positivo tal que para todo $0 < u \leq u_*$ implicar que $X + u\beta \in \chi$.

Proposição 1

Dado $X \in \chi$, então $\beta \in E^n$ é direção viável para X se e somente se $\beta \in \Omega$ e $\langle L_k, \beta \rangle \leq 0$ com $k \in A(X)$

Dem.

\rightarrow) Se $\beta \in E^n$ é direção viável para $X \in \chi$ então $\exists u_*$ tal que,

$X + u\beta \in \chi$ para $0 < u \leq u_*$, logo

para $k \in I$; $b_k = \langle L_k, X + u\beta \rangle = b_k + u \langle L_k, \beta \rangle$ com $0 < u \leq u_*$

o que implica que $\langle L_k, \beta \rangle = 0$ e portanto $\beta \in \Omega$; para

$k \in J$, somente as desigualdades ativas são restritivas pa

ra a viabilidade de β e neste caso $k \in A(X)$ teremos

$$\langle L_k, X \rangle + u \langle L_k, \beta \rangle = b_k + u \langle L_k, \beta \rangle \leq b_k$$

o que implica que $\langle L_k, \beta \rangle \leq 0$

\leftarrow) Se $\beta \in \Omega$ e $\langle L_k, \beta \rangle \leq 0 \quad k \in A(X)$, então é fácil verificar que é sempre possível encontrar u_* positivo tal que para todo $0 < u \leq u_*$ se verifica a desigualdade $u \langle L_k, \beta \rangle \leq b_k - \langle L_k, X \rangle$ para $k \in J - A(X)$

Proposição 2

Dado $X \in \chi$, o conjunto de direções viáveis $D(X)$ associado a X é um subconjunto de Ω definido por

$$D(X) = \{\beta / \beta \in \Omega; \langle L_k, \beta \rangle \leq 0 \quad k \in A(X)\}$$

Se $A(X) = \emptyset$, então $D(X) = \Omega$.

Dem.

Consequência imediata da Prop 1.

Proposição 3

Seja C um cone gerado por uma classe finita de vetores $\{\xi_1, \xi_2, \dots, \xi_p\}$ de E^n . Dado $\sigma \in E^n$, então a desigualdade $\langle \sigma, \xi \rangle \leq 0$ ocorre para todo $\xi \in C$ se e somente se ocorre para todo $\xi_i, i = \overline{1, p}$.

Dem.

$$\text{Se } \xi \in C \text{ então } \xi = \sum_{i=1}^p a_i \xi_i \quad a_i \geq 0 \quad i = \overline{1, p}$$

$$\text{Logo } \langle \sigma, \xi \rangle = \sum_{i=1}^p a_i \langle \sigma, \xi_i \rangle$$

→) Se $\langle \sigma, \xi \rangle \leq 0$ ocorre para todo $\xi \in C$, então como

$$\xi_i \in C \quad i = \overline{1, p} \rightarrow \langle \sigma, \xi_i \rangle \leq 0 \quad i = \overline{1, p}$$

←) Se $\langle \sigma, \xi_i \rangle \leq 0 \quad i = \overline{1, p}$, então $\xi \in C$ tem-se

$$\langle \sigma, \xi \rangle = \sum_{i=1}^p a_i \langle \sigma, \xi_i \rangle \leq 0 \quad \text{já que } a_i \geq 0 \quad i = \overline{1, p}$$

Proposição 4

Dada uma classe finita de vetores $\{\xi_1, \xi_2, \dots, \xi_p\}$ de um subespaço vetorial Ω de E^n , consideremos a classe C^* de vetores $\sigma \in \Omega$, tal que a desigualdade $\langle \sigma, \xi_i \rangle \leq 0$ ocorra para todo $i = \overline{1, p}$. Então a classe C^* constitui um cone convexo fechado, dito dual em Ω do cone C gerado pela classe de vetores $\{\xi_1, \xi_2, \dots, \xi_p\}$.

Dem.

Seja $\xi \in C$ e sejam σ_1 e σ_2 pertencente a C^* .

Então com $a_1 \geq 0$ e $a_2 \geq 0$ temos que

$$\langle \xi, a_1 \sigma_1 + a_2 \sigma_2 \rangle = a_1 \langle \xi, \sigma_1 \rangle + a_2 \langle \xi, \sigma_2 \rangle \leq 0$$

portanto, $a_1 \sigma_1 + a_2 \sigma_2 \in C^*$ o que prova que C^* é cone convexo.

Seja σ o limite de uma sequência $\{\sigma_q\}$ de vetores de C^* , então como a função $\langle \cdot, \cdot \rangle$ é contínua em E^n , temos que:

$$\langle \xi, \sigma \rangle = \lim_{q \rightarrow \infty} \langle \xi, \sigma_q \rangle \leq 0 \rightarrow \sigma \in C^*$$

portanto C^* é cone convexo fechado.

Neste ponto é conveniente lembrar que Ω e Γ são subespaços vetoriais ortogonais e complementares, então todo vetor $L \in E^n$ é univocamente expresso da forma $L = \bar{L} + \underline{L}$ onde $\bar{L} \in \Omega$ e $\underline{L} \in \Gamma$. Portanto na Prop 2, como $\beta \in \Omega$ a condição $\langle L_k, \beta \rangle \leq 0$ pode ser substituída por $\langle \bar{L}_k, \beta \rangle \leq 0$ $k \in A(X)$.

Notando por $C(X)$ o cone convexo fechado associado a $X \in \chi$ e gerado pelos vetores $\{\bar{L}_k/k \in A(X)\}$ podemos concluir das proposições anteriores que

Proposição 5

Dado $X \in \chi$, então $D(X) = C^*(X)$

Seja $\nabla f(X)$ o gradiente da função f no ponto $X \in \chi$ e notemos por $\bar{\nabla}f(X)$ e $\underline{\nabla}f(X)$ as projeções de $\nabla f(X)$ nos subespaços Ω e Γ , respectivamente.

Proposição 6

Seja $X \in \chi$, se existir $\beta \in D(X)$ tal que

$$\langle \bar{\nabla}f(X), \beta \rangle > 0$$

Então, existirá $t_* > 0$ tal que para todo $0 < t \leq t_*$

$$f(X + t\beta) > f(X)$$

Dem.

Considerando que $\beta \in D(X) \subseteq \Omega$, então é verdade que $\langle \bar{\nabla} f(X), \beta \rangle = \langle \nabla f(X), \beta \rangle$. Usando a definição de direção viável sabemos que $\exists u_* > 0$ tal que para todo $0 < u \leq u_* \rightarrow (X + u\beta) \in \chi$.

Da hipótese de que $f: E^n \rightarrow R$ possui derivada de primeira ordem contínua, implica que é diferencial e que portanto,

$$\lim_{v \rightarrow 0} \frac{f(X + v\beta) - f(X)}{v} = \langle \nabla f(X), \beta \rangle > 0$$

Logo, $\exists v_* > 0$ tal que para todo $0 < v < v_*$

temos

$$\frac{f(X + v\beta) - f(X)}{v} > 0$$

fazendo $t_* = \min \{u_*, v_*\}$ a proposição fica demonstrada

Neste estágio do trabalho estamos de posse da caracterização do conjunto de direções viáveis dada pela Proposição 5, e sabemos também através da Proposição 6, que a partir de um ponto $X \in \chi$ e de uma direção viável β com a propriedade $\langle \bar{\nabla} f(X), \beta \rangle > 0$, poderemos "caminhar" dentro do conjunto das soluções com acréscimo da função $f(X)$.

