



Universidade Estadual de Campinas
Instituto de Matemática Estatística e Computação Científica
DEPARTAMENTO DE MATEMÁTICA APLICADA



Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante

María Laura Schuverdt[†]

Doutorado em Matemática Aplicada - Campinas - SP

Orientador: Prof. Dr. José Mario Martínez

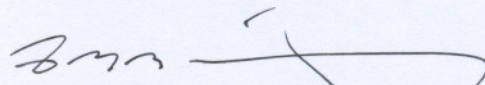
Co-orientador: Prof. Dr. Roberto Andreani

[†]Este trabalho teve apoio financeiro da FAPESP.

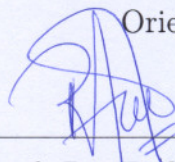
Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por **María Laura Schuverdt** e aprovada pela comissão julgadora.

Campinas, 8 de março de 2006.



Prof. Dr. **José Mario Martínez**
Orientador



Prof. Dr. **Roberto Andreani**
Co-orientador

Banca examinadora:

Prof. Dr. José Mario Martínez (IMECC/UNICAMP)

Prof. Dr. Alfredo Noel Iusem (IMPA)

Prof. Dr. Nelson Maculan Filho (COPPE/UFRJ)

Prof. Dr. Ernesto Julián Goldberg Birgin (IME/USP)

Profa. Dra. Sandra Augusta Santos (IMECC/UNICAMP)

Tese apresentada ao Instituto de Matemática Estatística e Computação Científica, UNICAMP, como requisito parcial para a obtenção do título de **Doutor em Matemática Aplicada**.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecário: Maria Júlia Milani Rodrigues – CRB8a / 2116

Schuverdt, Maria Laura

Sch77m Métodos de Lagrangiano aumentado com convergência utilizando a
condição de dependência linear positiva constante / Maria Laura Schuverdt --
Campinas, [S.P. :s.n.], 2006.

Orientadores : José Mario Martínez ; Roberto Andreani

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Computação Científica.

1. Programação não-linear. 2. Algoritmos. 3. Otimização matemática. I.
Martínez, José Mario. II. Andreani, Roberto. III. Universidade Estadual de
Campinas. Instituto de Matemática, Estatística e Computação Científica. IV.
Título.

Título em inglês: Augmented Lagrangian methods with convergence under the constant
positive linear dependence condition

Palavras-chave em inglês (Keywords): 1. Nonlinear programming. 2. Algorithms.
3. Mathematical optimization.

Área de concentração: Otimização

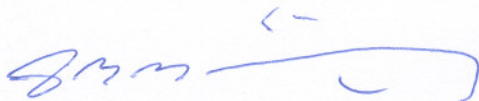
Titulação: Doutora em Matemática Aplicada

Banca examinadora: Prof. Dr. José Mario Martínez (IMECC-UNICAMP)
Prof. Dr. Alfredo Noel Iusem (IMPA-RJ)
Prof. Dr. Nelson Maculan Filho (COPPE-UFRJ)
Prof. Dr. Ernesto Julián Goldberg Birgin (IME-USP)
Profa. Dra. Sandra Augusta Santos (IMECC-UNICAMP)

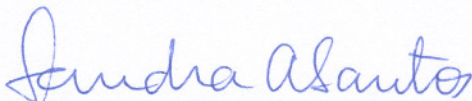
Data da defesa: 08/03/2006

Tese de Doutorado defendida em 08 de março de 2006 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



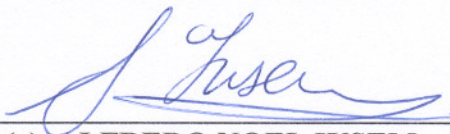
Prof. (a). Dr (a). JOSÉ MARIO MARTINEZ PEREZ



Prof. (a). Dr (a). SANDRA AUGUSTA SANTOS



Prof. (a). Dr (a). NELSON MACULAN FILHO



Prof. (a). Dr (a). ALFREDO NOEL IUSEM



Prof. (a) Dr. (a) ERNESTO JULIÁN GOLDBERG BIRGIN

Agradecimentos

Ao meu orientador José Mario Martínez por me receber no grupo, pela orientação exemplar, pela paciência e por me apresentar o Nino para trabalhar juntos.

Ao Nino, meu co-orientador, sem ele este trabalho não teria sido possível. Gracias por tu amistad, comprensión, paciencia y por todas las horas de charlas no solamente de matemática.

A mis padres y mis hermanas Ale y Ceci, por el apoyo incondicional, siempre, a pesar de los kilómetros. A toda mi familia rionegrina y a mi familia platense adoptiva: los Favero.

A Sandra Santos pela sua ajuda sempre e cada vez que eu precisei.

A la profesora Nelly Echebest, todos mis compañeros y profesores del Departamento de Matemática de la Universidad Nacional de la Plata por haberme enseñado a gustar de la matemática.

A Raúl y Valeria por la ayuda y el aguante de siempre y por brindarme todos los días su amistad. A Mirko por haber nacido y dejarme ser su amiga.

As Carmencitas por ter aberto as portas da sua casa e dos seus corações quando cheguei no Brasil e em especial a Giovana.

A todos meus amigos, os de lá e os daqui, que me deram a força necessária em todos os momentos e em especial a Lorian, Martín e Karime.

Aos meus colegas do doutorado, todos os professores do grupo de Otimização e demais professores e funcionários do IMECC-UNICAMP, que não citarei os nomes, mas que foram muito importantes neste período da minha vida.

Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, por fomentar este projeto, tendo um papel fundamental para a realização desta tese de doutorado.

Para finalizar, agradeço a todos vocês e a todas as pessoas que, direta ou indiretamente, me ajudaram a chegar até aqui.

Muchas gracias!

Resumo

Condições de qualificação são ferramentas úteis na análise de convergência de métodos de otimização. Neste trabalho provamos que a nova condição de dependência linear positiva constante (CPLD) é uma condição de qualificação e mostramos que ela é mais fraca que condições clássicas, como regularidade, Mangasarian-Fromovitz e posto constante. Além disso, apresentamos um algoritmo de Lagrangiano aumentado para resolver problemas gerais de programação matemática com convergência utilizando a CPLD. O algoritmo proposto é definido para resolver problemas com dois conjuntos de restrições: um, mais complexo, formado pelas restrições que são penalizadas e, outro, mais simples, pelas restrições que são satisfeitas por todos os iterados gerados no processo. O resultado de convergência global estabelece que se um ponto limite da sequência gerada pelo algoritmo satisfaz a condição CPLD então esse ponto é um ponto estacionário do problema original. O resultado de convergência global obtido é mais forte que resultados de convergência para problemas mais específicos obtidos utilizando condições de qualificação mais fortes, como a regularidade. Indicamos também as hipóteses adequadas sob as quais obtemos limitação do parâmetro de penalidade. A confiabilidade do algoritmo foi testada mediante uma exaustiva comparação com o algoritmo LANCELOT, mostrando que nosso método é mais robusto e eficiente. Além disso, e como aplicação do nosso algoritmo no caso em que restrições diferentes são incorporadas no problema, apresentamos a resolução de problemas de alocação nos quais existem muitas restrições não-lineares no conjunto complexo. Utilizando o método de Gradiente Projetado Espectral mostramos que problemas desse tipo com muitas variáveis e restrições são resolvidos de maneira eficiente num tempo razoável.

Palavras chave. Condições de qualificação, programação não-linear, métodos de Lagrangiano aumentado, convergência global, experimentos numéricos.

Abstract

Constraint qualifications are useful tools in the convergence analysis of optimization methods. In this work we prove that the new constant positive linear dependence condition (CPLD) is a constraint qualification and we show that it is weaker than classic constraint qualifications, like the regularity, the Mangasarian-Fromovitz and the constant rank conditions. Moreover, we introduce an augmented Lagrangian algorithm for solving general nonlinear programming problems whose convergence result uses the CPLD condition. The proposed algorithm is developed for problems with two sets of constraints: a complex one, formed by the penalized constraints and a simple one, formed by the constraints that are verified for all the iterates generated along the process. The global convergence result establishes that if a limit point of the sequence generated by the algorithm satisfies the CPLD condition then this point is a stationary point of the original problem. Thus, the global convergence result is stronger than the previous results for more specific problems obtained using stronger constraint qualification, as the regularity. We also indicate suitable conditions under which we prove boundedness of the penalty parameter. The reliability of the approach was tested by means of an exhaustive comparison against LANCELOT, demonstrating that our method is more robust and efficient. Moreover, as an application of our algorithm when different constraints are incorporated, we introduce the resolution of Location Problems in which there exist many nonlinear constraints in the complex set. We show that, employing the Spectral Projected Gradient method for solving the subproblems, this class of problems with many variables and constraints is efficiently solved with moderate computational effort.

Key words. Constraint qualification, nonlinear programming, augmented Lagrangian methods, global convergence, numerical experiments.

Sumário

Introdução	1
1 Condições de qualificação	7
1.1 Definições	8
1.2 Condições de otimalidade	17
1.2.1 O ponto de vista da penalidade externa	17
1.2.2 A condição AGP e relações	21
1.3 Penalidade externa e condições de qualificação	26
1.4 A condição CPLD implica a quase-normalidade	33
1.5 Relações entre as diferentes condições de qualificação	38
2 O método de Lagrangiano aumentado para resolver problemas gerais	45
2.1 O método de Lagrangiano Aumentado	46
2.1.1 A função Lagrangiano aumentado	48
2.1.2 O método proposto	50
2.2 Convergência para pontos viáveis	54
2.3 Convergência para pontos ótimos	59
2.4 Limitação do parâmetro de penalidade	62
2.4.1 Problema equivalente	62
2.4.2 Limitação para problemas com restrições de igualdade	65
2.4.3 Limitação no caso geral	77
2.5 Lagrangiano aumentado para problemas com restrições de canalização . . .	83
2.5.1 O método de Lagrangiano aumentado para restrições de canalização e a convergência global	84
2.5.2 Limitação dos parâmetros de penalidade	90
3 Experimentos numéricos	113
3.1 Restrições de canalização no conjunto simples	113
3.2 Restrições gerais no conjunto simples: o problema de alocação	117
Conclusões	123
Referências Bibliográficas	125

Introdução

Muitos problemas práticos de otimização podem ser formulados da seguinte maneira geral:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } x \in \Omega = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0\}, \end{aligned} \tag{1}$$

sendo $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ funções com derivadas primeiras contínuas sobre um conjunto aberto que contém Ω .

A solução global do problema (1) é um ponto $x_* \in \Omega$ tal que $f(x_*) \leq f(x)$ para todo $x \in \Omega$. Entretanto, calcular soluções globais é, freqüentemente, muito difícil e, mesmo reconhecer se um ponto $x \in \Omega$ é um minimizador global ou não é, em muitos casos, impossível. A maioria dos algoritmos populares de otimização não garantem a descoberta de minimizadores globais, eles se contentam em procurar candidatos a minimizadores, como pontos estacionários. Um ponto estacionário do problema (1) é um ponto $x_* \in \Omega$ para o qual existem escalares $\lambda_* \in \mathbb{R}^m$, $\mu_* \in \mathbb{R}_+^p$ tais que

$$\begin{aligned} \nabla f(x_*) + \sum_{i=1}^m [\lambda_*]_i \nabla h_i(x_*) + \sum_{j=1}^p [\mu_*]_j \nabla g_j(x_*) &= 0, \\ [\mu_*]_j g_j(x_*) &= 0, j = 1, \dots, p. \end{aligned}$$

Estas condições são conhecidas como condições de *Karush-Kuhn-Tucker* (condições KKT) e os escalares $[\lambda_*]_i$, $[\mu_*]_j$, como multiplicadores de Lagrange. Assim, quando um ponto viável satisfaz as condições KKT existe uma relação algébrica entre o gradiente da função objetivo e os gradientes das restrições.

A única dificuldade em relação às condições KKT é que nem todos os minimizadores locais as satisfazem. Para que isto aconteça é necessário que o ponto cumpra alguma condição de qualificação nas restrições. Assim, uma condição de qualificação é uma propriedade sobre os pontos viáveis de problemas de otimização que, quando satisfeita pelo minimizador global, garante que se satisfazem as condições KKT. Neste caso se garante a existência dos multiplicadores de Lagrange e algoritmos eficientes baseados em idéias de dualidade podem ser definidos.

Consideramos que boas condições de qualificação devem ser:

- não muito exigentes, ou seja, fracas;
- facilmente computáveis, preferencialmente mediante um cálculo algébrico;
- condições associadas com algum método ou algoritmo de programação matemática.

Existem várias condições de qualificação conhecidas na literatura: regularidade ou independência linear dos gradientes das restrições ativas, Mangasarian-Fromovitz [27], posto constante [24], quase-normalidade [23], pseudo-normalidade [6], entre outras.

Recentemente, em [32] Qi e Wei introduziram a condição de dependência linear positiva constante (CPLD) que demonstrou ser útil na análise de convergência de métodos de programação quadrática seqüencial. Um ponto viável satisfaz a condição CPLD se a dependência linear positiva de qualquer subconjunto dos gradientes das restrições ativas implica que esses gradientes continuam sendo positivo-linearmente dependentes na vizinhança do ponto. Logo, a verificação da condição CPLD requer somente o cálculo dos gradientes na vizinhança do ponto. Por outro lado, ela é mais fraca que a regularidade e que as condições de posto constante e de Mangasarian-Fromovitz. No trabalho onde a CPLD foi introduzida os autores conjecturaram que ela podia ser uma condição de qualificação. Esta questão foi apresentada como um problema em aberto na Seção 2 de [32]. Nesta tese provamos que a CPLD é de fato uma condição de qualificação e mostramos sua relação com algumas das condições de qualificação conhecidas na literatura associadas com problemas de otimização não convexa (ver [3]). Mostramos também como a condição de qualificação CPLD é utilizada na convergência global de um método de Lagrangiano aumentado.

O problema (1) pode ser definido de maneira mais genérica considerando dois conjuntos de restrições:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } x \in \Omega_1 = \{x \in \mathbb{R}^n \mid h_1(x) = 0, g_1(x) \leq 0\} \\ & \quad x \in \Omega_2 = \{x \in \mathbb{R}^n \mid h_2(x) = 0, g_2(x) \leq 0\}, \end{aligned} \tag{2}$$

sendo todas as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$, $h_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$, $g_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{p_1}$, $g_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{p_2}$ com derivadas primeiras contínuas sobre um conjunto aberto que contém $\Omega_1 \cap \Omega_2$.

Para um problema de otimização específico, a definição dos conjuntos viáveis Ω_1, Ω_2 pode ser feita de modo que o problema

$$\text{Minimizar } F(x) \text{ sujeita a } x \in \Omega_2 \tag{3}$$

seja consideravelmente mais fácil de ser resolvido que o problema (2) ou existe algum algoritmo eficiente para resolver esse problema que não pode ser aplicado ou deixa de ser eficiente quando restrições das do tipo do conjunto Ω_1 são adicionadas. Nesses casos, é natural resolver o problema original (2) mediante uma seqüência de problemas da forma (3) para alguma função adequada F , como sugere a idéia do método de Lagrangiano aumentado.

No algoritmo de Lagrangiano aumentado proposto nesta tese são penalizadas, utilizando a função de Powell-Hestenes-Rockafellar [22, 31, 33], todas as restrições do conjunto Ω_1 sendo que o processo garante que as restrições do conjunto Ω_2 são satisfeitas por todos os iterados gerados no processo. Assim, os conjuntos admissíveis Ω_1 e Ω_2 podem ser definidos de modo que as restrições mais fáceis de ser manipuladas, por exemplo restrições de canalização ou lineares, formem parte do conjunto Ω_2 , deixando as restrições mais difíceis de ser tratadas eficientemente, por exemplo as não-lineares, formando parte do conjunto Ω_1 .

O método de Lagrangiano aumentado consiste em uma seqüência de iterações externas. Em cada iteração externa, se F é a função Lagrangiano aumentado:

$$L(x, \lambda, \mu, \rho) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[[h_1(x)]_i + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[\max \left(0, [g_1(x)]_i + \frac{\mu_i}{\rho} \right) \right]^2 \right\}$$

um problema de minimização com restrições simples, da forma (3) é resolvido aproximadamente para obter o iterado atual:

$$\begin{aligned} &\text{Minimizar (aproximadamente)} \quad L(x, \lambda, \mu, \rho) \\ &\text{sujeita a} \quad x \in \Omega_2. \end{aligned}$$

Uma vez obtido este iterado, são atualizados os multiplicadores de Lagrange e os parâmetros de penalidade para logo repetir o processo. No algoritmo proposto utilizamos a projeção do estimado de primeira ordem como estimativa do multiplicador de Lagrange e aumentamos o parâmetro de penalidade quando a norma das restrições do conjunto Ω_2 não diminui com respeito a essa medida de inviabilidade no iterado anterior ([2]).

A maior contribuição teórica do método proposto é o uso da condição CPLD tanto na prova da viabilidade como na prova da convergência global:

1. Se x_* é um ponto limite da seqüência gerada pelo algoritmo introduzido então uma das seguintes possibilidades se verifica:
 - O ponto x_* é viável.
 - x_* é um ponto estacionário da soma dos quadrados das restrições h_1, g_1 do conjunto

Ω_1 , sujeita ao conjunto Ω_2 .

- x_* não verifica a condição CPLD associada ao conjunto Ω_2 .

2. Se x_* é um ponto limite viável então uma das seguintes propriedades se verifica:

- x_* é um ponto estacionário (KKT) do problema (2).
- x_* não verifica a condição CPLD associada ao conjunto viável original.

Observar que resultados de convergência global que utilizam a condição CPLD são mais fortes que resultados de convergência obtidos utilizando condições de qualificação mais fracas, como a regularidade.

Analisamos duas versões do algoritmo principal: com um único parâmetro para penalizar todas as restrições do conjunto Ω_1 ao mesmo tempo e com um parâmetro diferente para cada restrição. No primeiro caso provamos limitação da sequência de parâmetros de penalidade utilizando a redução ao caso com restrições de igualdade introduzida em [4]. As hipóteses utilizadas incluem, entre outras, a regularidade, a condição suficiente de segunda ordem e a condição de complementaridade estrita na solução, ver por exemplo [2]. No caso em que utilizamos um parâmetro diferente por restrição, se o conjunto Ω_1 é formado somente por restrições de igualdade e Ω_2 por restrições de canalização, provamos limitação da sequência de parâmetros de penalidade sem utilizar a condição de complementaridade estrita na solução (ver [1]).

Do ponto de vista prático e para analisar a confiabilidade do algoritmo proposto comparamos o desempenho do algoritmo com o LANCELOT [15] utilizando todos os problemas da coleção CUTE [13]. Neste caso utilizamos o algoritmo robusto GENCAN [9] na resolução dos subproblemas. Além disso, as aplicações mais interessantes do algoritmo proposto surgem quando o conjunto de restrições Ω_2 é arbitrário, possivelmente não convexo ou definido por restrições não-lineares. A família de problemas de alocação se apresenta como uma situação deste tipo. Utilizando o método de Gradiente Projetado Espectral [11] mostramos que problemas desse tipo com muitas variáveis e restrições são resolvidos de maneira eficiente num tempo razoável.

Este trabalho está organizado da seguinte maneira. No Capítulo 1, definimos algumas das condições de qualificação mais conhecidas na literatura e provamos que a condição de dependência linear positiva constante CPLD é uma condição de qualificação. Apresentamos também relações entre as diferentes condições de qualificação e com algumas condições de otimalidade.

No Capítulo 2, apresentamos o método de Lagrangiano aumentado proposto para resolver o problema (2) e provamos os resultados de convergência para pontos viáveis e

ótimos. Indicamos também as hipóteses adequadas sob as quais provamos limitação dos parâmetros de penalidade.

Finalmente, experimentos numéricos são apresentados no Capítulo 3.

Notações.

$B(x, \epsilon)$, vizinhança com centro em x e raio ϵ .

$[v]_i$, i -ésima componente do vetor v . Se não há possibilidade de confusão utilizamos a notação v_i .

Para $v \in \mathbb{R}^n$ e $a \in \mathbb{R}$, dizemos que $v > a$ (respectivamente $v \geq a$) se $[v]_i > a$ (respectivamente $[v]_i \geq a$) para todo $i = 1, \dots, m$.

$x = (x_1, \dots, x_n) \neq 0$ se existe ao menos um índice $i \in \{1, \dots, n\}$ para o qual $x_i \neq 0$.

$A_+ = \{t \in A \subset \mathbb{R} : t \geq 0\}$.

$A_{++} = \{t \in A \subset \mathbb{R} : t > 0\}$.

$\mathcal{P}_\Omega$ denota o operador projeção euclidiana sobre Ω .

$\|\cdot\| = \|\cdot\|_2$.

Se $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h = (h_1, \dots, h_m)$, denotamos $\nabla h(x) = [\nabla h_1(x) \dots \nabla h_m(x)] \in \mathbb{R}^{n \times m}$.

Capítulo 1

Condições de qualificação

Um problema geral de programação não-linear consiste em

$$\text{Minimizar } f(x)$$

sujeita a que se satisfaçam um conjunto de restrições de igualdade $h_i(x) = 0, i = 1, \dots, m$, e de desigualdade $g_j(x) \leq 0, j = 1, \dots, p$, sendo todas as funções contínuas, e em geral, deriváveis.

Gostaríamos de ter algoritmos de programação matemática que em todas as situações convergissem para a solução global do problema, mas calcular soluções globais é, frequentemente, muito difícil e, mesmo reconhecer se um ponto é minimizador global ou não, é, em muitos casos, impossível. Por isso, a grande maioria dos algoritmos que existem na literatura se contentam em procurar pontos estacionários (pontos que satisfazem as condições KKT) do problema geral de otimização. Os pontos estacionários são pontos candidatos a ser minimizadores do problema mas nem todos os minimizadores locais de problemas de otimização verificam as condições KKT. Para isto é necessário que o ponto satisfaça alguma condição de qualificação nas restrições. Logo, uma condição de qualificação é uma propriedade dos pontos viáveis de problemas de programação não-linear que, quando satisfeita por um minimizador local, garante que as condições KKT se verificam.

A condição de qualificação mais conhecida na literatura é a regularidade ou independência linear dos gradientes das restrições ativas. A regularidade joga um papel muito importante não somente em relação à obtenção de condições necessárias de otimalidade de problemas de programação não-linear mas também em relação à análise de convergência de algoritmos de otimização. Mas, a regularidade é uma condição muito restritiva: pode não ser satisfeita para alguns problemas simples. Por esta razão, continua-se procurando condições de qualificação mais fracas que a regularidade, que ao mesmo tempo tenham um papel relevante em relação à convergência de algoritmos de programação não-linear,

já que, certamente, se a condição de qualificação mais fraca é utilizada em um teorema de convergência, mais forte será o teorema. Assim, uma boa condição de qualificação deve ser fraca, ou seja, não exigir muito das restrições, deve ser simples de ser verificada e finalmente, é desejável que ela esteja associada a algum método eficiente de programação matemática. A investigação de novas condições de qualificação abre diferentes questões teóricas e práticas associadas com métodos de minimização.

1.1 Definições

Sejam $h : \mathbb{R}^n \rightarrow \mathbb{R}^m, g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ funções continuamente diferenciáveis, $h(x) = (h_1(x), \dots, h_m(x)), g(x) = (g_1(x), \dots, g_p(x))$. Definimos o conjunto viável X como:

$$X = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0\}.$$

Para cada ponto viável x , o conjunto de restrições de desigualdade ativas em x será denotado por

$$I(x) = \{i \in \{1, \dots, p\} \mid g_i(x) = 0\}.$$

Se $j \notin I(x)$, dizemos que a j -ésima restrição de desigualdade é inativa no ponto x .

Consideramos o problema geral de programação não-linear com conjunto viável X :

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeita a } x \in X \end{aligned} \tag{1.1}$$

para $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciável num conjunto aberto que contém X . A função f será chamada função objetivo.

Lembremos que um ponto x_* é um minimizador global do problema (1.1) se $f(x_*) \leq f(x)$ para todo $x \in X$. Dizemos que o ponto x_* é um minimizador local de (1.1) se existe $\epsilon > 0$ tal que $f(x_*) \leq f(x)$ para todo $x \in X \cap B(x_*, \epsilon)$, onde $B(x_*, \epsilon)$ denota uma vizinhança de x_* de raio ϵ .

Condições de otimalidade para problemas de programação matemática são condições necessárias para minimizador global, ou seja, são condições satisfeitas por todos os minimizadores globais do problema. Naturalmente, uma boa condição de otimalidade deve ser razoavelmente forte, assim ela fica mais próxima da condição ótima: a de ser minimizador global. Condições de otimalidade úteis devem ser de verificação simples, preferencialmente um cálculo mecânico. A condição de ser minimizador local é mais fraca que a condição de

ser minimizador global e, portanto, ela é uma condição de otimalidade, mas, é tão complicado verificar que um ponto é minimizador local quanto verificar se ele é minimizador global.

No caso de otimização sem restrições, se as derivadas parciais de f existem no minimizador local x_* , sabemos que $\nabla f(x_*) = 0$ é uma condição de otimalidade. No caso de otimização com restrições é necessário introduzir os denominados Multiplicadores de Lagrange para obter uma condição análoga àquela do caso de minimização sem restrições.

Definição 1.1.1. Dizemos que x_* é um ponto estacionário (ou KKT) do problema (1.1) se existem escalares $\lambda_* \in \mathbb{R}^m, \mu_* \in \mathbb{R}_+^p$ tais que

$$\nabla f(x_*) + \sum_{i=1}^m [\lambda_*]_i \nabla h_i(x_*) + \sum_{j \in I(x_*)} [\mu_*]_j \nabla g_j(x_*) = 0$$

$$h(x_*) = 0, g(x_*) \leq 0,$$

$$[\mu_*]_j g_j(x_*) = 0, j = 1, \dots, p.$$

Os escalares $[\lambda_*]_i, [\mu_*]_j$ da Definição 1.1.1 são denominados *Multiplicadores de Lagrange*.

Nem todo minimizador local de um problema de programação não-linear é um ponto estacionário, basta considerar, por exemplo, o problema: minimizar x sujeito a $x^2 = 0$, na solução $x_* = 0$.

Agora, quando um minimizador local satisfaz uma condição sobre as restrições podemos afirmar que as condições KKT verificam-se nesse ponto:

Definição 1.1.2. Uma condição de qualificação é uma propriedade sobre os pontos do conjunto viável X que, quando satisfeita por um minimizador local do problema, implica que se satisfazem as condições KKT neste ponto.

Cada Condição de Qualificação (CQ) está diretamente relacionada com uma condição de otimalidade: se x_* é um minimizador local de (1.1) então, ou ele é um ponto estacionário do problema ou ele não satisfaz a condição de qualificação CQ. Logo, a condição “ x_* é um ponto estacionário do problema ou x_* não satisfaz CQ” é uma condição de otimalidade. Conseqüentemente, como boas condições de otimalidade devem ser fortes, boas condições de qualificação devem ser fracas.

A condição de qualificação mais conhecida na literatura é a regularidade ou independência linear dos gradientes das restrições de igualdade e das restrições ativas no ponto:

Definição 1.1.3. [4] (LICQ) Dizemos que $x_* \in X$ é um ponto regular se os vetores $\{\nabla h_i(x_*)\}_{i \in \{1, \dots, m\}} \cup \{\nabla g_i(x_*)\}_{i \in I(x_*)}$ são linearmente independentes.

Quando um minimizador local é regular, além da existência, pode se garantir a unicidade dos multiplicadores de Lagrange na solução.

Existem casos nos quais as soluções não são pontos regulares mas eles satisfazem as condições KKT. Considerar, por exemplo, o problema: minimizar $(x-1)^2 + (y-1)^2$ sujeito a $h_1(x, y) = x + y$, $h_2(x, y) = -x - y$, na solução $x_* = (0, 0)$.

Problemas com restrições lineares de igualdade, como o do exemplo anterior, possuem uma propriedade notável: é sempre possível provar a existência dos multiplicadores de Lagrange na solução, mesmo para minimizadores locais que são não-regulares. Ver, por exemplo, [4]. Logo, o fato do conjunto admissível ser formado por restrições lineares de igualdade é uma condição de qualificação.

A seguinte condição de qualificação, introduzida por Mangasarian-Fromovitz em 1967, é mais fraca que a regularidade.

Definição 1.1.4. [27] (MFCQ) Dizemos que $x_* \in X$ satisfaz a condição de qualificação de Mangasarian-Fromovitz se os vetores $\{\nabla h_i(x_*)\}_{i \in \{1, \dots, m\}}$ são linearmente independentes e existe uma direção $d \in \mathbb{R}^n$ tal que

$$\nabla h_i(x_*)^T d = 0, \quad \forall i = 1, \dots, m, \quad \nabla g_j(x_*)^T d < 0, \quad \forall j \in I(x_*).$$

Quando um minimizador local satisfaz a condição MFCQ, não se pode garantir a unicidade dos multiplicadores de Lagrange na solução, mas pode se garantir que o conjunto formado por todos os multiplicadores de Lagrange é um conjunto compacto, ver [4], Exercício 3.3.19, pág. 354.

Observar que a condição MFCQ não se verifica quando as restrições de igualdade $h_i(x) = 0$ são substituídas por duas restrições de desigualdade ($h_i(x) \leq 0$ e $-h_i(x) \leq 0$), truque que é habitual fazer quando se quer considerar somente restrições de desigualdade. A seguinte condição de qualificação, introduzida por Janin em 1984, continua se verificando quando isto acontece.

Definição 1.1.5. [24] (CRCQ) Dizemos que $x_* \in X$ satisfaz a condição de qualificação de posto constante se existe uma vizinhança $B(x_*, \epsilon)$ de x_* tal que para todo $I_0 \subset \{1, \dots, m\}$, $J_0 \subset I(x_*)$, o conjunto de vetores gradientes

$$\{\nabla h_i(x)\}_{i \in I_0} \cup \{\nabla g_j(x)\}_{j \in J_0}$$

tem posto constante em $B(x_*, \epsilon)$.

Em outras palavras, x_* satisfaz a condição de posto constante se o posto de qualquer subconjunto dos gradientes das restrições que se satisfazem na igualdade não muda numa

vizinhança de x_* . É interessante observar que apesar da regularidade ser mais forte que ambas condições de qualificação MFCQ e CRCQ, elas não se implicam mutuamente.

Já vimos que a MFCQ não é satisfeita quando existem restrições do tipo $h(x) \leq 0, -h(x) \leq 0$ mas sim se verifica posto constante. Para mostrar que Mangasarian-Fromovitz também não implica posto constante consideremos o seguinte conjunto de restrições de desigualdade em $\mathbb{R}^2$: $g_1(x, y) = x + y^2 \leq 0, g_2(x, y) = x \leq 0$ na origem $x_* = (0, 0)$. Qualquer direção $d = (d_1, 0)$ com $d_1 < 0$ satisfaz a Definição 1.1.4, mas os gradientes são linearmente dependentes em x_* , e em toda vizinhança de x_* existe pelo menos um ponto onde os gradientes são linearmente independentes.

A condição de dependência linear positiva constante (CPLD) foi introduzida por Qi e Wei em [32] e utilizada para analisar algoritmos de programação quadrática seqüencial (SQP). O Teorema 4.2 de [32] estabelece que se um ponto limite de um método SQP geral satisfaz algumas condições que incluem a CPLD, então esse ponto é estacionário. Mas, os autores não provaram que a CPLD é uma condição de qualificação. Esta questão foi apresentada como um problema aberto na Seção 2 de [32]. Observa-se que se um problema de programação não-linear possui solução satisfazendo a CPLD (e as outras hipóteses do Teorema 4.2 de [32]) mas não satisfazendo KKT, essa solução não pode ser encontrada pelo algoritmo analisado. Portanto, é importante provar que a CPLD é de fato uma condição de qualificação, e eliminar a existência de tais minimizadores.

Primeiro definiremos o conceito de independência linear positiva.

Definição 1.1.6. [32] (PLD) *Sejam $A = \{a_1, \dots, a_m\}$, $B = \{b_1, \dots, b_p\}$ dois conjuntos finitos de vetores em $\mathbb{R}^n$. Dizemos que $A \cup B$ é positivo-linearmente dependente (PLD) se existem escalares $\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^p$ tais que $\beta \geq 0, (\alpha, \beta) \neq 0$ e*

$$\sum_{i=1}^m \alpha_i a_i + \sum_{j=1}^p \beta_j b_j = 0.$$

Caso contrário dizemos que $A \cup B$ é positivo-linearmente independente (PLI).

Claramente, como na dependência e independência linear de vetores, qualquer subconjunto de um conjunto PLI é sempre PLI e um conjunto que contém um subconjunto PLD é sempre PLD.

Assim, originalmente, a CPLD foi definida da seguinte maneira em [32]:

Definição 1.1.7. [32] (CPLD-original) *Dizemos que $x_* \in X$ satisfaz a condição CPLD-original se para todo $I_0 \subset \{1, \dots, m\}$, $J_0 \subset I(x_*)$, quando os gradientes $\{\nabla h_i(x_*)\}_{i \in J_0} \cup \{\nabla g_j(x_*)\}_{j \in I_0}$ são positivo-linearmente dependentes então, existe $B(x_*, \epsilon)$ vizinhança de x_**

tal que os gradientes $\{\nabla h_i(y)\}_{i \in J_0} \cup \{\nabla g_j(y)\}_{j \in I_0}$ são linearmente dependentes para todo $y \in B(x_*, \epsilon)$.

Observar que os autores requerem somente a dependência linear na vizinhança de x_* o que, em princípio, e como observam Qi e Wei em [32], é mais forte que o requerimento de dependência linear positiva na vizinhança de x_* :

Definição 1.1.8. Dizemos que $x_* \in X$ satisfaz a condição CPLD se para todo $I_0 \subset \{1, \dots, m\}$, $J_0 \subset I(x_*)$, quando os gradientes $\{\nabla h_i(x_*)\}_{i \in J_0} \cup \{\nabla g_j(x_*)\}_{j \in I_0}$ são positivo-linearmente dependentes então, existe $B(x_*, \epsilon)$ vizinhança de x_* tal que os gradientes $\{\nabla h_i(y)\}_{i \in J_0} \cup \{\nabla g_j(y)\}_{j \in I_0}$ são positivo-linearmente dependentes para todo $y \in B(x_*, \epsilon)$.

Claramente, como PLD implica dependência linear, CPLD implica CPLD-original. Mostraremos que, na verdade, ambas definições são equivalentes.

Podemos ver que as condições sobre as restrições definidas até aqui seguem uma linha: elas consideram somente os gradientes das restrições na análise. Existe porém outra linha de condições de qualificação fortemente mais teórica e que, além de analisar os gradientes, observa o valor numérico das funções na vizinhança do ponto.

Nosso aporte em relação às condições de qualificação surge da necessidade de provar não somente que a CPLD é uma condição de qualificação como também de mostrar o lugar que ela ocupa em relação às outras condições. Para isso precisamos definir a outra linha de condições de qualificação.

Com o propósito de estabelecer uma regra de multiplicador de Lagrange de primeira ordem para o problema (1.1) sem hipótese de qualificação sob as restrições, Hestenes [23] apresentou um resultado denominado Regra estendida do Multiplicador de Lagrange, no qual é introduzido um multiplicador adicional λ_0 associado com o gradiente da função objetivo:

Teorema 1.1.9. Regra estendida do Multiplicador de Lagrange:

Se x_* é um minimizador de f restrita ao conjunto X então existem multiplicadores $\lambda_0, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ tais que

(i) $\lambda_0, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ são não todos nulos,

(ii) $\lambda_0, \mu_1, \dots, \mu_p$ são não negativos,

(iii) $\lambda_0 \nabla f(x_*) + \sum_{i=1}^m \lambda_i \nabla h_i(x_*) + \sum_{j=1}^p \mu_j \nabla g_j(x_*) = 0$,

(iv) Se $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ são não todos nulos, então em toda vizinhança de x_* existe um ponto x tal que $\lambda_i h_i(x) > 0, \mu_j g_j(x) > 0$ para os multiplicadores λ_i, μ_j tais que $\lambda_i \neq 0, \mu_j \neq 0$.

Observações.

1. A tese do Teorema 1.1.9 é, na verdade, uma condição de otimalidade. Condições necessárias que envolvem um multiplicador não negativo associado ao gradiente da função objetivo são conhecidas como condições de Fritz-John, propostas inicialmente por John no artigo [25]. Na sua forma clássica, as condições foram estabelecidas considerando somente os pontos (i) – (iii) do Teorema 1.1.9. Ver [4], [30].
2. A condição (iv) do Teorema 1.1.9 diz que as restrições com multiplicador não nulo podem ser violadas arbitrariamente na vizinhança de x_* de maneira que o sinal da violação coincida com o sinal do correspondente multiplicador ($\lambda_i h_i(x) > 0, \mu_j g_j(x) > 0$). Além disso, se $g_j(x) \leq 0$ para todo x numa vizinhança de x_* então $\mu_j = 0$. Logo, em particular, se verifica a condição de complementaridade no minimizador local x_* :

$$\text{se } g_j(x_*) < 0 \text{ então } \mu_j = 0.$$

Portanto, a condição (iv) implica a condição de complementaridade e temos que, no caso em que $\lambda_0 \neq 0$, a Regra estendida do Multiplicador implica as condições KKT para o minimizador local x_* .

3. Os multiplicadores $\lambda_0, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ podem ser escolhidos tais que $\lambda_0 = 1$ se as condições (i) – (iv) não se verificam quando $\lambda_0 = 0$. Isto nos apresenta uma nova condição de qualificação:

Definição 1.1.10. [23] Dizemos que $x_* \in X$ é quase-normal se não existem escalares $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ tais que:

(i)

$$\sum_{i=1}^m \lambda_i \nabla h_i(x_*) + \sum_{j \in I(x_*)} \mu_j \nabla g_j(x_*) = 0.$$

(ii) $\mu_j \geq 0, \quad j \in I(x_*).$

(iii) $\lambda_1, \dots, \lambda_m, \mu_j, \quad j \in I(x_*)$ são não todos nulos.

(iv) Em toda vizinhança $B(x_*, \epsilon)$ de x_* existe um ponto $x \in B(x_*, \epsilon)$ tal que $\lambda_i h_i(x) > 0$ para todo i com $\lambda_i \neq 0$ e $\mu_j g_j(x) > 0$ para todo j com $\mu_j \neq 0$.

Evidentemente, se x_* é um minimizador local que é quase-normal, as condições do Teorema 1.1.9 não podem ser satisfeitas com $\lambda_0 = 0$, portanto, x_* satisfaz as condições KKT. Isto mostra que a quase-normalidade é uma condição de qualificação.

A Regra do Multiplicador de Lagrange foi estendida para problemas com o seguinte conjunto de viabilidade:

$$X = C \cap \{h_1(x) = 0, \dots, h_m(x) = 0\} \cap \{g_1(x) \leq 0, \dots, g_p(x) \leq 0\}$$

sendo C um conjunto abstrato de restrições. A extensão no caso em que o conjunto C é convexo fechado pode ser encontrada em [4] enquanto a extensão para o caso em que C é fechado mas não necessariamente convexo pode ser encontrada em [6]. A definição de quase-normalidade foi estendida para problemas com conjunto de restrições abstrato por Bertsekas e Ozdaglar [6, 7]. Os autores introduziram também uma outra condição de qualificação relacionada com a quase-normalidade e denominada pseudo-normalidade que é mais fácil de ser manipulada. Esta condição de qualificação é utilizada para analisar a relação entre a existência de multiplicadores de Lagrange e a admissão de função de penalidade exata para problemas de programação matemática com conjunto abstrato de restrições. Ver [6, 7] e [4, p. 337].

Definição 1.1.11. [6] Dizemos que $x_* \in X$ é pseudo-normal se não existem escalares $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ tais que:

(i)

$$\sum_{i=1}^m \lambda_i \nabla h_i(x_*) + \sum_{j \in I(x_*)} \mu_j \nabla g_j(x_*) = 0.$$

(ii) $\mu_j \geq 0, \quad j \in I(x_*).$

(iii) Em toda vizinhança $B(x_*, \epsilon)$ de x_* existe um ponto $x \in B(x_*, \epsilon)$ tal que

$$\sum_{i=1}^m \lambda_i h_i(x) + \sum_{j \in I(x_*)} \mu_j g_j(x) > 0.$$

Em [6] se demonstra que a pseudo-normalidade implica a quase-normalidade e se mostra um exemplo no qual a recíproca não é verdadeira.

Reescreveremos as Definições 1.1.4, 1.1.8, 1.1.10 e 1.1.11 em forma seqüencial. Para isto introduzimos a seguinte definição:

Definição 1.1.12. [3] Seja $x_* \in X$ com conjunto de restrições ativas $I(x_*)$.

1. Dizemos que x_* é um ponto MF-não-regular se existem escalares $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, $\mu_j \geq 0 \forall j \in I(x_*)$, $\sum_{i=1}^m |\lambda_i| + \sum_{j \in I(x_*)} \mu_j > 0$ tais que

$$\sum_{i=1}^m \lambda_i \nabla h_i(x_*) + \sum_{j \in I(x_*)} \mu_j \nabla g_j(x_*) = 0.$$

Em caso contrário dizemos que x_* é MF-regular.

2. *Propriedade seqüencial 1.* Dizemos que x_* satisfaz a propriedade seqüencial 1 associada aos escalares $\bar{\lambda}_1, \dots, \bar{\lambda}_m, \bar{\mu}_j, j \in I(x_*)$ se existe uma seqüência de pontos (não necessariamente viáveis) $y_k \in \mathbb{R}^n$ tal que $\lim_{k \rightarrow \infty} y_k = x_*$ e, para todo $i \in \{1, \dots, m\}$ tal que $\bar{\lambda}_i \neq 0$ e todo $j \in I(x_*)$ tal que $\bar{\mu}_j \neq 0$,

$$\bar{\lambda}_i h_i(y_k) > 0 \text{ e } \bar{\mu}_j g_j(y_k) > 0$$

para todo $k \in \mathbb{N}$.

3. *Propriedade seqüencial 2.* Dizemos que x_* satisfaz a propriedade seqüencial 2 associada aos escalares $\bar{\lambda}_1, \dots, \bar{\lambda}_m, \bar{\mu}_j, j \in I(x_*)$ se existe uma seqüência de pontos (não necessariamente viáveis) $y_k \in \mathbb{R}^n$ tal que $\lim_{k \rightarrow \infty} y_k = x_*$ e,

$$\sum_{j=1}^m \bar{\lambda}_i h_i(y_k) + \sum_{j \in I(x_*)} \bar{\mu}_j g_j(y_k) > 0. \quad (1.2)$$

Utilizando estas definições podemos reescrever as condições MFCQ, CPLD, quase-normalidade e pseudo-normalidade da seguinte forma.

