SOBRE UM MÉTODO DE DECOMPOSIÇÃO BASEADO EM PENALIZAÇÃO LAGRANGEANA

Fernando Hernán Paredes Cajas

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

Aprovada por:

Paulo Roberto Oliveira, Dr.Ing.
(Presidente)

Nelson Maculan Filho, D.Sc.

Susana Scheimberg de Makler, D.Sc.

Sérgio Granville, Ph.D.

Luis Contesse Becker, Dr.Ing.

RIO DE JANEIRO, RJ - BRASIL
JULHO DE 1991

PAREDES CAJAS, FERNANDO HERNÁN

Sobre um Método de Decomposição baseado em Penalização Lagrangeana [Rio de Janeiro] 1991.

VIII, 129 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 1991)

Tese-Universidade Federal do Rio de Janeiro, COPPE.

1. Decomposição em Programação Não-linear, Penalização Lagrangeana, Decomposição por meio de Penalização Lagrangeana.

I. COPPE/UFRJ II. Título(série).

A Silvia y a
nuestros hijos Gonzalo y
Carolina.

AGRADECIMIENTOS

Agradezco al profesor Paulo Roberto Oliveira, orientador de esta Tesis, por su predisposición positiva y sus valiosas observaciones, que contribuyeron a hacer realidad la versión definitiva de este trabajo.

También agradezco al profesor Luis Contesse B., co-orientador de esta Tesis, cuyo aporte y colaboración fué fundamental en la gestación y desarrollo de ésta nueva alternativa algorítmica de descomposición.

Agradezco al profesor Van Hién Nguyen por las sugerencias presentadas.

Agradezco a la Universidad Católica de Valparaíso, especialmente al Instituto de Matemáticas, por el patrocinio respectivo para realizar estos estudios.

Agradezco también al Programa de Ingeniería de Sistemas de la COPPE/UFRJ, por la acogida que me brindó durante mis estudios de Post-Grado y a CNPQ por la ayuda financiera.

Agradezco al Departamento de Ingeniería de Sistemas de la Pontificia Universidad Católica de Chile por las facilidades que me otorgara para desarrollar esta Tesis.

Agradezco a Jorge Villavicencio por su colaboración en el uso del código PENAMOR, así como también por proporcionarme detalles de su experiencia computacional con el problema de Gran Tamaño.

Agradezco a mi amigo Andrés Carrillo por su gran solidaridad y la edición computacional de gran parte de este trabajo.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências(D.Sc.)

SOBRE UM MÉTODO DE DECOMPOSIÇÃO BASEADO EM PENALIZAÇÃO LAGRANGEANA

Fernando Hernán Paredes Cajas

Julho de 1991

Orientador: *Paulo Roberto Oliveira.*
Co-Orientador: *Luis Contesse Becker.*
Programa: *Engenharia de Sistemas e Computação.*

RESUMO

Neste trabalho, desenvolvemos um Método de Decomposição para problemas de Programação Não-Linear de Grande Porte, baseado em uma certa decomposição das variáveis primais da função de Penalização Lagrangeana Aumentada. Começamos com uma revisão de algumas abordagens de decomposição que têm sido usadas até agora, e um breve estudo da convergência do Método de Penalização Lagrangeana. Nos restantes Capítulos, estudamos o Método proposto e apresentamos os resultados obtidos com sua Implementação Computacional sobre alguns problemas de prova conhecidos na literatura, e sobre um problema real de Despacho Econômico de Carga Elétrica.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

A PENALTY LAGRANGIAN DECOMPOSITION METHOD

Fernando Hernán Paredes Cajas

JULY 1991

Thesis Supervisor: *Paulo Roberto Oliveira*
Thesis Co-Supervisor: *Luis Contesse Becker*
Departament: *Systems Engineering and Computer Science*

ABSTRACT

In this thesis, we propose a decomposition algorithm for the solution of large scale nonlinear optimization problems, based on some primal decomposition of the Lagrangian Penalty Method. The work begins with a survey of several decomposition algorithms existing in the literature and a summary of the main convergence results of this last method. The work follows with a description of the proposed algorithm and the proof of its convergence. Finally, the application of this algorithm to some well known test problems, as well as to a real world Electrical Economic Dispatching Problem, is presented.

ÍNDICE

CAPÍTULO I

INTRODUÇÃO

A partir do trabalho de DANTZIG e WOLFE(1960), começaram a desenvolver-se as primeiras técnicas de decomposição em Programação Não Linear de Grande Porte. O método de Dantzig-Wolfe, como é sabido, foi desenvolvido para a decomposição de problemas de programação linear com coeficientes matriciais tipo angular. Neste método, o problema original é decomposto em vários subproblemas lineares menores e um problema mestre (coordenador). Em cada iteração, os subproblemas recebem um conjunto de parâmetros (preços simplex) do problema mestre. Os subproblemas enviam suas soluções correntes ao problema mestre o qual por sua vez volta a enviar um novo conjunto de preços aos subproblemas, e assim sucessivamente até obter a solução do problema original.

Os passos básicos seguidos pelos métodos de decomposição em programação não linear de Grande Porte são muitos parecidos com o método de Dantzig-Wolfe. Estes algoritmos têm geralmente estruturas de dois níveis. A resolução do problema original se reduz a resolver certo número de subproblemas de menor dimensão. No nível inferior, tais subproblemas são resolvidos em forma independente. No nível superior, as soluções dos subproblemas são coordenadas para obter a solução do problema original. Nas últimas décadas, os métodos de decomposição têm tido aplicação, por sua natureza, em projeto ótimo de Engenharia, como pode-se ver em SHAPOUR e LI(1986), por exemplo.

Desde a publicação do método de Dantzig-Wolfe têm-se desenvolvido muitos trabalhos de decomposição em Programação Não Linear, destacando-se no início os trabalhos de GEOFFRION(1970a,1970b,1972a,1972b), LASDON(1970), MESAROVIC et al.(1970), WISMER(1971), HIMMELBLAU(1973), e as referências contidas nestes.

A maioria das estratégias de decomposição estudadas até agora em Programação Não-linear supõem entre suas hipóteses convexidade do problema original, assim como eventualmente algum tipo de estrutura decomponível. Com o objetivo de obter uma estratégia de decomposição para problemas de Grande Porte, em geral não convexos, nesta tese se propõe um Método de Decomposição baseado em Penalização Lagrangeana e que não precisa necessariamente de que o problema original tenha estrutura decomponível, assim como também seja convexo. O método consiste em obter um ponto de sela da Lagrangeana Aumentada associada a um certo problema equivalente ao original. Para isto, aplica-se o Método de Penalizaçao Lagrangeana ou de Multiplicadores de Hestenes-Powell-Rockafellar, com decomposição primal da Lagrangeana Aumentada, utilizando os resultados de diferenciabilidade marginal obtidos em CONTESSE(1986b)(no contexto mais geral da otimização não diferenciável existem resultados análogos, como os assinalados em HIRIART-URRUTY(1978, pp. 304-305)). A convergência do método proposto resulta então essencialmente da conjuncão das convergências do Método de Multiplicadores, do Método de Gradiente Projetado com passo econômico utilizado no nível superior da decomposição e, finalmente, do Método de Broyden-Fletcher-Goldfarb-Shano(B.F.G.S.) para a minimização sob restrições lineares, utilizado na resolução dos problemas do nível inferior da decomposição.

Neste trabalho se estableceu completamente a convergência do método de decomposição proposto, utilizando fundamentalmente resultados dos trabalhos de CONTESSE(1986a,1986b,1987b,1990b) e CONTESSE e VILLAVICENCIO(1988). Também se fez uma implementação computacional do método resolvendo alguns problemas testes da literatura, obtendo resultados muitos satisfatórios. Além disso, um problema real de Grande Porte com 310 variáveis sobre um Modelo de Distribução de Energia Elétrica foi testado. Este problema comporta 180 variáveis, 38 restrições lineares, 20 restrições fortemente não lineares e cotas sobre todas as variáveis. Os resultados obtidos até agora foram comparados com os reportados em CONTESSE e VILLAVICENCIO(1990), o que prova que o método proposto é bastante promissor. O algoritmo, assim como alguns dos resultados numéricos iniciais foram apresentados pelo autor em CONTESSE, OLIVEIRA e PAREDES(1990).

A tese foi organizada na seguinte maneira:

No Capítulo II fazemos uma revisão das técnicas de decomposição que têm sido utilizadas em Programação Não-linear nas últimas décadas, fundamentalmente as baseadas em Paramétrizacão e Dualidade. Uma grande parte deste Capítulo foi apresentada pelo autor em OLIVEIRA e PAREDES(1988).

No Capítulo III se revisa brevemente a convergência do Método de Penalização Lagrangeana de HESTENES-POWELL-ROCKAFELLAR(1971,1973,1974), seguindo muito de perto o trabalho de CONTESSE(1987b).

No Capítulo IV se apresenta o método de decomposição proposto com alguns resultados teóricos associados e a relação existente com outros métodos.

No Capítulo V se estabelece a convergência completa do método de decomposição.

No Capítulo VI se apresentam os aspectos principais da implementação computacional, se descrevem em detalhe os distintos problemas de prova e os resultados numéricos obtidos.

No Capítulo VII se apresentam as conclusões e os comentários finais.

No final, depois das referências bibliográficas, inclue-se em um Apêndice, os resultados do trabalho de CONTESSE(1986b), assinalado acima.

CAPÍTULO II

SOBRE ALGUMAS METODOLOGIAS DE DECOMPOSIÇÃO EM PROGRAMAÇÃO NÃO-LINEAR

Neste Capítulo revisam-se algumas técnicas utilizadas na resolução de problemas de Programação Não-Linear (PNL) de grande porte, por meio da decomposição destes em subproblemas menores, cujas soluções possam ser obtidas mais facilmente. Estas, de uma certa forma prefixada, permitem a obtenção da solução do problema original. Uma grande parte deste Capítulo foi apresentada pelo autor em OLIVEIRA e PAREDES(1988).

Em Programação Não-Linear, o porte do problema é determinado (de maneira imprecisa) pela quantidade de variáveis, o número e complexidade das restrições, e a complexidade da função objetivo, LASDON (1970).

Uma vasta literatura tem-se desenvolvido no tema, destacando-se no início os trabalhos de GEOFFRION(1970a,1970b,1972a,1972b), LASDON(1970), MESAROVIC et al.(1970), WISMER(1971), HIMMELBLAU(1973), e as referências contidas nestes, entre outros.

II.1.-DECOMPOSIÇÃO PARAMÉTRICA.-

No método de decomposição paramétrica, ERMOLEV e ERMOLEVA(1973), assim como no método de decomposição primal por consignação de recursos, GEOFFRION(1970c), SILVERMAN(1972a,1972b), fixam-se certas variáveis do problema que são consideradas como parâmetros, de tal maneira que as estruturas obtidas sejam mais simples de tratar. Nesta seção veremos três casos: convexo, não-convexo, e diferenciável.

II.1.1.-Caso Convexo.

Seja o problema de programação convexa

$$\text{Mín } f_o(x,y) \tag{II.1}$$
s.a.
$$f_i(x,y)\le 0 \ , \ i=1,\ldots\ldots,m, \tag{II.2}$$

onde $x\in\mathbb{R}^n$, $y\in\mathbb{R}^p$, e as f_i, $i=0,\ldots\ldots,m$,são funções convexas. Fixemos x e consideremos o seguinte problema:

$$\text{Mín } f_o(\bar{x},y) \tag{II.3}$$
s.a.
$$f_i(\bar{x},y)\le 0 \ , \ i=1,\ldots\ldots,m \tag{II.4}$$

O problema (II.3)-(II.4) é convexo. Nos valores de $\bar{x}$ em que ele tem solução, se define a função ϕ:

$$\phi(\bar{x}) := \underset{y\in D(\bar{x})}{\text{Mín}}\ f_o(\bar{x},y), \tag{II.5}$$

onde, $D(\bar{x}) := \left\{ y\in\mathbb{R}^p / \ y \text{ satisfaz (II.4)} \right\}$

Teorema 1:(SHOR,(1985))

Se as funções f_i, $i=0,\ldots\ldots,m$, são convexas, então a função ϕ definida por (II.5) é convexa em algum subconjunto convexo W de $\mathbb{R}^n$. Se para algum $\bar{x}\in W$ se cumpre a condição de qualificação de SLATER para (II.4), então um subgradiente de ϕ em $\bar{x}$ pode-se calcular pela fórmula:

$$g_\phi(\bar{x}) = g^x_{L_U}(\bar{x},y(\bar{x})), \tag{II.6}$$

onde $L_U(x,y)=f_o(x,y)+\sum_{i=1}^{m} U_i f_i(x,y)$ é a Lagrangeana do problema (II.1)-(II.2), $y(\bar{x})$ é um dos valores ótimos de y no problema (II.3)-(II.4), $U=\{U_i\}_{i=1}^{m}$ os correspondentes multiplicadores de Lagrange (obtidos do teorema de Kuhn-Tucker) e $g^x_{L_U}(\bar{x},y(\bar{x}))$ é a projeção no subespaço $\mathbb{R}^n$ de um subgradiente da função $L_U(x,y)$, cuja projeção no subespaço $\mathbb{R}^p$ é o vetor 0(o subgradiente é considerado no ponto $(\bar{x},y(\bar{x}))$.

Corolário1:(Shor,(1985))

Se no teorema 1 supomos também que as $f_i(\cdot,\cdot)$, $i=0,\ldots\ldots,m$, são continuamente diferenciáveis em relação à variável y, então a fórmula (II.6) fica:

$$g_\phi(\bar{x})=g^x_{f_0}(\bar{x},y(\bar{x}))+\sum_{i=1}^{m}U_i(\bar{x})g^x_{f_i}(\bar{x},y(\bar{x})), \qquad \text{(II.7)}$$

onde $g^x_{f_i}(\bar{x},y(\bar{x}))$ representa a projeção no subespaço $\mathbb{R}^n$ de um subgradiente de f_i no ponto $\bar{z}=(\bar{x},y(\bar{x}))$ em $\mathbb{R}^n$, $i=0,1,\ldots m$.

Não se perde generalidade se supusermos que o problema (II.3)-(II.4) tem sempre solução. Daí é possível calcular um subgradiente $g_\phi(\bar{x})$ em qualquer $\bar{x}$ e consequentemente usar um algoritmo de subgradiente para minimizar ϕ. Assim, temos um algoritmo para obter a solução do problema (II.1)-(II.2), cuja k-ésima iteração define-se pelos seguintes passos:

a)Para $\bar{x}=\bar{x}_k$, obter a solução do problema (II.3)-(II.4) e calcular os multiplicadores de Kuhn-Tucker $U(\bar{x}_k)=\{U_i(\bar{x}_k)\}_{i=1}^{m}$;

b)Calcular o vetor $g_\phi(\bar{x}_k)$ por (II.6) ou (II.7), segundo seja o caso;

c)Define-se $\bar{x}_{k+1}=\bar{x}_k-h_{k+1}g_\phi(\bar{x}_k)$, com o passo h_k escolhido adequadamente.

O método de decomposição exposto anteriormente pode-se usar para obter a solução de vários problemas de otimização, em particular para a análise de grandes redes de transporte, SHOR(1985).

II.1.2. <u>Caso não convexo</u>.

Considera-se o seguinte problema (P.N.L.):

$$\begin{aligned}
&\text{Min } f(x,y,z)\\
&\text{s.a.}\\
&g_i(x,y)\leq 0,\quad i=1,\ldots\ldots,m\\
&h_j(x,z)=0,\quad j=1,\ldots\ldots,p\\
&l_k(y,z)\leq 0,\quad k=1,\ldots\ldots,r
\end{aligned}\tag{II.8}$$

onde $x\in\mathbb{R}^n, y\in\mathbb{R}^s, z\in\mathbb{R}^t$. Suponhamos que a função objetivo f e as restrições g_i, h_j sejam decomponíveis em seus argumentos x, y, z e que o problema de otimização seja fácil de resolver quando y e z forem mantidos fixos. Assim, sejam y, z fixos e consideremos o seguinte problema em x, o qual é resolvido por algum método padrão de (PNL),

(Primeiro nível da otimização):

$$\begin{aligned}
&\text{Min } f(x,y,z)\\
&\text{s.a.}\\
&g_i(x,y)\leq 0,\quad i=1,\ldots,m\\
&h_j(x,z)=0,\quad j=1,\ldots,p
\end{aligned}\tag{II.9}$$

Definen-se os seguintes conjuntos:

$$S(y):=\left\{x\in\mathbb{R}^n/\ g_i(x,y)\leq 0,\ i=1,\ldots\ldots,m\right\},$$

$$T(z):=\left\{x\in\mathbb{R}^n/\ h_j(x,z)=0,\ j=1,\ldots\ldots,p\right\},$$

então $U(y,z):=S(y)\cap T(z)$ é o conjunto viàvel para o problema (II.9).

O valor ótimo de (II.9) define-se por:

$$v(y,z):=\begin{cases} \min f(x,y,z) & \text{, se } U(y,z)\neq\emptyset, \\ \text{s.a.} & \\ x\in U(y,z) & \\ & \\ +\infty & \text{, em outro caso,} \end{cases} \qquad \text{(II.10)}$$

que é também chamada função valor ótimo.

Definamos os conjuntos:

$P(y,z):=\left\{\bar{x}\in U(y,z) / f(\bar{x},y,z)=v(y,z)\right\}$, o conjunto das soluções ótimas para (II.9),

$V=\left\{(y,z) / U(y,z)\neq\emptyset\right\}$, o conjunto de parâmetros viáveis para (II.9).

No segundo nível, uma pós-otimização é feita no conjunto de parâmetros viáveis:

$$\begin{aligned} &\text{Mín } v(y,z) \\ &\text{s.a.} \\ &(y,z)\in V \\ &l_k(y,z)\leq 0,\ k=1,\ldots\ldots,r. \end{aligned} \qquad \text{(II.11)}$$

Deste modo é possível utilizar condições de otimalidade, GAUVIN(1977), GAUVIN E TOLLE(1977), CLARKE(1983), para o problema de pós-otimização (II.11), e propor algoritmos para sua resolução, KIWIEL(1985).

A estrutura apresentada e a metodologia esboçada são generalizações do caso anterior (convexo), conforme pode ser facilmente verificado.

II.1.3. Um caso diferenciável.

Seja o problema

$$\mathrm{Min} \sum_{i=1}^{N} f_i(x_i,y)+w(y)$$

s.a.

$$
\begin{array}{llll}
g_i^{(1)}(x_1,y) & & \geq 0, i\in I^{(1)}, & \\
& g_i^{(2)}(x_2,y) & \geq 0, i\in I^{(2)}, & \text{(II.12)} \\
& \ddots & & \\
& & g_i^{(N)}(x_N,y)\geq 0, i\in I^{(N)}, & \\
y\in S, & & &
\end{array}
$$

onde $x_1,x_2,\ldots\ldots,x_N$ e y representam vetores, não necessariamente da mesma dimensão. As funções envolvidas são supostas duas vezes continuamente diferenciáveis e os conjuntos $I^{(1)},\ldots\ldots,I^{(N)}$ disjuntos. Observando-se a estrutura deste problema, é imediato que se o valor ótimo de y é conhecido, os valores ótimos das demais variáveis são os resultados de N otimizações independentes.

Assim o problema mestre é:

$$\text{Min} \sum_{k=1}^{N} v_k(y) + w(y)$$

s.a.

$y \in V^{(k)}$, $k=1,\ldots\ldots,N$

$y \in S$,

onde,

$$V^{(k)} := \left\{ y / v_k(y) := \min f_k(x_k, y) / g_i^k(x_k, y) \geq 0, \; i \in I^{(k)} \text{ existe e é finito} \right\}.$$

Em algumas aplicações escolhe-se S suficientemente pequeno de tal maneira que seja um subconjunto de $V^{(k)}$.

O k-ésimo subproblema é a determinação da k-ésima função valor ótimo, onde, por simplicidade de notação, omitimos o índice k, k=1,.......,N, de v, x , g_i e f

$$\begin{aligned} v(y) &:= \text{Min } f(x,y) \\ &\quad \text{s.a.} \\ &\quad g_i(x,y) \geq 0, \; i \in I \end{aligned} \qquad \text{(II.13)}$$

Seu domínio será representado por V.

A função (II.13) chamada também função de perturbação, tem sido extensivamente estudada por BANK et al.(1983), FIACCO(1984), GUIGNARD(1982), SORENSEN e WETS (1982).

Suponhamos que o problema (II.13) tem uma solução ótima $y=\bar{y}$. Sob certa condição de qualificação de restrições, existem multiplicadores não negativos $u_i(\bar{y})$, $i \in I$, de tal maneira que o par $(x(\bar{y}), u(\bar{y}))$ resolve as condições de Kuhn-Tucker:

$$\nabla_x f(x,y) - \sum_{i\in I} u_i \nabla_x g_i(x,y) = 0$$

$$u_i g_i(x,y) = 0, \quad i\in I, \tag{II.14}$$

para $y=\bar{y}$. Sob as chamadas "condições de unicidade do Jacobiano", LOOTSMA(1978), tem-se que a matriz Jacobiana do sistema (II.14) é não singular em $(x(\bar{y}),u(\bar{y}))$, e existe uma única solução $(x(y),u(y))$ de (II.14), diferenciável em uma certa vizinhança de $\bar{y}$. Uma destas condições é a estrita complementariedade de $x(\bar{y})$, outra é a independência linear dos gradientes $\nabla_x g_i(x(\bar{y}),\bar{y})$ correspondentes às restrições ativas em $x(\bar{y})$. Assim, tem-se que a diferenciabilidade de $(x(y),u(y))$ não é garantida se o conjunto ativo muda em $x(\bar{y})$. De fato não é difícil encontrar exemplos que demonstram que as descontinuidades nas derivadas de $(x(y),u(y))$ em $\bar{y}$ (sensibilidade do ótimo) ocorrem sob estas condições. Porém, se as condições de unicidade do Jacobiano são satisfeitas, então temos:

$$\frac{\partial\,[v(y)]}{\partial y_j} = \frac{\partial\,[f(x(y),y)]}{\partial y_j}$$

$$= \frac{\partial f(x(y),y)}{\partial y_j} + \sum_{l} \frac{\partial f(x(y),y)}{\partial x_l}\frac{\partial x_l(y)}{\partial y_j}$$

$$= \frac{\partial f(x(y),y)}{\partial y_j} + \sum_{i\in I_y}\sum_{l} u_i(y)\,\frac{\partial g_i(x(y),y)}{\partial x_l}\frac{\partial x_l(y)}{\partial y_j}$$

$$= \frac{\partial f(x(y),y)}{\partial y_j} + \sum_{i\in I_y} u_i(y)\,\frac{\partial g_i(x(y),y)}{\partial y_j} \quad ,$$

onde I_y é o conjunto de índices relativo às restrições que são ativas em $x(y)$. E' claro que o gradiente de v obtém-se facilmente sem necessitar da sensibilidade do ótimo $\frac{\partial x_i}{\partial y_j}$, ARORA e GOVIL(1977), SOBIESKI et al.(1982), SOBIESKI et al. (1985).

A questão de quando as referidas descontinuidades do gradiente de v alterariam seriamente o processo de decomposição tem sido estudado por Silva et al.(1986).

II.2 <u>Decomposição por Dualidade</u>.

II.2.1.-<u>Caso convexo geral</u>.

Seja o problema de programação convexa separável:

$$\text{Mín} \sum_{j=1}^{n} f_{oj}(x_j)$$

s.a.

$$\sum_{j=1}^{n} f_{ij}(x_j) \leq 0 \quad , \quad i=1,\ldots\ldots,m, \qquad \text{(II.15)}$$

$$x \in C$$

onde as funções f_{ij} , $i=o,\ldots\ldots\ldots,m$; $j=1,\ldots\ldots,n$, são convexas e definidas em $\mathbb{R}^{d_j}$ sobre $\mathbb{R}$, $x_j \in \mathbb{R}^{d_j}$, $j = 1,\ldots,n$, $\sum_{j=1}^{n} d_j = d$, $x=(x_1,x_2,\ldots\ldots\ldots,x_n)$, $C=\prod_{j=1}^{n} C_j \subset \mathbb{R}^d$ e os conjuntos C_j , $j=1,\ldots\ldots,n$, são convexos e fechados, subconjuntos de $\mathbb{R}^{d_j}$, respectivamente.

Seja a lagrangeana clássica associada ao problema (II.15):

$$L(x,y) = \sum_{j=1}^{n} [f_{oj}(x_j) + \sum_{i=1}^{m} y_i f_{ij}(x_j)] \quad ,x \in C, \ y \geq 0,$$

então o problema dual ao problema de (II.15) é:

$$\begin{array}{l} \text{Máx } g(y) \\ \text{s.a.} \\ y \geq 0, \end{array} \qquad \text{(II.16)}$$

onde $g(\cdot)$ é a função objetivo dual côncava:

$$g(y) = \inf_{x \in C} L(x,y)$$

$$= \sum_{j=1}^{n} \inf_{x_j \in C_j} \left[f_{oj}(x_j) + \sum_{i=1}^{m} y_i f_{ij}(x_j) \right]. \qquad \text{(II.17)}$$

Resolver o problema dual (II.16) se reduz a resolver em (II.17), separadamente, n subproblemas de minimização de dimensão bastante menor que a do problema original.

Sob certas condições de regularidade, ROCKAFELLAR(1970, teorema 28.2) não existe salto de Dualidade entre os problemas (II.16) e (II.17), isto é, o ínfimo em (II.15) é igual ao supremo em (II.16), e uma solução $\bar{y} \geq 0$ para (II.17) existe. Supondo que tal $\bar{y}$ existe, $\bar{x}$ resolve (II.16) se e somente se $(\bar{x},\bar{y})$ é ponto de sela para $L(\cdot,\cdot)$, ROCKAFELLAR(1970, teorema 28.1).

E' importante notar que esta aproximação clássica tem várias dificuldades potenciais. Para a maioria dos problemas não convexos, as soluções não correspondem a pontos de sela da Lagrangeana, de maneira que esta abordagem não se generaliza facilmente ao caso não convexo. Mais ainda, no caso convexo existem outras dificuldades. Para certos $y \geq 0$, $L(x,y)$ pode não alcançar seu mínimo em C. Além disto, $g(y)$ pode ter o valor $-\infty$. Por outra parte, é possível gerar uma sequência $\left\{ x^k \right\}$ que não convirja à solução de (II.15), ainda que aconteça que a sequência $\left\{ y^k \right\}$ associada seja uma sequência maximizante para (II.16). A maioria destes inconvenientes pode ser evitada, porém à custa de condições adicionais. Também se fica muito restrito na escolha de algum método de maximização para g. Cada cálculo de g(y) requer que a função L(x,y) seja minimizada em x sobre C. Ao mesmo tempo, qualquer método que precise muitas determinações de g será impraticável. Afortunadamente, a minimização em x de L(x,y) fornece sem custo extra um subgradiente para g em y, fato que motiva o método de Uzawa(1958), uma aproximação de máximo ascenso para a maximização de g. O algoritmo de decomposição de Dantzig-Wolfe (1960,1963) pode ser visto como a aproximação pela qual g é maximizada através de um método de plano de Corte, GEOFFRION(1971).

A estrutura particular de certos problemas pode induzir o plano da decomposição a utilizar. Por exemplo, GRANVILLE e SCHETMAN(1988) estudam o seguinte problema convexo:

$$\begin{aligned} &\text{Min } f_1(x)+f_2(y) \\ &\text{s.a.} \\ &h_1(x)+h_2(y)\le a \\ &g_1(x) \qquad\quad \le b_1 \\ &\qquad\qquad g_2(y)\le b_2, \end{aligned} \tag{II.18}$$

onde $x\in\mathbb{R}^{n_1}$, $y\in\mathbb{R}^{n_2}$

$h_i^T=(h_{i1},h_{i2},\ldots\ldots\ldots\ldots,h_{ik})$, $g_i^T=(g_{i1},g_{i2},\ldots\ ,g_{ik_i})$,

$b_i\in\mathbb{R}^{k_i}$, $i=1,2$, $a\in\mathbb{R}^k$.

Além disso, f_i , h_{ij} , g_{ij} são funções convexas próprias em seus respectivos domínios.

O primeiro passo do método de decomposição é escrever o problema (II.18) com a seguinte formulação equivalente:

$$
\begin{array}{lll}
\text{Min } f_1(x)+f_2(y) & & \\
\text{s.a.} & & \\
\qquad u_1+u_2=a & & \\
h_1(x)-u_1 \qquad \leq 0 & & \qquad (II.19)\\
g_1(x) \qquad \leq b_1 & & \\
\qquad h_2(y)-u_2 \leq 0 & & \\
\qquad g_2(y) \quad \leq b_2, & &
\end{array}
$$

onde $u_i \in \mathbb{R}^k$ $, i=1,2.$

Note que sem a primeira restrição, o problema (II.19) é inteiramente separável em relação às variáveis x e y.

Agora considera-se o seguinte problema associado:

$$
\text{Min}\left[f_1(x)+f_2(y)+\lambda^T(u_1+u_2-a)+1/2(u_1+u_2-a)^T D(u_1+u_2-a)\right]
$$

$$
\begin{array}{lll}
\text{s.a.} & & \\
h_1(x)-u_1 & \leq 0 & \\
g_1(x) & \leq b_1 & \qquad (II.20)\\
\qquad h_2(y)-u_2 & \leq 0 & \\
\qquad g_2(y) & \leq b_2 &
\end{array}
$$

onde $\lambda \in \mathbb{R}^k$ e D é uma matriz diagonal definida positiva em $\mathbb{R}^{k\times k}$.

A função:

$$
\alpha(x,y,\lambda)=f_1(x)+f_2(y)+\lambda^T(u_1+u_2-a)+1/2(u_1+u_2-a)^T D(u_1+u_2-a)
$$

é a função Lagrangeana Aumentada associada à primeira restrição do problema (II.19).

