

Universidade Estadual de Campinas

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA

Departamento de Matemática Aplicada

Tese de Doutorado

**Método Lagrangiano Aumentado
Regularizado para problemas
com Voracidade.**

por

André Luís Machado Martinez [†]

Doutorado em Matemática Aplicada - Campinas - SP

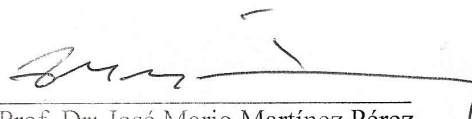
Orientador: Prof. Dr. José Mario Martínez Pérez

[†]Este trabalho contou com apoio financeiro da CAPES.

Método Lagrangiano Aumentado Regularizado para problemas com Voracidade

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **André Luís Machado Martínez** e aprovada pela comissão julgadora.

Campinas, 17 de junho de 2009



Prof. Dr: José Mario Martínez Pérez
Orientador

Banca Examinadora:

prof. Dr. José Mario Martínez Pérez.
prof. Dra. Sandra Augusta Santos.
prof. Dr. Benar Fux Svaiter.
prof. Dr. Alfredo Noel Iusem.
prof. Dr. Nelson Maculan Filho.

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

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

Bibliotecária: Crislene Queiroz Custódio – CRB8 / 7966

Martinez, André Luís Machado

M366m Método langrangiano aumentado regularizado para problemas com voracidade / André Luís Machado Martinez -- Campinas, [S.P. : s.n.], 2009.

Orientador : José Mario Martinez Pérez

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

1. Programação não-linear. 2. Lagrange, Funções de. 3. Otimização matemática. I. Martinez Pérez, José Mario. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

(cqc/imecc)

Título em inglês: Regularized augmented lagrangian method for problems with greediness

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

Área de concentração: Matemática Aplicada

Titulação: Doutor em Matemática Aplicada

Banca examinadora: Prof. Dr. José Mario Martinez Pérez (IMECC-UNICAMP)
Prof. Dra. Sandra Augusta Santos (IMECC-UNICAMP)
Prof. Dr. Benar Fux Svaiter (IMPA)
Prof. Dr. Alfredo Noel Iusem (IMPA)
Prof. Dr. Nelson Maculan Filho (UFRJ)

Data da defesa: 20/05/2009

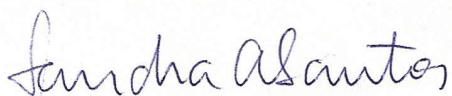
Programa de Pós-Graduação: Doutorado em Matemática Aplicada

Tese de Doutorado defendida em 20 de maio de 2009 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



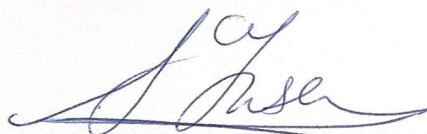
Prof(a). Dr(a). JOSÉ MARIO MARTÍNEZ PÉREZ



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



Prof(a). Dr(a). BENAR FUX SVAITER



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



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

*A minha esposa Cristiane, meus
pais André e Neuza, meu irmão
Ricardo e meus sobrinhos Pedro
e Matheus, dedico.*

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus por tudo. Agradeço as pessoas que mais amo neste mundo: minha menina (Cris), meus pais (André e Neuza), meu irmão (Ricardo), meus melhores amigos (Emerson, Carlos, Evandro, Jair, Wesley, Rodrigo Campiolo), meus tios e primos (Marcos, Silvana, Daniel, Leonice, Geraldo, Ermelinda, Acácio, Nelson, Mateus, Aline, entre outros), e em especial a todos que cursaram comigo o doutorado. A todas estas pessoas agradeço muito pela ajuda e apoio recebido.

Agradeço em especial ao professor José Mario Martínez pela paciência e dedicação ao me orientar no doutorado.

Destaco também a grande ajuda que recebi do Emerson para a elaboração desta tese, valeu monstro!

Agradeço a todos os professores que já tive, desde o primário até o curso de doutorado que de uma forma ou outra contribuíram para que chegasse até aqui.

Agradeço a CAPES pelo suporte financeiro.

RESUMO

Quando resolvemos problemas de programação não linear por meio de algoritmos que utilizam o Lagrangiano Aumentado, um fenômeno chamado voracidade pode ocorrer. Quando isto ocorre o método busca pontos muito infactíveis com valores de função muito pequenos, em geral, nas primeiras iterações, assim o parâmetro de penalidade cresce excessivamente, de tal forma que prejudica o condicionamento do problema. Neste trabalho é sugerida uma abordagem de regularização para superar esta dificuldade. Um método de Lagrangiano Aumentado é definido, com a adição de um termo regularizador que inibe a possibilidade do iterando se afastar demasiadamente do ponto de referência. Provamos convergência e apresentamos exemplos numéricos.

ABSTRACT

When one solves Nonlinear Programming problems by means of algorithms that use merit criteria combining the objective function and penalty feasibility terms, a phenomenon called greediness may occur. Unconstrained minimizers attract the iterates at early stages of the calculations and, so, the penalty parameter needs to grow excessively, in such a way that ill-conditioning harms the overall convergence. In this work a regularization approach is suggested to overcome this difficulty. An Augmented Lagrangian method is defined with the addition of a regularization term that inhibits the possibility that the iterates go far from a reference point. Convergence proofs and numerical examples are given.

SUMÁRIO

Agradecimentos	v
Resumo	vi
Abstract	vii
 Notações básicas	 x
 Introdução	 1
 1 O Método de Lagrangiano aumentado	 4
1.1 Método Lagrangiano aumentado	5
1.1.1 A função Lagrangiano aumentado	7
1.1.2 O algoritmo do método Lagrangiano aumentado	8
1.2 Convergência a pontos admissíveis	11
1.3 Convergência a pontos KKT	17
1.4 Limitação do parâmetro de penalidade	22
1.4.1 Restrições de igualdade	22
1.4.2 Restrições de igualdade e desigualdade	30

1.5	Minimizadores globais	34
1.5.1	Convergência global	36
2	Lagrangiano aumentado regularizado	41
2.1	Voracidade	41
2.2	O método Lagrangiano aumentado regularizado	44
2.3	Algoritmo Lagrangiano aumentado regularizado	45
2.4	Convergência	51
2.5	Limitação do parâmetro de penalidade	54
2.5.1	Restrições de igualdade	54
2.5.2	Restrições de igualdade e desigualdade	57
2.6	Minimizadores globais	58
2.6.1	Convergência global	61
3	Resultados numéricos	66
3.1	Algencan	66
3.1.1	Subproblema	67
3.1.2	Parâmetro de penalidade inicial e multiplicadores iniciais	67
3.1.3	Critérios de parada	68
3.2	RAlgencan	69
3.3	RAlgencan*	75
4	Conclusões	85
	Bibliografia	87

Notações básicas

$\|\cdot\|$ será sempre a norma Euclidiana, mas pode ser substituída por uma norma arbitrária.

Se $v \in \mathbb{R}^n$, denotamos v_+ o vetor cujas componetes são $\max\{0, v_1\}, \dots, \max\{0, v_n\}$.

$\mathbb{R}$ denota o conjunto dos números reais.

$\mathbb{N}$ é o conjunto dos números naturais, começando com 0.

$K \subset_{\infty} \mathbb{N}$ indica que K é um subconjunto infinito de $\mathbb{N}$, preservando a ordem.

Introdução

Neste trabalho trataremos de problemas de otimização que podem ser formulados da seguinte maneira:

$$\text{Minimizar } f(x) \text{ sujeito a } h(x) = 0, g(x) \leq 0, x \in \Omega, \quad (1)$$

onde Ω é definido como:

$$\Omega = \{x \in \mathbb{R}^n | \underline{h}(x) = 0, \underline{g}(x) \leq 0\}, \quad (2)$$

com

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^n \rightarrow \mathbb{R}^m, g : \mathbb{R}^n \rightarrow \mathbb{R}^p, \underline{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ e } \underline{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p. \quad (3)$$

Assumimos que as funções definidas em (3) possuem derivadas primeiras contínuas em $\mathbb{R}^n$.

Problemas deste tipo possuem diversas aplicações práticas e existem muitos métodos para resolvê-los. O método de Lagrangiano aumentado é uma versão atual de uma das idéias mais antigas para resolver o problema. Basicamente consiste em eliminar as restrições $h(x) = 0$ e $g(x) \leq 0$, incluindo-as na função objetivo de maneira que o problema assim transformado tenha soluções iguais ou parecidas às do problema original.

Dado um parâmetro de penalidade $\rho > 0$, a função que denominamos Lagrangiano aumentado se define como:

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(x) + \frac{\mu_i}{\rho} \right) \right]^2 \right\}.$$

As variáveis $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}_+^p$ se chamam “multiplicadores”. Para problemas que possuem apenas restrições de igualdade ($h(x) = 0$) a função Lagrangiano aumentado fica simplificada da seguinte maneira:

$$L_\rho(x, \lambda) = f(x) + \frac{\rho}{2} \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2.$$

Quando se inicia o método, os multiplicadores e o parâmetro de penalidade são dados, de modo que os passos fundamentais são:

1. Resolver

$$\min_{x \in \Omega} L_\rho(x, \lambda, \mu)$$

2. Decidir se o ponto obtido no primeiro passo pode ser aceito como solução do problema original, em caso afirmativo parar, caso contrário ir para o próximo passo.
3. De acordo com os resultados dos passos anteriores, atualizar os multiplicadores e o parâmetro de penalidade.
4. Voltar ao primeiro passo.

Os métodos de Lagrangiano Aumentado seguem, quase sempre, o esquema apresentado acima.

Geralmente no método de Lagrangiano aumentado define-se cautelosamente o parâmetro de penalidade para ser pequeno nas primeiras iterações externas, porque parâmetros grandes tendem a produzir subproblemas mal condicionados. Como consequência, as primeiras iterações tendem a privilegiar a otimalidade sobre a viabilidade. Assim o método pode se comportar

de maneira voraz em busca da otimalidade em alguns problemas e acabar divergindo. Esse fenômeno é denominado neste trabalho como voracidade.

Para contornar o problema da voracidade adicionamos o termo regularizador $\gamma \frac{\|x - \bar{x}\|_2^2}{2}$ à função Lagrangiano Aumentado, onde $\bar{x}$ é o ponto de referência, o melhor iterando até o momento em relação à factibilidade-complementariedade.

Então definimos a função Lagrangiano Aumentado Regularizado como:

$$L_{\rho, \gamma, \bar{x}}(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[\max(0, g_i(x) + \frac{\mu_i}{\rho}) \right]^2 \right\} + \gamma \frac{\|x - \bar{x}\|_2^2}{2}.$$

Com o incremento do termo regularizador o subproblema:

$$\min_{x \in \Omega} L_{\rho, \gamma, \bar{x}}(x, \lambda, \mu)$$

fica melhor condicionado.

CAPÍTULO 1

O Método de Lagrangiano aumentado

Neste capítulo apresentamos um estudo sobre o método de Lagrangiano aumentado baseado nos resultados teóricos de [2]. Estaremos considerando o problema de programação não linear definido em (1), (2) e (3).

As restrições de igualdade e desigualdade podem ser definidas de modo que as restrições mais fáceis de serem manipuladas, por exemplo restrições de canalização ou lineares, formem parte do conjunto simples Ω , deixando as restrições mais difíceis de serem tratadas eficientemente definirem as funções g e h , por exemplos as restrições não lineares. Assim dizemos que Ω é um subconjunto simples do $\mathbb{R}^n$ (geralmente Ω é uma caixa).

Para um problema de otimização, a definição de Ω pode ser feita de modo que o problema

$$\begin{array}{ll} \min & F(x) \\ \text{s.a} & x \in \Omega \end{array}$$

seja considerado mais fácil de resolver que o problema padrão, ou que exista algoritmo eficiente para resolver esse problema que não pode ser aplicado, ou deixa de ser eficiente, quando

aplicado ao problema original. Nesses casos, é natural resolver o problema original mediante uma sequência de problemas da forma acima para alguma função adequada F .

O método Lagrangiano aumentado consiste em uma sequência de iterações externas, nas quais um problema com restrições simples é resolvido aproximadamente para obter o iterando atual. Assim obtido este iterando são atualizados os multiplicadores de Lagrange e os parâmetros de penalidade, repetindo-se o processo.

1.1 Método Lagrangiano aumentado

De início consideramos o seguinte problema de programação não linear dado por:

$$\text{Minimizar } f(x) \text{ sujeito a } h(x) = 0, \quad (1.1)$$

onde as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ com $h(x) = (h_1(x), \dots, h_m(x))$ 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 $g_i(x) + z^2 = 0$.

Para o problema apenas com restrições de igualdade, dado um parâmetro “de penalidade” $\rho > 0$ e aproximações $\lambda_i \in \mathbb{R}$ para os multiplicadores de Lagrange, define-se a função Lagrangiano Aumentado por:

$$L_\rho(x, \lambda) = f(x) + \frac{\rho}{2} \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2$$

Método Lagrangiano aumentado

Sejam $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário, $\tau \in (0, 1]$, $\gamma > 1$, $\rho_1 > 0$, $\lambda_1 \in \mathbb{R}^m$. Definimos os passos fundamentais do método Lagrangiano aumentado por:

Passo 1. *Inicialização*

$$k \leftarrow 1$$

Passo 2. *Resolver o subproblema*

Usando x^{k-1} como aproximação inicial,

$$\min (\text{aproximadamente}) L_{\rho_k}(x, \lambda^k), \quad (1.2)$$

obtendo x^k como solução.

Passo 3. *Estimar os multiplicadores*

Definir uma nova estimativa para os multiplicadores de Lagrange, por exemplo a estimativa de primeira ordem:

$$\lambda_i^{k+1} = \lambda_i^k + \rho_k h_i(x^k), \quad \forall i = 1, \dots, m.$$

Passo 4. *Atualizar o parâmetro de penalidade*

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 5 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao passo 2.

Assim o método Lagrangiano aumentado consiste basicamente de três passos: no primeiro, o método procura uma solução aproximada da função Lagrangiano aumentado; no segundo, atualiza os multiplicadores de Lagrange, e no terceiro decide se incrementa ou não o parâmetro de penalidade.

1.1.1 A função Lagrangiano aumentado

Consideremos o seguinte problema de programação não-linear com restrições de igualdade e desigualdade dado por:

$$\text{Minimizar } f(x) \text{ sujeito a } h_i(x) = 0, \ i = 1, \dots, m; \ g_j(x) \leq 0, \ j = 1, \dots, p \quad (1.3)$$

Com a introdução de uma variável de folga z_j associada com cada restrição de desigualdade, convertamos o problema acima em um problema com restrições de igualdade do tipo:

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

Note que, o problema contém apenas restrições de igualdade e conhecemos sua função Lagrangiano aumentado:

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

Claramente, a função L depende das variáveis $x \in \mathbb{R}^n$ e $z \in \mathbb{R}^p$.

No segundo passo do método Lagrangiano Aumentado é resolvido o subproblema de encontrar um minimizador da função Lagrangiano Aumentado:

$$\min_{(x,z)} L_\rho(x, z, \lambda, \mu). \quad (1.4)$$

Note que a função L é quadrática na variável z_j para cada $j = 1, \dots, p$, logo a resolução de (1.4) pode ser feita em cada z_j independentemente de x . A minimização com respeito a z_j é equivalente a

$$\min \left[g_j(x) + z_j^2 + \frac{\mu_j}{\rho} \right]^2. \quad (1.5)$$

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:

$$2 \left[g_j(x) + z_j^2 + \frac{\mu_j}{\rho} \right] 2z_j.$$

Logo,

$$\widehat{z}_j = 0 \text{ ou } \widehat{z}_j^2 = - \left(g_j(x) + \frac{\mu_j}{\rho} \right).$$

Assim, podemos escrever a solução do problema da seguinte maneira:

$$z_j^2 = \max \left\{ 0, - \left(g_j(x) + \frac{\mu_j}{\rho} \right) \right\},$$

e então o z_j^* ótimo é:

$$z_j^* = \sqrt{\max \left\{ 0, - \left(g_j(x) + \frac{\mu_j}{\rho} \right) \right\}}. \quad (1.6)$$

Logo,

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

e

$$\left(g_j(x) + (z_j^*)^2 + \frac{\mu_j}{\rho} \right) = \max \left(0, g_j(x) + \frac{\mu_j}{\rho} \right).$$

Assim, a função Lagrangiano Aumentado fica:

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(x) + \frac{\mu_i}{\rho} \right) \right]^2 \right\}. \quad (1.7)$$

1.1.2 O algoritmo do método Lagrangiano aumentado

A seguir apresentamos o algoritmo do método Lagrangiano aumentado (sem critério de parada) para o problema definido em (1), (2) e (3) .

ALGORITMO 1.1

Seja $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário.