Proposição 1.1.13. *Seja $x \in X$.*

- (a) *x satisfaz a condição de qualificação de Mangasarian-Fromovitz se e somente se x é MF-regular.*
- (b) *x satisfaz a condição CPLD se e somente se x é MF-regular ou, sendo MF-não-regular com escalares λ_i, μ_j o conjunto de vetores*

$$\{\nabla h_i(y) \mid \lambda_i \neq 0, i = 1, \dots, m\} \cup \{\nabla g_j(y) \mid \mu_j > 0, j \in I(x)\}$$

é positivo-linearmente dependente para todo y em alguma vizinhança de x (de pontos não necessariamente viáveis).

- (c) *x é quase-normal se e somente se x é MF-regular ou, sendo MF-não-regular com escalares $\lambda_1, \dots, \lambda_m, \mu_j, j \in I(x)$, x não satisfaz a propriedade seqüencial 1 associada aos escalares $\lambda_1, \dots, \lambda_m, \mu_j, j \in I(x)$.*
- (d) *x é pseudo-normal se e somente se x é MF-regular ou, sendo MF-não-regular com escalares $\lambda_1, \dots, \lambda_m, \mu_j, j \in I(x)$, x não satisfaz a propriedade seqüencial 2 associada aos escalares $\lambda_1, \dots, \lambda_m, \mu_j, j \in I(x)$.*

Prova. (a) Suponhamos que $x \in X$ satisfaz a condição MFCQ da Definição 1.1.4. Sejam $\lambda \in \mathbb{R}^m, \mu_j \geq 0, j \in I(x)$ tais que

$$\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I(x)} \mu_j \nabla g_j(x) = 0.$$

Multiplicando pela direção d da Definição 1.1.4 obtemos $\mu_j = 0, j \in I(x)$ e $\lambda_i = 0, i = 1, \dots, m$ o que prova que os gradientes $\{\nabla h_i(x)\}_{i=1, \dots, m} \cup \{\nabla g_j(x)\}_{j \in I(x)}$ são PLI. Logo, x é um ponto MF-regular.

Suponhamos agora que os gradientes $\{\nabla h_i(x)\}_{i=1, \dots, m} \cup \{\nabla g_j(x)\}_{j \in I(x)}$ são PLI. Em particular, temos que os gradientes $\{\nabla h_i(x)\}_{i=1, \dots, m}$ são linearmente independentes. Além disso temos que provar que existe uma direção d como na Definição 1.1.4. Consideremos o seguinte problema de programação linear nas variáveis (z, d) :

$$\begin{aligned} & \text{Minimizar } z \\ & \text{sujeita a } \nabla h_i(x)^T d = 0, i = 1, \dots, m \\ & \quad \nabla g_j(x)^T d \leq z, j \in I(x). \end{aligned} \tag{1.3}$$

Mostraremos que o problema é ilimitado, ou seja que não tem solução finita. Observar que este problema tem solução já que $(z, d) = (0, 0)$ é um ponto viável.

Suponhamos que exista uma solução (z, d) . Logo, pelas condições KKT para o problema de programação linear sabemos que existem escalares λ_i, μ_j tais que $\mu_j \geq 0$ para todo $j \in I(x)$ e

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sum_{i=1}^m \lambda_i \begin{pmatrix} \nabla h_i(x) \\ 0 \end{pmatrix} + \sum_{j \in I(x)} \mu_j \begin{pmatrix} \nabla g_j(x) \\ -1 \end{pmatrix} = 0.$$

Logo, pela segunda equação temos que

$$\sum_{j \in I(x)} \mu_j = 1,$$

e portanto, existem $\lambda \in \mathbb{R}^m, \mu_j \geq 0, j \in I(x)$ não todos nulos tais que

$$\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I(x)} \mu_j \nabla g_j(x) = 0$$

o que implica que os gradientes são PLD contradizendo a hipótese. Portanto o problema (1.3) é ilimitado e existe uma direção d que satisfaz a Definição 1.1.4.

As implicações (b), (c) e (d) são imediatas das definições. $\square$

1.2 Condições de otimalidade

1.2.1 O ponto de vista da penalidade externa

A teoria dos multiplicadores de Lagrange pode ser desenvolvida de várias formas; nesta subseção seguiremos as idéias básicas do ponto de vista da penalidade.

O método de penalidade externa se baseia em obter uma solução do problema original como limite de soluções de problemas irrestritos. Nesses subproblemas a função a ser minimizada é a soma da função objetivo com alguma penalidade da violação das restrições.

Seja $\mathcal{C} \subset \mathbb{R}^n$ um conjunto fechado e $f : \mathcal{C} \rightarrow \mathbb{R}^n, h : \mathbb{R}^n \rightarrow \mathbb{R}^m, g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ todas funções contínuas. Consideramos o seguinte problema de programação não-linear:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeita a } h(x) = 0, \ g(x) \leq 0, \ x \in \mathcal{C}. \end{aligned} \tag{1.4}$$

Para cada $j = 1, \dots, p, x \in \mathbb{R}^n$ definimos

$$g(x)_+ = \max\{0, g(x)\}.$$

Seja $\{\rho_k\}$ uma seqüência de números positivos tal que

$$\lim_{k \rightarrow \infty} \rho_k = \infty.$$

Para cada $k \in \mathbb{N}$ definimos

$$F_k(x) = f(x) + \frac{\rho_k}{2} \left[\|h(x)\|^2 + \|g_+(x)\|^2 \right].$$

O método de penalidade externa associado com a seqüência $\{\rho_k\}$ se define da seguinte maneira:

Para cada $k \in K$, calcular x_k como sendo a solução ótima do problema

$$\begin{aligned} & \text{Minimizar } F_k(x) \\ & \text{sujeita a } x \in \mathcal{C}. \end{aligned}$$

O seguinte teorema garante a convergência para minimizadores globais do método de penalidade externa. Ver [28].

Teorema 1.2.1. *Suponhamos que o problema (1.4) tem pelo menos um ponto viável. Suponhamos também que a seqüência $\{x_k\}$ gerada pelo método de penalidade externa está bem definida para todo $k \in \mathbb{N}$ e que ela admite um ponto limite x_* . Então, x_* é um*

minimizador global de (1.4).

Prova. Seja $K \subset \mathbb{N}$ um subconjunto infinito tal que

$$\lim_{k \in K} x_k = x_*.$$

Mostraremos primeiro que o ponto limite x_* é viável, ou seja, que $x_* \in \mathcal{C}$, $h(x_*) = 0$ e $g(x_*) \leq 0$.

Como $x_k \rightarrow x_*$, $x_k \in \mathcal{C}$ para todo k e $\mathcal{C}$ é fechado temos que $x_* \in \mathcal{C}$.

Suponhamos que

$$\|h(x_*)\|^2 + \|g(x_*)_+\|^2 > 0.$$

Logo,

$$\lim_{k \in K} F_k(x_k) = \lim_{k \in K} f(x_k) + \frac{\rho_k}{2} [\|h(x_k)\|^2 + \|g(x_k)_+\|^2] = \infty.$$

Seja z um ponto viável de (1.4). Logo,

$$f(z) + \frac{\rho_k}{2} [\|h(z)\|^2 + \|g(z)_+\|^2] = f(z) \quad \forall k \in \mathbb{N},$$

e, para $k \in K$ suficientemente grande, temos que

$$f(x_k) + \frac{\rho_k}{2} [\|h(x_k)\|^2 + \|g(x_k)_+\|^2] > f(z) + \frac{\rho_k}{2} [\|h(z)\|^2 + \|g(z)_+\|^2].$$

Isto não pode acontecer, pois o ponto x_k foi definido como um minimizador global do subproblema.

Portanto, x_* é um ponto viável.

Provaremos agora que x_* é um minimizador global de (1.4).

Seja y um ponto viável arbitrário de (1.4). Pela definição do método,

$$f(x_k) + \frac{\rho_k}{2} [\|h(x_k)\|^2 + \|g(x_k)_+\|^2] \leq f(y) + \frac{\rho_k}{2} [\|h(y)\|^2 + \|g(y)_+\|^2] = f(y).$$

para todo $k \in K$.

Logo,

$$f(x_k) \leq f(y) \quad \forall k \in K.$$

Tomando limite nesta última desigualdade, obtemos que $f(x_*) \leq f(y)$. Como y é um ponto viável arbitrário, temos que x_* é um minimizador global de (1.4) como queríamos provar. $\square$

Observar que no teorema anterior não é requerida nenhuma hipótese sobre diferencia-

bilidade da função objetivo nem das restrições.

Teorema 1.2.2. *Consideremos o problema (1.4) e suponhamos que além de contínuas, as funções f, h e g são diferenciáveis. Seja x_* um minimizador global do problema (1.4) e suponhamos que x_* é um ponto interior de $\mathcal{C}$. Então, existem seqüências $\{x_k\}$, $\{\lambda_k\} \subset \mathbb{R}^m$, $\{\mu_k\} \subset \mathbb{R}_+^p$, tais que:*

1. $\lim_{k \rightarrow \infty} x_k = x_*$,
2. $\lim_{k \rightarrow \infty} \nabla f(x_k) + \nabla h(x_k)\lambda_k + \nabla g(x_k)\mu_k = 0$,
3. $\lim_{k \rightarrow \infty} \|h(x_k)\| = \lim_{k \rightarrow \infty} \|g(x_k)_+\| = 0$,
4. Existe $k_0 \in \mathbb{N}$ tal que, para todo $k \geq k_0$,

$$[g(x_*)]_i < 0 \Rightarrow [\mu_k]_i = 0.$$

Prova. Como x_* é um ponto interior de $\mathcal{C}$ e ele é um minimizador local de (1.4), existe $\delta > 0$ tal que x_* é uma solução global do problema

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeita a } h(x) = 0, \ g(x) \leq 0, \ \|x - x_*\| \leq \delta. \end{aligned}$$

Portanto, x_* é a única solução global de:

$$\begin{aligned} &\text{Minimizar } f(x) + \frac{1}{2}\|x - x_*\|^2 \\ &\text{sujeita a } h(x) = 0, \ g(x) \leq 0, \ \|x - x_*\| \leq \delta. \end{aligned} \tag{1.5}$$

Consideremos a aplicação do método de penalidade externa ao problema (1.5). Cada subproblema consistirá em:

$$\begin{aligned} &\text{Minimizar } f(x) + \frac{1}{2}\|x - x_*\|^2 + \frac{\rho_k}{2}[\|h(x)\|^2 + \|g(x)_+\|^2] \\ &\text{sujeita a } \|x - x_*\| \leq \delta. \end{aligned} \tag{1.6}$$

Estes subproblemas consistem em minimizar uma função contínua em um conjunto compacto, portanto para cada $k \in \mathbb{N}$ existe uma solução global x_k . Como x_* é a única solução global de (1.5), o Teorema 1.2.1 de convergência do método de penalidade externa garante que

$$\lim_{k \rightarrow \infty} x_k = x_*,$$

o que prova que existe uma seqüência $\{x_k\}$ satisfazendo a condição 1. Além disso, por ser x_* um minimizador local de (1.4) temos que se verifica a condição 3. Claramente,

$\|x_k - x_*\| < \delta$ para todo k suficientemente grande. Logo, pelas condições de otimalidade de primeira ordem para o problema (1.6), que pode ser considerado irrestrito, sabemos que existe $k_1 \in \mathbb{N}$ tal que para todo $k \geq k_1$,

$$\nabla f(x_k) + \nabla h(x_k)[\rho_k h(x_k)] + \nabla g(x_k)[\rho_k g(x_k)_+] + (x_k - x_*) = 0. \quad (1.7)$$

Definimos $\lambda_k = \rho_k h(x_k)$ e $\mu_k = \rho_k g(x_k)_+ \geq 0$. Tomando limites na igualdade (1.7), obtemos que

$$\lim_{k \rightarrow \infty} \nabla f(x_k) + \nabla h(x_k)\lambda_k + \nabla g(x_k)\mu_k = 0$$

o que prova a condição 2.

Finalmente, se $[g(x_*)]_i < 0$ então existe k_0 tal que, para todo $k \geq k_0$ deve ser $[g(x_k)]_i < 0$ e por definição de $[\mu_k]_i$ temos que

$$[\mu_k]_i = \rho_k [g(x_k)_+]_i = \rho_k \max\{0, g_i(x_k)\} = 0.$$

Assim, se verifica a condição 4, o que prova a tese do teorema. $\square$

Observar que este teorema nos proporciona uma condição de otimalidade seqüencial para o problema (1.1), no qual $\mathcal{C} = \mathbb{R}^n$:

Definição 1.2.3. *Um ponto $x_* \in \mathbb{R}^n$ satisfaz a condição de otimalidade AKKT-1 se existem seqüências $(x_k, \lambda_k, \mu_k, \epsilon_k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$ tais que*

$$\begin{aligned} x_k &\rightarrow x_* \\ \|\nabla f(x_k) + \nabla h(x_k)\lambda_k + \nabla g(x_k)\mu_k\| &\leq \epsilon_k \\ [\mu_k]_i &\geq 0, \quad i = 1, \dots, p \\ \text{se } g_i(x_k) < 0 \text{ então } [\mu_k]_i &= 0, \quad i = 1, \dots, p \\ \|h(x_k)\| &\leq \epsilon_k, \|g(x_k)_+\| \leq \epsilon_k, \end{aligned}$$

e $\epsilon_k \rightarrow 0$.

Entretanto, a condição associada com a complementaridade na definição acima (quarta condição) é muito exigente. Considerar por exemplo o problema

$$\begin{aligned} &\text{Minimizar } x \\ &\text{sujeita a } -x \leq 0, \end{aligned}$$

na solução $x_* = 0$. Suponhamos que um algoritmo gera uma seqüência $\{x_k\}$ que converge para x_* com $x_k > 0$ para todo k . Logo, se a condição de complementaridade da Definição 1.2.3 é satisfeita em cada x_k , o multiplicador μ_k associado deveria ser nulo. Mas nesse

caso $\nabla f(x_k) = 1$ e portanto a desigualdade $\|\nabla f(x_k) + \nabla h(x_k)\lambda_k + \nabla g(x_k)\mu_k\| \leq \epsilon_k$ não se cumpriria para nenhum $\epsilon_k < 1$.

Seria interessante então relaxar essa condição por uma menos exigente. Consideramos assim a seguinte condição de otimalidade mais fraca que a condição AKKT-1:

Definição 1.2.4. *Um ponto $x_* \in \mathbb{R}^n$ satisfaz a condição de otimalidade AKKT-2 se existem seqüências $(x_k, \lambda_k, \mu_k, \epsilon_k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$ tais que*

$$\begin{aligned} x_k &\rightarrow x_* \\ \|\nabla f(x_k) + \nabla h(x_k)\lambda_k + \nabla g(x_k)\mu_k\| &\leq \epsilon_k \\ [\mu_k]_i &\geq 0, \quad i = 1, \dots, p \\ \text{se } g_i(x_k) < -\epsilon_k &\text{ então } [\mu_k]_i = 0, \quad i = 1, \dots, p \\ \|h(x_k)\| &\leq \epsilon_k, \|g(x_k)_+\| \leq \epsilon_k, \end{aligned}$$

e $\epsilon_k \rightarrow 0$.

Claramente a condição AKKT-2 implica a condição AKKT-1. Resta perguntar se essas são condições de otimalidade razoavelmente fortes. Para isto analisaremos a relação com outra conhecida condição de otimalidade seqüencial.

1.2.2 A condição AGP e relações

A condição AGP (“Approximate Gradient Projection”) introduzida por Svaiter-Martínez em [29] é uma condição de otimalidade seqüencial estritamente mais forte que a conhecida condição de Fritz-John [27].

Definição 1.2.5. [29] *Dizemos que um ponto $x_* \in \mathbb{R}^n$ satisfaz a condição AGP(γ) (para $\gamma \in [-\infty, 0]$) se existe uma seqüência $\{x_k\}$ tal que $x_k \rightarrow x_*$*

$$d(x_k, \gamma) = \mathcal{P}_{\Omega(x_k, \gamma)}(x_k - \nabla f(x_k)) - x_k \rightarrow 0$$

sendo $\Omega(x, \gamma)$ o poliedro definido pelos pontos $z \in \mathbb{R}^n$ tais que:

$$g_i(x) + \nabla g_i(x)^T(z - x) \leq 0 \quad \text{se} \quad \gamma < g_i(x) < 0, \quad (\text{se } \gamma \neq 0)$$

$$\nabla g_i(x)^T(z - x) \leq 0 \quad \text{se} \quad g_i(x) \geq 0$$

$$h_i(x) + \nabla h_i(x)^T(z - x) = 0, \quad i = 1, \dots, m.$$

No Teorema 1 de [29] os autores provaram que se x_* é um minimizador local de (1.1) então se satisfaz a condição AGP(γ), o que mostra que a condição AGP(γ) é uma condição de otimalidade para o problema (1.1).

É interessante observar que a prova do Teorema 1 de [29] baseia-se na técnica de penalidade externa utilizada na prova do Teorema 1.2.2 (ver também a Proposição 3.3.5 de [4]).

Observar que, geometricamente, a condição $AGP(\gamma)$ diz que, se x_* é um minimizador local então existe uma seqüência convergindo para x_* para a qual o gradiente projetado sobre as aproximações lineares das restrições ativas e não ativas converge para zero.

O seguinte teorema mostra as relações entre as condições de otimalidade 1.2.3, 1.2.4 e 1.2.5.

Teorema 1.2.6. *Seja $x_* \in X$. Suponhamos que as funções f, h e g são continuamente diferenciáveis.*

- (a) *Se x_* satisfaz a condição $AGP(\gamma)$ então x_* satisfaz a condição $AKKT-2$.*
- (b) *Se x_* satisfaz a condição $AKKT-1$ então x_* satisfaz a condição $AGP(\gamma)$.*
- (c) *$AGP(0)$ é equivalente a $AKKT-1$.*

Prova.

(a) Seja $\{x_k\}$ uma seqüência convergindo para x_* que satisfaz a condição $AGP(\gamma)$. Seja $y_k = \mathcal{P}_{\Omega(x_k, \gamma)}(x_k - \nabla f(x_k))$. Logo, y_k é solução do seguinte problema de otimização:

$$\begin{aligned} &\text{Minimizar } \|y - (x_k - \nabla f(x_k))\|^2 \\ &\text{sujeita a } y \in \Omega(x_k, \gamma). \end{aligned}$$

Pelas condições KKT deste problema sabemos que existem $v_k \in \mathbb{R}^m, u_k \in \mathbb{R}_+^p$ tais que

$$\begin{aligned} \nabla f(x_k) + y_k - x_k + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i=1}^p [u_k]_i \nabla g_i(x_k) &= 0 \\ [u_k]_i (g_i(x_k) + \nabla g_i(x_k)^T (y_k - x_k)) &= 0 \quad \text{se } \gamma < g_i(x_k) < 0 \\ [u_k]_i (\nabla g_i(x_k)^T (y_k - x_k)) &= 0 \quad \text{se } g_i(x_k) \geq 0 \\ [u_k]_i &= 0 \quad \text{em outro caso.} \end{aligned} \tag{1.8}$$

Definimos

$$[\delta_k]_i = \begin{cases} \nabla g_i(x_k)^T (x_k - y_k) & \text{se } \gamma < g_i(x_k) < 0 \text{ (se } \gamma \neq 0) \\ g_i(x_k) + \nabla g_i(x_k)^T (x_k - y_k) & \text{se } g_i(x_k) \geq 0 \\ 0 & \text{em outro caso,} \end{cases}$$

e

$$\epsilon_k = \max\{\|x_k - y_k\|, \|h(x_k)\|, \|g(x_k)_+\|, \|\delta_k\|\}.$$

Mostraremos que a seqüência $(x_k, v_k, u_k, \epsilon_k)$ satisfaz a Definição 1.2.4.

Por (1.8) e definição de ϵ_k temos que

$$\|\nabla f(x_k) + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i=1}^p [u_k]_i \nabla g_i(x_k)\| = \|x_k - y_k\| \leq \epsilon_k,$$

e

$$\|h(x_k)\| \leq \epsilon_k, \quad \|g(x_k)_+\| \leq \epsilon_k.$$

Além disso, se $g_i(x_k) < -\epsilon_k \leq -[\delta_k]_i$ temos que $g_i(x_k) + [\delta_k]_i < 0$ e pela definição de $[\delta_k]_i$, $g_i(x_k) + [\delta_k]_i = g_i(x_k) + \nabla g_i(x_k)^T(x_k - y_k) < 0$ o que, pelas condições KKT (1.8), implica em $[u_k]_i = 0$.

Assim, somente resta provar que $\epsilon_k \rightarrow 0$.

Pela condição AGP(γ) sabemos que $\|x_k - y_k\| \rightarrow 0$, e pela viabilidade de x_* temos que $\|h(x_k)\| \rightarrow 0$ e $\|g(x_k)_+\| \rightarrow 0$. Vejamos que $[\delta_k]_i \rightarrow 0$ para todo $i = 1, \dots, p$.

Como a função g_i é continuamente diferenciável existem k_0 e uma constante C tais que, para todo $k \geq k_0$

$$\|\nabla g_i(x_k)\| \leq C. \quad (1.9)$$

Se $g_i(x_*) < 0$ temos que existe $k_1 \geq k_0$ tal que $g_i(x_k) < 0, \forall k \geq k_1$ e então, por definição temos que $[\delta_k]_i \leq |\nabla g_i(x_k)^T(x_k - y_k)| \leq \|\nabla g_i(x_k)\| \|x_k - y_k\| \leq C \|x_k - y_k\| \rightarrow 0$ quando $k \rightarrow \infty$.

Suponhamos que $g_i(x_*) = 0$. Se existem infinitos índices k para os quais $g_i(x_k) \geq 0$, por definição de $[\delta_k]_i$ e pela continuidade de ∇g_i , temos que $[\delta_k]_i \leq |g_i(x_k) + \nabla g_i(x_k)^T(x_k - y_k)| \leq \|g_i(x_k)\| + \|\nabla g_i(x_k)\| \|x_k - y_k\| \leq \|g_i(x_k)\| + C \|x_k - y_k\|$. Logo, $[\delta_k]_i \leq \min\{C \|x_k - y_k\|, \|g_i(x_k)\| + C \|x_k - y_k\|\} \rightarrow 0$, quando $k \rightarrow \infty$.

Portanto temos que $\epsilon_k \rightarrow 0$ e a sequência $(x_k, v_k, u_k, \epsilon_k)$ satisfaz a Definição 1.2.4, o que prova que x_* é um ponto AKKT-2.

(b) Seja $\{x_k\}$ uma sequência convergindo para x_* e que satisfaz a condição AKKT-1. Logo, pela Definição 1.2.3 temos que existem $v_k \in \mathbb{R}^m, u_k \in \mathbb{R}_+^p, \epsilon_k \in \mathbb{R}_+$ tais que $\epsilon_k \rightarrow 0$

e

$$\left\| x_k - \nabla f(x_k) - \left[x_k + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i: g_i(x_k) \geq 0} [u_k]_i \nabla g_i(x_k) \right] \right\| \leq \epsilon_k.$$

Então,

$$\left\| \mathcal{P}_{\Omega(x_k, \gamma)}(x_k - \nabla f(x_k)) - \mathcal{P}_{\Omega(x_k, \gamma)} \left(x_k + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i: g_i(x_k) \geq 0} [u_k]_i \nabla g_i(x_k) \right) \right\| \leq \epsilon_k. \quad (1.10)$$

Vejamos que $x_k = \mathcal{P}_{\Omega(x_k, \gamma)}(x_k + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i: g_i(x_k) \geq 0} [u_k]_i \nabla g_i(x_k))$. Consideremos

o seguinte problema de programação não-linear

$$\begin{aligned} & \text{Minimizar } \|y - (x_k + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i: g_i(x_k) \geq 0} [u_k]_i \nabla g_i(x_k))\|^2 \\ & \text{sujeito a } y \in \Omega(x_k, \gamma). \end{aligned}$$

Observemos que $y_k = x_k$ satisfaz as condições KKT deste problema para os multiplicadores escolhidos como $\lambda_k = v_k$ para as restrições de igualdade e $[\mu_k]_i = [u_k]_i$ se $g_i(x_k) \geq 0$, $[\mu_k]_i = 0$ no outro caso, para as restantes restrições de desigualdade. Logo, por (1.10) temos que $\|\mathcal{P}_{\Omega(x_k, \gamma)}(x_k - \nabla f(x_k)) - x_k\| \leq \epsilon_k \rightarrow 0$, como queríamos provar.

(c) Por (b) temos que, se x_* satisfaz a condição AKKT-1 então, em particular, x_* satisfaz AGP(0).

Resta-nos somente provar a recíproca. Como em (a), definimos $y_k = \mathcal{P}_{\Omega(x_k, \gamma)}(x_k - \nabla f(x_k))$ solução do seguinte problema de otimização:

$$\begin{aligned} & \text{Minimizar } \|y - (x_k - \nabla f(x_k))\|^2 \\ & \text{sujeito a } y \in \Omega(x_k, \gamma). \end{aligned}$$

Pelas condições KKT deste problema sabemos que existem $v_k \in \mathbb{R}^m, u_k \in \mathbb{R}_+^p$ tais que

$$\begin{aligned} \nabla f(x_k) + y_k - x_k + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i=1}^p [u_k]_i \nabla g_i(x_k) &= 0 \\ [u_k]_i (\nabla g_i(x_k))^T (y_k - x_k) &= 0 \quad \text{se } g_i(x_k) \geq 0 \\ [u_k]_i &= 0 \quad \text{em outro caso.} \end{aligned} \tag{1.11}$$

Definimos então

$$[\delta_k]_i = \begin{cases} g_i(x_k) + \nabla g_i(x_k)^T (x_k - y_k) & \text{se } g_i(x_k) \geq 0, \\ 0 & \text{c. c.,} \end{cases}$$

e

$$\epsilon_k = \max\{\|x_k - y_k\|, \|h(x_k)\|, \|g(x_k)_+\|, \|\delta_k\|\}.$$

Logo, por (1.11) temos que

$$\|\nabla f(x_k) + \sum_{i=1}^m [v_k]_i \nabla h_i(x_k) + \sum_{i=1}^p [u_k]_i \nabla g_i(x_k)\| = \|x_k - y_k\| \leq \epsilon_k,$$

e pela definição de ϵ_k temos que

$$\|h(x_k)\| \leq \epsilon_k, \quad \|g(x_k)_+\| \leq \epsilon_k.$$

Como em (a), para provar que $\epsilon_k \rightarrow 0$ temos que mostrar que $[\delta_k]_i \rightarrow 0$ para todo $i = 1, \dots, p$.

Se $g_i(x_*) < 0$ temos que existe k_0 tal que $g_i(x_k) < 0, \forall k \geq k_0$ e então, pela definição temos que $[\delta_k]_i = 0$.

Suponhamos que $g_i(x_*) = 0$. Se existem infinitos índices para os quais $g_i(x_k) \geq 0$, pela definição de $[\delta_k]_i$ e pela continuidade de ∇g_i , temos que $|[\delta_k]_i| \leq |g_i(x_k) + \nabla g_i(x_k)^T(x_k - y_k)| \leq \|g_i(x_k)\| + \|\nabla g_i(x_k)\| \|x_k - y_k\| \leq \|g_i(x_k)\| + C\|x_k - y_k\|$ para C definida em (1.9). Logo, $|[\delta_k]_i| \leq \|g_i(x_k)\| + C\|x_k - y_k\| \rightarrow 0$ quando $k \rightarrow \infty$.

Portanto temos que $\epsilon_k \rightarrow 0$. Logo, a seqüência $(x_k, v_k, u_k, \epsilon_k)$ satisfaz a Definição 1.2.3 e x_* é um ponto AKKT-1. $\square$

Apresentamos a seguir os contra-exemplos que mostram que as recíprocas das implicações (a) e (b) do teorema anterior não se cumprem.

Consideremos o seguinte problema de otimização com restrições de complementaridade em $\mathbb{R}^2$:

$$\begin{aligned} &\text{Minimizar } -x - y \\ &\text{sujeito a } xy = 0. \end{aligned}$$

Consideremos a seqüência $(x_k, y_k) = (1/k, 1/k)$ convergindo para o ponto $x_* = (0, 0)$ (que não é a solução do problema). Para $\lambda_k = k$ e $\epsilon_k = 1/k^2$ temos que a seqüência $((x_k, y_k), \lambda_k, \epsilon_k)$ satisfaz a condição AKKT-2.

Vejamus que nenhuma seqüência satisfaz a condição AGP(γ). Como temos só uma restrição de igualdade, para qualquer seqüência $\{(x_k, y_k)\} \subset \mathbb{R}^2$, temos que $\Omega((x_k, y_k), \gamma) = \{z = (z_1, z_2) \in \mathbb{R}^2 : x_k y_k + y_k(z_1 - x_k) + x_k(z_2 - y_k) = 0\}$. Logo,

$$d(x_k, \gamma) = \begin{pmatrix} 1 - \left(\frac{x_k y_k + x_k + y_k}{x_k^2 + y_k^2}\right) y_k \\ 1 - \left(\frac{x_k y_k + x_k + y_k}{x_k^2 + y_k^2}\right) x_k \end{pmatrix}$$

que não converge para 0 quando $(x_k, y_k) \rightarrow (0, 0)$. Portanto, o ponto $x_* = (0, 0)$ não satisfaz a condição AGP(γ), o que mostra que não se cumpre a recíproca de (a).

Consideremos agora o problema:

$$\begin{aligned} &\text{Minimizar } x \\ &\text{sujeito a } -x^2 \leq 0. \end{aligned}$$

Para a seqüência $x_k = 1/k$ temos que, para todo $\gamma \in [-\infty, 0)$, $\Omega(x_k, \gamma) = \{z : z \geq 1/2k\}$. Logo, para k suficientemente grande, $\|d(x_k, \gamma)\| \leq 1/k \rightarrow 0$ e portanto $x_* = 0$ satisfaz a condição AGP(γ).

Vejamos que não existe nenhuma sequência satisfazendo a condição AKKT-1. Seja $\{x_k\}$ é uma sequência tal que $x_k \rightarrow 0$ (que não é a solução do problema) e $x_k \neq 0$ para todo k suficientemente grande. Pela condição de complementaridade temos $\mu_k = 0$ e nesse caso, como $\nabla f = 1$, nunca acontece $\|\nabla f(x_k) + \nabla g(x_k)\mu_k\| \leq \epsilon_k$ para $\epsilon_k < 1$. Logo, nenhuma sequência satisfaz a Definição 1.2.3 e x_* não satisfaz a condição AKKT-1. Isto mostra que não se cumpre a recíproca de (b).

1.3 Penalidade externa e condições de qualificação

Mostraremos agora como o Teorema 1.2.2, e em particular a técnica de penalidade externa, podem ser utilizadas para provar que as condições definidas na Subseção 1.1 são efetivamente condições de qualificação. Em particular mostraremos que a condição CPLD é uma condição de qualificação, o resultado mais importante desta subseção.

O seguinte lema pode ser encontrado em [4], (pág. 689, Exercício B.1.7) e será utilizado na prova de que CRCQ e CPLD são condições de qualificação.

Lema 1.3.1. *Lema de Caratheodory. Sejam $u_1, \dots, u_m, v_1, \dots, v_p$ vetores em $\mathbb{R}^n$. Sejam $x \in \mathbb{R}^n$ e $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ escalares tais que $\mu_j \geq 0, \forall j = 1, \dots, p$ e*

$$x = \sum_{i=1}^m \lambda_i u_i + \sum_{j=1}^p \mu_j v_j. \quad (1.12)$$

Então, existem subconjuntos $I \subset \{1, \dots, m\}, J \subset \{1, \dots, p\}$ e escalares $\{\lambda'_i\}_{i \in I}, \{\mu'_j\}_{j \in J}, \mu'_j \geq 0, \forall j \in J$, tais que

$$x = \sum_{i \in I} \lambda'_i u_i + \sum_{j \in J} \mu'_j v_j$$

e os vetores $\{u_i\}_{i \in I} \cup \{v_j\}_{j \in J}$ são linearmente independentes.

Prova. Sem perda de generalidade suponhamos que $\lambda_i \neq 0, \mu_j > 0$ para todo $i = 1, \dots, m, j = 1, \dots, p$.

Suponhamos também que os vetores $u_1, \dots, u_m, v_1, \dots, v_p$ são linearmente dependentes. Logo, existem escalares $\alpha_i, i = 1, \dots, m, \beta_j, j = 1, \dots, p$ não todos nulos tais que

$$0 = \sum_{i=1}^m \alpha_i u_i + \sum_{j=1}^p \beta_j v_j. \quad (1.13)$$

Subtraindo (1.13) da igualdade (1.12) temos que, para todo $\gamma \geq 0$,

$$x = \sum_{i=1}^m (\lambda_i - \gamma \alpha_i) u_i + \sum_{j=1}^p (\mu_j - \gamma \beta_j) v_j.$$

Seja $\bar{\gamma}$ o menor valor de γ que faz com que um dos coeficientes $\lambda_i - \gamma \alpha_i, \mu_j - \gamma \beta_j$ seja nulo com $\mu_j - \bar{\gamma} \beta_j \geq 0$ para todo $j = 1, \dots, p$. Então:

$$x = \sum_{i=1}^m (\lambda_i - \bar{\gamma} \alpha_i) u_i + \sum_{j=1}^p (\mu_j - \bar{\gamma} \beta_j) v_j$$

com $\mu_j - \bar{\gamma} \beta_j \geq 0$ para todo $j = 1, \dots, p$. Logo, conseguimos escrever x como combinação linear de um subconjunto de vetores estritamente menor que o conjunto original. Este processo pode ser repetido até que todos os vetores da combinação linear sejam linearmente independentes. $\square$

Teorema 1.3.2. *Consideremos o problema (1.4) onde, além da continuidade, temos que as funções f, h, g são diferenciáveis. Seja x_* um minimizador local do problema (1.4). Suponhamos que as derivadas de f, h, g são contínuas em x_* , que x_* é um ponto interior de $\mathcal{C}$ e que satisfaz alguma das seguintes condições:*

- (a) regularidade, (Definição 1.1.3);
- (b) Mangasarian-Fromovitz, (Definição 1.1.4);
- (c) Posto constante, (Definição 1.1.5);
- (d) Dependência linear positiva constante (CPLD), (Definição 1.1.8);
- (e) Quase-normalidade, (Definição 1.1.10).

Então, existem $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}_+^p$, tais que:

- $\nabla f(x_*) + \nabla h(x_*)\lambda + \nabla g(x_*)\mu = 0$,
- $[g(x_*)]_i < 0 \Rightarrow \mu_i = 0$.

Prova. Pelo Teorema 1.2.2 sabemos que existem seqüências $\{\lambda_k\} \in \mathbb{R}^m, \{\mu_k\} \in \mathbb{R}_+^p$ tais que

$$\lim_{k \rightarrow \infty} \nabla f(x_k) + \nabla h(x_k)\lambda_k + \nabla g(x_k)\mu_k = 0, \quad (1.14)$$

$$\lim_{k \rightarrow \infty} \|h(x_k)\| = \lim_{k \rightarrow \infty} \|g(x_k)_+\| = 0, \quad (1.15)$$

e existe $k_0 \in \mathbb{N}$ tal que, para todo $k \geq k_0$,

$$[g(x_*)]_i < 0 \Rightarrow [\mu_k]_i = 0. \quad (1.16)$$

Por (1.14) e (1.16) temos que

$$\lim_{k \rightarrow \infty} \nabla f(x_k) + \sum_{i=1}^m [\lambda_k]_i \nabla h_i(x_k) + \sum_{i \in I(x_*)} [\mu_k]_i \nabla g_i(x_k) = 0, \quad (1.17)$$

(a) A regularidade é uma condição de qualificação:

Suponhamos primeiro que as seqüências de escalares $\{[\lambda_k]_i\}, i = 1, \dots, m$, $\{[\mu_k]_j\}, j \in I(x_*)$ que aparecem na expressão (1.17) estão contidas num conjunto compacto. Neste caso, existem $K \subset \mathbb{N}$, $\lambda \in \mathbb{R}^m$, $\mu_j \geq 0, j \in I(x_*)$ tais que

$$\lim_{k \in K} [\lambda_k]_i = \lambda_i, \quad \lim_{k \in K} [\mu_k]_j = \mu_j, \quad \forall i = 1, \dots, m, j \in I(x_*).$$

Passando ao limite em (1.15)-(1.17) para $k \in K$ obtemos o resultado desejado.

Consideremos agora o caso em que as seqüências $\{[\lambda_k]_i\}, i = 1, \dots, m$ e $\{[\mu_k]_j\}, j \in I(x_*)$ não estão contidas em um conjunto compacto.

Definimos, para todo k ,

$$M_k = \max\{|[\lambda_k]_i|, [\mu_k]_j, i = 1, \dots, m, j \in I(x_*)\}.$$

Neste caso, $\lim_{k \rightarrow \infty} M_k = \infty$. Então, por (1.17), temos que

$$\lim_{k \rightarrow \infty} \frac{\nabla f(x_k)}{M_k} + \sum_{i=1}^m \frac{[\lambda_k]_i}{M_k} \nabla h_i(x_k) + \sum_{j \in I(x_*)} \frac{[\mu_k]_j}{M_k} \nabla g_j(x_k) = 0.$$

Logo,

$$\lim_{k \rightarrow \infty} \sum_{i=1}^m \frac{[\lambda_k]_i}{M_k} \nabla h_i(x_k) + \sum_{j \in I(x_*)} \frac{[\mu_k]_j}{M_k} \nabla g_j(x_k) = 0.$$

Como o número de restrições ativas é finito ($\leq m + p$), existe um índice que realiza o máximo M_k infinitas vezes, digamos, para todo $k \in K \subset \mathbb{N}$. Além do mais, existe um subconjunto $K_1 \subset K$ tal que as seqüências $\{\frac{[\lambda_k]_i}{M_k}\}_{k \in K_1}, i = 1, \dots, m$ e $\{\frac{[\mu_k]_j}{M_k}\}_{k \in K_1}, j \in I(x_*)$ convergem, e convergem para um conjunto de coeficientes não todos nulos, pois ao menos um dos coeficientes tem módulo sempre igual a 1. Logo temos que existem $\tilde{\lambda} \in \mathbb{R}^m$,

$\tilde{\mu}_j \geq 0, j \in I(x_*)$ não todos nulos tais que

$$\lim_{k \in K_1} \frac{[\lambda_k]_i}{M_k} = \tilde{\lambda}_i, \quad \lim_{k \in K_1} \frac{[\mu_k]_j}{M_k} = \tilde{\mu}_j, \quad \forall i = 1, \dots, m, j \in I(x_*)$$

e

$$\sum_{i=1}^m \tilde{\lambda}_i \nabla h_i(x_*) + \sum_{j \in I(x_*)} \tilde{\mu}_j \nabla g_j(x_*) = 0. \quad (1.18)$$

Logo os gradientes $\{\nabla h_i(x_*)\}_{i \in \{1, \dots, m\}} \cup \{\nabla g_j(x_*)\}_{j \in I(x_*)}$ são linearmente dependentes o que contradiz a hipótese de regularidade.

(b) Mangasarian-Fromovitz é uma condição de qualificação:

Esta prova é trivial a partir da prova da parte (a).

Observar que no caso em que as seqüências $\{[\lambda_k]_i\}, \{[\mu_k]_j\}, i = 1, \dots, m, j \in I(x_*)$ são não limitadas, como $[\mu_k]_j \geq 0 \quad \forall k \in \mathbb{N}$, então o limite $\tilde{\mu}_j \geq 0$ para todo $j \in I(x_*)$. Logo, por (1.18), obtemos a dependência linear positiva dos vetores o que, por (a) da Proposição 1.1.13 contradiz a hipótese.

(c) Posto constante é uma condição de qualificação:

Por (1.17) temos que existe uma seqüência de escalares $\{\epsilon_k\}$ tal que $\epsilon_k \rightarrow 0$ e

$$\nabla f(x_k) + \sum_{i=1}^m [\lambda_k]_i \nabla h_i(x_k) + \sum_{i \in I(x_*)} [\mu_k]_i \nabla g_i(x_k) = \epsilon_k.$$

Assim, pelo Lema de Caratheodory sabemos que, para cada $k \in \mathbb{N}$, existem $I_k \subset \{1, \dots, m\}, J_k \subset I(x_*)$ e seqüências de vetores $\{\alpha_k\}, \{\beta_k\}$ tais que $[\beta_k]_j \geq 0$ para todo k ,

$$\nabla f(x_k) + \sum_{i \in I_k} [\alpha_k]_i \nabla h_i(x_k) + \sum_{j \in J_k} [\beta_k]_j \nabla g_j(x_k) = \epsilon_k,$$

e os vetores

$$\{\nabla h_i(x_k)\}_{i \in I_k} \cup \{\nabla g_j(x_k)\}_{j \in J_k}$$

são linearmente independentes.

Como o número possível de subconjuntos I_k, J_k é finito, existe $K \subset \mathbb{N}$ tal que, para todo $k \in K, I_k = I \subset \{1, \dots, m\}, J_k = J \subset I(x_*)$,

$$\nabla f(x_k) + \sum_{i \in I} [\alpha_k]_i \nabla h_i(x_k) + \sum_{j \in J} [\beta_k]_j \nabla g_j(x_k) = \epsilon_k, \quad (1.19)$$

e os vetores

$$\{\nabla h_i(x_k)\}_{i \in I} \cup \{\nabla g_j(x_k)\}_{j \in J} \quad \text{são linearmente independentes.} \quad (1.20)$$

Se as seqüências $\{\alpha_k\}, \{\beta_k\}$ estão contidas num conjunto compacto, então existe $K_1 \subset K$ tal que

$$\lim_{k \in K_1} \alpha_k = \alpha, \text{ e } \lim_{k \in K_1} \beta_k = \beta \geq 0.$$

Logo, passando ao limite em (1.19) obtemos que

$$\nabla f(x_*) + \sum_{i \in I} \alpha_i \nabla h_i(x_*) + \sum_{j \in J} \beta_j \nabla g_j(x_*) = 0,$$

o que prova o resultado desejado.