As proposições que se seguem nos ajudarão a determinar a direção viável ótima, aquela que maximiza a de

derivada direcional de $f(X)$ em $D(X)$, ou ficaremos sabendo que não existe β viável com $\langle \bar{\nabla} f(X), \beta \rangle > 0$

Proposição 7

Dado um cone convexo e fechado C em um subespaço vetorial $\Omega \subseteq E^n$, seja C^* seu dual em Ω .

Então todo vetor $\xi \in \Omega$ pode ser univocamente expresso da forma: $\xi = \alpha + \beta$ onde $\alpha \in C$ e $\beta \in C^*$ com $\langle \alpha, \beta \rangle = 0$.

Dem.

Consideremos um vetor $\xi_0 \in \Omega$ e seja α_0 o vetor em C mais próximo de ξ_0 , isto é, o vetor que minimiza a função,

$$g(\alpha) = |\xi_0 - \alpha|^2, \quad \alpha \in C$$

Se $\alpha \in C$, então $\alpha_0 + t\alpha$ com $0 \leq t < \infty$ também pertence a C , fazendo $\beta_0 = \xi_0 - \alpha_0$, temos que a função,

$$r(t) = g(\alpha_0 + t\alpha) = |\beta_0 - t\alpha|^2 = |\beta_0|^2 - 2t\langle \beta_0, \alpha \rangle + t^2 |\alpha|^2$$

definida para $t \geq 0$, tem um mínimo em $t = 0$, logo a derivada de $r(t)$ no ponto $t = 0$ será necessariamente,

$$r'(0) = -2 \langle \beta_0, \alpha \rangle \geq 0$$

Como α é arbitrário em C , podemos afirmar que o vetor $\beta_0 \in C^*$. Se $\alpha_0 = 0$, então $\langle \alpha_0, \beta_0 \rangle = 0$. Se $\alpha_0 \neq 0$ podemos fazer $\alpha = \alpha_0$ na função $r(t)$. Neste caso como $\alpha_0 + t\alpha_0 = (1+t)\alpha_0$ pertencerá a C sempre que $t \geq -1$, en-

tão a função $r(t)$ tem um mínimo para $t = 0$. Consequentemente,

$$r'(0) = -2 \langle \beta_0, \alpha_0 \rangle = 0$$

o que implica dizer que α_0 e β_0 são ortogonais.

A prova da unicidade é conseguida por redução ao absurdo. Vamos supor que existam $\alpha \in C$ e $\beta \in C^*$ tais que $\langle \alpha, \beta \rangle = 0$ com $\xi_0 = \alpha + \beta$. Então,

$$|\alpha|^2 = \langle \alpha, \xi_0 \rangle = \langle \alpha, \alpha_0 + \beta_0 \rangle = \langle \alpha, \alpha_0 \rangle + \langle \alpha, \beta_0 \rangle \leq \langle \alpha, \alpha_0 \rangle$$

$$|\alpha_0|^2 = \langle \alpha_0, \xi_0 \rangle = \langle \alpha_0, \alpha + \beta \rangle = \langle \alpha_0, \alpha \rangle + \langle \alpha_0, \beta \rangle \leq \langle \alpha, \alpha_0 \rangle$$

Essas duas desigualdades são possíveis somente se $\alpha = \alpha_0$, e então $\beta = \beta_0$.

Proposição 8

Dado $X \in \chi$, então $\bar{\nabla}f(x)$ é univocamente expresso da forma $\bar{\nabla}f(X) = \alpha_X + \beta_X$, onde $\alpha_X \in C(X)$ e $\beta_X \in D(X)$ com $\langle \alpha_X, \beta_X \rangle = 0$. Além disso, a direção β_X maximiza a derivada direcional de $f(X)$ em $D(X)$.

Dem.

Como $C(X)$ é um cone convexo fechado em Ω , gerado por uma classe finita de vetores de Ω , e como $D(X)$ pela Proposição 5, constitui o seu dual em Ω , segue que $\bar{\nabla}f(X) \in \Omega$ pode ser colocado pela Proposição 7 na forma apresentada. Pa

ra todo $\beta \in D(X)$, temos que:

$$\langle \bar{\nabla}f(X), \beta \rangle = \langle \beta_X, \beta \rangle + \langle \alpha_X, \beta \rangle$$

como, $\langle \alpha_X, \beta \rangle \leq 0$

$$\text{para } \beta \neq 0, \quad \frac{\langle \bar{\nabla}f(X), \beta \rangle}{|\beta|} \leq \frac{\langle \beta_X, \beta \rangle}{|\beta|} \leq |\beta_X|$$

lembrando que $\langle \nabla f(X), \beta \rangle = \langle \bar{\nabla}f(X), \beta \rangle$

$$\text{e que, } |\beta_X| = \frac{\langle \bar{\nabla}f(X), \beta_X \rangle}{|\beta_X|} \quad \text{se } \beta_X \neq 0$$

temos finalmente que

$$\frac{\langle \nabla f(X), \beta \rangle}{|\beta|} \leq \frac{\langle \nabla f(X), \beta_X \rangle}{|\beta_X|}$$

Se $\beta_X = 0$, então $\bar{\nabla}f(x) = \alpha_X \in C(X)$ o que implica que para todo β viável $\langle \bar{\nabla}f(X), \beta \rangle \leq 0$

É de extrema importância para o algoritmo que estamos desenvolvendo, a análise detalhada de casos em que dado $X \in \chi$ tenhamos $\beta_X = 0$, e de outras situações particulares que se seguem;

a) Se $\beta_X = 0$ então $\bar{\nabla}f(X) = \alpha_X$, isto é, $\bar{\nabla}f(X) \in C(X)$ e portanto para todo $\beta \in D(X)$ que é cone dual de $C(X)$ em Ω se verifica que $\langle \bar{\nabla}f(X), \beta \rangle \leq 0$.

Neste caso, $\beta_X = 0$, é interessante observar que $X \in \chi$ satisfaz as condições de Kuhn-Tucker para o nosso problema. De fato, como todo $X \in \chi$ atende às condições de regularidade e se verificam as relações;

$$\bar{\nabla}f(X) = \alpha_X \in C(X) \rightarrow \bar{\nabla}f(X) = \sum_{k \in \bar{A}(X)} a_k \cdot \bar{L}_k \text{ com } a_k \geq 0$$

para todo $k \in A(X)$. Ainda,

$$\bar{L}_k = L_k - \bar{L}_k \quad k \in A(X) \text{ e como } \bar{L}_k \in \Gamma \rightarrow \bar{L}_k = \sum_{k \in I} c_k \cdot L_k$$

Então o gradiente, $\nabla f(X) = \bar{\nabla}f(X) + \bar{\bar{\nabla}}f(X)$ pode ser colocado da seguinte forma

$$\nabla f(X) = \sum_{k \in \bar{A}(X)} \lambda_k L_k + \sum_{k \in I} \lambda_k L_k$$

com $\lambda_k \geq 0 \quad k \in A(X)$

O que mostra o atendimento das condições de Kuhn-Tucker.

- b. Um caso particular se verifica quando $\bar{\nabla}f(X) = 0$, pois tal fato implica que $\nabla f(X) \in \Gamma$ e portanto pode ser escrito da forma,

$$\nabla f(X) = \sum_{k \in I} \lambda_k L_k$$

o que representa a condição de Kuhn-Tucker.

- c. Obviamente, quando $\nabla f(X) = 0$ o ponto X , mais que satisfaz as condições de Kuhn-Tucker, indica a existência de um ponto crítico irrestrito.

d. Quando $X \in \chi$ é tal que $A(X) = 0$, então temos a identidade $D(X) = \Omega$ e portanto a direção viável β_X será o próprio $\bar{\nabla}f(X)$.