Os parâmetros para a execução do algoritmo são:

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

$$\eta > 1$$

$$-\infty < \lambda_{\min} < \lambda_{\max} < +\infty,$$

$$\mu_{\max} > 0,$$

$$\rho_1 > 0,$$

$$\overline{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}], \forall i = 1, \dots, m,$$

$$\overline{\mu}_i^1 \in [0, \mu_{\max}], \forall i = 1, \dots, p,$$

$\{\varepsilon_k\} \subset \mathbb{R}_+$ uma sequência de parâmetros de tolerância tal que

$$\lim_{k \rightarrow \infty} \varepsilon_k = 0. \quad (1.8)$$

Passo 1 *Inicialização*

$k \leftarrow 1$. Definir $V^0 = g(x^0)_+$;

Passo 2 *Resolução do subproblema*

Calcular (se possível) $x^k \in \mathbb{R}^n$ tal que existam $v^k \in \mathbb{R}^m, u^k \in \mathbb{R}^p$ satisfazendo

$$\left\| \nabla L_{\rho_k}(x^k, \overline{\lambda}^k, \overline{\mu}^k) + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^p u_i^k \nabla \underline{g}_i(x^k) \right\| \leq \varepsilon_k, \quad (1.9)$$

$$u_i^k \geq 0, \underline{g}_i(x^k) < \varepsilon_k \text{ para todo } i = 1, \dots, p, \quad (1.10)$$

$$\underline{g}_i(x^k) < -\varepsilon_k \Rightarrow u_i^k = 0 \text{ para todo } i = 1, \dots, p, \quad (1.11)$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k. \quad (1.12)$$

Se não é possível encontrar x^k cumprindo (1.9) - (1.12), parar a execução do algoritmo.

Passo 3 *Estimar multiplicadores*

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

$$\lambda_i^{k+1} = \overline{\lambda}_i^k + \rho_k h_i(x^k) \quad (1.13)$$

e

$$\overline{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]. \quad (1.14)$$

Geralmente, $\overline{\lambda}_i^{k+1}$ é a projeção de λ_i^{k+1} no intervalo $[\lambda_{\min}, \lambda_{\max}]$.

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

$$\mu_i^{k+1} = \max\{0, \overline{\mu}_i^k + \rho_k g_i(x^k)\} \quad (1.15)$$

$$\begin{aligned}
V_i^k &= \max \left\{ g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k} \right\} \\
\bar{\mu}_i^{k+1} &\in [0, \mu_{\max}].
\end{aligned} \tag{1.16}$$

(Geralmente, $\bar{\mu}_i^{k+1} = \min\{\mu_i^{k+1}, \bar{\mu}_{\max}\}$.)

Passo 4 *Atualizar o parâmetro de penalidade*

Se

$$\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

definir

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

Caso contrário, definir

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

Passo 5 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao Passo 2.

Observações

1. As condições (1.9)-(1.12) dizem que x^k é um ponto KKT aproximado para o subproblema

$$\begin{aligned}
\min \quad & L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \\
s.a \quad & \underline{h}(x) = 0 \\
& \underline{g}(x) \leq 0
\end{aligned}$$

Portanto, a cada iteração externa minimizamos (aproximadamente) o Lagrangiano aumentado no conjunto Ω .

2. No passo 3 atualizamos os multiplicadores de Lagrange. No passo 4 atualizamos o parâmetro de penalidade de acordo com a admissibilidade - complementariedade do iterando obtido comparado com o iterando anterior.

3. No algoritmo sem critério de parada, mesmo que x^k seja uma solução global do problema original, o algoritmo “continua”, neste caso repetindo o ponto x^k para sempre. Naturalmente, isto é um artifício teórico pois, na prática, se estabelecem critérios para interromper a execução em um número finito e moderado de iterações.

1.2 Convergência a pontos admissíveis

O ideal seria definir algoritmos que sempre encontram pontos admissíveis, mas isto é impossível, pois em casos extremos poderiam não existir pontos admissíveis. Conseqüentemente é importante estudar o comportamento dos algoritmos com respeito à admissibilidade.

Enunciemos agora a condição de qualificação CPLD e o lema de Carathéodory:

Lema de Carathéodory

Lema 1.2.1. *Supomos que*

$$u = \sum_{i=1}^m \lambda_i v^i + \sum_{j=1}^p \mu_j w^j,$$

onde $v^i \in \mathbb{R}^n$ para todo $i = 1, \dots, m$, $w^j \in \mathbb{R}^n$ para todo $j = 1, \dots, p$, e $\mu_j \geq 0$ para todo $j = 1, \dots, p$. Então, existem $I \subset \{1, \dots, m\}$, $J \subset \{1, \dots, p\}$, $\{\ddot{\lambda}_i\}_{i \in I} \subset \mathbb{R}$, $\{\ddot{\mu}_j\}_{j \in J} \subset \mathbb{R}_+$, tais que

$$u = \sum_{i=1}^m \ddot{\lambda}_i v^i + \sum_{j=1}^p \ddot{\mu}_j w^j$$

e os vetores $\{v^i\}_{i \in I} \cup \{w^j\}_{j \in J}$ são linearmente independentes.

Demonstração: Podemos supor sem perda de generalidade que $\lambda_i \neq 0$, $\mu_j > 0$ para todo $i = 1, \dots, m$, $j = 1, \dots, p$.

Suponhamos que os vetores $\{v^1, \dots, v^m, w^1, \dots, w^p\}$ são linearmente dependentes. Portanto, existem escalares α_i e β_j , não todos nulos, tais que:

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

Logo, pela hipótese, temos que

$$u = \sum_{i=1}^m (\lambda_i - t\alpha_i)v^i + \sum_{j=1}^p (\mu_j - t\beta_j)w^j,$$

para todo $t \in \mathbb{R}$. Para $t = 0$, nenhum dos coeficientes na igualdade acima se anula. Seja $t_{\min}$ o t de menor módulo que anula pelo menos um dos coeficientes $\lambda_i - t\alpha_i$ ou $\mu_j - t\beta_j$. Então:

$$u = \sum_{i=1}^m (\lambda_i - t_{\min}\alpha_i)v^i + \sum_{j=1}^p (\mu_j - t_{\min}\beta_j)w^j.$$

Claramente, $\mu_j - t_{\min}\beta_j \geq 0$ para todo j , mas conseguimos escrever u como combinação linear de, no máximo, $m + p - 1$ vetores. Este processo pode ser repetido até que os vetores que contribuem à combinação linear sejam todos linearmente independentes. $\square$

Condição de qualificação CPLD (ver [1]):

Definição 1.2.2. *Suponhamos que $h(x) = 0, g(x) \leq 0, I_A(x) = \{i \in \{1, \dots, p\} | g_i(x) = 0\}$. Dizemos que x satisfaz a Condição de Dependência Linear Positiva Constante (CPLD) se a existência de $I_h \subset \{1, \dots, m\}, I_g \subset I_A(x), \lambda_i \in \mathbb{R}$ para todo $i \in I_h, \mu_i \geq 0$ para todo $i \in I_g$ tais que*

$$\sum_{i \in I_h} \lambda_i \nabla h_i(x) + \sum_{i \in I_g} \mu_i \nabla g_i(x) = 0$$

com $\sum_{i \in I_h} |\lambda_i| + \sum_{i \in I_g} \mu_i > 0$, implica que os gradientes

$$\{\nabla h_i(z)\}_{i \in I_h}, \{\nabla g_i(z)\}_{i \in I_g}$$

são linearmente dependentes para todo z em uma vizinhança de x .

Nesta seção vamos supor que o algoritmo não pára no Passo 2, ou seja, sempre podemos encontrar x^k satisfazendo (1.9)-(1.12).

Além disso, vamos supor que existe pelo menos um ponto limite da sequência gerada pelo Algoritmo 1.1. Uma condição para que isto ocorra é que $\Omega \neq \emptyset$, e existe $\varepsilon > 0$ tal que o conjunto

$$\{x \in \mathbb{R}^n | \underline{g}_i(x) \leq \varepsilon, \|\underline{h}(x)\| \leq \varepsilon\}$$

é limitado. Esta condição pode ser forçada acrescentando restrições artificiais no conjunto Ω .

Vamos investigar as propriedades dos pontos limite de sequências geradas pelo Algoritmo 1.1. Primeiro vamos provar um resultado sobre admissibilidade desses pontos limites.

Tal resultado mostra que todo ponto limite é admissível, ou é um ponto KKT da soma de quadrados das inviabilidades, ou não satisfaz a condição CPLD com respeito às restrições que definem o conjunto Ω .

Teorema 1.2.3. *Seja $\{x^k\}$ uma sequência infinita gerada pelo Algoritmo 1.1. Seja x^* um ponto limite de $\{x^k\}$. Se a sequência de parâmetros de penalidade ρ_k é limitada, então x^* é admissível. Em caso contrário, pelo menos uma das seguintes possibilidades acontece:*

(i) *O ponto x^* é um ponto KKT do problema*

$$\begin{array}{ll} \min & \left[\sum_{i=1}^m h_i(x)^2 + \sum_{i=1}^p \max\{0, g_i(x)\}^2 \right] \\ \text{s.a} & x \in \Omega \end{array}$$

(ii) *x^* não satisfaz a condição CPLD associada ao conjunto Ω .*

Demonstração: Seja K um subconjunto infinito de $\mathbb{N}$ tal que

$$\lim_{k \in K} x^k = x^*.$$

Segue de (1.8), (1.10) e (1.12), que $\underline{g}(x^*) \leq 0$ e $\underline{h}(x^*) = 0$. Portanto,

$$x^* \in \Omega.$$

Consideremos duas possibilidades:

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

Caso (a).

Neste caso, a partir de certa iteração, o parâmetro de penalidade não é mais atualizado.

Portanto, segue do Passo 4 do Algoritmo 1.1 que:

$$\lim_{k \rightarrow \infty} \|h(x^k)\| = \lim_{k \rightarrow \infty} \|V^k\| = 0.$$

Assim,

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

Ainda mais, se $g_i(x^*) > 0$ temos que $g_i(x^k) > c > 0$ para $k \in K$ suficientemente grande, isso contradiz o fato de que $V_i^k \rightarrow 0$.

Logo,

$$g_i(x^*) \leq 0 \text{ para todo } i = 1, \dots, p.$$

Portanto x^* é admissível.

Assim o resultado está provado para o caso em que $\{\rho_k\}$ é limitado.

Caso (b).

Pela definição do Algoritmo 1.1, e a condição aproximada (1.9), temos:

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^m (\bar{\lambda}_i^k + \rho_k h_i(x^k)) \nabla h_i(x^k) + \sum_{i=1}^p \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \nabla g_i(x^k) \\ + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^p u_i^k \nabla \underline{g}_i(x^k) = \delta^k, \end{aligned} \quad (1.17)$$

como $\varepsilon_k \rightarrow 0$ (1.8),

$$\lim_{k \in K} \|\delta^k\| = 0.$$

Se $\underline{g}_j(x^*) < 0$, existe $k_1 \in \mathbb{N}$ tal que $\underline{g}_j(x^k) < -\varepsilon_k$ para todo $k \geq k_1, k \in K$. Portanto, pela complementariedade aproximada (1.11), $u_j^k = 0$ para todo $k \in K, k \geq k_1$.

Logo, por $x^* \in \Omega$ e (1.17), para todo $k \in K, k \geq k_1$ temos que

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^m (\bar{\lambda}_i^k + \rho_k h_i(x^k)) \nabla h_i(x^k) + \sum_{i=1}^p \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \nabla g_i(x^k) \\ + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{i | \underline{g}_i(x^*) = 0} u_i^k \nabla \underline{g}_i(x^k) = \delta^k. \end{aligned}$$

Dividindo por ρ_k obtemos:

$$\begin{aligned} \frac{\nabla f(x^k)}{\rho_k} + \sum_{i=1}^m \left(\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x^k) \right) \nabla h_i(x^k) + \sum_{i=1}^p \max \left\{ 0, \frac{\bar{\mu}_i^k}{\rho_k} + g_i(x^k) \right\} \nabla g_i(x^k) \\ + \sum_{i=1}^m \frac{v_i^k}{\rho_k} \nabla \underline{h}_i(x^k) + \sum_{i|\underline{g}_i(x^*)=0} \frac{u_i^k}{\rho_k} \nabla \underline{g}_i(x^k) = \frac{\delta^k}{\rho_k}. \end{aligned}$$

Pelo Lema de Carathéodory, existem

$$\widehat{I}_k \subset \{1, \dots, m\}, \widehat{J}_k \subset \{j | \underline{g}_j(x^*) = 0\},$$

$$v_i^k, i \in \widehat{I}_k, \widehat{u}_j^k \geq 0, j \in \widehat{J}_k$$

tais que os vetores

$$\{\nabla \underline{h}_i(x^k)\}_{i \in \widehat{I}_k} \cup \{\nabla \underline{g}_j(x^k)\}_{j \in \widehat{J}_k}$$

são linearmente independentes e

$$\begin{aligned} \frac{\nabla f(x^k)}{\rho_k} + \sum_{i=1}^m \left(\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x^k) \right) \nabla h_i(x^k) + \sum_{i=1}^p \max \left\{ 0, \frac{\bar{\mu}_i^k}{\rho_k} + g_i(x^k) \right\} \nabla g_i(x^k) \\ + \sum_{i \in \widehat{J}_k} \widehat{u}_i^k \nabla \underline{g}_i(x^k) + \sum_{i \in \widehat{I}_k} \widehat{v}_i^k \nabla \underline{h}_i(x^k) = \frac{\delta^k}{\rho_k}. \end{aligned}$$

Como o número de possíveis conjuntos $\widehat{I}_k, \widehat{J}_k$ é finito, existe um conjunto infinito de índices K_1 tal que

$$K_1 \subseteq \{k \in K | k \geq k_1\},$$

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

e

$$\widehat{J} = \widehat{J}_k \subset \{i | \underline{g}_i(x^*) = 0\} \quad (1.18)$$

para todo $k \in K_1$. Assim, para todo $k \in K_1$ obtemos:

$$\frac{\nabla f(x^k)}{\rho_k} + \sum_{i=1}^m \left(\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x^k) \right) \nabla h_i(x^k) + \sum_{i=1}^p \max \left\{ 0, \frac{\bar{\mu}_i^k}{\rho_k} + g_i(x^k) \right\} \nabla g_i(x^k) +$$

$$\sum_{i \in \hat{J}} \hat{u}_i^k \nabla \underline{g}_i(x^k) + \sum_{i \in \hat{I}} \hat{v}_i^k \nabla \underline{h}_i(x^k) = \frac{\delta^k}{\rho_k}. \quad (1.19)$$

e os gradientes

$$\{\nabla \underline{h}_i(x^k)\}_{i \in \hat{I}} \cup \{\nabla \underline{g}_j(x^k)\}_{j \in \hat{J}} \quad (1.20)$$

são linearmente independentes para todo $k \in K_1$.

Consideramos, novamente, dois casos:

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

Se a sequência $\{\|(\hat{v}^k, \hat{u}^k)\|\}_{k \in K_1}$ é limitada e $\hat{I} \cup \hat{J} \neq \emptyset$, existe $(\hat{v}, \hat{u})$, $\hat{u} \geq 0$ e um conjunto infinito de índices $K_2 \subseteq K_1$ tal que

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

Como $\{\rho_k\}$ é não limitada, temos que $\rho_k \rightarrow \infty$. Logo pela limitação de $\bar{\lambda}^k$ e $\bar{\mu}^k$, temos

$$\lim_{k \in K_2} \frac{\bar{\lambda}_i^k}{\rho_k} = \lim_{k \in K_2} \frac{\bar{\mu}_j^k}{\rho_k} = 0, \quad \forall i, j.$$

Como $\delta^k \rightarrow 0$, por (1.14), (1.16), tomando limite para $k \in K_2$ em (1.19), obtemos:

$$\sum_{i=1}^m h_i(x^*) \nabla h_i(x^*) + \sum_{i=1}^p \max\{0, g_i(x^*)\} \nabla g_i(x^*) + \sum_{i \in \hat{J}} \hat{u}_i \nabla \underline{g}_i(x^*) + \sum_{i \in \hat{I}} \hat{v}_i \nabla \underline{h}_i(x^*) = 0,$$

onde $\hat{J} \subset \{i \in \{1, \dots, p\} | g_i(x^*) = 0\}$. Portanto, por $x^* \in \Omega$ e (1.18), x^* é ponto KKT da soma de quadrados das inviabilidades. Se $\hat{I} \cup \hat{J} = \emptyset$ segue que $u_i^k = v_i^k = 0, \forall i$ e para k suficientemente grande, tomando o limite em (1.19) obtemos:

$$\sum_{i=1}^m h_i(x^*) \nabla h_i(x^*) + \sum_{i=1}^p \max\{0, g_i(x^*)\} \nabla g_i(x^*) = 0.$$

Portanto, por $x^* \in \Omega$, x^* é um ponto KKT da soma de quadrados de inviabilidades, como enunciado.