Agora consideremos o caso em que as seqüências $\{\alpha_k\}, \{\beta_k\}$ não estão contidas em um conjunto compacto. Definimos, para todo k ,

$$M_k = \max\{|\alpha_k|_i, |\beta_k|_j, i \in I, j \in J\}.$$

Neste caso, $\lim_{k \rightarrow \infty} M_k = \infty$. Logo, por (1.19) temos que

$$\frac{\nabla f(x_k)}{M_k} + \sum_{i \in I} \frac{[\alpha_k]_i}{M_k} \nabla h_i(x_k) + \sum_{j \in J} \frac{[\beta_k]_j}{M_k} \nabla g_j(x_k) = \frac{\epsilon_k}{M_k},$$

e então,

$$\lim_{k \rightarrow \infty} \sum_{i \in I} \frac{[\alpha_k]_i}{M_k} \nabla h_i(x_k) + \sum_{j \in J} \frac{[\beta_k]_j}{M_k} \nabla g_j(x_k) = 0.$$

Como o número de restrições é finito ($\leq m + p$), existe um índice que realiza o máximo M_k infinitas vezes, digamos, para todo $k \in K \subset \mathbb{N}$. Além do mais, existe um subconjunto $K_1 \subset K$ tal que as seqüências $\{\frac{[\alpha_k]_i}{M_k}\}_{k \in K_1}, i \in I$ e $\{\frac{[\beta_k]_j}{M_k}\}_{k \in K_1}, j \in J$ convergem, e convergem para um conjunto de coeficientes não todos nulos, pois ao menos um dos coeficientes tem módulo sempre igual a 1. Logo, tomando limites para $k \in K_1$, temos que existem $\tilde{\alpha} \in \mathbb{R}^m, \tilde{\beta} \in \mathbb{R}_+^p$ com componentes não todas nulas tais que

$$\sum_{i \in I} \tilde{\alpha}_i \nabla h_i(x_*) + \sum_{j \in J} \tilde{\beta}_j \nabla g_j(x_*) = 0.$$

Logo, temos que os gradientes $\{\nabla h_i(x_*)\}_{i \in I} \cup \{\nabla g_j(x_*)\}_{j \in J}$ são linearmente dependentes, e existem pontos de uma vizinhança de x_* nos quais esses gradientes são linearmente independentes, o que contradiz a hipótese.

(d) A condição de dependência linear positiva constante é uma condição de qualificação: Esta prova segue da prova de (c).

No caso em que as seqüências $\{\alpha_k\}$, $\{\beta_k\}$ não estão contidas em um conjunto compacto sabemos que existem $\tilde{\alpha} \in \mathbb{R}^m$, $\tilde{\beta} \in \mathbb{R}_+^p$ não todos nulos tais que

$$\sum_{i \in I} \tilde{\alpha}_i \nabla h_i(x_*) + \sum_{j \in J} \tilde{\beta}_j \nabla g_j(x_*) = 0.$$

Logo, como $\tilde{\beta} \in \mathbb{R}_+^p$ temos que os gradientes

$$\{\nabla h_i(x_*)\}_{i \in I} \cup \{\nabla g_j(x_*)\}_{j \in J}$$

são positivo-linearmente dependentes.

Mas, por (1.20) os gradientes $\{\nabla h_i(x_k)\}_{i \in I} \cup \{\nabla g_j(x_k)\}_{j \in J}$ não são linearmente dependentes e portanto também não são positivo-linearmente dependentes. Assim, temos que os gradientes $\{\nabla h_i(x)\}_{i \in I} \cup \{\nabla g_j(x)\}_{j \in J}$ são PLD em x_* e em toda vizinhança de x_* existem pontos nos quais esses gradientes não são PLD. Isto implica que x_* não satisfaz a condição CPLD, o que é uma contradição.

(e) A quase-normalidade é uma condição de qualificação:

Pela demonstração do Teorema 1.2.2 sabemos que os vetores λ_k e μ_k que aparecem na expressão (1.17) estão definidos como $\lambda_k = \rho_k h(x_k)$, $\mu_k = \rho_k g(x_k)_+$ para $\{\rho_k\}$ uma seqüência de números positivos tal que $\rho_{k+1} > \rho_k \geq 0$ e $\lim_{k \rightarrow \infty} \rho_k = 0$. Ver também [4].

Definimos

$$\delta_k = \sqrt{1 + \sum_{i=1}^m ([\lambda_k]_i)^2 + \sum_{j \in I(x_*)} ([\mu_k]_j)^2}.$$

Logo, por (1.17) temos que existe ϵ_k tal que $\epsilon_k \rightarrow 0$ e

$$\frac{\nabla f(x_k)}{\delta_k} + \sum_{i=1}^m \frac{[\lambda_k]_i}{\delta_k} \nabla h_i(x_k) + \sum_{j \in I(x_*)} \frac{[\mu_k]_j}{\delta_k} \nabla g_j(x_k) = \frac{\epsilon_k}{\delta_k}. \quad (1.21)$$

A seqüência $\{(\frac{1}{\delta_k}, \frac{\lambda_k}{\delta_k}, \frac{\mu_k}{\delta_k})\}$ é limitada, logo, existem $K \subset \mathbb{N}$ e $(\mu_0^*, \lambda_1^*, \dots, \lambda_m^*, \mu_1^*, \dots, \mu_p^*)$, não todos nulos tais que $\mu_j^* \geq 0, j = 0, \dots, p$ e

$$\lim_{k \in K, k \rightarrow \infty} \left(\frac{1}{\delta_k}, \frac{\lambda_k}{\delta_k}, \frac{\mu_k}{\delta_k} \right) = (\mu_0^*, \lambda_1^*, \dots, \lambda_m^*, \mu_1^*, \dots, \mu_p^*).$$

Tomando limite quando $k \in K$ cresce em (1.21) obtemos que

$$\mu_0^* \nabla f(x_*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x_*) + \sum_{j \in I(x_*)} \mu_j^* \nabla g_j(x_*) = 0. \quad (1.22)$$

Definimos os conjuntos $I_0 = \{i \in \{1, \dots, m\} | \lambda_i^* \neq 0\}$, $J_0 = \{j \in I(x_*) | \mu_j^* \neq 0\}$. Logo, para todo $k \in K$ suficientemente grande temos que

$$\lambda_i^* [\lambda_k]_i > 0, \forall i \in I_0, \quad \text{e} \quad \mu_j^* [\mu_k]_j > 0, \forall j \in J_0.$$

Pela definição de λ_k, μ_k e por ser $\rho_k > 0$ obtemos que

$$\lambda_i^* h_i(x_k) > 0, \forall i \in I_0, \quad \lambda_j^* g_j(x_k) > 0, \forall j \in J_0$$

para todo $k \in K$ suficientemente grande.

Logo, se $\mu_0^* = 0$, por (1.22) teríamos que x_* não seria quase-normal. Então, deve-se ter $\mu_0^* \neq 0$, o que mostra que x^* é um ponto KKT. $\square$

Observações.

1. A prova de que a condição CPLD-original é uma condição de qualificação foi feita na prova do item (d) do Teorema 1.3.2 .
2. Na verdade, no Teorema 1.3.2 provamos o seguinte resultado:

Teorema 1.3.3. *Seja $x_* \in \mathbb{R}^n$ um ponto limite de uma seqüência $\{x_k\} \subset \mathbb{R}^n$ para a qual existem seqüências $\{\lambda_k\} \subset \mathbb{R}^m$, $\{\mu_k\} \subset \mathbb{R}_+^p$, tais que:*

- $\lim_{k \rightarrow \infty} \nabla f(x_k) + \nabla h(x_k) \lambda_k + \nabla g(x_k) \mu_k = 0$,
- $\lim_{k \rightarrow \infty} \|h(x_k)\| = \lim_{k \rightarrow \infty} \|g(x_k)_+\| = 0$,
- *Existe $k_0 \in \mathbb{N}$ tal que, para todo $k \geq k_0$,*

$$[g(x_*)]_i < 0 \Rightarrow [\mu_k]_i = 0.$$

Suponhamos que x_ satisfaz alguma das seguintes condições de qualificação:*

- *LIAC,*
- *MFCQ,*
- *CRCQ,*

- CPLD,
- quase-normalidade.

Então x_* é um ponto estacionário de (1.4).

1.4 A condição CPLD implica a quase-normalidade

Na seção anterior provamos que, como conjecturaram Qi e Wei em [32], a condição CPLD é uma condição de qualificação. Do ponto de vista prático é interessante analisar qual é o *status* de uma condição de qualificação em relação às outras condições conhecidas na literatura.

Nesta seção provaremos que a CPLD implica a quase-normalidade, mostrando assim que a quase-normalidade é mais fraca que a condição CPLD. Nossa estratégia consiste em provar que a dependência linear positiva de qualquer subconjunto dos gradientes das restrições de igualdade e de desigualdade ativas implica que não se verifica a propriedade sequencial 1. Para isso precisamos de dois lemas preparatórios.

Lema 1.4.1. *Seja $D \subset \mathbb{R}^n$ um conjunto aberto, $x \in D$, $F : D \rightarrow \mathbb{R}^n$, $F = (f_1, \dots, f_n) \in C^1(D)$, $F'(x)$ não-singular. (Pelo Teorema da Função Inversa F é um homeomorfismo entre uma vizinhança aberta D_1 de x e $F(D_1)$, e $F^{-1} : F(D_1) \rightarrow D_1$ está bem definida, ver [4].)*

Assumimos que $q \in \{1, \dots, n-1\}$, $f : D \rightarrow \mathbb{R}$, $f \in C^1(D)$ é tal que $\nabla f(y)$ é uma combinação linear de $\nabla f_1(y), \dots, \nabla f_q(y)$ para todo $y \in D$.

Definimos, para todo $u \in F(D_1)$,

$$\varphi(u) = f[F^{-1}(u)]. \quad (1.23)$$

Então, para todo $u \in F(D_1)$, $j = q+1, \dots, n$,

$$\frac{\partial \varphi}{\partial u_j}(u) = 0. \quad (1.24)$$

Prova. Pelo Teorema da Função Implícita sabemos que existem $D_1 \subset D$ vizinhança de x e uma função contínua $F^{-1} : F(D_1) \rightarrow D_1$ tal que, se $F^{-1}(u) = (g_1(u), \dots, g_n(u))$ temos que

$$f_k(g_1(u), \dots, g_n(u)) = u_k \quad \forall k = 1, \dots, n. \quad (1.25)$$

Seja $u \in F^{-1}(D_1)$. Usando a regra da cadeia em (1.23) obtemos que

$$\frac{\partial \varphi}{\partial u_j}(u) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(F^{-1}(u)) \frac{\partial g_i}{\partial u_j}(u) \quad \forall j = 1, \dots, n. \quad (1.26)$$

Por hipótese sabemos que existem escalares $\lambda_1, \dots, \lambda_q$ tais que

$$\nabla f(y) = \sum_{k=1}^q \lambda_k \nabla f_k(y) \quad \forall y \in D. \quad (1.27)$$

Substituindo (1.27) em (1.26) temos que

$$\frac{\partial \varphi}{\partial u_j}(u) = \sum_{i=1}^n \left(\sum_{k=1}^q \lambda_k \frac{\partial f}{\partial x_i}(F^{-1}(u)) \right) \frac{\partial g_i}{\partial u_j}(u) = \sum_{k=1}^q \lambda_k \sum_{i=1}^n \frac{\partial f}{\partial x_i}(F^{-1}(u)) \frac{\partial g_i}{\partial u_j}(u)$$

e usando (1.25) obtemos que

$$\frac{\partial \varphi}{\partial u_j}(u) = 0 \quad \forall j = q+1, \dots, n,$$

como queríamos provar. □

Lema 1.4.2. *Sejam $f, f_1, \dots, f_q : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ funções continuamente diferenciáveis, $x \in D$ e D um conjunto aberto.*

Assumimos que $\nabla f_1(x), \dots, \nabla f_q(x)$ são linearmente independentes e que $\nabla f(y)$ é uma combinação linear de $\nabla f_1(y), \dots, \nabla f_q(y)$ para todo $y \in D$. Em particular,

$$\nabla f(x) = \sum_{i=1}^q \lambda_i \nabla f_i(x). \quad (1.28)$$

Então, existem $D_2 \subset \mathbb{R}^q$, uma vizinhança aberta de $(f_1(x), \dots, f_q(x))$, e uma função $\varphi : D_2 \rightarrow \mathbb{R}$, $\varphi \in C^1(D_2)$, tais que, para todo $y \in D_1$, temos que $(f_1(y), \dots, f_q(y)) \in D_2$ e

$$f(y) = \varphi(f_1(y), \dots, f_q(y)).$$

Além disso, para todo $i = 1, \dots, q$,

$$\lambda_i = \frac{\partial \varphi}{\partial u_i}(f_1(x), \dots, f_q(x)). \quad (1.29)$$

Prova. Definimos $n - q$ funções $f_{q+1}, \dots, f_n$ de tal forma que se verifiquem as hipóteses

do Lema 1.4.1. Então, aplicando (1.24) temos que a função φ definida em (1.23) não depende das variáveis $u_{q+1}, \dots, u_n$. Pela continuidade de $f_1, \dots, f_n$ existem $\tilde{D}_1 \subset D_1, B_1, B_2$ vizinhanças de $x, (f_1(x), \dots, f_q(x)), (f_{q+1}, \dots, f_n(x))$ respectivamente tais que

- (i) $B_1 \times B_2 \subset F(D_1)$,
- (ii) $\forall y \in \tilde{D}_1, (f_1(y), \dots, f_q(y)) \in B_1$,
- (iii) $\forall u \in B_1 \times B_2$, existe $\tilde{\varphi} : B_1 \rightarrow \mathbb{R}, \tilde{\varphi} \in C^1(B_1)$ tal que $\varphi(u) = \tilde{\varphi}(u_1, \dots, u_q)$.

Definimos $D_2 = B_1$. Então, $\forall y \in \tilde{D}_1$ temos, por (ii), (iii) e (1.23), que

$$(f_1(y), \dots, f_q(y)) \in D_2 \text{ e } \tilde{\varphi}(f_1(y), \dots, f_q(y)) = \varphi(F(y)) = f(y).$$

Logo, $f(y) = \varphi(f_1(y), \dots, f_q(y))$, $\forall y \in \tilde{D}_1$.

Além disso, pela definição de φ , obtemos que

$$\frac{\partial \varphi}{\partial u_j}(f_1(x), \dots, f_n(x)) = \frac{\partial \tilde{\varphi}}{\partial u_j}(f_1(x), \dots, f_q(x)) = \lambda_j, \quad \forall j = 1, \dots, q,$$

o que prova (1.29). □

Provaremos agora o resultado principal desta seção.

Teorema 1.4.3. *Se $x \in X$ satisfaz a condição CPLD então x é um ponto quase-normal.*

Prova. Assumimos que $x \in X$ satisfaz a condição CPLD. Se x é MF-regular, pela Proposição 1.1.13 a prova está feita. Suponhamos que x é MF-não-regular. Então, existem escalares $\lambda_1, \dots, \lambda_m \in \mathbb{R}, \mu_j \geq 0 \forall j \in I(x)$ tais que

$$\sum_{i=1}^m |\lambda_i| + \sum_{j \in I(x)} \mu_j > 0 \tag{1.30}$$

e

$$\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I(x)} \mu_j \nabla g_j(x) = 0. \tag{1.31}$$

Definimos

$$I_+(x) = \{i \in \{1, \dots, m\} \mid \lambda_i > 0\},$$

$$I_-(x) = \{i \in \{1, \dots, m\} \mid \lambda_i < 0\},$$

$$I_0(x) = \{j \in I(x) \mid \mu_j > 0\}.$$

Então, por (1.31),

$$\sum_{i \in I_+(x)} \lambda_i \nabla h_i(x) + \sum_{i \in I_-(x)} \lambda_i \nabla h_i(x) + \sum_{j \in I_0(x)} \mu_j \nabla g_j(x) = 0. \quad (1.32)$$

Por (1.30), $I_+(x) \cup I_-(x) \cup I_0(x) \neq \emptyset$.

Assumimos, primeiro, que $I_+(x) \neq \emptyset$. Seja $i_1 \in I_+(x)$. Então, por (1.32),

$$\lambda_{i_1} \nabla h_{i_1}(x) = - \sum_{i \in I_+(x) - \{i_1\}} \lambda_i \nabla h_i(x) - \sum_{i \in I_-(x)} \lambda_i \nabla h_i(x) - \sum_{j \in I_0(x)} \mu_j \nabla g_j(x).$$

Consideremos as duas possibilidades:

(a) $\nabla h_{i_1}(x) = 0$;

(b) $\nabla h_{i_1}(x) \neq 0$.

Se (a) se verifica, o conjunto unitário $\{\nabla h_{i_1}(x)\}$ é positivo-linearmente dependente. Então, pela condição CPLD, $\{\nabla h_{i_1}(y)\}$ é positivo-linearmente dependente para todo y numa vizinhança de x . Logo, $\nabla h_{i_1}(y) = 0$ para todo y nessa vizinhança. Como $h_{i_1}(x) = 0$ temos que $h_{i_1}(y) = 0$ para todo y perto de x . Portanto, uma seqüência $y_k \rightarrow x$ tal que $h_{i_1}(y_k) > 0$ para todo k não pode existir. Logo, a propriedade seqüencial 1 associada aos escalares $\lambda_1, \dots, \lambda_m, \mu_j, j \in I(x)$ não se verifica, cumprindo-se, portanto, a quase-normalidade.

Assumimos agora que (b) se verifica. Então, pelo Lema 1.3.1, existem

$$I_{++} \subset I_+(x) - \{i_1\}, \quad I_{--} \subset I_-(x), \quad I_{00} \subset I_0(x)$$

tais que os vetores

$$\{\nabla h_i(x)\}_{i \in I_{++}}, \quad \{\nabla h_i(x)\}_{i \in I_{--}}, \quad \{\nabla g_j(x)\}_{j \in I_{00}} \quad (1.33)$$

são linearmente independentes e

$$\lambda_{i_1} \nabla h_{i_1}(x) = - \sum_{i \in I_{++}} \bar{\lambda}_i \nabla h_i(x) - \sum_{i \in I_{--}} \bar{\lambda}_i \nabla h_i(x) - \sum_{j \in I_{00}} \bar{\mu}_j \nabla g_j(x) \quad (1.34)$$

com

$$\bar{\lambda}_i > 0, \quad \forall i \in I_{++},$$

$$\bar{\lambda}_i < 0, \quad \forall i \in I_{--}$$

e

$$\bar{\mu}_j > 0, \forall j \in I_{00}.$$

Pela independência linear dos vetores (1.33) e por argumentos de continuidade, os vetores

$$\{\nabla h_i(y)\}_{i \in I_{++}}, \{\nabla h_i(y)\}_{i \in I_{--}}, \{\nabla g_j(y)\}_{j \in I_{00}} \quad (1.35)$$

são linearmente independentes para todo y numa vizinhança de x . Mas, pela condição CPLD, os vetores

$$\lambda_{i_1} \nabla h_{i_1}(y), \{\nabla h_i(y)\}_{i \in I_{++}}, \{\nabla h_i(y)\}_{i \in I_{--}}, \{\nabla g_j(y)\}_{j \in I_{00}}$$

são positivo-linearmente dependentes numa vizinhança de x . Portanto, $\lambda_{i_1} \nabla h_{i_1}(y)$ deve ser uma combinação linear dos vetores (1.35) para todo y numa vizinhança de x .

Para simplificar, e sem perda de generalidade, escrevemos:

$$I_{++} = \{1, \dots, m_1\},$$

$$I_{--} = \{m_1 + 1, \dots, m_2\},$$

$$I_{00} = \{1, \dots, p_1\},$$

$$m_2 + p_1 = q.$$

Pelo Lema 1.4.2, existe uma função suave φ definida numa vizinhança de $(0, \dots, 0) \in \mathbb{R}^q$ tal que, para todo y numa vizinhança de x ,

$$\lambda_{i_1} h_{i_1}(y) = \varphi(h_1(y), \dots, h_{m_2}(y), g_1(y), \dots, g_{p_1}(y)). \quad (1.36)$$

Agora, suponhamos que $\{y_k\}$ é uma seqüência que converge a x tal que

$$h_i(y_k) > 0, i = 1, \dots, m_1,$$

$$h_i(y_k) < 0, i = m_1 + 1, \dots, m_2,$$

$$g_j(y_k) > 0, j = 1, \dots, p_1.$$

Então, por (1.29), (1.34) e (1.36), para k suficientemente grande devemos ter que $\lambda_{i_1} h_{i_1}(y_k) \leq 0$. Isto implica que a propriedade seqüencial 1 não se verifica.

As provas para os casos $I_-(x) \neq \emptyset$ e $I_0(x) \neq \emptyset$ são inteiramente análogas a este caso. Portanto, CPLD implica quase-normalidade, como queríamos provar. $\square$

1.5 Relações entre as diferentes condições de qualificação

Nesta seção provamos algumas relações entre a CPLD e outras condições de qualificação definidas na Seção 1.1.

Proposição 1.5.1. *As seguintes implicações são verdadeiras:*

- (a) *Regularidade* $\implies$ *MFCQ*;
- (b) *Regularidade* $\implies$ *CRCQ*;
- (c) *MFCQ* $\implies$ *CPLD*;
- (d) *CRCQ* $\implies$ *CPLD*;
- (e) *Pseudo-normalidade* $\implies$ *Quase-normalidade*;
- (f) *CPLD-original* $\iff$ *CPLD*.

Prova. As implicações (a) e (b) são imediatas a partir das definições.

(c) Se x_* satisfaz a condição MFCQ então o conjunto de gradientes

$$\{\nabla h_i(x_*)\}_{i \in \{1, \dots, m\}} \cup \{\nabla g_j(x_*)\}_{j \in I(x_*)}$$

é PLI. Logo qualquer subconjunto dele é PLI e é satisfeita a condição CPLD.

- (d) Trivial devido ao fato de que dependência linear positiva implica dependência linear.
- (e) Trivial devido ao fato de que se x_* satisfaz a propriedade seqüencial 1 então x_* satisfaz a propriedade seqüencial 2.
- (f) Assumimos que $x_* \in X$ satisfaz a condição CPLD-original.

Suponhamos que x_* não satisfaz a condição CPLD. Então, existem $I_0 \subset \{1, \dots, m\}$, $J_0 \subset I(x_*)$ tais que os gradientes $\{\nabla h_i(x_*)\}_{i \in I_0} \cup \{\nabla g_j(x_*)\}_{j \in J_0}$ são positivo-linearmente dependentes e existe uma seqüência $\{y_k\}$, $y_k \rightarrow x_*$ para a qual os gradientes $\{\nabla h_i(y_k)\}_{i \in I_0} \cup \{\nabla g_j(y_k)\}_{j \in J_0}$ são positivo-linearmente independentes. (E além do mais, sabemos que qualquer subconjunto desses gradientes é positivo-linearmente independente em y_k .)

Como os gradientes $\{\nabla h_i(x_*)\}_{i \in I_0} \cup \{\nabla g_j(x_*)\}_{j \in J_0}$ são positivo-linearmente dependentes sabemos que existem escalares $\lambda_i \in \mathbb{R}$, $i \in I_0$, $\mu_j \geq 0 \forall j \in J_0$, não todos nulos, tais que

$$\sum_{i \in I_0} \lambda_i \nabla h_i(x_*) + \sum_{j \in J_0} \mu_j \nabla g_j(x_*) = 0. \quad (1.37)$$

Definimos

$$I_+(x_*) = \{i \in I_0 \mid \lambda_i > 0\},$$

$$I_-(x_*) = \{i \in I_0 \mid \lambda_i < 0\},$$

$$I_0(x_*) = \{j \in J_0 \mid \mu_j > 0\}.$$

Temos que $I_+(x_*) \cup I_-(x_*) \cup I_0(x_*) \neq \emptyset$, e, por (1.37),

$$\sum_{i \in I_+(x_*)} \lambda_i \nabla h_i(x_*) + \sum_{i \in I_-(x_*)} \lambda_i \nabla h_i(x_*) + \sum_{j \in I_0(x_*)} \mu_j \nabla g_j(x_*) = 0. \quad (1.38)$$

Assumimos, primeiro, que $I_+(x_*) \neq \emptyset$. Seja $i_1 \in I_+(x_*)$. Então, por (1.38),

$$\lambda_{i_1} \nabla h_{i_1}(x_*) = - \sum_{i \in I_+(x_*) - \{i_1\}} \lambda_i \nabla h_i(x_*) - \sum_{i \in I_-(x_*)} \lambda_i \nabla h_i(x_*) - \sum_{j \in I_0(x_*)} \mu_j \nabla g_j(x_*). \quad (1.39)$$

Consideremos as duas possibilidades:

$$(a) \nabla h_{i_1}(x_*) = 0;$$

$$(b) \nabla h_{i_1}(x_*) \neq 0.$$

Se (a) se verifica, o conjunto unitário $\{\nabla h_{i_1}(x_*)\}$ é positivo-linearmente dependente. Então, pela condição CPLD-original, $\{\nabla h_{i_1}(y)\}$ é linearmente dependente para todo y numa vizinhança de x_* . Logo, $\nabla h_{i_1}(y) = 0$ para todo y nessa vizinhança e então, para $k \geq k_0$ temos que $\{\nabla h_{i_1}(y_k)\}$ é PLD, contradição.

Assumimos agora que (b) se verifica. Então, pelo Lema de Caratheodory, existem

$$I_{++} \subset I_+(x_*) - \{i_1\}, \quad I_{--} \subset I_-(x_*), \quad I_{00} \subset I_0(x_*)$$

e escalares

$$\bar{\lambda}_i > 0, \quad \forall i \in I_{++},$$

$$\bar{\lambda}_i < 0, \quad \forall i \in I_{--},$$

$$\bar{\mu}_j > 0, \quad \forall j \in I_{00},$$

tais que os vetores

$$\{\nabla h_i(x_*)\}_{i \in I_{++}}, \quad \{\nabla h_i(x_*)\}_{i \in I_{--}}, \quad \{\nabla g_j(x_*)\}_{j \in I_{00}} \quad \text{são linearmente independentes} \quad (1.40)$$

e

$$\lambda_{i_1} \nabla h_{i_1}(x_*) = - \sum_{i \in I_{++}} \bar{\lambda}_i \nabla h_i(x_*) - \sum_{i \in I_{--}} \bar{\lambda}_i \nabla h_i(x_*) - \sum_{j \in I_{00}} \bar{\mu}_j \nabla g_j(x). \quad (1.41)$$

Pela independência linear dos vetores (1.40) e por argumentos de continuidade, os vetores

$$\{\nabla h_i(y)\}_{i \in I_{++}}, \{\nabla h_i(y)\}_{i \in I_{--}}, \{\nabla g_j(y)\}_{j \in I_{00}} \quad (1.42)$$

são linearmente independentes para todo y numa vizinhança de x_* . Mas, pela condição CPLD-original, os vetores

$$\lambda_{i_1} \nabla h_{i_1}(y), \{\nabla h_i(y)\}_{i \in I_{++}}, \{\nabla h_i(y)\}_{i \in I_{--}}, \{\nabla g_j(y)\}_{j \in I_{00}}$$

são linearmente dependentes numa vizinhança de x_* . Portanto, $\lambda_{i_1} \nabla h_{i_1}(y)$ deve ser uma combinação linear dos vetores (1.42) para todo y numa vizinhança de x_* .

Em particular, para cada y_k sabemos que existem escalares $[\alpha_k]_i, [\beta_k]_j$ não todos nulos tais que

$$\lambda_{i_1} \nabla h_{i_1}(y_k) = - \sum_{i \in I_{++}} [\alpha_k]_i \nabla h_i(y_k) - \sum_{i \in I_{--}} [\alpha_k]_i \nabla h_i(y_k) - \sum_{j \in I_{00}} [\beta_k]_j \nabla g_j(y_k). \quad (1.43)$$

Se a sequência $\{(\alpha_k, \beta_k)\}$ é não limitada, dividindo a igualdade (1.43) por $\|(\alpha_k, \beta_k)\|$, considerando uma subsequência adequada e fazendo k tender a infinito para essa subsequência obtemos:

$$\sum_{i \in I_{++}} \alpha_i \nabla h_i(x_*) + \sum_{i \in I_{--}} \alpha_i \nabla h_i(x_*) + \sum_{j \in I_{00}} \beta_j \nabla g_j(x_*) = 0$$

com α_i, β_j não todos nulos. Logo, os gradientes

$$\{\nabla h_i(x_*)\}_{i \in I_{++}}, \{\nabla h_i(x_*)\}_{i \in I_{--}}, \{\nabla g_j(x_*)\}_{j \in I_{00}}$$

são linearmente dependentes, o que contradiz (1.40).

Assim, a sequência $\{(\alpha_k, \beta_k)\}$ é limitada e existem α_i, β_j não todos nulos tais que $[\alpha_k]_i \rightarrow \alpha_i, [\beta_k]_j \rightarrow \beta_j$. (Se fosse $\alpha = \beta = 0$ então $\nabla h_{i_1}(x_*) = 0$ e esse caso já foi analisado). Tomando o limite quando k cresce em (1.43) temos que

$$\lambda_{i_1} \nabla h_{i_1}(x_*) = - \sum_{i \in I_{++}} \alpha_i \nabla h_i(x_*) - \sum_{i \in I_{--}} \alpha_i \nabla h_i(x_*) - \sum_{j \in I_{00}} \beta_j \nabla g_j(x_*) \quad (1.44)$$

e subtraindo (1.44) de (1.41) obtemos que

$$\sum_{i \in I_{++}} (\bar{\lambda}_i - \alpha_i) \nabla h_i(x_*) + \sum_{i \in I_{--}} (\bar{\lambda}_i - \alpha_i) \nabla h_i(x_*) + \sum_{j \in I_{00}} (\bar{\mu}_j - \beta_j) \nabla g_j(x_*) = 0$$

o que, pela independência linear dos gradientes envolvidos, implica que

$$\bar{\lambda}_i = \alpha_i, \forall i \in I_{++} \cup I_{--}, \bar{\mu}_j = \beta_j, \forall j \in I_{00}.$$

Logo, como $\bar{\mu}_j > 0, \forall j \in I_{00}$, temos que existem $K \subset \mathbb{N}$ e k_0 tal que $\forall k \in K, k \geq k_0$,

$$[\beta_k]_j \geq 0,$$

o que, por (1.43), implica que os gradientes

$$\nabla h_{i_1}(y_k), \{\nabla h_i(y_k)\}_{i \in I_{++}}, \{\nabla h_i(y_k)\}_{i \in I_{--}}, \{\nabla g_j(y_k)\}_{j \in I_{00}}$$

são PLD. Contradição.

As provas para os casos $I_-(x_*) \neq \emptyset$ e $I_0(x_*) \neq \emptyset$ são inteiramente análogas a este caso. Portanto, CPLD-original implica CPLD, como queríamos provar. $\square$

Provamos que a condição CPLD é mais fraca que as condições MFCQ e CRCQ. É interessante investigar se a condição CPLD é estritamente mais fraca que ambas MFCQ e CRCQ juntas. O seguinte contra-exemplo mostra que, de fato, há pontos que satisfazem a CPLD mas não satisfazem nem MFCQ nem CRCQ.

Contra-exemplo 1. *CPLD não implica nem MFCQ nem CRCQ.*

Consideremos $n = 2, m = 0, p = 4$, o ponto $x_* = (0, 0)$ e as restrições de desigualdade

$$\begin{aligned} g_1(x_1, x_2) &= x_1 \\ g_2(x_1, x_2) &= x_1 + x_2^2 \\ g_3(x_1, x_2) &= x_1 + x_2 \\ g_4(x_1, x_2) &= -x_1 - x_2. \end{aligned}$$

Como $\nabla g_3(x_*) + \nabla g_4(x_*) = 0$, x_* não satisfaz a condição de qualificação de Mangasarian-Fromovitz.

Os gradientes $\nabla g_1(x_*), \nabla g_2(x_*)$ são linearmente dependentes mas em cada vizinhança de x_* há pontos nos quais $\nabla g_1(x), \nabla g_2(x)$ são linearmente independentes. Portanto, x_* não satisfaz CRCQ.

Mostramos que x_* satisfaz a condição CPLD-original. Assumimos que $\mu_j \geq 0$, $j = 1, \dots, 4$, $\sum_{j=1}^4 \mu_j > 0$, e

$$\mu_1 \nabla g_1(x_*) + \mu_2 \nabla g_2(x_*) + \mu_3 \nabla g_3(x_*) + \mu_4 \nabla g_4(x_*) = 0. \quad (1.45)$$

Se $\mu_2 = 0$, (1.45) implica que $\nabla g_1(x_*)$, $\nabla g_3(x_*)$, $\nabla g_4(x_*)$ são linearmente dependentes. Logo, como g_1, g_3, g_4 são lineares, $\nabla g_1(x)$, $\nabla g_3(x)$, $\nabla g_4(x)$ são linearmente dependentes para todo $x \in \mathbb{R}^2$.

Suponhamos que $\mu_2 > 0$. É simples ver que isto é impossível a menos que no mínimo dois dos multiplicadores μ_1, μ_3, μ_4 sejam positivos. Mas, como três vetores em $\mathbb{R}^2$ são sempre linearmente dependentes, temos que se verifica a condição CPLD-original e, portanto, também a condição CPLD.

A pergunta mais óbvia é se a quase-normalidade implica CPLD. O seguinte contra-exemplo mostra que isto não é verdadeiro.

Contra-exemplo 2. *Quase-normalidade não implica CPLD.*

Consideremos $n = 2, m = 2, p = 0, x_* = (0, 0)$ e as restrições de igualdade

$$\begin{aligned} h_1(x_1, x_2) &= x_2 e^{x_1}, \\ h_2(x_1, x_2) &= x_2. \end{aligned}$$

Temos que $\nabla h_1(x_*) = \nabla h_2(x_*)$. Então, $\lambda_1 \nabla h_1(x_*) + \lambda_2 \nabla h_2(x_*) = 0$ implica que $\lambda_1 = -\lambda_2$. A propriedade seqüencial 1 não se verifica, já que o sinal de $h_1(x)$ é o mesmo que o sinal de $h_2(x)$ para todo x . Portanto, x_* é quase-normal. Mas, em todo ponto de qualquer vizinhança de x_* há pontos nos quais ∇h_1 e ∇h_2 são linearmente independentes. Logo, a condição CPLD não é satisfeita em x_* .

No seguinte contra-exemplo provamos que CPLD não implica a pseudo-normalidade.

Contra-exemplo 3. *CPLD não implica pseudo-normalidade.*

Consideremos $n = 1, m = 0, p = 2, x_* = 0$ e as restrições de desigualdade

$$\begin{aligned} g_1(x_1) &= -x_1 \\ g_2(x_1) &= x_1 + x_1^2. \end{aligned}$$

O ponto x_* não é MF-regular já que $\nabla g_1(x_*) + \nabla g_2(x_*) = 0$. Mas $g_1(x_1) + g_2(x_1) = x_1^2 > 0$ para todo $x_1 \neq x_*$, portanto verifica-se a propriedade seqüencial 2. Logo, x_* não é pseudo-normal.

Agora, $\nabla g_1(x_*) \neq 0 \neq \nabla g_2(x_*)$ e $\nabla g_1(x), \nabla g_2(x)$ são linearmente dependentes para todo $x \in \mathbb{R}$. Portanto, x_* satisfaz a condição de qualificação de posto constante. Logo, x_* também satisfaz a condição CPLD.

O contra-exemplo 2 mostra uma situação na qual a quase-normalidade se verifica mas a CPLD não. É simples provar que, nesse exemplo, o ponto x_* não é pseudo-normal: tomemos $\lambda_1 = 1, \lambda_2 = -1$ e consideremos a sequência $y_k = (1/k, 1/k)$, para a qual é trivial ver que (1.2) se cumpre. Mas, o seguinte contra-exemplo mostra que a pseudo-normalidade não implica a CPLD.

Contra-exemplo 4. *Pseudo-normalidade não implica CPLD.*

Consideremos $n = 2, m = 0, p = 2$, $x_* = (0, 0)$, e as restrições de desigualdade

$$\begin{aligned} g_1(x_1, x_2) &= -x_1 \\ g_2(x_1, x_2) &= x_1 - x_1^2 x_2^2. \end{aligned}$$

Então,

$$\nabla g_1(x_1, x_2) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad \text{e} \quad \nabla g_2(x_1, x_2) = \begin{pmatrix} 1 - 2x_1 x_2^2 \\ -2x_1^2 x_2 \end{pmatrix}.$$

Logo, para todo $\mu_1 = \mu_2 > 0$, temos que:

$$\mu_1 \nabla g_1(0, 0) + \mu_2 \nabla g_2(0, 0) = \mu_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0.$$

Mas, se $\mu_1 = \mu_2 > 0$,

$$\mu_1 g_1(x_1, x_2) + \mu_2 g_2(x_1, x_2) = -\mu_1 x_1^2 x_2^2 < 0 \quad \forall (x_1, x_2) \neq (0, 0).$$

Logo, x_* é pseudo-normal.

Por outro lado, os gradientes $\nabla g_1(x), \nabla g_2(x)$ são linearmente independentes se $x_1 \neq 0 \neq x_2$. Isto implica que a condição CPLD-original não se verifica e portanto, também não se verifica a condição CPLD.

Na Figura 1.1 ilustramos as relações entre as diferentes condições de qualificação que foram mencionadas neste trabalho.

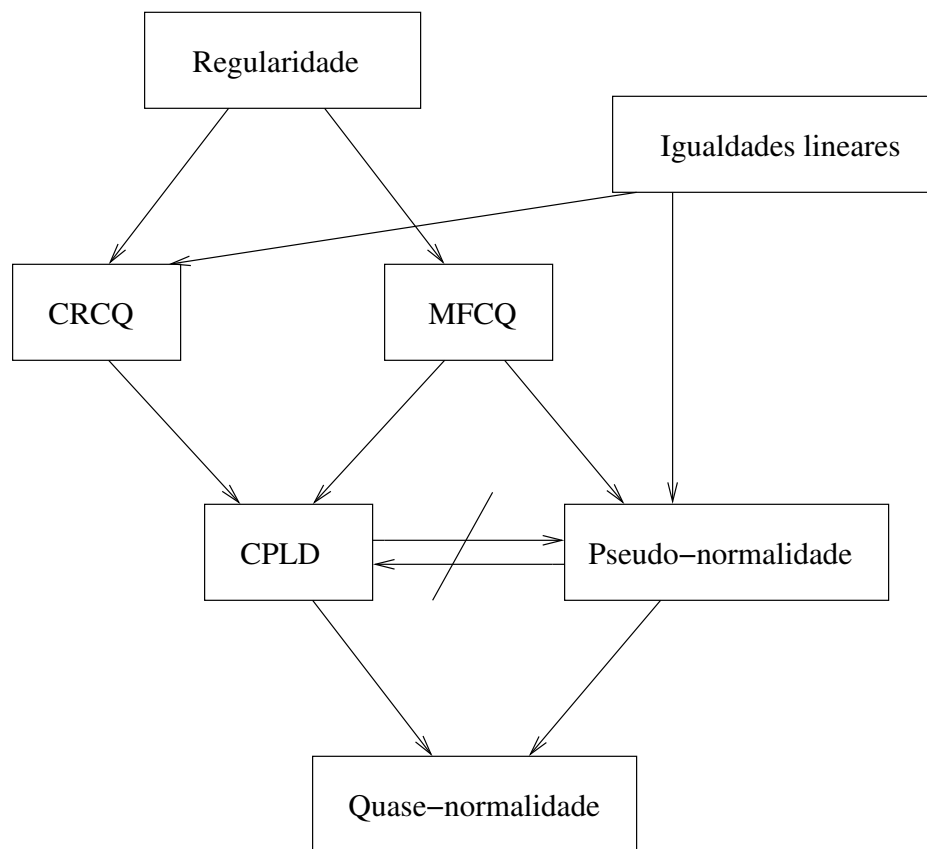


Figura 1.1: Relações entre as diferentes condições de qualificação.

Capítulo 2

O método de Lagrangiano aumentado para resolver problemas gerais

Neste capítulo consideramos o seguinte problema geral de programação não-linear:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeito a } x \in \Omega_1 \cap \Omega_2, \end{aligned} \tag{2.1}$$

para os conjuntos admissíveis Ω_1, Ω_2 definidos da seguinte forma:

$$\Omega_1 = \{x \in \mathbb{R}^n \mid h_1(x) = 0, g_1(x) \leq 0\},$$

$$\Omega_2 = \{x \in \mathbb{R}^n \mid h_2(x) = 0, g_2(x) \leq 0\},$$

sendo todas as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}, h_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}, h_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}, g_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{p_1}, g_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{p_2}$ contínuas com derivadas primeiras contínuas sobre um conjunto aberto que contém $\Omega_1 \cap \Omega_2$.

Os conjuntos admissíveis Ω_1 e Ω_2 podem ser definidos de modo que as restrições mais fáceis de ser manipuladas, por exemplo restrições de canalização ou lineares, formem parte do conjunto Ω_2 deixando as restrições mais difíceis de ser tratadas eficientemente, por exemplo as não-lineares, para formarem parte do conjunto Ω_1 . Assim, o conjunto de restrições Ω_2 será chamado de conjunto *simples* e o conjunto Ω_1 será chamado de *complexo*.

Para um problema de otimização específico, a definição dos conjuntos viáveis Ω_1, Ω_2 pode ser feita de modo que o problema

$$\text{Minimizar } F(x) \text{ sujeita a } x \in \Omega_2 \tag{2.2}$$

seja consideravelmente mais fácil de ser resolvido que o problema (2.1), ou que exista algum algoritmo eficiente para resolver esse problema que não pode ser aplicado, ou deixa

de ser eficiente, quando restrições do tipo das do conjunto Ω_1 são adicionadas. Nesses casos, é natural resolver o problema original (2.1) mediante uma seqüência de problemas da forma (2.2) para alguma função adequada F .

Levando em conta estas considerações proporemos um algoritmo de Lagrangiano aumentado para resolver o problema (2.1). O método de Lagrangiano aumentado, ver [4, 5, 20, 22, 31], consiste em uma seqüência de iterações externas. Em cada iteração externa um problema de minimização com restrições simples, da forma (2.2), é resolvido aproximadamente para obter o iterado atual e uma vez obtido este iterado são atualizados os multiplicadores de Lagrange e os parâmetros de penalidade, repetindo-se o processo.