O segundo passo do método de decomposição é considerar os dois subproblemas seguintes:

$$\left[P_\lambda^{(1)}(u_2)\right] \quad \begin{array}{ll} \text{Min } f_1(x)+\lambda^T(u_1+u_2-a)+1/2(u_1+u_2-a)^T D(u_1+u_2-a) & \\ \text{s.a.} & \\ h_1(x)-u_1\leq 0 & \\ g_1(x) \quad \leq b_1 & \end{array}$$

$$\left[P_\lambda^{(2)}(u_1)\right] \quad \begin{array}{ll} \text{Min } f_2(y)+\lambda^T(u_1+u_2-a)+1/2(u_1+u_2-a)^T D(u_1+u_2-a) & \\ \text{s.a.} & \\ h_2(y)-u_2\leq 0 & \\ g_2(y) \quad \leq b_2 & \end{array}$$

Eles desenvolvem a teoria supondo que as restrições destes subproblemas são afins e aplicam-na na resolução de problemas de programação linear com matriz de restrições com estrutura "bloco-angular" e "staircase".

Um dos algoritmos de decomposição proposto é o seguinte(o outro é análogo):

<u>Algoritmo</u>:

Passo 0.- Seja $\lambda^{(0)} \in \mathbb{R}^k$ uma estimativa inicial para $\bar{\lambda}$, solução do problema dual.

Seja $u^{(0)}=(u_1^{(0)},u_2^{(0)})\in\mathbb{R}^k x\mathbb{R}^k$ uma estimativa inicial para $\bar{u}=(\bar{u}_1,\bar{u}_2)$ parte da solução ótima do problema primal (II.19).

Faça $i=0$

Passo 1.- a)Resolver $\left[P_\lambda^{(1)}(u_2)\right]$ com $\lambda=\lambda^{(i)}$ e $u_2=u_2^{(i)}$.

Seja $(x^{(i+1)},u_1^{(i+1)})$ sua solução ótima.

b)Resolver $\left[P_\lambda^{(2)}(u_1)\right]$ com $\lambda=\lambda^{(i)}$ e $u_1=u_1^{(i+1)}$.

Seja $(y^{(i+1)},u_2^{(i+1)})$ sua solução ótima

Passo 2. - Fazer,

$$\lambda^{i+1}=\lambda^{i}+D(u_1^{(i+1)}+u_2^{(i+1)}-a)$$

$i=i+1$

Ir a passo 1.

II.2.2. -<u>Sobre um problema convexo-linear</u>

Seja o problema de programação convexa separável:

$$\begin{aligned} &\text{Mín} \sum_{i=1}^{n} f_i(x_i) \\ &\text{s.a.} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (II.21) \\ &\sum_{i=1}^{n} A_i x_i = a, \end{aligned}$$

onde para cada i , f_i é uma função convexa própria fechada de $\mathbb{R}^{n_i}$ com valores em $(-\infty,\infty]$, $a \in \mathbb{R}^m$ e cada A_i é uma transformação linear de $\mathbb{R}^{n_i}$ em $\mathbb{R}^m$. Este problema modela um processo econômico no qual n atividades (subsistemas) trabalham para m recursos a distribuir. A quantidade dos recursos disponíveis é definida pelo vetor a. A i-ésima atividade é controlada por n_i parâmetros representados pelo vetor $x_i \in \mathbb{R}^{n_i}$; seus requerimentos dos recursos são $A_i x_i$ e seu custo é $f_i(x_i)$. Resolver (II.21) equivale a determinar como escolher os parâmetros $x_1,\ldots\ldots\ldots,x_n$ para minimizar o custo do sistema enquanto se consomem os recursos disponíveis determinados pelo vetor a.

Um caso particular encontrado em diversas aplicações é o seguinte problema de programação linear bloco-angular:

$$\text{Mín} \sum_{i=1}^{n} \langle c_i, x_i \rangle$$

s.a.

$$
\begin{array}{lllll}
B_1x_1 & & & & =b_1 \\
 & B_2x_2 & & & =b_2 \\
 & & \cdot & & \\
 & & & \cdot & \\
 & & & B_nx_n & =b_n \\
 & & & & \\
\multicolumn{5}{l}{A_1x_1+A_2x_2+\dots\dots\dots+A_nx_n=a} \\
\multicolumn{5}{l}{x_1,x_2,\dots\dots\dots\dots,x_n\geq 0}
\end{array}
\qquad (II.22)
$$

O modelo (II.22) pode ser posto na forma (II.21) ao definir:

$$f_i(x_i)=\begin{cases} \langle c_i,x_i\rangle, \text{se } B_ix_i=b_i,\ x_i\geq 0 \\ +\infty \quad , \text{em outro caso,} \end{cases}$$

para $i=1,2,\dots\dots\dots,n$.

O problema (II.22) ocorre em modelos de planejamento de multiestado como também em problemas da programação linear estocástica a dois estados.

II.2.2.1.-<u>Analise de Dualidade</u>

Ao introduzir um vetor $p\in\mathbb{R}^m$ de perturbações no problema (II.21) tem-se a função:

$$F(x_1,x_2,\dots\dots,x_n,p)=\begin{cases} \sum_{i=1}^{n} f_i(x_i), \text{ se } \sum_{i=1}^{n} A_ix_i=a+p \\ \\ +\infty \text{ , em outro caso} \end{cases} \qquad (II.23)$$

Aplicando os resultados da Teoria da Dualidade, ROCKAFELLAR(1970,1974a), obtém-se de (II.23) a função objetivo dual,

$$g(y)=\langle a,y\rangle-\sum_{i=1}^{n} f_i^*(A_i^T y),$$

onde $f_i^*(x_i^*) := \operatorname{Sup}_{x_i}\left\{ \langle x_i^*, x_i\rangle - f_i(x_i) \right\}$ é a função conjugada de f_i.

Sob a suposição de $a\in\sum_{i=1}^{n} A_i(\text{ri dom } f_i)$, onde $\text{ri}(\cdot)$ representa o interior relativo e dom f_i é o domínio efetivo de f_i, sabe-se que a função g alcança seu máximo igual ao ínfimo em (II.21), ROCKAFELLAR(1970). Agora, se além disto se tem a condição de regularidade:

existe algum y_o tal que para cada i,

$$A_i^T y_o \in \text{ri dom } f_i^*,$$

então o subdiferencial ∂g é dado por:

$$\partial g(y)=a-\sum_{i=1}^{n} A_i x_i(y),$$

onde, para cada i, o ponto $x_i(y)$ resolve o problema:

$$\begin{aligned} &\text{Min } \left\{ f_i(x_i) - \langle A_i^T y, x_i\rangle \right\} \\ &\text{s.a.} \\ &x_i \in \mathbb{R}^{n_i} \end{aligned} \qquad \text{(II.24)}$$

Note que (II.24) requer a minimização em apenas n_i variáveis. E' assim fácil obter um elemento do $\partial g(y)$, sendo possível aplicar algum dos métodos da otimização não-suave.

Veremos a seguir alguns aspectos da aplicação do método de Feixe (Bundle Method), proposto por LEMARECHAL et al.(1981). Este método gera uma sequência de pontos $y^1, y^2, \ldots\ldots, y^k$, e utiliza um conjunto de subgradientes $v^1, v^2, \ldots\ldots, v^k$ tais que $v^j \in \partial g(y^j)$. E' calculada a seguir uma combinação convexa de $v^1, v^2, \ldots\ldots, v^k$ que se espera se aproxime do vetor zero. Se o aproximante encontrado não está muito perto de zero, um novo ponto y^{k+1} é determinado e um dos subgradientes, seja v^{k+1}, é então calculado. O método termina com y^k quando, para certos valores de ε e δ positivos previamente dados, existe um vetor d^k tal que:

$$\|d^k\| < \delta \qquad \text{e} \ -d^k \in \partial_\varepsilon g(y^k).$$

Assim y^k é um ε-maximizador aproximado de g, já que o ε-subdiferencial é definido por:

$$\partial_\varepsilon g(\hat{y}) := \left\{ v / \ \forall y, g(y) \leq g(\hat{y}) + \langle v, y-\hat{y} \rangle + \varepsilon \right\}.$$

O vetor d^k é a projeção do conjunto $G_k := co(v_1, v_2, \ldots\ldots, v_k)$ sobre o origem:

$$d^k = -\sum_{j=1}^{k} \lambda_j^k \ v^j \ , \quad \sum_{j=1}^{k} \lambda_j^k = 1, \ \lambda_j^k \geq 0. \qquad \text{(II.25)}$$

A principal dificuldade em usar um método de otimização não-suave para resolver (II.21), maximizando g, está no fato de que ainda que conhecêssemos um maximizador exato $\bar{y}$ para g, poderíamos não ser capazes de encontrar uma solução ótima primal $\bar{x}_1, \bar{x}_2, \ldots\ldots, \bar{x}_n$. De fato, ainda que $0 \in \partial g(\bar{y})$, realmente só se póde calcular um elemento de $\partial g(\bar{y})$.

Além disto poderíamos encontrar $\hat{x}_1(\bar{y}), \ldots\ldots, \hat{x}_n(\bar{y})$ tais que $w := a - \sum_{i=1}^{n} A_i \hat{x}_i(\bar{y}) \in \partial g(\bar{y})$ seja mais afastado de zero. Os pontos $\hat{x}_i(\bar{y})$ não constituiriam então uma solução primal porque seriam inviáveis.

O algoritmo de Feixes evita esta dificuldade pela forma com que é construído o subgradiente que se deve aproximar do vetor zero. A saber, se em (II.25), sustituirmos a expressão:

$$v^j = a - \sum_{i=1}^{n} A_i x_i (y^j) \quad , \quad j=1,\ldots\ldots,k,$$

obtém-se,

$$-d^k = a - \sum_{i=1}^{n} A_i x_i^k \quad ,$$

onde,

$$x_i^k = \sum_{j=1}^{k} \lambda_j^k \; x_i (y^j) .$$

A linearidade da restrição do problema (II.21) faz então com que os $x_1^k, x_2^k, \ldots\ldots\ldots x_n^k$, sejam próximos da viabilidade primal, se a norma de d^k for próxima de zero, o que evita a dificuldade anteriormente comentada. Testes computacionais são apresentados extensivamente por MEDHI(1987). ROBINSON(1987), fez um estudo completo da convergência do método proposto.

II.2.2.2 <u>Um problema sem restrições</u>.

NGUYEN(1988) propôs um método de decomposição baseado na Teoría da Dualidade de Fenchel, ROCKAFELLAR(1970), para o seguinte problema:

$$\text{Mín } f_o(x) + \sum_{i=1}^{m} f_i(x)$$

$$x \in \mathbb{R}^n.$$

Utilizando o Princípio do Problema Auxiliar, COHEN(1980), resolve este problema por meio da resolução de subproblemas de tamanho menor. Para provar a convergência, utiliza resultados da Teoria de

Epi-convergência, ATTOUCH(1984).

II.2.2.3 Referências Adicionais.

Por abordagens semelhantes às desta seção, têm sido desenvolvidos vários outros trabalhos como por exemplo, MIFFLIN(1985), GOL'SHTEIN(1986), TSENG e BERTSEKAS(1987).

II.2.3. Caso não-convexo.

Uma alternativa atrativa que evita algumas das dificuldades da aproximação dual clássica é oferecida pelos métodos de Lagrangeana Aumentada. Esta família provém do método de multiplicadores de HESTENES(1969), POWELL(1969), e ROCKAFELLAR(1973). Destes, o método proximal de multiplicadores, ROCKAFELLAR(1976b,1978) consiste no seguinte:

Seja o problema:

$$\text{Min} \sum_{j=1}^{n} f_{oj}(x_j) := f_o(x)$$

s.a.

$$f_i(x) := \sum_{j=1}^{n} f_{ij}(x_j) \leq 0 \quad , i=1,\dots\dots,m$$

$$x \in C$$

onde as f_{ij} , $i=0,\dots\dots,m$; $j=1,\dots\dots,n$, são funções de $\mathbb{R}^{d_j}$ em $\mathbb{R}$, $x_j \in \mathbb{R}^{d_j}$, $j=1,\dots\dots,n$.

$$\sum_{j=1}^{n} d_j = d \ , \ x^T = (x_1,\dots\dots,x_n), \ C = \prod_{j=1}^{n} C_j \subset \mathbb{R}^d, \ C_j \ , j=1,\dots n,$$

conjuntos fechados em $\mathbb{R}^{d_j}$ respectivamente.

Em cada passo tem-se $\bar{x}$ e $\bar{y}$ e deve-se minimizar a Lagrangeana Aumentada:

$$\rho(x)=f_o(x)+1/2\|x-\bar{x}\|^2+1/2\sum_{i=1}^{m}\text{máx}^2(0,\ f_i(x)+\bar{y}_i) \qquad (II.26)$$

sobre o conjunto C, para obter o próximo $\bar{x}$. Observa-se que estes métodos não levam diretamente a algoritmos de decomposição, porque $\rho(x)$ não é da forma $\sum_{j=1}^{n}\rho_j(x_j)$. Reconhece-se ser esta a principal desvantagem da aproximação via Lagrangeana Aumentada, quando é usada para propósitos de decomposição. Estratégias para abordagem deste problema têm sido estudadas na literatura.

STEPHANOPULOS e WESTERBERG(1975), propuserom linearizar a Lagrangeana Aumentada(originalmente UZAWA(1958) com a Lagrangeana Clássica) para induzir artificialmente a separabilidade. Vários métodos baseados nesta idéia são discutidos por FINDEISEN et al.(1980).

Uma abordagem deste tipo usando linearização tipo Frank-Wolfe(1956) para $\rho(x)$ foi usada por GUNN et al.(1988).

Também BERTSEKAS(1979) obteve uma forma de convexificar problemas não-convexos sem destruir a separabilidade. Sua abordagem não é porém de maior interesse se o problema é convexo, já que este procedimento exige resolver uma succesão de problemas de igual dificuldade ao problema original.

SPINGARN(1985), sustitui $\rho(x)$ (II.26), por uma forma decomponível através de uma função da forma $\sum_{j=1}^{n}\rho_j$, onde

$$\rho_j(x_j)=f_{oj}(x_j)+1/2\|x_j-\bar{x}_j\|^2+1/2\sum_{i=1}^{m}\text{máx}^2(0,f_{ij}(x_j)-\bar{u}_{ij}+\bar{y}_i).$$

LEMARECHAL(1989), utiliza uma abordagem de tipo proximal junto com a Lagrangeana Clássica de maneira a ter êxito na recuperação da solução primal. Ele estuda o seguinte problema primal:

$$\text{Mín} \sum_{i=1}^{N} f_i(x_i) := f(x)$$

s.a.

$$\sum_{i=1}^{N} g_i(x_i) := g(x) = 0 \in \mathbb{R}^n \qquad \text{(II.27)}$$

$$x_i \in D_i \subset \mathbb{R}^{n_i}, \quad i=1,\dots\dots,N.$$

Para λ fixo em $\mathbb{R}^n$, sua função de Lagrange pode ser minimizada resolvendo os N problemas locais seguintes:

$$P_i(\lambda): \qquad q_i(\lambda) := \min_{z \in D_i} \{f_i(z) + \lambda^T g_i(z)\}.$$

Seja $\bar{x}(\lambda) := (\bar{x}_i(\lambda),\ i=1,\dots\dots,N)$, onde $\bar{x}_i(\lambda)$ é a solução do subproblema $P_i(\lambda)$.

Já se viu na seção II.2.1 que esta solução pode não ser única. Também, para que $\bar{x}(\lambda)$ resolva o problema (II.27), é necessário que λ seja uma solução, digamos λ^* do problema dual associado:

$$\text{máx}\left\{ q(\lambda) := \sum_{i=1}^{N} q_i(\lambda),\ \lambda \in \mathbb{R}^n \right\}.$$

Tradicionalmente, o algoritmo de geração de colunas, DANTZIG(1963), tem sido usado para calcular λ^*. Sendo conhecidos os iterados $\lambda^1, \lambda^2, \dots\dots\dots, \lambda^k$ e os correspondentes $\gamma^j := g[\bar{x}(\lambda^j)]$, λ^{k+1} resolve o problema linear:

$$\underset{\lambda\in\mathbb{R}^n}{\text{Máx}} \quad \text{Mín}\left\{q(\lambda^j)+(\lambda-\lambda^j)^T\gamma^j, \; j=1,\ldots\ldots,k\right\}. \qquad \text{(II.28)}$$

Seu problema dual é: (lembre-se que $q-\lambda^T g=f$)

$$\underset{\alpha\in\mathbb{R}^k}{\text{mín}} \quad \sum_{j=1}^{k}\alpha^j f(x^j)$$

$$\alpha^j \geq 0 \;, \; j=1,\ldots\ldots\ldots,k \qquad \text{(II.29)}$$

$$\sum_{j=1}^{k}\alpha^j=1$$

$$\sum_{j=1}^{k}\alpha^j\gamma^j=0$$

E' sabido que a convergência deste método é muito lenta.

Algumas abordagens têm sido propostas, SHOR(1985), baseadas na otimização de subgradientes.

A fim de maximizar a função dual (II.28), Lemarechal a perturba obtendo:

$$\underset{\lambda\in\mathbb{R}^n}{\text{máx}} \quad \text{mín}\left\{-1/2\,|\lambda-\lambda^k|^2+f(x^j)+\lambda^T\gamma^j \;, \; j=1,\ldots\ldots,k\right\},$$

ou equivalentemente seu dual (II.29) como:

$$\min_{\alpha\in\mathbb{R}^k} \sum_{j=1}^{k}\alpha^j f(x^j)+1/2\|\sum_{j=1}^{k}\alpha^j\gamma^j\|^2$$

$$\alpha_j\geq 0,\ j=1,\ldots\ldots\ldots,k,$$

$$\sum_{j=1}^{k}\alpha_j=1,$$

o que é essencialmente o método de Feixes, LEMARECHAL et al.(1981) e KIWIEL(1983). Para k grande, $\sum_{j=1}^{k}\alpha^j f(x^j)$ estará próximo de $q(\lambda^k)$ e $\sum_{j=1}^{k}\alpha^j\gamma^j$ próximo de zero.

A combinação primal, a saber,

$$\tilde{x}^k := \sum_{j=1}^{k}\alpha^j\bar{x}(\lambda^j),$$

converge ao ótimo primal no caso convexo, ver MEDHI(1987).

Porém, no caso não convexo $\tilde{x}^k$ pode não satisfazer as restrições locais. Para evitar esta dificuldade Lemarechal propõe penalizar o desvio de $\tilde{x}^k$, isto é, a função objetivo dos subproblemas $P_i(\lambda)$ é perturbada por:

$$f_i(z)+\lambda^T g_i(z)+\pi\|z-\tilde{x}_i^k\|^2.$$

Agora o resultado satisfaz automáticamente as restrições locais. O problema resolvido por Lemarechal é o mesmo estudado em BERTSEKAS(1982), chamado "unit-commitment problem" obtendo bons resultados.

II.3. Outros Métodos.

Existe uma multiplicidade de abordagens que têm sido apresentados na literatura, muitas das quais dependendo da estrutura particular do problema, ver por exemplo, SHAPOUR e LI(1986), LUNA(1984), entre outros.

Mas por motivos de espaço, veremos a seguir apenas um destes métodos, por ter uma certa relação com o que se propõe nesta tese, a ser apresentado no próximo capítulo.

II.3.1.-Método do Lagrangeano Viável.(WISMER(1971),(1978))

Seja o problema de programação não-linear:

$$\text{Mín } f(x):=\sum_{i=1}^{N} f_i(x_i)$$

$$\text{s.a.} \qquad\qquad (II.30)$$

$$g_i(x)\leq 0 \ , \quad i=1,\ldots\ldots\ldots,N,$$

onde $x^T=(x_1,x_2,\ldots\ldots\ldots,x_N)\in\mathbb{R}^{\tilde{N}}$,

com $x_i\in\mathbb{R}^{n_i}$, $i=1,\ldots\ldots\ldots,N$, $\tilde{N}=\sum_{i=1}^{N} n_i$; $f_i:\mathbb{R}^{n_i}\longrightarrow\mathbb{R}$ e

$g_i:\mathbb{R}^{\tilde{N}}\longrightarrow\mathbb{R}^{m_i}$, $i=1,\ldots\ldots\ldots,N$ funções,

$m:=\sum_{i=1}^{N} m_i$, $m\leq N$.

Agora considere o seguinte problema em $\mathbb{R}^{2\tilde{N}}$ equivalente a (II.30):

$$\text{Min} \sum_{i=1}^{N} f_i(x_i)$$

s.a.

$$g_i(w_1, w_2, \ldots\ldots\ldots, w_{i-1}, x_i, w_{i+1}, \ldots\ldots, w_N) \leq 0$$

$$w_i = x_i, \qquad \text{(II.31)}$$

$$i = 1, \ldots\ldots\ldots, N$$

A Lagrangeana clássica para este problema é:

$$L(x,w,y,\lambda) := \sum_{i=1}^{N} \Big[f_i(x_i) + y_i^T g_i(w_1, w_2, \ldots, w_{i-1}, x_i, w_{i+1}, \ldots, w_N) + \\ + \lambda_i^T (x_i - w_i) \Big] \qquad \text{(II.32)}$$

$$:= \sum_{i=1}^{N} L_i(x_i, w, y_i, \lambda_i).$$

Vemos que fixando w a função Lagrangeana L é inteiramente decomponível. De fato, é possível aplicar as condições de Kuhn-Tucker a cada L_i(expressão entre colchetes), obtendo no primeiro nível da decomposição:

$$\frac{\partial L_i}{\partial x_i} = \frac{\partial f_i}{\partial x_i} + \left(\frac{\partial g_i}{\partial x_i} \right)^T y_i + \lambda_i = 0$$

$$\frac{\partial L_i}{\partial y_i} = g_i(w_1, w_2, \ldots\ldots\ldots, w_{i-1}, x_i, w_{i+1}, \ldots\ldots, w_N) \leq 0$$

$$\frac{\partial L_i}{\partial \lambda_i} = x_i - w_i = 0$$

$$y_i^T g_i = 0$$

$$y_i \geq 0 \ , \quad i=1,\ldots\ldots\ldots,N.$$

Este primeiro nível fornece as correspondentes soluções: $x_i(w)$, $y_i(w)$, $\lambda_i(w)$.

O segundo nível do esquema é a determinação do valor ótimo de w. Para este efeito aplica-se um método do tipo gradiente. Tem-se:

$$\frac{\partial L}{\partial w_i} = 0 \ , \quad i=1,\ldots\ldots,N.$$

Agora

$$\frac{\partial L}{\partial w_i} = \sum_{\substack{j=1 \\ j\neq i}}^{N} \left[\frac{\partial g_j}{\partial w_i} \right]^T y_j - \lambda_j \ , \quad i=1,\ldots\ldots\ldots,N.$$

Logo temos o diferencial correspondente, que é:

$$dL = \sum_{i=1}^{N} \left[\frac{\partial L}{\partial w_i} \right]^T dw_i \ ,$$

$$= \sum_{i=1}^{N} \left[\sum_{\substack{j=1 \\ j\neq i}}^{N} \left[\frac{\partial g_j}{\partial w_i} \right]^T y_j - \lambda_i \right]^T dw_i \ .$$

Para que se tenha dL<0, pode-se considerar:

$$dw_i = k \left[\sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\frac{\partial g_j}{\partial w_i} \right)^T y_j - \lambda_i \right] \quad , \quad i=1,\ldots\ldots\ldots,N,$$

para algum k<0.

Logo o algoritmo de otimização associado ao segundo nível é dado por:

$$\left. w_i^{q+1} = w_i^q + k \left[\sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\frac{\partial g_j}{\partial w_i} \right)^T y_j - \lambda_i \right] \quad , \quad i=1,\ldots\ldots\ldots,N. \right\} \; q=1,2,\ldots$$

$i=1,\ldots\ldots,N.$

CAPÍTULO III

O MÉTODO DE PENALIZAÇÃO LAGRANGEANA

Neste Capítulo revisamos brevemente o Método de Penalização Lagrangeana de HESTENES-POWELL-ROCKAFELLAR, ROCKAFELLAR(1971, 1973, 1974), seguindo muito de perto a revisão feita por CONTESSE(1987b), para a resolução de problemas de otimização não-lineares sob restrições de igualdade, desigualdade ou mistas. Começando com alguns antecedentes históricos ao método, continuamos com sua descrição; a seguir, detalhamos o principal resultado de convergência, dando uma demostração simplificada, CONTESSE(1987b). No final, tratamos aspectos teóricos relacionados com a implementação eficiente do método.

III.1 ALGUNS ANTECEDENTES HISTÓRICOS DO MÉTODO

Durante muito tempo os métodos de funções de penalização, FIACCO e McCORMICK(1968), foram reconhecidos como uma técnica eficiente para resolver problemas de otimização restrita. Como é sabido, eles requerem a solução de uma sequência de problemas auxiliares irrestritos ou com menos restrições que o problema original. Mas, é bem conhecido que à medida que cresce o parâmetro de penalização, os subproblemas auxiliares tornam crescentemente mal condicionados, de maneira tal que sua resolução numérica passa a ser difícil, se não impossível, FIACCO e McCORMICK(1968), LUENBERGER(1973). Por outro lado, no caso de problemas de minimização convexa as técnicas de relaxação lagrangeana, LUENBERGER(1973), definem um certo tipo de método de penalização finita, que são livres deste tipo de dificuldade. O primeiro método conhecido nesta família é o de Gradiente de UZAWA, ARROW et al.(1958). Uma extensão a problemas gerais não convexos com restrições de igualdade foi dada independentemente por HESTENES(1969), e POWELL(1969), e, um ano depois, por HAARHOFF e BUYS(1970). Esta extensão é amplamente conhecida como o Método de

Multiplicadores. Este novo método é baseado em uma utilização conjunta da penalização quadrática, assim como da relaxação lagrangeana. De fato, o método de multiplicadores pode ser visto como a aplicação do Método de Uzawa ao problema localmente convexificado definido pela minimização da função de penalização exterior quadrática para o problema original, sujeito às mesmas restrições originais. Um completo estudo deste método, também com uma extensa bibliografía para diferentes resultados relacionados, é feito em BERTSEKAS(1982). Em particular, os primeiros resultados de convergência local foram dados por BUYS(1972) e RUPP(1972). Resultados de convergência relacionados foram dados por WIERZBICKI(1971). Também, resultados de convergência de natureza global e para um parâmetro de penalidade variável foram dados independentemente em BERTSEKAS(1973) e, POLYAK e TRET'YAKOV(1973).

Algoritmos baseados em penalização lagrangeana para o caso de restrições de igualdade têm sido estudados por ARROW e SOLOW(1958), BERTSEKAS(1975), FLETCHER(1970,1973), FLETCHER e LILL(1971), KORT e BERTSEKAS(1972), LILL(1973), MIELE et al.(1971a,1971b,1972a,1972b), POLYAK(1970), TRIPATHI e NARENDRA(1972), WIERZBICKI(1970) e, WIERZBICKI e HATKO(1973), entre outros.

A extensão do método de multiplicadores ao caso de restrições de desigualdade, foi feita primeiro por ROCKAFELLAR(1971,1973). Isto justifica a denominação comumente encontrada na literatura de Método de Penalização Lagrangeana de Hestenes-Powell-Rockafellar. Resultados relacionados com esta forma mais geral do método foram também dados por BUYS(1972). Esta extensão pode ser vista como um método de relaxação lagrangeana para alguma função lagrangeana aumentada. Além disso, em ROCKAFELLAR(1974b), foi dada uma forma precisa de dualidade para este método de penalização. Concretamente Rockafellar prova que qualquer ponto estacionário restrito regular que satisfaz a condição suficiente de otimalidade de 2ª ordem para mínimo local estrito, junto com o multiplicador de Kuhn-Tucker associado, define um único ponto de sela local para esta função lagrangeana aumentada. Este resultado estende um resultado similar de ARROW, GOULD e HOWE(1973), para lagrangeanas gerais onde a condição de estrita complementaridade foi imposta. Sob esta última hipótese, o método geral de multiplicadores tem o mesmo comportamento local que o método original aplicado ao problema só com restrições de igualdade que resulta da transformação de todas as restrições de desigualdades ativas em restrições de igualdade e a eliminação de todas as restrições de desigualdade não ativas. Como consequência disto, todos os resultados de convergência local para o caso clássico são ainda válidos para este caso mais geral.

III.2 MÉTODO DE PENALIZAÇÃO LAGRANGEANA

Seja o problema geral de minimização

$$\text{P)} \quad \begin{aligned} &\text{Min } f(x) \\ &\text{s.a.} \\ &g_i(x) = 0, \quad i=1,2,\ldots\ldots\ldots,q, \\ &g_i(x) \le 0, \quad i=q+1,\ldots\ldots\ldots,p \end{aligned}$$

onde as funções $f(\cdot)$ e $g_i(\cdot)$, $i=1,\ldots\ldots\ldots,p$, são duas vezes localmente continuamente diferenciáveis.

Tal como foi dito na seção anterior, por meio da reformulação de P) como um problema clássico com restrições de igualdade, ROCKAFELLAR(1973), estende o método de multiplicadores ao caso geral. Logo, a nova lagrangeana aumentada fica definida por:

$$L(x,\lambda,\mu,\bar{r}) = L_0(x,\lambda,r) + \sum_{i=q+1}^{p} \psi(g_i(x);\mu_i,\tilde{r}_i) \qquad \text{(III.1)}$$

onde,

$$\psi(t,\mu_i,\tilde{r}_i)=\begin{cases}\mu_i t + \dfrac{\tilde{r}_i}{2}\, t^2, \text{ se } t \geq -\dfrac{\mu_i}{\tilde{r}_i}\\[2ex] -\dfrac{\mu_i^2}{2\tilde{r}_i}, \text{ se } t \leq -\dfrac{\mu_i}{\tilde{r}_i}\end{cases} \quad ,i=q+1,\dots\dots\dots p$$

$$L_o(x,\lambda,r):=f(x) + \sum_{i=1}^{q}\lambda_i g_i(x) + \sum_{i=1}^{q}\frac{r_i}{2} g_i^2(x),$$

e $\bar{r}:=(r,\tilde{r})$ é o vetor de penalização.