Finalmente, suponhamos que $\{ \|(\hat{v}^k, \hat{u}^k)\| \}_{k \in K_1}$ é não limitada. Seja $K_3 \subset K_1$ tal que $\lim_{k \in K_3} \|(\hat{v}^k, \hat{u}^k)\| = \infty$ e sejam $(\hat{v}, \hat{u}) \neq 0, \hat{u} \geq 0$ tal que

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

Dividindo ambos lados de (1.19) por $\|(\hat{v}^k, \hat{u}^k)\|$ e tomando limites para $k \in K_3$ obtemos:

$$\sum_{i \in \hat{J}} \hat{u}_i \nabla \underline{g}_i(x^*) + \sum_{i \in \hat{I}} \hat{v}_i \nabla \underline{h}_i(x^*) = 0.$$

Mas $\underline{g}_j(x^*) = 0$ para todo $j \in \hat{J}$. Logo, por (1.20), x^* não satisfaz a condição CPLD associada ao conjunto Ω . Com isso, a prova está completa. $\square$

Este resultado mostrou que pontos limites de sequências geradas pelo Algoritmo 1.1 podem ser admissíveis ou não. Essencialmente, o Teorema 1.2.3 diz que, se x^* não é admissível então é, muito provavelmente, minimizador local da inadmissibilidade.

1.3 Convergência a pontos KKT

Nesta seção provaremos um resultado de otimalidade para pontos limites admissíveis de sequências geradas pelo Algoritmo 1.1.

Provaremos que, sob a condição CPLD, os pontos limites são KKT do problema original.

O primeiro resultado é interessante e nos será útil, e ele independe da CPLD. Este resultado nos diz que, se uma restrição de desigualdade se verifica estritamente no ponto limite, então a estimativa do multiplicador associado se anula, para k suficientemente grande.

Proposição 1.3.1. *Seja $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 1.1. Suponhamos que x^* é um ponto limite desta sequência e $K \subset \mathbb{N}$ é tal que $\lim_{k \in K} x^k = x^*$. Então, para $k \in K$ suficientemente grande,*

$$\text{se } g_i(x^*) < 0 \Rightarrow \mu_i^{k+1} = 0 \text{ e se } \underline{g}_i(x^*) < 0 \Rightarrow u_i^k = 0. \quad (1.21)$$

Demonstração: Pelos critérios de parada do subproblema no passo 2 do Algoritmo 1.1, que definem x^k , temos que para todo $k \in K$, existem $u^k \in \mathbb{R}_+^p$, $\delta^k \in \mathbb{R}$ tais que $\|\delta^k\| \leq \varepsilon_k$ onde

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{i=1}^p \mu_i^{k+1} \nabla g_i(x^k) + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^p u_i^k \nabla \underline{g}_i(x^k) = \delta^k. \quad (1.22)$$

Por (1.15), $\mu^{k+1} \in \mathbb{R}_+^p$ para todo $k \in \mathbb{N}$.

Vamos supor que $\underline{g}_i(x^*) < 0$. Então, existe $k_1 \in \mathbb{N}$ tal que $\forall k \in K, k \geq k_1$, $\underline{g}_i(x^k) < -\varepsilon_k$. Então, por (1.11),

$$u_i^k = 0, \quad \forall k \in K, k \geq k_1.$$

Agora vamos provar que uma propriedade similar se verifica quando $g_i(x^*) < 0$. Neste caso, existem $k_2 \geq k_1$ e $c < 0$ tais que

$$g_i(x^k) < c < 0, \quad \forall k \in K, k \geq k_2. \quad (1.23)$$

Consideremos dois casos:

1. A sequência $\{\rho_k\}$ não é limitada.
2. A sequência $\{\rho_k\}$ é limitada.

No primeiro caso, temos que $\lim_{k \in K} \rho_k = \infty$. Como $\{\bar{\mu}_i^k\}$ é limitada, existe $k_3 \geq k_2$ tal que, para todo $k \in K$, $k \geq k_3$,

$$\bar{\mu}_i^k + \rho_k g_i(x^k) < 0.$$

A definição de μ^{k+1} (1.15) implica

$$\mu^{k+1} = 0, \quad \forall k \in K, k \geq k_3.$$

Consideremos agora o segundo caso em que a sequência $\{\rho_k\}$ é limitada. Neste caso

$$\lim_{k \rightarrow \infty} V_i^k = 0.$$

Portanto, como $g_i(x^k) < c < 0$ para $k \in K$ suficientemente grande,

$$\lim_{k \in K} \bar{\mu}_i^k = 0.$$

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

$$\bar{\mu}_i^k + \rho_k g_i(x^k) < 0.$$

Pela definição de μ_i^k , existe $k_4 \in \mathbb{N}$ tal que $\mu_i^{k+1} = 0$ para $k \in K, k \geq k_4$.

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

$$g_i(x^*) < 0 \Rightarrow \mu_i^{k+1} = 0. \quad (1.24)$$

Portanto de (1.23) e (1.24) segue a tese da proposição. $\square$

Teorema 1.3.2. *Seja $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 1.1. Suponhamos que x^* é um ponto limite admissível que satisfaz CPLD com respeito a todas as restrições do problema original. Então, x^* é ponto KKT do problema original.*

Demonstração: Segue da Proposição 1.3.1, que existem $K \subset \mathbb{N}$, $k_5 \in \mathbb{N}$, tais que para todo $k \in K, k \geq k_5$, temos:

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^m (\lambda_i^{k+1} \nabla h_i(x^k)) + \sum_{i|g_i(x^*)=0} \mu_i^{k+1} \nabla g_i(x^k) \\ + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{j|\underline{g}_j(x^*)=0} u_j^k \nabla \underline{g}_j(x^k) = \delta^k, \end{aligned} \quad (1.25)$$

com $\mu^{k+1} \in \mathbb{R}_+^p$, $u^k \in \mathbb{R}_+^p$ e $\|\delta^k\| \leq \varepsilon_k$.

Pelo Lema de Carathéodory, para todo $k \in K, k \geq k_5$, existem

$$\widehat{I}_k \subset \{1, \dots, m\}, \widehat{J}_k \subset \{j|g_j(x^*) = 0\}, \check{I}_k \subset \{1, \dots, \underline{m}\}, \check{J}_k \subset \{j|\underline{g}_j(x^*) = 0\},$$

$$\widehat{\lambda}_i^k, \forall i \in \widehat{I}_k, \widehat{\mu}_j^k \geq 0, \forall j \in \widehat{J}_k, \widehat{v}_i^k, \forall i \in \check{I}_k, \widehat{u}_j^k \geq 0, \forall j \in \check{J}_k,$$

tais que os vetores

$$\{\nabla h_i(x^k)\}_{i \in \widehat{I}_k} \cup \{\nabla g_i(x^k)\}_{i \in \widehat{J}_k} \cup \{\nabla \underline{h}_i(x^k)\}_{i \in \check{I}_k} \cup \{\nabla \underline{g}_i(x^k)\}_{i \in \check{J}_k},$$

são linearmente independentes.

Como o número de possíveis conjuntos de índices $\widehat{I}_k, \widehat{J}_k, \check{I}_k, \check{J}_k$ é finito, existe $K_1 \subset \infty \{k \in K | k \geq k_5\}$ tal que

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

para todo $k \in K_1$.

Logo de (1.25),

$$\begin{aligned} \nabla f(x^k) + \sum_{i \in \widehat{I}} \widehat{\lambda}_i^k \nabla h_i(x^k) + \sum_{i \in \widehat{J}} \widehat{\mu}_i^k \nabla g_i(x^k) \\ + \sum_{i \in \check{I}} \widehat{v}_i^k \nabla \underline{h}_i(x^k) + \sum_{i \in \check{J}} \widehat{u}_i^k \nabla \underline{g}_i(x^k) = \delta^k, \end{aligned} \quad (1.26)$$

e os vetores

$$\{\nabla h_i(x^k)\}_{i \in \widehat{I}} \cup \{\nabla g_i(x^k)\}_{i \in \widehat{J}} \cup \{\nabla \underline{h}_i(x^k)\}_{i \in \check{I}} \cup \{\nabla \underline{g}_i(x^k)\}_{i \in \check{J}} \quad (1.27)$$

são linearmente independentes para todo $k \in K_1$.

Definimos

$$S_k = \max\{\max\{|\widehat{\lambda}_i^k|, i \in \widehat{I}\}, \max\{|\widehat{\mu}_i^k|, i \in \widehat{J}\}, \max\{|\widehat{v}_i^k|, i \in \check{I}\}, \max\{|\widehat{u}_i^k|, i \in \check{J}\}\}.$$

Consideremos duas possibilidades:

(a) $\{S_k\}_{k \in K_1}$ tem uma subsequência limitada.

(b) $\lim_{k \in K_1} S_k = \infty$.

Se $\{S_k\}_{k \in K_1}$ admite uma subsequência limitada, existe $K_2 \subset \infty K_1$ tal que

$$\lim_{k \in K_2} \widehat{\lambda}_i^k = \widehat{\lambda}_i,$$

$$\lim_{k \in K_2} \widehat{\mu}_i^k = \widehat{\mu}_i \geq 0,$$

$$\lim_{k \in K_2} \widehat{v}_i^k = \widehat{v}_i,$$

$$\lim_{k \in K_2} \widehat{u}_i^k = \widehat{u}_i \geq 0.$$

Da Proposição 1.3.1 segue que x^* satisfaz a complementariedade. Como $\|\delta^k\| \leq \varepsilon_k$, e $\{x^k\}$ é limitada e de (1.8) segue, tomando limites em (1.25) para $k \in K_2$, que:

$$\nabla f(x^*) + \sum_{i \in \widehat{I}} \widehat{\lambda}_i \nabla h_i(x^*) + \sum_{i \in \widehat{J}} \widehat{\mu}_i \nabla g_i(x^*) + \sum_{i \in \widetilde{I}} \widehat{v}_i \nabla \underline{h}_i(x^*) + \sum_{i \in \widetilde{J}} \widehat{u}_j \nabla \underline{g}_j(x^*) = 0,$$

com $\widehat{\mu}_i \geq 0, \widehat{u}_i \geq 0$. Como x^* é admissível, temos que x^* é um ponto KKT do problema original.

Supomos agora que $\lim_{k \in K_1} S_k = \infty$. Dividindo ambos os lados de (1.26) por S_k obtemos:

$$\begin{aligned} \frac{\nabla f(x^k)}{S_k} + \sum_{i \in \widehat{I}} \frac{\widehat{\lambda}_i^k}{S_k} \nabla h_i(x^k) + \sum_{i \in \widehat{J}} \frac{\widehat{\mu}_i^k}{S_k} \nabla g_i(x^k) + \\ + \sum_{i \in \widetilde{I}} \frac{\widehat{v}_i^k}{S_k} \nabla \underline{h}_i(x^k) + \sum_{i \in \widetilde{J}} \frac{\widehat{u}_j^k}{S_k} \nabla \underline{g}_j(x^k) = \frac{\delta^k}{S_k}, \end{aligned} \quad (1.28)$$

onde

$$\left| \frac{\widehat{\lambda}_i^k}{S_k} \right| \leq 1, \left| \frac{\widehat{\mu}_i^k}{S_k} \right| \leq 1, \left| \frac{\widehat{v}_i^k}{S_k} \right| \leq 1, \left| \frac{\widehat{u}_j^k}{S_k} \right| \leq 1.$$

Portanto, existe $K_3 \subset K_1$ tal que

$$\lim_{k \in K_3} \frac{\widehat{\lambda}_i^k}{S_k} = \widehat{\lambda}_i, \lim_{k \in K_3} \frac{\widehat{\mu}_i^k}{S_k} = \widehat{\mu}_i \geq 0, \lim_{k \in K_3} \frac{\widehat{v}_i^k}{S_k} = \widehat{v}_i, \lim_{k \in K_3} \frac{\widehat{u}_j^k}{S_k} = \widehat{u}_j \geq 0.$$

Tomando limite em ambos os lados de (1.28) para $k \in K_3$, obtemos:

$$\sum_{i \in \widehat{I}} \widehat{\lambda}_i \nabla h_i(x^*) + \sum_{i \in \widehat{J}} \widehat{\mu}_i \nabla g_i(x^*) + \sum_{i \in \widetilde{I}} \widehat{v}_i \nabla \underline{h}_i(x^*) + \sum_{i \in \widetilde{J}} \widehat{u}_j \nabla \underline{g}_j(x^*) = 0,$$

Mas o módulo de pelo menos um dos coeficientes $\widehat{\lambda}_i, \widehat{\mu}_i, \widehat{v}_i, \widehat{u}_i$ é igual a 1. Logo, pela condição CPLD, os gradientes

$$\{\nabla h_i(x^k)\}_{i \in \widehat{I}} \cup \{\nabla g_i(x^k)\}_{i \in \widehat{J}} \cup \{\nabla \underline{h}_i(x^k)\}_{i \in \widetilde{I}} \cup \{\nabla \underline{g}_i(x^k)\}_{i \in \widetilde{J}}$$

são linearmente dependentes em uma vizinhança de x^* . Isto contradiz (1.27). Portanto, o teorema está provado. $\square$

1.4 Limitação do parâmetro de penalidade

Nesta seção mostramos que o parâmetro de penalidade fica limitado com certa frequência, isto é, sob condições suficientes, que muitas vezes são satisfeitas nos problemas práticos.

Quanto maior é o parâmetro de penalidade, mais difíceis de resolver ficam os subproblemas do método de Lagrangiano aumentado. A razão disto é que, ao ser multiplicado por um número muito grande, o termo penalizado domina a função objetivo do problema. Assim, $f(x)$ funciona como uma pequena perturbação do segundo termo e o algoritmo que resolve o subproblema desprezará tal perturbação.

Em ponto flutuante, a soma de um número muito grande e um número relativamente pequeno dá como resultado o número grande. Portanto quando se minimiza uma função do tipo Lagrangiano aumentado com um parâmetro grande de penalidade, a tendência é que na maioria das iterações, o método interno se comporte como se $f(x)$ não existisse.

Portanto, é importante que o método tenha a propriedade de que o parâmetro de penalidade não cresça indefinidamente.

1.4.1 Restrições de igualdade

Consideramos inicialmente o problema de programação não linear (1.1). Considere também o algoritmo Lagrangiano Aumentado para este problema, com $\Omega = \mathbb{R}^n$.

Consequentemente, o Algoritmo 1.1 aplicado ficará simplificado da seguinte maneira:

ALGORITMO 1.2

Seja $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário.

Os parâmetros para a execução do algoritmo são:

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

$$-\infty < \lambda_{\min} < \lambda_{\max} < +\infty,$$

$$\rho_1 > 0,$$

$$\bar{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}] \forall i = 1, \dots, m,$$

$\{\varepsilon_k\} \subset \mathbb{R}_+$ uma sequência de parâmetros de tolerância tal que

$$\lim_{k \rightarrow \infty} \varepsilon_k = 0. \quad (1.29)$$

Passo 1 *Inicialização*

$$k \leftarrow 1.$$

Passo 2 *Resolução do subproblema*

Calcular (se possível) $x^k \in \mathbb{R}^n$ tal que

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \leq \varepsilon_k, \quad (1.30)$$

Se não é possível encontrar x^k cumprindo essa propriedade, parar a execução do algoritmo.

Passo 3 *Estimar multiplicadores*

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

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k) \quad (1.31)$$

e

$$\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]. \quad (1.32)$$

Geralmente, $\bar{\lambda}_i^{k+1}$ é a projeção de λ_i^{k+1} no intervalo $[\lambda_{\min}, \lambda_{\max}]$.

Passo 4 *Atualizar o parâmetro de penalidade*

Se

$$\|h(x^k)\|_{\infty} \leq \tau \|h(x^{k-1})\|_{\infty},$$

definir

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

Caso contrário, definir

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

Passo 5 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao Passo 2.

Para mostrar a limitação do parâmetro de penalidade assumimos as seguintes hipóteses:

1. A sequência $\{x^k\}$ gerada pelo Algoritmo 1.2 converge a x^* .
2. O ponto x^* é admissível (ou seja, $h(x^*) = 0$).
3. O ponto x^* é regular (ou seja, a matriz $\nabla h(x^*)$ tem posto igual a m).
4. Os multiplicadores λ_i^* pertencem a $(\lambda_{\min}, \lambda_{\max})$ para todo $i = 1, \dots, m$. (Note que a existência de λ^* satisfazendo $\nabla f(x^*) + \nabla h(x^*)\lambda^* = 0$ está garantida pela regularidade de x^*).
5. As funções f e h admitem derivadas segundas contínuas em uma vizinhança de x^* .
6. Seja $Z \in \mathbb{R}^{n \times (n-m)}$ uma matriz cujas colunas formam uma base do núcleo de $\nabla h(x^*)^T$. Então, $Z^T \nabla^2 L_0(x^*, \lambda^*) Z$ é definida positiva, onde $L_0(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i h_i(x)$.