III. O Algoritmo

As proposições apresentadas sugerem naturalmente o seguinte procedimento iterativo para a busca de um ponto de χ que satisfaça as condições de Kuhn-Tucker. Dado $X \in \chi$

1. Calcular $\nabla f(X)$. Se $\nabla f(X) = 0$ pare. X é um ponto satisfazendo K - T.
2. Calcular $\bar{\nabla} f(X)$. Se $\bar{\nabla} f(X) = 0$ pare. X é um ponto satisfazendo K - T.
3. Calcular os vetores \bar{L}_k para $k \in A(X)$. Se $A(X) = \emptyset$, então fazer $\bar{\nabla} f(X) = \beta_X$ e ir para o item 5.
4. Calcular β_X . Se $\beta_X = 0$ pare. X é um ponto satisfazendo K - T.
5. Calcular u_* e t_* respectivamente

$$u_* = \max_u \{u \mid X + u\beta_X \in \chi\}$$

$$t_* \text{ maximiza } \{f(X + t\beta_X) \mid t \in (0, u_*]\}$$

6. Retornar ao item 1, com o novo ponto de χ dado por $X + t_*\beta_X$, se t_* é finito. Caso contrário pare, solução ilimitada.

III.1. A Convergência

Seja $\{X_n\}$ sequência de pontos de χ gerado pela aplicação sucessiva do algoritmo. Observemos que a se-

sequencia $\{X_n\}$ terminará após um número finito de interações, se ja X_F este ponto, somente se:

a. a direção $\beta_F = 0$; fazendo $\beta_F = \beta_{X_F}$

b. para $\beta_F \neq 0$, se tenha que $X_F + t\beta_F \in \chi$ para todo $t > 0$ e $t_* = \infty$.

No caso a. o ponto X_F é um ponto satisfazendo K. - T., e o caso b. ficamos sabendo que existe uma solução ilimitada.

Um algoritmo é dito ser teoricamente convergente para um dado tipo de problema se a partir de um ponto solução qualquer as seguintes condições são satisfeitas.

i. Se a sequencia $\{X_n\}$ terminar num ponto X_F após um número finito de interações, então implica que X_F ou é um ponto satisfazendo K - T, ou a partir de X_F uma solução ilimitada é evidenciada.

ii. Se a sequencia $\{X_n\}$ convergir para um ponto limite X_* , então implica que X_* é um ponto de χ satisfazendo K - T.

Devido as observações a. e b. anteriores, a condição i obviamente verifica. Assim sendo, resta-nos analisar as características do ponto limite X_* da sequencia $\{X_n\}$, isto é, para que nosso algoritmo seja convergente basta-nos mostrar que a condição ii. também é atendida. Observemos inicialmente que

pela continuidade da função $\langle \cdot, \cdot \rangle$ o ponto $X_* \in \chi$. Considere mos que $X_* \bar{\epsilon}$ um ponto não satisfazendo K. - T., isto $\bar{\epsilon}$, vamos supor que exista $\beta_* \neq 0$, onde para simplificar a nomenclatura fazemos $\beta_{X_*} = \beta_{X_*}$. Mostraremos primeiramente que existirá N^1 implicará que $\beta_* \in D(X_n)$ tal que para todo $n > N^1$

Como β_* $\bar{\epsilon}$ direção viável para X_* , então por defi nição existirá u_* tal que para todo $0 < u \leq u_*$ se verifica que $X_* + u\beta_* \in \chi$. Consideremos as restrições originais (1) para X_n na direção β_* ,

$$\langle L_k, X_n + u\beta_* \rangle = \langle L_k, X_n \rangle + u \langle L_k, \beta_* \rangle$$

como $\beta_* \in \Omega$, então para todo $k \in I$, e para todo n

$$\langle L_k, X_n \rangle + u \langle L_k, \beta_* \rangle \leq b_k \text{ para } u \text{ real.}$$

Para $k \in J$, temos que passando ao limite

$$\lim_{n \rightarrow \infty} \langle L_k, X_n + u\beta_* \rangle = \lim_{n \rightarrow \infty} \langle L_k, X_n \rangle + u \langle L_k, \beta_* \rangle =$$

$$= \langle L_k, X_* \rangle + u \langle L_k, \beta_* \rangle \leq b_k \quad k \in J$$

então para cada $k \in J$, $\exists N_k$ tal que para todo $n > N_k$ tem-se:

$$\langle L_k, X_n + u\beta_* \rangle \leq b_k \text{ para } 0 < u \leq u_k^*$$

Portanto, escolhendo $N^1 = \max\{N_k / k \in J\}$, temos que β_* é direção viável para todo X_n com $n > N^1$.

Ademais, $\lim_{n \rightarrow \infty} \langle \nabla f(X_n), \beta_* \rangle = \langle \nabla f(X_*), \beta_* \rangle > 0$, então $\exists N^2$ tal que para todo $n > N^2$ implica que $\langle \nabla f(X_n), \beta_* \rangle > 0$. Fazendo $N^3 = \max\{N^1, N^2\}$, tem-se que β_* é direção viável com acréscimo da função f para todo X_n com $n > N^3$

Sem prejuízo de generalidade, podemos supor que as direções β obtidas em cada passo do algoritmo serão vetores unitários $|\beta| = 1$. Neste caso a propriedade de máxima derivada direcional para β segundo a Proposição 8, poderá ser colocada da forma

$$\langle \nabla f(X_n), \beta_n \rangle \geq \langle \nabla f(X_n), \beta_* \rangle \quad \text{para } n > N^1$$

onde β_n representa a direção ótima para X_n . Então como f é diferenciável, temos que para $n > N^1$

$$\lim_{t \rightarrow 0} \frac{f(X_n + t\beta_n) - f(X_n)}{t} = \langle \nabla f(X_n), \beta_n \rangle \geq \langle \nabla f(X_n), \beta_* \rangle$$

Fazendo $n > N^3$, podemos escrever que,

$$\lim_{t \rightarrow 0} \frac{f(X_n + t\beta_n) - f(X_n)}{t} \geq \langle \nabla f(X_n), \beta_* \rangle > 0$$

Como $\lim_{n \rightarrow \infty} \langle \nabla f(X_n), \beta_* \rangle = \langle \nabla f(X_*), \beta_* \rangle > 0$, podemos finalmente dizer que $\exists t_* > 0$ tal que para todo $0 < t \leq t_*$ e para todo $n > N_*$ se verifica a desigualdade

$$f(X_n + t_* \beta_n) - f(X_n) > c \frac{t_*}{e} > 0$$

Como, X_{n+1} é igual a $(X_n + t_* \beta_n)$ temos finalmente que para todo $n > N_*$ a diferença entre $f(X_{n+1})$ e $f(X_n)$ é pelo menos maior que uma constante positiva, que junto com a hipótese de continuidade de $f(x)$ contraria a hipótese da convergência da sequência $\{X_n\}$ para o ponto X_* .

Assim sendo o algoritmo estudado é teoricamente convergente para o problema proposto sempre que a função $f: E^n \rightarrow R$ possuir as primeiras derivadas contínuas. Esta última hipótese de continuidade é fundamental para a demonstração da convergência. No problema apresentado a seguir, proposto por Wolfe [9], o algoritmo pode convergir para um ponto X_* não satisfazendo as condições K.-T.

$$\text{Max } f(X) = -\frac{4}{3}(X_1^2 - X_1 X_2 + X_2^2)^{3/4} - X_3$$

$$\text{sujeito a } X_1, X_2, X_3 \geq 0$$

O que realmente ocorre quando aplicamos nosso algoritmo partindo do ponto inicial $X_0 = (1/4, 0, 1)$. O vetor gradiente tem a fórmula geral

$$\nabla f(X) = \left(\frac{-2X_1 + X_2}{(X_1^2 - X_1 X_2 + X_2^2)^{1/4}}, \frac{X_1 - 2X_2}{(X_1^2 - X_1 X_2 + X_2^2)^{1/4}}, -1 \right)$$

Como $I = \emptyset$, temos neste caso que $\bar{\nabla} f(X) = \nabla f(X)$.