<u>O Algoritmo</u>:

<u>Passo 0</u>: Sejam λ^o, μ^o, x^o e para $\hat{r}$ suficientemente grande, seja $\bar{r}^o > \hat{r}$, dados; fazer $k = 0$;

<u>Passo 1</u>: Calcular o ponto x^k que é solução do subproblema:

$$\min_{x \in V(\hat{x})} L(x,\lambda^k,\mu^k,\bar{r}^k),$$

onde $V(\hat{x})$ é uma certa vizinhança de $\hat{x}$.

<u>Passo 2</u>: Atualizar λ^k,μ^k e $\bar{r}^k$ de acordo com as fórmulas de atualização dadas abaixo. Se não se tem convergência, fazer $k = k + 1$ e voltar ao passo 1. Caso contrário, parar.

<u>Fórmulas de atualização</u>:

$$\lambda_i^{k+1}= \lambda_i^k + r_i^k g_i(x^k) \quad , i=1,\dots\dots\dots q \qquad \text{(III.2)}$$

$$\mu_i^{k+1}= \text{máx } (0, \mu_i^k +\tilde{r}_i^k g_i(x^k)), \quad i=q+1,\dots\dots\dots p \qquad \text{(III.3)}$$

Seja $\bar{r}^{\,k+1}= (r^{k+1}, \tilde{r}^{\,k+1})$, então atualizar $\bar{r}^k$ por meio de uma heurística se necessário, ou fazer $\bar{r}^{k+1} = \bar{r}^k$.

III.3. RESULTADO DE CONVERGÊNCIA LOCAL

Nesta seção provamos a convergência local do método de penalização lagrangeana. Na primeira parte do teorema 1, se provam as relações clássicas de convergência do método de multiplicadores em uma certa forma direta. A idéia da demonstração é baseada no fato que o método pode ser visto como um método de gradiente de passo fixo para resolver localmente o problema dual associado com P). Além disto, a demonstração confirma o argumento de LUENBERGER(1973), que observou que estas iterações de gradiente são próximas às iterações de Newton quando o parâmetro de penalização converge a infinito.

Teorema 1.

Seja $\hat{x}$ um ponto de mínimo local estrito para o problema P), o qual satisfaz a condição de regularidade:

i) $\left\{Dg_i(\hat{x}),\ \forall\ i \in \{1,\ldots\ldots,q\} \cup I(\hat{x})\right\}$ é um conjunto linearmente independente, onde $I(\hat{x})$ representa o conjunto de índices das desigualdades ativas em $\hat{x}$, isto é,

$$I(\hat{x}) := \left\{i \in \{q+1,\ldots\ldots\ldots,p\} \ /\ g_i(\hat{x}) = 0\right\}$$

a condição suficiente de otimalidade local de 2^a ordem:

ii) $D_x\mathcal{L}(\hat{x},\hat{\lambda},\hat{\mu}) = 0$

iii) $d^T D^2_{xx}\ \mathcal{L}(\hat{x},\hat{\lambda},\hat{\mu})d > 0,\ \forall\ d \in S(0,1)$:

$$Dg_i(\hat{x})d = 0,\ \forall\ i \in \{1,\ldots\ldots,q\} \cup I(\hat{x});$$

onde $S(0,1) := \{d \in \mathbb{R}^n \ /\ \| d \| = 1\}$;

e a condição de folga de complementaridade estrita:

iv) $\hat{u}_i > 0,\ \forall\ i \in I(\hat{x})$,

onde $\mathcal{L}(x,\lambda,\mu)$ representa a função lagrangeana clássica para o problema P), $(\hat{\lambda},\ \hat{\mu}) \in \mathbb{R}^q \times \mathbb{R}^{p-q}_+$ o vetor de multiplicadores de Karush-Kuhn-Tucker associado com $\hat{x}$ (o subíndice + representa o ortante positivo correspondente).

Sob estas hipóteses, $\exists\ \hat{r} > 0$ e $\hat{\varepsilon} > 0$ tal que
$\forall\ 0 < \varepsilon \leq \hat{\varepsilon}$ e $r \geq \hat{r}$, $\exists\ \delta(r) := \delta(r,\varepsilon) > 0$ e $M > 0$, $M < \|\hat{r}\|$ tal que,
$\forall(\lambda^o,\mu^o) \in \mathbb{R}^p$, $\|(\lambda^o,\mu^o)-(\hat{\lambda},\hat{\mu})\| < \delta(r)$, temos:

$$\|(\lambda^{k+1},\mu^{k+1}) - (\hat{\lambda},\hat{\mu})\| \leq \left(\frac{M}{\|r\|}\right)\|(\lambda^k,\mu^k) - (\hat{\lambda},\hat{\mu})\|, \qquad \text{(III.3a)}$$

e

$$\|x^k - \hat{x}\| \leq \left(\frac{M}{\|r\|}\right)\|(\lambda^{k+1},\mu^{k+1}) - (\hat{\lambda},\hat{\mu})\|, \qquad \text{(III.3b)}$$

onde o vetor de multiplicadores $(\lambda^{k+1},\mu^{k+1})$ é iterativamente atualizado pelas fórmulas (III.2)-(III.3), e x^k é dado pela solução única do subproblema lagrangeano de minimização local:

$$\begin{aligned} &\text{Mín } L(x,\lambda^k,\mu^k,\bar{r}^k) \\ &\text{s.a.} \\ &\|x-\hat{x}\|<\varepsilon,\ \forall\ k \geq 0 \end{aligned}$$

<u>Demonstração</u>: (conforme CONTESSE 1987b)

Na primeira parte da demonstração, supomos que não existem desigualdades no problema P). Além disto, por simplicidade, admitimos que o vetor $\bar{r}$ tem todas as suas componentes iguais a um único escalar que representaremos por r. Por outra parte, consideraremos inicialmente a hipótese de convexidade local:

$$D^2_{xx}\ \mathcal{L}(\hat{x},\hat{\lambda}) \text{ é definida positiva}, \qquad \text{(III.4)}$$

onde $\mathcal{L}(x,\lambda)$ representa a função Lagrangeana clássica, para o problema P) só com restrições de igualdade.

Logo, por continuidade $\exists\ \hat{\varepsilon} > 0$ e $\hat{\delta} > 0$ tal que:

$D^2_{xx}\ \mathcal{L}(x,\lambda)$ é definida positiva, $\forall x \in B(\hat{x},\hat{\varepsilon})$, $\forall \lambda \in B(\hat{\lambda},\hat{\delta})$ (III.5)

Também, já que $Dg(\hat{x})$ é de posto máximo, podemos supor, por continuidade, que $\hat{\varepsilon}$ é suficientemente pequeno de tal maneira que:

$Dg(\hat{x})$ é de posto máximo, $\forall\ x \in B(\hat{x},\hat{\varepsilon})$ (III.6)

Agora, $\forall\ r \geq 0$, é claro que $(\hat{x},\hat{\lambda})$ é solução do sistema de equações não-lineares:

$$D_x L_o(x,\lambda,r) = 0, \qquad \text{(III.7)}$$

onde $L_o(x,\lambda,r)$ é a função de penalização lagrangeana clássica representada em (III.1).

De (III.4) e do Teorema da função implícita, $\forall r \geq 0$ e $\forall 0 < \varepsilon \leq \hat{\varepsilon}$, $\exists$ uma bola $B(\hat{\lambda},\delta(r))$ centrada em $\hat{\lambda}$, com raio $\delta(r) > 0$, $\delta(r) \leq \hat{\delta}$, e $x(\lambda,r)$ continuamente diferenciável em relação a λ, tal que $x(\lambda,r) \in \text{int}(B(\hat{x},\varepsilon))$, $\forall\ \lambda \in B(\hat{\lambda},\delta(r))$. Logo de (III.5) e (III.7), $x(\lambda,r)$ é um mínimo local estrito de $L_o(\cdot,\lambda,r)$ no $\text{int}(B(\hat{x},\varepsilon))$, $\forall \lambda \in B(\hat{\lambda},\delta(r))$.

Consequentemente, podemos definir, ao menos localmente, a função dual aumentada associada ao problema P) restrito só com igualdades, por meio de :

$$\psi(\lambda,r) = \underset{x \in \text{int}(B(\hat{x},\varepsilon))}{\text{Mín}} L_o(x,\lambda,r) = L_o(x(\lambda,r),\lambda,r), \qquad \text{(III.8)}$$

$\forall\ \lambda \in B(\hat{\lambda},\delta(r))$ e $\forall\ r \geq 0$. Agora, derivando (III.8) em relação a λ, temos:

$D_\lambda \psi(\lambda,r) = D_x L_o(x(\lambda,r),\lambda,r) \cdot D_\lambda x(\lambda,r) + D_\lambda L_o(x(\lambda,r),\lambda,r)$ e utilizando (III.7) tem-se:

$$D_\lambda \psi(\lambda,r) = D_\lambda L_o(x(\lambda,r),\lambda,r) = g(x(\lambda,r))^T, \qquad \text{(III.9)}$$

$\forall\ \lambda \in B(\hat{\lambda},\delta(r))$. Desta última relação se obtém que $\psi(\cdot,r)$ é duas vezes continuamente derivável em relação a λ, e

$$D^2_{\lambda\lambda}\ \psi(\lambda,r) = Dg(x(\lambda,r)) \cdot D_\lambda x(\lambda,r),\ \forall\lambda\in B(\hat{\lambda},\delta(r)).$$

Por outra parte, já que $x(\lambda,r)$ é solução de (III.7) $\forall\lambda\in B(\hat{\lambda},\delta(r))$, derivando esta expressão (III.7) em relação a λ temos:

$$D^2_{xx}\ L_o(x(\lambda,r),\lambda,r)D_\lambda x(\lambda,r)\ +\ Dg(x(\lambda,r))^T = 0$$

e, já que $D^2_{xx}L_o(x(\lambda,r),\lambda,r)$ é regular:

$$D_\lambda x(\lambda,r)\ =\ -D^2_{xx}L_o(x(\lambda,r),\lambda,r)^{-1}\ Dg(x(\lambda,r))^T, \qquad \text{(III.10)}$$

de maneira que:

$$D^2_{\lambda\lambda}\psi(\lambda,r) = -Dg(x(\lambda,r)D^2_{xx}L_o(x(\lambda,r),\lambda,r)^{-1}Dg(x(\lambda,r))^T. \qquad \text{(III.11)}$$

Logo, de (III.6) e da definição de $\delta(r) > 0$, $D^2_{\lambda\lambda}\ \psi(\lambda,r)$ é definida negativa, $\forall\lambda\in B(\hat{\lambda},\delta(r))$, e assim tem-se que a função dual aumentada $\psi(\cdot,r)$ é estritamente concava sobre $B(\hat{\lambda},\delta(r))$. Além disto, de (III.9) deduzamos que a fórmula de atualização (III.2) para os multiplicadores associados às restrições de igualdade:

$$\tilde{\lambda}(\lambda,r)\ =\ \lambda\ +\ rg(x(\lambda,r)), \qquad \text{(III.12)}$$

é uma iteração de ascenso com passo constante r para a maximização da função dual $\psi(\cdot,r)$ em relação a λ. Daqui temos então que:

$$\|\tilde{\lambda}(\lambda,r)\ -\ \hat{\lambda}\|\ =\ \|\lambda\ -\ \hat{\lambda}\ +\ rD^T_\lambda\psi(\lambda,r)\|$$

$$= \|\lambda-\hat{\lambda}+r\int_0^1\ D^2_{\lambda\lambda}\ \psi(\lambda(t),r)(\lambda-\hat{\lambda})dt\|$$

onde $\lambda(t):=\hat{\lambda}+t(\lambda-\hat{\lambda})$, $\forall\ t \in [0,1]$, ou equivalentemente:

$$\|\tilde{\lambda}(\lambda,r)-\hat{\lambda}\| = \left\|\left[\int_0^1 (I + rD^2_{\lambda\lambda}\ \psi(\lambda(t),r))dt\ \right](\lambda-\hat{\lambda})\right\|,$$

que implicam em:

$$\|\tilde{\lambda}(\lambda,r)-\hat{\lambda}\| \le \|\lambda-\hat{\lambda}\| \operatorname*{Máx}_{t\in[0,1]} \|I + rD^2_{\lambda\lambda}\ \psi(\lambda(t),r)\| . \qquad \text{(III.13)}$$

Por outra parte, já que $\tilde{\lambda}(\cdot,r)$ é uma função contínua de λ tal que $\tilde{\lambda}(\hat{\lambda},r) = \hat{\lambda}$, nós podemos considerar $\delta(r)>0$ suficientemente pequeno tal que:

$$\tilde{\lambda}(\lambda,r) \in B(\hat{\lambda},\hat{\delta}),\ \forall\lambda\in B(\hat{\lambda},\delta(r)) \qquad \text{(III.14)}$$

Consideremos agora

$$D^2_{\lambda\lambda}\ \psi(\lambda,r) = H(x(\lambda,r),\tilde{\lambda}(\lambda,r),r),$$

onde,

$$H(x,\lambda,r):=-Dg(x)\left[D^2_{xx}\ \mathscr{L}(x,\lambda)+rDg(x)^T Dg(x)\right]^{-1} Dg(x)^T.$$

Da fórmula de Sherman-Morrison (FLETCHER (1987), p.76), tem-se:

$$I + rH(x,\lambda,r) = \left[I - rW(x,\lambda)\right]^{-1},$$

com:

$$W(x,\lambda) = -Dg(x)\left[D^2_{xx}\ \mathscr{L}(x,\lambda)\right]^{-1} Dg(x)^T$$

a qual é válida $\forall x\in B(\hat{x},\hat{\varepsilon})$ e $\forall\lambda\in B(\hat{\lambda},\hat{\delta})$. Temos assim:

$$\|I+rD^2_{\lambda\lambda}\ \psi(\lambda,r)\| = \operatorname*{máx}_{i=1,..q} \frac{1}{1-r\nu_i\left[W(x(\lambda,r),\tilde{\lambda})\right]},$$

$\forall\lambda\in B(\lambda,\hat{\delta}(r))$, onde $\nu_i[A]$ representa o i-ésimo valor próprio de A, $\forall\ i\in\{1,......q\}$. Mas, de (III.5) e (III.6), tem-se que

$W(x,\lambda)$ é definida negativa $\forall x \in B(\hat{x},\hat{\varepsilon})$, $\forall \lambda \in B(\hat{\lambda},\hat{\delta})$, de maneira que:

$$\left\| I + rD^2_{\lambda\lambda}\ \psi(\lambda,r) \right\| \leq \frac{M}{r} \tag{III.15}$$

$\forall\ \lambda \in B(\hat{\lambda},\delta(r))$, onde a constante $M > 0$ é definida por:

$$M := \sup_{\substack{x\in B(\hat{x},\hat{\varepsilon}),\ \lambda\in B(\hat{\lambda},\hat{\delta}) \\ i=1,\ldots\ldots..q}} \left\{ \frac{1}{|v_i\,[W(x,\lambda)]\,|} \right\}$$

e é independente de r, $r \geq 0$. Em particular , $\hat{r}$ pode escolher-se suficientemente grande, de modo que se tenha: $M < \|\hat{r}\|$.

Logo, já que $\lambda(t) \in B(\hat{\lambda},\hat{\delta})$, $\forall\ \lambda \in B(\hat{\lambda},\hat{\delta})$, $\forall t \in [0,1]$, de (III.13) temos que:

$$\|\tilde{\lambda}(\lambda,r)-\hat{\lambda}\| \leq \frac{M}{r}\ \|\lambda-\hat{\lambda}\|, \tag{III.16}$$

$\forall \lambda \in B(\hat{\lambda},\delta(r))$, onde $\frac{M}{r} < 1$, $\forall r \geq \hat{r}$.

A seguir, provamos a relação de convergência das variáveis primais. Primeiramente, temos que:

$$x(\tilde{\lambda},r)-\hat{x}=\int_0^1 D_\lambda x(\tilde{\lambda}(t),r)(\tilde{\lambda}-\hat{\lambda})dt, \tag{III.17}$$

onde:

$$\tilde{\lambda}(t):=(1-t)\hat{\lambda}+t\tilde{\lambda},\ \ \forall t \in [0,1] \tag{III.18}$$

com $\tilde{\lambda}:=\tilde{\lambda}(\lambda,r):=\lambda+rg(x(\lambda,r))$ (como em (III.12)).

Então, já que:

$$D_\lambda x(\tilde{\lambda}(t),r)=x_\lambda(\tilde{\lambda}(t),r)[I+rDg(x(\tilde{\lambda}(t),r))x_\lambda(\tilde{\lambda}(t),r)]t,$$

de (III.17), (III.11) e (III.10), temos que:

$$\|x(\tilde{\lambda},r)-\hat{x}\|\leq(\underset{t\in[0,1]}{\text{Máx}}\|x_\lambda(\tilde{\lambda}(t),r)\|)(\underset{t\in[0,1]}{\text{Máx}}\|I+rD^2_{\lambda\lambda}\psi(\tilde{\lambda}(t),r)\|)\|\tilde{\lambda}-\hat{\lambda}\|. \tag{III.19}$$

Por outra parte, de (III.10), tem-se que:

$$\|x_\lambda(\lambda,r)\|\leq S\|D^2_{xx}\ L_o(x(\lambda,r),\lambda,r)^{-1}\|, \tag{III.20}$$

onde $S:=\sup_{x\in B(\hat{x},\hat{\varepsilon})}\|Dg(x)\|$,

com S constante estritamente positiva devido à hipótese (III.6).

Observemos que a inversa da hessiana de L_o pode ser escrita como:

$$D^2_{xx}L_o(x,\tilde{\lambda},r)^{-1}=[(D^2_{xx}\mathscr{L}(x,\tilde{\lambda}))^{1/2}(D^2_{xx}\mathscr{L}(x,\tilde{\lambda}))^{1/2}+rDg^T(x)Dg(x)]^{-1}$$

ou, equivalentemente, devido à simetria de $(D^2_{xx}\mathscr{L}(x,\lambda))^{1/2}$:

$$D^2_{xx}L_o(x,\tilde{\lambda},r)^{-1}=[D^2_{xx}\mathscr{L}(x,\tilde{\lambda})]^{-1/2}[I+rM(x,\tilde{\lambda})]^{-1}[D^2_{xx}\mathscr{L}(x,\tilde{\lambda})]^{-1/2}, \tag{III.21}$$

onde:

$$M(x,\lambda):=[Dg(x)D^2_{xx}\mathscr{L}(x,\lambda)^{-1/2}]^T[Dg(x)D^2_{xx}\mathscr{L}(x,\lambda)^{-1/2}].$$

A seguir, repetindo o desenvolvimento que fizemos para a matriz
$[I-rW(x,\lambda)]^{-1}$, obtemos:

$$\|[I+rM(x,\lambda)]^{-1}\|\leq 1,\ \forall x\in B(\hat{x},\hat{\varepsilon}),\ \forall\lambda\in B(\hat{\lambda},\hat{\delta}). \tag{III.22}$$

Então, de (III.22), (III.21), (III.20), (III.18) e (III.14), tem-se que:

$\|x_\lambda(\tilde{\lambda}(t),r)\|\le C,\ \forall\lambda\in B(\hat{\lambda},\delta(r)),\ \forall t\in[0,1],$

para uma certa constante C independente de r:

$$C:=S\ \text{Máx}\left\{\|D_{xx}^{2}\mathcal{L}(x,\lambda)^{-1}\| /\ x\in B(\hat{x},\hat{\varepsilon}),\ \lambda\in B(\hat{\lambda},\hat{\delta})\right\}.$$

Logo, de (III.19) e (III.15), temos que:

$$\|x(\tilde{\lambda},r)-\hat{x}\|\le \frac{K}{r}\|\tilde{\lambda}-\hat{\lambda}\|, \qquad \text{(III.23)}$$

$\forall\lambda\in B(\hat{\lambda},\delta(r))$, para uma certa constante K.

Redefinindo M por Máx(M,K), verificamos que com (III.16) e (III.23) as relações clássicas de convergência, sob a hipótese de convexidade local (III.4), se satisfazem. Tomando $\lambda^{o}\in\mathbb{R}^{q}$ tal que $\|\lambda^{o}-\hat{\lambda}\|\le\delta(r)$, podemos provar indutivamente que:

$$\|\lambda^{k+1}-\hat{\lambda}\|\le \frac{M}{r}\|\lambda^{k}-\hat{\lambda}\|, \qquad \text{(III.24)}$$

e

$$\|x^{k+1}-\hat{x}\|\le \frac{M}{r}\ \|\lambda^{k+1}-\hat{\lambda}\|, \qquad \text{(III.25)}$$

$\forall$ k = 0,1,2.....,$\forall r>0$ suficientemente grande, isto é, $r\ge r_o$, para algum $r_o>0$ tal que $\frac{M}{r_o}<1$.

No caso geral, sob a hipótese iii) do teorema, e como consequência do Lema de Debreu-Finler (ARROW, GOULD e HOWE,(1973)), $\exists\hat{r}>0$ tal que:

$$d^{T}D_{xx}^{2}L_{o}(\hat{x},\hat{\lambda},r)d>0,\ \forall d\in\mathbb{R}^{n},\ d\ne 0,\ \forall r\ge\hat{r}.$$

Sob estas condições, aplicar o método de Multiplicadores ao problema original restrito só com igualdades, é equivalente a aplicar este mesmo método ao problema localmente convexificado:

$$\text{Min } f(x) + \frac{\hat{r}}{2} \sum_{i=1}^{q} g_i(x)$$

s.a.

$$g_i(x) = 0, \qquad i=1,\ldots\ldots,q,$$

onde os multiplicadores são atualizados por:

$$\lambda^{k+1} = \lambda^k + (r - \hat{r})g(x^k), \quad k=0,1,2,\ldots\ldots$$

Como $r > \hat{r}$ e lembrando que $g(\hat{x}) = 0$, de (III.12) e (III.16) temos que:

$$\|\lambda^{k+1} - \hat{\lambda}\| = \|\lambda^k + rg(x^k) - \hat{\lambda} - \hat{r}(g(x^k) - g(\hat{x}))\| \le \frac{M}{(r-\hat{r})} \|\lambda^k - \hat{\lambda}\|,$$

$\forall k=0,1,2.\ldots.$ e $\forall r \ge \hat{r} + r_o$, para algum $r_o > 0$.
Logo:

$$\|\lambda^k + rg(x^k) - \hat{\lambda}\| \le \frac{M}{(r-\hat{r})} \|\lambda^k - \hat{\lambda}\| + \hat{r}\|g(x^k) - g(\hat{x})\|, \quad \forall k=0,1,2.\ldots,$$

$$\forall r \ge \hat{r} + r_o \qquad \text{(III.26)}$$

Por outra parte, como $g(\cdot)$ é continuamente diferenciável, existe uma constante $L > 0$ tal que:

$$\|g(x) - g(z)\| \le L\|x - z\|, \quad \forall x, z \in B(\hat{x}, \hat{\varepsilon}). \qquad \text{(III.27)}$$

Temos assim de (III.27), (III.26) e (III.25), que:

$$\|\lambda^k + rg(x^k) - \hat{\lambda}\| \le \frac{M}{(r-\hat{r})} \|\lambda^k - \hat{\lambda}\| + \frac{M\,\hat{r}}{(r-\hat{r})} M\,L\|\lambda^k - \hat{\lambda}\|, \quad \forall k=0,1.\ldots,$$

ou:

$$\|\lambda^k + rg(x^k) - \hat{\lambda}\| \le \left(\frac{\bar{M}}{r}\right) \left[\frac{1 + \hat{r}}{1 - \frac{\hat{r}}{r}}\right] \|\lambda^k - \hat{\lambda}\|.$$

$\forall k=0,1,2,\ldots, \forall r \geq \hat{r}+r_o$, onde $\tilde{M} = \text{máx}[M, M*L]$,

ou ainda:

$$\|\lambda^k + r g(x^k) - \hat{\lambda}\| \leq \frac{K}{r} \|\lambda^k - \hat{\lambda}\| \quad , \forall k=0,1,2\ldots \qquad \text{(III.28)}$$

onde,

$$K = \frac{\bar{M}\,(1+\hat{r})}{1 - \dfrac{\hat{r}}{(\hat{r}+r_o)}} .$$

Escolhendo $r_o > 0$ suficientemente grande, de maneira que $\frac{K}{r} < 1$, $\forall\ r \geq \hat{r}+r_o$, a relação (III.28) garante a relação (III.24) de convergência para os multiplicadores no caso em que não se tenha a hipótese de convexidade local (III.4). Deste resultado se conclui também a convergência (III.25) das variáveis primais sem a hipótese (III.4). Provamos portanto o resultado clássico de convergência.

Consideramos a seguir o seguinte problema auxiliar de minimização:

$$\begin{aligned} &\text{Mín } f(x) \\ &\text{s.a.} \\ &g_i(x)=0, \ \forall i \in \{1,2,\ldots\ldots,q\} \cup I(\hat{x}), \end{aligned} \qquad \text{(III.29)}$$

onde $I(\hat{x}) = \left\{ i \in \{q+1,\ldots\ldots,p\} \,/\, g_i(\hat{x})=0 \right\}$.

Aplicando os resultados da primeira parte da demonstração ao problema (III.29), sob as hipóteses i), ii) e iii), existe $\hat{r}>0$ tal que $r>\hat{r}$ e $\forall \varepsilon > 0$ suficientemente pequeno, existe $\delta(r,\varepsilon)>0$ e uma função continuamente diferenciável $x(\cdot,\cdot,r)$ definida sobre $B((\hat{\lambda},\hat{\mu}),\delta(r,\varepsilon))$, tal que $x(\lambda,\mu,r)$ é a única solução do problema:

$$\begin{aligned} &\text{Mín } L_o(x,(\lambda,\mu),r), \qquad \text{(III.30)} \\ &\|x-\hat{x}\| < \varepsilon \end{aligned}$$

$\forall(\lambda,\mu)\in B((\hat{\lambda},\hat{\mu}),\delta(r,\varepsilon))$, onde $L_o(x,(\lambda,\mu),r)$ representa a função lagrangeana aumentada clássica associada ao problema (III.29), com $x(\hat{\lambda},\hat{\mu},r)=\hat{x}$.

Além disto, existe uma constante $M>0$ tal que as relações de convergência (III.3a) e (III.3b), adequadamente modificadas para este problema, são satisfeitas. Para terminar a demonstração, temos somente a provar que $\forall\varepsilon\in[0,\hat{\varepsilon}]$, para algum $\hat{\varepsilon}>0$, com $\hat{u}_i>0$ $\forall i\in I(\hat{x})$, $\delta(r,\varepsilon)$ pode ser redefinido suficientemente pequeno tal que $x(\lambda,\mu,r)$ é de fato um único ponto de mínimo do problema:

$$\begin{array}{ll} \text{Mín } L_o(x,\lambda,\mu,r) & \\ \text{s.a.} & \text{(III.31)} \\ \|x-\hat{x}\|<\varepsilon & \end{array}$$

$\forall(\lambda,\mu)\in B((\hat{\lambda},\hat{\mu}),\delta(r,\varepsilon))$. Por simplicidade, μ representa agora o vetor de multiplicadores associado a todas as restrições de desigualdade do problema original e $x(\lambda,\mu,r)$ a solução ótima do problema (III.30) em termos destas novas variáveis.

Primeiramente, da continuidade das restrições de desigualdades ativas em $\hat{x}$, podemos definir $\varepsilon>0$ suficientemente pequeno tal que $\forall\beta>0$, $\beta<\hat{\mu}_i$ $\forall i\in I(\hat{x})$, se tenha:

$$|g_i(x)|=|g_i(x)-g_i(\hat{x})|<\frac{\hat{\mu}_i-\beta}{r}, \quad \forall i\in I(\hat{x}), \text{ logo,}$$

$$g_i(x)>\frac{-\hat{\mu}_i+\beta}{r}, \quad \forall i\in I(\hat{x}) \qquad \text{(III.32)}$$

$\forall x\in B(\hat{x},\varepsilon)$. Por outra parte, do fato que $g_i(\hat{x})<0$, $\forall i\in\{q+1,\ldots,p\}\setminus I(\hat{x})$, é possível redefinir $\varepsilon>0$ e $\beta>0$ suficientemente pequenos, tais que:

$$g_i(x)<-\frac{\beta}{r} \quad \forall i\in\{q+1,\ldots\ldots,p\}\setminus I(\hat{x}), \qquad \text{(III.33)}$$

$\forall x\in B(\hat{x},\varepsilon)$.

Portanto, de (III.32), (III.33) e a definição de $L_o(x,\lambda,\mu,r)$, é claro que $x(\lambda,\mu,r)$ é o único ponto de mínimo do problema (III.31), $\forall(\lambda,\mu)\in B((\hat{\lambda},\hat{\mu}),\delta(r,\varepsilon))$: supondo que $\delta(r,\varepsilon)>0$, é suficientemente pequeno de tal modo que por uma parte $\delta(r,\varepsilon)<\beta$ e, por outra, $x(\lambda,\mu,r)\in B(\hat{x},\varepsilon)$, $\forall(\lambda,\mu)\in B((\hat{\lambda},\hat{\mu}),\delta(r,\varepsilon))$.

Finalmente temos:

$0<\mu_i(\lambda,\mu,r)=\mu_i+rg_i(x(\lambda,\mu,r))$, $\forall i\in I(\hat{x})$, e

$0=\mu_i(\lambda,\mu,r)$, $\forall i\in\{q+1,\ldots\ldots\ldots,p\}\setminus I(\hat{x})$,

$\forall(\lambda,\mu)\in B((\hat{\lambda},\hat{\mu}),\delta(r,\varepsilon))$, de maneira que a convergência neste caso é uma consequência do resultado de convergência da primeira parte da demonstração. Logo, tem-se a prova completa do teorema.■ ■

<u>Observação 1</u>:

Do passo (III.15) da demostração do teorema 1, tem-se que $\|I+rD^2_{\lambda\lambda}\psi(\lambda,r)\|\leq\frac{M}{r}$, ou equivalentemente

que, $D^2_{\lambda\lambda}\psi(\lambda,r)=-\frac{1}{r}\,I+o\left(\frac{1}{r}\right)$.

Este resultado confirma o argumento de LUENBERGER(1973), de que, se r cresce para +∞ a iteração (III.12) se aproxima do passo de Newton para resolver o Problema Dual.