Lema 1.4.1. *A sequência $\{\lambda^k\}$ converge a λ^* .*

Demonstração: Suponhamos que λ^k não converge a λ^* . Portanto λ^{k+1} não converge a λ^* . Logo, existem $c > 0$ e um conjunto infinito de índices $K \subset \mathbb{N}$ tais que

$$\|\lambda^{k+1} - \lambda^*\| \geq c, \quad \forall k \in K. \quad (1.33)$$

Pelo critério de resolução aproximada dos subproblemas (1.30) e a fórmula de λ^{k+1} (1.31) temos:

$$\lim_{k \rightarrow \infty} \nabla f(x^k) + \nabla h(x^k)\lambda^{k+1} = 0. \quad (1.34)$$

Supomos primeiro que a subsequência $\{\lambda^{k+1}\}_{k \in K}$ esteja limitada. Neste caso, ela tem uma subsequência convergente a, digamos, $\hat{\lambda}$. Portanto, passando (1.34) ao limite para esta subsequência,

$$\nabla f(x^*) + \nabla h(x^*)\hat{\lambda} = 0.$$

Como a matriz $\nabla h(x^*)$ tem posto completo, devemos ter $\widehat{\lambda} = \lambda^*$, o que contradiz (1.33).

Supomos que a subsequência $\{\lambda^{k+1}\}_{k \in K}$ não é limitada. Então, dividindo (1.34) por $\|\lambda^{k+1}\|$ e tomando outra subsequência adequada, chegamos que o ponto x^* não é regular, o que contradiz a hipótese. $\square$

Lema 1.4.2. *Para k suficientemente grande, $\bar{\lambda}^k = \lambda^k$.*

Demonstração: Pelo Lema 1.4.1, λ^k tende a λ^* . Mas, como λ^* pertence a $(\lambda_{\min}, \lambda_{\max})^m$, temos que $\lambda^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ para k suficientemente grande. Portanto, para k suficientemente grande, $\bar{\lambda}^{k+1} = \lambda^{k+1}$. $\square$

Lema 1.4.3. *Existem $c > 0$, $\bar{\rho} > 0$, $\delta > 0$ tais que, se $\pi \in [0, \frac{1}{\bar{\rho}}]$, $\|x - x^*\| \leq \delta$ e $\|\lambda - \lambda^*\| \leq \delta$, a matriz*

$$\int_0^1 \begin{pmatrix} \nabla^2 L_0(x^* + t(x - x^*), \lambda^* + t(\lambda - \lambda^*)) & \nabla h(x^* + t(x - x^*)) \\ \nabla h(x^* + t(x - x^*))^T & -\pi I \end{pmatrix} dt$$

é não singular e

$$\left\| \left[\int_0^1 \begin{pmatrix} \nabla^2 L_0(x^* + t(x - x^*), \lambda^* + t(\lambda - \lambda^*)) & \nabla h(x^* + t(x - x^*)) \\ \nabla h(x^* + t(x - x^*))^T & -\pi I \end{pmatrix} dt \right]^{-1} \right\| \leq c.$$

Demonstração: Vejamos primeiro que a matriz é não singular se $\pi = 0$, $x = x^*$, $\lambda = \lambda^*$.

Suponhamos que $u^1 \in \mathbb{R}^n$, $u^2 \in \mathbb{R}^m$ são tais que

$$\begin{pmatrix} \nabla^2 L_0(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)^T & 0 \end{pmatrix} \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Logo, u^1 está no núcleo de $\nabla h(x^*)^T$, portanto, existe $v \in \mathbb{R}^{n-m}$ tal que

$$u^1 = Zv.$$

Portanto,

$$\nabla^2 L_0(x^*, \lambda^*) Z v + \nabla h(x^*) u^2 = 0.$$

Pré-multiplicando por $v^T Z^T$ obtemos:

$$v^T Z^T \nabla^2 L_0(x^*, \lambda^*) Z v + v^T Z^T \nabla h(x^*) u^2 = 0.$$

Pela definição de Z , temos que $Z^T \nabla h(x^*) = 0$, portanto:

$$v^T Z^T \nabla^2 L_0(x^*, \lambda^*) Z v = 0.$$

Mas, pela hipótese 6, $Z^T \nabla^2 L_0(x^*, \lambda^*) Z$ é definida positiva, logo $v = 0$. Isto implica que $u^2 = 0$.

Em consequência,

$$\nabla h(x^*) u^2 = 0.$$

Como as colunas de $\nabla h(x^*)$ são por hipótese, linearmente independentes, deduzimos que $u^2 = 0$.

Portanto, a matriz $\begin{pmatrix} \nabla^2 L_0(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)^T & -\pi I \end{pmatrix}$ é não singular se $\pi = 0$.

Assim, considerando

$$F(x, \lambda, \pi) = \int_0^1 \begin{pmatrix} \nabla^2 L_0(x^* + t(x - x^*), \lambda^* + t(\lambda - \lambda^*)) & \nabla h(x^* + t(x - x^*)) \\ \nabla h(x^* + t(x - x^*))^T & -\pi I \end{pmatrix} dt.$$

temos que $F(x^*, \lambda^*, 0)$ é inversível. Portanto da continuidade das derivadas segundas e da inversa de matrizes segue a tese do lema. $\square$

Lema 1.4.4. *Seja $\bar{\rho}$ como no Lema 1.4.3. Supomos que existe $k_0 \in \mathbb{N}$ tal que $\rho_k \geq \bar{\rho}$ para todo $k \geq k_0$. Então, existe $M > 0$ tal que, para todo $k \geq k_0$,*

$$\|x^k - x^*\| \leq M \max \left\{ \frac{\|\bar{\lambda}^k - \lambda^*\|}{\rho_k}, \|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \right\},$$

e

$$\|\lambda^{k+1} - \lambda^*\| \leq M \max \left\{ \frac{\|\bar{\lambda}^k - \lambda^*\|}{\rho_k}, \|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \right\}.$$

Demonstração: Definimos, para todo $k \in \mathbb{N}$,

$$z^k = \nabla L_{\rho_k}(x^k, \bar{\lambda}^k) \in \mathbb{R}^n, \quad (1.35)$$

$$y^k = \frac{\lambda^* - \bar{\lambda}^k}{\rho_k} \in \mathbb{R}^m, \quad (1.36)$$

$$\pi_k = \frac{1}{\rho_k} \in \mathbb{R}_+.$$

Observe que, pela definição de λ^{k+1} , dada em (1.31)

$$z^k = \nabla L_0(x^k, \lambda^{k+1}).$$

Logo para todo $k \in \mathbb{N}$,

$$z^k = \nabla f(x^k) + \nabla h(x^k) \lambda^{k+1},$$

$$y^k = h(x^k) + (\lambda^* - \lambda^{k+1}) \pi_k.$$

Definimos, para todo $\pi \in [0, \frac{1}{\rho}]$,

$$F_\pi : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^{n+m}$$

por

$$F_\pi(x, \lambda) = \begin{pmatrix} \nabla f(x) + \nabla h(x) \lambda \\ h(x) + (\lambda^* - \lambda) \pi \end{pmatrix}.$$

Claramente,

$$F_{\pi_k}(x^k, \lambda^{k+1}) = \begin{pmatrix} z^k \\ y^k \end{pmatrix}$$

e, pelas hipóteses,

$$F_{\pi_k}(x^*, \lambda^*) = 0,$$

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

Portanto,

$$F_{\pi_k}(x^k, \lambda^{k+1}) - F_{\pi_k}(x^*, \lambda^*) = \begin{pmatrix} z^k \\ y^k \end{pmatrix}.$$

Mas pelo Teorema Fundamental do Cálculo Integral,

$$\begin{aligned} & F_{\pi_k}(x^k, \lambda^{k+1}) - F_{\pi_k}(x^*, \lambda^*) = \\ &= \left[\int_0^1 \begin{pmatrix} \nabla^2 L_0(x^* + t(x^k - x^*), \lambda^* + t(\lambda^k - \lambda^*)) & \nabla h(x^* + t(x^k - x^*)) \\ \nabla h(x^* + t(x^k - x^*))^T & -\pi I \end{pmatrix} dt \right] \begin{pmatrix} x^k - x^* \\ \lambda^{k+1} - \lambda^* \end{pmatrix}. \end{aligned}$$

Logo pelo Lema 1.4.3, se $\rho_k \geq \bar{\rho}$, e k é suficientemente grande, teremos $c > 0$, $\pi_k \leq \frac{1}{\bar{\rho}}$, $\|x^k - x^*\| \leq \delta$ e $\|\lambda^{k+1} - \lambda^*\| \leq \delta$. Portanto,

$$\left\| \begin{pmatrix} x^k - x^* \\ \lambda^{k+1} - \lambda^* \end{pmatrix} \right\| \leq c \left\| \begin{pmatrix} z^k \\ y^k \end{pmatrix} \right\|. \quad (1.37)$$

Agora segue da desigualdade (1.37) e da definição de z^k (1.35) que

$$\begin{aligned} \|x^k - x^*\| &\leq c \left(\frac{\|\bar{\lambda}^k - \lambda^*\|}{\rho_k} + \|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \right) \\ &\leq 2c \max \left\{ \frac{\|\bar{\lambda}^k - \lambda^*\|}{\rho_k}, \|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \right\}. \end{aligned}$$

Analogamente obtemos

$$\|\lambda^{k+1} - \lambda^*\| \leq 2c \max \left\{ \frac{\|\bar{\lambda}^k - \lambda^*\|}{\rho_k}, \|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \right\}.$$

□

Teorema 1.4.5. *Suponhamos, além das hipóteses desta seção, que existe uma sequência $\{\eta_k\}$ que converge a zero, tal que*

$$\varepsilon_k \leq \eta_k \|h(x^k)\|_\infty \text{ para todo } k \in \mathbb{N}, \quad (1.38)$$

onde ε_k foi definido em (1.29). Então, a sequência de parametros de penalidade $\{\rho_k\}$ está limitada.

Demonstração: Supomos que $\lim_{k \rightarrow \infty} \rho_k = \infty$. Como $h(x^*) = 0$, pela continuidade das primeiras derivadas de h , existe $L > 0$ tal que para todo $k \in \mathbb{N}$,

$$\|h(x^k)\|_\infty \leq L\|x^k - x^*\|.$$

Por (1.38), e pelo Lema 1.4.4 e o fato de que $\bar{\lambda}^k = \lambda^k$ para k suficientemente grande, temos que

$$\|h(x^k)\|_\infty \leq LM \max \left\{ \frac{\|\lambda^k - \lambda^*\|_\infty}{\rho_k}, \eta_k \|h(x^k)\|_\infty \right\}.$$

Como $\eta_k \rightarrow 0$, isto implica que

$$\|h(x^k)\|_\infty \leq LM \frac{\|\lambda^k - \lambda^*\|_\infty}{\rho_k}, \quad (1.39)$$

para k suficientemente grande.

Pela fórmula de λ^{k+1} , temos que

$$\lambda^k = \lambda^{k-1} + \rho_{k-1} h(x^{k-1}),$$

portanto,

$$\|h(x^{k-1})\|_\infty = \frac{\|\lambda^k - \lambda^{k-1}\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda^{k-1} - \lambda^*\|_\infty}{\rho_{k-1}} - \frac{\|\lambda^k - \lambda^*\|_\infty}{\rho_{k-1}} \quad (1.40)$$

Pelo Lema 1.4.4, a hipótese deste teorema e por $\bar{\lambda}^k = \lambda^k$, temos:

$$\|\lambda^k - \lambda^*\|_\infty \leq M \left(\frac{\|\lambda^{k-1} - \lambda^*\|_\infty}{\rho_{k-1}} \eta_{k-1} \|h(x^{k-1})\|_\infty \right).$$

Isto implica que

$$\frac{\|\lambda^{k-1} - \lambda^*\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda^k - \lambda^*\|_\infty}{M} - \eta_{k-1} \|h(x^{k-1})\|_\infty.$$

Logo, por (1.40), para $\rho_{k-1} \geq 2M$,

$$(1 + \eta_{k-1}) \|h(x^{k-1})\|_\infty \geq \|\lambda^k - \lambda^*\|_\infty \left(\frac{1}{M} - \frac{1}{\rho_{k-1}} \right) \geq \frac{1}{2M} \|\lambda^k - \lambda^*\|_\infty.$$

Assim,

$$\|\lambda^k - \lambda^*\|_\infty \leq 3M \|h(x^{k-1})\|_\infty,$$

para k suficientemente grande. Por (1.39), temos que

$$\|h(x^k)\|_\infty \leq \tau \|h(x^{k-1})\|_\infty$$

para todo $k \geq k_1$.

Isto implica que $\rho_{k+1} = \rho_k$ para todo $k \geq k_1$ e em consequência, a sequência $\{\rho_k\}$ está limitada.

□

1.4.2 Restrições de igualdade e desigualdade

Consideramos o problema de programação não linear definido em (1.3). Consideramos também o algoritmo Lagrangiano aumentado para este problema, ou seja, Algoritmo 1.1. Para mostrar a limitação do parâmetro de penalidade assumimos as seguintes hipóteses:

1. A sequência $\{x^k\}$ gerada pelo algoritmo converge a x^* .
2. O ponto x^* é admissível (ou seja, $h(x^*) = 0$, e $g(x^*) \leq 0$).
3. O ponto x^* é regular. Ou seja, o conjunto formado pelos gradientes das restrições de igualdade e os gradientes das restrições de desigualdade tais que $g_i(x^*) = 0$ é linearmente independente. Como no caso de restrições de igualdade, isto implica que as condições KKT se verificam em x^* . Portanto, existem multiplicadores de Lagrange $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}_+^p$ associados a estas restrições.
4. Os multiplicadores λ_i^* pertencem a $[\lambda_{min}, \lambda_{max}]$ para todo $i = 1, \dots, m$. Analogamente, cada multiplicador μ_i^* pertence a $[0, \mu_{max}]$ para todo $i = 1, \dots, p$.
5. As funções f , h e g admitem derivadas segundas contínuas em uma vizinhança de x^* .
6. Para todo $i = 1, \dots, p$ tal que $g_i(x^*) = 0$, temos $\mu_i^* > 0$.

7. Seja $A \in \mathbb{R}^{(m+q) \times n}$ a matriz cujas linhas são os gradientes das restrições de igualdade seguidas dos gradientes das q restrições ativas ($g_i(x^*) = 0$) em x^* . Seja $Z \in \mathbb{R}^{n \times (n-m-q)}$ uma matriz cujas colunas formam uma base do núcleo de A . Então, $Z^T \nabla^2 L_0(x^*, \lambda^*, \mu^*) Z$ é definida positiva, onde $L_0(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x)_+$.

Teorema 1.4.6. *Suponhamos, além das hipóteses desta seção, que existe uma sequência $\{\eta_k\}$ que converge a zero, tal que*

$$\varepsilon_k \leq \eta_k \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \text{ para todo } k \in \mathbb{N},$$

onde ε_k foi definido em (1.8). Então, a sequência de parâmetros de penalidade $\{\rho_k\}$ está limitada.

Demonstração: Suponhamos, por absurdo, que $\rho_k \rightarrow \infty$. A técnica para provar este teorema consiste em reduzir o problema definido em (1), (2) e (3) e o próprio Algoritmo 1.1 ao caso com restrições de igualdade do Algoritmo 1.2. Sem perda de generalidade, supomos que as restrições de desigualdade ativas em x^* são as q primeiras, de modo que

$$g_i(x^*) = 0 \text{ se } i \leq q, \quad g_i(x^*) < 0 \text{ se } i > q.$$

Consideremos o problema auxiliar:

$$\min \quad f(x) \tag{1.41}$$

$$\text{suja} \quad H(x) = 0, \tag{1.42}$$

onde

$$H(x) = \begin{pmatrix} h(x) \\ g_1(x) \\ \vdots \\ g_q(x) \end{pmatrix}.$$

Pelas hipóteses desta seção, o problema auxiliar satisfaz todas as suposições da seção anterior. Mais ainda, os multiplicadores associados a x^* no problema auxiliar são os mesmos que os

associados ao problema original.

Portanto, pelo Teorema 1.4.5, resta provar que a aplicação do Algoritmo 1.1 ao problema original é equivalente à aplicação do Algoritmo 1.2 para o problema auxiliar e que a condição $\varepsilon_k \leq \eta_k \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\}$ é equivalente a $\varepsilon_k \leq \eta_k \|H(x^k)\|_\infty$.

Seja i tal que $g_i(x^*) < 0$. Pela Proposição 1.3.1, temos que $\mu_i^k = 0$ para k suficientemente grande.