Além disso, ocorrerá que para todo ponto X_n da sequência gerada pelo algoritmo a direção viável ótima coincidirá com o gradiente, $\beta_n = \nabla f(X_n)$. Como resultado do processo podemos verificar que a sequência convergirá para o ponto

$$X_* = (0, 0, 1 - \frac{1}{4}(1 - \frac{1}{\sqrt{2}} + \frac{1}{2} + \frac{1}{2\sqrt{2}} + \dots)) = (0, 0, \frac{2 - \sqrt{2}}{4}) \approx (0, 0, .146)$$

que evidentemente não pode ser um ponto de K.-T., já que $\beta = (0, 0, -1)$ é direção viável para X_* com acréscimo da função $f(X)$. Observemos que $\nabla f(X)$ não satisfaz a condição de continuidade no ponto X_* , que como dissemos é de grande importância para garantir a convergência do algoritmo. Pode-se verificar ainda no exemplo, que a convergência $\{X_n\} \rightarrow X_*$ se faz através de um processo de zig-zag entre os planos $X_1 = 0$ e $X_2 = 0$, que é um fenômeno indesejável e com possibilidade de ocorrência em algoritmos do tipo direção-viável. Mesmo que o algoritmo seja convergente, como é o nosso caso, a aparecimento do zig-zag retarda muito a convergência do processo e deve portanto ser evitado. Técnicas para reparar este inconveniente são conhecidas desde 1960, quando Zoutendijk, que foi o primeiro a publicar trabalho nesta área, sugere em seu texto ^[12] uma solução para o problema. O mesmo Zoutendijk, em sua publicação ^[13], apresenta melhorias da sua idéia original. Outras técnicas, como a de Wolfe ^[9] e de McCormick ^[5]

são dirigidas para evitar o fenômeno do zig-zag.

III.2. A Implementação

A continuação, algumas considerações serão feitas com relação a implementação do algoritmo apresentado.

Vimos que na aplicação é requerido em diferentes itens o cálculo da projeção de vetores no subespaço Ω e a determinação de β_X . Trataremos de fazer alguns comentários a respeito do esforço operacional exigido.

Consideremos uma base $\{L_1, L_2, \dots, L_m\}$ do subespaço vetorial Γ . Como já dissemos todo o vetor de E^n , por exemplo $\nabla f(X)$, pode ser colocado na forma

$$\nabla f(X) = \bar{\nabla} f(X) + \bar{\bar{\nabla}} f(X)$$

onde $\bar{\nabla} f(X) \in \Omega$ e $\bar{\bar{\nabla}} f(X) \in \Gamma$ com $\langle \bar{\nabla} f(X), \bar{\bar{\nabla}} f(X) \rangle = 0$

$$\text{Então, } \bar{\bar{\nabla}} f(X) = \sum_{i=1}^m u_i L_i$$

Fazendo, L matriz $n \times m$ cujas colunas são os vetores L_i ($i=1, m$) da base de Γ , podemos escrever

$$\nabla f(X) = \bar{\nabla} f(X) + L.U$$

onde U é vetor $m \times 1$, $U^T = (u_1, u_2, \dots, u_m)$.

Multiplicando a igualdade anterior por L^T , resulta

$$L^T \cdot \nabla f(X) = L^T \cdot L \cdot U$$

pois que $L^T \cdot \bar{\nabla} f(X) = 0$. Lembrando que $L^T \cdot L$ é não-singular, podemos explicitar U

$$U = (L^T \cdot L)^{-1} \cdot L^T \cdot \nabla f(X)$$

Portanto, $\bar{\bar{\nabla}} f(X) = LU = L \cdot (L^T \cdot L)^{-1} \cdot L^T \cdot \nabla f(X)$

e também,

$$\bar{\nabla} f(X) = (I - L \cdot (L^T \cdot L)^{-1} \cdot L^T) \nabla f(X)$$

A matriz $P = (I - L \cdot (L^T \cdot L)^{-1} \cdot L^T)$ é denominado a matriz de projeção de Ω .

Logo para obter $\bar{\nabla} f(X)$ e \bar{L}_k $k \in A(X)$, basta executar o produto matricial

$$\bar{\nabla} f(X) = P \cdot \nabla f(X)$$

$$\bar{L}_k = P \cdot L_k \quad k \in A(X)$$

Notemos que a matriz de projeção P permanece fixa durante o processo iterativo do algoritmo.

A determinação de β_X já exige em sua obtenção um maior esforço, pois que seremos obrigados a resolver um problema de programação quadrática como veremos a seguir.

Recordemos que a decomposição de $\bar{\nabla} f(X)$ se faz na forma,

$$\bar{\nabla} f(X) = \alpha_X + \beta_X$$

Onde α_X , minimiza a função $g(\alpha) = |\bar{\nabla}f(X) - \alpha|^2$ com $\alpha \in C(X)$. Como $C(X)$ representa o cone convexo fechado gerado pelos vetores $\{\bar{L}_k / k \in A(X)\}$, podemos escrever α_X da forma

$$\alpha_X = \sum_{k \in A(X)} y_k \bar{L}_k \quad y_k \geq 0$$

Designando por $L(X)$ a matriz cujas colunas são os vetores \bar{L}_k com $k \in A(X)$ e abandonando a título de simplificação o índice $k \in A(X)$, a determinação de α é equivalente a solução do seguinte problema,

$$\text{Minimizar } g(Y) = |\bar{\nabla}f(X) - L(X)Y|^2$$

$$\text{sujeito a } Y \geq 0 \quad Y = (y_1, y_2, \dots, y_p)$$

Onde p representa o número de índices existentes no conjunto $A(X)$. Desenvolvendo a função $g(Y)$, teremos que o nosso problema pode ser resumido em,

$$\text{Minimizar } Q(Y) = Y^T (L^T(X) \cdot L(X)) Y - 2\bar{\nabla}f(X)^T \cdot L(X) \cdot Y$$

(2)

$$\text{sujeito a } Y \geq 0$$

Notemos que, a matriz da forma quadrática $(L^T(X) \cdot L(X))$ é definida não-negativa e portanto se,

$$\bar{\nabla}f(X)^T \cdot L(X) \leq 0 \rightarrow \text{o ponto } Y_* = 0 \text{ é solução ótima}$$

$$\text{ma } \rightarrow \alpha_X = 0 \rightarrow \bar{\nabla}f(X) = \beta_X.$$

O que era de se esperar, pois se $\bar{\nabla}f(X)^T \cdot L(X) \leq 0$

então implica que $\bar{\nabla}f(X) \in D(X)$ e portanto $\bar{\nabla}f(X) = \beta_X$. Por outro lado, se $Q(Y_*) = 0$ então implica que $\bar{\nabla}f(X) \in C^*$ e portanto $\beta_X = 0$.

Para resolver o problema (2), podemos lançar mão de um algoritmo desenvolvido por C.E.LEMKE em 1962 e apresentado em seu trabalho "A Method for Solution of Quadratic Programs", Management Science, 8, 1962, pp.442-453. (Ver anexo A)

O algoritmo de LEMKE para resolver (2) é provado ser convergente num número finito de iterações. E apresenta ainda a vantagem de admitir em processo iterativo bastante simples.