III.4. IMPLEMENTAÇÃO EFICIENTE DO MÉTODO DE PENALIZAÇÃO LAGRANGEANA

O método de penalização lagrangeana tem a dificuldade aparente que em cada iteração correspondente às atualizações dos multiplicadores por meio de (III.2) e (III.3), as minimizações da função lagrangeana aumentada devem ser feitas exatamente. Entretanto vários pesquisadores, POLYAK e TRET'YAKOV(1973), BUYS(1972), HAARHOFF e BUYS(1970), BERTSEKAS(1975,1976a,1976b), MIELE et al.(1971a), ROCKAFELLAR(1974b), entre outros, provaram que não se precisava resolver os problemas lagrangeanos de maneira exata. As soluções obtidas desta forma passaram a chamar-se soluções assintoticamente exatas. Os esquemas utilizados para obtê-las são baseados na idéia de aplicar precisão moderada nas primeiras minimizações da lagrangeana aumentada, para ir aumentando a precisão através das iterações seguintes.

Por exemplo, no caso de um problema só com restrições de igualdade, BUYS(1972) propôs o seguinte critério de parada para as minimizações lagrangeanas:

$$\|\nabla L_0(x^k,\lambda^k,r^k)\|\le\varepsilon_k \ ,$$

onde $\langle\varepsilon_k\rangle$ é uma sucessão decrescente de números reais positivos convergente a zero. BERTSEKAS(1976b) provou que este esquema destrói a rapidez de convergência linear do método de multiplicadores. Ele modificou este critério de modo a preservar a rapidez de convergência, substituindo-o por:

$$\|\nabla L_0(x^k,\lambda^k,r^k)\|\le \min\left\{\xi_k,\eta_k\left[\sum_{j=1}^{q} g_j^2(x^k)\right]^{1/2}\right\}$$

onde $\langle\xi_k\rangle$ y $\langle\eta_k\rangle$ são sucessões decrescentes de números reais positivos convergentes a zero.

No caso geral de restrições de igualdade e desigualdade, últimamente CONTESSE(1990a), propôs um esquema análogo ao de Bertsekas, o qual é válido na ausência da hipótese de estrita complementariedade.

Para o problema P) da seção (III.2), o critério de parada consiste em:

$$\|\nabla L_o(x^k,\lambda^k,\mu^k,r^k)\| \le \frac{\xi_k}{R},$$

e

$$\|\nabla L_o(x^k,\lambda^k,\mu^k,r^k)\| \le \eta_k \left[\sum_{j=1}^{q} g_j^2(x^k) + \sum_{j\in J(x^k)} g_j^2(x^k)\right]^{1/2}$$

onde:

$$J(x^k) := \left\{ j\in(q+1,\dots\dots p) / \ \mu_j^k + r_j^k g_j(x^k) \ge 0 \right\}$$

para (ξ_k) y (η_k) sucessões decrescentes de números reais positivos convergentes a zero e $R=\max(r_j^k)$.

CAPÍTULO IV

UM MÉTODO DE DECOMPOSIÇÃO BASEADO EM PENALIZAÇÃO LAGRANGEANA

A maioria dos Métodos Lagrangeanos de decomposição existentes na Programação Não-Linear, requer hipóteses de convexidade sobre o problema original. Por outra parte, na ausência destas últimas, vimos no Capítulo III que o Método de Penalização Lagrangeana de Hestenes-Powell-Rockafellar realiza uma convexificação local do problema, sendo assim possível estender a aplicabilidade destes métodos ao caso dos problemas não-convexos, e eliminando o Salto de Dualidade usualmente neles presente.

Neste capítulo, desenvolvemos um método de decomposição para resolver problemas da Programação Não-Linear de Grande Porte que consiste em dividir a resolução do problema original resolvendo subproblemas de tamanho menor, por meio de uma decomposição primal da função Lagrangeana Aumentada associada.

IV.1.-FORMULAÇÃO DO PROBLEMA.

Seja o problema:

$$\begin{aligned} &\text{Mín} \sum_{i=1}^{N} f_i(x_i) \\ &\text{s. a.} \\ &g_i(x)=0, \quad i=1,\ldots\ldots,p, \\ &g_i(x)\le 0, \quad i=p+1,\ldots\ldots,m, \\ &x^T=(x_1,x_2,\ldots\ldots,x_N) \in C, \end{aligned} \tag{IV.1}$$

onde $C = \prod_{i=1}^{N} C_i$ e $C_i \subseteq \mathbb{R}^{n_i}$ é um conjunto fechado (usualmente $C_i = \mathbb{R}^{n_i}$, ou bem, definido por restrições de cotas nas variáveis); $f_i : \mathbb{R}^{n_i} \longrightarrow \mathbb{R}$, $i=1,\ldots\ldots,N$; $g_i : \mathbb{R}^{\tilde{N}} \longrightarrow \mathbb{R}^{m_i}$,

$M := \sum_{i=1}^{m} m_i$, $\tilde{N} := \sum_{i=1}^{N} n_i$, $m \leq N$.

Além disso, suponhamos que se satisfaz a seguinte hipótese,

Hipótese (H):

a) O problema (IV.1) tem uma solução (local) $\hat{x}$;

b) As funcões f_i, $i=1,\ldots\ldots,N$, g_j, $j=1,\ldots\ldots,m$, são duas vezes localmente continuamente diferenciáveis;

c) A matriz Jacobiana $(\nabla g_i(\hat{x}))_{i \in I(\hat{x})}$ tem posto máximo para

$I(\hat{x}) := \{ 1,\ldots,p \} \cup \{ i \in \{ p+1,\ldots\ldots,m \} : g_i(\hat{x}) = 0 \}$;

d) Satisfaz-se em $\hat{x}$ a condição suficiente de otimalidade local de segunda ordem, sob estrita complementaridade.

No capítulo III, vimos que a solução $\hat{x}$ do problema (IV.1) junto com o multiplicador de Kuhn-Tucker associado define um ponto de sela para a função Lagrangeana Aumentada de Hestenes-Powell-Rockafellar, sempre que se defina um parâmetro de penalização suficientemente grande.

Com o objetivo de aproveitar as vantagens do método e ao mesmo tempo obter separabilidade, define-se o seguinte problema equivalente a (IV.1):

$$\text{Min} \sum_{i=1}^{N} f_i(x_i)$$

s. a.

$g_i(w;x_i) = 0$, $i=1,\ldots.p$,

$g_i(w;x_i) \leq 0$, $i=p+1,\ldots,m$, (IV.2)

$w_i = x_i$, $i=1,\ldots\ldots,N$ (IV.2a)

x, $w \in C$,

onde $g_i(w;x_i)$ representa a função g_i original calculada em$(w_1, w_2, \ldots\ldots, w_{i-1}, x_i, w_{i+1}, \ldots\ldots, w_N)$, $i=1,\ldots.m$.

IV.2.- A DECOMPOSIÇÃO DO PROBLEMA ORIGINAL.

Seja a função de Penalização Lagrangeana para o problema (IV.2) :

$$L(x,w,\bar{y},\bar{r}) := \sum_{i=1}^{N} [f_i(x_i) + y_i^T*(w_i - x_i) + (r_i/2)*\|w_i - x_i\|^2] \qquad \text{(IV.3)}$$

$$+ \sum_{i=1}^{m} \theta(g_i(w;x_i), \tilde{y}_i, \tilde{r}_i),$$

onde $\bar{y} := (y, \tilde{y})$, $\bar{r} := (r, \tilde{r})$,

$\theta(g_i(w;x_i), \tilde{y}_i, \tilde{r}_i)$:

$$=(\tilde{r}_i/2) * g_i^2(w;x_i) + \tilde{y}_i^T * g_i(w;x_i), \qquad \text{(IV.3a)}$$

para $i=1,\ldots\ldots,p$, e

$\theta(g_i(w;x_i), \tilde{y}_i, \tilde{r}_i)$:

$$:= \begin{cases} (\tilde{r}_i/2)*g_i^2(w;x_i) + \tilde{y}_i^T * g_i(w;x_i), & \text{se } g_i(w;x_i) \geq -\tilde{y}_i/\tilde{r}_i, \\ -\tilde{y}_i^2/(2*\tilde{r}_i), & \text{em outro caso,} \end{cases} \qquad \text{(IV.3b)}$$

para $i= p+1,\ldots\ldots,m$.

Em (IV.3a) e (IV.3b) definem-se:

$\tilde{y}_i^2 := \|\tilde{y}_i\|^2$ e $g_i^2(w\ ;x_i) := \|g_i(w;x_i)\|^2$.

Agora, seja:

$$L(x,w,\bar{y},\bar{r}) := \sum_{i=1}^{N} L_i(x_i,w,\bar{y},\bar{r}), \qquad \text{(IV.4)}$$

onde,

$$L_i(x_i,w,\bar{y},\bar{r}) := \begin{cases} f_i(x_i) + y_i^T * (w_i - x_i) + (r_i/2) * \|w_i - x_i\|^2 \\ \quad + \theta(\, g_i(w;x_i), \tilde{y}_i\, \tilde{r}_i), \text{ se } 1 \le i \le m, \\ f_i(x_i) + y_i^T * (w_i - x_i) + \frac{r_i}{2} \|w_i - x_i\|^2, \, i = m+1, \ldots, N. \end{cases}$$

Vemos de (IV.4) que ao fixar as variáveis w, $\bar{y}$, e $\bar{r}$, se obtém separabilidade nas variáveis primais x_i. Assim, o método de decomposição que apresentamos a seguir, está baseado na busca de um ponto de sela da função lagrangeana L definida em (IV.3), por meio da decomposição nas variáveis primais. Portanto propõe-se realizar a minimização primal, resolvendo a cada iteração correspondente à atualização dos multiplicadores, os seguintes dois subproblemas:

$$\begin{aligned} \psi(w,\bar{y},\bar{r}) :&= \operatorname*{Min}_{x \in C} \sum_{i=1}^{N} L_i(x_i,w,\bar{y},\bar{r}) \\ &= \sum_{i=1}^{N} \operatorname*{Min}_{x_i \in C_i} L_i(x_i,w,\bar{y},\bar{r}), \end{aligned} \qquad \text{(IV.5)}$$

e,

$$\operatorname*{Min}_{w \in C} \psi(w,\bar{y},\bar{r}). \qquad \text{(IV.6)}$$

IV.3. -<u>O Algoritmo</u>.

Tal como foi dito na seção anterior, na resolução do problema (IV.2) aplicaremos o Método de Penalização

Lagrangeana e tendo em conta a decomposição nas variáveis primais propõe-se o seguinte algoritmo que por sua forma conceitual será chamado Método de Decomposição Primal-Primal-Dual:

<u>Algoritmo P-P-Dual</u>:

<u>Passo 0</u>. - Sejam $\bar{y}^o, w^o, x^o$, e para $\hat{r}$ suficientemente grande, seja $\bar{r}^o > \hat{r}$, dados; fazer $k=0$.

<u>Passo 1</u>. - Solução de (IV.5)-(IV.6) para x, w:
Dados $\bar{y}=\bar{y}^k$, $\bar{r}=\bar{r}^k$ e $w=w^k$, resolver (IV.5) em x; com $\bar{y}, \bar{r}$ e a solução obtida, resolver (IV.6) em w; o processo é repetido até que seja obtida uma solução de (IV.5)-(IV.6), (solução assintoticamente exata de acordo com os critérios a serem vistos ao final da seção VI.1).
Os dados iniciais são: $w=w^k, \bar{y}=\bar{y}^k, \bar{r}=\bar{r}^k$, e para o processo iterativo da resolução de (IV.5), usar $x=x^k$ como ponto de partida.
A solução de (IV.5)-(IV.6) é denominada w^k, x^k.

<u>Passo 2</u>. - Atualização de $\bar{y}$:

Seja $\bar{y}^{k+1}=(y^{k+1}, \tilde{y}^{k+1})$, então

$$y_i^{k+1}=y_i^k+r_i^k(w_i^k-x_i^k) \quad , \quad i=1,\ldots\ldots\ldots,N,$$

e

$$\tilde{y}_{ij}^{k+1}=\begin{cases} \tilde{y}_{ij}^k+\tilde{r}_i^k g_{ij}(w^k;x_i^k), & i=1,\ldots\ldots\ldots,p, \\ \text{máx}(0, \tilde{y}_{ij}^k+\tilde{r}_i^k g_{ij}(w^k;x_i^k)), & i=p+1,\ldots\ldots\ldots,m, \end{cases}$$

$j=1,\ldots\ldots,m_i$.

<u>Passo 3</u>. - Seja $\bar{r}^{k+1}=(r^{k+1}, \tilde{r}^{k+1})$, então atualizar $\bar{r}^k$ por meio de uma heurística, se necessário, ou fazer $\bar{r}^{k+1}=\bar{r}^k$.

<u>Passo 4</u>.- Se x^k satisfatório, parar. Senão, fazer $k=k+1$; $w^k=w^{k-1}$; $x^k=x^{k-1}$ e voltar a Passo 1.

No próximo Capítulo será provado que a hipótese (H) também é satisfeita pelo problema equivalente (IV.2) e portanto o Método de Penalização Lagrangeana a ele aplicado converge localmente. Também se prova a convergência local do Método de Decomposição.

IV.4 <u>Aspectos teóricos associados e relação com outros métodos</u>.

IV.4.1.- <u>Um resultado teórico fundamental</u>.

Para fazer uma implementação do método de decomposição proposto e ao mesmo tempo poder aplicar um método do tipo gradiente na resolução do subproblema (IV.6) é necessário provar que a função $\psi(\cdot,\bar{y}^k,\bar{r}^k)$ é continuamente diferenciável. O teorema seguinte prova esta condição. Baseamos sua demonstração em uma aplicação direta do resultado de CONTESSE(1986b), que apresentamos como o Corolário 3 no Apêndice.

<u>Teorema 1</u>.-

Sob a hipótese (H) da seção (IV.1), existem $\rho>0$, $\delta>0$, $\hat{r}>0$ tais que a função $\psi(\cdot,y,r)$ definida como:

$$\Psi(w,y,r)= \min_{x\in B(\hat{x},\rho)} L(x,w,y,r),$$

é continuamente diferenciável em relação a w sobre $B(\hat{w},\rho)$, onde $\hat{w}=\hat{x}$ é a solução ótima do problema (IV.1), e $B(\alpha,\rho)$ representa uma vizinhança do ponto α de raio ρ. Além disto, $\nabla_w\psi(w,y,r)=\nabla_w L(\tilde{x},w,y,r)$, onde $\tilde{x}$ minimiza $L(\cdot,w,y,r)$ sobre $B(\hat{x},\rho)$, $\forall y\in B(\hat{y},\delta)$, $\forall r\geq\hat{r}$, onde $\hat{y}$ representa o vetor de multiplicadores ótimos do problema (IV.1).

<u>Demonstração</u>:

Com efeito, da hipótese (H) da seção IV.1, existem $\rho>0$,

$\delta>0$, $\hat{r}>0$ tais que $L(\cdot,\cdot,y,r)$ é estritamente convexa em $B(\hat{w},\rho)\times B(\hat{w},\rho)$, $\forall y\in B(\hat{y},\delta)$, $\forall r\geq\hat{r}$, portanto convexa em $B(\hat{w},\rho)\times B(\hat{w},\rho)$.

Por outra parte $L(\cdot,\cdot,y,r)$ é diferenciável em relação a w em $B(\hat{w},\rho)\times B(\hat{w},\rho)$, $\forall y\in B(\hat{y},\rho)$, $\forall r\geq\hat{r}$ o qual se conlui diretamente de (IV.3) por serem todas as funções $f_i(\cdot)$, $g_i(\cdot,\cdot)$ e $\theta(\cdot,\cdot,\cdot)$ diferenciáveis.

Pela mesma razão anterior, tem-se que $\nabla_w L(\cdot,\cdot,y,r)$ é contínua sobre $B(\hat{w},\rho)\times B(\hat{w},\rho)$,. $\forall y\in B(\hat{y},\delta)$, $\forall r\geq\hat{r}$.

Finalmente $L(\cdot,w,y,r)$ é semicontínua inferior em relação a x (de fato contínua) $\forall w\in B(\hat{w},\rho)$, $\forall y\in B(\hat{y},\delta)$ $\forall r\geq\hat{r}$, porque todas as funções em (IV.3) são contínuas.

Apliquemos agora o Corolário 3 do Apêndice:

$\forall w\in B(\hat{w},\rho)$, a multiaplicação de soluções ótimas,

$$M(w):=\left\{x\in B[\hat{w},\rho]\,/\,L(x,\tilde{w},y,r)=\text{Min } L(\tilde{x},\tilde{w},y,r),\ \tilde{x}\in B[\hat{w},\rho]\right\}$$

$\forall y\in B(\hat{y},\delta)$, $\forall r\geq\hat{r}$,
é um conjunto compacto não vazio, sendo na verdade um singleton (por ser L estritamente convexa em relação a x). Tem-se então que $\psi(\cdot,y,r)$ é continuamente diferenciável em $B(\hat{w},\rho)$, e

$$\nabla_w\Psi(\tilde{w},y,r)=\nabla_w L(\tilde{x},\tilde{w},y,r),\ \forall\ \tilde{w}\in B(\hat{w},\rho),\ \forall y\in B(\hat{y},\delta),\ \forall r\geq\hat{r},$$

onde $\tilde{x}$ é a solução do problema:

$$\underset{x\in B[\hat{w},\rho]}{\text{Min}}\ L(x,\tilde{w},y,r).\ \blacksquare\blacksquare$$

O exemplo a seguir mostra uma aplicação do Teorema 1.

<u>Exemplo 1</u>. -

Considere o problema:

$$P)\quad \begin{aligned} &\text{Min } x^2+z^2 \\ &\text{s.a.} \\ &2x+z-1=0, \end{aligned}$$

cuja solução é: (0.4,0.2).

O problema equivalente associado a considerar para a decomposição é,

$P_w)$ Min x^2+z^2

s.a.

$x-w=0$

$2w+z-1=0.$

A Lagrangeana Aumentada para $P_w)$ é

$$L(x,z,w,y,r)=x^2+z^2+y_1(x-w)+\frac{r_1}{2}(x-w)^2+y_2(2w+z-1)+\frac{r_2}{2}(2w+z-1)^2.$$

Agora, supondo w, y, r fixos, tem-se:

$$\begin{aligned}\psi(w) &:= \min_{x,z} L(x,z,w,y,r)\\ &= \left[\min_x \underbrace{x^2+y_1(x-w)+\frac{r_1}{2}(x-w)^2}_{L_1}\right]+\\ &\qquad +\left[\min_z \underbrace{z^2+y_2(2w+z-1)+\frac{r_2}{2}(2w+z-1)^2}_{L_2}\right]\end{aligned} \tag{IV.7}$$

As soluções para cada um destes subproblemas são, respectivamente :

$$x(w):=\frac{r_1 w-y_1}{2+r_1},$$

$$z(w):=\frac{r_2-y_2-2wr_2}{r_2+2}, \tag{IV.8}$$

Substituindo (IV.8) em (IV.7), tem-se

$$\psi(w)=x^2(w)+z^2(w)+y_1(x(w)-w)+\frac{r_1}{2}(x(w)-w)^2+y_2(2w+z(w)-1)+$$
$$+\frac{r_2}{2}(2w+z(w)-1)^2.$$

A derivação direta de ψ fornece

$$\psi'(w)=2x\,x'(w)+2z\,z'(w)+y_1(x'(w)-1)\quad +r_1(x-w)(x'(w)-1)+$$

$$+y_2(2+z'(w))+r_2(2w+z(w)-1)\;(2+z'(w)); \qquad \text{(IV.9)}$$

enquanto a aplicação do Teorema 1 leva à expressão, já simplificada,

$$\psi'(w)=-y_1-r_1(x(w)-w)+2y_2+2r_2(2w+z(w)-1). \qquad \text{(IV.10)}$$

IV.4.2. Relações com outros métodos.

O método proposto pode ser classificado em um dos tipos vistos na seção II.2. O problema primal:

$$\underset{x,w}{\text{Mín}}\ \underset{y}{\text{Máx}}\ L(x,w,y,r), \qquad \text{(IV.11)}$$

é substituido pelo dual,

$$\underset{y}{\text{Máx}}\ \underset{x,w}{\text{Mín}}\ L(x,w,y,r). \qquad \text{(IV.12}$$

Lembre-se que nosso método resolve (IV.12) por meio de uma decomposição nas variáveis primais, a saber:

$$\underset{y}{\text{Máx}}\ \underset{w}{\text{Mín}}\ \underset{x}{\text{Mín}}\ L(x,w,y,r) \qquad \text{(IV.13)}$$

O Método do Lagrangeano Viável, seção II.3.1, resolve implicitamente

$$\underset{w}{\text{Mín}}\ \underset{y}{\text{Máx}}\ \underset{x}{\text{Mín}}\ L(x,w,y), \qquad \text{(IV.14)}$$

onde L é a Lagrangeana Clássica. Este problema é, usando resultados da Teoria da Dualidade, equivalente a:

$$\underset{w}{\text{Mín}}\ \underset{x}{\text{Mín}}\ \underset{y}{\text{Máx}}\ L(x,w,y), \qquad \text{(IV.15)}$$

o qual também é equivalente a

$$\underset{x,w}{\text{Mín}}\ \underset{y}{\text{Máx}}\ L(x,w,y),$$

Logo, a diferença fundamental entre estes métodos é que, em primeiro lugar o que propusemos pode resolver problemas não convexos por usar a Lagrangeana Aumentada, e em segundo lugar pela forma pelas quais se resolvem os problemas primais (IV.11) e (IV.15) respectivamente.

Pode-se dizer que até agora só tem sido possível provar a continuidade da função:

$$\psi(w):=\underset{y}{\text{Máx}}\ \underset{x}{\text{Mín}}\ L(x,w,y)$$

presente no problema (IV.14). Mais a propriedade de contínua diferenciabilidade dela, está ainda em aberto.

CAPÍTULO V

CONVERGÊNCIA LOCAL DO MÉTODO DE DECOMPOSIÇÃO P-P-DUAL

Neste Capítulo se demonstra a convergência local do método de decomposição P-P-Dual. Primeiramente, que o método de multiplicadores aplicado ao problema (IV.2) converge localmente. A seguir, se demonstra que o método aplicado para resolver os distintos subproblemas (IV.5)-(IV.6) converge localmente. Trata-se de um método do tipo gradiente com projeção sobre um convexo e com passo econômico de Armijo.

Finalmente, a convergência do método de decomposição resulta de maneira natural dos resultados anteriores.

V.1.-Convergência local do método de multiplicadores aplicado ao problema (IV.2).

O seguinte demonstra que a regularidade do problema é herdada do problema (IV.1).

Lema 1.-

Se os gradientes das restrições ativas no ótimo do problema (IV.1) são linearmente independentes, então o mesmo acontece com o problema (IV.2).

Demonstração:

O que faremos é obter a transformação linear (regular) que permite passar de (IV.1) a (IV.2).

Seja $\hat{x}$ o ótimo do problema (IV.1). Sem perda de generalidade podemos supor que todas as restrições são ativas em $\hat{x}$ (apenas para simplicidade de notação).

Deriva disto que a matriz jacobiana

$$D=\begin{bmatrix} D_{x_1}g_1 & D_{x_2}g_1 \cdots\cdots\cdots\cdots & D_{x_N}g_1 \\ D_{x_1}g_2 & D_{x_2}g_2 \cdots\cdots\cdots\cdots & D_{x_N}g_2 \\ & & \\ & & \\ D_{x_1}g_m & D_{x_2}g_m \cdots\cdots\cdots\cdots & D_{x_N}g_m \end{bmatrix}_{M\times\tilde{N}}, \qquad (V.1)$$

tem posto máximo M. Em (V.1)

$D_{x_i}g_j := D_{x_i}g_j(\hat{x})$ representa a matriz jacobiana de g_j em relação a x_i e calculada em $\hat{x}$. E' importante notar que $D_{x_i}g_j$ é definida por:

$$D_{x_i}g_j := \begin{bmatrix} D_{x_i}g_{j1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ D_{x_i}g_{jm_j} \end{bmatrix}, \quad i=1,\ldots,N,\ j=1,\ldots m.$$

Escrevamos agora a matriz jacobiana associada às restrições do problema (IV.2) calculada em $(\hat{x},\hat{x})$:

$$
\tilde{D}=\left[\begin{array}{ccccc|cccc}
D_{x_1}g_1 & 0 & 0\ldots\ldots 0 & 0 & & 0 & D_{x_2}g_1 & D_{x_3}g_1\ldots\ldots & D_{x_N}g_1 \\
0 & D_{x_2}g_2 & 0\ldots\ldots 0 & 0 & & D_{x_1}g_2 & 0 & D_{x_3}g_2\ldots\ldots & D_{x_N}g_2 \\
 & \cdot & & & & & & & \\
 & & \cdot & & & & & & \\
 & & & \cdot & & & & & \\
 & & & & D_{x_N}g_m & D_{x_1}g_m & D_{x_2}g_m\ldots\ldots & \ldots & 0 \\
\hline
 & & I & & & & & -I & \\
\end{array}\right]_{(M+\tilde{N})\times(2\tilde{N})}
$$

onde I representa a matriz identidade de ordem $\tilde{N} \times \tilde{N}$.

Seja agora a matriz inversível

$$
U=\begin{bmatrix} I & 0 \\ I & I \end{bmatrix}, \quad I \text{ de ordem } \tilde{N} \times \tilde{N}.
$$

E' imediato que a matriz

$$
H:=\tilde{D}\cdot U=\left[\begin{array}{c|c}
D & \begin{matrix} 0 & & \\ & 0 & \bar{D}_U \\ \bar{D}_L & \cdot & \\ & \cdot & 0 \end{matrix} \\
\hline
0 & -I
\end{array}\right],
$$

é de posto máximo. Aqui D é a matriz definida em (V.1) e $\bar{D}_L$, $\bar{D}_U$ representam a parte inferior e superior de D, respectivamente.

Assim $\tilde{D}$ é uma matriz de posto máximo, já que se tem:

$$\tilde{D}=HU^{-1}, \text{ com } U^{-1} = \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix},$$

o que termina por demonstrar o Lema. ■ ■

O próximo Lema monstra que o problema (IV.2) satisfaz a condição suficiente de otimalidade de 2^a ordem, desde que (IV.1) a satisfaça.

Lema 2. -

Seja $(\hat{x},\hat{\mu})$ o ponto de Karush-Kuhn-Tucker associado ao problema (IV.1) e

$$L_o(x,\mu):=f(x)+\sum_{i=1}^{m}\mu_i^T g_i(x)$$

a função Lagrangeana clássica respectiva. Se (IV.1), verifica para $(\hat{x},\hat{\mu})$ a condição suficiente de otimalidade de 2^a ordem sob estrita complementaridade, então (IV.2) satisfaz o mesmo para $(\hat{x},\hat{x},\hat{\mu},\hat{\lambda})$, onde $\hat{\lambda}$ é o vetor de multiplicadores ótimos associado às restrições de cópia (IV.2.a).

Demonstração:

Seja

$$\tilde{L}_o(x,w,\mu,\lambda):=f(x)+\sum_{i=1}^{N}\lambda_i^T(x_i-w_i)+\sum_{i=1}^{m}\mu_i^T g_i(w;x_i)$$

a função Lagrangeana clássica associada ao problema (IV.2).

Lembremos que:

$$g_i(w;x_i):=g_i(w_1,w_2,\ldots\ldots\ldots\ldots,w_{i-1},x_i,w_{i+1},\ldots\ldots,w_N),$$

$i=1,2,\ldots\ldots\ldots N.$

A matriz hessiana de $\tilde{L}_o$ relativamente a (x,w), calculada em $(\hat{x},\hat{w},\hat{\mu},\hat{\lambda})$, que se representa por $H(\hat{x},\hat{w},\hat{\mu},\hat{\lambda})$, é dada por

$$H(\hat{x},\hat{w},\hat{\mu},\hat{\lambda})=\begin{bmatrix} \nabla_x^2 f(\hat{x}) & \square \\ \square & \square \end{bmatrix} + \sum_{i=1}^{m} \hat{\mu}_i^T \nabla_{x,w}^2 g_i(\hat{w};\hat{x}_i). \qquad (V.2)$$

Sem perda de generalidade, suponhamos que em $\hat{x}$ todas as restrições de desigualdade são ativas. Temos por um lado as condições de otimalidade de (IV.1):

$$\forall d \in \mathbb{R}^{\tilde{N}},\ d \neq 0 \text{ tal que } d^T \nabla g_{il}(\hat{x})=0,\ i=1,\ldots\ldots,m,\ l=1,\ldots\ldots m_i$$

tem-se

$$d^T \nabla_x^2 L_o(\hat{x},\hat{\mu}) d > 0, \qquad (V.3)$$

onde

$$\nabla_x^2 L_o(\hat{x},\hat{\mu}) = \nabla_x^2 f(\hat{x}) + \sum_{i=1}^{m} \hat{\mu}_i^T \nabla_x^2 g_i(\hat{x}). \qquad (V.4)$$

Por outra parte temos a matriz $\nabla^2_{x,v} g_{il}(\hat{w};\hat{x}_i)$:

$$
:=\left[\begin{array}{c|cc|c|cc}
\begin{matrix} 0 & & \square \\ & \ddots \; \dfrac{\partial^2 g_{il}}{\partial x_i^2} \; \ddots & \\ \square & & 0 \end{matrix} & \dfrac{\partial^2 g_{il}}{\partial v_1 \partial x_i} & \dfrac{\partial^2 g_{il}}{\partial v_2 \partial x_i} \cdots & 0 & \cdots \dfrac{\partial^2 g_{il}}{\partial v_{N-1} \partial x_i} & \dfrac{\partial^2 g_{il}}{\partial v_N \partial x_i} \\
\hline
\dfrac{\partial^2 g_{il}}{\partial x_i \partial v_1} & \dfrac{\partial^2 g_{il}}{\partial v_1^2} & \dfrac{\partial^2 g_{il}}{\partial v_2 \partial v_1} \cdots & 0 & \cdots\cdots & \dfrac{\partial^2 g_{il}}{\partial v_N \partial v_1} \\
\dfrac{\partial^2 g_{il}}{\partial x_i \partial v_2} & \dfrac{\partial^2 g_{il}}{\partial v_1 \partial v_2} & \dfrac{\partial^2 g_{il}}{\partial v_2^2} \cdots & 0 & \cdots\cdots & \dfrac{\partial^2 g_{il}}{\partial v_N \partial v_2} \\
\vdots & & & & & \vdots \\
\hline
\square & 0 & 0 & 0 & & 0 \\
\hline
\vdots & \vdots & \vdots & 0 & & \vdots \\
\square \;\; \dfrac{\partial^2 g_{il}}{\partial x_i \partial v_N} \;\; \square & \dfrac{\partial^2 g_{il}}{\partial v_1 \partial v_N} & \dfrac{\partial^2 g_{il}}{\partial v_2 \partial v_N} \cdots & 0 & \cdots\cdots & \dfrac{\partial^2 g_{il}}{\partial v_N^2}
\end{array}\right]_{2\tilde{N}\times 2\tilde{N}}
$$

$$l=1,\ldots\ldots\ldots\ldots,m_i$$

(V.5)

onde todos os elementos são calculados em $(\hat{w},\hat{x}_i)$.