Logo, por (1.9) com $\Omega = \mathbb{R}^n$ e das atualizações de λ^{k+1} e μ^{k+1} ,

$$\left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{i=1}^q \mu_i^{k+1} \nabla g_i(x^k) \right\| \leq \varepsilon_k$$

para k suficientemente grande. Pela independência linear dos gradientes $\nabla h_i(x^*)$, $i = 1, \dots, m$, $\nabla g_j(x^*)$, $j = 1, \dots, q$, e a convergência de x^k , temos, como no Lema 1.4.1, que

$$\lim_{k \rightarrow \infty} \lambda^k = \lambda^* \text{ e } \lim_{k \rightarrow \infty} \mu^k = \mu^*.$$

Em particular, como $\mu_i^* > 0$ para todo $i \leq q$,

$$\mu_i^k > 0,$$

para k suficientemente grande.

Como $\lambda^* \in (\lambda_{\min}, \lambda_{\max})$ e $\mu_i^* < \mu_{\max}$, temos também que

$$\bar{\mu}_i^k = \mu_i^k, i = 1, \dots, p$$

e

$$\bar{\lambda}_i^k = \lambda_i^k, i = 1, \dots, m$$

para k suficientemente grande.

Vejamos agora que a fórmula de atualização de μ_i^{k+1} dada pelo Algoritmo 1.1 coincide com a fórmula de atualização do mesmo multiplicador no Algoritmo 1.2 aplicado ao problema auxiliar.

Com efeito, por (1.15) e $\bar{\mu}_i^k = \mu_i^k$, temos, para k suficientemente grande:

$$\mu_i^{k+1} = \max\{0, \mu_i^k + \rho_k g_i(x^k)\},$$

Mas, pela complementariedade estrita e $\mu^k \rightarrow \mu^*$, $\mu_i^k > 0$ para k suficientemente grande. Portanto, para k suficientemente grande,

$$\mu_i^{k+1} = \mu_i^k + \rho_k g_i(x^k), \quad i = 1, \dots, q.$$

Ou seja, em termos do problema auxiliar,

$$\mu_i^{k+1} = \mu_i^k + \rho_k H(x^k), \quad i = 1, \dots, q.$$

Assim, os multiplicadores correspondentes às restrições ativas de desigualdade se atualizam no Algoritmo 1.1, pela mesma regra aplicada no Algoritmo 1.2 para o problema auxiliar.

Vejamos agora o significado de V_i^k . Pelo Algoritmo 1.1, temos

$$V_i^k = \max \left\{ g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k} \right\}$$

para todo $i = 1, \dots, p$.

Agora, se $i > q$, como $g_i(x^*) < 0$, g_i é contínua e $\bar{\mu}_i^k = 0$, temos que $V_i^k = 0$ para k suficientemente grande.

Suponhamos agora que $i \leq q$. Se $g_i(x^k)$ fosse menor que $-\frac{\bar{\mu}_i^k}{\rho_k}$, pela atualização de μ^{k+1} no Algoritmo 1.1, teríamos que $\mu^{k+1} = 0$. Isto é impossível para k suficientemente grande devido à complementariedade estrita. Logo, para k suficientemente grande, $V_i^k = g_i(x^k)$.

Assim, para k suficientemente grande,

$$\|H(x^k)\| = \left\| \begin{pmatrix} h(x^k) \\ V^k \end{pmatrix} \right\|_{\infty}.$$

Isto implica que o teste para atualizar o parâmetro de penalidade no Algoritmo 1.1 é o mesmo que para o problema auxiliar no Algoritmo 1.2. Mais ainda, a condição

$$\bar{\varepsilon}_k \leq \eta_k \max\{\|V^k\|_{\infty}, \|h(x^k)\|_{\infty}\}$$

também é equivalente à condição $\varepsilon_k \|H(x^k)\|_{\infty}$ imposta no Teorema 1.4.5.

Portanto, para k suficientemente grande, a sequência gerada pelo Algoritmo 1.1 pode ser

pensada como sendo gerada pelo Algoritmo 1.2, junto com seus multiplicadores associados e parâmetros de penalidade. Isto implica que o parâmetro de penalidade é limitado, o que contradiz a não limitação assumida inicialmente na prova deste teorema. $\square$

1.5 Minimizadores globais

Nesta seção apresentamos o método de Lagrangiano aumentado para minimizadores globais.

Consideraremos o problema de programação não linear definido em (1), (2) e (3).

Vamos supor que a cada iteração externa o subproblema a ser resolvido possui mínimo global em Ω .

A seguir apresentamos o Algoritmo do Lagrangiano Aumentado para minimizadores globais (sem critério de parada).

ALGORITMO 1.3

Seja $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário.

Os parâmetros para a execução do algoritmo são:

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

$$-\infty < \lambda_{\min} < \lambda_{\max} < +\infty,$$

$$\mu_{\max} > 0,$$

$$\rho_1 > 0,$$

$$\overline{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}] \forall i = 1, \dots, m,$$

$$\mu_i^1 \in [0, \mu_{\max}], \forall i = 1, \dots, p,$$

Passo 1 Inicialização

$$k \leftarrow 1.$$

$$\text{Definir } V^0 = g(x^0)_+;$$

Passo 2 *Resolução do subproblema.*

Encontrar $x^k \in \Omega$ solução global do subproblema

$$\min f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[\max(0, g_i(x) + \frac{\mu_i}{\rho}) \right]^2 \right\} \quad (1.43)$$

sujeito a $x \in \Omega$.

Passo 3 *Estimar multiplicadores*

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

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k) \quad (1.44)$$

e

$$\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]. \quad (1.45)$$

Geralmente, $\bar{\lambda}_i^{k+1}$ é a projecção de λ_i^{k+1} no intervalo $[\lambda_{\min}, \lambda_{\max}]$.

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

$$\mu_i^{k+1} = \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \quad (1.46)$$

$$V_i^k = \max \left\{ g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k} \right\}$$

$$\bar{\mu}_i^{k+1} \in [0, \mu_{\max}]. \quad (1.47)$$

(Geralmente, $\bar{\mu}_i^{k+1} = \min\{\mu_i^{k+1}, \bar{\mu}_{\max}\}$.)

Passo 4 *Atualizar o parâmetro de penalidade*

$$R^k = \max\{\|h(x^k)\|_{\infty}, \|V^k\|_{\infty}\}$$

Se

$$R^k \leq \tau R^{k-1},$$

definir

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

Caso contrário, definir

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

Passo 5 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao Passo 2.

Observações

1. A condição (1.43) exige que x^k seja a solução global do subproblema

$$\begin{aligned} \min \quad & L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \\ \text{s.a} \quad & \underline{h}(x) = 0 \\ & \underline{g}(x) \leq 0 \end{aligned}$$

Portanto, a cada iteração externa encontramos o mínimo global do Lagrangiano aumentado no conjunto Ω .

2. Os passos 3 e 4 não se alteraram.

1.5.1 Convergência global

Mostraremos que pontos limites de sequências geradas pelo Algoritmo 1.3, são admissíveis se a região admissível é não vazia, e mais, são minimizadores globais do problema.

Teorema 1.5.1. *Suponhamos que a região admissível do problema de programação não linear é não vazia. Suponhamos além disso, que a sequência $\{x^k\}$ gerada pelo Algoritmo 1.3 possui um ponto limite x^* . Então x^* é admissível.*

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

$$\lim_{k \in K} x^k = x^*.$$

Dividiremos a demonstração em dois casos:

I- A sequência ρ_k é limitada;

II- A sequência ρ_k não é limitada;

Caso I.

Neste caso, a partir de certa iteração, o parâmetro de penalidade não é mais atualizado.

Portanto, segue do Passo 4 do Algoritmo 1.3 que:

$$\lim_{k \rightarrow \infty} \|h(x^k)\| = \lim_{k \rightarrow \infty} \|V^k\| = 0.$$

Assim,

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

Ainda mais, se $g_j(x^*) > 0$ temos que $g_j(x^k) > c > 0$ para $k \in K$ suficientemente grande, isso contradiz o fato de que $V_j^k \rightarrow 0, \forall j$.

Logo,

$$g_i(x^*) \leq 0 \text{ para todo } i = 1, \dots, p.$$

Agora como $x^k \in \Omega, \forall k$, e Ω é fechado então $x^* \in \Omega$, portanto x^* é admissível.

Assim o resultado está provado para o caso em que $\{\rho_k\}$ é limitado.

Caso II.

Agora como $x^k \in \Omega, \forall k$, e Ω é fechado então $x^* \in \Omega$. Suponhamos que x^* não é admissível então existe $z \in \Omega$ tal que

$$\|h(x^*)\|^2 + \|g(x^*)_+\|^2 > \|h(z)\|^2 + \|g(z)_+\|^2 = 0.$$

Como h e g são contínuas e, além disso, $\bar{\lambda}^k$ e $\bar{\mu}^k$ são limitados e $\rho_k \rightarrow +\infty$, quando $k \rightarrow +\infty$, então existe $c > 0$ tal que para $k \geq k_0, k \in K$

$$\left\| h(x^k) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(x^k) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| > \left\| h(z) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(z) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| + c. \quad (1.48)$$

Portanto, para esses índices k ,

$$f(x^k) + \frac{\rho_k}{2} \left[\left\| h(x^k) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(x^k) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| \right] > \quad (1.49)$$

$$f(z) + \frac{\rho_k}{2} \left[\left\| h(z) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(z) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| \right] + \frac{\rho_k c}{2} + f(x^k) - f(z).$$

De (1.49) obtemos:

$$L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) > L_{\rho_k}(z, \bar{\lambda}^k, \bar{\mu}^k) + \frac{(\rho_k)c}{2} + f(x^k) - f(z).$$

Agora, como $\lim_{k \in K} x^k = x^*$ e f é contínua, temos que, para $k \in K$ suficientemente grande:

$$\frac{(\rho_k - 1)c}{2} + f(x^k) - f(z) > 0.$$

Logo,

$$L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) > L_{\rho_k}(z, \bar{\lambda}^k, \bar{\mu}^k),$$

o que contradiz o fato de que x^k é uma solução global do subproblema definido por ρ_k . Esta contradição veio de supor que x^* não era admissível. $\square$

Teorema 1.5.2. *Suponhamos que a região admissível do problema de programação não linear definido em (1), (2) e (3) é não vazia e que a sequência $\{x^k\}$ está bem definida pelo Algoritmo 1.3, e x^* um ponto limite da sequência. Então, x^* é um minimizador global do problema.*

Demonstração: Seja $K \subseteq \mathbb{N}$ tal que

$$\lim_{k \in K} x^k = x^*.$$

Segue do teorema anterior que x^* é admissível.

Dividiremos novamente a demonstração em dois casos:

I- A sequência ρ_k é limitada;

II- A sequência ρ_k não é limitada;

Caso I,

Suponhamos que para $k \geq k_0$ não atualizamos mais o parâmetro de penalidade. Como as sequências $\{\bar{\lambda}^k\}$ e $\{\bar{\mu}^k\}$ estão em um compacto, existem $K_1 \subseteq K$, $\lambda^* \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^* \in [0, \mu_{\max}]^p$ tais que

$$\lim_{k \in K_1} x^k = x^*, \lim_{k \in K_1} \bar{\lambda}^k = \lambda^*, \lim_{k \in K_1} \bar{\mu}^k = \mu^*.$$

Mas, pelo passo 3 do Algoritmo 1.3, $\lim_{k \rightarrow \infty} \max\{g_i(x^k), -\bar{\mu}_i^k/\rho_{k_0}\} = 0$. Portanto, se $g_i(x^*) < 0$ devemos ter, forçosamente, que $\mu_i^* = 0$. Em outras palavras,

$$g_i(x^*)\mu_i^* = 0 \text{ para todo } i = 1, \dots, p.$$

Agora, seja $z \in \Omega$ um ponto admissível arbitrário, então teríamos pela definição do Algoritmo 1.3 para $k \geq k_0$

$$L_{\rho_{k_0}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_{k_0}}(z, \bar{\lambda}^k, \bar{\mu}^k)$$

Como $g(z) \leq 0$ e $\bar{\mu}^k/\rho_k \geq 0$,

$$\left\| \left(g(z) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\|^2 \leq \left\| \frac{\bar{\mu}^k}{\rho_k} \right\|^2,$$

portanto como $h(z) = 0$,

$$L_{\rho_{k_0}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq f(z) + \frac{\rho_{k_0}}{2} \left[\left\| \frac{\bar{\lambda}^k}{\rho_{k_0}} \right\|^2 + \left\| \frac{\bar{\mu}^k}{\rho_{k_0}} \right\|^2 \right],$$

para todo $k \geq k_0$. Tomando limite na desigualdade acima obtemos:

$$f(x^*) + \frac{\rho_{k_0}}{2} \left\| \left(g(x^*) + \frac{\mu^*}{\rho_{k_0}} \right)_+ \right\|^2 \leq f(z) + \frac{\rho_{k_0}}{2} \left\| \frac{\mu^*}{\rho_{k_0}} \right\|^2.$$

Ou seja,

$$f(x^*) + \frac{\rho_{k_0}}{2} \sum_{i=1}^p \left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}} \right)_+^2 \leq f(z) + \frac{\rho_{k_0}}{2} \sum_{i=1}^p \left(\frac{\mu_i^*}{\rho_{k_0}} \right)^2. \quad (1.50)$$

Mas

$$\left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}} \right)_+^2 = \max \left\{ 0, g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}} \right\}^2.$$

Portanto, se $g_i(x^*) = 0$,

$$\left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}}\right)_+^2 = \left(\frac{\mu_i^*}{\rho_{k_0}}\right)_+^2$$

e, se $g_i(x^*) < 0$, logo $\mu_i^* = 0$,

$$\left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}}\right)_+^2 = 0.$$

Portanto de (1.50) obtemos $f(x^*) \leq f(z)$. Como z é um ponto admissível arbitrário, resulta que x^* é solução global.

Caso II,

Seja $z \in \Omega$ um ponto admissível. Pela definição do Algoritmo 1.3, temos que

$$L_{\rho_{k_0}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_{k_0}}(z, \bar{\lambda}^k, \bar{\mu}^k)$$

Mas $h(z) = 0$ e $g(z)_+ = 0$, portanto

$$\left\|h(z) + \frac{\bar{\lambda}^k}{\rho_k}\right\|^2 = \left\|\frac{\bar{\lambda}^k}{\rho_k}\right\|^2$$

e

$$\left\|\left(g(z) + \frac{\bar{\mu}^k}{\rho_k}\right)_+\right\|^2 \leq \left\|\frac{\bar{\mu}^k}{\rho_k}\right\|^2.$$

Logo,

$$\begin{aligned} f(x^k) &\leq L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \\ &\leq f(z) + \frac{\rho_k}{2} \left[\left\|\frac{\bar{\lambda}^k}{\rho_k}\right\|^2 + \left\|\frac{\bar{\mu}^k}{\rho_k}\right\|^2 \right], \end{aligned}$$

Tomando limites para $k \in K$ e usando que $\rho_k \rightarrow \infty$, e que $\bar{\lambda}^k, \bar{\mu}^k$ são limitados, obtemos:

$$f(x^*) \leq f(z).$$

Como z é um ponto admissível arbitrário, resulta que x^* é solução global.

Do caso I e II segue a tese do teorema. $\square$

CAPÍTULO 2

Lagrangiano aumentado regularizado

Neste capítulo introduzimos um procedimento que utiliza a regularização para inibir a voracidade. A função objetivo utilizada nos subproblemas de Algencan ¹ (implementação de grande porte do Algoritmo 1.1) é modificada com o incremento do termo regularizador. Mostraremos que sobre este novo método continuam válidas as propriedades de convergência e de limitação do parâmetro de penalidade apresentadas no Capítulo 1. Este capítulo complementa os estudos feitos em [6].