IV. Conclusões

Embora nosso trabalho tenha enfatizado as questões teóricas do algoritmo proposto, algumas considerações de ordem prática podem ser apresentadas com o intuito de oferecer subsídios que permitam situá-lo, computacionalmente, entre os existentes. É claro que não havendo experiência numérica, nossas observações serão dirigidas para o esforço exigido.

É oportuno lembrar que, por ser nosso algoritmo do tipo de direção viável ótima, este possui a importante propriedade de ser convergente. E o é, sem exigir condições de regularidade muito restritivas, bastando somente que $f(X)$ seja diferenciável no conjunto de soluções.

Quanto a esforço computacional, o algoritmo proposto requer o cálculo da matriz de projeção e a minimização de uma forma quadrática na determinação de β_x . O cálculo da matriz de projeção representa um esforço não repetitivo, é calculada na fase inicial do processo e permanecerá fixa até o fim.

Para determinar β_x , o que deve ser feito a cada passo, lançamos mão do algoritmo de LEMKE que opera de forma muito semelhante ao simplex, invertendo matrizes (Ver anexo A) que se modificam simplesmente pela troca de um vetor.

Dentro do grupo de direção viável ótima, nosso método trabalha com matrizes básicas sempre inferior aos demais, isto porque, nossa ordem será igual a $m_2 + m_j$. Assim a vantagem do método, em relação aos demais, aumenta com o número de igualdades m_1 do problema original. Resta saber, se esta vantagem compensa a otimização de uma forma quadrática em relação ao método de Zontendijk que trabalha com o simplex adaptado e sujei

to a problemas de degeneração.

Quanto ao grupo de não ótima, o método de Rosen é bastante simples e certamente o que exige menos esforço computacional, mesmo trabalhando com matrizes e ordem superior a nossa, o processo é simplificado pelas formas de recorrência. Fica entretanto o problema da convergência que não pode ser resolvido sem romper a simplicidade do método e assim mesmo só garantindo convergência para problemas concavos e sob condições especiais de regularidade. Com relação ao método Convex Simplex, que garante a convergência, se utiliza sempre de matrizes de ordem $m_1 + m_2$ a cada iteração. Outra vez, vai depender do valor de m_1 a vantagem ou não do nosso algoritmo em relação a este, como no caso do método de Zontendijk ocorrendo degeneração a convergência pode ficar comprometida. Já o método Reduced Gradient não garante a convergência e seu esforço computacional é equivalente a C.S.M., sendo portanto superado por este.

V. Obras Citadas

- [¹] - Frank, M., and P. Wolfe, "An Algorithm for Quadratic Programming, Naval Research Logistics Quarterly, III, 95-110, 1965.
- [²] - Lemke, C.E., "The Constrained Gradient Method of Linear Programming", Journal of Society for Industrial and Applied Mathematics, 9, 1-17, 1961.
- [³] - Lemke, C.E., "A Method for Solution of Quadratic Programs, Management Science, 8, 442-453, 1962.
- [⁴] - Luenberger, D., "Introduction to Linear and Nonlinear Programming", Addison - Wiley, 1972.
- [⁵] - McCormick, G.P., "Anti - zigzagging by bending", Management Science 15, 315-320, 1969.
- [⁶] - Polak, E., Computational Methods in Optimizations, Academic Press. New York, 1971.
- [⁷] - Rosen, J.B., "The Gradient Projection Method for Nonlinear Programming, Part I. Linear Constraints", Journal of the Society for Industrial and Applied Mathematics, 8, 181-217, 1960.
- [⁸] - Wolfe, P., "An Extended Simplex Method", Notices of the American Mathematics Society, IX, nº 4, 308 (Abstract), 1962.
- [⁹] - Wolfe, P., "On the convergence of Gradient Methods under Constraints", IBM Research Report RZ-204 Zurich Research Laboratories, Ruschlikon, Zurich, Switzerland, 1966.

- |¹⁰| - Zangwill, W.I., "The Convex Simplex Method", Management Science - A, XIV, n^o 3, 221-238, 1967.
- |¹¹| - Zoutendijk, G., "Maximizing a Function in a Convex Region", Journal of the Royal Statistical Society (B) 21, 338-355, 1959.
- |¹²| - Zoutendijk, G., "Methods of Feasible Direction", Amsterdam, Elsevier, 1960.
- |¹³| - Zoutendijk, G., "Nonlinear Programming: A Numerical Survey", SIAM Journal on Control, IV, n^o 1, 194-210, 1966.

ANEXO AA Method of Solution for Quadratic ProgramsC.E. LEMKEA₁ - Introdução e Transformações

Considere-se o problema de programação quadrática da forma:

$$\text{Min. } - a^T X + \frac{1}{2} X^T Q_0 X, \quad X \in E^m \quad (A)$$

sujeito a $A^T X \leq d_0$, $d_0 \in E^n$

onde, Q_0 é matriz definida positiva.

Sabe-se que o mínimo irrestrito da função objetivo acontece no ponto $X_0 = Q_0^{-1} a$, onde o gradiente $Q_0 X - a$ é anulado. Se além disso X_0 satisfaz as restrições então é solução de (A).

Para que X viável seja solução, as condições de Kuhn-Tucker são necessárias, e no caso (A) também suficientes; devemos ter $(Q_0 X - a) + AY = 0$; $Y \geq 0$ e $Y^T(d_0 - A^T X) = 0$ (1) Então, da primeira igualdade teremos que:

$$X = Q_0^{-1} (a - AY) \quad (2)$$

que nos dá a solução de (A) a partir do conhecimento dos multiplicadores lagrangianos. Usando (2) nas restrições de (A) e utilizando as relações definidas por:

$$Q = A^T \cdot Q_0^{-1} \cdot A, \quad d = d_0 - A^T \cdot Q_0^{-1} \cdot a$$

além de incorporar uma variável de folga, podemos rescrever as condições de otimalidade,

$$- QY + Z = d \quad Y, Z \geq 0 \quad \langle Y^T, Z \rangle = 0 \quad (3)$$

onde Q , é matriz definida não-negativa. Resolvendo (3), podemos então através de (2) obter a solução de (A).

Entretanto, é fácil constatar que o problema (3) representa as condições de Kuhn-Tucker para o seguinte problema quadrático na variável Y .

$$\text{Min } f(Y) = d^T Y + \frac{1}{2} Y^T Q Y \quad \text{sujeito } \bar{a} Y \geq 0 \quad (4)$$

Assim sendo, basta resolver (4) que através (2) temos a solução de (A). Note-se que um ponto viável para (4) é obviamente $Y = 0$, que é também o único vértice do conjunto viável. $Y = 0$ é solução de (4) se e somente se $d \geq 0$. Observemos que não se fará:

- i. hipótese de convexidade estrita para $f(Y)$.
- ii. hipótese de não degeneração nas restrições de (A)

Vamos requerer unicamente a existência de uma solução finita para (4).