2.1 O método de Lagrangiano Aumentado

Para fixar idéias consideremos o seguinte problema simples de programação não-linear:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeito a } h(x) = 0, \end{aligned} \tag{2.3}$$

para as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ $h(x) = (h_1(x), \dots, h_m(x))$ e assumimos que as funções admitem derivadas primeiras contínuas.

Este tipo de problema é capaz de representar qualquer problema de otimização com restrições, pois restrições de desigualdade do tipo $g_i(x) \leq 0$ podem ser substituídas por $g_i(x) + z^2 = 0$.

A função de Lagrangiano aumentado mais utilizada na literatura associada ao problema (2.3) é definida por:

$$L(x, \lambda, \rho) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \frac{\rho}{2} \sum_{i=1}^m h_i(x)^2.$$

Observemos que L é a soma da função Lagrangiano com a penalidade quadrática das restrições de igualdade. Poderiam utilizar-se outras funções de penalização [5, 8, 20] mas neste trabalho consideramos somente a função de penalidade quadrática. As componentes λ_i de λ são conhecidas como *estimativas dos multiplicadores de Lagrange* e ρ é conhecido como *parâmetro de penalidade* e é o parâmetro que determina a severidade da penalidade.

No caso em que as funções são mal escaladas podem ser considerados vários parâmetros de penalidade, um parâmetro diferente para cada restrição ou por grupo de restrições que satisfaçam alguma regra e que será atualizado dependendo do progresso individual ou, respectivamente, grupal.

No caso em que existem restrições simples, como restrições lineares ou de canalização,

estas podem não ser penalizadas pois é possível trabalhar com elas eficientemente e, em geral, existem bons algoritmos para resolver problemas diretamente com este tipo de restrições. Logo são penalizadas somente as restrições não-lineares.

Um método de Lagrangiano aumentado para resolver o problema (2.3) utilizando a função L pode ser definido da seguinte forma:

Método de Lagrangiano aumentado

Sejam $x_0 \in \mathbb{R}^n$ um ponto inicial arbitrário, $\tau \in (0, 1]$, $\gamma > 1$, $\rho_1 > 0$, $\lambda_1 \in \mathbb{R}^m$. Inicializar $k \leftarrow 1$.

Passo 1. *Resolver o subproblema*

Usando x_{k-1} como aproximação inicial,

$$\text{Minimizar (aproximadamente) } L(x, \lambda_k, \rho_k), \quad (2.4)$$

obtendo x_k como solução.

Passo 2. *Estimar os multiplicadores*

Definir uma nova estimativa do multiplicador de Lagrange λ_{k+1} , por exemplo, utilizar o estimado de primeira ordem:

$$\lambda_{k+1} = \lambda_k + \rho_k h(x_k).$$

Passo 3. *Atualizar o parâmetro de penalidade*

Considerar alguma medida de inviabilidade para analisar o progresso das restrições que foram penalizadas, por exemplo:

se

$$\|h(x_k)\|_\infty \leq \tau \|h(x_{k-1})\|_\infty$$

definir

$$\rho_{k+1} = \rho_k,$$

senão, definir

$$\rho_{k+1} = \gamma \rho_k.$$

Passo 4. *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Ir ao Passo 1.

Logo, um método de Lagrangiano aumentado é definido em, basicamente, três passos:

no primeiro, o método procura uma solução aproximada da função Lagrangiano aumentado; no segundo, atualiza a estimativa do multiplicador de Lagrange utilizando alguma regra de atualização e finalmente, no terceiro, define alguma medida de inviabilidade para incrementar ou não o parâmetro de penalidade.

Observar que, no caso em que nem todas as restrições são penalizadas, o subproblema (2.4) deixa de ser irrestrito.

2.1.1 A função Lagrangiano aumentado

Nossa proposta para resolver o problema (2.1) é utilizar um método de Lagrangiano aumentado no qual são penalizadas somente as restrições do conjunto complexo Ω_1 . Neste conjunto há restrições de desigualdade, logo precisamos definir uma função de Lagrangiano aumentado para restrições de desigualdade. Neste caso utilizaremos a conhecida função de Powell-Hestenes-Rockafellar (PHR) [22, 31, 33] que no trabalho [8] demonstrou ter o melhor desempenho ao ser comparada com mais de 60 funções de Lagrangiano aumentado.

Faremos a seguir a dedução da função PHR já que é interessante observar como ela se deduz da aplicação da função de penalidade quadrática em um problema equivalente que tem somente restrições de igualdade. Para fixar idéias, consideremos o seguinte problema de programação não-linear incluindo ambas restrições, de igualdade e de desigualdade:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeito a } h_i(x) = 0, \quad i = 1, \dots, m \\ &\quad g_j(x) \leq 0, \quad j = 1, \dots, p. \end{aligned} \tag{2.5}$$

Mediante a introdução de uma variável adicional z_j associada com cada restrição de desigualdade, é possível converter o problema (2.5) em um problema com restrições de igualdade:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeito a } h_i(x) = 0, \quad i = 1, \dots, m, \\ &\quad g_j(x) + z_j^2 = 0, \quad j = 1, \dots, p. \end{aligned} \tag{2.6}$$

Logo, o problema (2.6) é um problema apenas com restrições de igualdade e sabemos como é definida a função de Lagrangiano aumentado que utiliza a função de penalidade quadrática para penalizar as restrições:

$$\bar{L}(x, z, \lambda, \mu, \rho) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \frac{\rho}{2} \sum_{i=1}^m (h_i(x))^2 + \sum_{j=1}^p \mu_j [g_j(x) + z_j^2] + \frac{\rho}{2} \sum_{j=1}^p [g_j(x) + z_j^2]^2.$$

Claramente, a função $\bar{L}$ depende da variável $x \in \mathbb{R}^n$ e da variável adicional $z \in \mathbb{R}^p$.

No primeiro passo do método de Lagrangiano aumentado é resolvido o subproblema de achar um minimizador da função Lagrangiano aumentado:

$$\text{Minimizar } \bar{L}(x, z, \lambda_k, \mu_k, \rho_k). \quad (2.7)$$

Observando a função $\bar{L}$ vemos que ela é quadrática na variável z_j^2 para cada $j = 1, \dots, p$, logo a minimização (2.7) pode ser feita em cada z_j independentemente de x . A minimização com respeito a z_j é equivalente a

$$\text{Minimizar } \mu_j[g_j(x) + z_j^2] + \frac{\rho}{2}[g_j(x) + z_j^2]^2. \quad (2.8)$$

A função a ser minimizada é uma quadrática na variável z_j^2 e o candidato a minimizador global é o escalar $\hat{z}_j$ que anula a derivada:

$$2\mu_j\hat{z}_j + \rho[g_j(x) + \hat{z}_j^2]2\hat{z}_j = 0.$$

Logo,

$$\hat{z}_j = 0 \quad \text{ou} \quad \hat{z}_j^2 = -\mu_j/\rho - g_j(x).$$

Assim, podemos escrever a solução do problema da seguinte maneira

$$z_j^2 = \max\{0, -\mu_j/\rho - g_j(x)\}$$

e então o z_j ótimo é:

$$z_j^* = \sqrt{\max\{0, -\mu_j/\rho - g_j(x)\}}. \quad (2.9)$$

Logo,

$$g_j(x) + (z_j^*)^2 = \max\{g_j(x), -\mu_j/\rho\},$$

e

$$\mu_j[g_j(x) + (z_j^*)^2] + \frac{\rho}{2}[g_j(x) + (z_j^*)^2]^2 = \frac{1}{2\rho}([\max\{0, \mu_j + \rho g_j(x)\}]^2 - \mu_j^2).$$

Assim, quando penalizamos as restrições do conjunto Ω_1 utilizando um parâmetro diferente para cada restrição obtemos a seguinte função de Lagrangiano aumentado:

$$\begin{aligned} L(x, \lambda, \mu, \rho) = f(x) + \sum_{i=1}^{m_1} \lambda_i [h_1(x)]_i + \sum_{i=1}^{m_1} \frac{[\rho]_i}{2} ([h_1(x)]_i)^2 \\ + \sum_{i=1}^{p_1} \frac{1}{2[\rho]_{m_1+i}} ([\max\{0, \mu_i + [\rho]_{m_1+i}[g_1(x)]_i\}]^2 - \mu_i^2), \end{aligned} \quad (2.10)$$

para $x \in \mathbb{R}^n, \rho \in \mathbb{R}_{++}^{m_1+p_1}, \lambda \in \mathbb{R}^{m_1}, \mu \in \mathbb{R}_+^{p_1}$.

Observar que em (2.10) colocamos parâmetros de penalização diferentes para cada restrição mas na definição dos algoritmos consideraremos também o caso em que todas as restrições do conjunto complexo são penalizadas utilizando um único parâmetro.

2.1.2 O método proposto

A seguir apresentamos o algoritmo de Lagrangiano aumentado proposto para resolver o problema (2.1), [2]:

Algoritmo 2.1.1. *Seja $x_0 \in \mathbb{R}^n$ um ponto inicial arbitrário.*

Os parâmetros utilizados pelo algoritmo são os seguintes:

$$\tau \in [0, 1), \gamma > 1,$$

$$-\infty < [\bar{\lambda}_{\min}]_i < [\bar{\lambda}_{\max}]_i < \infty \quad \forall i = 1, \dots, m_1,$$

$$0 \leq [\bar{\mu}_{\min}]_i < [\bar{\mu}_{\max}]_i < \infty \quad \forall i = 1, \dots, p_1,$$

$$\rho_0, \rho_1 \in \mathbb{R}_{++}^{m_1+p_1},$$

$$[\bar{\lambda}_1]_i \in [[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i] \quad \forall i = 1, \dots, m_1,$$

$$[\bar{\mu}_0]_i, [\bar{\mu}_1]_i \in [0, [\bar{\mu}_{\max}]_i] \quad \forall i = 1, \dots, p_1.$$

Finalmente, $\{\varepsilon_k\} \subset \mathbb{R}_+$ é uma seqüência de escalares positivos tal que

$$\lim_{k \rightarrow \infty} \varepsilon_k = 0. \tag{2.11}$$

Passo 1. *Inicialização*

Fazer $k \leftarrow 1$. Para $j = 1, \dots, p_1$, calcular

$$[\sigma_0]_j = \max \left\{ [g_1(x_0)]_j, -\frac{[\bar{\mu}_0]_j}{[\rho_0]_{m_1+j}} \right\}. \tag{2.12}$$

Passo 2. *Resolver o subproblema*

Calcular (se for possível) $x_k \in \mathbb{R}^n$ tal que existem $v_k \in \mathbb{R}^{m_2}, u_k \in \mathbb{R}^{p_2}$ satisfazendo

$$\|\nabla L(x_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [g_2(x_k)]_i\|_\infty \leq \varepsilon_k, \quad (2.13)$$

$$[g_2(x_k)]_j \leq \varepsilon_k \text{ para todo } j = 1, \dots, p_2, \quad (2.14)$$

$$u_k \geq 0, \quad (2.15)$$

$$\text{se } [g_2(x_k)]_j < -\varepsilon_k \text{ então } [u_k]_j = 0 \text{ para todo } j = 1, \dots, p_2, \quad (2.16)$$

$$\|h_2(x_k)\|_\infty \leq \varepsilon_k. \quad (2.17)$$

Se não for possível calcular x_k satisfazendo (2.13)–(2.17), parar a execução do algoritmo.

Passo 3. Estimar os multiplicadores de Lagrange

Para todo $i = 1, \dots, m_1$, calcular

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i [h_1(x_k)]_i \quad (2.18)$$

e

$$[\bar{\lambda}_{k+1}]_i \in [[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]. \quad (2.19)$$

Para todo $j = 1, \dots, p_1$, calcular

$$[\mu_{k+1}]_j = \max\{0, [\bar{\mu}_k]_j + [\rho_k]_{m_1+j} [g_1(x_k)]_j\}, \quad (2.20)$$

$$[\sigma_k]_j = \max\left\{[g_1(x_k)]_j, -\frac{[\bar{\mu}_k]_j}{[\rho_k]_{m_1+j}}\right\}, \quad (2.21)$$

e

$$[\bar{\mu}_{k+1}]_j \in [0, [\bar{\mu}_{\max}]_j]. \quad (2.22)$$

Passo 4. Adaptar os parâmetros de penalidade

Utilizar uma das seguintes regras para adaptar os parâmetros de penalidade:

Regra 1. Se

$$\max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\} \leq \tau \max\{\|h_1(x_{k-1})\|_\infty, \|\sigma_{k-1}\|_\infty\},$$

definir

$$[\rho_{k+1}]_i = [\rho_k]_i, \quad i = 1, \dots, m_1 + p_1.$$

Senão, definir

$$[\rho_{k+1}]_i = \gamma [\rho_k]_i, \quad i = 1, \dots, m_1 + p_1.$$

Regra 2. Para cada $i = 1, \dots, m_1$, se

$$|[h_1(x_k)]_i| \leq \tau |[h_1(x_{k-1})]_i|,$$

definir

$$[\rho_{k+1}]_i = [\rho_k]_i.$$

Senão, definir

$$[\rho_{k+1}]_i = \gamma [\rho_k]_i.$$

Para cada $j = 1, \dots, p_1$, se

$$|[\sigma_k]_j| \leq \tau |[\sigma_{k-1}]_j|, \tag{2.23}$$

definir

$$[\rho_{k+1}]_{m_1+j} = [\rho_k]_{m_1+j}.$$

Senão, definir

$$[\rho_{k+1}]_{m_1+j} = \gamma [\rho_k]_{m_1+j}.$$

Passo 5. Começar uma nova iteração externa

Fazer $k \leftarrow k + 1$. Ir ao Passo 2.

Observações

1. A maior contribuição teórica do Algoritmo 2.1.1 é a resolução inexata dos subproblemas do Passo 2:

$$\text{Minimizar } L(x, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) \text{ sujeito a } x \in \Omega_2. \tag{2.24}$$

Em [5] foram analisados métodos de Lagrangiano aumentado considerando a resolução exata dos subproblemas.

Resolução inexata dos subproblemas a serem resolvidos em métodos de Lagrangiano aumentado foram considerados em [1] e em [16] (com diferentes resultados de convergência global) para resolver problemas com restrições de igualdade no conjunto complexo e de canalização no conjunto simples, e em [14] para resolver problemas com restrições de igualdade no conjunto complexo e lineares no conjunto simples.

2. Observar que as condições (2.13)–(2.17) aproximam as condições KKT do problema (2.24) no sentido da condição de otimalidade definida em 1.2.4. Logo, se a sequência $\{x_k\}$ tem um ponto limite, esse ponto satisfaz a condição de otimalidade AKKT-2 da Definição 1.2.4.

Na prova de convergência assumimos que somos capazes de calcular as seqüências x_k, v_k, u_k satisfazendo (2.13)–(2.17). A existência destas seqüências depende das características do conjunto Ω_2 e da função objetivo de (2.24). A primeira pergunta que nos fazemos é se é possível encontrar uma seqüência $\{x_k\}$ satisfazendo todas as condições ao mesmo tempo. Se o problema (2.24) tem solução, é natural pensar que existem algoritmos de otimização que conseguem aproximar tanto a solução quanto os multiplicadores de Lagrange correspondentes de modo a satisfazer todas as condições (2.13)–(2.17) com exceção, talvez, da condição (2.16).

Suponhamos que existe $\hat{x}$ ponto estacionário do problema (2.24). Logo, existem os multiplicadores de Lagrange $\hat{v} \in \mathbb{R}^{m_2}, \hat{u} \in \mathbb{R}_+^{p_2}$. Suponhamos também que é possível aproximar $(\hat{x}, \hat{v}, \hat{u})$ por uma seqüência convergente

$$(\hat{x}_\ell, \hat{v}_\ell, \hat{u}_\ell) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}_+^{p_1}, \ell \in \mathbb{N},$$

gerada por algum algoritmo numérico. Logo, temos que

$$\lim_{\ell \rightarrow \infty} (\hat{x}_\ell, \hat{v}_\ell, \hat{u}_\ell) = (\hat{x}, \hat{v}, \hat{u}). \quad (2.25)$$

Em princípio, é possível que nenhum elemento da seqüência $(\hat{x}_\ell, \hat{v}_\ell, \hat{u}_\ell)$ satisfaça a condição (2.16). Isto poderia acontecer no caso em que, para algum índice j ,

$$[g(\hat{x})]_j < -\varepsilon_k < 0,$$

e portanto $[\hat{u}]_j = 0$, mas $[\hat{u}_\ell]_j > 0$ para todo $\ell \in \mathbb{N}$. Ou seja, embora $\lim_{\ell \rightarrow \infty} [\hat{u}_\ell]_j = [\hat{u}]_j = 0$, os elementos da seqüência $[\hat{u}_\ell]_j$ são estritamente positivos para todo $\ell \in \mathbb{N}$.

Afortunadamente, esta situação pode ser superada da seguinte forma: para cada $\ell \in \mathbb{N}$, definimos $[\hat{u}'_\ell]_j$ como sendo:

$$[\hat{u}'_\ell]_j = \begin{cases} 0 & \text{se } [g(\hat{x}_\ell)]_j < -\varepsilon_k \\ [\hat{u}_\ell]_j & \text{em caso contrário.} \end{cases}$$

Claramente, substituindo $\hat{u}_\ell$ por $\hat{u}'_\ell$ satisfaz-se a condição (2.16) e tem-se que:

$$\lim_{\ell \rightarrow \infty} (\hat{x}_\ell, \hat{v}_\ell, \hat{u}'_\ell) = (\hat{x}, \hat{v}, \hat{u}).$$

Conseqüentemente, a seqüência $(\hat{x}_\ell, \hat{v}_\ell, \hat{u}'_\ell)$ satisfaz as condições (2.13), (2.14) e (2.17) para ℓ suficientemente grande.

3. As fórmulas (2.18) e (2.20) definem estimativas de primeira ordem do multiplicador

de Lagrange. Essa é uma maneira, entre várias, de predizer o multiplicador de Lagrange correto na solução e talvez não seja a mais precisa. Predições de ordem superior podem ser definidas (ver [4, 20]) mas usualmente elas requerem cálculos adicionais mais custosos.

4. A fórmula (2.21) nada mais é que o valor numérico da restrição de desigualdade $[g_1(x)]_j$ quando ela é convertida para restrição de igualdade mediante a introdução da variável adicional z_j utilizando o valor ótimo de z_j obtido em (2.9):

$$[g_1(x_k)]_j + (z_j^*)^2 = [\sigma_k]_j.$$

Logo, a medida de inviabilidade considerada em (2.23) é análoga à medida de inviabilidade $\|h(x)\|_\infty$ utilizada para restrições de igualdade, só que adequada para restrições de desigualdade.

2.2 Convergência para pontos viáveis

Gostaríamos de definir algoritmos que em todas as situações encontram pontos viáveis, mas isto é impossível, pois em casos extremos poderiam não existir pontos viáveis. Conseqüentemente, é importante estudar o comportamento dos algoritmos com respeito à viabilidade. Assim, gostaríamos de saber quais são os possíveis pontos que podem ser pontos limites do algoritmo.

Para obter os resultados de convergência do algoritmo proposto precisamos assumir que para todo $k \in \mathbb{N}$ é sempre possível encontrar um ponto x_k satisfazendo (2.13)-(2.17), ou seja, precisamos supor que o algoritmo não pára no Passo 2. É possível, em muitos casos, encontrar condições suficientes que garantam esta hipótese, mas, temos que advertir que em algumas situações o algoritmo necessariamente pode parar no Passo 2, por exemplo, no caso em que o conjunto simples é vazio.

Precisamos assumir também que existe pelo menos um ponto limite da seqüência gerada pelo algoritmo. Uma condição suficiente para isto é a existência de um escalar $\varepsilon > 0$ para o qual o conjunto

$$\{x \in \mathbb{R}^n \mid g_2(x) \leq \varepsilon, \|h_2(x)\| \leq \varepsilon\}$$

é limitado. Podemos assegurar o cumprimento desta condição mediante a adição de restrições artificiais no conjunto Ω_2 . Entretanto, ao fazer isso devemos ter a precaução de não destruir a estrutura do conjunto Ω_2 .

Provamos que ou um ponto limite da seqüência gerada pelo Algoritmo 2.1.1 é viável, ou, se a condição CPLD associada ao conjunto simples Ω_2 se verifica então o ponto limite

é um ponto estacionário de um problema de quadrados mínimos com pesos envolvendo as restrições do conjunto complexo, sujeito ao conjunto simples.

Teorema 2.2.1. *Seja $\{x_k\}$ a seqüência gerada pelo Algoritmo 2.1.1. Seja x_* um ponto limite de $\{x_k\}$. Então, se a seqüência de parâmetros de penalidade $\{\rho_k\}$ é limitada, o ponto limite x_* é viável. Senão, uma das seguintes possibilidades se verifica:*

(i) *Existe $w \in [0, 1]^{m_1+p_1}$, $w \neq 0$, tal que x_* é um ponto KKT do problema*

$$\begin{aligned} \text{Minimizar } \frac{1}{2} \left[\sum_{i=1}^{m_1} [w]_i [h_1(x)]_i^2 + \sum_{i=1}^{p_1} [w]_{m_1+i} \max\{0, [g_1(x)]_i\}^2 \right] \\ \text{sujeito a } x \in \Omega_2. \end{aligned} \quad (2.26)$$

Além disso, se em todas as iterações do Algoritmo 2.1.1 é utilizada a Regra 1 e $[\rho_0]_i = [\rho_0]_1$ para todo $i = 1, \dots, m_1 + p_1$, vale a propriedade descrita acima com $[w]_i = 1, i = 1, \dots, m_1 + p_1$.

(ii) *x_* não verifica a condição CPLD associada ao conjunto Ω_2 .*

Prova. Seja K um subconjunto infinito de $\mathbb{N}$ tal que

$$\lim_{k \in K} x_k = x_*.$$

Por (2.11), (2.14) e (2.17) temos que $g_2(x_*) \leq 0$ e $h_2(x_*) = 0$. Logo,

$$x_* \in \Omega_2. \quad (2.27)$$

Consideramos duas possibilidades:

- (a) A seqüência $\{\rho_k\}$ é limitada.
- (b) A seqüência $\{\rho_k\}$ é não limitada.

Consideremos primeiro (a). Neste caso, a partir de alguma iteração, os parâmetros de penalidade não são mais atualizados. Conseqüentemente, ambas regras de adaptação implicam que

$$\lim_{k \rightarrow \infty} [h_1(x_k)]_i = \lim_{k \rightarrow \infty} [\sigma_k]_j = 0, \quad \forall i = 1, \dots, m_1, \quad j = 1, \dots, p_1.$$

Logo,

$$h_1(x_*) = 0. \quad (2.28)$$

Além disso, se $[g_1(x_*)]_j > 0$ então $[g_1(x_k)]_j > 0$ para todo $k \in K$ suficientemente grande e isto contradiz o fato de que $[\sigma_k]_j \rightarrow 0$.

Portanto,

$$[g_1(x_*)]_i \leq 0 \quad \forall i = 1, \dots, p_1. \quad (2.29)$$

Por (2.27), (2.28) e (2.29) temos que $x_* \in \Omega_1 \cap \Omega_2$. Então, a função objetivo de (2.26) se anula em x_* e obtemos que x_* é um ponto estacionário de (2.26).

Consideremos agora o caso (b), no qual a seqüência de parâmetros de penalidade $\{\rho_k\}_{k \in K}$ é não limitada. Por (2.10) e (2.13) temos que existe $\{\delta_k\}$ tal que $\|\delta_k\|_\infty \leq \varepsilon_k$,

$$\begin{aligned} \nabla f(x_k) + \sum_{i=1}^{m_1} ([\bar{\lambda}_k]_i + [\rho_k]_i [h_1(x_k)]_i) \nabla [h_1(x_k)]_i + \sum_{i=1}^{p_1} \max\{0, [\bar{\mu}_k]_i + [\rho_k]_{m_1+i} [g_1(x_k)]_i\} \nabla [g_1(x_k)]_i \\ + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{j=1}^{p_2} [u_k]_j \nabla [g_2(x_k)]_j = \delta_k, \end{aligned} \quad (2.30)$$

e, por (2.11),

$$\lim_{k \in K} \|\delta_k\|_\infty = 0. \quad (2.31)$$

Se $[g_2(x_*)]_j < 0$, existe $k_1 \in \mathbb{N}$ tal que $[g_2(x_k)]_j < -\varepsilon_k$ para todo $k \geq k_1, k \in K$. Logo, por (2.16), $[u_k]_j = 0$ para todo $k \in K, k \geq k_1$. Assim, por (2.30), temos que, para todo $k \in K, k \geq k_1$

$$\begin{aligned} \nabla f(x_k) + \sum_{i=1}^{m_1} ([\bar{\lambda}_k]_i + [\rho_k]_i [h_1(x_k)]_i) \nabla [h_1(x_k)]_i + \sum_{i=1}^{p_1} \max\{0, [\bar{\mu}_k]_i + [\rho_k]_{m_1+i} [g_1(x_k)]_i\} \nabla [g_1(x_k)]_i \\ + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0} [u_k]_j \nabla [g_2(x_k)]_j = \delta_k. \end{aligned}$$

Dividindo por $\|\rho_k\|_\infty$ obtemos que:

$$\begin{aligned} \frac{\nabla f(x_k)}{\|\rho_k\|_\infty} + \sum_{i=1}^{m_1} \left(\frac{[\bar{\lambda}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_i}{\|\rho_k\|_\infty} [h_1(x_k)]_i \right) \nabla [h_1(x_k)]_i + \\ \sum_{i=1}^{p_1} \max \left\{ 0, \frac{[\bar{\mu}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_{m_1+i}}{\|\rho_k\|_\infty} [g_1(x_k)]_i \right\} \nabla [g_1(x_k)]_i + \\ \sum_{i=1}^{m_2} \frac{[v_k]_i}{\|\rho_k\|_\infty} \nabla [h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0} \frac{[u_k]_j}{\|\rho_k\|_\infty} \nabla [g_2(x_k)]_j = \frac{\delta_k}{\|\rho_k\|_\infty}. \end{aligned}$$

Pelo Lema de Caratheodory, (Lema 1.3.1), sabemos que existem conjuntos

$$\widehat{I}_k \subset \{1, \dots, m_2\}, \widehat{J}_k \subset \{j \mid [g_2(x_*)]_j = 0\},$$

e escalares

$$[\widehat{v}_k]_i, \quad i \in \widehat{I}_k \quad \text{e} \quad [\widehat{u}_k]_j \geq 0, \quad j \in \widehat{J}_k$$

tais que os vetores

$$\{\nabla[h_2(x_k)]_i\}_{i \in \widehat{I}_k} \cup \{\nabla[g_2(x_k)]_j\}_{j \in \widehat{J}_k}$$

são linearmente independentes e

$$\begin{aligned} & \frac{\nabla f(x_k)}{\|\rho_k\|_\infty} + \sum_{i=1}^{m_1} \left(\frac{[\bar{\lambda}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_i}{\|\rho_k\|_\infty} [h_1(x_k)]_i \right) \nabla[h_1(x_k)]_i + \\ & \sum_{i=1}^{p_1} \max \left\{ 0, \frac{[\bar{\mu}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_{m_1+i}}{\|\rho_k\|_\infty} [g_1(x_k)]_i \right\} \nabla[g_1(x_k)]_i + \\ & \sum_{i \in \widehat{I}_k} [\widehat{v}_k]_i \nabla[h_2(x_k)]_i + \sum_{j \in \widehat{J}_k} [\widehat{u}_k]_j \nabla[g_2(x_k)]_j = \frac{\delta_k}{\|\rho_k\|_\infty}. \end{aligned} \quad (2.32)$$

Como existe um número finito de possíveis subconjuntos $\widehat{I}_k, \widehat{J}_k$, existe um subconjunto infinito de índices K_1 tal que

$$K_1 \subset \{k \in K \mid k \geq k_1\},$$

$$\widehat{I}_k = \widehat{I},$$

e

$$\widehat{J}_k = \widehat{J} \subset \{j \mid [g_2(x_*)]_j = 0\} \quad (2.33)$$

para todo $k \in K_1$. Então, por (2.32), para todo $k \in K_1$ temos que:

$$\begin{aligned} & \frac{\nabla f(x_k)}{\|\rho_k\|_\infty} + \sum_{i=1}^{m_1} \left(\frac{[\bar{\lambda}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_i}{\|\rho_k\|_\infty} [h_1(x_k)]_i \right) \nabla[h_1(x_k)]_i + \\ & \sum_{i=1}^{p_1} \max \left\{ 0, \frac{[\bar{\mu}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_{m_1+i}}{\|\rho_k\|_\infty} [g_1(x_k)]_i \right\} \nabla[g_1(x_k)]_i + \\ & \sum_{i \in \widehat{I}} [\widehat{v}_k]_i \nabla[h_2(x_k)]_i + \sum_{j \in \widehat{J}} [\widehat{u}_k]_j \nabla[g_2(x_k)]_j = \frac{\delta_k}{\|\rho_k\|_\infty}, \end{aligned} \quad (2.34)$$

e os gradientes

$$\{\nabla[h_2(x_k)]_i\}_{i \in \hat{I}} \cup \{\nabla[g_2(x_k)]_j\}_{j \in \hat{J}} \text{ são linearmente independentes.} \quad (2.35)$$

Consideramos dois casos:

1. A sequência $\{\|(\hat{v}_k, \hat{u}_k)\|_\infty, k \in K_1\}$ é limitada.
2. A sequência $\{\|(\hat{v}_k, \hat{u}_k)\|_\infty, k \in K_1\}$ é não limitada.

Se a sequência $\{\|(\hat{v}_k, \hat{u}_k)\|_\infty\}_{k \in K_1}$ é limitada, existem $w \in [0, 1]^{m_1+p_1}$, $w \neq 0$, $(\hat{v}, \hat{u}), \hat{u} \geq 0$ e um subconjunto infinito de índices $K_2 \subset K_1$ tais que

$$\lim_{k \in K_2} (\hat{v}_k, \hat{u}_k) = (\hat{v}, \hat{u})$$

e

$$\lim_{k \in K_2} \frac{[\rho_k]_i}{\|\rho_k\|_\infty} = w_i, \quad i = 1, \dots, m_1 + p_1.$$

Por (2.19), (2.22) e (2.31), tomando limites para $k \in K_2$ em (2.34), obtemos que:

$$\begin{aligned} & \sum_{i=1}^{m_1} [w]_i [h_1(x_*)]_i \nabla[h_1(x_*)]_i + \sum_{i=1}^{p_1} [w]_{m_1+i} \max\{0, [g_1(x_*)]_i\} \nabla[g_1(x_*)]_i + \\ & \sum_{i \in \hat{I}} \hat{v}_i \nabla[h_2(x_*)]_i + \sum_{j \in \hat{J}} \hat{u}_j \nabla[g_2(x_*)]_j = 0. \end{aligned}$$

Portanto, por (2.27) e (2.33), x_* é um ponto estacionário de (2.26) como queríamos provar.

Observar que se a Regra 1 é utilizada para todo k e todos os parâmetros são iguais na primeira iteração, eles permanecem iguais para todo k . Logo, todas as componentes de w devem ser iguais e se verifica a tese para a Regra 1.

Assumimos agora que $\{\|(\hat{v}_k, \hat{u}_k)\|_\infty\}_{k \in K_1}$ é não limitada. Seja $K_3 \subset K_1$ tal que $\lim_{k \in K_3} \|(\hat{v}_k, \hat{u}_k)\| = \infty$ e $(\hat{v}, \hat{u}) \neq 0, \hat{u} \geq 0$ tais que

$$\lim_{k \in K_3} \frac{(\hat{v}_k, \hat{u}_k)}{\|(\hat{v}_k, \hat{u}_k)\|_\infty} = (\hat{v}, \hat{u}).$$

Dividindo ambos lados de (2.34) por $\|(\hat{v}_k, \hat{u}_k)\|_\infty$ e tomando limites para $k \in K_3$, obtemos que:

$$\sum_{i \in \hat{I}} \hat{v}_i \nabla[h_2(x_*)]_i + \sum_{j \in \hat{J}} \hat{u}_j \nabla[g_2(x_*)]_j = 0.$$

Então, os gradientes

$$\{\nabla[h_2(x_*)]_i\}_{i \in \hat{I}} \cup \{\nabla[g_2(x_*)]_j\}_{j \in \hat{J}}$$

são positivo-linearmente dependentes. Logo, por (2.35), x_* não satisfaz a condição CPLD associada ao conjunto Ω_2 . Assim, a prova está completa. $\square$

Observar que, no caso em que as restrições são divididas nos conjuntos Ω_1 e Ω_2 de modo que todo ponto viável do conjunto Ω_2 satisfaz a condição CPLD, o Teorema 2.2.1 diz que um ponto limite ou é viável ou é um ponto estacionário do problema de achar o mínimo das inviabilidades do conjunto complexo sujeita à viabilidade do conjunto simples. Logo, se os únicos pontos estacionários dessa classe são os pontos viáveis, garantimos convergência para pontos viáveis.

2.3 Convergência para pontos ótimos

Uma vez caracterizados os possíveis pontos limites do algoritmo, estamos interessados em pesquisar sob quais condições um ponto limite viável da sequência gerada pelo Algoritmo 2.1.1 é um ponto estacionário do problema original. O resultado principal estabelece que um ponto limite viável é um ponto KKT se ele satisfaz a condição de qualificação de dependência linear positiva constante CPLD.

Teorema 2.3.1. *Seja $\{x_k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 2.1.1. Assumimos que $x_* \in \Omega_1 \cap \Omega_2$ é um ponto limite que satisfaz a condição de qualificação CPLD associada ao conjunto viável $\Omega_1 \cap \Omega_2$. Então, x_* é um ponto KKT do problema original (2.1).*

Prova. Para todo $k \in \mathbb{N}$, pela definição de L e por (2.11), (2.13), (2.15), (2.18) e (2.20), existem $v_k \in \mathbb{R}^{m_2}$ e $u_k \in \mathbb{R}_+^{p_2}$, tais que

$$\left\| \nabla f(x_k) + \sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{i=1}^{p_1} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i \right. \\ \left. + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{j=1}^{p_2} [u_k]_j \nabla [g_2(x_k)]_j \right\|_{\infty} \rightarrow 0. \quad (2.36)$$

Por (2.20), temos que $\mu_{k+1} \in \mathbb{R}_+^{p_1}$ para todo $k \in \mathbb{N}$.

Seja $K \subset \mathbb{N}$ tal que

$$\lim_{k \in K} x_k = x_*.$$

Suponhamos que $[g_2(x_*)]_i < 0$. Então, existe $k_1 \in \mathbb{N}$ tal que $\forall k \in K, k \geq k_1$, $[g_2(x_k)]_i < -\varepsilon_k$. Então, por (2.16), temos que

$$[u_k]_i = 0 \quad \forall k \in K, k \geq k_1.$$

Provemos que uma propriedade similar se verifica quando $[g_1(x_*)]_i < 0$. Neste caso, existe $k_2 \geq k_1$ tal que

$$[g_1(x_k)]_i < 0 \quad \forall k \in K, k \geq k_2. \quad (2.37)$$

Consideramos duas possibilidades:

1. A sequência $\{[\rho_k]_{m_1+i}\}$ é não limitada.
2. A sequência $\{[\rho_k]_{m_1+i}\}$ é limitada.

No primeiro caso temos que $\lim_{k \in K} [\rho_k]_{m_1+i} = \infty$. Como $\{[\bar{\mu}_k]_i\}$ é limitada, existe $k_3 \geq k_2$ tal que, para todo $k \in K, k \geq k_3$,

$$[\bar{\mu}_k]_i + [\rho_k]_{m_1+i} [g_1(x_k)]_i < 0.$$

Pela definição de μ_{k+1} temos que

$$[\mu_{k+1}]_i = \max\{0, [\bar{\mu}_k]_i + [\rho_k]_{m_1+i} [g_1(x_k)]_i\} = 0 \quad \forall k \in K, k \geq k_3.$$

Consideremos agora o caso em que $\{[\rho_k]_{m_1+i}\}$ é limitada. Neste caso,

$$0 = \lim_{k \rightarrow \infty} [\sigma_k]_i = \lim_{k \rightarrow \infty} \max \left\{ [g_1(x_k)]_i, -\frac{[\bar{\mu}_k]_i}{[\rho_k]_{m_1+i}} \right\}.$$

Logo, por (2.37), temos que deve ser

$$\lim_{k \in K} [\bar{\mu}_k]_i = 0.$$

Então, para $k \in K$ suficientemente grande,

$$[\bar{\mu}_k]_i + [\rho_k]_{m_1+i} [g_1(x_k)]_i < 0$$

e pela definição de μ_{k+1} temos que existe $k_4 \in \mathbb{N}$ tal que $[\mu_{k+1}]_i = 0$ for $k \in K, k \geq k_4$.

Portanto, existe $k_5 \geq \max\{k_3, k_4\}$ tal que, para todo $k \in K, k \geq k_5$,

$$[g_1(x_*)]_i < 0 \Rightarrow [\mu_{k+1}]_i = 0 \text{ e } [g_2(x_*)]_i < 0 \Rightarrow [u_k]_i = 0. \quad (2.38)$$

Observar que, até aqui, não utilizamos a condição CPLD, logo a condição (2.38) se verifica sem hipótese de qualificação sobre as restrições.

Como o ponto limite x_* é viável temos que

$$\lim_{k \in K, k \rightarrow \infty} \|h_1(x_k)\|_\infty = \lim_{k \in K, k \rightarrow \infty} \|h_2(x_k)\|_\infty = 0, \quad (2.39)$$

$$\lim_{k \in K, k \rightarrow \infty} \|g_1(x_k)_+\|_\infty = \lim_{k \in K, k \rightarrow \infty} \|g_2(x_k)_+\|_\infty = 0. \quad (2.40)$$

Logo, provamos que existem seqüências $\{(x_k, (\lambda_{k+1}, v_k), (\mu_{k+1}, u_k))\} \in \mathbb{R}^n \times \mathbb{R}^{m_1+m_2} \times \mathbb{R}^{p_1+p_2}$ satisfazendo (2.36), (2.38), (2.39) e (2.40). Assim, temos que x_* satisfaz a hipótese do Teorema 1.3.3 o que prova o resultado desejado. $\square$

Resultados de convergência global que utilizam a condição de qualificação CPLD são mais fortes que resultados de convergência global de métodos de Lagrangiano aumentado para resolver problemas mas específicos que utilizam condições de qualificação mais fortes como, por exemplo, a regularidade. Em particular, a hipótese AS3 utilizada no método de Lagrangiano aumentado proposto em [16] para resolver problemas com restrições de igualdade como conjunto complexo, e de canalização como conjunto simples, quando aplicada somente para pontos viáveis, é a clássica hipótese de regularidade. Da mesma maneira, em [14], Conn, Gould, Sartenaer e Toint provam convergência global de um método de Lagrangiano aumentado para resolver problemas com restrições de igualdade e lineares utilizando a regularidade. O Teorema 2.3.1 mostra que uma condição de qualificação bem menos restritiva que LICQ pode ser utilizada para obter o mesmo resultado.

Corolário 2.3.2. *Sejam $\{x_k\}_{k \in \mathbb{N}}$ uma seqüência gerada pelo Algoritmo 2.1.1 e $x_* \in \Omega_1 \cap \Omega_2$ um ponto limite que satisfaz a condição de qualificação de Mangasarian-Fromovitz. Então, se $\{x_k\}_{k \in K}$ é uma subsequência que converge a x_* , o conjunto*

$$\{\|\lambda_{k+1}\|_\infty, \|\mu_{k+1}\|_\infty, \|v_k\|_\infty, \|u_k\|_\infty\}_{k \in K} \text{ é limitado.} \quad (2.41)$$

Prova. A prova segue da prova do teorema anterior. Definimos

$$B_k = \max\{\|\lambda_{k+1}\|_\infty, \|\mu_{k+1}\|_\infty, \|v_k\|_\infty, \|u_k\|_\infty\}_{k \in K}.$$

Se (2.41) não se cumpre, temos que $\lim_{k \in K} B_k = \infty$. Por (2.36) e (2.38) temos que existe $\delta_k \in \mathbb{R}^n$ tal que $\|\delta_k\|_\infty \leq \varepsilon_k$ e

$$\begin{aligned} \nabla f(x_k) + \sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{[g_1(x_*)]_i=0}^{p_1} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i + \\ \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0} [u_k]_j \nabla [g_2(x_k)]_j = \delta_k, \end{aligned}$$

Dividindo ambos lados desta igualdade por B_k e tomando limites para uma subsequência

apropriada, obtemos que os gradientes

$$\{\nabla[h_1(x_*)]_i\}_{i=1}^{m_1} \cup \{\nabla[g_1(x_*)]_i\}_{[g_1(x_*)]_i=0} \cup \{\nabla[h_2(x_*)]_i\}_{i=1}^{m_2} \cup \{\nabla[g_2(x_*)]_i\}_{[g_2(x_*)]_i=0}$$

são positivo-linearmente dependentes. Logo, x_* não satisfaz a condição de qualificação de Mangasarian-Fromovitz. $\square$

2.4 Limitação do parâmetro de penalidade

Quando os parâmetros de penalidade de métodos de Lagrangiano aumentado são muito grandes os subproblemas tornam-se mal condicionados e difíceis de serem resolvidos eficientemente. Além disso, nosso método é um método de Lagrangiano aumentado e não de penalidade externa no qual é necessário que o parâmetro cresça ilimitadamente para obter a convergência. Nesta seção vamos obter condições apropriadas sob as quais garantimos que o parâmetro de penalidade do Algoritmo 2.1.1, utilizando a Regra 1, mantêm-se limitado. A técnica será a redução ao caso com restrições somente de igualdade, [2].