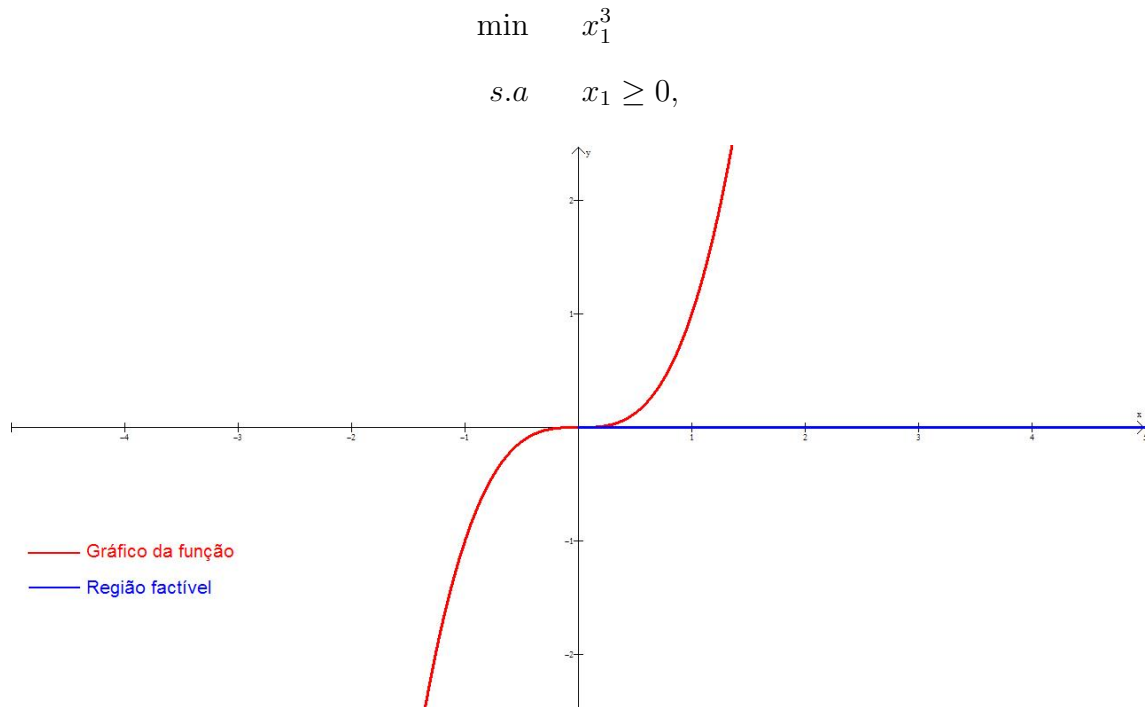
2.1 Voracidade

No método Lagrangiano aumentado definimos cautelosamente o parâmetro de penalidade para ser pequeno nas primeiras iterações externas, porque sabemos que parâmetros grandes tendem a produzir subproblemas mal condicionados. Como consequência, as primeiras iterações tendem a privilegiar a otimalidade sobre a viabilidade. Assim o método pode se comportar de maneira

¹Algencan esta disponível em www.ime.usp.br/~egbirgin/tango

voraz em busca da otimalidade em alguns problemas e acabar se perdendo, e consequentemente divergir. Esse fenômeno é denominado aqui como voracidade.

Para tornar mais claro este fenômeno, consideremos o exemplo a seguir, no qual o número de variáveis é 1 e o número de restrições também é 1.



O problema acima foi testado pelo método computacional Algencan. A restrição de caixa $x_1 \geq 0$ foi implementada como uma restrição de desigualdade e a caixa artificial fornecida ao método foi $-10^{20} \leq x_1 \leq 10^{20}$, com ponto inicial $x_1 = -7$ o resultado foi o seguinte:

Output of Algencan:

Initial $\rho = 10$ incrementado 8 vezes.

Outer iterations = 9

Final Objective function value = -1.0000E+60

Final Sup-norm of the constraints = 1.0000E+20

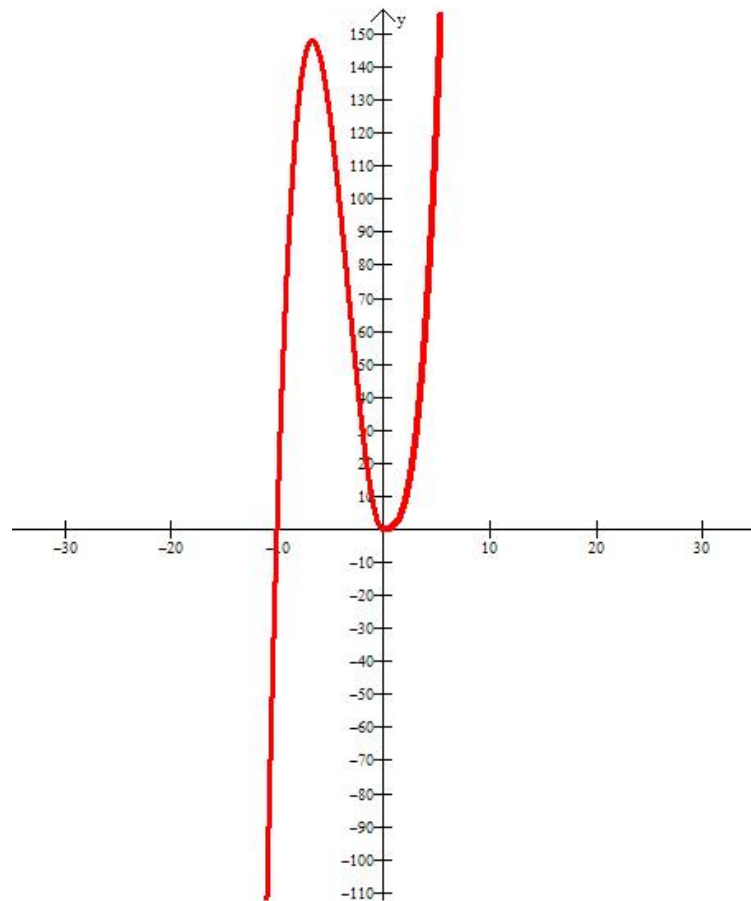
Final Sup-norm of the projected gradient of the Lagrangian = 0.0000E+00

Flag = 4 (sem decréscimo de infactibilidade depois de 9 iterações, problema possivelmente infactível).

O problema apresentou voracidade. Como podemos observar, na primeira iteração o subproblema a ser resolvido foi o seguinte:

$$\begin{aligned} \min \quad & x_1^3 + 10[\max\{0, -x_1\}]^2 \\ \text{s.a} \quad & -10^{20} \leq x_1 \leq 10^{20}, \end{aligned}$$

Veja abaixo o gráfico da função Lagrangiano Aumentado para a primeira iteração:



A solução deste problema encontra-se no bordo da caixa artificial, de onde o método não conseguiu sair, mesmo aumentando o parâmetro de penalidade por várias iterações.

A primeira idéia que surge é penalizar mais desde o início e, com isso, contornar a voracidade. Agora sabemos que, quanto maior o parâmetro de penalidade, mais difícil se torna resolver o problema original. Apresentamos uma maneira de contornar este problema sem aumentar demasiadamente o parâmetro de penalidade, e manter assim as boas propriedades do método para problemas que não apresentam voracidade.

2.2 O método Lagrangiano aumentado regularizado

Nesta seção consideramos o problema de programação não linear na sua forma padrão, dado em (1), (2) e (3).

Neste texto introduziremos no método de Lagrangiano aumentado um termo regularizador, cujas principais motivações são:

- Melhor aproveitar a aproximação inicial x^0 ;
- Evitar que o método se distancie muito da factibilidade nas suas primeiras iterações;
- Contornar o problema da voracidade;
- Melhorar o condicionamento do problema, tornando assim mais “leves” as primeiras iterações.

Para isto somamos o termo regularizador

$$\gamma \frac{\|x - \bar{x}\|_2^2}{2} \tag{2.1}$$

ao Lagrangiano Aumentado, onde $\gamma \geq 0$ e $\bar{x}$ é o ponto de referência, o melhor iterando até o momento em relação à factibilidade-complementariedade. Assim definimos a função

Lagrangiano Aumentado Regularizado como:

$$L_{\rho, \gamma, \bar{x}}(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[\max(0, g_i(x) + \frac{\mu_i}{\rho}) \right]^2 \right\} + \gamma \frac{\|x - \bar{x}\|_2^2}{2}. \quad (2.2)$$

2.3 Algoritmo Lagrangiano aumentado regularizado

A seguir apresentamos o algoritmo do Lagrangiano aumentado regularizado (sem critério de parada):

ALGORITMO 2.1.

Seja $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário.

Os parâmetros para a execução do algoritmo são:

$$\tau \in [0, 1), \gamma_1 \geq 0, \gamma_{\max} > 0, \eta > 1, \beta > 0$$

$$-\infty < \lambda_{\min} < \lambda_{\max} < +\infty,$$

$$\mu_{\max} > 0,$$

$$\rho_1 > 0,$$

$$\bar{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}] \forall i = 1, \dots, m,$$

$$\mu_i^1 \in [0, \mu_{\max}], \forall i = 1, \dots, p,$$

$\{\varepsilon_k\} \subset \mathbb{R}_+$ uma sequência de parâmetros de tolerância tal que

$$\lim_{k \rightarrow \infty} \varepsilon_k = 0. \quad (2.3)$$

Passo 1 Inicialização

$$k \leftarrow 1.$$

$$\text{Definir } V^0 = g(x^0)_+;$$

$$\bar{x}^0 = x^0;$$

$$\bar{R} = R^0 = \max\{\|h(x^0)\|_\infty, \|V^0\|_\infty\};$$

Passo 2 *Resolução do subproblema*

Calcular (se possível) $x^k \in \mathbb{R}^n$ tal que existam $v^k \in \mathbb{R}^m, u^k \in \mathbb{R}^p$ satisfazendo

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \gamma_k(x^k - \bar{x}^{k-1}) + \sum_{i=1}^m v_i^k \nabla h_i(x^k) + \sum_{i=1}^p u_i^k \nabla g_i(x^k)\| \leq \varepsilon_k, \quad (2.4)$$

$$u_i^k \geq 0, g_i(x^k) < \varepsilon_k \text{ para todo } i = 1, \dots, p, \quad (2.5)$$

$$g_i(x^k) < -\varepsilon_k \Rightarrow u_i^k = 0 \text{ para todo } i = 1, \dots, p, \quad (2.6)$$

$$\|h(x^k)\| \leq \varepsilon_k. \quad (2.7)$$

Se não é possível encontrar x^k cumprindo (2.4) - (2.7), parar a execução do algoritmo.

Passo 3 *Estimar multiplicadores*

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

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k). \quad (2.8)$$

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

$$\begin{aligned} \mu_i^{k+1} &= \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \\ V_i^k &= \max\left\{g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k}\right\}. \end{aligned} \quad (2.9)$$

Passo 4 *Atualizar o parâmetro de penalidade*

$$R_k = \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\}.$$

Se

$$R_k \leq \tau R_{k-1} \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

definir

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

Caso contrário, definir

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

Passo 5 *Atualizar o parâmetro de regularização e o ponto de referência*

Se

$$R_k = \min\{R_0, \dots, R_k\}, \quad (2.10)$$

definimos $\bar{x}^k = x^k$.

Caso contrário, definimos $\bar{x}^k = \bar{x}^{k-1}$.

Para todo $i = 1, \dots, m$ computar:

$$\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]. \quad (2.11)$$

Para todo $i = 1, \dots, p$ computar:

$$\bar{\mu}_i^{k+1} \in [0, \mu_{\max}]. \quad (2.12)$$

Escolhemos $\gamma_{k+1} \in [0, \gamma_{\max}]$ de maneira que

$$\gamma_{k+1} \leq \min\{\gamma_k, \beta R_k\}. \quad (2.13)$$

Passo 6 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao Passo 2.

Observações

1. As condições (2.4)-(2.7) dizem que x^k é um ponto KKT aproximado para o subproblema

$$\begin{aligned} \min \quad & L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \\ s.a \quad & \underline{h}(x) = 0 \\ & \underline{g}(x) \leq 0 \end{aligned}$$

Portanto, a cada iteração externa, minimizamos (aproximadamente) o Lagrangiano Aumentado Regularizado no conjunto Ω .

2. Os passos 3 e 4 são como no Algoritmo 2.1. No passo 5, o ponto de referência e o parâmetro de regularização são atualizados. A idéia é que o ponto de referência seja o melhor iterando até o momento em relação à factibilidade-complementariedade que é medida por R_k . A condição (2.13) impõe que a sequência do parâmetro de regularização seja monótona decrescente, e convirja à zero se alguma subsequência de $\{R_k\}_{k=0}^{\infty}$ convergir a zero. Assim o efeito da regularização tende a desaparecer quando o Algoritmo se comporta bem em termos da obtenção de pontos factíveis.
3. Se atualizamos o ponto de referência, os multiplicadores $\bar{\lambda}$ e $\bar{\mu}$ em (2.11) e (2.12) serão calculados da seguinte forma:

$$\bar{\lambda}_i^{k+1} = \max\{\lambda_{\min}, \min\{\lambda_{\max}, \lambda_i^{k+1}\}\},$$

para $i = 1, \dots, m$ e

$$\bar{\mu}_i^{k+1} = \min\{\mu_{\max}, \mu_i^{k+1}\},$$

para $i = 1, \dots, p$.

Quando não atualizamos o ponto de referência, definimos $\bar{\lambda}^{k+1} = \bar{\lambda}^k$ e $\bar{\mu}^{k+1} = \bar{\mu}^k$.

Para provar a convergência global do Algoritmo 2.1 usamos o seguinte fato: se $K \subseteq \mathbb{N}$,

$$\lim_{k \in K} R_k = 0 \Rightarrow \lim_{k \rightarrow \infty} \gamma_k = 0. \quad (2.14)$$

Esta propriedade é verdadeira porque $\lim_{k \in K} R_k = 0$ implica, por (2.13), que $\lim_{k \in K} \gamma_{k+1} = 0$. Agora, como $\gamma_{k+1} \leq \gamma_k$ para todo k , portanto $\lim_{k \rightarrow \infty} \gamma_k = 0$. Isto significa que a condição

$$\gamma_{k+1} \leq \gamma_k, \quad \forall k, \quad (2.15)$$

imposta em (2.13) é necessária para provar o teorema de convergência. No entanto, a condição (2.15) contradiz o senso comum porque, talvez gostaríamos de aumentar o parâmetro de regularização se a factibilidade se deteriorasse. Em outras palavras, são razoáveis estratégias de regularização que exigem somente $\gamma_{k+1} \leq \beta R_k$, mas não que $\gamma_{k+1} \leq \gamma_k$.

Veremos que, com uma adequada modificação do Algoritmo 2.1, a condição de monotonicidade sobre γ_k pode ser eliminada.

ALGORITMO 2.2

Este algoritmo coincide com o Algoritmo 2.1, exceto que:

- Para a solução do subproblema, emprega um algoritmo que usa $\bar{x}^{k-1}$ como ponto inicial, e então garantimos que

$$L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(\bar{x}^{k-1}, \bar{\lambda}^k, \bar{\mu}^k) \quad (2.16)$$

- A condição (2.13) é reescrita por

$$\gamma_{k+1} \leq \beta R_k. \quad (2.17)$$

A condição (2.16) é bastante natural para um algoritmo de minimização monótona que preserve viabilidade das restrições para a resolução dos subproblemas.

O que permite relaxar (2.17) é o seguinte resultado fundamental.

Lema 2.3.1. *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 2.2 e suponha que existe um conjunto infinito de índices K tal que*

$$\lim_{k \in K} R_k = 0.$$

Então,

$$\lim_{k \rightarrow \infty} R_k = 0. \quad (2.18)$$

Demonstração: Se a sequência $\{\rho_k\}$ é limitada, então $R_k \leq \tau R_{k-1}$ para todo k suficientemente grande, a tese esta provada.

Assumimos, agora, que $\lim_{k \rightarrow \infty} \rho_k = \infty$.

Por (2.16), temos para todo $k \in \mathbb{N}$,

$$f(x^k) + \frac{\rho_k}{2} \left\{ \sum_{i=1}^m \left[h_i(x_k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(x_k) + \frac{\bar{\mu}_i^k}{\rho_k} \right) \right]^2 \right\} + \frac{\gamma_k}{2} \|x^k - \bar{x}^{k-1}\|^2$$

$$\leq f(\bar{x}^{k-1}) + \frac{\rho_k}{2} \left\{ \sum_{i=1}^m \left[h_i(\bar{x}_{k-1}) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(\bar{x}^{k-1}) + \frac{\bar{\mu}_i^k}{\rho_k} \right) \right]^2 \right\}.$$

Dividindo por ρ_k obtemos:

$$\begin{aligned} & \frac{1}{\rho_k} f(x^k) + \frac{1}{2} \left\{ \sum_{i=1}^m \left[h_i(x_k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right) \right]^2 \right\} + \frac{\gamma_k}{2} \|x^k - \bar{x}^{k-1}\|^2 \\ & \leq \frac{1}{\rho_k} f(\bar{x}^{k-1}) + \frac{1}{2} \left\{ \sum_{i=1}^m \left[h_i(\bar{x}^{k-1}) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(\bar{x}^{k-1}) + \frac{\bar{\mu}_i^k}{\rho_k} \right) \right]^2 \right\}. \end{aligned} \quad (2.19)$$

Pela definição de $\bar{x}^{k-1}$ podemos escrever:

$$\{\bar{x}^s, \bar{x}^{s+1}, \bar{x}^{s+2}, \dots\} = \{x^{k_0}, x^{k_1}, x^{k_2}, \dots\}$$

onde $k_0 \leq k_1 \leq k_2 \leq \dots$. No entanto, sendo $\lim_{k \in K} R_k = 0$, temos que

$$\lim_{j \rightarrow \infty} R_{k_j} = 0. \quad (2.20)$$

Claramente, (2.20) implica que

$$\lim_{j \rightarrow \infty} \|h(x^{k_j})\|^2 + \sum_{i=1}^p \max\{0, g_i(x^{k_j})\}^2 = 0.$$

Portanto,

$$\lim_{k \rightarrow \infty} \|h(\bar{x}^{k-1})\|^2 + \sum_{i=1}^p \max\{0, g_i(\bar{x}^{k-1})\}^2 = 0.$$

Portanto, o lado direito da desigualdade (2.19) tende a zero quando k tende ao infinito.