A₂ - O Algoritmo

De (3), os pares Y e Z são ótimos se e somente se:

- i. $Y \geq 0$
- ii. $Z = \nabla f(Y)$ em (4), $z \geq 0$
- iii. $\langle Y, Z \rangle = 0$

Em cada iteração, teremos associado ao ponto Y viável um conjunto de n direções independentes. Como no simplex, essas direções serão as linhas da inversa de uma matriz $n \times n$ denominada básica. A cada passo esta matriz será modificada pela simples substituição de uma coluna. Seja B esta matriz básica, $B = (b_1, b_2, \dots, b_n)$ onde b_i representa a i ésima coluna. Notamos por $(B^{-1})^T = (b^1, b^2, \dots, b^n)$ onde b^i repre

seja a i -ésima linha da matriz B^{-1} . Então, tem-se:

$$b_i^T b^i = 1, \quad b_i^T b^j = 0 \quad i \neq j$$

Quando $y_i = 0$ e $b_i = e_i$, diremos que b_i é uma coluna restrita. Inicialmente $Y = 0$ e $B = I$, todas as colunas serão restritas. As demais colunas são ditas livres. Seja R , um subconjunto de índices tal que,

$$b_i \text{ é restrito} \leftrightarrow i \in R$$

Notaremos por b_r a coluna a ser substituída e por \underline{b}_r a substituída. O critério iterativo será baseado em $W = B^{-1}Z = B^{-1}(d + QY)$. Podemos agora definir o algoritmo:

0. Dados iniciais,

$$Y = 0$$

$$B = I = B^{-1}$$

$$b^i = b_i = e_i$$

$$W = Z = d$$

$$R = \{1, 2, \dots, n\}$$

1. Determinação de b_r

Calcular $W = B^{-1}Z$, se o par Y, Z não são ótimos ocorrerá:

I. $\exists i \notin R$ com $w_i \neq 0$

II. $\forall i \in R$ com $w_i = 0$, mas $W \geq 0$ não ocorre. Se nenhum dos casos ocorre, então temos a solução. Se I. ocorre, $\exists r \notin R$ tal que $w_r \neq 0$ enquanto b_r é uma coluna livre. No caso II, $\exists r \in R$ tal que $w_r < 0$ com b_r restrita. Em cada caso b_r é a coluna selecionada a sair da base, se houver mais de uma podemos escolher

qualquer uma.

2. Determinação de \underline{b}_r

Dada a direção viável b^r com decréscimo da função $f(Y)$, o novo ponto viável será:

$$\underline{Y} = Y - \sigma b^r$$

como σ minimizando $f(Y)$ na direção b^r e mantendo a viabilidade $Y \geq 0$. A determinação de σ definirá \underline{b}_r , como é evidenciado a seguir,

$$f(Y - \sigma b^r) = f(Y) + \frac{1}{2} \sigma^2 (b^r)^T \cdot Q \cdot b^r - \sigma (d^T \cdot b^r + Y^T Q b^r)$$

$$\text{mas, } (d^T \cdot b^r + Y^T Q b^r) = (d + Y^T Q) b^r = W_r$$

$$f(Y - \sigma b^r) = f(Y) + \frac{1}{2} \sigma^2 (b^r)^T \cdot Q \cdot b^r - \sigma W_r$$

$$\text{se } q = Q b^r \neq 0, \text{ seja } \sigma_0 = \frac{W_r}{q^T \cdot b^r}$$

$$f(Y - \sigma b^r) = f(Y) - \frac{1}{2} q^T b^r [\sigma_0^2 - (\sigma_0 - \sigma)^2]$$

$$\text{então, } \sigma = \sigma_0 = \frac{W_r}{q^T \cdot b^r} \text{ é a solução}$$

Para manter a viabilidade de $Y - \sigma b^r$, temos que:

$$Y - t \sigma_0 b^r \geq 0, \quad e_i^T \cdot Y - t \cdot \sigma_0 \cdot e_i^T \cdot b^r \geq 0 \quad i = \overline{1, n}$$

$$\text{para } \sigma_0 \cdot l_i^T \cdot b^r > 0, \quad t \leq \frac{e_i^T \cdot Y}{\sigma_0 \cdot e_i^T \cdot b^r}$$

$$\text{para } \sigma_0 \cdot l_i^T \cdot b^r \leq 0, \quad t \geq 0$$

$$\text{então seja } t_0 = \text{Min}_i \left\{ \frac{e_i^T \cdot Y}{\sigma_0 \cdot e_i^T \cdot b^r} \mid \sigma_0 \cdot e_i^T \cdot b^r > 0 \right\}$$

Finalmente,

$$\sigma = \sigma_0 \quad \text{se } t_0 > 1$$

$$\sigma = \sigma_0 \cdot t_0 \quad \text{se } t_0 \leq 1$$

Quando $\sigma = \sigma_0 \cdot t_0$, existirá índice K com $\sigma = \frac{e_k^T \cdot Y}{e_k^T b^r}$.

Por outro lado, se $q = 0$ teremos

$$f(Y - \sigma b^r) = f(Y) - \sigma W_r$$

Portanto, para todo σ com $\sigma \cdot W_r > 0$ a função $f(Y)$ de crescerá naquela direção. Da hipótese de ótimo finito, o valor absoluto de σ não poderá se tornar infinito. Assim a determinação de σ será definida em sinal por $\sigma \cdot W_r > 0$ e em valor absoluto pela viabilidade de $Y - \sigma b^r$. Resumindo,

$$\sigma = \text{Min}_i \left\{ \frac{e_i^T Y}{e_i^T b^r} \mid W_r l_i^T b^r > 0 \right\}; \quad \sigma = t_0$$

Então o valor de σ definirá quatro casos:

$$\text{Caso Ia ou IIa} : \sigma = \sigma_0 \implies \underline{b}^r = Q; b^r = q$$

$$\text{Caso Ib ou IIb} : \sigma = \frac{e_k^T Y}{e_k^T b^r} \implies \underline{b}^r = e_R$$

3. Modificações dos dados

\underline{B}^{-1} é definida pelas relações

$$\underline{b}^r = \frac{1}{\underline{b}^r \cdot b^r} b^r$$

$$\underline{b}^i = b^i - (\underline{b}^r \cdot b^i) b^r \quad (i \neq r)$$

$$\underline{Y} = Y - \sigma b^r$$

$$\underline{Z} = Z - \sigma q$$

\underline{R} é obtido de R por

i. casos Ia ou IIb, inalterado.

ii. caso Ib, incluir r.

iii. caso IIa, excluir r.

A₃ - A convergência.

Partindo da suposição que existe solução para o problema (A) e portanto para (4), este estará na fronteira ou no interior do quadrante positivo. Será ponto interior se e somente se existir $Y > 0$ tal que $d + QY = 0$. Se no processo iterativo $Z = 0$ e $R \neq 0$, então teremos uma solução interior. No algoritmo para caso $R \neq 0$ não ocorrendo caso I ou caso II é dado por:

$$\forall i \in R \quad W_i = 0 \quad W \geq 0$$

evidenciando as condições de Kuhn-Tucker para (4), pois que o quadrante $Z = \nabla f(Y)$ será dado por:

$$Z = \sum_{i \in R} W_i e_i \quad W_i \geq 0 \quad i \in R$$

Lema 1 - Quando caso I ocorrer, este poderá se repetir somente um número finito de vezes, após o que ocorrerá o caso II ou o ponto ótimo.

Dem.