2.4.1 Problema equivalente

O seguinte problema de programação com restrições de igualdade é equivalente ao problema (2.1):

$$\text{Minimizar } F(y) \text{ sujeito a } H_1(y) = 0, G_1(y) = 0, H_2(y) = 0, G_2(y) = 0, \quad (2.42)$$

sendo

$$y = (x, w, z) \in \mathbb{R}^{n+p_1+p_2},$$

$$F(y) = f(x), H_1(y) = h_1(x), H_2(y) = h_2(x),$$

$$[G_1(y)]_i = [g_1(x)]_i + w_i^2, i = 1, \dots, p_1,$$

$$[G_2(y)]_i = [g_2(x)]_i + z_i^2, i = 1, \dots, p_2.$$

Consideraremos a aplicação do Algoritmo 2.1.1 ao problema (2.42). Neste caso as restrições do conjunto complexo são:

$$H_1(y) = 0, G_1(y) = 0,$$

enquanto as do conjunto simples são:

$$H_2(y) = 0, \quad G_2(y) = 0.$$

Consideramos também que, quando o Algoritmo 2.1.1 é aplicado ao problema (2.42), os parâmetros $\tau, \gamma, \rho_0, \rho_1$ são os mesmos que foram utilizados na aplicação do algoritmo ao problema original (2.1). A estimativa inicial do vetor de multiplicadores de Lagrange para as restrições do conjunto Ω_1 no problema (2.42) é $(\bar{\lambda}_1, \bar{\mu}_1) \in \mathbb{R}^{m_1+p_1}$. A cota inferior do intervalo no qual é projetada a estimativa do multiplicador de Lagrange é $(\bar{\lambda}_{\min}, \bar{\mu}_{\min}) \in \mathbb{R}^{m_1+p_1}$ e a cota superior é $(\bar{\lambda}_{\max}, \bar{\mu}_{\max}) \in \mathbb{R}^{m_1+p_1}$. Finalmente, na aplicação do Algoritmo ao problema (2.42), usamos o mesmo ponto inicial x_0 utilizado na aplicação ao problema (2.1) e completamos o vetor $y_0 = (x_0, w_0, z_0)$ definindo

$$[w_0]_i = \sqrt{\max \left\{ 0, -[g_1(x_0)]_i - \frac{[\bar{\mu}_0]_i}{[\rho_0]_{m_1+i}} \right\}}, i = 1, \dots, p_1,$$

$$[z_0]_i = \sqrt{\max \{ 0, -[g_2(x_0)]_i \}}, i = 1, \dots, p_2.$$

Teorema 2.4.1. *Assumimos que $\{x_k\}_{k \in \mathbb{N}}$ é a seqüência gerada pelo Algoritmo 2.1.1 quando aplicado ao problema (2.1). Suponhamos que $\lim_{k \rightarrow \infty} x_k = x_* \in \Omega_1 \cap \Omega_2$ e que ao menos uma das seguintes condições se verifica:*

1. *Não há restrições de desigualdade na definição do conjunto Ω_2 (ou seja $p_2 = 0$).*
2. *O ponto x_* satisfaz a condição de qualificação de Mangasarian-Fromovitz. Neste caso, seja $C > 0$ tal que*

$$[u_k]_i \leq C \text{ para todo } i = 1, \dots, p_2, k \in \mathbb{N}. \quad (2.43)$$

(A constante C existe pelo Corolário 2.3.2)

Para cada $k \in \mathbb{N}$ definimos:

$$[w_k]_i = \sqrt{\max \left\{ 0, -[g_1(x_k)]_i - \frac{[\bar{\mu}_k]_i}{[\rho_k]_{m_1+i}} \right\}}, i = 1, \dots, p_1, \quad (2.44)$$

$$[z_k]_i = \sqrt{\max \{ 0, -[g_2(x_k)]_i \}}, i = 1, \dots, p_2, \quad (2.45)$$

$$y_k = (x_k, w_k, z_k).$$

Então, a seqüência $\{y_k\}$ pode ser gerada quando o Algoritmo 2.1.1 é aplicado ao problema (2.42) com a mesma seqüência de parâmetros de penalidade e a tolerância $\{\hat{\epsilon}_k\}$

definida por

$$\hat{\varepsilon}_k = \max\{\varepsilon_k, 2C\sqrt{\varepsilon_k}\}.$$

Prova. Penalizando as restrições do conjunto complexo para o problema (2.42) temos que a função Lagrangiano aumentado é definida por:

$$L_E(y, \lambda, \mu, \rho) = F(y) + \sum_{i=1}^{m_1} \lambda_i [H_1(y)]_i + \sum_{i=1}^{m_1} \frac{[\rho]_i}{2} [H_1(y)]_i^2 + \sum_{i=1}^{p_1} \mu_i [G_1(y)]_i + \sum_{i=1}^{p_1} \frac{[\rho]_{m_1+i}}{2} [G_1(y)]_i^2.$$

Usando que x_k satisfaz (2.13)-(2.17), temos que provar que y_k satisfaz:

(i)

$$\left\| \nabla L_E(y_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) + \sum_{i=1}^{m_2} [v_k]_i \nabla [H_2(y_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [G_2(y_k)]_i \right\|_{\infty} \leq \hat{\varepsilon}_k,$$

(ii)

$$\|H_2(y_k)\|_{\infty} \leq \hat{\varepsilon}_k,$$

(iii)

$$\|G_2(y_k)\|_{\infty} \leq \hat{\varepsilon}_k.$$

Definimos a função $\phi : \mathbb{R}^n \times \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \rightarrow \mathbb{R}^{n+p_1+p_2}$ por

$$\phi(x_k, w_k, z_k) = \nabla L_E(y_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) + \sum_{i=1}^{m_2} [v_k]_i \nabla [H_2(y_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [G_2(y_k)]_i. \quad (2.46)$$

Considerando as primeiras n componentes de ϕ (aquelas correspondentes à variável x) temos que

$$\begin{aligned} [\phi(x_k, w_k, z_k)]_x &= \nabla f(x_k) + \sum_{i=1}^{m_1} [\bar{\lambda}_k]_i \nabla [h_1(x_k)]_i + \sum_{i=1}^{m_1} [\rho_k]_i [h_1(x_k)]_i \nabla [h_1(x_k)]_i + \\ &+ \sum_{i=1}^{p_1} [\bar{\mu}_k]_i \nabla [g_1(x_k)]_i + \sum_{i=1}^{p_1} [\rho_k]_{m_1+i} ([g_1(x_k)]_i + [w_k]_i^2) \nabla [g_1(x_k)]_i + \\ &\sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [g_2(x_k)]_i. \end{aligned}$$

Utilizando (2.44), (2.18) e (2.20) temos que

$$[\phi(x_k, w_k, z_k)]_x = \nabla f(x_k) + \sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{i=1}^{p_1} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i +$$

$$\sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [g_2(x_k)]_i.$$

Logo, por (2.13),

$$\begin{aligned} \|\phi(x_k, w_k, z_k)\|_\infty &= \left\| \nabla f(x_k) + \sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{i=1}^{p_1} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i + \right. \\ &\quad \left. + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [g_2(x_k)]_i \right\|_\infty \leq \varepsilon_k \leq \hat{\varepsilon}_k. \end{aligned}$$

Considerando a componente $n+i, i=1, \dots, p_1$ de ϕ e utilizando a definição de w_k temos que

$$[\phi(x_k, w_k, z_k)]_{w_i} = 2[\bar{\mu}_k]_i [w_k]_i + 2[\rho_k]_{m_1+i} [w_k]_i ([g_1(x_k)]_i + [w_k]_i^2) = 0 \leq \hat{\varepsilon}_k. \quad (2.47)$$

Considerando a componente $n+p_1+i, i=1, \dots, p_2$ de ϕ e utilizando a definição de z_k temos que

$$[\phi(x_k, w_k, z_k)]_{z_i} = 2[z_k]_i [u_k]_i. \quad (2.48)$$

Temos que provar que $2[z_k]_i [u_k]_i \leq \hat{\varepsilon}_k, \forall i=1, \dots, p_2$.

Se $[g_2(x_k)]_i < -\varepsilon_k$, por (2.16), temos que $[u_k]_i = 0$.

Se $[g_2(x_k)]_i \geq 0$, por definição de $[z_k]_i$, temos que $[z_k]_i = 0$.

Consideremos o caso em que $-\varepsilon_k \leq [g_2(x_k)]_i < 0$. Como $[u_k]_i, [z_k]_i \geq 0$, por (2.43) e (2.45), temos que:

$$0 \leq 2[z_k]_i [u_k]_i = 2\sqrt{-[g_2(x_k)]_i} [u_k]_i \leq 2\sqrt{\varepsilon_k} [u_k]_i \leq 2\sqrt{\varepsilon_k} C \leq \hat{\varepsilon}_k.$$

Logo, $\|\phi(x_k, y_k, z_k)\|_\infty \leq \max\{\varepsilon_k, 2C\sqrt{\varepsilon_k}\} = \hat{\varepsilon}_k$ o que mostra que se verifica (i).

A condição (ii) é imediata de (2.17) e da definição de $\hat{\varepsilon}_k$ e a condição (iii) segue de (2.14), da definição de $\hat{\varepsilon}_k$ e de $[z_k]_i$. $\square$

2.4.2 Limitação para problemas com restrições de igualdade

Devido à identificação da sequência gerada pelo Algoritmo 2.1.1 quando aplicado ao problema (2.1) com uma sequência gerada pelo mesmo algoritmo aplicado a um problema com restrições somente de igualdade, a prova de limitação do parâmetro se reduz a essa prova considerando restrições somente de igualdade.

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}, c_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}, c_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$. Consideremos o seguinte

problema de otimização com restrições de igualdade

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } c_1(x) = 0, c_2(x) = 0. \end{aligned} \tag{2.49}$$

Definimos a função Lagrangiano do problema (2.49)

$$L_0(x, \lambda, v) = f(x) + \sum_{i=1}^{m_1} \lambda_i [c_1(x)]_i + \sum_{i=1}^{m_2} v_i [c_2(x)]_i,$$

a função de Lagrangiano aumentado penalizando todas as restrições com um único parâmetro de penalidade

$$L(x, \lambda, v, \rho) = f(x) + \sum_{i=1}^{m_1} \lambda_i [c_1(x)]_i + \sum_{i=1}^{m_2} v_i [c_2(x)]_i + \frac{\rho}{2} \left(\sum_{i=1}^{m_1} [c_1(x)]_i^2 + \sum_{i=1}^{m_2} [c_2(x)]_i^2 \right)$$

e a função

$$L_1(x, \lambda, \rho) = f(x) + \sum_{i=1}^{m_1} \lambda_i [c_1(x)]_i + \frac{\rho}{2} \sum_{i=1}^{m_1} [c_1(x)]_i^2$$

considerando somente as restrições $c_1(x) = 0$.

Hipótese 2.4.2. *Assumimos que*

1. *As funções f, c_1 e c_2 admitem derivadas segundas contínuas numa vizinhança de x_* .*
2. *x_* é um minimizador local de (2.49).*
3. *Os gradientes $\nabla[c_1(x_*)]_1, \dots, \nabla[c_1(x_*)]_{m_1}, \nabla[c_2(x_*)]_1, \dots, \nabla[c_2(x_*)]_{m_2}$ são linearmente independentes.*
4. *Verifica-se a condição suficiente de segunda ordem para minimizadores locais: se λ_*, v_* são os multiplicadores de Lagrange associados com x_* , então:*

$$z^T \nabla_{xx}^2 L_0(x_*, \lambda_*, v_*) z > 0$$

para todo $z \neq 0$ tal que $\nabla c_1(x_)^T z = 0$ e $\nabla c_2(x_*)^T z = 0$.*

Para obter limitação do parâmetro, seguiremos as idéias da Proposição 2.4 de [5]. Para isso precisamos de alguns lemas preparatórios.

Lema 2.4.3. [5] *Assumimos que se verifica a Hipótese 2.4.2. Então existe um escalar $\bar{\rho} > 0$ tal que*

$$z^T \nabla_{xx}^2 L(x_*, \lambda_*, v_*, \rho) z > 0 \quad \forall z \neq 0, \nabla c_2(x_*)^T z = 0$$

para todo $\rho \geq \bar{\rho}$.

Prova. Suponhamos, pelo contrário, que para todo k existe x_k tal que $\|x_k\| = 1$, $\nabla c_2(x_*)^T x_k = 0$ e

$$x_k^T \nabla_{xx}^2 L(x_*, \lambda_*, v_*, k) x_k \leq 0. \quad (2.50)$$

A sequência $\{x_k\}$ é limitada e portanto tem uma subsequência convergente. Logo, existem um subconjunto infinito $K \subset \mathbb{N}$ e um ponto $\bar{x}$ tais que

$$\lim_{k \in K} x_k = \bar{x}.$$

Como $\|x_k\| = 1$ e $\nabla c_2(x_*)^T x_k = 0$ temos que $\bar{x} \neq 0$ e $\nabla c_2(x_*)^T \bar{x} = 0$.

Pela definição de L temos que

$$\begin{aligned} \nabla_{xx}^2 L(x, \lambda, v, \rho) &= \nabla^2 f(x) + \sum_{i=1}^{m_1} \lambda_i \nabla^2 [c_1(x)]_i + \sum_{i=1}^{m_2} v_i \nabla^2 [c_2(x)]_i + \\ &\rho \left[\sum_{i=1}^{m_1} [c_1(x)]_i \nabla^2 [c_1(x)]_i + \sum_{i=1}^{m_2} [c_2(x)]_i \nabla^2 [c_2(x)]_i + \right. \\ &\left. \sum_{i=1}^{m_1} \nabla [c_1(x)]_i \nabla [c_1(x)]_i^T + \sum_{i=1}^{m_2} \nabla [c_2(x)]_i \nabla [c_2(x)]_i^T \right]. \end{aligned}$$

Logo, usando a viabilidade de x_* , por (2.50) temos que

$$\lim_{k \rightarrow \infty} x_k^T \nabla_{xx}^2 L(x_*, \lambda_*, v_*, k) x_k = \bar{x}^T \left(\nabla^2 f(x_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla^2 [c_1(x_*)]_i + \sum_{i=1}^{m_2} [v_*]_i \nabla^2 [c_2(x_*)]_i \right) \bar{x} + \quad (2.51)$$

$$\limsup_{k \in K} k x_k^T \left(\sum_{i=1}^{m_1} \nabla [c_1(x_k)]_i \nabla [c_1(x_k)]_i^T + \sum_{i=1}^{m_2} \nabla [c_2(x_k)]_i \nabla [c_2(x_k)]_i^T \right) x_k \leq 0.$$

Mas,

$$x_k^T \left(\sum_{i=1}^{m_1} \nabla [c_1(x_k)]_i \nabla [c_1(x_k)]_i^T + \sum_{i=1}^{m_2} \nabla [c_2(x_k)]_i \nabla [c_2(x_k)]_i^T \right) x_k \geq 0, \forall k,$$

logo (2.51) implica que a sequência

$$\{x_k^T \left(\sum_{i=1}^{m_1} \nabla [c_1(x_k)]_i \nabla [c_1(x_k)]_i^T + \sum_{i=1}^{m_2} \nabla [c_2(x_k)]_i \nabla [c_2(x_k)]_i^T \right) x_k\}_{k \in K}$$

converge para zero e então

$$\bar{x}^T \left(\sum_{i=1}^{m_1} \nabla[c_1(x_*)]_i \nabla[c_1(x_*)]_i^T + \sum_{i=1}^{m_2} \nabla[c_2(x_*)]_i \nabla[c_2(x_*)]_i^T \right) \bar{x} = 0$$

Assim, temos que $\nabla c_1(x_*)^T \bar{x} = \nabla c_2(x_*)^T \bar{x} = 0$ e pela hipótese deve ocorrer

$$\bar{x}^T \nabla_{xx}^2 L_0(x_*, \lambda_*, v_*) \bar{x} = \bar{x}^T \left(\nabla^2 f(x_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla^2[c_1(x_*)]_i + \sum_{i=1}^{m_2} [v_*]_i \nabla^2[c_2(x_*)]_i \right) \bar{x} > 0$$

o que contradiz (2.51). $\square$

Lema 2.4.4. [5] *Assumimos que se verifica a Hipótese 2.4.2. Seja $\bar{\rho}$ o escalar obtido no Lema 2.4.3. Então, para todo $\gamma \in [0, \frac{1}{\bar{\rho}}]$ a matriz*

$$J = \begin{pmatrix} \nabla_{xx}^2 L_0(x_*, \lambda_*, v_*) & \nabla c_1(x_*) & \nabla c_2(x_*) \\ \nabla c_1(x_*)^T & -\gamma I & 0 \\ \nabla c_2(x_*)^T & 0 & 0 \end{pmatrix}$$

é não-singular.

Prova. Seja $\gamma = 0$. Se J é singular então existe um vetor $\bar{x} = (y, z, w) \in \mathbb{R}^{n+m_1+m_2}$, $(y, z, w) \neq 0$ tal que $J\bar{x} = 0$. Logo:

$$\nabla_{xx}^2 L_0(x_*, \lambda_*, v_*)y + \nabla c_1(x_*)z + \nabla c_2(x_*)w = 0, \quad (2.52)$$

$$\nabla c_1(x_*)^T y = 0, \quad (2.53)$$

$$\nabla c_2(x_*)^T y = 0. \quad (2.54)$$

Pré-multiplicando (2.52) por y^T e usando (2.53) e (2.54) obtemos que

$$y^T \nabla_{xx}^2 L_0(x_*, \lambda_*, v_*)y = 0.$$

Logo, pelo item 4 da hipótese deve ocorrer $y = 0$. Assim, por (2.52) e o item 3 da hipótese obtemos que $z = w = 0$, o que contradiz o fato de que $(y, z, w) \neq 0$. Logo J é não-singular.

Suponhamos que $\gamma \in (0, \frac{1}{\bar{\rho}}]$. Como antes, suponhamos que existe $\bar{x} = (y, z, w) \in \mathbb{R}^{n+m_1+m_2}$, $(y, z, w) \neq 0$ tal que $J\bar{x} = 0$. Logo:

$$\nabla_{xx}^2 L_0(x_*, \lambda_*, v_*)y + \nabla c_1(x_*)z + \nabla c_2(x_*)w = 0, \quad (2.55)$$

$$\nabla c_1(x_*)^T y - \gamma z = 0, \quad (2.56)$$

$$\nabla c_2(x_*)^T y = 0. \quad (2.57)$$

Por (2.56) temos que $z = \frac{1}{\gamma} \nabla c_1(x_*)^T y$. Substituindo este valor em (2.55), pré-multiplicando por y^T e usando (2.57) obtemos que

$$y^T \left[\nabla_{xx}^2 L_0(x_*, \lambda_*, v_*) + \frac{1}{\gamma} \left(\nabla c_1(x_*) \nabla c_1(x_*)^T + \nabla c_2(x_*) \nabla c_2(x_*)^T \right) \right] y = 0.$$

Logo, para todo $\gamma = \frac{1}{\rho}$ com $\rho \geq \bar{\rho}$, por (2.57) e pelo Lema 2.4.3 temos que deve ocorrer $y = 0$. Assim, da definição de z temos que $z = 0$ e pelo item 3 da hipótese temos $w = 0$, o que contradiz o fato de que $(y, z, w) \neq 0$. Logo, J é não-singular, como queríamos provar. $\square$

Utilizaremos na prova do próximo teorema o Segundo Teorema da Função Implícita (ver [5, p. 12]) enunciado abaixo para completeza do texto.

Para um subconjunto $A \subset \mathbb{R}^n$ e um escalar $r > 0$ definimos o conjunto

$$S(A, r) = \{x \in \mathbb{R}^n : \|x - a\| < r, a \in A\}$$

Se $h : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$ é uma função definida nas variáveis (x, y) sendo $x = (x_1, \dots, x_m) \in \mathbb{R}^m$, $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ denotamos

$$\nabla_y h(x, y) = \begin{pmatrix} \frac{\partial h(x, y)}{\partial y_1} \\ \vdots \\ \frac{\partial h(x, y)}{\partial y_n} \end{pmatrix}.$$

Segundo Teorema da Função Implícita, [5]. *Seja B um subconjunto aberto de $\mathbb{R}^{m+n}$, $\bar{X}$ um subconjunto compacto de $\mathbb{R}^m$, e $h : B \rightarrow \mathbb{R}$ uma função tal que para algum índice $p \geq 1$, $h \in \mathcal{C}^p(B)$. Assumimos que existe $\nabla_y h(x, y)$ e é contínua sobre B . Assumimos que o vetor $\bar{y} \in \mathbb{R}^n$ é tal que $(\bar{x}, \bar{y}) \in B$, $h(\bar{x}, \bar{y}) = 0$, e que a matriz $\nabla_y h(\bar{x}, \bar{y})$ é não-singular para todo $\bar{x} \in \bar{X}$. Então, existem escalares $\varepsilon > 0, \delta > 0$ e uma função $\phi : S(\bar{X}; \varepsilon) \rightarrow S(\bar{y}; \delta)$ tal que $\phi \in \mathcal{C}^p(S(\bar{X}; \varepsilon))$, $\bar{y} = \phi(\bar{x})$ para todo $\bar{x} \in \bar{X}$ e $h(x, \phi(x)) = 0$ para todo $x \in S(\bar{X}; \varepsilon)$. A função ϕ é única no sentido de que se $x \in S(\bar{X}; \varepsilon)$, $y \in S(\bar{y}; \delta)$ e $h(x, y) = 0$, então $y = \phi(x)$.*

Teorema 2.4.5. *Assumimos que se verifica a Hipótese 2.4.2. Seja $\bar{\rho} > 0$ o escalar obtido no Lema 2.4.3. Então, existem escalares positivos δ, ε, M tais que para todo $(\lambda, \rho, \alpha) \in$*

$D \subset \mathbb{R}^{m_1+1+n+m_2}$ definido por

$$D = \left\{ (\lambda, \rho, \alpha) \mid \sqrt{\frac{\|\lambda - \lambda_*\|_2^2}{\rho^2} + \|\alpha\|_2^2} < \delta, \rho \geq \bar{\rho} \right\},$$

(a) existe um único vetor $x_\alpha(\lambda, \rho) \in B(x_*, \varepsilon)$ para o qual existe $v_\alpha(\lambda, \rho)$ tal que

$$\begin{aligned} \nabla L_1(x_\alpha(\lambda, \rho), \lambda, \rho) + \nabla c_2(x_\alpha(\lambda, \rho))v_\alpha(\lambda, \rho) &= \alpha_1, \\ c_2(x_\alpha(\lambda, \rho)) &= \alpha_2, \end{aligned}$$

sendo $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^{n+m_2}$.

(b) A função x_α é continuamente diferenciável no interior de $B(x_*, \varepsilon)$ e

$$\|x_\alpha(\lambda, \rho) - x_*\|_2 \leq M \sqrt{\frac{\|\lambda - \lambda_*\|_2^2}{\rho^2} + \|\alpha\|_2^2}. \quad (2.58)$$

Se definimos

$$\tilde{\lambda}_\alpha(\lambda, \rho) = \lambda + \rho c_1(x_\alpha(\lambda, \rho))$$

então

$$\|\tilde{\lambda}_\alpha(\lambda, \rho) - \lambda_*\|_2 \leq M \sqrt{\frac{\|\lambda - \lambda_*\|_2^2}{\rho^2} + \|\alpha\|_2^2}. \quad (2.59)$$

(c) Se definimos $\bar{L}_\rho(x, \lambda, v) = f(x) + \sum_{i=1}^{m_1} \lambda_i [c_1(x)]_i + \frac{\rho}{2} \|c_1(x)\|^2 + \sum_{i=1}^{m_2} v_i [c_2(x)]_i = L_1(x, \lambda, \rho) + \sum_{i=1}^{m_2} v_i [c_2(x)]_i$ então

$$z^T \nabla_{xx}^2 \bar{L}_\rho(x_\alpha(\lambda, \rho), \lambda, v_\alpha(\lambda, \rho)) z > 0, \quad (2.60)$$

para todo $z \neq 0$ tal que $\nabla c_2(x_\alpha(\lambda, \rho))^T z = 0$.

Prova. Para $\rho > 0$ consideremos o seguinte sistema nas variáveis $(x, \tilde{\lambda}, \lambda, v, \rho, \alpha)$:

$$T(x, \tilde{\lambda}, \lambda, v, \rho, \alpha) = \begin{bmatrix} \nabla f(x) + \nabla c_1(x) \tilde{\lambda} + \nabla c_2(x) v - \alpha_1 \\ c_1(x) + \frac{\lambda - \tilde{\lambda}}{\rho} \\ c_2(x) - \alpha_2 \end{bmatrix} = 0.$$

Definindo as novas variáveis:

$$t = \frac{\lambda - \lambda_*}{\rho}, \quad \gamma = \frac{1}{\rho},$$

obtemos o seguinte sistema nas variáveis $(x, \tilde{\lambda}, v, t, \gamma, \alpha)$:

$$T(x, \tilde{\lambda}, v, t, \gamma, \alpha) = \begin{bmatrix} \nabla f(x) + \nabla c_1(x)\tilde{\lambda} + \nabla c_2(x)v - \alpha_1 \\ c_1(x) + t + \gamma\lambda_* - \gamma\tilde{\lambda} \\ c_2(x) - \alpha_2 \end{bmatrix} = 0. \quad (2.61)$$

Para $t = 0, \gamma \in [0, \frac{1}{\rho}]$ e $\alpha = 0$, o sistema (2.61) tem solução $x = x_*, \tilde{\lambda} = \lambda_*, v = v_*$. O Jacobiano do sistema com relação a $(x, \tilde{\lambda}, v)$, avaliado nesse ponto é

$$\begin{pmatrix} \nabla_{xx}^2 L_0(x_*, \lambda_*, v_*) & \nabla c_1(x_*) & \nabla c_2(x_*) \\ \nabla c_1(x_*)^T & -\gamma I & 0 \\ \nabla c_2(x_*)^T & 0 & 0 \end{pmatrix}. \quad (2.62)$$

Pelo Lema 2.4.4 sabemos que este Jacobiano é invertível para todo $\gamma \in [0, \frac{1}{\rho}]$ e podemos aplicar o Segundo Teorema da Função Implícita: existem $\epsilon > 0, \delta > 0$ e únicas funções continuamente diferenciáveis

$$\hat{x}_\alpha(t, \gamma), \hat{\lambda}_\alpha(t, \gamma), \hat{v}_\alpha(t, \gamma)$$

definidas sobre

$$S(K, \delta) = \{(t, \gamma, \alpha) \mid \|(t, \gamma, \alpha) - z\|_2 < \delta, \forall z \in K\}$$

sendo $K = \{(0, \gamma, 0) \mid \gamma \in [0, \frac{1}{\rho}]\}$ tais que

$$(\|\hat{x}_\alpha(t, \gamma) - x_*\|_2^2 + \|\hat{\lambda}_\alpha(t, \gamma) - \lambda_*\|_2^2 + \|\hat{v}_\alpha(t, \gamma) - v_*\|_2^2)^{\frac{1}{2}} < \epsilon$$

para todo $(t, \gamma, \alpha) \in S(K, \delta)$ e

$$\begin{aligned} \nabla f(\hat{x}_\alpha(t, \gamma)) + \nabla c_1(\hat{x}_\alpha(t, \gamma))\hat{\lambda}_\alpha(t, \gamma) + \nabla c_2(\hat{x}_\alpha(t, \gamma))\hat{v}_\alpha(t, \gamma) &= \alpha_1, \\ c_1(\hat{x}_\alpha(t, \gamma)) + t + \gamma\lambda_* - \gamma\hat{\lambda}_\alpha(t, \gamma) &= 0, \\ c_2(\hat{x}_\alpha(t, \gamma)) &= \alpha_2. \end{aligned} \quad (2.63)$$

Além disso, δ e ϵ podem ser escolhidos tais que, $\forall z \neq 0$ tal que $\nabla c_2(\hat{x}_\alpha(t, \gamma))^T z = 0$ temos

$$z^T \nabla_{xx}^2 \bar{L}_\rho(\hat{x}_\alpha(t, \gamma), \hat{\lambda}_\alpha(t, \gamma), \hat{v}_\alpha(t, \gamma)) z > 0$$

sendo $\bar{L}_\rho(x, \lambda, v)$ a função definida no item (c).

Derivando (2.63) com relação a $t, \gamma, \alpha_1, \alpha_2$ temos que

$$\begin{pmatrix} \nabla_t \hat{x}_\alpha(t, \gamma)^T & \nabla_\gamma \hat{x}_\alpha(t, \gamma)^T & \nabla_{\alpha_1} \hat{x}_\alpha(t, \gamma)^T & \nabla_{\alpha_2} \hat{x}_\alpha(t, \gamma)^T \\ \nabla_t \hat{\lambda}_\alpha(t, \gamma)^T & \nabla_\gamma \hat{\lambda}_\alpha(t, \gamma)^T & \nabla_{\alpha_1} \hat{\lambda}_\alpha(t, \gamma)^T & \nabla_{\alpha_2} \hat{\lambda}_\alpha(t, \gamma)^T \\ \nabla_t \hat{v}_\alpha(t, \gamma)^T & \nabla_\gamma \hat{v}_\alpha(t, \gamma)^T & \nabla_{\alpha_1} \hat{v}_\alpha(t, \gamma)^T & \nabla_{\alpha_2} \hat{v}_\alpha(t, \gamma)^T \end{pmatrix} = A(t, \gamma, \alpha) \begin{pmatrix} 0 & 0 & I & 0 \\ -I & \hat{\lambda}_\alpha(t, \gamma) - \lambda_* & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix},$$

onde

$$A(t, \gamma, \alpha) = \begin{pmatrix} \nabla_{xx}^2 L_0(\hat{x}_\alpha(t, \gamma), \hat{\lambda}_\alpha(t, \gamma), \hat{v}_\alpha(t, \gamma)) & \nabla c_1(\hat{x}_\alpha(t, \gamma)) & \nabla c_2(\hat{x}_\alpha(t, \gamma)) \\ \nabla c_1(\hat{x}_\alpha(t, \gamma))^T & -\gamma I & 0 \\ \nabla c_2(\hat{x}_\alpha(t, \gamma))^T & 0 & 0 \end{pmatrix}^{-1}.$$

Logo, para todo (t, γ, α) tais que $\sqrt{\|t\|_2^2 + \|\alpha\|_2^2} < \delta$ e $\gamma \in [0, 1/\bar{\rho}]$ obtemos que

$$\begin{bmatrix} \hat{x}_\alpha(t, \gamma) - x_* \\ \hat{\lambda}_\alpha(t, \gamma) - \lambda_* \\ \hat{v}_\alpha(t, \gamma) - v_* \end{bmatrix} = \int_0^1 A(\xi t, \xi \gamma, \xi \alpha) \begin{bmatrix} 0 & 0 & I & 0 \\ -I & \hat{\lambda}_{\xi \alpha}(\xi t, \xi \gamma) - \lambda_* & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \begin{bmatrix} t \\ \gamma \\ \alpha_1 \\ \alpha_2 \end{bmatrix} d\xi.$$

Como a matriz (2.62) é não-singular $\forall \gamma \in [0, \frac{1}{\bar{\rho}}]$ temos que, para todo δ suficientemente pequeno a matriz $A(t, \gamma, \alpha)$ é uniformemente limitada sobre $\{(t, \gamma, \alpha) | \sqrt{\|t\|_2^2 + \|\alpha\|_2^2} < \delta, \gamma \in [0, 1/\bar{\rho}]\}$. Seja então μ tal que $\|A(t, \gamma, \alpha)\|_2 \leq \mu$, para todo (t, γ, α) tal que $\sqrt{\|t\|_2^2 + \|\alpha\|_2^2} < \delta, \gamma \in [0, 1/\bar{\rho}]$, e ser for necessário, consideramos δ suficientemente pequeno tal que $\mu\delta < 1$. Então

$$(\|\hat{x}_\alpha(t, \gamma) - x_*\|_2^2 + \|\hat{\lambda}_\alpha(t, \gamma) - \lambda_*\|_2^2)^{\frac{1}{2}} \leq \mu(\|(t, \alpha)\|_2 + \max_{0 \leq \xi \leq 1} \|\hat{\lambda}_{\xi \alpha}(\xi t, \xi \gamma) - \lambda_*\|_2 \gamma). \quad (2.64)$$

Logo, temos que, $\forall (t, \gamma, \alpha)$ tal que $(\|t\|_2^2 + \|\alpha\|_2^2)^{\frac{1}{2}} < \delta, \gamma \in [0, \frac{1}{\bar{\rho}}]$,

$$\|\hat{\lambda}_\alpha(t, \gamma) - \lambda_*\|_2 \leq \mu(\|t\|_2^2 + \|\alpha\|_2^2)^{\frac{1}{2}} + \mu\gamma \max_{0 \leq \xi \leq 1} \|\hat{\lambda}_{\xi \alpha}(\xi t, \xi \gamma) - \lambda_*\|_2.$$

Utilizando essa desigualdade com $\xi t, \xi \gamma, \xi \alpha, \xi \in [0, 1]$ em lugar de t, γ, α obtemos que

$$\max_{0 \leq \xi \leq 1} \|\hat{\lambda}_{\xi \alpha}(\xi t, \xi \gamma) - \lambda_*\|_2 \leq \frac{\mu}{1 - \mu\gamma} (\|t\|_2^2 + \|\alpha\|_2^2)^{\frac{1}{2}}. \quad (2.65)$$

Combinando as desigualdades (2.64) e (2.65) temos que, se $\gamma < \delta$:

$$(\|\hat{x}_\alpha(t, \gamma) - x_*\|_2^2 + \|\hat{\lambda}_\alpha(t, \gamma) - \lambda_*\|_2^2)^{\frac{1}{2}} \leq \left(\mu + \frac{\mu^2 \gamma}{1 - \mu \gamma} \right) (\|t\|_2^2 + \|\alpha\|_2^2)^{\frac{1}{2}} \leq \frac{\mu}{1 - \mu \delta} (\|t\|_2^2 + \|\alpha\|_2^2)^{\frac{1}{2}}.$$

Seja δ suficientemente pequeno tal que

$$(\|\hat{x}_\alpha(t, \gamma) - x_*\|_2^2 + \|\hat{\lambda}_\alpha(t, \gamma) - \lambda_*\|_2^2)^{\frac{1}{2}} \leq 2\mu(\|t\|_2^2 + \|\alpha\|_2^2)^{\frac{1}{2}}. \quad (2.66)$$

Utilizando uma mudança de variáveis e escrevendo

$$x_\alpha(\lambda, \rho) = \hat{x}_\alpha(t, \gamma), \tilde{\lambda}_\alpha(\lambda, \rho) = \hat{\lambda}_\alpha(t, \gamma), v_\alpha(\lambda, \rho) = \hat{v}_\alpha(t, \gamma)$$

temos que $\forall(\lambda, \rho, \alpha) \in D$ tal que

$$\sqrt{\frac{\|\lambda - \lambda_*\|_2^2}{\rho^2} + \|\alpha\|_2^2} < \delta, \quad \rho \geq \bar{\rho}$$

se verificam as igualdades

$$\nabla f(x_\alpha(\lambda, \rho)) + \nabla c_1(x_\alpha(\lambda, \rho))\tilde{\lambda}_\alpha(\lambda, \rho) + \nabla c_2(x_\alpha(\lambda, \rho))v_\alpha(\lambda, \rho) =$$

$$\nabla L_1(x_\alpha(\lambda, \rho), \lambda, \rho) + \nabla c_2(x_\alpha(\lambda, \rho))v_\alpha(\lambda, \rho) = \alpha_1,$$

$$c_2(x_\alpha(\lambda, \rho)) = \alpha_2.$$

Além disso, por (2.66)

$$\|x_\alpha(\lambda, \rho) - x_*\|_2 \leq M \sqrt{\frac{\|\lambda - \lambda_*\|_2^2}{\rho^2} + \|\alpha\|_2^2}$$

e, para

$$\tilde{\lambda}_\alpha(\lambda, \rho) = \lambda + \rho c_1(x_\alpha(\lambda, \rho))$$

temos que, por (2.63),

$$\|\tilde{\lambda}_\alpha(\lambda, \rho) - \lambda_*\|_2 \leq M \sqrt{\frac{\|\lambda - \lambda_*\|_2^2}{\rho^2} + \|\alpha\|_2^2}$$

sendo $M = 2\mu$. □

Hipótese 2.4.6. *Assumimos que $\{x_k\}$ é a seqüência gerada quando o Algoritmo 2.1.1*

utilizando a Regra 1 de adaptação é aplicado ao problema (2.49) considerando os conjuntos de restrições

$$\Omega_1 = \{x \in \mathbb{R}^n | c_1(x) = 0\}, \quad e \quad \Omega_2 = \{x \in \mathbb{R}^n | c_2(x) = 0\},$$

e que

$$\lim_{k \rightarrow \infty} x_k = x_*.$$

Assumimos que em todas as iterações do algoritmo $[\rho_0]_i = [\rho_0]_1$ para todo $i = 1, \dots, m_1$ e definimos $\rho_k = [\rho_k]_i$ para todo $i = 1, \dots, m_1$.

Hipótese 2.4.7. Assumimos que para todo $k \in \mathbb{N}$, $i = 1, \dots, m_1$, $[\bar{\lambda}_{k+1}]_i$ é a projeção de $[\lambda_{k+1}]_i$ no intervalo $[[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]$.

Corolário 2.4.8. Suponhamos que se verificam as Hipóteses 2.4.2, 2.4.6 e 2.4.7, e seja $\bar{\rho}$ como no Lema 2.4.3. Suponhamos que existe $k_0 \in \mathbb{N}$ tal que $\rho_k \geq \bar{\rho}$ para todo $k \geq k_0$. Definimos arbitrariamente α_k de modo que $\|\alpha_k\|_2 \leq \varepsilon_k$. Então, existe $M > 0$ tal que, para todo $k \in \mathbb{N}$,

$$\|x_k - x_*\|_2 \leq M \sqrt{\frac{\|\bar{\lambda}_k - \lambda_*\|_2^2}{\rho_k^2} + \|\alpha_k\|_2^2} \quad (2.67)$$

e, se $\lambda_{k+1} = \bar{\lambda}_k + \rho_k c_1(x_k)$ então

$$\|\lambda_{k+1} - \lambda_*\|_2 \leq M \sqrt{\frac{\|\bar{\lambda}_k - \lambda_*\|_2^2}{\rho_k^2} + \|\alpha_k\|_2^2}. \quad (2.68)$$

Prova. Como a sequência $\{\bar{\lambda}_k\}$ é limitada e $\|\alpha_k\|_2 \rightarrow 0$ temos que existe $k_1 \geq k_0$ tal que

$$\sqrt{\frac{\|\bar{\lambda}_k - \lambda_*\|_2^2}{\rho_k^2} + \|\alpha_k\|_2^2} < \delta \quad \forall k \geq k_1.$$

Assim, pela parte (a) do Teorema 2.4.5 existe um único x_k na vizinhança de x_* para o qual existe v_k tal que, se $\alpha_k = (\alpha_k^1, \alpha_k^2)$ então

$$\nabla f(x_k) + \nabla c_1(x_k) \bar{\lambda}_k + \rho_k \sum_{i=1}^{m_1} [c_1(x_k)]_i \nabla [c_1(x_k)]_i + \nabla c_2(x_k) v_k = \alpha_k^1$$

$$c_2(x_k) = \alpha_k^2,$$

$$\|x_k - x_*\|_2 \leq M \sqrt{\frac{\|\bar{\lambda}_k - \lambda_*\|_2^2}{\rho_k^2} + \|\alpha_k\|_2^2}$$

e se

$$\tilde{\lambda}_k = \bar{\lambda}_k + \rho_k c_1(x_k) = \lambda_{k+1}.$$

então

$$\|\tilde{\lambda}_k - \lambda_*\|_2 \leq M \sqrt{\frac{\|\bar{\lambda}_k - \lambda_*\|_2^2}{\rho_k^2} + \|\alpha_k\|_2^2}, \quad (2.69)$$

como queríamos provar. $\square$

Lema 2.4.9. *Suponhamos que se verificam as Hipóteses 2.4.2, 2.4.6 e 2.4.7. Então,*

$$\lim_{k \rightarrow \infty} \lambda_k = \lambda_*. \quad (2.70)$$

Prova. Como $\|\alpha_k\|_2 \leq \varepsilon_k \rightarrow 0$ e $\{\bar{\lambda}_k\}$ é limitada, temos que o lado direito da desigualdade (2.68) do Corolário 2.4.8 tende para zero. Isto implica que (2.70) se cumpre. $\square$

Hipótese 2.4.10. *Para todo $i = 1, \dots, m_1$,*

$$[\lambda_*]_i \in ([\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i).$$

Lema 2.4.11. *Suponhamos que se verificam as Hipóteses 2.4.2, 2.4.6, 2.4.7 e 2.4.10. Então*

$$\bar{\lambda}_k = \lambda_k$$

para k suficientemente grande.

Prova. Seja $\epsilon = \frac{1}{2} \min_{i=1, \dots, m_1} \{[\lambda_*]_i - [\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i - [\lambda_*]_i\}$. Pela Hipótese 2.4.10 temos que $\epsilon > 0$.

Pelo Lema 2.4.9, temos que $\lambda_{k+1} \rightarrow \lambda_*$ e portanto, existe k_1 tal que, $\forall k \geq k_1$

$$\|\lambda_{k+1} - \lambda_*\|_\infty \leq \epsilon.$$

Logo, como $|[\lambda_{k+1}]_i - [\lambda_*]_i| \leq \epsilon, \forall i = 1, \dots, m_1$, pela definição de ϵ temos que

$$[\lambda_{k+1}]_i \in [[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]$$

o que, pela Hipótese 2.4.7 implica que $\bar{\lambda}_{k+1} = \lambda_{k+1}$ para todo k suficientemente grande. $\square$

Teorema 2.4.12. *Suponhamos que as Hipóteses 2.4.2, 2.4.6, 2.4.7 e 2.4.10 são satisfeitas pela seqüência gerada quando o Algoritmo 2.1.1 é aplicado ao problema (2.49). Assumimos também que existe uma seqüência $\eta_k \rightarrow 0$ tal que*

$$\varepsilon_k \leq \eta_k \|c_1(x_k)\|_\infty \quad (2.71)$$

para todo $k \in \mathbb{N}$.

Então, a sequência de parâmetros de penalidade $\{\rho_k\}$ é limitada.

Prova. Assumimos, por contradição, que

$$\lim_{k \rightarrow \infty} \rho_k = \infty. \quad (2.72)$$

Seja $\bar{k}$ suficientemente grande tal que $\rho_{k-1} \geq \max\{2M, \bar{\rho}\}$ para todo $k \geq \bar{k}$.