Assim,

$$\lim_{k \rightarrow \infty} \frac{1}{2} \left\{ \sum_{i=1}^m \left[h_i(x_k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right]^2 + \sum_{i=1}^p \left[\max \left(0, g_i(x^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right) \right]^2 \right\} = 0. \quad (2.21)$$

Como $\rho_k \rightarrow \infty$ e $\bar{\mu}^k$ é limitado, (2.21) implica que

$$\lim_{k \rightarrow \infty} \|h(x^k)\| = 0 \quad (2.22)$$

e

$$\lim_{k \rightarrow \infty} \max\{0, g_i(x^k)\} = 0 \quad \forall i = 1, \dots, p.$$

Então,

$$\lim_{k \rightarrow \infty} V_i^k = 0 \quad \forall i = 1, \dots, p. \quad (2.23)$$

Por (2.22) e (2.23) obtemos (2.18). $\square$

Lema 2.3.2. *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 2.2 e suponha que existe um conjunto infinito de índices K tal que*

$$\lim_{k \in K} R_k = 0.$$

Então,

$$\lim_{k \rightarrow \infty} \gamma_k = 0.$$

Demonstração: O resultado segue de (2.17) e do Lema 2.3.1. $\square$

2.4 Convergência

Nesta seção mostraremos a convergência global dos Algoritmos 2.1 e 2.2. Essencialmente, mostraremos que a adição do termo regularizador não interfere nas propriedades mostradas no Capítulo 1.

Supomos que os Algoritmos 2.1 e 2.2 não param no Passo 2 e que a sequência gerada $\{x^k\}$ é limitada. Assim a sequência possui um ponto limite. Uma condição para que isto ocorra é que $\Omega \neq \emptyset$ e existe $\varepsilon > 0$ tal que o conjunto

$$\{x \in \mathbb{R}^n \mid \underline{g}_i(x) \leq \varepsilon, \|\underline{h}(x)\| \leq \varepsilon\}$$

seja limitado. Esta condição pode ser forçada acrescentando restrições artificiais no conjunto Ω .

O primeiro resultado é interessante e útil, ele independe da CPLD. Este resultado nos diz que, se uma restrição de desigualdade se verifica estritamente no ponto limite, então a estimativa do multiplicador associado se anula, para k suficientemente grande.

Proposição 2.4.1. *Seja $\{x^k\}_{k \in \mathbb{N}}$ uma sequência gerada pelo Algoritmo 2.1 ou 2.2. Suponhamos que x^* é um ponto limite desta sequência e $K \subset \mathbb{N}$ é tal que $\lim_{k \in K} x^k = x^*$. Então, para $k \in K$ suficientemente grande,*

$$g_i(x^*) < 0 \Rightarrow \mu_i^{k+1} = 0 \text{ e } \underline{g}_i(x^*) < 0 \Rightarrow u_i^k = 0. \quad (2.24)$$

Demonstração: É obtida seguindo a mesma sequência de argumentos usada na Proposição 1.3.1. $\square$

Lema 2.4.2. *Seja $\{x^k\}$ uma sequência limitada gerada pelo Algoritmo 2.1 ou 2.2. Então,*

$$\lim_{k \rightarrow \infty} \frac{\gamma_k}{\rho_k} \|x^k - \bar{x}^{k-1}\| = 0 \quad (2.25)$$

Demonstração: A sequência $\{\bar{x}^k\}$ é limitada por definição, uma vez que $\{\bar{x}^k\} \subset \{x^k\}$. Logo $\|x^k - \bar{x}^{k-1}\|$ é limitado.

Do Passo 5 dos Algoritmos 2.1 e 2.2 segue que a sequência $\{\gamma_k\}$ é limitada.

Temos duas possibilidades a considerar:

1. A sequência $\{\rho_k\}$ é limitada.

Então o parâmetro para de ser incrementado para k suficientemente grande. Portanto, segue do passo 4 do Algoritmo 2.1 e 2.2 que, $\lim_{k \rightarrow \infty} R_k = 0$. Então sendo $\gamma_{k+1} \leq \beta R_k$, segue que $\lim_{k \rightarrow \infty} \gamma_k = 0$. Como $\|x^k - \bar{x}^{k-1}\|$ é limitado, obtemos (2.25).

2. A sequência $\{\rho_k\}$ não é limitada.

As limitações de $\{\gamma_k\}$ e $\|x^k - \bar{x}^{k-1}\|$ implicam que (2.25) ocorre. $\square$

Teorema 2.4.3. *Seja $\{x^k\}$ uma sequência limitada gerada pelo Algoritmo 2.1 ou 2.2, e seja x^* um ponto limite desta sequência. Então:*

1. *Se a sequência de parâmetros de penalidade ρ_k é limitada, x^* é admissível. Em caso contrário, pelo menos uma das seguintes possibilidades acontece:*

(i) O ponto x^* é um ponto KKT do problema

$$\begin{aligned} \min \quad & \left[\sum_{i=1}^m h_i(x)^2 + \sum_{i=1}^p \max\{0, g_i(x)\}^2 \right] \\ \text{s.a} \quad & x \in \Omega \end{aligned}$$

(ii) x^* não satisfaz a condição CPLD associada às restrições do conjunto Ω .

2. Se x^* é um ponto limite admissível que satisfaz CPLD com respeito a todas as restrições do problema definido em (1), (2) e (3). Então, x^* é ponto KKT do problema.

Demonstração: Seja $k \subseteq \mathbb{N}$ tal que

$$\lim_{k \in K} x^k = x^*.$$

A primeira parte do teorema é obtida seguindo a mesma sequência de argumentos usada no Teorema 1.2.3, com a diferença que teremos que utilizar o Lema 2.4.2 e o fato de que a sequência $\{x^k\}$ é limitada.

Vamos provar a segunda parte do teorema. Segue de (2.4) - (2.7) e da Proposição 2.4.1, que para $k \in K$ suficientemente grande temos que:

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^m (\lambda_i^{k+1} \nabla h_i(x^k) + \sum_{i|g_i(x^*)=0} \mu_i^{k+1} \nabla g_i(x^k) + \gamma_k(x^k - \bar{x}^{k-1}) \\ + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{j|\underline{g}_j(x^k)=0} u_j^k \nabla \underline{g}_j(x^k) = \underline{\delta}^k, \end{aligned}$$

onde $\mu^{k+1} \in \mathbb{R}_+^p$, $w^k \in \mathbb{R}_+^p$ e $\|\underline{\delta}^k\| \rightarrow 0$.

Como x^* é factível, temos que $\lim_{k \in K} R_k = 0$. Assim por (2.13) para o Algoritmo 2.1 temos:

$$\lim_{k \rightarrow \infty} \gamma_k = 0. \quad (2.26)$$

No caso do Algoritmo 2.2, (2.26) segue do Lema 2.3.2. Portanto, podemos definir

$$\delta^k = \underline{\delta}^k - \gamma_k(x^k - \bar{x}^{k-1}),$$

e então, para k suficientemente grande,

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^m (\lambda_i^{k+1} \nabla h_i(x^k) + \sum_{i|g_i(x^*)=0} \mu_i^{k+1} \nabla g_i(x^k) \\ + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{j|\underline{g}_j(x^k)=0} u_j^k \nabla \underline{g}_j(x^k) = \delta^k, \end{aligned}$$

onde $\mu^{k+1} \in \mathbb{R}_+^p$, $w^k \in \mathbb{R}_+^p$ e $\|\delta^k\| \rightarrow 0$.

A partir deste ponto, seguem os mesmos argumentos do Teorema 1.3.2, e obtemos o resultado.

□

2.5 Limitação do parâmetro de penalidade

2.5.1 Restrições de igualdade

Consideramos inicialmente o problema (1.1), que possui apenas restrições de igualdade. Considere também o algoritmo Lagrangiano aumentado regularizado para este problema, com $\Omega = \mathbb{R}^n$.

Consequentemente, o Algoritmo 2.1 aplicado ficará simplificado da seguinte maneira:

ALGORITMO 2.3

Seja $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário.

Os parâmetros para a execução do algoritmo são:

$$\tau \in [0, 1), \gamma_1 \geq 0, \gamma_{\max}, \eta > 1, \beta > 0$$

$$-\infty < \lambda_{\min} < \lambda_{\max} < +\infty,$$

$$\rho_1 > 0,$$

$$\overline{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}] \forall i = 1, \dots, m,$$

$\{\varepsilon_k\} \subset \mathbb{R}_+$ uma sequência de parâmetros de tolerância tal que

$$\lim_{k \rightarrow \infty} \varepsilon_k = 0. \tag{2.27}$$

Passo 1 *Inicialização*

$k \leftarrow 1$. Definir $R_0 = \|h(x^0)\|_\infty$

$h_1 = 0, h_2 = 0$.

Passo 2 *Resolução do subproblema*

Caso seja possível encontrar $x^k \in \mathbb{R}^n$ tal que

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k) + \gamma_k(x^k - \bar{x}^{k-1})\| \leq \varepsilon_k, \quad (2.28)$$

Se não é possível encontrar x^k cumprindo essa propriedade parar a execução do algoritmo.

Passo 3 *Estimar multiplicadores*

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

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k). \quad (2.29)$$

Passo 4 *Atualizar o parâmetro de penalidade*

$$R_k = \|h(x^k)\|_\infty$$

Se

$$R_k \leq \tau R_{k-1},$$

definir

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

Caso contrário, definir

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

Passo 5 *Atualizar o parâmetro de regularização e o ponto de referência*

Se $R_k = \min\{R_0, \dots, R_k\}$ definimos $\bar{x}^k = x^k$.

Caso contrário, definimos $\bar{x}^k = \bar{x}^{k-1}$.

Escolhemos $\gamma_{k+1} \in [0, \gamma_{\max}]$ de maneira que

$$\gamma_{k+1} \leq \min\{\gamma_k, \beta R_k\}. \quad (2.30)$$

Para todo $i = 1, \dots, m$ computar:

$$\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]. \quad (2.31)$$

Passo 6 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao Passo 2.

Para mostrar a limitação do parâmetro de penalidade assumimos as seguintes hipóteses:

1. A sequência $\{x^k\}$ gerada pelo algoritmo converge a x^* .
2. O ponto x^* é admissível (ou seja, $h(x^*) = 0$).
3. O ponto x^* é regular (ou seja, a matriz $\nabla h(x^*)$ tem posto igual a m).
4. Os multiplicadores λ_i^* pertencem a $[\lambda_{\min}, \lambda_{\max}]$ para todo $i = 1, \dots, m$. (Note que a existência de λ^* satisfazendo $\nabla f(x^*) + \nabla h(x^*)\lambda^*$ está garantida pela regularidade de x^*).
5. As funções f e h admitem derivadas segundas contínuas em uma vizinhança de x^* .
6. Seja $Z \in \mathbb{R}^{n \times (n-m)}$ uma matriz cujas colunas formam uma base do núcleo de $\nabla h(x^*)^T$. Então, $Z^T \nabla^2 L_0(x^*, \lambda^*) Z$ é definida positiva.
7. Existe $k_0 \in \mathbb{N}$, tal que para $k \geq k_0$, $R_{k+1} = \min\{R_0, \dots, R_{k+1}\}$, ou seja, para k suficientemente grande atualizamos os multiplicadores como no Algoritmo 1.1.

Lema 2.5.1. *Existe sequência $\bar{\varepsilon}_k \rightarrow 0$ tal que*

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k)\| \leq \bar{\varepsilon}_k, \quad \forall k \in \mathbb{N}.$$

Demonstração: Sendo x^* admissível, então $\lim_{k \rightarrow \infty} R_k = 0 \Rightarrow \lim_{k \rightarrow \infty} \gamma_k = 0$, portanto sendo $\{x^k\}$ limitada então $\lim_{k \rightarrow \infty} \|\gamma_k(x^k - \bar{x})\| = 0$

Do passo 2 do Algoritmo 2.1 temos

$$\|\nabla L_{\rho_k}(x^k, \lambda^k) + \gamma_k(x^k - \bar{x})\| \leq \varepsilon_k,$$

segue da desigualdade triangular que

$$\|\nabla L_{\rho_k}(x^k, \lambda^k)\| - \|\gamma_k(x^k - \bar{x})\| \leq \varepsilon_k,$$

então

$$0 \leq \|\nabla L_{\rho_k}(x^k, \lambda^k)\| \leq \underbrace{\varepsilon_k + \|\gamma_k(x^k - \bar{x})\|}_{\bar{\varepsilon}_k}$$

Segue de (2.27) que $\lim_{k \rightarrow \infty} \varepsilon_k = 0$, portanto $\lim_{k \rightarrow \infty} \bar{\varepsilon}_k = 0$

□

Com este lema podemos reproduzir todos os resultados da subseção 1.4.1.

2.5.2 Restrições de igualdade e desigualdade

Consideremos o problema (1.3) que possui restrições de igualdade e desigualdade. Consideremos também o algoritmo Lagrangiano aumentado regularizado para o problema, com $\Omega = \mathbb{R}^n$. Para mostrar a limitação do parâmetro de penalidade assumimos as seguintes hipóteses:

1. A sequência $\{x^k\}$ gerada pelo algoritmo converge a x^* .
2. O ponto x^* é admissível (ou seja, $h(x^*) = 0$, e $g(x^*) \leq 0$).
3. O ponto x^* é regular.
4. Os multiplicadores λ_i^* pertence a $[\lambda_{min}, \lambda_{max}]$ para todo $i = 1, \dots, m$. Analogamente, cada multiplicador μ_i^* pertence a $[0, \mu_{max}]$ para todo $i = 1, \dots, p$.

5. As funções f , h e g admitem derivadas segundas contínuas em uma vizinhança de x^* .
6. Para todo $i = 1, \dots, p$ tal que $g_i(x^*) = 0$, temos $\mu_i^* > 0$.
7. Seja $A \in \mathbb{R}^{(m+q) \times n}$ a matriz cujas linhas são os gradientes das restrições de igualdade seguidas dos gradientes das q restrições ativas ($g_i(x^*) = 0$) em x^* . Seja $Z \in \mathbb{R}^{n \times (n-m-q)}$ uma matriz cujas colunas formam uma base do núcleo de A . Então, $Z^T \nabla^2 L_0(x^*, \lambda^*, \mu^*) Z$ é definida positiva.
8. Existe $k_0 \in \mathbb{N}$, tal que para $k \geq k_0$, $R_{k+1} = \min\{R_0, \dots, R_{k+1}\}$, ou seja, para k suficientemente grande atualizamos os multiplicadores como no Algoritmo 1.1.

Lema 2.5.2. *Existe sequência $\bar{\varepsilon}_k \rightarrow 0$ tal que*

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)\| \leq \bar{\varepsilon}_k, \quad \forall k \in \mathbb{N}.$$

Demonstração: Análoga a prova do Lema 2.5.1. □

Com este lema podemos reproduzir todos os resultados da subseção 1.4.2.

2.6 Minimizadores globais

Nesta seção mostramos que o método de Lagrangiano aumentado regularizado, com poucas alterações possui as mesmas propriedades que o método sem regularização para minimizadores globais.

Consideraremos o problema de programação não linear definido em (1), (2) e (3).

Supomos que a cada iteração externa o subproblema a ser resolvido possui mínimo global em Ω .

A seguir apresentamos o algoritmo do Lagrangiano aumentado regularizado para minimizadores globais (sem critério de parada).

ALGORITMO 2.4

Seja $x^0 \in \mathbb{R}^n$ um ponto inicial arbitrário.