Seja s o número de colunas livres e vamos supor que o caso I ocorra. Se ocorre Ib, então o número s passará a $s-1$ na próxima iteração. Quando ocorrer Ia o número s permanecerá fixo. Assim sendo, se mostrarmos que quando o caso Ia ocorre com s colunas livres ele somente poderá se repetir s iterações consecutivas, a recorrência sucessiva do caso I nos levará a situação de $s = 0$ ou seja $Y = 0$ o que é impossível, a menos que $d \geq 0$ que corresponderia a não ter dado início ao processo. Mostremos então, que Ia não pode se repetir consecutivamente mais que s vezes. Seja b_r o vetor que sai, então $\underline{b}^r = Q b^r$ entra, ocorrendo Ia verificaremos que $W_r = 0$ o qual permanecerá nulo enquanto Ia continuar ocorrer. Seja B a base e vamos supor que para algum i ; $b_i = K Q b^i$, como para $j \neq i$ $b_i^T b^j = 0$ $b_i^T \cdot Q \cdot b^i = 0$. Então tendo-se no caso Ia para algum $r \neq i$ $\underline{b}_r = Q b^r$ teremos a nova linha $\underline{b}_i = b^i - (b_r^T b^i) \underline{b}_r = b^i$. Assim o novo $\underline{Z} = Z - \sigma_0 Q b^r = \sum_{i=1}^n [W_i - \sigma_0 (b_i^T Q b^r)] b^r$ apresentará o coeficiente de

b_r nulo, pela definição de σ_0 .

Portanto, para i tal que se tenha $W_i = 0$ e $b_i = K Q b^i$, teremos os novos valores $\underline{W}_i = 0$ e $\underline{b}_i = b^i$, então os W correspondentes às colunas livres serão anulados e permanecerão nulos.

Lema 2 - Quando caso III ocorrer, acontecerá um decréscimo em $f(Y)$ dentro de um número finito de iterações ou ótimo será alcançado.

Dem.

Vamos supor que caso II ocorra na forma IIa, então teremos que $f(\underline{Y}) = f(Y) - \frac{1}{2} (b^r{}^T Q b^r) \sigma_0^2$, representará um decréscimo efetivo $f(\underline{Y}) < f(Y)$. Se na próxima iteração caso I ocorre, o Lema 1 mostra que ocorrerá o caso II dentro de um número finito de iterações ou ocorrerá o ótimo. Como para cada caso IIa $f(\underline{Y}) < f(Y)$, este somente ocorrerá um número finito de vezes ou ótimo é alcançado. Consideremos agora caso IIb, nesta situação é possível que $f(\underline{Y}) = f(Y)$. Mostraremos que o caso I sempre acontece em seguida de um caso IIb. Expressando

$$Z = \sum_{i=1}^n [W_i - (W_r / e_k^T \cdot b^r) e_k^T b^i] b_i + (W_r / e_k^T b^r) e_k \quad \text{em termos}$$

da nova base e especificando os novos coeficientes de \underline{W} . Como e_k não pode ser expresso somente em termos de colunas restritas, então algum $e_k^T b_i$ com $i \notin R$ é não nulo e portanto o correspondente coeficiente de b_i é não nulo, o que caracteriza o caso I. Por outro lado, se caso Ia ocorre então $f(\underline{Y}) < f(Y)$, e se caso Ib ocorrer pode acontecer $f(\underline{Y}) = f(Y)$ mas certamente nessa hipótese $\underline{s} = s-1$. Assim sendo, o único modo de não ocorrer decréscimo em $f(Y)$ em sucessivas iterações é acontecer somente caso Ib e caso IIb, com $\sigma = 0$ permanentemente. Entretanto, como caso Ib sempre segue caso IIb, e nesta situação s é diminuído de uma unidade a cada ocorrência de Ib, então a sucessão terminará ou no ótimo, ou numa situação de decréscimo efetivo (Caso Ia ou IIa), ou $s = 0$ que segundo Lema 1 é impossível. Finalmente quando caso IIb ocorrer teremos garantido ou o ótimo ou um decréscimo efetivo em $f(Y)$, que por sua vez será seguido do ótimo ou do caso II dentro de um número finito de iterações.

ANEXO BEXEMPLOSB₁ a Apresentação

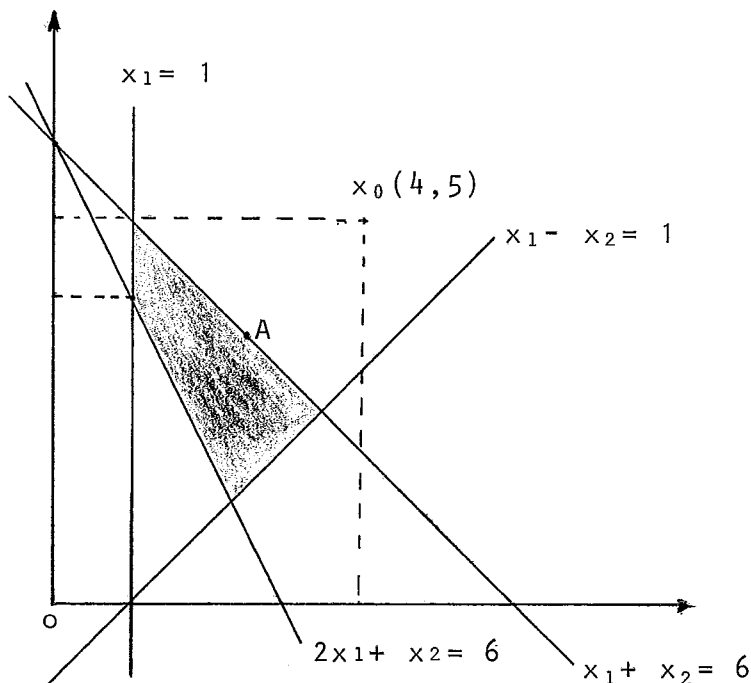
Com o propósito de ilustrar a aplicação do algoritmo proposto, utilizaremos um problema definido em R^3 , mas de fácil representação em R^2 . Definiremos o conjunto das soluções viáveis pelas restrições:

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ x_1 - x_2 &\leq 1 \\ -2x_1 - x_2 &\leq -6 \\ -x_1 &\leq -1 \\ -x_2 &\leq 0 \\ -x_3 &\leq 0 \end{aligned} \quad (I)$$

Rescrevendo a primeira restrição de igualdade,

$$x_1 + x_2 \leq 6, \text{ pois } x_3 \geq 0$$

teremos um conjunto viável em R^2 ($x_3 = 0$) representado pelo gráfico que se segue. O eixo \vec{x}_3 deve ser considerado perpendicular à figura.



Considerando-se as restrições (I), a matriz de projeção P , dada pela expressão:

$$P = I - L (L^T \cdot L)^{-1} L^T$$

terá a matriz L formada pelo único vetor da primeira restrição,

$$L = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{logo } P = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad \text{cujo escalar } \frac{1}{3} \text{ po}$$

de ser abandonado, já que estamos interessados tão somente nas direções.

Problema 1

Determinar um ponto satisfazendo K-T para o problema:

$$\text{Min } f(X) = (x_1 - 4)^2 + (x_2 - 5)^2 + x_3^2$$

X satisfazendo (I)