Como $c_1(x_*) = 0$, pela continuidade das derivadas primeiras de c_1 existe $L > 0$ tal que, para todo $k \in \mathbb{N}$,

$$\|c_1(x_k)\|_2 \leq L\|x_k - x_*\|_2.$$

Conseqüentemente, por (2.67), temos que

$$\|c_1(x_k)\|_2 \leq LM \sqrt{\frac{\|\bar{\lambda}_k - \lambda_*\|_2^2}{\rho_k^2} + \|\alpha_k\|_2^2} \leq LM \left(\frac{\|\bar{\lambda}_k - \lambda_*\|_2}{\rho_k} + \|\alpha_k\|_2 \right)$$

para k suficientemente grande. Logo, como $\|\alpha_k\|_2 \leq \varepsilon_k$, por (2.71) e pelo Lema 2.4.11, temos que

$$\|c_1(x_k)\|_2 \leq \frac{LM}{1 - \eta_k LM} \frac{\|\lambda_k - \lambda_*\|_2}{\rho_k} \quad (2.73)$$

para k suficientemente grande de modo que $1 - \eta_k LM > 0$.

Por (2.19) e pelo Lema 2.4.11, temos que $\lambda_k = \lambda_{k-1} + \rho_{k-1}c_1(x_{k-1})$ para k suficientemente grande. Logo,

$$\|c_1(x_{k-1})\|_2 = \frac{\|\lambda_k - \lambda_{k-1}\|_2}{\rho_{k-1}} \geq \frac{\|\lambda_{k-1} - \lambda_*\|_2}{\rho_{k-1}} - \frac{\|\lambda_k - \lambda_*\|_2}{\rho_{k-1}} \quad (2.74)$$

para k suficientemente grande.

Assim, por (2.68), Lema 2.4.11 e (2.71),

$$\|\lambda_k - \lambda_*\|_2 \leq M \left(\frac{\|\lambda_{k-1} - \lambda_*\|_2}{\rho_{k-1}} + \eta_{k-1} \|c_1(x_{k-1})\|_2 \right)$$

para k suficientemente grande. Então,

$$\frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda_k - \lambda_*\|_\infty}{M} - \eta_{k-1} \|c_1(x_{k-1})\|_\infty$$

para k suficientemente grande. Conseqüentemente, por (2.74),

$$(1 + \eta_{k-1}) \|c_1(x_{k-1})\|_2 \geq \|\lambda_k - \lambda_*\|_2 \left(\frac{1}{M} - \frac{1}{\rho_{k-1}} \right) \geq \frac{1}{2M} \|\lambda_k - \lambda_*\|_2$$

para k suficientemente grande. Assim, isolando $\|\lambda_k - \lambda_*\|_2$ obtemos que

$$\|\lambda_k - \lambda_*\|_2 \leq 2M(1 + \eta_{k-1})\|c_1(x_{k-1})\|_2$$

para k suficientemente grande. Por (2.73), isto implica que

$$\|c_1(x_k)\|_2 \leq \frac{2LM^2(1 + \eta_{k-1})}{1 - LM\eta_k} \cdot \frac{1}{\rho_k} \|c_1(x_{k-1})\|_2.$$

Por equivalência entre normas temos que existe uma constante $\beta > 0$ tal que

$$\|c_1(x_k)\|_\infty \leq \beta \frac{2LM^2(1 + \eta_{k-1})}{1 - LM\eta_k} \cdot \frac{1}{\rho_k} \|c_1(x_{k-1})\|_\infty.$$

Logo, como $\rho_k \rightarrow \infty$ e $\eta_k \rightarrow 0$, temos que existe $k_1 \in \mathbb{N}$, $k_1 \geq \bar{k}$ tal que

$$\beta \frac{2LM^2(1 + \eta_{k-1})}{1 - LM\eta_k} \frac{1}{\rho_k} < \tau$$

para todo $k \geq k_1$ o que mostra que

$$\|c_1(x_k)\|_\infty \leq \tau \|c_1(x_{k-1})\|_\infty, \quad \forall k \geq k_1$$

Isto implica que $\rho_{k+1} = \rho_k$ para todo $k \geq k_1$. Assim, (2.72) é falso. $\square$

2.4.3 Limitação no caso geral

Nesta subseção apresentamos o resultado final de limitação do parâmetro de penalidade associado ao Algoritmo 2.1.1 - Regra 1. Como no Teorema 2.4.12, a hipótese mais importante é que a precisão utilizada para resolver os subproblemas deve tender a zero mais rápido que a medida de inviabilidade das restrições do conjunto complexo. Este tipo de requerimento é usual em métodos de Lagrangiano aumentado [4, 5, 14, 16, 18, 19, 21].

Hipótese 2.4.13. *Assumimos que*

1. *As funções f, h_1, g_1, h_2 e g_2 admitem derivadas segundas contínuas numa vizinhança de x_* .*
2. *A sequência $\{x_k\}$ é a gerada pela aplicação do Algoritmo 2.1.1 ao problema original (2.1) e*

$$\lim_{k \rightarrow \infty} x_k = x_*.$$

3. O ponto x_* é viável ($h_1(x_*) = 0, g_1(x_*) \leq 0, h_2(x_*) = 0, g_2(x_*) \leq 0$.)

4. Os gradientes

$$\{\nabla[h_1(x_*)]_i\}_{i=1}^{m_1}, \{\nabla[g_1(x_*)]_i\}_{[g_1(x_*)]_i=0}, \{\nabla[h_2(x_*)]_i\}_{i=1}^{m_2}, \{\nabla[g_2(x_*)]_i\}_{[g_2(x_*)]_i=0}$$

são linearmente independentes.

5. Verifica-se a condição de complementaridade estrita em x_* : se $\mu_* \in \mathbb{R}_+^{p_1}$ e $u_* \in \mathbb{R}_+^{p_2}$ são os multiplicadores de Lagrange associados às restrições $g_1(x) \leq 0$ e $g_2(x) \leq 0$, respectivamente, então:

$$[g_1(x_*)]_i = 0 \Rightarrow [\mu_*]_i > 0$$

e

$$[g_2(x_*)]_i = 0 \Rightarrow [u_*]_i > 0.$$

6. Verifica-se a condição suficiente de segunda ordem em x_* : se definimos o subespaço tangente T como o conjunto de pontos $z \in \mathbb{R}^n$ tais que

$$\nabla h_1(x_*)^T z = \nabla h_2(x_*)^T z = 0,$$

$$\nabla[g_1(x_*)]_i^T z = 0$$

para todo i tal que $[g_1(x_*)]_i = 0$ e

$$\nabla[g_2(x_*)]_i^T z = 0$$

para todo i tal que $[g_2(x_*)]_i = 0$.

Então, para todo $0 \neq z \in T$,

$$\begin{aligned} z^T & \left[\nabla^2 f(x_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla^2 [h_1(x_*)]_i + \sum_{i=1}^{p_1} [\mu_*]_i \nabla^2 [g_1(x_*)]_i \right. \\ & \left. + \sum_{i=1}^{m_2} [v_*]_i \nabla^2 [h_2(x_*)]_i + \sum_{i=1}^{p_2} [u_*]_i \nabla^2 [g_2(x_*)]_i \right] z > 0. \end{aligned}$$

Proposição 2.4.14. Assumimos que se verifica a Hipótese 2.4.13. Definimos

$$w_* = (\sqrt{-[g_1(x_*)]_1}, \dots, \sqrt{-[g_1(x_*)]_{p_1}}),$$

$$z_* = (\sqrt{-[g_2(x_*)]_1}, \dots, \sqrt{-[g_2(x_*)]_{p_2}})$$

e

$$y_* = (x_*, w_*, z_*).$$

Consideremos o problema (2.42) com

$$c_1(y) = (H_1(y), G_1(y)), c_2(y) = (H_2(y), G_2(y)).$$

Então, o problema (2.42) satisfaz a Hipótese 2.4.2.

Prova. Provaremos primeiro que os gradientes são linearmente independentes. Para isso, lembremos que

$$\begin{aligned} [H_1(y)]_i &= [h_1(x)]_i, i = 1, \dots, m_1 \\ [G_1(y)]_i &= [g_1(x)]_i + w_i^2, i = 1, \dots, p_1 \\ [H_2(y)]_i &= [h_2(x)]_i, i = 1, \dots, m_2 \\ [G_2(y)]_i &= [g_2(x)]_i + z_i^2, i = 1, \dots, p_2. \end{aligned}$$

Logo, temos que

$$\nabla[H_1(y_*)]_i = \begin{pmatrix} \nabla[h_1(x_*)]_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}, i = 1, \dots, m_1$$

$$\nabla[G_1(y_*)]_i = \begin{pmatrix} \nabla[g_1(x_*)]_i \\ 0 \\ \vdots \\ 2[w_*]_i \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}, i = 1, \dots, p_1$$

$$\nabla[H_2(y_*)]_i = \begin{pmatrix} \nabla[h_2(x_*)]_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}, i = 1, \dots, m_2$$

$$\nabla[G_2(y_*)]_i = \begin{pmatrix} \nabla[g_2(x_*)]_i \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 2[z_*]_i \\ \vdots \\ 0 \end{pmatrix}, i = 1, \dots, p_2.$$

Sejam $\lambda_1 \in \mathbb{R}^{m_1}, \mu_1 \in \mathbb{R}^{p_1}, \lambda_2 \in \mathbb{R}^{m_2}, \mu_2 \in \mathbb{R}^{p_2}$ tais que

$$\sum_{i=1}^{m_1} [\lambda_1]_i \nabla[H_1(y_*)]_i + \sum_{i=1}^{p_1} [\mu_1]_i \nabla[G_1(y_*)]_i + \sum_{i=1}^{m_2} [\lambda_2]_i \nabla[H_2(y_*)]_i + \sum_{i=1}^{p_2} [\mu_2]_i \nabla[G_2(y_*)]_i = 0.$$

Então

$$\sum_{i=1}^{m_1} [\lambda_1]_i \nabla[h_1(x_*)]_i + \sum_{i=1}^{p_1} [\mu_1]_i \nabla[g_1(x_*)]_i + \sum_{i=1}^{m_2} [\lambda_2]_i \nabla[h_2(x_*)]_i + \sum_{i=1}^{p_2} [\mu_2]_i \nabla[g_2(x_*)]_i = 0,$$

$$[\mu_1]_i [w_*]_i = 0, i = 1, \dots, p_1,$$

$$[\mu_2]_i [z_*]_i = 0, i = 1, \dots, p_2.$$

Pela definição de w_* e z_* temos que se $[\mu_1]_i = 0$ para todo i tal que $[w_*]_i = [g_1(x_*)]_i \neq 0, i = 1, \dots, p_1$ e $[\mu_2]_i = 0$ para todo i tal que $[z_*]_i = [g_2(x_*)]_i \neq 0, i = 1, \dots, p_2$, então

$$\sum_{i=1}^{m_1} [\lambda_1]_i \nabla[h_1(x_*)]_i + \sum_{[g_1(x_*)]_i=0} [\mu_1]_i \nabla[g_1(x_*)]_i + \sum_{i=1}^{m_2} [\lambda_2]_i \nabla[h_2(x_*)]_i + \sum_{[g_2(x_*)]_i=0} [\mu_2]_i \nabla[g_2(x_*)]_i = 0$$

e, pelo item 4 da Hipótese 2.4.13 temos que $[\lambda_1]_i = 0, \forall i = 1, \dots, m_1, [\mu_1]_i = 0, \forall i$ tal que $[g_1(x_*)]_i = 0, [\lambda_2]_i = 0, \forall i = 1, \dots, m_2$, e $[\mu_2]_i = 0, \forall i$ tal que $[g_2(x_*)]_i = 0$. Logo, os gradientes $\nabla[c_1(y_*)]_1, \dots, \nabla[c_1(y_*)]_{m_1+p_1}, \nabla[c_2(y_*)]_1, \dots, \nabla[c_2(y_*)]_{m_2+p_2}$ são linearmente independentes. Assim, se satisfaz o item 3 da Hipótese 2.4.2.

Mostraremos a seguir que y_* é um ponto estacionário do problema (2.42). Pelo Teorema de convergência global do algoritmo sabemos que x_* é um ponto estacionário de primeira ordem do problema (2.1): existem multiplicadores $\lambda_*, \mu_*, v_*, u_*$ satisfazendo o item 5 da hipótese tais que

$$\nabla f(x_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla [h_1(x_*)]_i + \sum_{i=1}^{p_1} [\mu_*]_i \nabla [g_1(x_*)]_i + \sum_{i=1}^{m_2} [v_*]_i \nabla [h_2(x_*)]_i + \sum_{i=1}^{p_2} [u_*]_i \nabla [g_2(x_*)]_i = 0. \quad (2.75)$$

Logo, por definição de w_* e z_* e a condição de complementaridade temos que $[\mu_*]_i [w_*]_i = 0$ para todo $i = 1, \dots, p_1$, $[u_*]_i [z_*]_i = 0$ para todo $i = 1, \dots, p_2$ e assim, por (2.75):

$$\begin{aligned} \nabla F(y_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla [H_1(y_*)]_i + \sum_{i=1}^{p_1} [\mu_*]_i \nabla [G_1(y_*)]_i + \\ \sum_{i=1}^{m_2} [v_*]_i \nabla [H_2(y_*)]_i + \sum_{i=1}^{p_2} [u_*]_i \nabla [G_2(y_*)]_i = 0, \end{aligned}$$

onde $F(y) = f(x)$. Assim, provamos que y_* é um ponto estacionário do problema (2.42).

Provaremos agora que se satisfaz o item 4 da Hipótese 2.4.2. Seja $\theta = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^{n+p_1+p_2}$, $\theta \neq 0$ tal que $\nabla c_1(y_*)^T \theta = \nabla c_2(y_*)^T \theta = 0$. Temos que provar que

$$\theta^T \nabla^2 L_0(y_*, \lambda_*, \mu_*, v_*, u_*) \theta > 0. \quad (2.76)$$

Definimos $D(\alpha)$ como a matriz diagonal com elementos α_i na diagonal.

Pela definição de L_0 temos que, se definimos

$$\begin{aligned} H(x_*, \lambda_*, \mu_*, v_*, u_*) = \nabla^2 f(x_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla^2 [h_1(x_*)]_i + \sum_{i=1}^{p_1} [\mu_*]_i \nabla^2 [g_1(x_*)]_i \\ \sum_{i=1}^{m_2} [v_*]_i \nabla^2 [h_2(x_*)]_i + \sum_{i=1}^{p_2} [u_*]_i \nabla^2 [g_2(x_*)]_i \end{aligned}$$

então

$$\nabla^2 L_0(y_*, \lambda_*, \mu_*, v_*, u_*) = \begin{pmatrix} H(x_*, \lambda_*, \mu_*, v_*, u_*) & 0 & 0 \\ 0 & 2D(\mu_*) & 0 \\ 0 & 0 & 2D(u_*) \end{pmatrix}.$$

Assim,

$$\theta^T \nabla^2 L_0(y_*, \lambda_*, \mu_*, v_*, u_*) \theta = \theta_1^T H(x_*, \lambda_*, \mu_*, v_*, u_*) \theta_1 + 2 \sum_{i=1}^{p_1} [\mu_*]_i [\theta_2]_i^2 + 2 \sum_{i=1}^{p_2} [u_*]_i [\theta_3]_i^2. \quad (2.77)$$

Logo, se $[g_1(x_*)]_i = 0$ para algum $i \in \{1, \dots, p_1\}$ então $[w_*]_i = 0$ e, pela definição de

$[G_1(y_*)]_i$ temos que

$$\nabla[G_1(y_*)]_i^T \theta = \nabla[g_1(x_*)]_i^T \theta_1 = 0.$$

Da mesma maneira, se $[g_2(x_*)]_i = 0$ para algum $i \in \{1, \dots, p_2\}$ então $[z_*]_i = 0$ e, pela definição de $[G_2(y_*)]_i$, temos que

$$\nabla[G_2(y_*)]_i^T \theta = \nabla[g_2(x_*)]_i^T \theta_1 = 0.$$

Logo, se $\theta_1 \neq 0$, pelo item 6 da Hipótese 2.4.13 obtemos que

$$\theta_1^T H(x_*, \lambda_*, \mu_*, v_*, u_*) \theta > 0,$$

e pela não negatividade dos multiplicadores μ_*, u_* e (2.77) temos que se verifica (2.76).

Consideremos agora o caso em que $\theta_1 = 0$. Como $\theta \neq 0$ suponhamos que existe i tal que $[\theta_2]_i \neq 0$. Como $\nabla c_1(y_*)^T \theta = 0$ deve ocorrer $2[w_*]_i[\theta_2]_i = 0$ o que implica que $[w_*]_i = [g_1(x_*)]_i = 0$. Logo, pela hipótese de complementaridade estrita na solução temos que $[\mu_*]_i > 0$ e assim

$$\theta^T \nabla^2 L_0(y_*, \lambda_*, \mu_*, v_*, u_*) \theta = 2 \sum_{i=1}^{p_1} [\mu_*]_i [\theta_2]_i^2 + 2 \sum_{i=1}^{p_2} [u_*]_i [\theta_3]_i^2 > 0.$$

como queríamos provar. A prova é análoga no caso em que $[\theta_3]_i \neq 0$ para algum índice i .

Assim verifica-se a condição do item 4 que, junto com a condição de ser y_* um ponto estacionário, mostra que ele é um minimizador local de (2.42). $\square$

Teorema 2.4.15. *Assumimos que se verifica a Hipótese 2.4.13 e que:*

1. *A Regra 1 é utilizada em todas as iterações do algoritmo e que $[\rho_0]_i = [\rho_0]_1$ para todo $i = 1, \dots, m_1 + p_1$.*

2. *Existe uma sequência $\eta_k \rightarrow 0$ tal que*

$$\varepsilon_k \leq \eta_k \max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\} \forall k \in \mathbb{N}.$$

3. *$[\lambda_*]_i \in ([\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i) \forall i = 1, \dots, m_1$ e $[\mu_*]_i \in ([\bar{\mu}_{\min}]_i, [\bar{\mu}_{\max}]_i) \forall i = 1, \dots, p_1$.*

4. *$[\bar{\lambda}_{k+1}]_i$ é a projeção de $[\lambda_{k+1}]_i$ no conjunto $[[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]$ e $[\bar{\mu}_{k+1}]_j$ é a projeção de $[\mu_{k+1}]_j$ no conjunto $[[\bar{\mu}_{\min}]_j, [\bar{\mu}_{\max}]_j]$ para todo $i = 1, \dots, m_1, j = 1, \dots, p_1, k \in \mathbb{N}$.*

Então, a sequência de parâmetros de penalidade $\{\rho_k\}$ é limitada.

Prova. Consideremos o problema equivalente (2.42). Pela hipótese do teorema e pela Proposição 2.4.14 sabemos que se verificam as hipóteses do Teorema 2.4.12.

Pelo Teorema 2.4.1 temos que a sequência $\{y_k\}$ pode ser gerada quando o Algoritmo 2.1.1 é aplicado ao problema (2.42). Assim, por definição de $[w_k]_i$ (2.44) temos que

$$\|g_1(x_k) + w_k^2\|_\infty = \|\sigma_k\|_\infty$$

e portanto

$$\|c_1(y_k)\|_\infty = \|(H_1(y_k), G_1(y_k))\|_\infty = \max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\}.$$

Como $\varepsilon_k \rightarrow 0$ temos que existe $\hat{\eta}_k \rightarrow 0$ tal que

$$\hat{\varepsilon}_k \leq \hat{\eta}_k \|c_1(y_k)\|_\infty$$

para todo k suficientemente grande e para $\hat{\varepsilon}_k$ definido no Teorema 2.4.1. Assim, podemos aplicar o Teorema 2.4.12 e a limitação da sequência de parâmetros de penalidade se segue desse teorema. $\square$

Observar que para obter a limitação do parâmetro de penalidade necessitamos que a precisão utilizada para resolver cada subproblema dependa do nível de inviabilidade na solução aproximada.

Assim, sob hipóteses habituais podemos garantir que os parâmetros de penalidade são limitados quando é utilizada a Regra 1 no Algoritmo 2.1.1. A pergunta sobre ser possível obter um resultado semelhante no caso em que se utiliza a Regra 2 é ainda aberta. A esse respeito, podemos provar limitação dos parâmetros para um algoritmo mais específico que mostraremos a seguir.

2.5 Lagrangiano aumentado para problemas com restrições de canalização

Nesta seção consideramos o seguinte problema geral de programação não-linear:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } h(x) = 0, \ x \in \Omega, \end{aligned} \tag{2.78}$$

para as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h(x) = (h_1(x), \dots, h_m(x))$ e $\Omega = \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$ e assumimos que as funções f e h admitem derivadas

primeiras contínuas sobre um conjunto aberto que contém Ω .

Este tipo de problema é capaz de representar qualquer problema de otimização com restrições, pois restrições de desigualdade do tipo $g_i(x) \leq 0$ podem ser substituídas por $g_i(x) + z = 0, z \geq 0$. Porém temos que advertir que isto pode não ser muito apropriado no caso em que há várias restrições de desigualdade já que a dimensão do problema aumenta consideravelmente.

Problemas com restrições de canalização são interessantes de ser tratados separadamente porque, do ponto de vista teórico, possuem condições KKT mais simples, o que facilita as provas de convergência e, do ponto de vista prático, existem algoritmos eficientes que aproveitam a estrutura das restrições.

A contribuição teórica mais importante associada ao algoritmo que propomos para resolver o problema (2.78) é que os resultados de limitação dos parâmetros de penalidade podem ser obtidos sem a hipótese de complementaridade estrita na solução tanto para vários parâmetros quanto para um único parâmetro.

2.5.1 O método de Lagrangiano aumentado para restrições de canalização e a convergência global

Seguindo a notação utilizada neste capítulo, definimos:

$$\Omega_1 = \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \dots, m\},$$

$$\Omega_2 = [l, u],$$

e a função Lagrangiano aumentado

$$L(x, \lambda, \rho) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \frac{1}{2} \sum_{i=1}^m \rho_i (h_i(x))^2$$

para todo $x \in \Omega_2, \lambda \in \mathbb{R}^m$ e $\rho \in \mathbb{R}_{++}^m$.

Assim, o algoritmo de Lagrangiano aumentado proposto para resolver o problema (2.78) é definido por:

Algoritmo 2.5.1. *Sejam $x_0 \in \Omega_2, \tau \in [0, 1), \gamma > 1, -\infty < \bar{\lambda}_{\min} < \bar{\lambda}_{\max} < \infty, \rho_1 \in \mathbb{R}_{++}^m, \bar{\lambda}_1 \in [\bar{\lambda}_{\min}, \bar{\lambda}_{\max}]^m$ dados. Seja $\{\varepsilon_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+$ uma seqüência de números positivos que converge para zero.*

Passo 1. *Inicialização*

Seja $k \leftarrow 1$.

Passo 2. Resolver o subproblema

Usando x_{k-1} como aproximação inicial,

Minimizar (aproximadamente) $L(x, \bar{\lambda}_k, \rho_k)$ sujeito a $x \in \Omega_2$.

O minimizador aproximado $x_k \in \Omega_2$ deve ser tal que

$$\|\mathcal{P}_{\Omega_2}[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_{\infty} \leq \varepsilon_k. \quad (2.79)$$

Passo 3. Estimar os multiplicadores

Para cada $i = 1, \dots, m$ definir

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k) \quad (2.80)$$

Se $h(x_k) = 0$ e $\mathcal{P}_{\Omega_2}[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k = 0$ terminar o processo (nesse caso, x_k é um ponto KKT e λ_{k+1} é o vetor de multiplicadores de Lagrange).

Calcular

$$\bar{\lambda}_{k+1} \in [\bar{\lambda}_{\min}, \bar{\lambda}_{\max}]^m. \quad (2.81)$$

Passo 4. Atualizar os parâmetros de penalidade

Utilizar uma das seguintes regras:

Regra 1. Se

$$\|h(x_k)\|_{\infty} \leq \tau \|h(x_{k-1})\|_{\infty},$$

definir

$$[\rho_{k+1}]_i = [\rho_k]_i, \quad i = 1, \dots, m.$$

Senão, definir

$$[\rho_{k+1}]_i = \gamma [\rho_k]_i, \quad i = 1, \dots, m.$$

Regra 2. Para cada $i = 1, \dots, m$, se

$$|h_i(x_k)| \leq \tau \|h(x_k)\|_{\infty},$$

definir

$$[\rho_{k+1}]_i = [\rho_k]_i.$$

Senão, definir

$$[\rho_{k+1}]_i = \gamma[\rho_k]_i.$$

Passo 5. *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$ e ir ao Passo 2.

Observações.

1. Suponhamos que $[\rho_0]_i = \rho_0$ para todo $i = 1, \dots, m$ no caso em que é utilizada a Regra 1. Assim, se para algum i

$$|h_i(x_k)| > \tau \|h(x_{k-1})\|_\infty$$

então

$$\|h(x_k)\|_\infty > \tau \|h(x_{k-1})\|_\infty.$$

Logo, quando o algoritmo adapta pelo menos um parâmetro de penalidade utilizando a Regra 2, ele adapta o único parâmetro utilizando a Regra 1. Nesta situação, outros parâmetros podem se manter sem cambiar quando se utiliza a Regra 2. Conseqüentemente, os parâmetros de penalidade para um algoritmo que utiliza a Regra 2 tendem a ser menores que o único parâmetro de um algoritmo que utiliza a Regra 1.

2. Como no caso geral, os resultados de convergência global não dependem da eleição do multiplicador de Lagrange $\bar{\lambda}_{k+1}$ em (2.81). Aqui calculamos a estimativa de primeira ordem do multiplicador de Lagrange e a projetamos num intervalo fechado prefixado desde o início do processo. Logo, utilizamos a estimativa natural do multiplicador de Lagrange no caso em que ele esteja dentro desse intervalo.
3. O algoritmo está baseado na resolução do subproblema (2.79). Como Ω_2 é um conjunto compacto, existem minimizadores e pontos estacionários e eles podem ser obtidos com qualquer precisão arbitrária utilizando algoritmos razoáveis. Logo, o algoritmo está bem definido.

Com relação à análise de viabilidade, provamos que utilizando a Regra 1 o algoritmo sempre converge para pontos estacionários do problema de minimizar $\|h(x)\|_2^2$ sujeito a $l \leq x \leq u$ que é o resultado análogo ao obtido no Teorema 2.2.1. No caso em que a Regra 2 é utilizada provamos que um ponto limite pode ser a solução de um problema de quadrados mínimos com pesos envolvendo as restrições. Nessa prova utilizaremos a

seguinte propriedade,[4]:

$$\|\mathcal{P}_{\Omega_2}(u + tv) - u\|_2 \leq \|\mathcal{P}_{\Omega_2}(u + v) - u\|_2 \quad \forall u \in \Omega_2, v \in \mathbb{R}^n, t \in [0, 1]. \quad (2.82)$$

Teorema 2.5.1. *Seja $\{x_k\}_{k \in \mathbb{N}}$ a seqüência gerada pelo Algoritmo 2.5.1.*

- *Suponhamos que a Regra 1 é utilizada e que $[\rho_0]_i = \rho_0$ para todo $i = 1, \dots, m$. Se x_* é um ponto limite de $\{x_k\}$ então x_* é um ponto estacionário do problema*

$$\begin{aligned} & \text{Minimizar } \frac{1}{2} \sum h_i(x)^2 \\ & \text{sujeito a } x \in \Omega_2. \end{aligned} \quad (2.83)$$

- *Se a Regra 2 é utilizada, uma das seguintes possibilidades se verifica:*

1. *A seqüência admite um ponto limite viável.*
2. *A seqüência admite um ponto limite x_* inviável para o qual existe $w \in \mathbb{R}_+^m$ tal que x_* é um ponto estacionário do problema:*

$$\begin{aligned} & \text{Minimizar } \sum_{i=1}^m w_i h_i(x)^2 \\ & \text{sujeito a } x \in \Omega_2, \end{aligned} \quad (2.84)$$

e

$$\sum_{i=1}^m w_i h_i(x_*)^2 > 0. \quad (2.85)$$

Prova. Se a Regra 1 é utilizada a prova é a mesma do Teorema 2.2.1 já que todos os pontos da caixa satisfazem a condição CPLD.

Suponhamos então que é utilizada a Regra 2. Assumimos que todos os pontos da seqüência $\{x_k\}$ são inviáveis. Conseqüentemente, existe $\varepsilon > 0$ tal que

$$\|h(x_k)\|_\infty \geq \varepsilon \quad (2.86)$$

para todo $k \in \mathbb{N}$. Isto implica que

$$\lim_{k \rightarrow \infty} \|\rho_k\|_\infty = \infty.$$

Seja K um subconjunto infinito de $\mathbb{N}$ tal que

$$\|\rho_k\|_\infty > \|\rho_{k-1}\|_\infty \quad \forall k \in K. \quad (2.87)$$

Seja K_1 um subconjunto infinito de K e $j \in \{1, \dots, m\}$ tais que

$$\|\rho_k\|_\infty = [\rho_k]_j \quad \forall k \in K_1. \quad (2.88)$$

Então, por (2.87),

$$[\rho_k]_j = \gamma[\rho_{k-1}]_j \quad \forall k \in K_1.$$

Pela definição do algoritmo, temos que, para todo $k \in K_1$,

$$|h_j(x_{k-1})| > \tau \|h(x_{k-2})\|_\infty.$$

Logo, por (2.86),

$$|h_j(x_{k-1})| > \tau\varepsilon \quad \forall k \in K_1. \quad (2.89)$$

Além disso, pela definição do algoritmo, por (2.87) e (2.88), obtemos:

$$[\rho_{k-1}]_j \geq \frac{\|\rho_{k-1}\|_\infty}{\gamma} \quad \forall k \in K_1. \quad (2.90)$$

Seja K_2 um subconjunto infinito de índices de $\{k-1\}_{k \in K_1}$ tal que

$$\lim_{k \in K_2} x_k = x_*.$$

Por (2.89) temos que

$$h_j(x_*) \neq 0. \quad (2.91)$$

Por (2.79) e pela equivalência entre normas em $\mathbb{R}^n$, obtemos:

$$\lim_{k \rightarrow \infty} \|\mathcal{P}_{\Omega_2}[x_k - \nabla f(x_k) - \sum_{i=1}^m ([\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k)) \nabla h_i(x_k)] - x_k\|_2 = 0. \quad (2.92)$$

Como $\|\rho_k\|_\infty \rightarrow \infty$ temos que $\|\rho_k\|_\infty > 1$ para $k \in K_2$ suficientemente grande. Logo, usando (2.82) com

$$u = x_k, \quad v = -\nabla f(x_k) - \sum_{i=1}^m ([\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k)) \nabla h_i(x_k), \quad t = 1/\|\rho_k\|_\infty,$$

temos, por (2.92), que

$$\lim_{k \in K_2} \left\| \mathcal{P}_{\Omega_2} \left[x_k - \frac{\nabla f(x_k)}{\|\rho_k\|_\infty} - \sum_{i=1}^m \left(\frac{[\bar{\lambda}_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_i}{\|\rho_k\|_\infty} h_i(x_k) \right) \nabla h_i(x_k) \right] - x_k \right\|_2 = 0. \quad (2.93)$$

Mas

$$\frac{[\rho_k]_i}{\|\rho_k\|_\infty} \leq 1 \quad \forall i = 1, \dots, m.$$

Conseqüentemente, existem $K_3 \subset K_2$ e $w \in \mathbb{R}_+^m$ tais que

$$\lim_{k \in K_1} \frac{[\rho_k]_i}{\|\rho_k\|_\infty} = w_i \quad \forall i = 1, \dots, m.$$

Além disso, por (2.90),

$$w_j > 0. \quad (2.94)$$

Como $\{\bar{\lambda}_k\}_{k \in K_1}$ é limitada, tomando limite para $k \in K_1$ em (2.93), obtemos:

$$\|\mathcal{P}_{\Omega_2}[x_* - \sum_{i=1}^m w_i h_i(x_*) \nabla h_i(x_*)] - x_*\|_2 = 0.$$

Logo, x_* é um ponto estacionário de (2.84). Por (2.91) e (2.94), verifica-se a condição (2.85). Como consequência, x_* satisfaz a condição 2. $\square$

Observação. Denominemos ponto *degenerado* a um ponto x para o qual existe $w \in \mathbb{R}_+^m$ satisfazendo (2.85) tal que x é um ponto estacionário do problema (2.84). Assim, qualquer ponto estacionário inviável que é solução do problema (2.83) é degenerado. Além disso, se x é inviável, por (2.85) e pelas condições KKT do problema (2.84), temos que os gradientes das restrições de igualdade e de canalização ativas em x são linearmente dependentes. A recíproca não é verdadeira: consideremos o seguinte conjunto de restrições

$$h(x) = x = 0 \in \mathbb{R}^1, \quad -1 \leq x \leq 1. \quad (2.95)$$

Nos pontos $z = -1$ e $z = 1$ os gradientes das restrições de igualdade e de canalização ativas em x são linearmente dependentes mas esses pontos não são degenerados. Em [16] os autores assumem que, em todos os pontos da sequência gerada pelo algoritmo, os gradientes das restrições de igualdade e de canalização ativas em x são linearmente independentes (Hipótese AS3 de [16], ver também [14]). Utilizando esta hipótese os autores provam que os pontos limites são viáveis. Pelas considerações acima, vemos que a Hipótese AS3 é mais forte que a hipótese sobre não existência de pontos inviáveis degenerados. Em outras palavras, a hipótese sobre o problema que garante que AS3 se verifica é que os gradientes das restrições de igualdade e de canalização ativas são linearmente independentes em todos os pontos da caixa (não necessariamente nos pontos viáveis). Esta hipótese não se verifica no exemplo (2.95) e é mais forte que a hipótese sobre a não existência de pontos degenerados no conjunto viável.

A prova de convergência para pontos ótimos se segue das idéias apresentadas na prova

do Teorema 2.3.1.

Teorema 2.5.2. *Seja $\{x_k\}_{k \in \mathbb{N}}$ uma seqüência gerada pelo Algoritmo 2.5.1. Assumimos que $x_* \in \Omega_1 \cap \Omega_2$ é um ponto limite que satisfaz a condição de qualificação CPLD associada ao conjunto viável $\Omega_1 \cap \Omega_2$. Então, x_* é um ponto KKT do problema original (2.78).*

2.5.2 Limitação dos parâmetros de penalidade

Por simplicidade e sem perda de generalidade consideramos o problema

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } h(x) = 0, \\ & x_1 \geq 0, \dots, x_n \geq 0. \end{aligned} \tag{2.96}$$

As seguintes definições e notações podem ser encontradas em [16].

* Definimos $N = \{1, \dots, n\}$ e $\Omega_n = \{x \in \mathbb{R}^n : x_1 \geq 0, \dots, x_n \geq 0\}$ e denotamos

$$\mathcal{P}_n = \text{operador projeção euclidiana sobre } \Omega_n.$$

* Denotamos $g(x) = \nabla f(x)$, $H(x) = \nabla^2 f(x)$ gradiente e Hessiana, respectivamente, da função $f(x)$; $A(x) = \nabla h(x)^T$ Jacobiano de $h(x)$ e $H_i(x) = \nabla^2 h_i(x)$ Hessiana de $h_i(x)$.

Definimos a função de Lagrange $l(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i h_i(x)$ e denotamos $g_l(x, \lambda) = \nabla_x l(x, \lambda)$, $H_l(x, \lambda) = \nabla_{xx}^2 l(x, \lambda)$ gradiente e matriz Hessiana respectivamente, tomados com relação ao primeiro argumento da função l .

* Suponhamos que $\{x_k\} \subset \Omega_n$ e $\{\bar{\lambda}_k\}$ são seqüências infinitas de vetores em $\mathbb{R}^n$ e $\mathbb{R}^m$ respectivamente, e que $\{\rho_k\}$ é uma seqüência infinita de vetores em $\mathbb{R}^m$ tal que $[\rho_k]_i > 0$ para todo k e todo i .

Para cada ponto x_k , temos duas possibilidades para cada componente $[x_k]_i$, se $i \in N$:

$$\begin{aligned} & \text{(i) } 0 \leq [x_k]_i \leq [\nabla L(x_k, \bar{\lambda}_k, \rho_k)]_i, \text{ ou} \\ & \text{(ii) } [\nabla L(x_k, \bar{\lambda}_k, \rho_k)]_i < [x_k]_i. \end{aligned}$$

Denominamos variável *dominada em x_k* àquela variável que satisfaz (i) e variável *flutuante em x_k* àquela variável que satisfaz (ii).

No caso (i) temos que

$$[\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k]_i = -[x_k]_i, \tag{2.97}$$

enquanto no caso (ii) temos que

$$[\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k]_i = -[\nabla L(x_k, \bar{\lambda}_k, \rho_k)]_i. \quad (2.98)$$

Mais ainda, se x_* é um ponto limite da seqüência $\{x_k\}$, definimos os seguintes subconjuntos de N :

$$I_1 = \{i \in N : [x_k]_i \text{ é flutuante } \forall k \text{ suficientemente grande e } [x_*]_i > 0\}, \quad (2.99)$$

$$I_2 = \{i \in N : [x_k]_i \text{ é dominada } \forall k \text{ suficientemente grande}\} \quad (2.100)$$

e,

$$I_3 = N - (I_1 \cup I_2). \quad (2.101)$$

* Denotamos por $\hat{y}$ ao vetor formado pelas componentes i de y tais que $i \in I_1$ e $\hat{A}$ à matriz formada de A tomando as colunas i tais que $i \in I_1$.

* Se J_1 e J_2 são subconjuntos de $\{1, \dots, n\}$, $B_{[J_1, J_2]}$ é a submatriz da matriz B formada pelas linhas e colunas indexadas pelos conjuntos J_1 e J_2 respectivamente. $C_{[J_1]}$ é a submatriz da matriz C formada pelas colunas indexadas por J_1 e $y_{[J_1]}$ é o vetor formado tomando do vetor y as componentes i tais que $i \in J_1$.

* Definimos

$$\lambda(x) = -(\hat{A}(x)^\dagger)^T \hat{g}(x) \quad (2.102)$$

para todos os pontos onde a pseudo-inversa de $\hat{A}(x)$, dada por $\hat{A}(x)^\dagger = \hat{A}(x)^T (\hat{A}(x) \hat{A}(x)^T)^{-1}$, está bem definida.

A função $\lambda(x)$ é diferenciável. No seguinte lema (Lema 2.2 de [16]) calcula-se a derivada de $\lambda(x)$.

Lema 2.5.3. *Suponhamos que as funções $f(x)$ e $h_i(x)$ são duas vezes continuamente diferenciáveis em Ω_n . Se $\hat{A}(x) \hat{A}(x)^T$ é não-singular, então $\lambda(x)$ é diferenciável e sua derivada está dada por*

$$\lambda'(x) = -(\hat{A}(x)^\dagger)^T \hat{H}_l(x, \lambda(x)) - (\hat{A}(x) \hat{A}(x)^T)^{-1} R(x) \quad (2.103)$$

sendo a i -ésima linha de $R(x)$ dada por $(\hat{g}(x) + \hat{A}(x)^T \lambda(x))^T \hat{H}_i(x)$.

Prova. Definimos

$$r(x) = \hat{g}(x) + \hat{A}(x)^T \lambda(x).$$

Por (2.102) temos que

$$\hat{A}(x)r(x) = \hat{A}(x)\hat{g}(x) - \hat{A}(x)\hat{A}(x)^T(\hat{A}(x)^\dagger)^T\hat{g}(x) = 0.$$

Logo, derivando as equações

$$\begin{aligned} r(x) - \hat{A}(x)^T \lambda(x) &= \hat{g}(x) \\ \hat{A}(x)r(x) &= 0 \end{aligned}$$

obtemos que

$$\begin{aligned} r'(x) - \hat{A}(x)^T \lambda'(x) - \sum_{i=1}^m \hat{H}_i(x) \lambda_i(x) &= \hat{H}(x) \\ R(x) + \hat{A}(x)r'(x) &= 0. \end{aligned} \tag{2.104}$$

Pre-multiplicando a primeira equação de (2.104) por $\hat{A}(x)$ e substituindo $\hat{A}(x)r'(x) = -R(x)$ da segunda equação obtemos

$$-R(x) - \hat{A}(x)\hat{A}(x)^T \lambda'(x) - \sum_{i=1}^m \hat{A}(x)\hat{H}_i(x) \lambda_i(x) = \hat{A}(x)\hat{H}(x). \tag{2.105}$$

Pre-multiplicando por $(\hat{A}(x)\hat{A}(x)^T)^{-1}$, usando que $\hat{A}(x)^\dagger = \hat{A}(x)^T(\hat{A}(x)\hat{A}(x)^T)^{-1}$ e isolando $\lambda'(x)$ de (2.105) temos que

$$\begin{aligned} \lambda'(x) &= -(\hat{A}(x)^\dagger)^T(\hat{H}(x) + \sum_{i=1}^m \hat{H}_i(x) \lambda_i(x)) - (\hat{A}(x)\hat{A}(x)^T)^{-1}R(x) = \\ &= -(\hat{A}(x)^\dagger)^T \hat{H}_l(x, \lambda(x)) - (\hat{A}(x)\hat{A}(x)^T)^{-1}R(x), \end{aligned}$$

como queríamos provar. □

Hipótese 2.5.4. *Assumimos que as funções $f(x)$ e $h_i(x)$ são duas vezes continuamente diferenciáveis e as derivadas segundas são Lipschitz contínuas em todo ponto do conjunto Ω_n .*

Hipótese 2.5.5. *Suponhamos que (x_*, λ_*) é um ponto estacionário do problema (2.96) e que*

$$J_1 = \{ i \in N \mid [g_l(x_*, \lambda_*)]_i = 0 \text{ e } [x_*]_i > 0 \}$$

$$J_2 = \{ i \in N \mid [g_l(x_*, \lambda_*)]_i = 0 \text{ e } [x_*]_i = 0 \}.$$

Então, assumimos que a matriz

$$\begin{pmatrix} H_l(x_*, \lambda_*)_{[J,J]} & (A(x_*)_{[J]})^T \\ A(x_*)_{[J]} & 0 \end{pmatrix}$$

é não-singular para qualquer conjunto J , sendo $J = J_1 \cup K$ para $K \subset J_2$.

Lembremos que um ponto x que satisfaz o conjunto de restrições

$$h(x) = 0, \quad x_1 \geq 0, \dots, x_n \geq 0 \quad (2.106)$$

é regular, Definição 1.1.3, se os gradientes $\{\nabla h_i(x)\}$ das restrições de igualdade junto com os gradientes $\{e_i\}_{\{i \in N : x_i=0\}}$ das restrições de desigualdade que são ativas no ponto são linearmente independentes. Observar que a Hipótese 2.5.5 implica a regularidade.