Seja agora, $(h,v)\in\mathbb{R}^{2\tilde{N}}$, $(h,v)\neq 0$, pertencente ao hiperplano tangente em $(\hat{x},\hat{w})$ definido pelas restrições (IV.2.a) do problema (IV.2), isto é,

$$(h,v)\begin{bmatrix} e_i \\ -e_i \end{bmatrix}=0, \quad i=1,\ldots\ldots\ldots N,$$

onde e_i representa o i-ésimo vetor da base padrão de $\mathbb{R}^{\tilde{N}}$. E' evidente que:

$h_i=v_i$, $i=1,\ldots\ldots,N$, ou seja $h=v$.

Portanto, temos que demonstrar que para d definido por $d:=h=v$, se tem:

$$(d^T,d^T)H(\hat{x},\hat{w},\hat{\mu},\hat{\lambda})\begin{bmatrix} d \\ d \end{bmatrix}>0,\ \forall d\neq 0 \text{ tal que,}$$

$$(d^T,d^T)\ \nabla g_i(\hat{w};\hat{x}_i)=0,\ i=1,\ldots\ldots\ldots,m.$$

Com efeito, seja $d\in\tilde{\mathbb{R}}^N$, $d\neq 0$. Por um lado temos que

$$(d^T,d^T)\begin{bmatrix} \nabla^2_x f(\hat{x}) & \square \\ \square & \square \end{bmatrix}\begin{bmatrix} d \\ d \end{bmatrix}=d^T\nabla^2_x f(\hat{x})d. \qquad (V.6)$$

Por outro lado $\forall i=1,\ldots\ldots\ldots,m$,

$$(d^T,d^T)\ \nabla^2_{x,v}g_{il}(\hat{w};\hat{x}_i)\begin{bmatrix} d \\ d \end{bmatrix}=d^T\nabla^2_x g_{il}(\hat{x}_i)d,\ l=1,\ldots\ldots,m_i, \qquad (V.7)$$

o que demonstramos a seguir.

De (V.5) é fácil ver que:

$$\nabla^2_{x,v}g_{il}(\hat{w};\hat{x}_i)\begin{bmatrix} d \\ d \end{bmatrix}:=(y_1,y_2,\ldots\ldots\ldots,y_N,y_{N+1},\ldots,y_{2N})^T, \quad (V.8)$$

onde

$$y_j=\begin{cases} 0,\ j=1,\ldots,N\ ,\ j\neq i\ ,\ j=N+1 \\[2ex] \displaystyle\sum_{\substack{k=1\\k\neq i}}\frac{\partial^2 g_{il}}{\partial v_k\partial x_j}d_k+\frac{\partial^2 g_{il}}{\partial x_j^2}d_i,\ j=i \\[2ex] \displaystyle\sum_{\substack{k=1\\k\neq i}}\frac{\partial^2 g_{il}}{\partial x_k\partial x_{j-N}}d_k+\frac{\partial^2 g_{il}}{\partial x_i\partial w_{j-N}}d_i,\ j=N+1,\ldots,2N,\ j\neq N+i, \\ \qquad l=1,\ldots\ldots,m_i \end{cases}$$

Agora usando o fato de que todas as derivadas (parciais) estão calculadas no ponto $(\hat{x},\hat{w})$ e como $\hat{w}=\hat{x}$, tem-se então que:

$$y_j=\begin{cases} 0,\ j=1,\ldots,N\ ,\quad j\neq i\ ,\ j=N+i \\ \displaystyle\sum_{k=1}^{N} \frac{\partial^2 g_{il}}{\partial x_k \partial x_j}\, d_k\ ,\ j=i \\ \displaystyle\sum_{k=1}^{N} \frac{\partial^2 g_{il}}{\partial x_k \partial x_{j-N}}\, d_k\ ,\ j=N+1,\ldots,2N,\ j\neq N+i. \end{cases} \tag{V.9}$$

Além disso, por (V.8) temos que:

$$(d^T,d^T)\nabla^2_{x,w} g_{il}(\hat{w};\hat{x}_i)\begin{bmatrix} d \\ d \end{bmatrix}=(d^T,d^T)y.$$

Logo, usando (V.9), obtém-se:

$$(d^T,d^T)y \quad =\sum_{k=1}^{N} d_i \frac{\partial^2 g_{il}}{\partial x_k \partial x_i}\, d_k+\sum_{\substack{r=1\\ r\neq i}}^{N} d_r \sum_{k=1}^{N}\frac{\partial^2 g_{il}}{\partial x_k \partial x_r}\, d_k$$

$$=\sum_{k=1}^{N}\sum_{r=1}^{N} \frac{\partial^2 g_{il}}{\partial x_k \partial x_r}\, d_k d_r$$

$$=d^T\ \nabla^2_x g_{il}(\hat{x}_i)d,\ \ i=1,\ldots,m;\ \ l=1,\ldots,m_i,$$

o que demonstra (V.7).

De (V.6) e (V.7), é imediato que:

$$(d^T,d^T)\ H(\hat{x},\hat{w},\hat{\mu},\hat{\lambda})\begin{bmatrix} d \\ d \end{bmatrix}=d^T\nabla^2_x L_0(\hat{x},\hat{\mu})d,$$

$\forall d\in\mathbb{R}^{\tilde{N}}$, $d\neq 0$, e

como por hipótese tínhamos a condição de otimalidade

(V.3), tem-se então

$$(d^T,d^T)\ H(\hat{x},\hat{w},\hat{\mu},\hat{\lambda})\begin{bmatrix} d \\ d \end{bmatrix}>0,\ \forall d\in\mathbb{R}^N, d\neq 0,\ \text{tal que}$$

$$(d^T,d^T)\nabla g_{i\,l}(\hat{w};x_i)=0,\ i=1,2,\ldots\ldots m,\ l=1,\ldots,m_i\ \text{já que}\ \hat{w}=\hat{x},$$

o que termina a demonstração. ■ ■

Sabemos que sob a hipótese (H) da seção (IV.1), o método de penalização lagrangeana ou de multiplicadores converge, logo os Lemas 1 e 2, implicam diretamente no seguinte teorema:

<u>Teorema 1</u>.

Suponhamos que se satisfaz a hipótese (H) da seção (IV.1) incluída a condição de estrita complementaridade para o problema (IV.1). Então o método de penalização lagrangeana aplicado ao problema (IV.2) converge localmente.

V.2.-<u>Convergência local do método aplicado para resolver os distintos subproblemas (IV.5)-(IV.6).</u>

Para resolver cada um dos subproblemas (IV.5)-(IV.6) aplicaremos um método do tipo gradiente projetado para a minimização de uma função continuamente diferenciável sobre um domínio convexo compacto.

Consideremos assim o seguinte problema de minimização restrita:

$$\mathcal{P})\ \min_{x\in C\subset\mathbb{R}^n} f(x)$$

onde $f(\cdot)$ é uma função continuamente diferenciável (eventualmente convexa) e C, um domínio convexo compacto de $\mathbb{R}^n$.

A seguir, estudamos o seguinte algoritmo de gradiente projetado para a determinação de um ponto estacionário restrito de $\mathcal{P}$):

<u>Algoritmo</u>:

Passo 0: Sejam $\varepsilon \in]0,1/2[$, $\alpha>0$ dados.
Inicialmente: $x^o \in C$, ponto inicial viável dado, $k=0$.

Passo 1: Definir $d_k:=d(x^k):=z(x^k)-x^k$,
com, $z(x^k):=P_C(x^k-\alpha\nabla f(x^k))$, onde $P_C(\cdot)$ representa a projeção sobre C.
-Se $d_k=0$, fim: x^k é ponto estacionário de $\mathcal{P}$).
-Senão, determinar $1\geq t_k>0$, tal que

$$f(x^k+t_kd_k)\leq f(x^k)+\varepsilon t_k\nabla f(x^k)\cdot d_k.$$

Ir a passo 2.

Passo 2 Definir $x^{k+1}:=x^k+t_kd_k$, fazer $k=k+1$ e voltar a Passo 1.

O seguinte teorema demonstra a convergência local do algoritmo anterior.

<u>Teorema 2</u>.

O algoritmo, ou bem produz um ponto estacionário após um número finito de iterações, ou bem gera uma sucessão infinita de pontos $\{x^k\}$, a qual admite ao menos um ponto de acumulação, e tal que todo ponto de acumulação é um ponto estacionário de $\mathcal{P}$).

<u>Demonstração</u>: (conforme CONTESSE (1990b))

Seja $x\in C$. Seja $z(x)=P_C(x-\alpha\nabla f(x))$; a condição de otimalidade da projeção de $x-\alpha\nabla f(x)$ sobre C é dada por:

$$(y-z(x))\cdot([x-\alpha\nabla f(x)]-z(x))\leq 0,\ \forall y\in C. \tag{V.10}$$

Logo se a direção d(x) for nula, valerá a propriedade:

$$d(x)=z(x)-x=0 \Longleftrightarrow \nabla f(x)\cdot(y-x)\geq 0,\ \forall y\in C, \tag{V.11}$$

que é a condição necessária de otimalidade para $\mathcal{P}$).

Assim, se $d_{\bar{k}}=0$, para um certo $\bar{k}$, $x^{\bar{k}}$ é um ponto estacionário restrito de $\mathcal{P}$).

Por outro lado, da desigualdade (V.10) tem-se:

$\|d(x)\|^2+\alpha\nabla f(x)d(x)\leq 0$, onde $d(x):=z(x)-x$. Logo, como $d(x)\neq 0$ para um $x\in C$ que não seja solução, temos que:

$$\nabla f(x)\cdot d(x)<0. \tag{V.12}$$

Assim, se $d_k\neq 0$, existe $t_k>0$, tal que:

$$f(x^k+t_kd_k)-f(x^k)\leq \varepsilon t_k\nabla f(x^k)\cdot d_k<0, \tag{V.13}$$

$$\text{e, } (x^k+t_kd_k)\in C, \tag{V.14}$$

já que $t_k\in\,]0,1[$.

Suponhamos agora que se gera uma sucessão infinita $(x^k)_{k\in\mathbb{N}}$. Como C é compacto, esta admite pelo menos um ponto de acumulação em C. Seja x^* um qualquer destes pontos. Existe portanto $\mathbb{N}'\subset\mathbb{N}$ tal que:

$$\left\{x^k\right\}_{k\in\mathbb{N}'}\longrightarrow x^* \tag{V.15}$$

Como a sucessão $\left\{x^{k+1}\right\}_{k\in\mathbb{N}'}\subset C$, ela também admite pelo menos um ponto de acumulação, digamos $x^{**}\in C$.

Analogamente existe $\mathbb{N}''\subset\mathbb{N}'$ tal que,

$$\left\{x^{k+1}\right\}_{k\in\mathbb{N}''}\longrightarrow x^{**}. \tag{V.16}$$

Observemos agora que por ser f contínua definida no compacto C, a propriedade de descida (V.13), $f(x^{k+1})<f(x^k)$,

$\forall k\in\mathbb{N}$, nos permite concluir que,

$$\left\{f(x^k)\right\}_{k\in\mathbb{N}} \text{ converge em } \bar{\mathbb{R}}. \tag{V.17}$$

De novo, por continuidade temos

$$f(x^*) = \lim_{\substack{k\to\infty \\ k\in\mathbb{N}''}} f(x^k) = \lim_{\substack{k\to\infty \\ k\in\mathbb{N}''}} f(x^{k+1})\in\mathbb{R}. \tag{V.18}$$

Provaremos que x^* é um ponto estacionário, partindo, por absurdo, da hipótese contrária, isto é, de acordo com (V.11):

$$d(x^*)=z(x^*)-x^*\neq 0. \tag{V.19}$$

De (V.13) obtemos diretamente:

$$f(x^{k+1})-f(x^k)\leq \varepsilon t_k \nabla f(x^k)\cdot d(x^k), \quad \forall k\in\mathbb{N}''.$$

Passando ao limite quando $k \xrightarrow{\mathbb{N}''} \infty$,levando em conta a continuidade de $P_c(\cdot)$, de $f(\cdot)$ e de $\nabla f(\cdot)$, e utilizando (V.15), (V.16), (V.18) e (V.19) chega-se a

$$0=f(x^{**})-f(x^*)\leq \varepsilon t^* \nabla f(x^*)\cdot d(x^*)<0, \text{ onde}$$

$$t^*=\frac{\|x^{**}-x^*\|}{\|d(x^*)\|}$$

o que é uma contradição. Portanto x^* é um ponto estacionário restrito de $\mathcal{P}$).

<u>Corolário 1</u>.-

Se $f(\cdot)$, além de ser continuamente diferenciável é convexa, ou bem o algoritmo produz uma solução ótima para $\mathcal{P}$) após um número finito de iterações, ou bem todo ponto de acumulação é uma solução ótima de $\mathcal{P}$).

<u>Demonstração</u>: Imediata.

Corolário 2.-

Se $f(\cdot)$, além de ser continuamente diferenciável é estritamente convexa, ou bem o algoritmo produz a única solução ótima de $\mathcal{P}$) após um número finito de iterações, ou bem gera uma sucessão convergente cujo limite é a solução ótima de $\mathcal{P}$).

Demonstração: Imediata.

Observação 1:

Realizamos uma adaptação do método do Gradiente Projetado utilizando um passo de minimização de Armijo na direção viável d(x). Porém, o resultado de convergência é válido para qualquer dos passos de minimização usuais na literatura (Wolfe, Goldstein, minimização exata, etc.).

Aplicação ao caso de Restrições de cotas sobre as variáveis.

O algoritmo descrito anteriormente é particularmente atrativo no caso de restrições de cotas sobre as variáveis.

Neste caso a direção de descida viável se obtém de maneira muito econômica.

Com efeito, suponhamos que C é definido por

$$C:=\left\{x\in\mathbb{R}^n / \ L\leq x\leq U\right\}.$$

Então se $d^k=-\nabla f(x^k)$, z^k fica definido por:

$$z_i^k=\begin{cases} L_i, & \text{si } x_i^k+\alpha d_i^k\leq L_i \\ U_i, & \text{si } x_i^k+\alpha d_i^k\geq U_i \\ x_i^k+\alpha d_i^k & \text{si } L_i<x_i^k+\alpha d_i^k<U_i. \end{cases}$$

Os teoremas 1 e 2 demonstram de forma imediata o seguinte teorema de convergência local do método de decomposição P-P-Dual.

Teorema 3.

Sob as condições dos teoremas 1 e 2 o método de Decomposição P-P-Dual converge localmente.

Demonstração:

Sejam $\bar{y}^o$, $\bar{r}^o$, os vetores de multiplicadores e de penalidades respectivos iniciais, escolhidos adequadamente. Na iteração k, (x^k, w^k) é a solução do subproblema Lagrangeano (IV.5)-(IV.6), pelo Teorema 2. Logo, pelo Teorema 1, tem-se que:

$$(x^k, w^k, \bar{y}^k, \bar{r}^k) \xrightarrow[k \to \infty]{} (\hat{x}, \hat{x}, \hat{y}, \hat{r}),$$

ponto de sela da Lagrangeana Aumentada do problema (IV.2), o que termina a demonstração. ■ ■

CAPÍTULO VI

EXPERIÊNCIA COMPUTACIONAL

VI.1-ALGUNS ASPECTOS ACERCA DA IMPLEMENTAÇÃO COMPUTACIONAL.

Pela forma do problema real de Grande Porte que foi resolvido usando o método de decomposição proposto, é que consideramos o seguinte problema onde separamos em (IV.1) as restrições de desigualdade lineares (separáveis em relação a todas suas variáveis) das demais:

$$\text{Min} \sum_{i=1}^{N} f_i(x_i) := f(x)$$

s.a.

$$\begin{aligned} & g_i(x)=0, \quad i=1,\ldots\ldots\ldots,p \\ & g_i(x)\leq 0, \quad i=p+1,\ldots\ldots\ldots,m \qquad\qquad (\text{VI.1}) \\ & A_i x_i \geq b_i, \quad i=1,\ldots\ldots\ldots,N \end{aligned}$$

$I_i \leq x_i \leq S_i$, $i=1,\ldots\ldots,N$, onde A_i é uma matriz e b_i um vetor de dimensões adequadas.

Também supomos que este problema satisfaz a hipótese (H) da seção (IV.1). Agora, o problema equivalente a (VI.1) análogo a (IV.2) é:

$$\text{Min} \sum_{i=1}^{N} f_i(x_i) := f(x)$$

s.a.

$$\begin{aligned} & g_i(w;x_i)=0, \quad i=1,\ldots\ldots\ldots,p, \\ & g_i(w;x_i)\leq 0, \quad i=p+1,\ldots\ldots\ldots,m \qquad\qquad (\text{VI.2}) \\ & w_i=x_i, \quad i=1,\ldots\ldots\ldots,N, \qquad\qquad (\text{VI.2a}) \\ & A_i x_i \geq b_i, \; I_i \leq x_i \leq S_i, \; I_i \leq w_i \leq S_i, \; i=1,\ldots\ldots\ldots,N. \end{aligned}$$

É importante notar que o parâmetro w não se inclui nas restrições de desigualdade lineares por serem completamente separáveis.

Nosso método resolve o problema (VI.2) por Penalização Lagrangeana. Logo, em cada iteração correspondente à atualização dos multiplicadores das restrições não lineares, resolve-se o problema:

$$\begin{aligned} &\text{Mín } L(x,w,\bar{y},\bar{r}) \\ &\text{s.a.} \\ &A_i x_i \geq b_i \\ &I_i \leq x_i \leq S_i,\ I_i \leq w_i \leq S_i,\ i=1,\ldots\ldots\ldots,N, \end{aligned} \qquad \text{(VI.3)}$$

onde L representa a função Lagrangeana Aumentada associada às restrições não lineares do problema (VI.2) e de igualdade (VI.2a). Mas o problema (VI.3) é resolvido por decomposição fixando o vetor w. Portanto tem-se que resolver:

$$\begin{aligned} &\text{Mín } \psi(w,\bar{y},\bar{r}) \\ &I_i \leq w_i \leq S_i,\ i=1,\ldots\ldots,N \end{aligned} \qquad \text{(VI.4)}$$

onde

$$\psi(w,\bar{y},\bar{r}) = \sum_{i=1}^{N} \begin{aligned} &\text{Mín } L_i(x_i,w,\bar{y},\bar{r}) \\ &\text{s.a.} \\ &A_i x_i \geq b_i \\ &I_i \leq x_i \leq S_i \end{aligned} \qquad \text{(VI.5)}$$

As iterações que são feitas na resolução dos distintos subproblemas do tipo (VI.4) são chamadas iterações menores, e as iterações feitas na atualização dos multiplicadores serão chamadas iterações maiores.

Os distintos subproblemas que aparecem em (VI.5) foram resolvidos pelo Código PENAMOR, CONTESSE e VILLAVICENCIO(1990). Este código resolve problemas da forma:

$$
\begin{aligned}
&\text{Min } f(x)\\
&g_j(x)=0, \quad j=1,\ldots\ldots\ldots,mne,\\
&g_j(x)\le 0, \quad j=mne+1,\ldots\ldots,mn,\\
&a_i^T x=b_i, \quad i=1,\ldots\ldots\ldots\ldots,mle,\\
&a_i^T x\ge b_i, \quad i=mle+1,\ldots\ldots\ldots,ml\\
&l_k\le x_k\le \mu_k, \quad k=1,\ldots\ldots\ldots\ldots,n,
\end{aligned}
$$

onde pode-se ter mne ou mn assim como mle ou ml sejam zero. Também pode ocorrer que existam variáveis não limitadas, isto é, que $l_k=-\infty$ ou $\mu_k=+\infty$.

O método usado neste código é o Método de Penalização Lagrangeana Amortecida, CONTESSE(1985,1986a,1987a), CONTESSE e VILLAVICENCIO(1987,1988,1989). Sua primeira implementação foi feita por VILLAVICENCIO(1987). Os diferentes subproblemas Lagrangeanos no método de Penalização Lagrangeana Amortecida são resolvidos por uma adaptação do Método B.F.G.S. ao caso de restrições lineares, por meio de ativação de restrições.

Os distintos subproblemas do tipo (VI.4) foram resolvidos pelo algoritmo de gradiente projetado proposto na seção V.2. utilizando uma busca linear baseada no passo de Armijo.

O controle das iterações maiores é feito pelo programa principal. Este programa inicializa todas as variáveis necessárias , controla a solução do subproblema (VI.3), atualiza as aproximações do vetor de multiplicadores $\bar{y}^k$ e o vetor de parâmetros $\bar{r}^k$.

O programa considera uma solução (x^k,w^k) como ótima quando se satisfazem alguns dos seguintes critérios. O primeiro requer que se satifaçam as condições de otimalidade de Kunh-Tucker para o problema (VI.2):

$$\|\nabla f(x^k)+\nabla_{x,w} g(w^k;x_i^k)\tilde{y}^k+By^k-Dz^k\|\le 10^{-\sigma},$$

$$\max\left\{|g_i(w^k;x_i^k)|,\ i=1,.........p\right\}\leq 10^{-6},$$

$$\max\left\{g_i(w^k;x_i^k),\ i=p+1,...........m\right\}\leq 10^{-6},$$
$$z_i^k\geq -10^{-6},$$

para todo $i \in AL$, onde AL representa o conjunto de índices das restrições lineares de desigualdades que são ativas em (w^k,x^k), D representa a matriz cujas colunas são as correspondentes normais às restrições lineares ativas, e B representa a matriz cujas colunas são as normais às restrições (VI.2a).

O segundo critério requer que x^{k+1}, w^{k+1} e $\bar{y}^{k+1}$ não sejam muito diferentes de x^k, w^k e $\bar{y}^k$ respectivamente, o que é feito através do teste:

$$\|x^{k+1}-x^k\|+\|w^{k+1}-w^k\|+\|\bar{y}^{k+1}-\bar{y}^k\|\leq 10^{-5}\Big(\|x^{k+1}\|+\|x^k\|+\|w^{k+1}\|+\|w^k\|+$$
$$+\|\bar{y}^{k+1}\|+\|\bar{y}^k\|\Big).$$

Seguindo BERTSEKAS(1976a), e POWELL(1969), $\bar{r}^{k+1}$ se define a partir de $\bar{r}^k$ por meio das seguintes regras:

Se $|\bar{y}_j^{k+1}-\bar{y}_j^k|>cof*|\bar{y}_j^k-\bar{y}_j^{k-1}|$ e

$|\bar{y}_j^k-\bar{y}_j^{k-1}|>cof*|\bar{y}_j^{k-1}-\bar{y}_j^{k-2}|$,

então

$$\bar{r}_j^{k+1}=\min\left\{PMF * \bar{r}_j^k, RCOTA\right\}$$

de outro modo,

$$\bar{r}_j^{k+1}=\bar{r}_j^k .$$

Também, se se define $RE(w^k;x_i^k)$ como o vetor das restrições não lineares e lineares (VI.2a) do problema (VI.2), $\bar{r}^{k+1}$ é modificado por meio de:

Se $\|RE(w^{k+1};x_i^{k+1})\|>cof*\|RE(w^k;x_i^k)\|$ e

$$\|RE(w^k;x_i^k)\|>cof*\|RE(w^{k-1};x_i^{k-1})\|,$$

então,

$$\bar{r}_j^{k+1}=\min\left\{PMF*\bar{r}_j^k,\ RCOTA\right\},$$

de outro modo,

$$\bar{r}_j^{k+1}=\bar{r}_j^k.$$

Em ambas as regras anteriores se usa PMF=2.0 , cof=0.75 e RCOTA é uma cota para o vetor de parâmetros.

Por outra parte, o ponto inicial (x_o^{k+1},w_o^{k+1}) para resolver os distintos subproblemas do tipo (VI.3) na iteração k+1 foi considerado sempre o ótimo $(\hat{x}^k,\hat{v}^k)$ (assintoticamente exato, como BERTSEKAS(1975,1976a,1976b), e como vimos também, no capítulo III) associado ao respectivo subproblema na iteração k. Também seguindo Bertsekas, os multiplicadores se mantiveram limitados. É importante notar que estes critérios devem usar-se a fim de garantir convergência ao nível de coordenação dos multiplicadores.

Quanto ao controle das iterações menores, este se realiza por meio de três critérios.

O primeiro tem a ver com a convergência do algoritmo do gradiente projetado proposto na seção V.2.:

Se $d_k=0$, fim da minimização do problema (VI.4).

O segundo critério está baseado em que não se precisa resolver exatamente os distintos subproblemas do tipo (VI.3), isto é, só é necessário resolvê-los em forma assintoticamente exata BERTSEKAS(1975,1976a,1976b). Isto torna mais eficiente este ponto. Utilizamos um critério proposto por CONTESSE(1990a), o qual é análogo aos propostos por Bertsekas, e válido na ausência da hipótese de estrita complementaridade:

$$\|\nabla_{x,v} L(\cdot)-Dz\| \le \text{GAM}^k/R$$

e,

$$\|\nabla_{x,v} L(x^k,w^k,\bar{y}^k,\bar{r}^k)-Dz\| \le \text{GAM}^k *$$

$$* \sqrt{\sum_{j=1}^{p} g_j^2(w^k;x_j)+\sum_{i=1}^{N} h_i^2(w^k;x_i^k)+\sum_{j\in J(x^k,v^k)} g_j^2(w^k;x_j^k)}$$

onde,

$$h_i(w^k;x_i^k):=w_i^k-x_i^k \quad ,i=1,\ldots\ldots\ldots\ldots N, \text{ e}$$

$$J(x^k,w^k)=\left\{j\in(p+1,\ldots\ldots\ldots,m)/\ \tilde{y}_j+\tilde{r}_j g_j(w^k;x_j^k)\ge 0\right\}, \text{ e}$$

$z_i \ge 0$, $\forall i\in AL$, conjunto de índices definido anteriormente,

$\text{GAM}^{k+1}=\text{COFGAM} * \text{GAM}^k$, $\text{GAM}^0=0.5$,

$\text{COFAM}=0.75$ e $R=\text{máx}(\bar{r}_j)$.

Nas duas desigualdades anteriores a primeira expressão $\|\cdot\|$ equivale a:

$$\|\nabla_x L(x^k,w^k,\bar{y}^k,\bar{r}^k)-Dz\|+\sum_{I_i<v_i^k<S_i}\left|\frac{\partial L}{\partial w_i}\right|+\sum_{v_i^k=I_i}\left[-\frac{\partial L}{\partial v_i}\right]_+ +\sum_{v_i^k=S_i}\left[\frac{\partial L}{\partial w_i}\right]_+ ,$$

onde,

$[\alpha]_+ := \text{Máx }(0,\alpha)$ e $|\cdot|$ representa o valor absoluto.

O último somando da expressão anterior deve-se ao critério de Kuhn-Tucker do problema (IV.2) em relação às cotas deste.

O terceiro critério está baseado em parar as iterações quando a diferença entre dois iterados é muito pequena:

$$\|w^{k+1}-w^k\|+\|x^{k+1}-x^k\|\leq \text{TOL}*(\|w^{k+1}\|+\|w^k\|+\|x^{k+1}\|+\|x^k\|),$$

com $\text{TOL}=10^{-5}$.

VI.2 Experiências Numéricas

Nesta seção se apresentam os resultados obtidos ao aplicar o método de decomposição proposto. Na primeira parte mostramos as experiências sobre quatro problemas de prova conhecidos na literatura.

Estes primeiros testes numéricos, assim como o algoritmo P-P-Dual com alguns aspectos teóricos relacionados, foram apresentados pelo autor em CONTESSE, OLIVEIRA e PAREDES(1990).

Na segunda parte mostramos os resultados obtidos sobre um problema real de Despacho Econômico de Energia Elétrica no Sistema Interconetado Chileno. Este é um problema de Grande Porte que foi resolvido pelos métodos diretos PENAMOR e MINOS/AUGMENTED, MURTAGH e SAUNDERS(1982); seus últimos resultados foram reportados por CONTESSE e VILLAVICENCIO(1990).

VI.2.1.-Quatro problemas testes da literatura

VI.2.1.1.-Problema Ad-hoc.

Este é um problema quadrático em duas variáveis, onde a função objetivo é côncava e o conjunto viável é convexo. Foi utilizado por CONTESSE e VILLAVICENCIO(1987), para provar o comportamento do Algoritmo de Penalização Amortecida. Este é um bom problema de prova, já que tem quatro pontos estacionários de Karush-Kuhn-Tucker, os quais podem ser pontos de atração para algoritmos. Porém tem só um ponto mínimo local, que é de fato global: é dado por $(x_1,x_2)=(1,-1)$, cujo multiplicador ótimo associado é

$(u_1,u_2,u_3)=(0,5/4,1)$. Os restantes pontos estacionários são: $(x_1,x_2)=(0,0)$, com multiplicador $(u_1,u_2,u_3)=(0,0,5/4)$; $(x_1,x_2)=(0,-1)$, com multiplicador $(u_1,u_2,u_3)=(0,1/4,0)$; e $(x_1,x_2)=(0,-3/4)$, com multiplicador $(u_1,u_2,u_3)=(0,0,0)$.