O ponto X^* satisfazendo K-T fornecerá na realidade um mínimo global de $f(X)$, por ser esta convexa num conjunto convexo. Este problema é equivalente a determinar no conjunto viável em \mathbb{R}^2 , ver gráfico, o ponto de menor distância a $X_0 = (4,5)$. É fácil ver que esse ponto é dado pelas coordenadas $X_A = (5/2, 7/2)$ e portanto $X^* = (5/2, 7/2, 0)$. Aplicaremos o algoritmo buscando o Max $g(X)$ com $g(X) = -f(X)$, o que nos dará :

$$\nabla g(x) = 2 \begin{bmatrix} 4 - x_1 \\ 5 - x_2 \\ -x_3 \end{bmatrix}$$

1. Seja $x_1 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ ponto viável de (1)

$$A(X_1) = \{3\}, L_3 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}, \nabla g_1 = \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix}, g_1 = -17$$

$$\bar{L}_3 = \begin{bmatrix} -3 \\ 0 \\ 3 \end{bmatrix}, \nabla g_1 = \begin{bmatrix} 3 \\ 6 \\ -9 \end{bmatrix}, \text{ como } \langle \bar{L}_3, \nabla g_1 \rangle < 0$$

$$\text{então } \beta_1 = \nabla g_1$$

O valor de t^* que maximiza $g(X + t \beta)$ é dado pela expressão geral

$$t^* = - \frac{\beta_1 (x_1 - 4) + \beta_2 (x_2 - 5) + \beta_3 x_3}{\beta_1^2 + \beta_2^2 + \beta_3^2}$$

$$t_1^* = \frac{1}{3} \text{ e } \mu_1^* = \frac{2}{9}$$

$$X_2 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} + \frac{2}{9} \begin{bmatrix} 3 \\ 6 \\ -9 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 8 \\ 10 \\ 0 \end{bmatrix}$$

$$2. A(X_2) = \{5\}, L_5 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \nabla g_2 = \frac{1}{3} \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}, g_2 = -\frac{41}{9}$$

$$\bar{L}_5 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad \bar{\nabla} g_2 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \quad \text{como } \langle \bar{L}_5, \bar{\nabla} g_2 \rangle > 0$$

a componente α_2 representa a projeção de $\bar{\nabla} g_2$ em \bar{L}_5

$$\alpha_2 = \frac{\langle \bar{L}_5, \bar{\nabla} g_2 \rangle}{|\bar{L}_5|^2} \cdot \bar{L}_5 = \begin{bmatrix} 3/2 \\ 3/2 \\ -3 \end{bmatrix}$$

$$\beta_2 = \bar{\nabla} g_2 - \alpha_2 = \begin{bmatrix} -1/2 \\ 1/2 \\ 0 \end{bmatrix}, \quad t_2^* = \frac{1}{3}, \quad \mu_2^* = \frac{10}{3}$$

$$X_3 = \frac{1}{3} \begin{bmatrix} 8 \\ 10 \\ 0 \end{bmatrix} + \frac{1}{5} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 \\ 7 \\ 0 \end{bmatrix}$$

$$3. \quad A(X_3) = \{5\}, \quad L_5 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \nabla g_3 = \frac{1}{2} \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix}, \quad g_3 = -\frac{9}{2}$$

$$\bar{L}_5 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad \bar{\nabla} g_3 = \begin{bmatrix} 3/2 \\ 3/2 \\ -3 \end{bmatrix}, \quad \text{como } \bar{\nabla} g_3 = \frac{3}{2} \bar{L}_5$$

temos que X_3 é o K-T ponto, como havíamos dito.

Problema 2

Determinar um ponto satisfazendo K-T, para o problema:

$$\text{Min } f(X) = (x_1 - 4)^2 + (x_2 - 5)^2 - 20x_3^2$$

X satisfazendo (I)

Não sendo $f(X)$ convexo o ponto X^* satisfazendo as condições de K-T não garante o mínimo global. Tal fato será confirmado pelo algoritmo que fornecerá diferentes soluções quando partirmos de diferentes pontos. Fazendo $g(X) = -f(X)$, teremos:

$$\nabla g(X) = 2 \begin{bmatrix} 4 - x_1 \\ 5 - x_2 \\ 20x_3 \end{bmatrix}$$

$$1. \quad X_1 = \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix}, \quad A(X_1) = \{4, 5\}, \quad L_4 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \quad L_5 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix},$$

$$\nabla g_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, \quad g_1 = -9$$

$$\bar{L}_4 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \bar{L}_5 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad \bar{\nabla} g_1 = \begin{bmatrix} 6 \\ -3 \\ -3 \end{bmatrix}$$

$$\text{Min } Q(Y) = Y^T \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix} Y - 2 \begin{bmatrix} -18 & 9 \end{bmatrix} Y \quad Y \geq 0$$

que resolvido pelo algoritmo de LEMKE, dará:

$$Y^* = \begin{bmatrix} 0 \\ 3/2 \end{bmatrix}, \text{ logo } \alpha_1 = \frac{3}{2} \quad \bar{L}_5 = \begin{bmatrix} 3/2 \\ 3/2 \\ -3 \end{bmatrix}$$

$$\beta_1 = \bar{\nabla} g_1 - \alpha_1 = \frac{9}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

O valor de t^* será dado pela expressão geral:

$$t^* = - \frac{\beta_1 (x_1 - 4) + \beta_2 (x_2 - 5) - 20 \beta_3 x_3}{\beta_1^2 + \beta_2^2 - 20 \beta_3^2}$$

$$\mu_1^* > t_1^* = \frac{1}{3}$$

$$X_2 = \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 \\ 7 \\ 0 \end{bmatrix}$$

$$2. \quad A(X_2) = \{5\}, \quad L_5 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \nabla g_2 = \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad g_2 = -\frac{9}{2}$$

$$\bar{L}_5 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad \nabla g_2 = \begin{bmatrix} 3/2 \\ 3/2 \\ -3 \end{bmatrix}; \text{ como } \bar{\nabla} g_2 = \frac{3}{2} \bar{L}_5$$

temos que X_2 é um K-T ponto e representa um mínimo local para $f(X)$, que será evidenciado pela aplicação do algoritmo partindo agora do ponto inicial

$$X_1 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$$

$$1. \quad A(X_1) = \{3, 4\}, \quad L_3 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}, \quad L_4 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \quad \nabla g_1 = \begin{bmatrix} 3 \\ 1 \\ 20 \end{bmatrix},$$

$$g_1 = -10$$

$$\bar{L}_3 = \begin{bmatrix} -3 \\ 0 \\ 3 \end{bmatrix}, \quad L_4 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \bar{\nabla} g_1 = \begin{bmatrix} -15 \\ -21 \\ 36 \end{bmatrix}$$

$$\text{Min } Q(Y) = Y^T \begin{bmatrix} 18 & 9 \\ 9 & 6 \end{bmatrix} Y - 2 [153, 45] Y, \quad Y \geq 0$$

que resolvido pelo algoritmo de LEMKE, darã

$$Y^* = \frac{1}{2} \begin{bmatrix} 17 \\ 0 \end{bmatrix}, \quad \text{logo } \alpha_1 = \frac{17}{2}, \quad \bar{L}_3 = \frac{1}{2} \begin{bmatrix} -51 \\ 0 \\ 51 \end{bmatrix}$$

$$\beta_1 = \bar{\nabla} g_1 - \alpha_1 = \frac{1}{2} \begin{bmatrix} 21 \\ -42 \\ 21 \end{bmatrix}$$

$$\text{como } t_1^* > \mu_1^* = \frac{8}{63}$$

$$X_2 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} + \frac{4}{83} \begin{bmatrix} 21 \\ -42 \\ 21 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 7 \\ 4 \\ 7 \end{bmatrix}$$

$$2. \quad A(X_2) = \{2,3\}, \quad L_2 = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}, \quad L_3 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}, \quad g_2 = -\frac{834}{9}$$

$$\nabla g_2 = \frac{1}{3} \begin{bmatrix} 5 \\ 11 \\ 140 \end{bmatrix}, \quad \bar{L}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad \bar{L}_3 = \begin{bmatrix} -3 \\ 0 \\ 3 \end{bmatrix}, \quad \bar{\nabla} g_2 = \begin{bmatrix} -141 \\ -123 \\ 264 \end{bmatrix}$$

como, $\bar{\nabla} g_2 = 123 \bar{L}_3 + 6 \bar{L}_2$ temos que X_2 é um K-T ponto, correspondendo neste caso ao mínimo global.