Além disso e como observam os autores em [16], se $J_2 = \emptyset$ então a conhecida condição suficiente de segunda ordem para um minimizador de (2.96), (ver por exemplo o livro de Fletcher, [20]) implica a condição definida na Hipótese 2.5.5. Quando J_2 é não vazio, a conexão entre a Hipótese 2.5.5 e a condição definida no livro do Fletcher é menos evidente, mais certamente a condição suficiente de segunda ordem forte definida em Luenberguer [26, p. 234–235] implica a Hipótese 2.5.5.

O seguinte lema estabelece uma equivalência com a regularidade definida acima.

Lema 2.5.6. *Seja x_* ponto limite de uma seqüência $\{x_k\} \subset \Omega_n$ que satisfaz as restrições (2.106). Sejam $\{\bar{\lambda}_k\}$ e $\{\rho_k\}$ seqüências de vetores em $\mathbb{R}^m$ tais que $[\rho_k]_i \geq [\rho_{k-1}]_i > 0$ para todo k e todo $i = 1, \dots, m$ e suponhamos que*

$$\lim_{k \rightarrow \infty} \mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k = 0. \quad (2.107)$$

As seguintes condições são equivalentes:

(i) x_* é regular,

(ii) a matriz $\hat{A}(x_*)$ tem posto não menor a m .

Prova.

(i) $\Rightarrow$ (ii): Por definição de I_1 temos que, se $i \in N - I_1$ e $[x_k]_i$ não é uma variável flutuante para k suficientemente grande, existe uma subsequência tal que se cumpre (2.97). Logo, por (2.107),

$$[x_*]_i = \lim_{k \rightarrow \infty} [x_k]_i = 0. \quad (2.108)$$

Seja $p = |I_1|$. Suponhamos que o posto da matriz $\hat{A}(x_*) \in \mathbb{R}^{m \times p}$ é $s < m$.

Então, o sistema linear

$$\hat{A}(x_*)^T \alpha = 0$$

é um sistema de p equações, m variáveis e matriz do sistema com posto $s < m$ e, portanto, existe uma solução α_0 não nula tal que

$$\hat{A}(x_*)^T \alpha_0 = 0.$$

Denotamos por $[B]_i$, à linha i da matriz B e definimos

$$\beta_i = -[A(x_*)^T]_i \cdot \alpha_0 \text{ se } [x_*]_i = 0.$$

Logo, por (2.108), temos que existem α_0, β não todos nulos tais que

$$\sum_{i=1}^m [\alpha_0]_i \nabla h_i(x_*) + \sum_{i: [x_*]_i=0} \beta_i e_i = A(x_*)^T \alpha_0 + \sum_{i: [x_*]_i=0} \beta_i e_i = 0$$

o que contradiz a hipótese de regularidade no ponto x_* .

(ii) $\Rightarrow$ (i): Suponhamos que $\hat{A}(x_*)$ tem posto não menor a m .

Se α_i, β_i são escalares tais que

$$\sum_{i=1}^m \alpha_i \nabla h_i(x_*) + \sum_{i: [x_*]_i=0} \beta_i e_i = 0$$

então,

$$A(x_*)^T \alpha + \sum_{i: [x_*]_i=0} \beta_i e_i = 0. \quad (2.109)$$

Como $[x_*]_i > 0$ se $i \in I_1$, por (2.109), temos que

$$\hat{A}(x_*)^T \alpha = 0$$

e portanto, como $\text{posto}(\hat{A}(x_*)^T) \geq m$ deve ocorrer $\alpha = 0$. Logo, por (2.109) temos $\beta_i = 0$ para todo i tal que $[x_*]_i = 0$, o que prova a regularidade de x_* . $\square$

Os Lemas 2.5.7 e 2.5.8 provados a seguir são os Lemas 4.3 e 5.1 respectivamente, da referência [16].

Lema 2.5.7. *Suponhamos que se verifica a Hipótese 2.5.4. Seja $\{x_k\} \subset \Omega_n$ uma seqüência*

que converge ao ponto x_* , ponto regular do problema (2.96). Seja $\lambda_* = \lambda(x_*)$, para $\lambda(\cdot)$ definida em (2.102). Sejam $\{\bar{\lambda}_k\}$ uma seqüência de vetores em $\mathbb{R}^m$ e $\{\rho_k\}$ uma seqüência não-decrescente de escalares positivos. Suponhamos que $[\rho_k]_i = \rho_k$ para todo $i = 1, \dots, m, k \in \mathbb{N}$ e que

$$\|\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\| \leq \varepsilon_k \quad (2.110)$$

para $\{\varepsilon_k\}$ uma seqüência de escalares positivos que converge para 0.

Então, se

$$\lambda_{k+1} = \bar{\lambda}_k + \rho_k h(x_k), \quad (2.111)$$

existem constantes positivas a_1 e a_2 e um inteiro k_0 tais que

$$\|\lambda_{k+1} - \lambda_*\| \leq a_1 \varepsilon_k + a_2 \|x_k - x_*\| \quad (2.112)$$

$$\|h(x_k)\| \leq a_1 \frac{\varepsilon_k}{\rho_k} + \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} + a_2 \frac{\|x_k - x_*\|}{\rho_k} \quad (2.113)$$

para todo $k \geq k_0$.

Prova. Pela regularidade de x_* e por continuidade existe k_1 tal que para todo $k \geq k_1$, existe $(\hat{A}(x_k))^\dagger$, é limitada e converge para $(\hat{A}(x_*))^\dagger$. Portanto, existe $a_1 > 0$ tal que $\|(\hat{A}(x_k))^\dagger\|^T \leq a_1$.

Pela definição de λ_{k+1} e de $A(x_k)$ temos que

$$\nabla L(x_k, \bar{\lambda}_k, \rho_k) = \nabla f(x_k) + \nabla h(x_k) \lambda_{k+1} = \nabla f(x_k) + A(x_k)^T \lambda_{k+1}. \quad (2.114)$$

Usando (2.98) e (2.114) temos que, se $i \in I_1$, então

$$\begin{aligned} [\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k]_i &= -[\nabla L(x_k, \bar{\lambda}_k, \rho_k)]_i = \\ &= -[\nabla f(x_k) + A(x_k)^T \lambda_{k+1}]_i = -(g_i(x_k) + [A(x_k)^T]_{i,\cdot} [\lambda_{k+1}]_i) \end{aligned}$$

onde, como antes, $[A(x_k)^T]_{i,\cdot}$ indica a linha i da matriz $A(x_k)^T$.

Logo, pela definição de $\hat{A}(x_k)$, temos que

$$[\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k]_{[I_1]} = \hat{g}(x_k) + \hat{A}(x_k)^T \lambda_{k+1}.$$

Então, por (2.110)

$$\|\hat{g}(x_k) + \hat{A}(x_k)^T \lambda_{k+1}\| \leq \varepsilon_k.$$

Logo, pela definição de $\lambda(x)$ dada em (2.102) e por propriedades da pseudo-inversa

$$\begin{aligned}\|\lambda_{k+1} - \lambda(x_k)\| &= \|\lambda_{k+1} + (\hat{A}(x_k)^\dagger)^T \hat{g}(x_k)\| = \\ &= \|(\hat{A}(x_k)^\dagger)^T ((\hat{A}(x_k))^T \lambda_{k+1} + \hat{g}(x_k))\| \leq a_1 \varepsilon_k.\end{aligned}\tag{2.115}$$

Pelo Lema 2.5.3 e pelo Teorema do Valor Médio da Integral

$$\lambda(x_k) - \lambda(x_*) = \int_0^1 \lambda'(x(s)) ds (x_k - x_*)$$

para $\lambda'(x)$ dada pela equação (2.103) e $x(s) = x_k + s(x_* - x_k)$.

Pela diferenciabilidade de $\lambda(x)$ temos que, para x_k suficientemente perto de x_* , existe $a_2 > 0$ tal que

$$\|\lambda(x_k) - \lambda(x_*)\| \leq a_2 \|x_k - x_*\|.\tag{2.116}$$

Então, combinando (2.115) e (2.116) obtemos

$$\|\lambda_{k+1} - \lambda_*\| \leq \|\lambda_{k+1} - \lambda(x_k)\| + \|\lambda(x_k) - \lambda_*\| \leq a_1 \varepsilon_k + a_2 \|x_k - x_*\|$$

que é a desigualdade (2.112).

Usando (2.111) temos

$$h(x_k) = \frac{\lambda_{k+1} - \bar{\lambda}_k}{\rho_k} = \frac{(\lambda_{k+1} - \lambda_*) + (\lambda_* - \bar{\lambda}_k)}{\rho_k}.\tag{2.117}$$

Tomando norma em (2.117), usando a desigualdade triangular e (2.112) obtemos (2.113).

□

Lema 2.5.8. *Suponhamos que se verifica a Hipótese 2.5.4. Seja $\{x_k\} \subset \Omega_n$ uma seqüência que converge ao ponto KKT x_* para o qual verifica-se a Hipótese 2.5.5. Seja λ_* o multiplicador de Lagrange associado. Sejam $\{\bar{\lambda}_k\}$ uma seqüência de vetores em $\mathbb{R}^m$ e $\{\rho_k\}$ uma seqüência não-decrescente de escalares positivos. Suponhamos que $[\rho_k]_i = \rho_k$ para todo $i = 1, \dots, m, k \in \mathbb{N}$ e que*

$$\|\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\| \leq \varepsilon_k\tag{2.118}$$

para $\{\varepsilon_k\}$ uma seqüência de escalares positivos que converge para 0.

Então, se

$$\lambda_{k+1} = \bar{\lambda}_k + \rho_k h(x_k),\tag{2.119}$$

existem constantes positivas $\bar{\rho}, a_3, a_4, a_5, a_6$ e um inteiro k_0 tais que, se $\rho_{k_0} \geq \bar{\rho}$ tem-se que

$$\|x_k - x_*\| \leq a_3 \varepsilon_k + a_4 \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.120)$$

$$\|\lambda_{k+1} - \lambda_*\| \leq a_5 \varepsilon_k + a_6 \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.121)$$

$$\|h(x_k)\| \leq a_5 \varepsilon_k \frac{1}{\rho_k} + \left(1 + \frac{a_6}{\rho_k}\right) \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.122)$$

para todo $k \geq k_0$.

Prova. Seja k suficientemente grande tal que se tem as desigualdades (2.112) e (2.113) do Lema 2.5.7:

$$\|\lambda_{k+1} - \lambda_*\| \leq a_1 \varepsilon_k + a_2 \|x_k - x_*\| \quad (2.123)$$

$$\|h(x_k)\| \leq a_1 \frac{\varepsilon_k}{\rho_k} + \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} + a_2 \frac{\|x_k - x_*\|}{\rho_k}. \quad (2.124)$$

Seja k suficientemente grande tal que os conjuntos I_1, I_2 definidos em (2.99) e (2.100) estão bem determinados.

Seja I_3 o conjunto definido em (2.101), formado pelos índices do conjunto N que oscilam entre dominados e flutuantes, e seja $K \subset \mathbb{N}$ um subconjunto infinito tal que

- (i) $I_3 = I_4 \cup I_5$ com $I_4 \cap I_5 = \emptyset$.
- (ii) As variáveis no conjunto I_4 são flutuantes para todo $k \in K$.
- (iii) As variáveis no conjunto I_5 são dominadas para todo $k \in K$.

Notar que há só um número finito de tais subconjuntos K ($\leq 2^{|I_3|}$) e que para todo k suficientemente grande, cada $k \in \mathbb{N}$ está num desses subconjuntos. É assim suficiente provar o lema para $k \in K$.

Agora, para $k \in K$ definimos

$$I_F = I_1 \cup I_4 \quad \text{e} \quad I_D = I_2 \cup I_5.$$

Então, as variáveis em I_F são flutuantes enquanto as variáveis em I_D são dominadas. Por (2.123) temos que $\lambda_{k+1} \rightarrow \lambda_*$ e portanto, por continuidade,

$$\nabla L(x_k, \bar{\lambda}_k, \rho_k) = \nabla f(x_k) + \nabla h(x_k) \lambda_{k+1} \rightarrow \nabla f(x_*) + \nabla h(x_*) \lambda_* = g_l(x_*, \lambda_*).$$

Logo, por (2.118) e (2.97),

$$[x_*]_i = 0 \quad \forall i \in I_D \quad (2.125)$$

e, por (2.118) e (2.98) obtemos

$$[g_l(x_*, \lambda_*)]_i = [\nabla f(x_*) + \nabla h(x_*)\lambda_*]_i = 0 \quad \forall i \in I_F. \quad (2.126)$$

Usando a expansão de Taylor temos que

$$g(x_k) = g(x_*) + H(x_*)(x_k - x_*) + \bar{r}_1(x_k, x_*)$$

e

$$\begin{aligned} A(x_k)^T \lambda_{k+1} &= A(x_*)^T \lambda_{k+1} + \sum_{j=1}^m [\lambda_{k+1}]_j H_j(x_*)(x_k - x_*) + \bar{r}_2(x_k, x_*, \lambda_{k+1}) = \\ &= A(x_*)^T \lambda_* + A(x_*)^T (\lambda_{k+1} - \lambda_*) + \sum_{j=1}^m [\lambda_*]_j H_j(x_*)(x_k - x_*) + \\ &+ \sum_{j=1}^m ([\lambda_{k+1}]_j - [\lambda_*]_j) H_j(x_*)(x_k - x_*) + \bar{r}_2(x_k, x_*, \lambda_{k+1}) \end{aligned}$$

para

$$\bar{r}_1(x_k, x_*) = \int_0^1 [H(x_k + s(x_* - x_k)) - H(x_*)](x_k - x_*) ds$$

e

$$\bar{r}_2(x_k, x_*, \lambda_{k+1}) = \int_0^1 \sum_{j=1}^m [\lambda_{k+1}]_j [H_j(x_k + s(x_* - x_k)) - H_j(x_*)](x_k - x_*) ds.$$

Então,

$$\begin{aligned} \nabla L(x_k, \bar{\lambda}_k, \rho_k) &= g(x_k) + A(x_k)^T \lambda_{k+1} = g(x_*) + A(x_*)^T \lambda_* + \\ &+ [H(x_*) + \sum_{j=1}^m [\lambda_*]_j H_j(x_*)](x_k - x_*) + A(x_*)^T (\lambda_{k+1} - \lambda_*) + \\ &+ \sum_{j=1}^m ([\lambda_{k+1}]_j - [\lambda_*]_j) H_j(x_*)(x_k - x_*) + \bar{r}_1(x_k, x_*) + \bar{r}_2(x_k, x_*, \lambda_{k+1}) = \end{aligned} \quad (2.127)$$

$$\begin{aligned} &g_l(x_*, \lambda_*) + H_l(x_*, \lambda_*)(x_k - x_*) + A(x_*)^T (\lambda_{k+1} - \lambda_*) + \\ &+ r_1(x_k, x_*, \lambda_{k+1}) + r_2(x_k, x_*, \lambda_{k+1}, \lambda_*), \end{aligned}$$

para

$$\begin{aligned} r_1(x_k, x_*, \lambda_{k+1}) &= \bar{r}_1(x_k, x_*) + \bar{r}_2(x_k, x_*, \lambda_{k+1}) = \\ &= \int_0^1 [H_l(x_k + s(x_* - x_k), \lambda_{k+1}) - H_l(x_*, \lambda_{k+1})](x_k - x_*) ds \end{aligned}$$

e

$$r_2(x_k, x_*, \lambda_{k+1}, \lambda_*) = \sum_{j=1}^m ([\lambda_{k+1}]_j - [\lambda_*]_j) H_j(x_*)(x_k - x_*).$$

Pela limitação e continuidade Lipschitz das matrizes Hessianas de f e h_i numa vizinhança de x_* , e pela convergência de λ_{k+1} a λ_* , obtemos que

$$\begin{aligned} \|r_1(x_k, x_*, \lambda_{k+1})\| &\leq a_7 \|x_k - x_*\|^2 \\ \|r_2(x_k, x_*, \lambda_{k+1}, \lambda_*)\| &\leq a_8 \|x_k - x_*\| \|\lambda_{k+1} - \lambda_*\| \end{aligned} \quad (2.128)$$

para constantes positivas a_7, a_8 .

Usando outra vez a expansão de Taylor e o fato de que $h(x_*) = 0$,

$$h(x_k) = A(x_*)(x_k - x_*) + r_3(x_k, x_*), \quad (2.129)$$

para

$$r_3(x_k, x_*) = \int_0^1 s \int_0^1 (x_k - x_*)^T H_i(x_* + ts(x_k - x_*))(x_k - x_*) dt ds.$$

Pela limitação das matrizes Hessianas de h_i numa vizinhança de x_* obtemos que

$$\|r_3(x_k, x_*)\| \leq a_9 \|x_k - x_*\|^2 \quad (2.130)$$

para alguma constante $a_9 > 0$.

Combinando (2.127) e (2.129) obtemos que

$$\begin{aligned} &\begin{pmatrix} H_l(x_*, \lambda_*) & A(x_*)^T \\ A(x_*) & 0 \end{pmatrix} \begin{pmatrix} x_k - x_* \\ \lambda_{k+1} - \lambda_* \end{pmatrix} = \\ &\begin{pmatrix} \nabla L(x_k, \bar{\lambda}_k, \rho_k) - g_l(x_*, \lambda_*) \\ h(x_k) \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ r_3 \end{pmatrix} \end{aligned} \quad (2.131)$$

Introduzindo a notação $y_{[J]}$ e usando que $[x_*]_i = 0$ se $i \in I_D$, $[g_l(x_*, \lambda_*)]_i = 0$ se $i \in I_F$, podemos reescrever (2.131) como

$$\begin{aligned} &\begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F, I_F]} & H_l(x_*, \lambda_*)_{[I_F, I_D]} & A(x_*)_{[I_F]}^T \\ H_l(x_*, \lambda_*)_{[I_D, I_F]} & H_l(x_*, \lambda_*)_{[I_D, I_D]} & A(x_*)_{[I_D]}^T \\ A(x_*)_{[I_F]} & A(x_*)_{[I_D]} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ (x_k)_{[I_D]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} = \\ &\begin{pmatrix} \nabla L(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} \\ (\nabla L(x_k, \bar{\lambda}_k, \rho_k) - g_l(x_*, \lambda_*))_{[I_D]} \\ h(x_k) \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ (r_1 + r_2)_{[I_D]} \\ r_3 \end{pmatrix}. \end{aligned} \quad (2.132)$$

Removendo a segunda coluna em (2.132) obtemos

$$\begin{aligned}
& \begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F, I_F]} & A(x_*)_{[I_F]}^T \\ H_l(x_*, \lambda_*)_{[I_D, I_F]} & A(x_*)_{[I_D]}^T \\ A(x_*)_{[I_F]} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} = \\
& \begin{pmatrix} \nabla L(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} \\ (\nabla L(x_k, \bar{\lambda}_k, \rho_k) - g_l(x_*, \lambda_*))_{[I_D]} \\ h(x_k) \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ (r_1 + r_2)_{[I_D]} \\ r_3 \end{pmatrix} - \\
& \begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F, I_D]}(x_k)_{[I_D]} \\ H_l(x_*, \lambda_*)_{[I_D, I_D]}(x_k)_{[I_D]} \\ A(x_*)_{[I_D]}(x_k)_{[I_D]} \end{pmatrix}. \tag{2.133}
\end{aligned}$$

Eliminando a segunda equação em (2.133) obtemos o sistema

$$\begin{aligned}
& \begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F, I_F]} & A(x_*)_{[I_F]}^T \\ A(x_*)_{[I_F]} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} = \\
& \begin{pmatrix} \nabla L(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} - H_l(x_*, \lambda_*)_{[I_F, I_D]}(x_k)_{[I_D]} \\ h(x_k) - A(x_*)_{[I_D]}(x_k)_{[I_D]} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ r_3 \end{pmatrix},
\end{aligned}$$

que podemos reescrever como

$$\begin{aligned}
& \begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F, I_F]} & A(x_*)_{[I_F]}^T \\ A(x_*)_{[I_F]} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} = \\
& \begin{pmatrix} \nabla L(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} - H_l(x_*, \lambda_*)_{[I_F, I_D]}(x_k)_{[I_D]} \\ h(x_k) - A(x_*)_{[I_D]}(x_k)_{[I_D]} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ r_3 \end{pmatrix}. \tag{2.134}
\end{aligned}$$

Observemos que por (2.97) e (2.118)

$$\|(x_k)_{[I_D]}\| \leq \varepsilon_k, \tag{2.135}$$

e, por (2.98) e (2.118)

$$\|\nabla L(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]}\| \leq \varepsilon_k. \tag{2.136}$$

Usando (2.125) e (2.135)

$$\|x_k - x_*\| \leq \|(x_k - x_*)_{[I_F]}\| + \varepsilon_k. \tag{2.137}$$

Seja $\Delta x_k = \|(x_k - x_*)_{[I_F]}\|$.

Combinando (2.123) e (2.137) obtemos

$$\|\lambda_{k+1} - \lambda_*\| \leq a_{10}\varepsilon_k + a_2\Delta x_k \quad (2.138)$$

para $a_{10} = a_1 + a_2$. Além disso, por (2.128), (2.130), (2.137) e (2.138)

$$\left\| \begin{pmatrix} (r_1 + r_2)_{[I_F]} \\ r_3 \end{pmatrix} \right\| \leq a_{11}(\Delta x_k)^2 + a_{12}\Delta x_k\varepsilon_k + a_{13}\varepsilon_k^2$$

para $a_{11} = a_7 + a_9 + a_8a_2$, $a_{12} = 2(a_7 + a_9) + a_8(a_{10} + a_2)$ e $a_{13} = a_7 + a_9 + a_8a_{10}$.

Mais ainda, por (2.135)-(2.136)

$$\left\| \begin{pmatrix} \nabla L(x_k, \bar{\lambda}_k, \rho_k)_{[I_F]} - H_l(x_*, \lambda_*)_{[I_F, I_D]}(x_k)_{[I_D]} \\ h(x_k) - A(x_*)_{[I_D]}(x_k)_{[I_D]} \end{pmatrix} \right\| \leq a_{14}\varepsilon_k + \|h(x_k)\| \quad (2.139)$$

para $a_{14} = 1 + \left\| \begin{pmatrix} H_l(x_*, \lambda_*)_{[I_F, I_D]} \\ A(x_*)_{[I_D]} \end{pmatrix} \right\|$.

Pela Hipótese 2.5.5, a matriz da esquerda de (2.134) é não-singular. Seja M a norma da sua inversa. Multiplicando ambos lados da equação por esta inversa e tomando normas obtemos

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\| \leq M(a_{14}\varepsilon_k + \|h(x_k)\| + a_{11}(\Delta x_k)^2 + a_{12}\Delta x_k\varepsilon_k + a_{13}\varepsilon_k^2). \quad (2.140)$$

Usando (2.124) e (2.137)

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\| \leq M \left(a_{14}\varepsilon_k + \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} + a_{10}\frac{\varepsilon_k}{\rho_k} + a_2\frac{\Delta x_k}{\rho_k} + a_{11}(\Delta x_k)^2 + a_{12}\Delta x_k\varepsilon_k + a_{13}\varepsilon_k^2 \right). \quad (2.141)$$

Agora, suponhamos que k é suficientemente grande e tal que

$$\varepsilon_k \leq \min \left\{ 1, \frac{1}{4Ma_{12}} \right\}. \quad (2.142)$$

Seja

$$\bar{\rho} = \max \left\{ 1, 4Ma_2 \right\}. \quad (2.143)$$

Então, se $\rho_k \geq \bar{\rho}$, por (2.141)-(2.143) obtem-se

$$\Delta x_k \leq \frac{\Delta x_k}{2} + M \left(a_{15}\varepsilon_k + \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} + a_{11}(\Delta x_k)^2 \right) \quad (2.144)$$

para $a_{15} = a_{10} + a_{13} + a_{14}$.

Como Δx_k converge para zero, temos que

$$|\Delta x_k| \leq \frac{1}{4Ma_{11}} \quad (2.145)$$

para todo k suficientemente grande. Portanto, das desigualdades (2.144) e (2.145) obtemos que

$$\Delta x_k \leq 4M \left(a_{15}\varepsilon_k + \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \right). \quad (2.146)$$

Como $\Delta x_k = \|(x_k - x_*)_{[I_F]}\|$, substituindo (2.146) em (2.137) obtemos que

$$\|x_k - x_*\| \leq \Delta x_k + \varepsilon_k \leq a_3\varepsilon_k + a_4 \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.147)$$

para as constantes $a_3 = 1 + 4Ma_{15}$ e $a_4 = 4M$, o que prova a desigualdade (2.120).

Usando (2.147) e (2.123) obtemos que

$$\|\lambda_{k+1} - \lambda_*\| \leq a_5\varepsilon_k + a_6 \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.148)$$

para as constantes $a_5 = a_1 + a_2a_3$ e $a_6 = a_2a_4$, o que prova a fórmula (2.121).

Finalmente, para obter (2.122), por (2.119), temos que

$$h(x_k) = \frac{\lambda_{k+1} - \bar{\lambda}_k}{\rho_k} = \frac{(\lambda_{k+1} - \lambda_*)}{\rho_k} + \frac{(\lambda_* - \bar{\lambda}_k)}{\rho_k}.$$

Tomando norma, e usando a desigualdade triangular e (2.148) obtemos que

$$\begin{aligned} \|h(x_k)\| &\leq \frac{1}{\rho_k} \left(a_5\varepsilon_k + a_6 \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \right) + \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} = \\ &a_5 \frac{\varepsilon_k}{\rho_k} + \left(\frac{a_6}{\rho_k} + 1 \right) \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \end{aligned}$$

que é a desigualdade (2.122).

Assim, a demonstração está completa. $\square$

Hipótese 2.5.9. *Assumimos que, para todo $k \in \mathbb{N}$, $\bar{\lambda}_k$ é a projeção de λ_k no conjunto $[\bar{\lambda}_{\min}, \bar{\lambda}_{\max}]^m$.*

Lema 2.5.10. *Seja $\{x_k\} \subset \Omega_n$ uma seqüência que converge ao ponto x_* , ponto estacionário e regular do problema (2.96). Seja λ_* o multiplicador de Lagrange associado e suponhamos que*

$$[\lambda_*]_i \in (\bar{\lambda}_{\min}, \bar{\lambda}_{\max}) \quad (2.149)$$

para todo $i \in \{1, \dots, m\}$.

Sejam $\{\bar{\lambda}_k\}$, $\{\rho_k\}$, $\{\varepsilon_k\}$, $\{\lambda_k\}$ seqüências como no Lema 2.5.7.

Se a Hipótese 2.5.9 se verifica então existe k_0 tal que para todo $k \geq k_0$

$$\bar{\lambda}_k = \lambda_k.$$

Prova. Seja k_1 tal que para todo $k \geq k_1$ tem-se a desigualdade (2.112) do Lema 2.5.7:

$$\|\lambda_{k+1} - \lambda_*\| \leq a_1 \varepsilon_k + a_2 \|x_k - x_*\|. \quad (2.150)$$

Seja $\epsilon = \frac{1}{2} \min_i \{[\lambda_*]_i - \bar{\lambda}_{\min}, \bar{\lambda}_{\max} - [\lambda_*]_i\} > 0$.

Como $\|x_k - x_*\| \rightarrow 0$ e $\varepsilon_k \rightarrow 0$ existe k_2 tal que para todo $k \geq k_2$

$$\|x_k - x_*\| \leq \frac{\epsilon}{2a_2} \quad (2.151)$$

e

$$\varepsilon_k \leq \frac{\epsilon}{2a_1}. \quad (2.152)$$

Seja k_0 tal que $k_0 \geq k_1, k_0 \geq k_2$. Seja $k \geq k_0$.

Então, por (2.150)-(2.152) obtemos que

$$\|\lambda_{k+1} - \lambda_*\| \leq \epsilon.$$

Logo, como $|[\lambda_{k+1}]_i - [\lambda_*]_i| \leq \epsilon$, por definição de ϵ deve ser

$$[\lambda_{k+1}]_i \in [\bar{\lambda}_{\min}, \bar{\lambda}_{\max}].$$

Portanto, por (2.5.9) deve ocorrer $\bar{\lambda}_{k+1} = \lambda_{k+1}$, como queríamos provar. $\square$

Na demonstração de limitação dos parâmetros de penalidade necessitaremos utilizar a propriedade de equivalência entre normas em $\mathbb{R}^p$, [4]: existem $\alpha, \beta > 0$ tais que, para todo $y \in \mathbb{R}^p$

$$\alpha \|y\|_\infty \leq \|y\| \leq \beta \|y\|_\infty. \quad (2.153)$$

O seguinte resultado é o teorema de limitação do parâmetro de penalidade para o Algoritmo 2.5.1 utilizando a Regra 1 de adaptação.

Teorema 2.5.11. *Seja $\{x_k\}$ a seqüência gerada pelo Algoritmo 2.5.1 quando aplicado ao problema (2.96) utilizando a Regra 1. Suponhamos que $\{x_k\}$ converge a um ponto limite x_* , que as Hipóteses 2.5.4, 2.5.5, 2.5.9 verificam-se e que o multiplicador de Lagrange λ_* satisfaz (2.149). Suponhamos também que ε_k é escolhido como*

$$\varepsilon_k = \min\{\varepsilon'_k, \|h(x_k)\|_\infty\} \quad (2.154)$$

para uma seqüência decrescente $\{\varepsilon'_k\}$ que converge para 0. Então a seqüência de parâmetros de penalidade $\{\rho_k\}$ é limitada.

Prova. A seqüência $\{x_k\}$ é gerada satisfazendo a condição (2.79), logo, por (2.153) temos que

$$\|\mathcal{P}_{\Omega_n}[x_k - \nabla f(x_k) - \sum_{i=1}^m [\lambda_{k+1}]_i \nabla h_i(x_k)] - x_k\| \leq \beta \varepsilon_k$$

e podemos utilizar os lemas provados anteriormente.

Seja k_0 o inteiro obtido do Lema 2.5.10. Então, para todo $k \geq k_0$ temos que $\bar{\lambda}_k = \lambda_k$.

Por Lema 2.5.8 temos que existem constantes positivas $\bar{\rho}, a_5, a_6$ e um inteiro k_1 tais que, se $\rho_{k_1} \geq \bar{\rho}$ tem-se as desigualdades (2.120)-(2.122):

$$\|\lambda_{k+1} - \lambda_*\| \leq a_5 \varepsilon_k + a_6 \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.155)$$

$$\|h(x_k)\| \leq a_5 \varepsilon_k \frac{1}{\rho_k} + \left(1 + \frac{a_6}{\rho_k}\right) \frac{\|\bar{\lambda}_k - \lambda_*\|}{\rho_k} \quad (2.156)$$

para todo $k \geq k_1$.

Suponhamos, pelo contrário, que $\rho_k \rightarrow \infty$. Seja $k_2 \geq \max\{k_0, k_1\}$ tal que $\frac{a_5}{\alpha \rho_k} < 1$ para todo $k \geq k_2$.

Por (2.153) e (2.154) temos que $\varepsilon_k \leq \|h(x_k)\|_\infty \leq \frac{1}{\alpha} \|h(x_k)\|$. Logo, por (2.156) obtemos que

$$\|h(x_k)\| \leq \frac{a_5}{\alpha \rho_k} \|h(x_k)\| + \left(1 + \frac{a_6}{\rho_k}\right) \frac{\|\lambda_k - \lambda_*\|}{\rho_k}.$$

Logo,

$$\|h(x_k)\| \leq \left(1 + \frac{a_6}{\rho_k}\right) \left(\frac{1}{1 - \frac{a_5}{\alpha\rho_k}}\right) \frac{\|\lambda_k - \lambda_*\|}{\rho_k}. \quad (2.157)$$

Por outro lado, como $\lambda_k = \lambda_{k-1} + \rho_{k-1}h(x_{k-1})$ temos

$$\|h(x_{k-1})\| = \frac{\|\lambda_k - \lambda_{k-1}\|}{\rho_{k-1}} \geq \frac{\|\lambda_{k-1} - \lambda_*\|}{\rho_{k-1}} - \frac{\|\lambda_k - \lambda_*\|}{\rho_{k-1}}. \quad (2.158)$$

Usando a desigualdade (2.155) para $k - 1$ no lugar de k obtemos que

$$\frac{\|\lambda_{k-1} - \lambda_*\|}{\rho_{k-1}} \geq \frac{\|\lambda_k - \lambda_*\|}{a_6} - \frac{a_5}{a_6} \varepsilon_{k-1}$$

e substituindo em (2.158) temos

$$\|h(x_{k-1})\| \geq \|\lambda_k - \lambda_*\| \left(\frac{1}{a_6} - \frac{1}{\rho_{k-1}} \right) - \frac{a_5}{a_6} \varepsilon_{k-1}.$$

Logo, isolando $\|\lambda_k - \lambda_*\|$ desta última desigualdade e usando que, por (2.154) e (2.153), $\varepsilon_{k-1} \leq \frac{1}{\alpha} \|h(x_{k-1})\|$ obtemos

$$\|\lambda_k - \lambda_*\| \leq \frac{1}{a_6^{-1} - \rho_{k-1}^{-1}} \left(1 + \frac{a_5}{\alpha a_6}\right) \|h(x_{k-1})\|. \quad (2.159)$$

Substituindo (2.159) em (2.157) obtemos

$$\|h(x_k)\| \leq m_k \|h(x_{k-1})\|$$

para

$$m_k = \frac{1}{\rho_k \left(1 - \frac{a_5}{\alpha\rho_k}\right)} \left(1 + \frac{a_6}{\rho_k}\right) \left(1 + \frac{a_5}{\alpha a_6}\right) \frac{1}{a_6^{-1} - \rho_{k-1}^{-1}}.$$

Logo, por (2.153) temos que

$$\|h(x_k)\|_\infty \leq \frac{\beta m_k}{\alpha} \|h(x_{k-1})\|_\infty$$

Como $\lim_{k \rightarrow \infty} \frac{\beta m_k}{\alpha} = 0$, existe $k_3 \geq k_2$ tal que $\frac{\beta m_k}{\alpha} < \tau$ e então $\|h(x_k)\|_\infty \leq \tau \|h(x_{k-1})\|_\infty$, $\forall k \geq k_3$. Logo, temos que $\rho_{k+1} = \rho_k$ para todo $k \geq k_3$ o que produz uma contradição. $\square$

Observação. Observar que como ε'_k tende para 0, ε_k também o faz, como é requerido pelo algoritmo. Porém, quando a regra (2.154) é adotada, ε_k não é definido antes da resolução

dos subproblemas. Ou seja, o algoritmo interno utilizado para resolver cada subproblema pára (com uma solução aproximada x_k) somente quando se satisfaz a condição

$$\|\mathcal{P}_{\Omega_2}[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \leq \min\{\varepsilon'_k, \|h(x_k)\|_\infty\}.$$

Para obter o resultado de limitação do parâmetro de penalidade para o Algoritmo 2.5.1 utilizando a Regra 2 precisamos de novas definições e resultados.

Lema 2.5.12. *Suponhamos que se verifica a Hipótese 2.5.4. Seja $\{x_k\} \subset \Omega_n$ uma seqüência que converge ao ponto x_* , ponto regular do problema (2.96). Seja $\lambda_* = \lambda(x_*)$, para $\lambda(\cdot)$ definida em (2.102). Sejam $\{\bar{\lambda}_k\}$ e $\{\rho_k\}$ seqüências de vetores em $\mathbb{R}^m$ tais que $\{[\rho_k]_i\}$ é uma seqüência não-decrescente de escalares positivos para cada $i \in \{1, \dots, m\}$. Suponhamos que*

$$\|\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\| \leq \varepsilon_k \quad (2.160)$$

para $\{\varepsilon_k\}$ uma seqüência de escalares positivos que converge para 0.

Então, se

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k) \quad i = 1, \dots, m,$$

existem constantes positivas b_1 e b_2 e um inteiro k_0 tais que

$$\|\lambda_{k+1} - \lambda_*\| \leq b_1 \varepsilon_k + b_2 \|x_k - x_*\| \quad (2.161)$$

para todo $k \geq k_0$.

Prova. Pela definição de λ_{k+1} temos que

$$\nabla L(x_k, \bar{\lambda}_k, \rho_k) = \nabla f(x_k) + A(x_k)^T \lambda_{k+1}$$

e portanto a demonstração da desigualdade (2.161) é como a demonstração da desigualdade (2.112) feita no Lema 2.5.7. $\square$

Dadas as seqüências $\{x_k\}$, $\{\bar{\lambda}_k\}$ e $\{\rho_k\}$ geradas pelo Algoritmo 2.5.1 utilizando a Regra 2, definimos os conjuntos

$$I_\infty = \{i \in \{1, \dots, m\} : [\rho_k]_i \rightarrow \infty\}, \text{ e}$$

$$I_a = \{i \in \{1, \dots, m\} : [\rho_k]_i \text{ é limitada}\}.$$

Definimos também

$$\rho_k = \min_{i \in I_\infty} \{[\rho_k]_i\} \quad (2.162)$$

e

$$\eta_k = \sum_{i \in I_a} |h_i(x_k)|. \quad (2.163)$$

O seguinte lema é análogo ao Lema 2.5.8 deste trabalho e portanto, ao Lema 5.1 de [16], adequado para o caso em que há vários parâmetros de penalidade.

Lema 2.5.13. *Suponhamos que se verifica a Hipótese 2.5.4. Seja $\{x_k\} \subset \Omega_n$ uma seqüência que converge a um ponto x_* para o qual verifica-se a Hipótese 2.5.5, e seja λ_* o multiplicador de Lagrange associado. Sejam $\{\bar{\lambda}_k\}$ e $\{\rho_k\}$ seqüências de vetores em $\mathbb{R}^m$ tais que $\{[\rho_k]_i\}$ é uma seqüência não-decrescente de escalares positivos para todo $i \in \{1, \dots, m\}$. Suponhamos também que*

$$\|\mathcal{P}_n[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\| \leq \varepsilon_k$$

para $\{\varepsilon_k\}$ uma seqüência de escalares positivos que converge para 0.

Então, se

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k) \quad i = 1, \dots, m, \quad (2.164)$$

existem constantes positivas $\bar{\rho}$, α_1 , α_2 , α_3 , α_4 , α_5 e um inteiro k_0 tais que, se $[\rho_{k_0}]_i \geq \bar{\rho}$ para todo $i \in I_\infty$, então

$$\|x_k - x_*\| \leq \alpha_1 \varepsilon_k + \alpha_2 \eta_k + \alpha_3 \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}, \quad (2.165)$$

$$\|\lambda_{k+1} - \lambda_*\| \leq \alpha_4 \varepsilon_k + \alpha_5 \|h(x_k)\| \quad (2.166)$$

para todo $k \geq k_0$.

Prova. Seja k suficientemente grande tal que se tem a desigualdade (2.161) do Lema 2.5.12:

$$\|\lambda_{k+1} - \lambda_*\| \leq b_1 \varepsilon_k + b_2 \|x_k - x_*\|, \quad (2.167)$$

com o qual vemos que $\lambda_{k+1} \rightarrow \lambda_*$ quando k cresce.

Como no Lema 2.5.12, temos que

$$\nabla L(x_k, \bar{\lambda}_k, \rho_k) = g(x_k) + \nabla h(x_k) \lambda_{k+1}$$

e, podemos então obter a desigualdade (2.140) do Lema 2.5.8:

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\| \leq M(b_{14}\varepsilon_k + \|h(x_k)\| + b_{11}(\Delta x_k)^2 + b_{12}\Delta x_k\varepsilon_k + b_{13}\varepsilon_k^2) \quad (2.168)$$

para $b_{11} = a_7 + a_9 + a_8b_2$, $b_{12} = 2(a_7 + a_9) + a_8(b_{10} + b_2)$, $b_{13} = a_7 + a_9 + a_8b_{10}$, $b_{10} = b_1 + b_2$ e para as constantes a_7 , a_8 e a_9 obtidas por (2.128) e (2.130).

Usando (2.164) e a desigualdade (2.167) obtemos o seguinte:

$$\begin{aligned} |h_i(x_k)| &= \frac{|[\lambda_{k+1}]_i - [\bar{\lambda}_k]_i|}{[\rho_k]_i} \leq \frac{|[\lambda_{k+1}]_i - [\lambda_*]_i| + |[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} \leq \\ &\leq \frac{b_1\varepsilon_k + b_2\|x_k - x_*\| + |[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}. \end{aligned} \quad (2.169)$$

Logo, como

$$\|h(x_k)\|^2 = \sum_{i \in I_\infty} |h_i(x_k)|^2 + \sum_{i \in I_a} |h_i(x_k)|^2,$$

usando que $(a + b)^2 \leq 2(a^2 + b^2)$, $\sqrt{a^2 + b^2} \leq a + b$ para a e b constantes não negativas e (2.169) para os índices $i \in I_\infty$, das definições (2.162) e (2.163) obtemos

$$\|h(x_k)\| \leq \eta_k + \sqrt{2} \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + 2\sqrt{|I_\infty|} \frac{(b_1\varepsilon_k + b_2\|x_k - x_*\|)}{\rho_k}. \quad (2.170)$$

Lembremos que

$$\|x_k - x_*\| \leq \Delta x_k + \varepsilon_k \quad (2.171)$$

para

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\|.$$

Substituindo (2.170) em (2.168) e usando (2.171) obtemos

$$\begin{aligned} \left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\| &\leq M \left(b_{14}\varepsilon_k + \eta_k + \sqrt{2} \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + \right. \\ &2\sqrt{|I_\infty|} \frac{(b_1\varepsilon_k + b_2\|x_k - x_*\|)}{\rho_k} + b_{11}\Delta x_k^2 + b_{12}\Delta x_k\varepsilon_k + b_{13}\varepsilon_k^2 \Big) \leq \\ &M \left(b_{14}\varepsilon_k + \eta_k + \sqrt{2} \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + \tilde{b}_{10} \frac{\varepsilon_k}{\rho_k} + b_2\sqrt{|I_\infty|} \frac{\Delta x_k}{\rho_k} + \right. \end{aligned}$$

$$b_{11}\Delta x_k^2 + b_{12}\Delta x_k\varepsilon_k + b_{13}\varepsilon_k^2 \Big) \quad (2.172)$$

para $\tilde{b}_{10} = 2\sqrt{|I_\infty|}(b_1 + b_2)$.