O problema é:

$$\text{Min } f(x_1,x_2):=-x_1^2-0.5x_2^2-0.75x_2$$
s.a.
$$-x_1\le 0$$
$$-x_2-1\le 0$$
$$x_1^2+x_2\le 0.$$
Ponto inicial: (1,0).

O problema equivalente para aplicação do método de decomposição é:

$$\text{Min } f(x)=-x_1^2-0.5x_2^2-0.75x_2$$
s.a.
$$-x_1\le 0$$
$$-x_2-1\le 0$$
$$g(w;x_1):=x_1^2+w\le 0$$
$$w-x_2=0$$
$$-w-1\le 0.$$

VI.2.1.2.-<u>Problema de Biggs.</u>

E' um problema com função objetivo não-linear de três variáveis. Tem uma restrição linear limitada inferior e superiormente. O mesmo acontece com suas variáveis. Encontra-se publicado em BUYS(1972), e em HOCK e SCHITTKOWSKI(1981, problema N^o 36). O problema é:

$$\text{Min } f(x_1,x_2,x_3)=-x_1x_2x_3$$
s.a.
$$0\le x_1+2x_2+2x_3\le 72$$
$$0\le x_1\le 20$$

$0 \le x_2 \le 11$
$0 \le x_3 \le 42$

Ponto inicial:(10,10,10).
Soluçao:(20,11,15)

O problema equivalente é:

Mín x_4
s.a.
$w_1 - x_1 = 0$
$w_2 - x_2 = 0$
$w_3 - x_3 = 0$
$g_1(w;x_1) := -x_1 - 2w_2 - 2w_3 \le 0$
$g_2(w;x_2) := w_1 + 2x_2 + 2w_3 - 72 \le 0$
$g_3(w;x_3,x_4) := -w_1 w_2 x_3 - x_4 \le 0$
$0 \le x_1, w_1 \le 20$
$0 \le x_2, w_2 \le 11$
$0 \le x_3, w_3 \le 42.$

VI.2.1.3.-<u>Problema de Paviani</u>.

Este se encontra publicado em HIMMELBLAU(1972), e em HOCK e SCHITTKOWSKI(1981, problema N^o 63).

O problema é:

Mín $f(x_1,x_2,x_3) = 1000 - x_1^2 - 2x_2^2 - x_3^2 - x_1x_2 - x_1x_3$
s.a.
$x_1^2 + x_2^2 + x_3^2 - 25 = 0$
$8x_1 + 14x_2 + 7x_3 - 56 = 0$
$x_1, x_2, x_3 \ge 0$

Ponto inicial:(7,0,0).
Solução:(3.51485,0.21694,3.54901)

O problema equivalente é:

Min x_4

s.a.

$w_1-x_1=0$

$w_2-x_2=0$

$w_3-x_3=0$

$w_4-x_4=0$

$g_1(w;x_1):=x_1^2+w_2^2+w_3^2-25=0$

$g_2(w;x_2,x_3):=8w_1+14x_2+7x_3-56=0$

$g_3(w;x_4):=1000-w_1^2-2w_2^2-w_3^2-w_1(w_2+w_3)-x_4\leq 0$

$x,w\geq 0.$

VI.2.1.4.-<u>Problema de Rosen-Suzuki</u>.

Este problema se encontra publicado em HOCK e SCHITTKOWSKI(1981, problema N°43).

O problema é:

Min $f(x_1,x_2,x_3,x_4)=x_1^2+x_2^2+2x_3^2+x_4^2-5x_1-5x_2-21x_3+7x_4$

s.a.

$2x_1^2+x_2^2+x_3^2+2x_1-x_2-x_4-5\leq 0$

$x_1^2+x_2^2+x_3^2+x_4^2+x_1-x_2+x_3-x_4-8\leq 0$

$x_1^2+2x_2^2+x_3^2+2x_4^2-x_1-x_4-10\leq 0.$

Ponto inicial: (0,0,0,0)

Solução: (0,1,2,-1)

O problema equivalente é:

Min $f(x_1,x_2,x_3,x_4)=x_1^2+x_2^2+2x_3^2+x_4^2-5x_1-5x_2-21x_3+7x_4$

s.a.

$w_1-x_1=0$

$w_2-x_2=0$

$w_3-x_3=0$

$w_4-x_4=0$

$g_1(w;x_1)=2x_1^2+2x_1+w_2^2+w_3^2-w_2-w_4-5\leq 0$

$g_2(w;x_2)=x_2^2-x_2+w_1^2+w_1+w_3^2+w_3-w_4-8+w_4^2\leq 0$

$g_3(w;x_3,x_4)=w_1^2+2w_2^2+x_3^2+2x_4^2-w_1-x_4-10\leq 0.$

VI.2.2. Problema de Despacho Econômico de Carga Elétrica de Natureza Real.

Considera-se o problema de geração de energia elétrica com um sistema interconetado de dois nós (zonas geográficas: Norte e Sul). Cada nó conta com um conjunto de centrais térmicas caracterizadas por um determinado custo variável de operação(independente do nível de geração de potência), sujeitas a restrições de máxima geração de potência, e com um conjunto de centrais hidroelétricas, com custos variáveis nulos, também sujeitas a restrições de máxima geração e, além disto, a limitações de máxima geração de energia total.

Associada a cada nó, existe uma curva de demanda zonal mensal (Potência vs. tempo), chamada Curva de Duração, que para efeitos práticos se conhece sob a forma da função escalonada monótona decrescente, como se mostra na Fig.1.

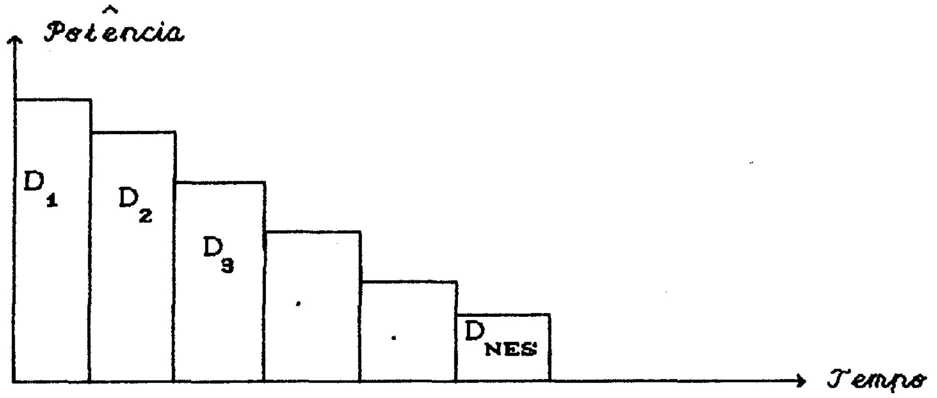

Fig.1

O sistema de transmissão entre ambos os nós é de capacidade limitada e se produzem perdas de potência através dele, regulamentadas pela seguinte função não-linear:

$$P(\text{perda})=\alpha P^2(\text{injectada})+\beta P(\text{injectada})+\gamma, \qquad \text{(VI.6)}$$

onde os coeficientes α, β, γ são por sua vez funções quadráticas da potência (Base e Ponta) gerada em uma das centrais hidroelétricas, chamada COLBÚN e ligada ao Nó Sul.

O sistema interconetado deve funcionar de forma tal a satisfazer as demandas e ao mesmo tempo minimizar o custo de peração e de falha (por falha se entende a parte da demanda de energia que não é possível satisfazer). As falhas são valoradas por uma função "CUSTO DE FALHA" para cada um dos nós, função linear por partes, que leva em conta o nível de profundidade da falha, é dizer, representa uma penalização parcelada da potência a falhar, um exemplo das quais se mostra na figura 2.

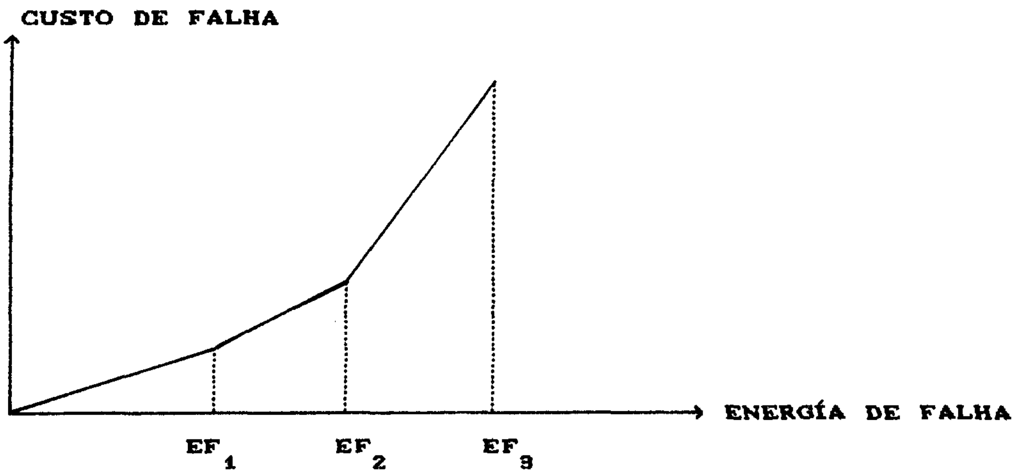

Fig. 2

Pela natureza do problema os custos variáveis de falha são considerados maiores que os custos variáveis de operação. Logo o sistema considera além das centrais hidroelétricas e térmicas, centrais de falha que não são mas que centrais térmicas de custo maior que só funcionarão se a demanda não for satisfeita pelas outras centrais.

Modelo de Programação Matemática

a) Parámetros.

n : número total de centrais.

nh : " " " " hidroelétricas.

nt : " " " " térmicas.

nh_1 : " " " " hidroelétricas no nó N_1(Sul).

nt_1 : " " " " térmicas no nó N_2(Norte).

NES : número de escalões da CURVA DURAÇÃO.

H_j : Número de horas que indicam o comprimento do j-ésimo escalão.

CV_k : Custo variável de operação da k-ésima central térmica.

CW_i : Custo variável de falha i-ésimo, i=1,2,3 (seções 1,2 e 3 do primeiro nó); i=4,5,6 (seções 1,2 e 3 do segundo nó). Para um mesmo nó: $i<j \longrightarrow CW_i \leq CW_j$ (Custos marginais de falha crescentes).

nf : número total de centrais de falla.

nf_1 : número total de centrais de falha no nó N_1.

$EF_i(N_k)$: Níveis de energia de quebra para a função de falha do nó N_k, k=1,2; i=1,2,3.

P_i^B, P_i^P : Potência máxima de geração em Base e em Ponta, respectivamente, para a i-ésima central hidroelétrica, i=1,. . . . . ,nh.

P_k : Potência máxima de geração para a i-ésima central térmica, k=1,. . . . . ,nt.

T : Capacidade máxima de recepção na linha de transmissão.

D_{ij} : Potência demandada no escalão j-ésimo do i-ésimo nó; i=1,2; j=1,. . . . . ,NES.

E_i^P : Máxima capacidade de energia em ponta na i-ésima central hidroelétrica, i=1,. . . . . ,nh.

b) <u>Variáveis</u>

h_{ij}^B, h_{ij}^P : Potência em base e em ponta, respectivamente, gerada pela i-ésima central hidroelétrica no j-ésimo escalão; i=1,. . . . . . . ,nh; j=1,. . . . . . ,NES.

t_{ij} : Potência gerada pela i-ésima central térmica no j-ésimo escalão, i=1,. . . . . . . ,nt; j=1,. . . . . ,NES.

$P_{r_{i}j}$: Potência recebida no j-ésimo escalão do nó i; i=1,2; j=1,. . . . . . ,NES.

$P_{t_{i}j}$: Potencia transmitida no j-ésimo escalão do nó i; i=1,2; j=1,. . . . . . . ,NES.

f_{ij} : Potência de falha gerada no j-ésimo escalão, medida sobre a seção 1,2 ou 3 de algum nó, de acordo com o valor de i(i=1,. . . . . ,6); j=1,. . . . . . . ,NES.

c) Modelo de Otimização.

Minimizar a função objetivo:

$$\sum_{i=1}^{nt} CV_i \left[\sum_{j=1}^{NES} t_{ij} H_j \right] + \sum_{i=1}^{6} CW_i \left[\sum_{j=1}^{NES} f_{ij} H_j \right]$$

s.a.

Restrições de demanda:

$$\sum_{i \in N_1} \left[h^{B}_{ij} + h^{P}_{ij} \right] + \sum_{i \in N_1} t_{ij} + f_{1j} + f_{2j} + f_{3j} + P_{r_1 j} - P_{t_1 j} \geq D_{1j},$$

$$\sum_{i \in N_2} \left[h^{B}_{ij} + h^{P}_{ij} \right] + \sum_{i \in N_2} t_{ij} + f_{4j} + f_{5j} + f_{6j} + P_{r_2 j} - P_{t_2 j} \geq D_{2j},$$

$$j = 1, \ldots\ldots\ldots, NES.$$

Restrições de energia hidroelétrica:

$$\sum_{j=1}^{NES} h^{P}_{ij} H_j \leq E^{P}_{i}, \quad i=1, \ldots\ldots\ldots\ldots\ldots\ldots, nh.$$

Restriçoes de capacidade de geração:

$$P_{r_1 j} + P_{r_2 j} \leq T, \quad j=1, \ldots\ldots\ldots, NES. \qquad (VI.7)$$

Relaçoes de perda de potência na transmissão:

$$P_{r_1 j} = P_{t_2 j} - (\alpha_j P^2_{t_2 j} + \beta_j P_{t_2 j} + \gamma_j),$$

$$P_{r_2 j} = P_{t_1 j} - (\alpha_j P^2_{t_1 j} + \beta_j P_{t_1 j} + \gamma_j), \quad j=1, \ldots\ldots\ldots, NES, \qquad (VI.8)$$

onde, α_j, β_j, e γ_j tal como dissemos em (VI.6), estão definidos por meio de:

$$\alpha_j=\tilde{\alpha}_1 \mathrm{COLB}_j^2+\tilde{\alpha}_2 \mathrm{COLB}_j+\tilde{\alpha}_3,$$

$$\beta_j=\tilde{\beta}_1 \mathrm{COLB}_j^2+\tilde{\beta}_2 \mathrm{COLB}_j+\tilde{\beta}_3, \qquad \text{(VI.9)}$$

$$\gamma_j=\tilde{\gamma}_1 \mathrm{COLB}_j^2+\tilde{\gamma}_2 \mathrm{COLB}+\tilde{\gamma}_3,$$

com,

$$\mathrm{COLB}_j:=h^B_{\mathrm{COLB},j}+h^P_{\mathrm{COLB},j},$$

e o índice COLB representa a

central hidroelétrica de COLBUN.

Os valores de $\tilde{\alpha}_k, \tilde{\beta}_k$ e $\tilde{\gamma}_k$, $k=1,2,3$, serão dados mais adiante, junto com os dados do problema.

<u>Restriçoes de energia de falha no nó</u> N_1:

$$\sum_{j=1}^{NES} f_{1j} H_j \leq EF_1(N_1),$$

$$\sum_{j=1}^{NES} f_{2j} H_j \leq EF_2(N_1)-EF_1(N_1).$$

<u>Restriçoes de energia de falha no nó</u> N_2:

$$\sum_{j=1}^{NES} f_{4j} H_j \leq EF_1(N_2)$$

$$\sum_{j=1}^{NES} f_{5j} H_j \leq EF_2(N_2)-EF_1(N_2).$$

<u>Restriçoes de Capacidade de Geração</u>: (Cotas)

$h^{B}_{i\,j} \le P^{B}_{i}$, $i=1,\ldots\ldots,nh;$ $j=1,\ldots\ldots\ldots,NES.$

$h^{P}_{i\,j} \le P^{P}_{i}$, $i=1,\ldots\ldots,nh;$ $j=1,\ldots\ldots\ldots,NES.$

$t_{k\,j} \le P_{k}$, $k=1,\ldots\ldots,nt;$ $j=1,\ldots\ldots\ldots,NES.$

<u>Restriçoes de não negatividade</u>: (Cotas)

$h^{B}_{i\,j} \ge 0$, $i=1,\ldots\ldots\ldots,nh,$

$h^{P}_{i\,j} \ge 0$, $i=1,\ldots\ldots\ldots,nh,$
$t_{i\,j} \ge 0$, $i=1,\ldots\ldots\ldots,nt,$

$P_{r_i\,j} \ge 0$, $i=1,2,$

$P_{t_i\,j} \ge 0$, $i=1,2,$

$f_{i\,j} \ge 0$, $i=1,\ldots\ldots,6;$

$j=1,\ldots\ldots\ldots,NES.$

d) <u>Dados do problema</u>.

Além das centrais hidroelétricas, térmicas, e de falha(térmicas caras), o modelo considera duas centrais adicionais chamadas centrais de PASSO, cuja característica é que sempre estão gerando energia hidroelétrica. Estas centrais de PASSO(uma para cada nó) se representaran por $PASS_1$ e $PASS_2$ respectivamente. Portanto tem-se os seguintes valores para os dados:

n	nh	nt	nh_1	nt_1	nf	nf_1	T[MW]
24	6	12	4	11	6	3	500

j	1	2	3	4	5	6
D_{1j} [MW] (Demanda nó N_1)	1371	1310	1207	1095	1041	963
D_{2j} [MW] (Demanda nó N_2)	360	341	319	306	294	294
H_j [H/1000] (Comprimento curva duração, ambos nós)	0.016	0.027	0.053	0.091	0.143	0.074

j	7	8	9	10
D_{1j} [MW] (Demanda nó N_1	865	782	727	671
D_{2j} [MW] (Demanda nó N_2)	282	266	253	237
H_j [H/1000] (Comprimento curva duração, ambos nós)	0.074	0.069	0.139	0.058

	Níveis de "energia" de quebra para a função de Falha do Nó respectivo.		
i	1	2	3
$EF_i(N_1)$ [GWH]	20.97	69.90	699.00
$EF_i(N_2)$ [GWH]	6.35	21.17	211.70

Custos variáveis de falha						
	Nó N_1			Nó N_2		
i	1	2	3	4	5	6
CW_i [mills/KWH]	122.499512	=	=	=	=	=

CARACTERÍSTICAS CENTRAIS hidroelétricas nó N_1			
i	P_i^B[MW]	P_i^P[MW]	E_i^P[GWH]
1 ($PASS_1$)	153.3	0.0	0.0
2 (RAPEL)	0.0	352.5	129.0
3 (CISL)	38.8	101.7	69.7
4 (COLB)	90.7	425.2	274.2

CARACTERÍSTICAS CENTRAIS HIDROELÉTRICAS nó N_2			
i	P_i^B[MW]	P_i^P[MW]	E_i^P[GWH]
5 ($PASS_2$)	75.4	0.0	0.0
6 (LAJA)	205.98	553.52	28.51

CARACTERÍSTICAS CENTRAIS TERMOELÉTRICAS			
k	P_k[MW]	CV_k[Mills/KWH]	NÓ
1($RENC_1$)	46.0	38.2660	N_1
2($RENC_2$)	46.0	38.2660	N_1
3($HUAS_1$)	6.7	44.3800	N_1
4($HUAS_2$)	6.7	44.3800	N_1
5(LAG_1)	28.9	62.6850	N_1
6(LAG_2)	23.8	63.0250	N_1
7($FGFO_1$)	21.0	76.9290	N_1
8($FGFO_2$)	21.0	76.9290	N_1

9($TGDI_1$)	23.5	103.5580	N_1
10($TGDI_2$)	23.5	103.5580	N
11($TGDI_3$)	23.5	103.5580	N_1
12(TGCO)	22.5	103.5580	N_2

	COEFICIENTES RELAÇÕES (VI.9)		
K	1	2	3
$\tilde{\alpha}_K$	43.466E-12	50.168E-09	1.216E-04
$\tilde{\beta}_K$	12.589E-10	4.044E-05	-32.893E-06
$\tilde{\gamma}_k$	24.306E-06	23.345E-04	.439E+01

Apresentaremos a seguir o problema equivalente usado para aplicar o método de decomposição. Se observamos o modelo de otimização descrito anteriormente vemos que a estrutura do problema é de quasi-separabilidade dos nós N_1 e N_2. Os únicos pacotes de restrições que contém variáveis de ambos os nós, são os assinalados em (VI.7) e (VI.8). Assim só estas variáveis se doubrarão, a fim de obter separabilidade total relativamente aos nós N_1 e N_2. Obtemos portanto:

$$\text{Min} \sum_{i=1}^{nt} CV_i \sum_{j=1}^{NES} t_{ij} H_j + \sum_{i=1}^{6} CW_i \sum_{j=1}^{NES} f_{ij} H_j$$

Demanda:

$$\sum_{i \in N_1} \left[h^B_{ij} + h^P_{ij} \right] + \sum_{i \in N_1} t_{ij} + f_{1j} + f_{2j} + f_{3j} + P_{r_1 j} - P_{t_1 j} \geq D_{1j},$$

$$\sum_{i \in N_2} \left[h^B_{ij} + h^P_{ij} \right] + \sum_{i \in N_2} t_{ij} + f_{4j} + f_{5j} + f_{6j} + P_{t_2 j} - P_{r_2 j} \geq D_{2j},$$

$j=1,\ldots\ldots\ldots,NES.$

Energia Hidroelétrica:

$$\sum_{j=1}^{NES} h^P_{ij} H_j \leq E^P_i, \quad i=1,\ldots\ldots\ldots,nh.$$

Capacidade de transmissão:

$$P_{r_1 j} + w^{r\,2}_j \leq T, \quad j=1,\ldots\ldots\ldots,NES.$$

Perda de Potência na transmissão:

$$P_{r_1 j} = w^{t\,2}_j - \left[\alpha_j \left(w^{t\,2}_j \right)^2 + \beta_j w^{t\,2}_j + \gamma_j \right]$$

$$w^{r\,2}_j = P_{t_1 j} - \left[\alpha_j \left(P_{t_1 j} \right)^2 + \beta_j P_{t_1 j} + \gamma_j \right]$$

$j=1,\ldots\ldots\ldots\ldots\ldots,NES.$

Energia de falha nó N_1:

$$\sum_{j=1}^{NES} f_{1j}H_j \leq EF_1(N_1),$$

$$\sum_{j=1}^{NES} f_{2j}H_j \leq EF_2(N_1)-EF_1(N_1).$$

Energia de falha nó N_2:

$$\sum_{j=1}^{NES} f_{4j}H_j \leq EF_1(N_2),$$

$$\sum_{j=1}^{NES} f_{5j}H_j \leq EF_2(N_2)-EF_1(N_2).$$

Capacidade de geração e não-negatividade:

$$0 \leq h^B_{ij} \leq P^B_i,$$

$$0 \leq h^P_{ij} \leq P^P_i, \quad i=1,\ldots\ldots\ldots,nh; \; j=1,\ldots\ldots\ldots,NES.$$

$$0 \leq t_{kj} \leq P_k, \quad k=1,\ldots\ldots\ldots,nt; \; j=1,\ldots\ldots\ldots,NES.$$

$$P_{r_i j} \geq 0 \; , \; P_{t_i j} \geq 0, \quad i=1,2; \; j=1,\ldots\ldots\ldots,NES.$$

$$f_{ij} \geq 0, \quad i=1,\ldots\ldots,6; \; j=1,\ldots\ldots\ldots,NES.$$

Restrições adicionais ao modelo:

$$w_j^{2}-P_{r_2 j}=0,$$

$${w_j^{t}}^{2}-P_{t_2 j}=0,$$

$${w_j^{r_2}}^{2} \geq 0 \; , \; {w_j^{t}}^{2} \geq 0, \quad j=1,\ldots\ldots\ldots,NES.$$

Os subproblemas Lagrangeanos que resultam depois da decomposição são: (y_1 e AW1 representam os multiplicadores e as penalidades respectivamente)

<u>Subproblema Nó</u> N_1:

$$\mathrm{Min}\sum_{i=1}^{11} CV_i \sum_{j=1}^{NES} t_{ij} H_j + \sum_{i=1}^{9} CW_i \sum_{j=1}^{NES} f_{ij} H_j +$$

$$+\sum_{j=1}^{NES} y_1(j) * \left[P_{r_1 j} + \alpha_j (w_j^{t\,2})^2 + \beta_j w_j^{t\,2} + \gamma_j - w_j^{t\,2} \right] +$$

$$+1/2 \sum_{j=1}^{NES} \left[P_{r_1 j} + \alpha_j (w_j^{t\,2})^2 + \beta_j w_j^{t\,2} + \gamma_j - w_j^{t\,2} \right]^2 * AW1(j) +$$

$$+\sum_{j=1}^{NES} y_1(NES+j) * \left[w_j^{r\,2} + \alpha_j (P_{t_1 j})^2 + \beta_j P_{t_1 j} + \gamma_j - P_{t_1 j} \right] +$$

$$+1/2 \sum_{j=1}^{NES} \left[w_j^{r\,2} + \alpha_j (P_{t_1 j})^2 + \beta_j P_{t_1 j} + \gamma_j - P_{t_1 j} \right]^2 * AW1(NES+j) +$$

$$+ \sum_{j=1}^{NES} \theta \left[g_{4*NES+j}, y_1(4*NES+j),\ AW1(4*NES+j) \right],$$

com,

$$g_{4*NES+j} := P_{r_1 j} + w_j^{r\,2} - T, \quad \text{e}$$

$$\theta(\cdot):=\begin{cases} y_1(4*\mathrm{NES}+j)*g_{4*\mathrm{NES}+j}+1/2g^2_{4*\mathrm{NES}+j}*AW1(4*\mathrm{NES}+j), \\ \qquad\qquad \text{se } g_{.} \geq \dfrac{-y_1(4*\mathrm{NES}+j)}{AW1(4*\mathrm{NES}+j)}, \\ \\ \dfrac{-y_1^2(4*\mathrm{NES}+j)}{2*AW1(4*\mathrm{NES}+j)}, \text{ em outro caso.} \end{cases}$$

s. a.

$$\sum_{i\in N_1}\left[h^B_{ij}+h^P_{ij}\right]+\sum_{i\in N_1}t_{ij}+f_{1j}+f_{2j}+f_{3j}+P_{r_1 j}-P_{t_1 j}\geq D_{1j},$$

$j=1,\ldots,\mathrm{NES},$

$$\sum_{j=1}^{\mathrm{NES}} h^P_{ij}H_j\leq E^P_i\ ,\quad i=2,\ldots,4,$$

$$\sum_{j=1}^{\mathrm{NES}} f_{1j}H_j\leq EF_1(N_1),$$

$$\sum_{j=1}^{\mathrm{NES}} f_{2j}H_j\leq EF_2(N_1)-EF_1(N_1),$$

$0\leq h^B_{ij}\leq P^B_i\ ,\quad i=1,\ldots,4,\ \ i\neq 2,\ \ j=1,\ldots,\mathrm{NES},$

$0\leq h^P_{ij}\leq P^P_i\ ,\quad i=2,\ldots,4,\ \ j=1,\ldots,\mathrm{NES},$

$0\leq t_{kj}\leq P_k\ ,\quad k=1,\ldots,11,\ \ j=1,\ldots,\mathrm{NES}$

$0\leq P_{r_1 j},\quad j=1,\ldots,\mathrm{NES},$

$0\leq P_{t_1 j}\ ,\quad j=1,\ldots,\mathrm{NES},$

$0\leq f_{ij}\ ,\quad i=1,\ldots,3,\ \ j=1,\ldots,\mathrm{NES}.$

Subproblema Nó N_2:

$$\text{Min } CV_{12}\sum_{j=1}^{NES} t_{12,j} + \sum_{i=4}^{6} CW_i \sum_{j=1}^{NES} f_{ij} H_j +$$

$$+\sum_{j=1}^{NES} y_1(2*NES+j)*({w_j^r}^2 - P_{r_2 j}) + 1/2\sum_{j=1}^{NES}\left[{w_j^r}^2 - P_{r_2 j}\right]^2 *AW1(2*NES+j) +$$

$$+\sum_{j=1}^{NES} y_1(3*NES+j)*({w_j^t}^2 - P_{t_2 j}) + 1/2\sum_{j=1}^{NES}\left[{w_j^t}^2 - P_{t_2 j}\right]^2 *AW1(3*NES+j)$$

s.a.

$$\sum_{i \in N_2}\left[h^B_{ij} + h^P_{ij}\right] + \sum_{i \in N_2} t_{ij} + f_{4j} + f_{5j} + f_{6j} + P_{r_2 j} - P_{t_2 j} \geq D_{2j}, \quad j=1,\ldots,NES,$$

$$\sum_{j=1}^{NES} h^P_{6j} H_j \leq E^P_6 ,$$

$$\sum_{j=1}^{NES} f_{4j} H_j \leq EF_1(N_2),$$

$$\sum_{j=1}^{NES} f_{5j} H_j \leq EF_2(N_2) - EF_1(N_2),$$

$$0 \leq h^B_{ij} \leq P^B_i , \quad i=5,6 , \quad j=1,\ldots\ldots,NES,$$

$$0 \leq h^P_{6j} \leq P^P_6 , \quad j=1,\ldots\ldots,NES,$$

$$0 \leq t_{12,j} \leq P_{12} , \quad j=1,\ldots\ldots,NES,$$

$$0 \leq P_{r_2 j} , \quad 0 \leq P_{t_2 j} , \quad j=1,\ldots\ldots,NES,$$

$$0 \leq f_{ij} , \quad i=4,5,6, \quad j=1,\ldots\ldots,NES.$$

O ponto inicial viável primal foi construido por um algoritmo heurístico chamado "Método do Rasurado", idealizado para este problema particular, com a finalidade de obter o melhor ponto viável primal com relação somente às restrições lineares do modelo. E' o que apresentamos a seguir:

<u>Algoritmo do Rasurado</u>: Fazer para cada Nó em separado:

PASSO 1.- Para cada escalão, se fixa a geração das "centrais de passo" em sua potência máxima, e se desconta esta contribuição da "Curva de Duração".

PASSO 2.- Para cada central hidroelétrica regulável(isto é, aquela que gera potência em Ponta), começando desde o escalão mais alto, se coloca sua potência até igualar eventualmente a altura do escalão seguinte, de forma sucessiva, e de modo tal que não ultrapasse a capacidade de geração da central, nem a energia disponível.

PASSO 3.- Se depois de colocar as centrais hidroelétricas reguláveis, ainda não se satisfizer a demanda, então se usa a potência restante das centrais térmicas (sob o nível máximo). Isto é feito a partir do escalão mais alto e de modo a igualar a altura do escalão seguinte, de forma sucessiva.

PASSO 4.- Se depois de colocar as centrais térmicas ainda não se satisfizer a demanda, a potência em falta se completa então com as "centrais de falha", começando com a primeira.

<u>OBSERVAÇÃO</u>: Pela natureza do problema, se não existissem as restrições não lineares (VI.8), o ponto gerado por este algoritmo seria o ótimo do problema. Logo, tem-se aqui um método alternativo de resolução para resolver o problema de programação linear particular.

VI.2.3.-<u>RESULTADOS</u>.

Os resultados foram obtidos em um computador Micro VAX-3400, sob VMS (versão V5.3-1). O algoritmo de decomposição foi implementado em linguagem Fortran 77 . Todos os cálculos, assim como o registro dos dados e variáveis, foram feitos em dupla precisão. Nas tabelas N°1, N°2 e N°3, resumimos os resultados obtidos na resolução dos problemas VI.2.1.1.-VI.2.1.4 e VI.2.2, utilizando o algoritmo de decomposição e os códigos, MINOS/AUGMENTED e PENAMOR. O problema de Grande Porte VI.2.2 foi resolvido para NES=2,3,4,5 e 10.

PROBLEMA	VALOR ÓTIMO		
	PENAMOR	MINOS	P-P-DUAL
Ad-Hoc	-0.75	-0.75	-0.7499
Biggs	-3300	-3300	-3299.915
Paviani	961.715	961.715	961.710
Rosen-Suzuki	-44	-44	-44
GRANDE PORTE NES=2	0	0	0
GRANDE PORTE NES=3	0	0	0
GRANDE PORTE NES=4	0	0	0
GRANDE PORTE NES=5	0	0	0
GRANDE PORTE NES=10	233.761	233.7608	235.005

TABELA N°1

PROBLEMA	ITERAÇÕES MAIORES			ITERAÇÕES MENORES		
	PENAM.	MINOS	P-P-DUAL	PENAM.	MINOS	P-P-DUAL
AD-HOC	11	4	4	12	2	9
Biggs	1	1	12	3	3	18
Paviani	4	23	14	13	52	82
Rosen-Susuki	20	13	16	28	35	53
GRANDE PORTE NES=2	3	8	6	112	80	7
GRANDE PORTE NES=3	5	9	6	134	111	6
GRANDE PORTE NES=4	2	9	6	136	144	11
GRANDE PORTE NES=5	2	9	7	144	165	14
GRANDE PORTE NES=10	9	13	9	467	388	52

TABELA N°2

PROBLEMA	TEMPO C.P.U. (seg)		
	PENAMOR	MINOS	P-P-DUAL
AD-HOC	0.77	0.76	0.37
Biggs	0.22	0.80	1.55
Paviani	0.50	1.95	7.93
ROSEN-SUSUKI	0.78	1.60	5.70
GRANDE PORTE NES=2	10.03	5.94	5.90
GRANDE PORTE NES=3	17.75	9.23	11.42
GRANDE PORTE NES=4	25.04	12.91	23.98
GRANDE PORTE NES=5	36.58	17.19	47.42
GRANDE PORTE NES=10	5.01.17	1.12.21	22.06.18(*)

TABELA N 3

(*) Este tempo se reduz a 18.52.20 (seg) no caso em que as minimizações na variável primal x sejam feitas por meio de processamento em paralelo. Mesmo assim, este último tempo se reduz a 12.36.65 (seg) se só queremos uma precisão da ordem do 5% para a solução ótima, ou equivalentemente de 2% para o valor ótimo.

CAPÍTULO VII

CONCLUSÕES E COMENTÁRIOS FINÁIS

Neste trabalho se desenvolveu, tanto nos aspectos conceituais como computacionais, um Método de Decomposição por meio de Penalização Lagrangeana para a resolução de problemas de programação não-linear de grande porte não necessariamente convexos e com uma estrutura qualquer.

Os resultados obtidos na aplicação a problemas de prova usuais na literatura e a um problema de Despacho Econômico de Carga Elétrica de natureza real, satisfazem amplamente as expectativas computacionais iniciais para um método desta natureza.

Em particular, nos exemplos resolvidos, o método resultante apresenta propriedades satisfatórias de convergência local, sem requerer necessariamente uma boa estimativa da solução ótima. Mas, das experiênçias numéricas conclui-se que o método produz rapidamente uma boa aproximação da solução ótima, que é melhorada mais lentamente nas iterações posteriores. Isto é concordante com a natureza mesma do algoritmo que utiliza uma minimização parcial do tipo descida mais pronunciada, que sabemos apresentar um comportamento similar. Isto foi corroborado no caso do problema real de grande porte com 10 escalões, onde após três iterações maiores se obteve uma aproximação da solução ótima com um erro relativo da ordem de 5%, o que se traduz em um erro relativo da ordem de 2% no valor ótimo(ver tabela N°1). Por outro lado, em relação aos tempos totais de cálculo para os problemas reportados na tabela N°3, é possível concluir que o Método é competitivo com MINOS e PENAMOR em vários destes problemas. Porém, este não é o caso para a resolução numericamente exata do problema de grande porte. Una das razões fundamentais é que à diferença das características descritas mais acima, tanto MINOS como PENAMOR utilizam um método de segunda ordem para a resolução dos problemas lagrangeanos. Para paliar esta dificuldade, foi necessário utilizar um escalonamento das variáveis, definido a partir de uma aproximação da inversa da matriz hessiana no ponto atual.

Sem prejuizo do anterior, o Método aqui desenvolvido é sempre atrativo no sentido que requer a resolução de subproblemas lagrangeanos de dimensão menor, que podem ser eventualmente resolvidos por meio de processamento paralelo. Concretamente, no caso do problema real segundo a tabela N°3, isto significaría reduzir o tempo total de cálculo em torno de 25% para a solução exata, e de 50% para a solução aproximada em 5%.

Através da experimentação computacional, evidenciou-se que a incorporação do critério assintótico descrito no final da seção VI.1, melhorou ostensivamente a convergência computacional do Método.

Em relação com o problema geral de grande porte cabe assinalar que existem outras alternativas de decomposição que podem significar melhoras significativas nos tempos totais de cálculo. De maneira geral, esperamos que uma boa alternativa de decomposição deve estabelecer um compromisso adequado entre o grau de dificuldade para a resolução dos problemas do nível inferior e o condicionamento numérico do problema lagrangeano na variável w. De qualquer maneira, não devemos perder de vista que toda decomposição deve traduzir-se em uma interpretação natural da resolução do problema real.

Finalmente, é importante notar que fica aberta a possibilidade, heurística por enquanto, de aplicar um método de segunda ordem tipo métrica variável para resolver os subproblemas na variável w.

REFERÊNCIAS BIBLIOGRÁFICAS

ARORA, J.S. e GOVIL, A.K.(1977),"An efficient method for optimal structural design by sub-structuring",Computing Structures 7,pp. 507-515

ARROW, K.J., HURWICZ, L. e UZAWA, H.(1958),"Studies in Linear and Nonlinear Programming", Stanford University Press, Stanford California.

ARROW, K.J. e SOLOW, R.M.(1958),"Gradient methods for constrained maxima with weakened assumptions",Studies in Linear and Nonlinear Programming", K. Arrow, L. Hurwicz and H. Uzawa, eds., Stanford University Press, Stanford, California.

ARROW, K.J., GOULD, F.J. e HOWE, S.M.(1973),"A general saddle-point result for constrained minimization", Mathematical Programming, Vol. 5, pp. 225-234.

ATTOUCH, H.(1984),"Variational Convergence for Functions and Operators", Applicable Maths-Series, Pitman, London.

BANK, B., GUDLAT, J., KLATTE, D., KUMMER, B. e TAMMER,K.(1983),"Nonlinear Parametric Optimization", (Birkhaüser,Basel).

BARTHELEMY, J.F.M.(1988),"Enginneering Applications of heuristics Multilevel Optimization Methods", Gamm-Seminar sur Optimization and Modelization", University of Siegen, Oct. 5-7.

BERTSEKAS, D. P.(1973),"Convergence rate of penalty and multipliers methods",Proc. 1973 IEEE Confer. Decision Control, San Diego, California, pp. 260-264.

————————(1975),"Combined Primal-Dual and Penalty Methods for Constrained Minimization", Siam J. Control, Vol. 13, N°3, pp. 521-544.

————————(1976a),"Multiplier Methods: A Survey", Automatica, Vol. 12, pp.133-145.

————————(1976b),"On Penalty and multiplier Method for Constrained Minimization", Siam Journal on Control and Optimization, Vol. 14, N°2, pp. 216-235.

————————(1979),"Convexification procedures and decomposition methods for nonconvex optimization problems", Journal of Opt. Theory and App., N°29, pp. 169-197.

——————(1982),"Constrained Optimization and Lagrange Multiplier Methods",Academic Press, New York.

BUYS, J. D.(1972),"Dual Algorithms for Constrained Optimization Problems", Ph.D. Thesis, Rijksuniversiteit de Leiden.

CLARKE, F. H.(1983),"Nonsmooth Analysis and Optimization", New York, Wiley-Interscience.

COHEN, G.(1978),"Optimization by Decomposition and Coordination: A Unified Approach", IEEE Transactions on Automatic Control, Vol. AC-23, N°2, pp. 222-232.

——————(1980),"Auxiliary Problem Principle and Decomposition of Optimization Problems", Journal of Opt. Theory and App., Vol. 32, N°3, pp. 277-305.

CONTESSE. L.(1985),"A damped penalty method in Nonlinear Optimization", 12th International Symposium on Mathematical Programming, Massachusetss Institute of Technology, 5-9 Agosto, Cambridge.

——————(1986a),Una variante del método de Penalización Lagrangeana para la Optimización de Modelos no Lineales", Anales III CLAIO, Tomo 1, pp. 285-293, Santiago-Chile.

——————(1986b),"Sur la differentiabilite de fonctions marginales", Documento de Trabajo 86/6, Depto. de Ingeniería de Sistemas, Pontificia Universidad Católica de Chile, presentado en el I Encuentro Internacional de Optimización realizado conjuntamente con III CLAIO, 18-22 de Agosto, Santiago,Chile.

——————(1986c),"On the Boundedness of Certain Point-to-Set Maps and Its Application in Optimization",in Recent Advances in System Modelling and Optimization, Lecture Notes on Control and Optimization, Springer Verlag.

——————(1987a),"Método de Penalización Lagrangeana Amortiguada para la resolución de modelos no lineales", Apuntes de Ingeniería, N°26, pp. 63-83.

——————(1987b),"A Damped Lagrangian Penalty Approach to

Second Order Constrained Stationary Points", The first International Conference on Advances in Communication and Control Systems, Washington D.C., 18-20 June.

————————(1990a),"Extended Convergence Results for the Method of Multipliers for Non-strictly binding Constraints", Documento de Trabajo 90/1, Depto. de Ingeniería de Sistemas, Pontificia Universidad Católica de Chile.

————————(1990b),"Método del gradiente proyectado para la minimización de una función continuamente diferenciable sobre un dominio convexo compacto", Documento de Trabajo 90/3, Departamento de Ingeniería de Sistemas, Pontificia Universidad Católica de Chile.

CONTESSE, L. e VILLAVICENCIO, J.(1987),"Aplicación parcial del método de Penalización Lagrangeana Amortiguada ante restricciones lineales", Apuntes de Ingeniería, N°26, pp.63-83.

————————————————(1988),"Una variante del método de Penalización Lagrangeana bajo restricciones lineales", Investigación Operativa, Vol 1, N°1, pp. 35-42.

————————————————(1989),"Resolución del Modelo Económico de Despacho de Carga Eléctrica Mediante el Método de Penalización Lagrangeana con Cotas", Revista ICHIO, Vol. 3, N°1, pp.80-112.

————————————————(1990),"PENAMOR: Una Implementación Computacional del Algoritmo de Penalización Lagrangeana Amortiguada", XIX JAIIO, V CLAIO, Buenos Aires, Argentina, 10-14 de Septiembre.

CONTESSE, L., OLIVEIRA, P.R. e PAREDES, F.(1990),"Sobre un Método de Descomposición mediante Penalización Lagrangeana", 19 Jornadas Argentinas de Informática e Investigación Operativa (XIX JAIIO), V Congreso Latinoamericano de Investigación Operativa (V CLAIO), Buenos Aires, Argentina 10-14 de Septiembre.

DANTZIG, G. B.(1963),"Linear Programming and Extensions", Princeton University Press, Princeton,N.J.

DANTZIG, G. B. e WOLFE P.(1960),"Decomposition principle for linear programs", Operations Research 8, pp. 101-111.

DURAN, M. A. e GROSSMAN(1986),"An Outer-Approximation Algorithm for a Class of Mixed-Integer Nonlinear Programs", Mathematical Programming 36, pp. 307-339.

ERMOLEV, Y. M. e ERMOLEVA, L. G.(1973),"The method of parametric decomposition", Kibernetica 9, 2, pp. 66-69; Cybernetics 9, 2, pp. 262-266.

FIACCO, A. V.(1984),"Sensitivity, Stability and Parametric Analysis", Mathematical Programming Study 21.

FIACCO, A. V. e McCORMICK, P.(1968),"Nonlinear Programming: Sequential Unconstrained Minimization Techniques", New York, Wiley.

FINDEISEN, W., BAILEY, F. N., BRDYS, M., MALINOWSKI, K., TATJEWSKI, P. e WOZNIAK, A.(1980),"Control and Coordination in Hierarchical Systems", John Wiley and Sons.

FLETCHER, R.(1970),"A class of methods for nonlinear programming with termination and convergence properties", Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam.

————————(1973),"A class of methods for nonlinear programming, III: Rates of Convergence, Numerical Methods for Non-linear Optimization", F. A. Lootsma, ed., New York, Academic Press.

————————(1987),"Practical Methods of Optimization", John Wiley & Sons.

FLETCHER, R. e LILL, S.(1971),"A class of methods for nonlinear programming, II: Computational experience", Nonlinear Programming, J. B. Rosen, O. 1. Mangasarian and K. Ritter, eds., New York, Academic Press.

FRANK, M. e WOLFE, P.(1956),"An algorithm for quadratic programming", Naval Research Logistics Quarterly 3, pp. 95-110.

GABRIELE, G. A. e RAGSDELL, K. M.(1980),"Large Scale Nonlinear Programming Using the Generalizad Reduced Gradient Method", ASME TRANS. Journal of Mechanical Design, Vol. 102, pp. 566-573.

GAUVIN, J.(1977),"The method of parametric decomposition in Mathematical Programming: The case nonconvex", in

Nonsmooth Optimization, Proceedings of a IIASA Workshop, Lemarechal C. e Mifflin R., eds., March 28-April 8.

GAUVIN, J. e TOLLE, J. W.(1977),"Differential Stability in Nonlinear Programming", Siam Journal on Control and Optimization, Vol. 15, N°2, pp.294-311.

GEOFFRION, A. M.(1970a),"Elements of Large-Scale Mathematical Programming Part I: Concepts", Management Science, Vol. 16, N°11,pp. 652-675.

————————(1970b),"Elements of Large-Scale Mathematical Programming Part II: Synthesis of Algorithms and Bibliography", Management Science, Vol. 16, N°11, pp. 676-691.

————————(1970c),"Primal Resource Directive Approaches for Optimizing Nonlinear Decomposable Systems", Operations Research, Vol. 18, N°3, pp. 294-311.

————————(1971),"Large-Scale linear and nonlinear programming", in Optimization Methods for Large-Scale Systems with Applications, Wismer, D. A., ed., New York, McGraw-Hill.

————————(1972a),"Duality in nonlinear Programming a simple application-oriented development", in Perspectives on Optimization, Geoffrion, A. M., ed., Addison-Wesley, Reading, M.A., pp. 65-101.

————————(1972b),"Generalized Benders Decomposition", Journal Optimization Theory and Appl., 10, pp. 237-260.

GOL'SHTEIN, E. G.(1986),"The block method of convex programming", Soviet Math. Dokl., Vol. 33, N°3, pp. 584-587.

GRANVILLE, S. e SCHECHTMAN, J.(1988),"An Application of the Augmented Lagrangian Approach to Decomposition in Linear Programmming", Investigación Operativa, Vol. 1, N°1, pp. 55-78.

GUIGNARD, M.(1982),"Optimality and Stability in Mathematical Programming", Mathematical Programming Study 19.

GUNN, E. A., THORBURN, M. e RAI, A. K.(1988),"A Decomposition Method based on the Augmented Lagrangian", Infor, Vol. 26, N°2, pp. 91-113.

HAARHOFF, P. C. e BUYS, J. D.(1970),"A new method for the optimization of a nonlinear function subject to nonlinear constraints", Computer Journal, Vol. 13, pp. 178-184.

HESTENES, M. R.(1969),"Multiplier and Gradient Method", Journal of Optimization Theory and Applications, Vol. 4, N°5, pp. 303-320.

HIMMELBLAU, D. M.(1972),"Applied Nonlinear Programming", Mc Graw-Hill Book Company, New York.

————(1973),"Decomposition of Large-Scale Problems", North Holland Publishing Company.

HIRIART-URRUTY, J.B.(1978),"Gradients Generalises de Fonctions Marginales", Siam J. Control and Optimization, Vol. 16, N°2.

HOCK, W. e SCHITTKOWSKI, K.(1981),"Test Examples for Nonlinear Programming Codes", Lecture Notes in Economics and Mathematical Systems, Vol. 187, Springer-Verlag, Berlin.

HOGAN, W. (1973),"Directional Derivatives for Extremal-Value Functions with Applications to the Completely Convex Case", Operations Research 21, pp. 188-209.

KIWIEL, K. C.(1983),"An aggregate subgradient method for nonsmooth convex minimization", Mathematical Programming 27, N°3, pp. 320-341.

————(1985),"Methods of Descent for Nondifferentiable Optimization", Dold, A. e Eckmann, B., eds., Lecture Notes in Mathematics, New York, Springer Verlag.

KOCIS, G. R. e GROSSMANN, J. E.(1987),"Global Optimization of Nonconvex MINLP Problems in Process Synthesis", Technical Report department of Chemical Engineering, Carnegie-Mellon University.

KORT, B. W. e BERTSEKAS, D. P.(1972),"A new penalty function method for constrained minimization", Proc. IEEE

Decision and Control Conference, New Orleans.

LASDON, L. S.(1970),"Optimization Theory for Large Systems", London, The Macmillan Company.

————————(1973),"Decomposition in resource allocation", in Decomposition of Large-Scale Problems, Himmelblau, D.M.,ed., North Holland, pp.277-293.

————————(1989),"Methods of Operations Research", XIV Symposium on Operations Research, University of Ulm, Sept 6-8, Proceedings ed. by Rieder, U.,Gesmer, P., Peyerinhoff, A. e Radermacher, F. J.

LEMARECHAL, C.,STRODIOT,J.J. e BIHAIN, A.(1981),"On a bundle algorithm for nonsmooth optimization", in Nonlinear Programming 4, Mangasarian,O.L., Meyer, R.R. e Robinson, S.M., eds., New York, Academic Press.

LILL, S.A.(1973),"Generalization of an exact method for solving equality constrained problems to deal with inequality constraints", in Numerical Methods for Nonlinear Optimization, Lootsma, F.A., ed., New York, Academic Press.

LOOTSMA, F.A. (1978),"The Algol 60 procedure minifun for solving non-linear optimization problem", in Design and Implementation of Optimiazation Software, Greenberg, H.J.,ed., pp. 397-445.

LUENBERGER, D.G. (1969),"Optimization by Vector Space Methods", John Wiley & Sons, Inc.

————————(1973),"Introduction to linear and Nonlinear Programming", Massachusets, Addison-Wesley, Reading.

LUNA, H.P.L. (1984),"A survey on Informational Decentralization and Mathematical Programming Decomposition", in Informational Decentralization and Mathematical Programming Decomposition, Cottle, R.W., Beunsousan, M.L. e Korte, B., eds., Elsevier Science Publishers, B.V., North Holland.

MEDHI, D. (1987),"Decomposition of structure large scale Optimization problems and paralell Computing", Thesis, Computer Sciences Department, University of Wisconsin, Madison.

MESAROVIC, M.D., MACKO, D., TAKAHARA, Y.(1970),"Theory of Hierarchical Multilevel Systems", Academic Press, New York and London.

MIELE, A., CRAGG, E.E., IYER, R. R. e LEVY, A.V.(1971a),"Use of the augmented penalty function in mathematical programming problems", Part I, J. Optimization Theory and App. pp. 115-130.

MIELE, A., CRAGG, E.E. e LEVY, A.V.(1971b),"Use of the augmented penalty function in mathematical programming problems", Part II, J. Optimization Theory and App.8, pp.131-153.

MIELE, A., MOSELEY, P.E., LEVY, A.V. e COGGINS, G.M.(1972a)"On the method of multiplier for mathematical programming problems", J. Optimization Theory App.,10, pp. 1-33.

MIELE, A., MOSELEY, P.E., e CRAGG, E.E.(1972b)"A modification of the method of multipliers for mathematical programming problems", in Techniques of Optimization, Balakrishnan, A.V., ed., New York, Academic Press.

MICHELON, P. e MACULAN, N.(1988),"Decomposition Lagrangienne pour Problèmes de Programation non Lineáire en Nombres Entiers à Contraintes Linéaires", Technical Report, Coppe e Instituto de Matematica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

MIFFLIN, R.(1985),"The solution of a nested nonsmooth Optimization problem".

MURTAGH, B.A. e SAUNDERS, M.A.(1982),"A projected Lagrangian Algorithm and its implementation for Sparse Nonlinear Constraints", Mathematical Programming Study, Vol 16, pp. 84-118.

NGUYEN, V.H.(1988),"Decomposition based on Nonsmooth Optimization Methods", Workshop on Mathematical Programming, October 10-14, Rio de Janeiro, Brazil.

OLIVEIRA, P.R. e PAREDES, F.(1988),"Programação Não-Suave e Decomposição em Programação Não-Linear", IV Congreso Latino-Ibero-Americano de Pesquisa Operacional e Engenharia de Sistemas (IV CLAIO), XXI Simpósio

Brasileiro de Pesquisa Operacional (XXI SBPO), Rio de Janeiro, Outubro, 18-22.

POWELL, M.J.D.(1969),"A method for nonlinear constraints in minimization problem", in Optimization, Fletcher, R., ed., New York, Academic Press, pp. 283-298.

POLYAK, B.T.(1970),"Iteration methods using Lagrange Multipliers for the solution of extremal problems with constraints of the equality type", U.S.S.R. Comput. Math. and Math. Phys., 10, N°5, pp. 1098-1106.

POLYAK, V.T, e TRET'YAKOV, N.V.(1973),"The method of penality estimates for conditional extremun problems", Zh. vychisl. Mat. mat. Fiz., 13, 1, pp. 34-46.

ROBINSON, S.M.(1987),"Bundle-Based decomposition: Conditions for Convergence", Working Paper, International Institute for Applied Systems Analysis A-2361, Laxemburg, Austria.

ROCKAFELLAR, R.T.(1970),"Convex Analysis", Princeton University Press, Princeton, N.Y.

———————(1971),"New Applications of Duality in Convex Programming", Proc. Confer. Probab., 4th. Brasov, Romenia, pp. 73-81.

———————(1973),"The multiplier Method of Hestenes and Powell applied to Convex Programming", Journal of Optimization Theory and app., Vol 12, pp. 555-562.

———————(1974a),"Conjugate Duality and Optimization", CBMS Regional Conference Series in Applied Mathematics N°16, Society for Industrial and Applied Mathematics, Philadelphia, P.A.

———————(1974b),"Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming", SIAM, Journal on Control and Optimization, Vol 12, N°2, pp. 268-285.

———————(1976a),"Solving a nonlinear program by way of a dual problem", Symposia Mathematica 19, pp. 135-160.

———————(1976b),"Augmented Lagrangians and applications of the proximal point algorithm in Convex programming", Mathematics of Operations Research 1, pp. 97-116.

______________(1978),"Monotone Operators and Augmented Lagrangian methods in nonlinear programming", in Nonlinear Programming 3, Mangasarian, O.L., Meyer, R.R. e ROBINSON, S.M., eds., New York, Academic Press, pp. 21-25.

RUPP, R.D.(1972),"Approximation of the classical isoperimetric problem", Journal of Optimization Theory and App., Vol 9, pp. 251-264.

SHAPOUR, A. e LI, W.S.(1986),"Optimal Design using decomposition methods", Technical Research Report, TR-86-60, Systems Research Center, University of Mariland.

SHOR, N.Z.(1985),"Minimization Methods for Non-Differentiable Functions", Springer Verlag.

SILVA, B.M.E., RAGSDELL, K.M., SANDGREN,E., BHATT, V., MARJADI, D. e Wu, J.J.(1986),"Design of ultra-light vehicles using decomposition in a paralell Computing enviroment", Report 86 DPC021, Design Productivity Center, College of Engineering, University of Missouri, Columbia, MO 65211.

SILVERMAN, G.J.(1972a),"Primal Decomposition of Mathematical Programs by Resource Allocation: I. Basic Theory and a Direction Finding Procedure", Operations Research, 20, 1, pp. 58-74.

______________(1972b),"Primal Decomposition of Mathematical Programs by Resource Allocation:II. Computational Algorithm with an Application to the Modular Design Problem", Operations Research, 20, 1, pp. 75-93.

SOBIESKI, J.S., BARTHELEMY, J.F.M. e RILEY, K.M.(1982),"Sensitivity of optimun solutions of problem parameters", Amer. Inst. Aeronautics and Astronautics 20, pp. 1291-2199.

SOBIESKI, J.S., JAMES, B.B. e DOVI, A.R. (1985),"Structural Optimization by multilevel decomposition", American Inst. Aeronautics and Astronautics 23, pp. 1775-1782.

SORENSEN, D.C. e WETS, R.(1982),"Nondifferentiable and

Variational Technique in Optimization", Mathematical Programming Study 17.

SPINGARN, J.E.(1985),"Application of the method of Partial inverses to convex programming: Decomposition", Mathematical Programming, 32, pp. 199-223.

STEPHANOPOULOS, G. e WESTERBERG, W.(1975),"The use of Hestenes Method of multiplier to resolve dual gaps in engineering systems Optimization", Journal of Optim. Theory and App., 15, pp. 285-309.

TATJEWSKI, P. e WOZNIAK, A.(1980),"Control and Coordination in Hierarchical Systems", John Wiley and Sons.

TSENG, P. e BERTSEKAS, D.P.(1987),"Relaxation methods for problems with strictly convex separable costs and linear constraints", Mathematical Programming 38, N^{o}3, pp. 303-319.

TRIPATHI, S.S. e NARENDRA, K.S. (1972),"Constrained Optimization Problems Using Multipliers Methods", J. Opt. Theory and App.,9, pp. 59-70.

UZAWA, H.(1958),"Iterative methods for concave programming", in Studies in Linear and Nonlinear Programming, Arrow, K., Harcwicz e Uzawa, H., eds., Stanford University Press, Stanford, CA.

VILLAVICENCIO, J.(1987),"Una implementación eficiente del Método de Multiplicadores Amortiguado ante la Presencia de restricciones lineales", Tesis, Ingeniero Civil Matemático, Universidad de Chile, Santiago-Chile.

WIERZBICKI, A.P.(1970),"Convergence properties of a penalty shifting algorithm for nonlinear programming problems with inequality constraints", Arch. Automatic i Telemech.

——————————(1971),"A penalty function shifting method in constrained static optimization and its convergence properties", Arch. Automatic i Telemech 16, pp. 395-416.

WIERZBICKI, A.P. e HATKO, A. (1973),"Computational methods in Hilbert space for optimal control problems with delays", Proc. Sth. IFIP Conference on Optimization

Techniques, Rome Springer-Verlag, Berlin.

WISMER, D.A.(1971),"Decomposition of Large-Scale Problems", North Holland Publishing Company.

WISMER, D.A. e CHATTERGY, R.(1978),"Introduction to Nonlinear Optimization", North Holland, New York, pp. 325-349.

APÊNDICE

Neste capítulo apresentamos um estudo desenvolvido por CONTESSE(1986b), sobre a diferenciabilidade das funcões marginais.

Sejam X e Y dois espaços de Banach, $\Omega: Y \longrightarrow X$ uma multiaplicação e $f: X\times Y \longrightarrow \mathbb{R}$ uma função numérica. Considera-se a seguinte família parâmetrizada de problemas de otimização:

$$P(y): \underset{x\in\Omega(y)}{\text{Min}}\ f(x,y).$$

Associada a esta família, tem-se a função marginal definida por:

$$v(y):=\begin{cases} \inf\left\{f(x,y)/x\in\Omega(y)\right\}, \text{ se } \Omega(y)\neq\phi, \\ +\infty, \text{ em outro caso,} \end{cases}$$

e a multiaplicação de soluções otimais $M: Y \longrightarrow X$, definida por:

$$M(y):=\left\{x\in\Omega(y)/\ f(x,y)\leq v(y)\right\}, \forall\ y\in Y.$$

Neste contexto estudamos as propriedades de diferenciabilidade da função marginal $v(\cdot)$, baseadas nas propriedades de diferenciabilidade da função econômica $f(\cdot,\cdot)$, em relação do parâmetro y, no caso dos problemas completamente convexos. Logo, temos o Lema seguinte:

<u>Lema 1</u>.-Sea $f: X\times Y \longrightarrow \mathbb{R}$ convexa em XxY em relação do par (x,y), parcialmente diferenciável com respecto de y, em $M(\bar{y})\times\langle\bar{y}\rangle$, para um certo $\bar{y}\in Y$ tal que $M(\bar{y})\neq\phi$, com $M(\bar{y}):=\left\{x\in X/\ f(x,\bar{y})=\inf_{z\in Y} f(z,\bar{y})\right\}$. Sob estas condições,

tem-se:

$$D_y f(x,\bar{y}) := \left.\frac{\partial f(x,y)}{\partial y}\right|_{y=\bar{y}} = V(\bar{y}), \forall\ x\in M(\bar{y}),$$

onde, $V(\bar{y})$ representa uma certa aplicação linear continua fixa de Y em $\mathbb{R}$, é dizer, $V(\bar{y})\in\mathcal{L}(Y,\mathbb{R})$.