Seja k suficientemente grande tal que

$$\varepsilon_k \leq \min \left\{ 1, \frac{1}{4Mb_{12}} \right\} \quad (2.173)$$

e

$$\Delta x_k \leq \frac{1}{4Mb_{11}}. \quad (2.174)$$

Seja $\bar{\rho} = \max\{1, 8\sqrt{|I_\infty|}Mb_2\}$. Se k é suficientemente grande tal que $[\rho_k]_i \geq \bar{\rho}$ para todo $i \in I_\infty$, de (2.172)-(2.174) obtemos que

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\| \leq 4M \left((b_{14} + \tilde{b}_{10} + b_{13})\varepsilon_k + \eta_k + \sqrt{2} \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} \right)$$

e, substituindo em (2.171) temos que

$$\|x_k - x_*\| \leq \alpha_1 \varepsilon_k + \alpha_2 \eta_k + \alpha_3 \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i}$$

para as constantes $\alpha_1 = 1 + 4M(b_{14} + \tilde{b}_{10} + b_{13})$, $\alpha_2 = 4M$ e $\alpha_3 = 4M\sqrt{2}$ que é a desigualdade (2.165).

Agora, para obter a desigualdade (2.166), usando (2.173) e (2.174) na desigualdade (2.168) obtemos

$$\left\| \begin{pmatrix} (x_k - x_*)_{[I_F]} \\ \lambda_{k+1} - \lambda_* \end{pmatrix} \right\| \leq M(b_{15}\varepsilon_k + \|h(x_k)\|) + \frac{\Delta x_k}{2M} + b_{13}\varepsilon_k$$

para $b_{15} = b_{13} + b_{14}$.

Logo

$$\Delta x_k = \|(x_k - x_*)_{[I_F]}\| \leq \frac{\Delta x_k}{2} + M(b_{15}\varepsilon_k + \|h(x_k)\|)$$

e portanto

$$\Delta x_k \leq 2M(b_{15}\varepsilon_k + \|h(x_k)\|). \quad (2.175)$$

Por (2.175) e (2.171) obtemos que

$$\|x_k - x_*\| \leq \alpha_6 \varepsilon_k + \alpha_7 \|h(x_k)\| \quad (2.176)$$

para as constantes $\alpha_6 = 1 + 2b_{15}M$ e $\alpha_7 = 2M$.

Substituindo (2.176) em (2.167) obtemos a desigualdade

$$\|\lambda_{k+1} - \lambda_*\| \leq \alpha_4 \varepsilon_k + \alpha_5 \|h(x_k)\|$$

para $\alpha_4 = b_1 + b_2 \alpha_6$ e $\alpha_5 = b_2 \alpha_7$ que é a desigualdade (2.166). $\square$

Lema 2.5.14. *Seja $\{x_k\} \subset \Omega_n$ uma seqüência que converge ao ponto x_* , ponto estacionário e regular do problema (2.96). Seja λ_* o multiplicador de Lagrange associado e suponhamos que satisfaz a condição (2.149) do Lema 2.5.10.*

Sejam $\{\bar{\lambda}_k\}$, $\{\rho_k\}$, $\{\varepsilon_k\}$, $\{\lambda_k\}$ seqüências como no Lema 2.5.12 que satisfazem a desigualdade (2.160).

Então, se $\{\bar{\lambda}_k\}$ é calculado como na Hipótese 2.5.9, existe k_0 tal que para todo $k \geq k_0$

$$\bar{\lambda}_k = \lambda_k.$$

Prova. Como as seqüências $\{x_k\}$, $\{\lambda_k\}$ satisfazem a desigualdade (2.161) do Lema 2.5.12:

$$\|\lambda_{k+1} - \lambda_*\| \leq b_1 \varepsilon_k + b_2 \|x_k - x_*\|$$

a demonstração é a mesma do Lema 2.5.10. $\square$

Teorema 2.5.15. *Seja $\{x_k\}$ a seqüência gerada pelo Algoritmo 2.5.1 utilizando a Regra 2 de adaptação do parâmetro de penalidade quando aplicado ao problema (2.96). Suponhamos que $\{x_k\}$ é uma seqüência infinita e que converge a um ponto limite x_* , que as Hipóteses 2.5.4, 2.5.5, 2.5.9 verificam-se e que o multiplicador de Lagrange λ_* satisfaz (2.149). Suponhamos também que ε_k é escolhido como*

$$\varepsilon_k = \min\{\varepsilon_{k-1}, \|h(x_k)\|_\infty, \varepsilon'_k\} \quad (2.177)$$

para $\{\varepsilon'_k\}$ uma seqüência decrescente que converge para 0, para $\varepsilon_0 > 0$. Então a seqüência de parâmetros de penalidade $\{\rho_k\}$ é limitada.

Prova. A seqüência $\{x_k\}$ é gerada satisfazendo a condição (2.79), logo, por (2.153) temos que

$$\|\mathcal{P}_{\Omega_n}[x_k - \nabla f(x_k) - \sum_{i=1}^m [\lambda_{k+1}]_i \nabla h_i(x_k)] - x_k\| \leq \beta \varepsilon_k$$

e podemos utilizar os lemas provados anteriormente.

Seja k_0 o inteiro obtido do Lema 2.5.14. Então, para todo $k \geq k_0$ temos que $\bar{\lambda}_k = \lambda_k$.

Pelos Lemas 2.5.12 e 2.5.13 temos que existem constantes positivas $b_1, b_2, \bar{\rho}, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ e um inteiro k_1 tais que se verifica a desigualdade (2.161) do Lema 2.5.12:

$$\|\lambda_{k+1} - \lambda_*\| \leq b_1 \varepsilon_k + b_2 \|x_k - x_*\|, \quad (2.178)$$

e, se $[\rho_{k_1}]_i \geq \bar{\rho}$ para todo $i \in I_\infty$, tem-se as desigualdades (2.165) e (2.166) do Lema 2.5.13:

$$\|x_k - x_*\| \leq \alpha_1 \varepsilon_k + \alpha_2 \eta_k + \alpha_3 \sum_{i \in I_\infty} \frac{|[\bar{\lambda}_k]_i - [\lambda_*]_i|}{[\rho_k]_i} \quad (2.179)$$

$$\|\lambda_{k+1} - \lambda_*\| \leq \alpha_4 \varepsilon_k + \alpha_5 \|h(x_k)\| \quad (2.180)$$

para todo $k \geq k_1$.

Suponhamos que $I_\infty \neq \emptyset$. Seja $i_0 \in I_\infty$.

Para cada $i \in I_a$ existe $k_1(i)$ tal que $\forall k \geq k_1(i)$, $[\rho_{k+1}]_i = [\rho_k]_i$. Seja k suficientemente grande tal que para todo $i \in I_a$

$$|h_i(x_k)| \leq \tau \|h(x_{k-1})\|_\infty.$$

Então,

$$\eta_k = \sum_{i \in I_a} |h_i(x_k)| \leq |I_a| \tau \|h(x_{k-1})\|_\infty. \quad (2.181)$$

Seja $k \geq \tilde{k} = \max_{i \in I_a} \{k_0, k_1, k_1(i)\}$.

Usando a fórmula de adaptação do multiplicador (2.80) temos que

$$|h_{i_0}(x_k)| = \frac{|[\lambda_{k+1}]_{i_0} - [\lambda_k]_{i_0}|}{[\rho_k]_{i_0}} \leq \frac{|[\lambda_{k+1}]_{i_0} - [\lambda_*]_{i_0}| + |[\lambda_k]_{i_0} - [\lambda_*]_{i_0}|}{[\rho_k]_{i_0}}$$

e, usando a desigualdade (2.178) obtemos

$$|h_{i_0}(x_k)| \leq \frac{b_1 \varepsilon_k + b_2 \|x_k - x_*\| + |[\lambda_k]_{i_0} - [\lambda_*]_{i_0}|}{[\rho_k]_{i_0}}.$$

Logo, por (2.179) obtemos

$$\begin{aligned} |h_{i_0}(x_k)| &\leq \frac{1}{[\rho_k]_{i_0}} \left[(b_1 + b_2 \alpha_1) \varepsilon_k + b_2 \alpha_2 \eta_k + \right. \\ &\quad \left. b_2 \alpha_3 \sum_{i \in I_\infty} \frac{|[\lambda_k]_i - [\lambda_*]_i|}{[\rho_k]_i} + |[\lambda_k]_{i_0} - [\lambda_*]_{i_0}| \right]. \end{aligned} \quad (2.182)$$

Agora, usando (2.180) para λ_k no lugar de λ_{k+1} , (2.177) e (2.153) obtemos que

$$|[\lambda_k]_i - [\lambda_*]_i| \leq \|\lambda_k - \lambda_*\| \leq (\alpha_4 + \alpha_5\beta)\|h(x_{k-1})\|_\infty \quad i = 1, \dots, m. \quad (2.183)$$

Como $\varepsilon_k \leq \varepsilon_{k-1} \leq \|h(x_{k-1})\|_\infty$, combinando (2.181)-(2.183) temos que

$$|h_{i_0}(x_k)| \leq m_k(i_0)\|h(x_{k-1})\|_\infty$$

para

$$m_k(i_0) = \frac{1}{[\rho_k]_{i_0}} \left[b_1 + b_2\alpha_1 + b_2\alpha_2|I_a|\tau + b_2\alpha_3 \sum_{i \in I_\infty} \frac{(\alpha_4 + \alpha_5\beta)}{[\rho_k]_i} + \alpha_4 + \alpha_5\beta \right].$$

Como $m_k(i_0) \rightarrow 0$ quando k cresce, existe $\tilde{k}(i_0) \geq \tilde{k}$ tal que

$$|h_{i_0}(x_k)| \leq \tau\|h(x_{k-1})\|_\infty$$

para todo $k \geq \tilde{k}(i_0)$ e portanto $[\rho_{k+1}]_{i_0} = [\rho_k]_{i_0}$, o que produz uma contradição. $\square$

Observação. Como no Teorema 2.5.11, a eleição de ε_k é adaptativa. Em outras palavras, a precisão necessária para resolver cada subproblema depende do nível de inviabilidade na solução aproximada. Assim, o critério de parada de cada subproblema é

$$\|\mathcal{P}_{\Omega_2}[x_k - \nabla L(x_k, \bar{\lambda}_k, \rho_k)] - x_k\|_\infty \leq \min\{\varepsilon_{k-1}, \|h(x_k)\|_\infty, \varepsilon'_k\}$$

para ε_{k-1} definido em (2.177).

Capítulo 3

Experimentos numéricos

No capítulo anterior definimos um algoritmo de Lagrangiano aumentado para resolver problemas gerais de otimização não-linear que é globalmente convergente utilizando a condição de qualificação CPLD recentemente introduzida na literatura. Assim, do ponto de vista teórico conseguimos definir um algoritmo que elimina o uso de condições de qualificação fortes. Neste capítulo analisaremos o desempenho do algoritmo do ponto de vista prático. Ver [2].

Nesse respeito apresentaremos dois conjuntos de experimentos numéricos. Por um lado avaliamos a confiabilidade do nosso Algoritmo 2.1.1 comparando-o com LANCELOT [15] para o caso em que o conjunto simples é uma caixa (restrições de canalização) e utilizamos todos os problemas da coleção CUTE [13] como base da comparação.

Por outro lado, as aplicações mais interessantes do Algoritmo 2.1.1 surgem quando o conjunto simples é arbitrário, talvez não convexo ou definido por restrições não lineares. A família de problemas de alocação se apresenta como uma situação deste tipo. Existe um algoritmo específico (SPG) que resulta ser muito eficiente para resolver esta classe de problemas mas que não pode ser utilizado quando restrições adicionais são incorporadas. Afortunadamente, o método de Lagrangiano aumentado proposto é capaz de tratar com restrições adicionais e diferentes, aproveitando a eficiência de SPG para resolver os subproblemas.

Todos os experimentos numéricos apresentados nesta seção foram realizados pelo Prof. Dr. Ernesto Birgin da Universidade de São Paulo e podem ser encontrados no artigo [2].

3.1 Restrições de canalização no conjunto simples

Para analisar o desempenho do Algoritmo 2.1.1 para o caso em que o conjunto simples Ω_2 é uma caixa comparamos nosso método com LANCELOT [15].

Suponhamos que $\Omega_2 = [l, u], l < u$. Assim, em cada iteração do método é resolvido um

problema com restrições de canalização. Nossa implementação utiliza o método GENCAN [9] para resolver estes subproblemas. GENCAN é um algoritmo de restrições ativas para resolver problemas de otimização de grande porte com restrições de canalização que, dentro das faces utiliza um método de Newton truncado com busca linear e, para abandonar as faces, utiliza iterações do tipo gradiente projetado espectral como as definidas em [10, 11]. Assim, o método que resulta será denominado ALGENCAN (Augmented Lagrangian utilizando GENCAN).

Nos experimentos testamos duas versões de ALGENCAN:

- um único parâmetro para penalizar todas as restrições ao mesmo tempo como sugerido na Regra 1;
- um parâmetro diferente para penalizar cada restrição do conjunto complexo como sugerido na Regra 2.

Consideramos também duas versões de LANCELOT:

- a versão original com todas as opções *default*;
- uma versão utilizando Hessianos verdadeiros e Gradientes Conjugados.

Esta última versão do LANCELOT é mais adequada já que a versão de GENCAN utilizada na comparação não utiliza nenhum tipo de preconditionador.

Para analisar os resultados numéricos consideramos todos os problemas da coleção CUTE [13]; assim testamos um total de 873 problemas.

Em cada passo do algoritmo de Lagrangiano aumentado, GENCAN é utilizado para encontrar um ponto satisfazendo

$$\|P_{\Omega_2}[x_k - \nabla L(x_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k)] - x_k\|_{\infty} \leq \varepsilon_k.$$

Para ambas versões de ALGENCAN e baseados nos experimentos numéricos apresentados em [1, 8], consideramos $\tau = 0.5$, $\gamma = 10$, $\bar{\mu}_{\min} = \bar{\lambda}_{\min} = -10^{20}$, $\bar{\mu}_{\max} = \bar{\lambda}_{\max} = 10^{20}$, $\varepsilon_k = 10^{-4}$ para todo k , e ρ_0 e ρ_1 com todas as componentes iguais a 10. Tanto no cálculo de $\bar{\mu}_0 = \bar{\mu}_1$ e de $\bar{\lambda}_1$, como no cálculo de $\bar{\mu}_k$ e de $\bar{\lambda}_k$ no Passo 3 de ALGENCAN, consideramos as projeções de μ_k e λ_k nos conjuntos $[\bar{\mu}_{\min}, \bar{\mu}_{\max}]^{p_1}$ e $[\bar{\lambda}_{\min}, \bar{\lambda}_{\max}]^{m_1}$, respectivamente. Para ALGENCAN e LANCELOT utilizamos o iterado inicial x_0 proporcionado por CUTE e o vetor nulo como estimativa inicial dos multiplicadores de Lagrange.

Como critério de parada utilizamos

$$\begin{aligned} \|P_{\Omega_2}[x_k - \nabla L(x_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k)] - x_k\|_{\infty} &\leq 10^{-4} \\ \max(\|h_1(x_k)\|_{\infty}, \|\sigma_k\|_{\infty}) &\leq 10^{-4}. \end{aligned} \tag{3.1}$$

Observar que a condição $\|\sigma_k\|_\infty \leq 10^{-4}$ garante que, para todo $i = 1, \dots, p_1$, $[g_1(x_k)]_i \leq 10^{-4}$ e que, quando $[g_1(x_k)]_i < -10^{-4}$ temos que $|\sigma_k|_i = \frac{[\bar{\mu}_k]_i}{[\rho_k]_{m_1+i}} \leq 10^{-4}$ o que implica que $[\mu_{k+1}]_i = \max\{0, [\bar{\mu}_k]_i + [\rho_k]_{m_1+i}[g_1(x_k)]_i\} = 0$. Logo, a condição $\|\sigma_k\|_\infty \leq 10^{-4}$ diz que se verificam, aproximadamente, a viabilidade e a complementaridade no ponto final. Ou seja a condição (3.1) estabelece que paramos a execução do algoritmo quando um iterado satisfaz aproximadamente as condições KKT para o problema original:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeito a } h_1(x) = 0, g_1(x) \leq 0, \\ & x \in [l, u]. \end{aligned}$$

Paramos também a execução do algoritmo quando o tempo de CPU excede 10 minutos.

Todos os experimentos foram feitos num Processador Sun Fire 880 com 8 900 Mhz UltraSPARC III, 32 Gb de memoria RAM, running SunOS 5.8. Os códigos foram programados em Fortran 77 e compilados com Forte Developer 7 Fortran 95 7.0 2002/03/09. Utilizamos a opção -O3 para otimizar os códigos.

Comparamos quatro métodos: duas versões de LANCELOT e duas versões de ALGENCAN. Dado um problema fixo, para cada método M , denotamos por x_{final}^M ao ponto final obtido quando o método M é aplicado para resolver o problema dado. Dizemos que x_{final}^M é viável se

$$\max\{\|h_1(x_{\text{final}}^M)\|_\infty, \|g_1(x_{\text{final}}^M)_+\|_\infty\} \leq 10^{-4}.$$

Observar que, pela estrutura do conjunto simples e a forma pela qual os subproblemas são resolvidos, as restrições de canalização são sempre satisfeitas.

Definimos

$$f_{\text{best}} = \min_M \{f(x_{\text{final}}^M) \mid x_{\text{final}}^M \text{ é viável}\}.$$

Dizemos que o método M encontrou a solução do problema se x_{final}^M é viável e

$$f(x_{\text{final}}^M) \leq f_{\text{best}} + 10^{-3}|f_{\text{best}}| + 10^{-6}.$$

Finalmente, seja t^M o tempo de CPU que o método M utiliza para encontrar x_{final}^M . Definimos

$$t_{\text{best}} = \min_M \{t^M \mid \text{o método } M \text{ encontra a solução}\}$$

e dizemos que o método M é *um dos métodos mais rápidos* para o problema se

$$t^M \leq t_{\text{best}} + 0.01t_{\text{best}}.$$

Estas definições são as mesmas utilizadas em [8] para comparar diferentes funções

de Lagrangiano aumentado. Estamos interessados em comprar os diferentes algoritmos de Lagrangiano aumentados com respeito a Robustez e Eficiência. Dizemos que um algoritmo particular é *robusto* para resolver um problema se *encontra a solução* de acordo com o critério definido anteriormente. Dizemos que o algoritmo é *eficiente* se é um dos métodos mais rápidos que resolve o problema.

Os perfis de desempenho para comparar ALGENCAN e LANCELOT utilizam o tempo r como medida de desempenho, [17]. Ver Figura 3.1. Em termos de perfis de desempenho [17] eficiência e robustez são, essencialmente, os valores da função do perfil em 1 e ∞ , respectivamente. Da figura observamos que ALGENCAN com um único parâmetro de penalidade é mais eficiente que os outros três algoritmos de Lagrangiano aumentado. Com respeito a robustez (lado direito da função do perfil de desempenho), vemos que esta versão de ALGENCAN encontrou a solução em 731 (83.73%) problemas, enquanto a versão ALGENCAN que considera vários parâmetros de penalidade o fez em 707 (80.99%) problemas, a versão original de LANCELOT resolveu 712 (81.56%) problemas e a versão de LANCELOT que utiliza Hessianos verdadeiros e Gradientes Conjugados resolveu 669 (76.63%) problemas.

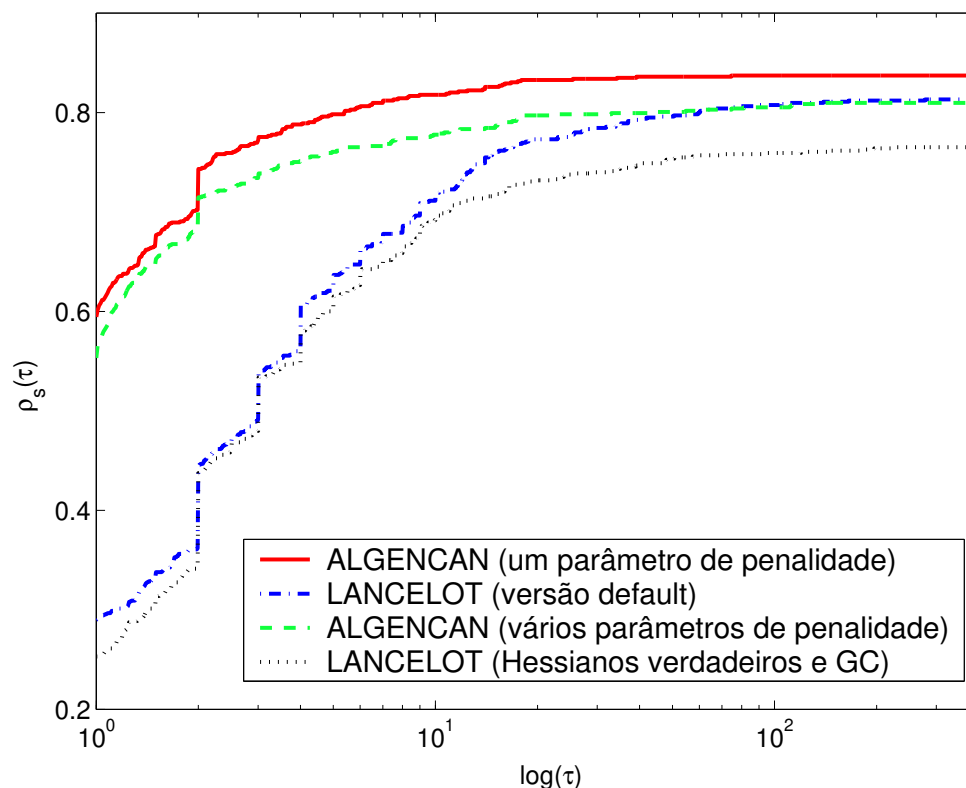


Figura 3.1: Comparação entre ALGENCAN e LANCELOT utilizando perfil de desempenho.

3.2 Restrições gerais no conjunto simples: o problema de alocação

Consideramos agora uma variante da família de *problemas de alocação* introduzidos em [11]. Sejam fixos n_p polígonos disjuntos em $\mathbb{R}^2$, $P_1, P_2, \dots, P_{n_p}$. O problema de alocação consiste em encontrar um ponto $z^1 \in P_1$ que minimize a soma das distâncias aos outros polígonos. Assim, uma formulação matemática do problema é:

$$\begin{aligned} \min_{z^i, i=1, \dots, n_p} \quad & \sum_{i=2}^{n_p} \|z^i - z^1\|_2 \\ \text{sujeito a} \quad & z^i \in P_i, \quad i = 1, \dots, n_p. \end{aligned}$$

Este problema é resolvido eficientemente utilizando o método do gradiente projetado espectral, SPG, conforme [11]. SPG é um algoritmo de gradiente projetado espectral não-monótono para resolver problemas de otimização convexa de grande porte. SPG somente pode ser aplicado para resolver problemas com restrições convexas.

Neste trabalho consideramos uma variante do problema de alocação descrito acima que torna o método SPG inaplicável. Além dos n_p polígonos, consideramos n_c círculos e uma elipse com intersecção não vazia com o primeiro polígono P_1 . Assim, o problema consiste em encontrar $z^1 \in P_1$ que minimize a soma das distâncias para os outros polígonos e círculos de modo que z^1 seja o único ponto que se encontra no interior ou na borda da elipse. Logo, se $a, b, c \in \mathbb{R}$ são constantes positivas e $g(x) = (x_1/a)^2 + (x_2/b)^2 - c$ é a elipse, o problema que desejamos resolver pode ser formulado da seguinte maneira:

$$\begin{aligned} \min_{z^i, i=1, \dots, n_p+n_c} \quad & \sum_{i=2}^{n_p} \|z^i - z^1\|_2 + \sum_{i=1}^{n_c} \|z^{n_p+i} - z^1\|_2 \\ \text{sujeito a} \quad & g(z^1) \leq 0, \\ & g(z^i) \geq 0, \quad i = 2, \dots, n_p + n_c, \\ & z^i \in P_i, \quad i = 1, \dots, n_p, \\ & z^{n_p+i} \in C_i, \quad i = 1, \dots, n_c. \end{aligned}$$

Observar que devido à introdução da restrição não convexa $g(z^i) \geq 0, i = 2, \dots, n_p + n_c$ é impossível utilizar o método SPG.

Para resolver os problemas, consideramos como conjunto complexo:

$$\Omega_1 = \{g(z^1) \leq 0, g(z^i) \geq 0, i = 2, \dots, n_p + n_c\},$$

e como conjunto simples:

$$\Omega_2 = \{z^i \in P_i, i = 1, \dots, n_p, \text{ e } z^{n_p+i} \in C_i, i = 1, \dots, n_c\}.$$

Assim, são penalizadas aquelas restrições adicionais que fazem que o método SPG não possa ser aplicado. Desta maneira implementamos um método de Lagrangiano aumentado que utiliza SPG [10, 11] para resolver os subproblemas eficientemente e aproveitando a estrutura do conjunto simples. Esta implementação é denominada ALSPG. Em geral, seria interessante aplicar ALSPG a qualquer problema no qual o conjunto escolhido como simples seja um conjunto convexo para o qual é fácil (não muito caro) calcular a projeção de pontos arbitrários.

Foram gerados 18 problemas variando o número de círculos n_c e de polígonos n_p e escolhendo aleatoriamente a localização dos círculos e dos polígonos e o número de vértices de cada polígono. Na tabela 3.1 exibimos as principais características de cada problema sendo:

- n_c o número de círculos,
- n_p o número de polígonos,
- *totnvs* o número total de vértices de todos os polígonos,
- n a dimensão do problema,
- p_1 o número de restrições de desigualdade do conjunto complexo,
- p_2 o número de restrições de desigualdade do conjunto simples.

Na figura 3.2 ilustramos a solução de um problema pequeno: 24 variáveis, 81 restrições no nível simples e 12 restrições no nível complexo.

Na tabela 3.2 apresentamos os resultados obtidos na aplicação de ALSPG aos problemas da tabela 3.1. Utilizamos $\epsilon = 10^{-4}$ como tolerância tanto para viabilidade quanto para otimalidade. Na tabela 3.2, *OuIt* indica o número de iterações externas (de ALSPG) e *InIt* o número de iterações internas (de SPG); *Fcnt* e *Gcnt* indicam o número de avaliações da função Lagrangiano aumentado e do gradiente, respectivamente; *Time* indica o tempo (em segundos) e finalmente f o valor da função objetivo do problema original. ALSPG satisfaz o critério de convergência em todos os problemas.

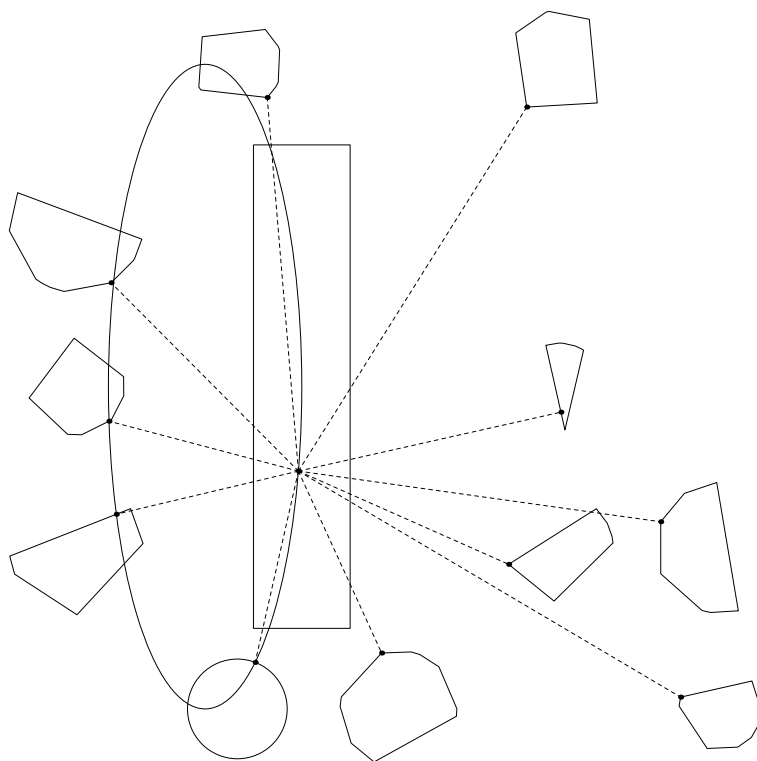


Figura 3.2: Ilustração para um problema de alocação de 12 conjuntos.

Problema	n_c	n_p	$totnvs$	n	p_1	p_2
1	2.929	4.935	61.755	15.728	7.864	64.684
2	4.403	7.346	91.896	23.498	11.749	96.299
3	6.370	10.732	13.165	34.204	17.102	140.535
4	13.344	22.158	27.927	71.004	35.502	290.271
5	19.924	33.265	415.691	106.378	53.189	435.615
6	29.998	49.754	621.824	159.504	79.752	651.822
7	26.288	43.775	546.943	140.126	70.063	573.231
8	39.459	65.891	823.823	210.700	105.350	863.282
9	59.494	98.875	1.236.031	316.738	158.369	1.295.525
10	65.146	109.360	1.367.104	349.012	174.506	1.432.250
11	98.138	163.854	2.048.239	523.984	261.992	2.146.377
12	147.511	245.983	3.074.109	786.988	393.494	3.221.620
13	130.344	218.296	2.729.187	697.280	348.640	2.859.531
14	195.725	327.418	4.093.497	1.046.286	523.143	4.289.222
15	293.913	490.522	6.130.282	1.568.870	784.435	6.424.195
16	260.705	436.037	5.449.661	1.393.484	696.742	5.710.366
17	391.023	653.802	8.172.819	2.089.650	1.044.825	8.563.842
18	587.182	980.615	12.257.368	3.135.594	1.567.797	12.844.550

Tabela 3.1: Problemas de alocação e suas principais características. A geração do problema se baseia numa partição. O número de círculos (n_c) e de polígonos (n_p) dependem do número de pontos na malha. O número de vértices dos polígonos é um número aleatório e o número total de vértices de todos os polígonos é $totnvs$. Finalmente, o número de variáveis do problema é $n = 2(n_c + n_p)$, o número de restrições de desigualdade do conjunto complexo é $p_1 = n_c + n_p$ e o número de restrições de desigualdade do conjunto simples é $p_2 = n_c + totnvs$.

Problema	OuIt	InIt	Fcnt	Gcnt	Time (seg)	f
1	9	334	488	343	12,28	4,5601D+02
2	9	677	949	686	41,30	5,6228D+02
3	11	245	390	256	23,63	6,8965D+02
4	7	69	120	76	18,11	4,6149D+02
5	9	283	476	292	111,45	5,6461D+02
6	10	88	170	98	58,68	6,9408D+02
7	7	71	122	78	40,29	4,6326D+02
8	7	65	121	72	58,09	5,6419D+02
9	8	105	165	113	136,42	6,9382D+02
10	8	71	135	79	111,41	4,6272D+02
11	7	66	121	73	158,23	5,6425D+02
12	9	87	163	96	308,75	6,9381D+02
13	7	55	98	62	185,13	4,6280D+02
14	7	91	151	98	421,58	5,6440D+02
15	10	83	166	93	602,97	6,9384D+02
16	8	68	138	76	458,96	4,6271D+02
17	7	80	140	87	765,60	5,6433D+02
18	10	85	171	95	1234,53	6,9408D+02

Tabela 3.2: Desempenho de ALSPG na resolução do problemas de alocação. O principal inconveniente na aplicação de ALSPG na resolução de problemas de dimensão maior à do problema 18 (aproximadamente 3×10^6 variáveis, 1.5×10^6 restrições de desigualdade do conjunto complexo, e 1.2×10^7 restrições de desigualdade do conjunto simples) é a limitação de memória.

Conclusões

Nesta tese apresentamos um método de Lagrangiano aumentado para resolver problemas gerais de otimização com convergência utilizando a condição de dependência linear positiva constante.

Muitas vezes acontece que um algoritmo particular é extremamente eficiente para tratar com um tipo de restrição, mas falha quando uma classe diferente de restrições é incorporada. No método de Lagrangiano aumentado, restrições adicionais são naturalmente incorporadas à função objetivo do subproblema, o que preserva a estrutura das restrições que não são penalizadas. Por este motivo acreditamos que a técnica de Lagrangiano aumentado continuará sendo muito utilizada.

Este fato nos motivou a analisar métodos de Lagrangiano aumentado para resolver problemas com restrições gerais, especificamente com dois conjuntos de restrições, um deles simples e o outro complexo. Do ponto de vista teórico, nosso maior objetivo foi eliminar, tanto quanto possível, condições de qualificação restritivas. Com isto como objetivo utilizamos, tanto na prova de viabilidade quanto na prova de otimalidade, a condição de qualificação de dependência linear positiva constante (CPLD) introduzida por Qi e Wei em [32]. A condição CPLD foi definida e utilizada em [32] para analisar a convergência de métodos de programação quadrática seqüencial. Esta condição é razoavelmente geral no sentido de que ela é mais fraca que as clássicas condições de Mangasarian-Fromovitz [27], posto constante [24] e regularidade [4], além de ser satisfeita por restrições lineares e restrições repetidas. No trabalho onde a CPLD foi introduzida, os autores conjecturaram que ela podia ser uma condição de qualificação. Nesta tese provamos que ela é, com efeito, uma condição de qualificação, e mostramos suas relações com outras conhecidas condições de qualificação associadas com problemas de programação não convexos.

O resultado de convergência global do método proposto estabelece que se um ponto limite viável satisfaz a condição CPLD então ele é um ponto estacionário do problema original. Este resultado é mais forte que resultados prévios obtidos para problemas mais específicos. Em particular, Conn, Gould, Sartenauer e Toint provam em [14] convergência global de um método de Lagrangiano aumentado para resolver problemas com restrições de igualdade e lineares utilizando a regularidade, condição mais forte que a CPLD. Provamos

aqui que uma condição de qualificação melhor pode ser utilizada com o mesmo propósito. O problema de descobrir condições de qualificação ainda mais fracas que sejam úteis nas provas de convergência global de métodos de otimização está ainda em aberto.

Como o método proposto é um método de Lagrangiano aumentado e não um de penalidade quadrática, é interessante observar o comportamento da sequência de parâmetros de penalidade. Analisamos duas versões do algoritmo principal: com um único parâmetro para penalizar todas as restrições do conjunto Ω_1 ao mesmo tempo e com um parâmetro diferente para cada restrição. No primeiro caso provamos, sob hipóteses razoáveis, limitação da sequência de parâmetros de penalidade. Utilizamos na prova a redução ao caso com restrições de igualdade introduzida em [4]. As hipóteses utilizadas incluem, entre outras, a regularidade, a condição suficiente de segunda ordem e a condição de complementaridade estrita na solução, [2]. Apesar de serem estas as condições habituais consideradas na literatura, acreditamos que o mesmo resultado pode ser obtido utilizando condições mais fracas, ou que nos permitam eliminar a hipótese restritiva de complementaridade estrita na solução.

No caso em que utilizamos um parâmetro diferente por restrição, se o conjunto Ω_1 é formado somente por restrições de igualdade e Ω_2 por restrições de canalização, provamos limitação da sequência de parâmetros de penalidade sem utilizar a condição de complementaridade estrita na solução, [1]. A questão de se é possível obter limitação dos parâmetros de penalidade quando há vários deles está ainda em aberto para o caso de Ω_1, Ω_2 mais gerais.

Do ponto de vista prático testamos a confiabilidade do método de duas maneiras diferentes: por um lado e no caso em que o conjunto Ω_2 é uma caixa, mediante uma exaustiva comparação com o algoritmo LANCELOT [15] utilizando todos os problemas da coleção CUTE [13]; e por outro lado mediante a resolução dos problemas de alocação.

No primeiro caso, nossa implementação foi baseada no algoritmo robusto GENCAN para minimização com restrições de canalização, [9]. Mostramos que nosso algoritmo é tão robusto e eficiente quanto o LANCELOT para este conjunto de problemas.

No segundo caso, as aplicações mais interessantes do nosso algoritmo surgem quando o conjunto Ω_2 é arbitrário, talvez não convexo ou definido por restrições não-lineares. A família de problemas de alocação se apresenta como uma situação deste tipo. Existe um algoritmo específico (SPG) [11] que resulta ser muito eficiente para resolver esta classe de problemas mas que não pode ser utilizado quando restrições adicionais são incorporadas. Felizmente, o algoritmo de Lagrangiano aumentado proposto é capaz de tratar com restrições adicionais e diferentes aproveitando a eficiência de SPG para resolver os sub-problemas. Assim, foi possível resolver problemas de programação não-linear com mais de 3.000.000 variáveis e 14.000.000 restrições em menos de meia hora de tempo CPU.

Referências Bibliográficas

- [1] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. Augmented Lagrangian Methods under the Constant Positive Linear Dependence Constraint Qualification. Por aparecer em *Mathematical Programming*.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. On Augmented Lagrangian methods with general lower-level constraints, Technical Report MCDO-050303, Department of Applied Mathematics, UNICAMP, Brazil, (2005).
- [3] R. Andreani, J. M. Martínez e M. L. Schuverdt. On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification, *Journal of Optimization Theory and Applications* 125, pp. 473–485 (2005).
- [4] D. P. Bertsekas. *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.
- [5] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982; republished by Athena Scientific, 1996.
- [6] D. P. Bertsekas e A. E. Ozdaglar. Pseudonormality and a Lagrange multiplier theory for constrained optimization, *Journal of Optimization Theory and Applications* 114, pp. 287–343 (2002).
- [7] D. P. Bertsekas e A. E. Ozdaglar. The relation between pseudonormality and quasiregularity in constrained optimization, *Optimization Methods and Software*, Vol. 19, pp. 493–506 (2004).
- [8] E. G. Birgin, R. Castillo e J. M. Martínez. Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems, *Computational Optimization and Applications* 31, pp. 31–56 (2005).
- [9] E. G. Birgin e J. M. Martínez. Large-scale active-set box-constrained optimization method with spectral projected gradients, *Computational Optimization and Applications* 23, pp. 101–125 (2002).

- [10] E. G. Birgin, J. M. Martínez e M. Raydan. Nonmonotone spectral projected gradient methods on convex sets, *SIAM Journal on Optimization* 10, pp. 1196–1211 (2000).
- [11] E. G. Birgin, J. M. Martínez e M. Raydan. Algorithm 813: SPG - Software for convex-constrained optimization, *ACM Transactions on Mathematical Software* 27, pp. 340–349 (2001).
- [12] E. G. Birgin, J. M. Martínez e M. Raydan. Inexact Spectral Projected Gradient methods on convex sets, *IMA Journal on Numerical Analysis* 23, pp. 539–559 (2003).
- [13] I. Bongartz, A. R. Conn, N. I. M. Gould e Ph. L. Toint. CUTE: constrained and unconstrained testing environment, *ACM Transactions on Mathematical Software* 21, pp. 123–160 (1995).
- [14] A. R. Conn, N. I. M. Gould, A. Sartenaer e Ph. L. Toint. Convergence properties of an Augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints, *SIAM Journal on Optimization* 6, pp. 674–703 (1996).
- [15] A. R. Conn, N. I. M. Gould e Ph. L. Toint. *LANCELOT: A Fortran package for large scale nonlinear optimization*. Springer-Verlag, Berlin, 1992.
- [16] A. R. Conn, N. I. M. Gould e Ph. L. Toint. A globally convergent Augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM Journal on Numerical Analysis* 28, pp. 545–572 (1991).
- [17] E. D. Dolan e J. J. Moré. Benchmarking optimization software with performance profiles, *Mathematical Programming* 91, pp. 201–213 (2002).
- [18] Z. Dostál. Inexact semi-monotonic Augmented Lagrangians with optimal feasibility convergence for convex bound and equality constrained quadratic programming, *SIAM Journal on Numerical Analysis* 43, pp 96–115 (2005).
- [19] Z. Dostál, A. Friedlander e S. A. Santos. Augmented Lagrangian with adaptive precision control for quadratic programming with simple bounds and equality constraints, *SIAM Journal on Optimization* 13, pp. 1120–1140 (2003).
- [20] R. Fletcher. *Practical Methods of Optimization*, Academic Press, London, 1987.
- [21] W. W. Hager. Analysis and implementation of a dual algorithm for constrained optimization, *Journal of Optimization Theory and Applications* 79, pp. 37–71 (1993).
- [22] M. R. Hestenes. *Multiplier and gradient methods*, Journal of Optimization Theory and Applications 4, pp. 303–320 (1969).

- [23] M. R. Hestenes. *Optimization Theory, The finite dimensional case*, John Wiley, New York, 1975.
- [24] R. Janin. Direction derivative of the marginal function in nonlinear programming, *Mathematical Programming Study* 21, pp. 127–138 (1984).
- [25] F. John. Extremum problems with inequalities as subsidiary conditions, *Studies and Essays: Courant Anniversary Volume*, Edited by K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, Willey-Interscience, New York, pp. 187–204 (1984).
- [26] D. G. Luenberger. *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, London, 1973.
- [27] O. L. Mangasarian e S. Fromovitz. The Fritz-John necessary optimality conditions in presence of equality and inequality constraints, *Journal of Mathematical Analysis and Applications* 17, pp. 37–47 (1967).
- [28] J. M. Martínez. *Condições de otimalidade em programação não-linear*, Minicurso ditado na Primeira Escola Latinoamericana de Otimização, 4 a 8 de julho de 2005, Bahía Blanca, Argentina.
- [29] J. M. Martínez e B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming, *Journal of Optimization Theory and Applications* 118, pp. 117–133 (2003).
- [30] E. J. McShane. The Lagrange Multiplier Rule. *American Mathematical Monthly* 80, pp. 922–925 (1973).
- [31] M. J. D. Powell. A method for nonlinear constraints in minimization problems, *Optimization*, R. Fletcher (ed.), Academic Press, New York, NY, pp. 283–298, (1969).
- [32] L. Qi e Z. Wei. On the constant positive linear dependence condition and its application to SQP methods, *SIAM Journal on Optimization* 10, pp. 963–981 (2000).
- [33] R. T. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming, *Journal of Optimization Theory and Applications* 12, pp. 555–562 (1973).
- [34] R. T. Rockafellar. Lagrange multipliers and optimality, *SIAM Review* 35, pp. 183–238 (1993).