<u>Demonstração</u>:

Sejam $x^1, x^2 \in M(\bar{y})$ e seja $d \in X$. Como $f(\cdot,\cdot)$ é diferenciável em relação de y nos pontos $(x^1,\bar{y})$ e $(x^2,\bar{y})$, tem-se:

$$f(x^1,\bar{y}+\lambda d)=f(x^1,\bar{y})+\lambda D_y f(x^1,\bar{y})d+o_1(\lambda),$$

$$f(x^2,\bar{y}-\lambda d)=f(x^2,\bar{y})-\lambda D_y f(x^2,\bar{y})d+o_2(\lambda),$$

onde, $\lambda^{-1}o_1(\lambda)\longrightarrow 0$ e $\lambda^{-1}o_2(\lambda)\longrightarrow 0$, quando $\lambda\longrightarrow 0$, de maneira que, pela convexidade de $f(\cdot,\cdot)$ em relação de (x,y), temos:

$$\begin{aligned} f\left(\frac{x^1+x^2}{2},\bar{y}\right) &\le 1/2\ f(x^1,\bar{y}+\lambda d)+1/2\ f(x^2,\bar{y}-\lambda d) \\ &= 1/2\ f(x^1,\bar{y})+1/2\ f(x^2,\bar{y})+\lambda/2\left[D_y f(x^1,\bar{y})-D_y f(x^2,\bar{y})\right]+ \\ &+ \underbrace{1/2\ o_1(\lambda)+1/2\ o_2(\lambda)}_{o_3(\lambda)}, \end{aligned}$$

Agora, como $M(\bar{y})$ é convexo, se cumpre:

$$f\left(\frac{x^1+x^2}{2},\bar{y}\right) = 1/2\left[f(x^1,\bar{y})+f(x^2,\bar{y})\right].$$

Logo, usando este fato na desigualdade acima tem-se:

$$1/2\ [\ D_y f(x^1,\bar{y})-D\ f(x^2,\bar{y})\ ]d + 1/2\ \frac{o_3(\lambda)}{\lambda} \ge 0,\ \forall\lambda>0.$$

Sob estas condições, se $\lambda\longrightarrow 0_+$, obtemos:

$$[\,D_y f(x^1,y)-D_y f(x^2,y)]d\geq 0$$

Como $d\in X$ é arbitrário, então temos:

$$D_y f(x^1,\bar{y})-D_y f(x^2,\bar{y})=0,$$

o que prova o Lema. ■ ■

<u>Observação 1</u>:

O resultado do Lema 1, fica naturalmente verdadeiro, se f é restrita a um conjunto convexo $C\times O\subseteq X\times Y$, onde C é convexo em X e O é um convexo aberto em Y, tal que $\bar{y}\in O$. Neste caso, $M(\bar{y})$ fica definido por:

$$M(\bar{y}):=\left\{ x\in C/\ f(x,\bar{y})=\inf_{z\in C} f(z,\bar{y})\right\}.$$

<u>Lema 2</u>:

Sejam $\Omega: Y\longrightarrow X$, definida por $\Omega(y)=C$, $\forall y\in Y$, $f: C\times Y\longrightarrow \mathbb{R}$ tales que $M(\cdot)$ é a valores não vazios em uma vizinhança de um certo ponto $\bar{y}\in Y$. Então, se $f(\bar{x},\cdot)$ é semicontínua superior (s.c.s.) em relaçao de y no ponto $\bar{y}$, para um certo $\bar{x}\in M(\bar{y})$, a função marginal $v(\cdot)$ é s.c.s. no ponto $\bar{y}$.

<u>Demonstração</u>:

Seja $\left\{y^k\in Y, k\in\mathbb{N}\right\}$ uma sucessão convergente a $\bar{y}$. Sem perda de generalidade, suponhamos que $M(y^k)\neq\emptyset$, $\forall\ k\in\mathbb{N}$. temos então que:

$$v(y^k)=f(x^k,y^k)=\underbrace{f(x^k,y^k)-f(\bar{x},y^k)}_{\leq 0}+f(\bar{x},y^k),\ \forall\ k\in\mathbb{N},$$

onde x^k é um ponto qualquer em $M(y^k)$, $\forall\ k\in\mathbb{N}$, de maneira que:

$$v(y^k)\leq f(\bar{x},y^k),\ \forall\ k\in\mathbb{N},$$

logo temos:

$$\limsup_{k\longrightarrow\infty} v(y^k) \le \limsup_{k\longrightarrow\infty} f(\bar{x},y^k) \le f(\bar{x},\bar{y})=v(\bar{y}),$$

o qual prova o Lema.■ ■

<u>Teorema 1</u>:

Sejam X e Y dois espaços de Banach, C um subconjunto convexo fechado (não vazío) de X, $B(\bar{y})$ uma bola aberta em Y centrada em um certo ponto $\bar{y} \in Y$ e $f:C\times B(\bar{y}) \longrightarrow \mathbb{R}$, uma função numérica tal que:

i)$f(\cdot,\cdot)$ é convexa sobre $C \times B(\bar{y})$,

ii)$f(\cdot,\cdot)$ é contínua sobre $C\times\{\bar{y}\}$ ou mais geralmente, semicontínua inferior (s.c.i.) sobre $C\times\{\bar{y}\}$,

iii)$f(\cdot,\cdot)$ é parcialmente diferenciável em relação de y sobre $C\times B(\bar{y})$,

iv)a derivada parcial $D_y f(\cdot,\cdot)$ é contínua sobre $C\times\{\bar{y}\}$.

Nestas condições, se define-se $v(y):=\{\inf f(x,y)/x\in C\}$ e $M(y):=\left\{x\in C/f(x,y) = \inf_{z\in C} f(z,y)\right\}$, $\forall y\in B(\bar{y})$, então, sob a hipótese suplementaria:

H):$M(\cdot)$ é não vazía e localmente uniformemente compacta em torno de $\bar{y}$, a função $v(\cdot)$ é diferenciável no ponto $\bar{y}$, e além disso, tem-se:

$$Dv(\bar{y})=D_y f(\bar{x},\bar{y}),\ \forall\ \bar{x}\in M(\bar{y}),$$

onde,

$$D_y f(\bar{x},\bar{y}):=\frac{\partial}{\partial y} f(x,y)\Big|_{(x,y)=(\bar{x},\bar{y})}.$$

<u>Demonstração</u>:

Sob a hipótese ii) e a fechadura de C, tem-se que a multiaplicação $M(\cdot)$ é fechada (o que se conhece na literatura também como s.c.s.) no ponto $\bar{y}$. Agora consideremos uma sucessão qualquera $\left\{y^k\right\}_{k\in\mathbb{N}} \subset B(\bar{y})$, convergente a $\bar{y}$. Pela hipótese H), existem $\mathbb{N}'\subset \mathbb{N}$ e $\mathbb{N}''\subset \mathbb{N}$, com $\|\mathbb{N}'\|=\|\mathbb{N}''\|=+\infty$, e duas sucessões $\left\{x^k\right\}_{k\in\mathbb{N}'} \subset M(y^k)$ e $\left\{x^k\right\}_{k\in\mathbb{N}''} \subset M(y^k)$, ambas convergentes a $\bar{x}\in M(\bar{y})$ e $\tilde{x}\in M(y)$ respectivamente, tais que:

$$\lim_{\substack{k\to\infty \\ k\in\mathbb{N}'}} \frac{f(x^k,y^k)-f(\bar{x},\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} =$$

$$= \limsup_{\substack{k\to\infty \\ k\in\mathbb{N}}} \frac{v(y^k)-v(\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} \tag{VIII.1}$$

e,

$$\lim_{\substack{k\to\infty \\ k\in\mathbb{N}''}} \frac{f(x^k,y^k)-f(\bar{x},\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} =$$

$$\liminf_{\substack{k\to\infty \\ k\in\mathbb{N}}} \frac{v(y^k)-v(y)-D_y f(\tilde{x},\bar{y})(y^k-\bar{y})}{\|y^k-y\|} . \tag{VIII.2}$$

Por outra parte, como $x^k\in M(y^k)$, $\forall k\in\mathbb{N}'$, tem-se que:

$$\frac{f(x^k,y^k)-f(\bar{x},\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|}$$

$$\leq \frac{f(\bar{x},y^k)-f(\bar{x},\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} ,\ \forall k\in\mathbb{N}'.$$

Análogamente, já que $\bar{x}\in M(\bar{y})$, tem-se que:

$$\frac{f(x^k,y^k)-f(x^k,\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} \leq$$

$$\leq \frac{f(x^k,y^k)-f(\bar{x},\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} ,\ \forall k\in\mathbb{N}''.$$

Aplicando o Teorema do Valor Meio em relação á segunda variável da função f, e somando e restando termos adequados nos numeradores das expressões ezquerdas de (VIII.1) e (VIII.2) respectivamente, temos que:

$$\limsup_{\substack{k\to\infty \\ k\in\mathbb{N}}} \frac{v(y^k)-v(\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} \leq$$

$$\lim_{\substack{k\to\infty \\ k\in\mathbb{N}'}} \frac{[D_y f(\bar{x},y(\vartheta_k))-D_y f(\bar{x},\bar{y})](y^k-\bar{y})}{\|y^k-\bar{y}\|} \qquad \text{(VIII.3)}$$

Com, $y(\vartheta_k):=(1-\vartheta_k)\bar{y}+\vartheta_k y^k$, para um certo $\vartheta_k \in]0,1[$, $\forall k\in\mathbb{N}'$, e

$$\liminf_{\substack{k\to\infty \\ k\in\mathbb{N}}} \frac{v(y^k)-v(\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|} \geq$$

$$\lim_{\substack{k\to\infty \\ k\in\mathbb{N}''}} \frac{[D_y f(x^k,y(\vartheta_k))-D_y f(\bar{x},\bar{y})](y^k-\bar{y})}{\|y^k-\bar{y}\|} \qquad \text{(VIII.4)}$$

com, $y(\vartheta_k):=(1-\vartheta_k)\bar{y}+\vartheta_k y^k$, para um certo $\vartheta_k\in]0,1[$, $\forall k\in\mathbb{N}''$.

Por outra parte, já que:

$-\|A\|\|y\|\leq Ay\leq\|A\|\|y\|$, $\forall A\in\mathcal{L}(Y,\mathbb{R})$, $\forall y\in Y$, que não é mais que a desigualdade de Cauchy-Schwartz no caso de $Y=\mathbb{R}^n$, tem-se de (VIII.3) e (VIII.4) :

$$\lim_{\substack{k\to\infty \\ k\in\mathbb{N}'}} -\|[D_y f(x^k,y(\vartheta_k))-D_y f(\bar{x},\bar{y})]\| \leq \liminf_{\substack{k\to\infty \\ k\in\mathbb{N}}} \Delta(k) \leq$$

$$\leq \limsup_{\substack{k\to\infty \\ k\in\mathbb{N}'}} \Delta(k) \leq \lim_{\substack{k\to\infty \\ k\in\mathbb{N}'}} \|D_y f(\bar{x},y(\vartheta_k))-D_y f(\bar{x},\bar{y})\|,$$

onde,

$$\Delta(k):=\frac{v(y^k)-v(\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|}\,,\ \forall k\in\mathbb{N}.$$

Portanto, já que $y(\vartheta_k)\xrightarrow[k\to\infty]{k\in\mathbb{N}'}\bar{y}$, e

$y(\vartheta_k)\xrightarrow[k\to\infty]{k\in\mathbb{N}''}\bar{y}$, e usando o Lema 1 e hipótese iv),

temos que:

$$0=-\|D_y f(\tilde{x},\bar{y})-D_y f(\bar{x},\bar{y})\|\le\liminf_{\substack{k\to\infty\\k\in\mathbb{N}}}\Delta(k)\le$$

$$\le\limsup_{\substack{k\to\infty\\k\in\mathbb{N}}}\Delta(k)\le\|D_y f(\bar{x},\bar{y})-D_y f(\bar{x},\bar{y})\|=0,$$

é dizer,

$$\lim_{\substack{k\to\infty\\k\in\mathbb{N}}}\frac{v(y^k)-v(\bar{y})-D_y f(\bar{x},\bar{y})(y^k-\bar{y})}{\|y^k-\bar{y}\|}=0.$$

Como a sucessão $\langle y^k\in Y/\ k\in\mathbb{N}\rangle$ é convergente a $\bar{y}$ qualquer, então temos demonstrado que $Dv(\bar{y})=Df(\bar{x},\bar{y})_y$. Além disso, do Lema 1 tem-se que $D_y f(\bar{x},\bar{y})$ é uma constante como função de x sobre $M(\bar{y})$, o que prova o resultado.■ ■

<u>Observação 2</u>:

As hipóteses de fechadura de C e de continuidade de $f(\cdot,\cdot)$ sobre $C\times\langle\bar{y}\rangle$ podem ser reemplazadas de maneira natural pela hipótese, mais fraça, que $M(\cdot)$ seja fechada no ponto $\bar{y}$. Por outra parte, na hipótese iv), só é necessário supor que $D_y f(\cdot,\cdot)$ seja contínua sobre $M(\bar{y})\times\langle\bar{y}\rangle$. A seguir, vamos a considerar só uma parte do resultado do Teorema 1 no caso particular em que o espaço X é de dimensão finita.

<u>Corolario 1</u>:

Seja X um espaço normado de dimensão finita, Y um espaço de Banach qualquer, C um subconjunto (não vazio) convexo e fechado de X, $B(\bar{y})$ uma bola aberta em Y em torno de um ponto $\bar{y}\in Y$ e $f:C\times B(\bar{y})\longrightarrow\mathbb{R}$ uma função numérica tal que:

i) $f(\cdot,\cdot)$ é convexa sobre $C\times B(\bar{y})$,

ii) $f(\cdot,\cdot)$ é contínua sobre $C\times B(\bar{y})$ ou, mais geralmente, s.c.i. sobre $C\times\{\bar{y}\}$,

iii) $f(\cdot,\cdot)$ é parcialmente diferenciável em relação de y sobre $C\times B(\bar{y})$,

iv) a derivada parcial $D_y f(\cdot,\cdot)$ é contínua sobre $C\times B(\bar{y})$ ou, mais geralmente, sobre $M(\bar{y})\times B(\bar{y})$,

v) $f(\cdot,y)$ é s.c.i. em relação de x, $\forall y$ em uma certa vizinhança de $\bar{y}$.

Sob estas condições, se $M(\bar{y})$ é compacto e não vazío (é dizer, se $f(\cdot,\bar{y})$ é ínf-compacta em relação de x sobre C), a função $v(y):=\inf_{x\in C} f(x,y)$, é diferenciável no ponto $\bar{y}$, e além disto tem-se:

$$Dv(\bar{y})=D_y f(x,\bar{y}),\ \forall x\in M(\bar{y}).$$

<u>Demonstração</u>:

E' uma consequênça direta do Teorema 1, do fato que, sob a hipótese de que $M(\bar{y})$ é compacto não vazío, tem-se que $M(\cdot)$ é não vazía e localmente uniformemente compacta em torno de $\bar{y}$.

<u>Observação 3</u>:

Seja o problema de otimização convexa:

$P_1)$ Mín $f(x,y)$

s.a.

$(x,y)\in C\times D$

$C\subset X$

$D\subset Y$,

onde X e Y são dois espaços de Banach, C e D dois subconjuntos convexos de X e Y respectivamente e $f(\cdot,\cdot)$ uma função convexa contínua em relação de (x,y), contínuamente diferenciável respecto de y. Então, sob a hipótese H) do Teorema 1, se este problema é redefinido como um problema a dois estados:

$$P_2)\quad \underset{y\in D}{\text{Mín}} \left[\underset{x\in C}{\text{Mín}}\ f(x,y)\right],$$

o problema exterior que resulta, é não só um problema convexo, mas agora, diferenciável. Este é o caso, em particular, se X é de dimensão finita e $M(\bar{y})$ é não vazío e compacto (Corolário 1). Neste último caso, o problema interior será um problema convexo não diferenciável, mais em dimensão finita.

A seguir, desenvolvemos estas idéias no caso de um problema de minimização convexa com restrições. Concretamente, seja o problema convexo:

$$\begin{aligned} P)\ & \text{Mín } f(x,y) \\ & \text{s.a.} \\ & g(x,y)\leq 0 \\ & (x,y)\in C\times D \\ & C\times D \subset X\times Y, \end{aligned}$$

onde X e Y são dois espaços de Banach, $X=\mathbb{R}^n$, C um subconjunto fechado de X, D um subconjunto convexo de Y, $f(\cdot,\cdot)$ e $g(\cdot,\cdot)$ funções convexas contínuas em relação de (x,y) em $C\times D$, a valores em $\mathbb{R}^n$e $\mathbb{R}^m$ respectivamente, contínuamente diferenciáveis em relação de y em $X\times Y$.

Este problema pode ser redefinido como um problema convexo a dois estados, sob a forma:

$$P)\,\underset{y\in D}{\text{Mín}}\ v(y)$$

onde $v(y)$ é a função marginal associada à família de problemas parâmetrizados:

$$P)_y \quad \text{Mín } f(x,y)$$
$$\text{s.a.}$$
$$g(x,y)\leq 0$$
$$x\in C.$$

Portanto, se $X(y)$ representa o conjunto de soluções otimais do problema dual $P)$ e $U(y)$ o conjunto de soluçoes otimais do problema dual associado:

$$D)_y \quad \text{Máx } H(u,y)$$
$$u\geq 0,$$

onde $H(u,y):=\inf_{x\in C} f(x,y)+u^t g(x,y)$, temos o resultado sequinte:

<u>Corolário 2</u>.

Seja $\bar{y}\in D$ um valor de y tal que as restrições do problema $P)_y$ satisfaçem a condição de Regularidade de Slater, a saber, $\exists\ \tilde{x}\in C: g\ (\tilde{x},\bar{y})<0,\ \forall i\in\{1,\ldots\ldots,m\}$. Então, se $X(\bar{y})$ é compacto não vazío, existe uma vizinhança $V(\bar{y})$ de $\bar{y}$ tal que, $\forall y\in V(\bar{y})$, a derivada direcional de $v(\cdot)$, no ponto y, em uma direção qualquer $d\in Y$, está definida por:

$$\dot{v}(y;d)=\underset{u\in U(y)}{\text{Máx}}\ D_y L(x,u,y)d=\underset{u\in U(y)}{\text{Máx}}\left[D_y f(x,y)+\sum_{i=1}^{m} u_i D_y g_i(x,y)\right]d,$$

$\forall x\in X(y)$.

<u>Demonstração</u>:

Sob o Lema de Hogan, (CONTESSE(1986c,teorema4), com $D(y)=C\ \forall y\in Y$), as multiaplicações $X(\cdot)$ e $U(\cdot)$ são localmente uniformemente compactas e não vazías em torno de $\bar{y}$, e fechadas no ponto $\bar{y}$. Por outra parte, tem-se a igualdade:

$$v(y)=d(y):=\underset{u\geq 0}{\text{Sup}}\ H(u,y),$$

é dizer,

$$v(y)=\underset{u\geq 0}{\text{Sup}}\left[\inf_{x\in C} L(x,u,y)\right],$$

onde $L(x,u,y):=f(x,y)+u^Tg(x,y)$, $\forall y$ em uma certa vizinhança $V(\bar{y})$ de $\bar{y}$. Sem perda de generalidade, pode-se supor que $V(\bar{y})$ é suficientemente pequena, de maneira que se tenha: $X(y)$ e $U(y)$ compactos não vazios $\forall y\in V(\bar{y})$.

Logo, sob estas condições, $X(\cdot)$ e $U(\cdot)$ são localmente uniformemente compactos e não vazías no ponto y, $\forall y\in V(\bar{y})$.

Portanto, já que $\forall u\in\mathbb{R}^m_+$ (é dizer, $u\geq 0$), $L(x,u,y)$ é convexa em relação de (x,y), aplicando o Corolário 1, tem-se que $H(u,y)$ é diferenciável em relação de y sobre $\mathbb{R}^m_+ \times V(\bar{y})$ e:

$$D_yH(u,y)=D_yL(x,u,y),\quad \forall x\in X(y). \tag{VIII.5}$$

Agora, demonstraremos que $H(\cdot,\cdot)$ é ,de fato, contínuamente diferenciável em relação de y sobre $H^m_+ \times V(\bar{y})$. Seja $(\tilde{u},\tilde{y})\in \mathbb{R}^m_+\times V(\bar{y})$ e $\left\{(u^k,y^k)/\ k\in\mathbb{N}\right\}$ uma sucessão convergente a $(\tilde{u},\tilde{y})$, tal que $X(y^k)\neq\emptyset$, $\forall k$. Seja $\left\{x^k\in C/\ k\in\mathbb{N}\right\}$ uma sucessão tal que $x^k\in X(y^k)$, $\forall k$ (pode-se notar que estas sucessões estão bem definidas). Já que $X(\cdot)$ é localmente compacta no ponto $\tilde{y}$, existe uma subsucessão $\left\{x^k\in X(y^k)/k\in\mathbb{N}'\right\}$ convergente a, dizemos, $\tilde{x}\in C$. Logo, como $X(\cdot)$ é fechada no ponto $\tilde{y}$, tem-se que $\tilde{x}\in X(\tilde{y})$. Portanto:

$$\lim_{\substack{k\to\infty\\ k\in\mathbb{N}'}} D_yH(u^k,y^k)=\lim_{\substack{k\to\infty\\ k\in\mathbb{N}'}} D_yL(x^k,u^k,y^k)$$

$$\underset{\substack{\uparrow\\ D_yL(\cdot)\\ \text{contínua}}}{=} D_yL(\tilde{x},\tilde{u},\tilde{y}) \quad \underset{\substack{\uparrow\\ \tilde{x}\in X(\tilde{y})\\ \text{(ver (VIII.5))}}}{=} \quad D_yH(\tilde{u},\tilde{y}).$$

Agora demonstraremos que a sucessão original é também convergente a este límite. Pelo contrario, se não fora este o caso, existiría uma subsucessão $\left\{x^k / \ k\in\mathbb{N}''\right\}$, $\mathbb{N}''\subset\mathbb{N}\backslash\mathbb{N}'$, $\|\mathbb{N}''\|=+\infty$, e $\eta>0$ tais que:

$$\|D_yH(u^k,y^k)-D_yH(\tilde{u},\tilde{y})\|>\eta>0, \qquad \text{(VIII.6)}$$

onde $D_yH(u^k,y^k)=D_yL(x^k,u^k,y^k)$, $\forall k\in\mathbb{N}''$.

Agora, fazendo com a sucessão $\left\{x^k, k\in\mathbb{N}''\right\}$ o mesmo o que se fez com a sucessão original, existirá uma outra subsucessão $\left\{x^k / \ k\in\mathbb{N}'''\right\}$, $\mathbb{N}'''\subset\mathbb{N}''$, convergente a um certo $\hat{x}\in X(y)$, tal que:

$$\lim_{k\to\infty} D_yH(u^k,y^k) = D_yL(\hat{x},\tilde{u},\tilde{y})$$

$$= D_yH(\tilde{u},\tilde{y})$$

$$\uparrow$$

$(\hat{x}\in X(\tilde{y})$, Corolário 1),

o que é uma contradição com (VIII.6). Logo temos demonstrado que:

$$(u^k,y^k)\xrightarrow[k\to\infty]{k\in\mathbb{N}}(\tilde{u},\tilde{y})\Longrightarrow D_yH(u^k,y^k)\xrightarrow[k\to\infty]{k\in\mathbb{N}}D_yH(\tilde{u},\tilde{y}),$$

é dizer, $H(\cdot,\cdot)$ é contínuamente diferenciável em relação de y sobre $\mathbb{R}^m_+ \times V(\bar{y})$. Por outra parte, $H(\cdot,\cdot)$ é trivialmente contínua sobre este mesmo conjunto. Logo, já que $\forall\tilde{y}\in V(\bar{y})$, $U(\cdot)$ é localmente uniformemente compacta em torno de $\tilde{y}$, para estudar as propriedades de diferenciabilidade de $v(\cdot)$ no ponto $\tilde{y}$, pode-se supor, sem perda de generalidade, que $v(\cdot)$ é definida em torno de $\tilde{y}$ por:

$$v(y) = \operatorname*{Sup}_{u\in K\cap\mathbb{R}^m_+} H(u,y)$$

onde K é uma bola fechada em $\mathbb{R}^m$ (logo compacta) tal que

$U(y) \subset \text{int}(K)$, $\forall y$ em uma certa vizinhança de $\tilde{y}$. Neste contexto, podemos aplicar o Teorema de Danskin, HOGAN(1973), para ter finalmente que:

$$\dot{v}(\tilde{y};d) = \max_{u \in U(\tilde{y})} \left[D_y H(u,\tilde{y}) \right] d$$

Danskin

$$= \max_{u \in U(\tilde{y})} \left[D_y f(x,\tilde{y}) + \sum_{i=1}^{m} u_i D_y g_i(x,\tilde{y}) \right] d; \quad \forall x \in X(\tilde{y}),$$

Teorema 1

qualquer seja $\tilde{y} \in V(\bar{y})$. ∎∎

Observação 4:

Depois da demonstração do Corolário 2, , podemos ter o resultado seguinte:

Corolário 3:

Se no Teorema 1 [respectivamente no Corolário 1] considera-se que f é contínua sobre $C \times B(\bar{y})$, en vez de $C \times \{\bar{y}\}$, a função $v(\cdot)$ é então contínuamente diferenciável sobre uma certa vizinhança $V(\bar{y})$ de $\bar{y}$, e além disto tem-se:

$$Dv(\tilde{y}) = D_y f(x,\tilde{y}), \quad \forall x \in M(\tilde{y}), \quad \forall \tilde{y} \in V(\bar{y})$$

Demonstração:

Pelas hipóteses do teorema 1 [respectivamente do Corolário 1] modificadas como acima, existe uma vizinhanza $V(\bar{y})$ de $\bar{y}$, contida em $B(\bar{y})$, tal que, por uma parte, $M(\cdot)$ é não vazia e localmente uniformemente compacta em torno de $\bar{y}$ e, por outro lado:

$$Dv(\tilde{y}) = D_y f(x,\tilde{y}), \quad \forall x \in M(y), \quad \forall \tilde{y} \in V(\bar{y}).$$

Monstremos agora que $v(\cdot)$ é contínuamente diferenciável sobre $V(\bar{y})$. Seja $\tilde{y}\in V(\bar{y})$ e seja $\langle y^k / k\in\mathbb{N}\rangle$ uma sucessão convergente a $\tilde{y}$, tal que $M(y^k)\neq\emptyset$, $\forall k$. Seja $\langle x^k\in C / k\in\mathbb{N}\rangle$ uma sucessão tal que $x^k\in M(y^k)$, $\forall k\in\mathbb{N}$. Notemos que estas duas sucessões estão bem definidas. Já que $M(\cdot)$ é localmente compacta, existe uma subsucessão $\langle x^k / k\in\mathbb{N}''\rangle$ que é convergente a, dizemos, $\tilde{x}\in C$. Logo, do fato que $M(\cdot)$ é fechada no ponto $\tilde{y}$, tem-se que $\tilde{x}\in M(\tilde{y})$. Portanto:

$$\lim_{\substack{k\to\infty\\ k\in\mathbb{N}'}} Dv(y^k) \underset{\substack{\uparrow\\ x^k\in M(y^k),\ \forall k\in\mathbb{N}''}}{=} \lim_{\substack{k\to\infty\\ k\in\mathbb{N}'}} D_y f(x^k,y^k) \underset{\substack{\uparrow\\ D_y f(\ ,\ )\\ \text{contínua sobre}\\ C\times B(\bar{y})}}{=} D_y f(\tilde{x},\tilde{y}) \underset{\substack{\uparrow\\ \tilde{x}\in M(\tilde{y})}}{=} Dv(\tilde{y}).$$

Monstremos agora que a sucessão $\langle Dv(y^k) / k\in\mathbb{N}\rangle$ é também convergente a $Dv(\tilde{y})$. Suponhamos que isto não é verdadeiro. Então, existe uma subsucessão $\langle x^k / k\in\mathbb{N}''\rangle$, $\mathbb{N}''\subset\mathbb{N}\setminus\mathbb{N}'$, $|\mathbb{N}''| =+\infty$, e um real $\eta>0$ tais que

$$\|Dv(y^k)-Dv(\tilde{y})\|>\eta>0, \qquad \text{(VIII.7)}$$

onde $Dv(y^k)=D_y f(x^k,y^k)$, com $x^k\in M(y^k)$, $\forall k\in\mathbb{N}''$.

Agora se razonamos sobre a sucessão $\langle x^k / k\in\mathbb{N}''\rangle$ da mesma forma como se fez com a sucessão original, existirá uma outra subsucessão $\langle x^k / k\in\mathbb{N}'''\rangle$, $\mathbb{N}'''\subset\mathbb{N}''$, convergente a um certo $\hat{x}\in M(\tilde{y})$, tal que:

$$\lim_{\substack{k\to\infty\\ k\in\mathbb{N}'''}} Dv(y^k) = D_y f(\hat{x},\tilde{y}) \underset{\substack{\uparrow\\ \hat{x}\in M(\tilde{y}),\\ \text{[Teorema 1 ou Corolário 1]}}}{=} Dv(\tilde{y})$$

o que contradiz (VIII.7). Logo, temos demonstrado que:

$$y^k \xrightarrow[k\to+\infty]{k\in\mathbb{N}} \tilde{y} \Longrightarrow Dv(y^k) \xrightarrow[k\to+\infty]{k\in\mathbb{N}} Dv(\tilde{y})$$

é dizer, $v(\cdot)$ é continuamente diferenciável em $V(\bar{y})$.