Os parâmetros para a execução do algoritmo são:

$$\tau \in [0, 1), \gamma_1 \geq 0, \gamma_{\max} > 0, \eta > 1, \beta_1 > 0, \beta_2 > 0$$

$$-\infty < \lambda_{\min} < \lambda_{\max} < +\infty,$$

$$\mu_{\max} > 0,$$

$$\rho_1 > 0,$$

$$\bar{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}], \forall i = 1, \dots, m,$$

$$\bar{\mu}_i^1 \in [0, \mu_{\max}], \forall i = 1, \dots, p,$$

Passo 1 *Inicialização*

$$k \leftarrow 1.$$

$$\text{Definir } V^0 = g(x^0)_+;$$

$$\bar{x} = x^0;$$

$$\bar{R} = R^0 = \max\{\|h(x^0)\|_\infty, \|V^0\|_\infty\};$$

Passo 2 *Resolução do subproblema.*

Encontrar $x^k \in \Omega$ solução global do subproblema

$$\text{Minimizar } L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x, \bar{\lambda}^k, \bar{\mu}^k), \text{ sujeito a } x \in \Omega. \quad (2.32)$$

Passo 3 *Estimar multiplicadores*

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

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k). \quad (2.33)$$

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

$$\begin{aligned}\mu_i^{k+1} &= \max\{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\} \\ V_i^k &= \max\left\{g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k}\right\}.\end{aligned}\tag{2.34}$$

Passo 4 *Atualizar o parâmetro de penalidade*

$$R_k = \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\}$$

Se

$$R_k \leq \tau R_{k-1},$$

definir

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

Caso contrário, definir

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

Passo 5 *Atualizar o parâmetro de regularização e o ponto de referência*

Se $R_k = \min\{R_0, \dots, R_k\}$ definimos $\bar{x}^k = x^k$.

Caso contrário, definimos $\bar{x}^k = \bar{x}^{k-1}$.

Para todo $i = 1, \dots, m$ computar:

$$\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}].\tag{2.35}$$

Para todo $i = 1, \dots, p$ computar:

$$\bar{\mu}_i^{k+1} \in [0, \mu_{\max}].\tag{2.36}$$

Escolhemos $\gamma_{k+1} \in [0, \gamma_{\max}]$ de maneira que

$$\gamma_{k+1} \leq \min\left\{\gamma_k, \beta_1 R_k, \beta_2 \frac{1}{\rho_{k+1}}\right\}\tag{2.37}$$

Passo 6 *Começar uma nova iteração*

Atualizar $k \leftarrow k + 1$. Voltar ao Passo 2.

Observações

1. A condição (2.32) exige que x^k seja a solução global do subproblema

$$\begin{aligned} \min \quad & L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \\ \text{s.a} \quad & \underline{h}(x) = 0 \\ & \underline{g}(x) \leq 0 \end{aligned}$$

Portanto, a cada iteração externa encontramos o mínimo global do Lagrangiano aumentado regularizado no conjunto Ω .

2. No passo 5 o ponto de referência e o parâmetro de regularização são atualizados. A diferença é que também diminuimos o compromisso com o ponto de referência se o parâmetro de penalidade aumentar, ou seja, se o parâmetro de penalidade já se preocupa com a admissibilidade, o parâmetro de regularização pode diminuir.

Observação 2.6.1. *Encontrar minimizadores globais parece não combinar muito com o fenômeno da voracidade. Para tanto devemos ter todos os subproblemas limitados inferiormente em Ω , e sabemos que isto pode não ocorrer quando o problema possui voracidade. Mas mostraremos que o método regularizado possui as mesmas propriedades com relação à convergência a minimizadores globais se comparado ao método de Lagrangiano aumentado.*

2.6.1 Convergência global

Nesta seção provamos que os pontos limites de sequências geradas pelo Algoritmo 2.4, são admissíveis se a região admissível é não vazia, e mais são minimizadores globais do problema definido em (1), (2) e (3).

Teorema 2.6.2. *Suponhamos que a região admissível do problema definido em (1), (2) e (3) é não vazia. Suponhamos além disso, que a sequência $\{x^k\}$ gerada pelo Algoritmo 2.4 é limitada, e seja x^* um ponto limite desta sequência. Então x^* é um ponto admissível para o problema.*

Demonstração: Seja $K \subseteq \mathbb{N}$ tal que

$$\lim_{k \in K} x^k = x^*.$$

Dividiremos a demonstração em dois casos:

I- A sequência ρ_k é limitada;

II- A sequência ρ_k não é limitada;

Caso I.

Análogo ao Caso I do Teorema 1.5.1.

Caso II.

Agora como $x^k \in \Omega, \forall k$, e Ω é fechado então $x^* \in \Omega$, suponhamos que x^* não é admissível então existe $z \in \Omega$ tal que

$$\|h(x^*)\|^2 + \|g(x^*)_+\|^2 > \|h(z)\|^2 + \|g(z)_+\|^2 = 0.$$

Como h e g são contínuas e, além disso, $\bar{\lambda}^k$ e $\bar{\mu}^k$ são limitados e $\rho_k \rightarrow +\infty$, quando $k \rightarrow +\infty$, então existe $c > 0$ tal que para $k \geq k_0, k \in K$

$$\left\| h(x^k) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(x^k) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| > \left\| h(z) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(z) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| + c. \quad (2.38)$$

Portanto, para esses índices k ,

$$f(x^k) + \frac{\rho_k}{2} \left[\left\| h(x^k) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(x^k) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| \right] > \quad (2.39)$$

$$f(z) + \frac{\rho_k}{2} \left[\left\| h(z) + \frac{\bar{\lambda}^k}{\rho_k} \right\| + \left\| \left(g(z) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\| \right] + \frac{\rho_k c}{2} + f(x^k) - f(z).$$

Como $\rho_k \rightarrow +\infty$, segue de (2.37), que $\gamma_k \rightarrow 0$, quando $k \rightarrow +\infty$, sendo $\{x^k\}$ limitada então $\{\bar{x}^k\}$ também é limitada, assim existe k_1 , tal que para todo $k \geq k_1$, $k \in K$ temos

$$\frac{\|x^k - \bar{x}^{k-1}\|_2^2}{2} \gamma_k + \frac{c}{2} > \frac{\|z - \bar{x}^{k-1}\|_2^2}{2} \gamma_k \quad (2.40)$$

Combinando (2.39) e (2.40) obtemos:

$$L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) > L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(z, \bar{\lambda}^k, \bar{\mu}^k) + \frac{(\rho_k - 1)c}{2} + f(x^k) - f(z).$$

Agora, como $\lim_{k \in K} x^k = x^*$ e f é contínua, temos que, para $k \in K$ suficientemente grande:

$$\frac{(\rho_k - 1)c}{2} + f(x^k) - f(z) > 0.$$

Logo,

$$L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) > L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(z, \bar{\lambda}^k, \bar{\mu}^k),$$

que contradiz o fato de que x^k é uma solução global do subproblema definido por ρ_k . Esta contradição veio de supor que x^* não era admissível. $\square$

Teorema 2.6.3. *Suponhamos que a região admissível do problema definido em (1), (2) e (3) é não vazia e que a sequência $\{x^k\}$ está bem definida pelo Algoritmo 2.4, e é limitada sendo x^* um ponto limite. Então, x^* é um minimizador global do problema.*

Demonstração: Seja $K \subseteq \mathbb{N}$ tal que

$$\lim_{k \in K} x^k = x^*.$$

Segue do teorema anterior que x^* é admissível.

Dividiremos novamente a demonstração em dois casos:

- I- A sequência ρ_k é limitada;
- II- A sequência ρ_k não é limitada;

Caso I,

Suponhamos que para $k \geq k_0$ não atualizamos mais o parâmetro de penalidade. Como as

sequências $\{\bar{\lambda}^k\}$ e $\{\bar{\mu}^k\}$ estão em um compacto, existem $K_1 \subseteq K$, $\lambda^* \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^* \in [0, \mu_{\max}]^p$ tais que

$$\lim_{k \in K_1} x^k = x^*, \lim_{k \in K_1} \bar{\lambda}^k = \lambda^*, \lim_{k \in K_1} \bar{\mu}^k = \mu^*.$$

Mas, pelo Passo 3 do Algoritmo 2.4, $\lim_{k \rightarrow \infty} \max\{g_i(x^k), -\bar{\mu}_i^k/\rho_{k_0}\} = 0$. Portanto, se $g_i(x^*) < 0$ devemos ter, forçosamente, que $\mu_i^* = 0$. Em outras palavras,

$$g_i(x^*)\mu_i^* = 0 \text{ para todo } i = 1, \dots, p.$$

Agora, seja $z \in \Omega$ um ponto admissível arbitrário, então teríamos, pela definição do Algoritmo 2.4 para $k \geq k_0$

$$L_{\rho_{k_0}, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_{k_0}, \gamma_k, \bar{x}^{k-1}}(z, \bar{\lambda}^k, \bar{\mu}^k)$$

Como $g(z) \leq 0$ e $\bar{\mu}^k/\rho_k \geq 0$,

$$\left\| \left(g(z) + \frac{\bar{\mu}^k}{\rho_k} \right)_+ \right\|^2 \leq \left\| \frac{\bar{\mu}^k}{\rho_k} \right\|^2,$$

portanto como $h(z) = 0$,

$$L_{\rho_{k_0}, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq f(z) + \frac{\rho_{k_0}}{2} \left[\left\| \frac{\bar{\lambda}^k}{\rho_{k_0}} \right\|^2 + \left\| \frac{\bar{\mu}^k}{\rho_{k_0}} \right\|^2 \right] + \frac{\|z - \bar{x}^{k-1}\|_2^2}{2} \gamma_k, \quad (2.41)$$

para todo $k \geq k_0$. Como $\{x^k\}$ é limitada, logo $\{\bar{x}^k\}$ também é, sendo x^* admissível então $\lim_{k \in K_1} \gamma_k = 0$, então tomando limite na desigualdade (2.41) obtemos:

$$f(x^*) + \frac{\rho_{k_0}}{2} \left\| \left(g(x^*) + \frac{\mu^*}{\rho_{k_0}} \right)_+ \right\|^2 \leq f(z) + \frac{\rho_{k_0}}{2} \left\| \frac{\mu^*}{\rho_{k_0}} \right\|^2.$$

Ou seja,

$$f(x^*) + \frac{\rho_{k_0}}{2} \sum_{i=1}^p \left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}} \right)_+^2 \leq f(z) + \frac{\rho_{k_0}}{2} \sum_{i=1}^p \left(\frac{\mu_i^*}{\rho_{k_0}} \right)^2. \quad (2.42)$$

Mas

$$\left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}} \right)_+^2 = \max \left\{ 0, g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}} \right\}^2.$$

Portanto, se $g_i(x^*) = 0$,

$$\left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}}\right)_+^2 = \left(\frac{\mu_i^*}{\rho_{k_0}}\right)_+^2$$

e, se $g_i(x^*) < 0$, logo $\mu_i^* = 0$,

$$\left(g_i(x^*) + \frac{\mu_i^*}{\rho_{k_0}}\right)_+^2 = 0.$$

Portanto de (2.42) obtemos $f(x^*) \leq f(z)$. Como z é um ponto admissível arbitrário, resulta que x^* é solução global.

Caso II,

Seja $z \in \Omega$ um ponto admissível. Pela definição do Algoritmo 2.4, temos que

$$L_{\rho_{k_0}, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq L_{\rho_{k_0}, \gamma_k, \bar{x}^{k-1}}(z, \bar{\lambda}^k, \bar{\mu}^k)$$

Mas $h(z) = 0$ e $g(z)_+ = 0$, portanto

$$\left\|h(z) + \frac{\bar{\lambda}^k}{\rho_k}\right\|^2 = \left\|\frac{\bar{\lambda}^k}{\rho_k}\right\|^2$$

e

$$\left\|\left(g(z) + \frac{\bar{\mu}^k}{\rho_k}\right)_+\right\|^2 \leq \left\|\frac{\bar{\mu}^k}{\rho_k}\right\|^2.$$

Assim,

$$\begin{aligned} f(x^k) &\leq L_{\rho_k, \gamma_k, \bar{x}^{k-1}}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \\ &\leq f(z) + \frac{\rho_k}{2} \left[\left\|\frac{\bar{\lambda}^k}{\rho_k}\right\|^2 + \left\|\frac{\bar{\mu}^k}{\rho_k}\right\|^2 \right] + \frac{\|z - \bar{x}^{k-1}\|_2^2}{2} \gamma_k \end{aligned}$$

Tomando limites para $k \in K$ e usando que $\rho_k \rightarrow \infty$, $\gamma_k \rightarrow 0$ e que $\bar{\lambda}^k$, $\bar{\mu}^k$, $\{x^k\}$, $\{\bar{x}^k\}$ são limitados, obtemos:

$$f(x^*) \leq f(z).$$

Como z é um ponto admissível arbitrário, resulta que x^* é solução global.

Dos casos I e II segue a tese do teorema.

□

%

CAPÍTULO 3

Resultados numéricos

Neste capítulo apresentamos os detalhes sobre o desempenho dos métodos computacionais RAlgencan e RAlgencan* que são implementações de grande porte dos algoritmos 2.2 e 2.1 respectivamente, feitas baseadas no método Algencan (versão de março de 2007) ¹.

3.1 Algencan

Nesta seção apresentamos detalhes da implementação do método computacional Algencan.

Algencan é um método computacional para resolver o problema de programação não linear

¹Algencan esta disponível em www.ime.usp.br/~egbirgin/tango

formulado como

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0 \\ & g(x) \leq 0 \\ & a^{\min} \leq x \leq a^{\max}, \end{aligned}$$

onde, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ possuem derivadas primeiras contínuas.

Em linhas gerais, Algencan é uma implementação do Algoritmo 1.1.

3.1.1 Subproblema

Para resolver o subproblema de minimizar o Lagrangiano aumentado no Algoritmo 1.1 é utilizado um método para minimizar em caixa. O esquema básico deste método é o de GENCAN. Ele utiliza estratégia de restrições ativas abandonando as faces quando necessário, usando o gradiente espectral projetado.

3.1.2 Parâmetro de penalidade inicial e multiplicadores iniciais

Na primeira iteração a aproximação para os multiplicadores é nula. A escolha de ρ_1 tem bastante influência no desempenho prático do método de Lagrangiano aumentado.

Devemos salientar que Algencan não escala os problemas.

Se a infactibilidade é nula no ponto inicial então $\rho_0 = 10$, senão

$$\rho_0 = \max \left\{ 10^{-6}, \min \left\{ 10, \frac{20 \max\{1, |f(x^0)|\}}{\|h(x^0)\|^2 + \|g_+(x^0)\|^2} \right\} \right\}$$

3.1.3 Critérios de parada

Na resolução do subproblema GENCAN pára se encontrar x^k tal que $a^{\min} \leq x^k \leq a^{\max}$ e tal que

$$\|P(x^k - \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)) - x^k\|_{\infty} \leq \varepsilon_k.$$

P aqui é o operador projeção na caixa $a^{\min} \leq x \leq a^{\max}$, o critério teórico exigido no Algoritmo 1.1 para o iterando x^k (condições (1.9)-(1.12)) será satisfeito.

Algencan para se a solução não é atingida com um número máximo de iterações para o subproblema, utilizamos o valor padrão 5000 iterações.

A execução de Algencan se interrompe declarando convergência, quando as seguintes condições são verificadas, utilizamos $\varepsilon = 10^{-4}$

$$\|P(x^k - \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)) - x^k\|_{\infty} \leq \varepsilon, \quad (3.1)$$

$$\|h(x^k)\|_{\infty} \leq \varepsilon, g_i(x^k) \leq \varepsilon, \text{ para todo } i = 1, \dots, p, \quad (3.2)$$

$$\mu_i^{k+1} = 0 \text{ para todo } i \text{ tal que } g_i(x^k) < -\varepsilon. \quad (3.3)$$

Devido às definições (1.13) e (1.15) de λ^{k+1} e μ^{k+1} , a condição (3.1) é equivalente a

$$\|P(x^k - \nabla L_0(x^k, \lambda^k, \mu^k)) - x^k\|_{\infty} \leq \varepsilon.$$

Algencan ainda pode parar sem declarar convergência, se uma das condições abaixo ocorrer:

- Se por 9 iterações consecutivas não houve progresso na admissibilidade de x^k ;
- Chegou-se a um máximo de 50 iterações externas;
- Se o parâmetro de penalidade ficar demasiadamente grande, $\rho_k \geq 10^{20}$.

3.2 RAlgencan

Implementamos o Algoritmo 2.2 no quadro geral de Algencan, o qual denominamos RAlgencan, mantendo as opções-padrão, bem como as rotinas utilizadas na resolução dos subproblemas. Usamos a versão de Algencan disponível em março de 2007. As tolerâncias de viabilidade e otimalidade foram $\varepsilon_{feas} = 10^{-8}$ e $\varepsilon_{opt} = 10^{-4}$. Como ponto inicial para o algoritmo que resolve o subproblema k usamos o ponto de referência $\bar{x}^{k-1}$. Desta forma garantimos o cumprimento de (2.16).

Para o Algoritmo 2.2 se o ponto inicial é viável e os multiplicadores são escolhidas de forma que a complementaridade ocorre, temos $R_0 = 0$ e a condição (2.10) pode ser muito restritiva para atualizar o ponto de referência. Neste caso, é conveniente substituir (2.10) por

$$R_k = \min\{R_{tol}, R_1, \dots, R_k\}, \quad (3.4)$$

onde $R_{tol} > 0$ é uma tolerância inicial. O uso de (3.4), no lugar de (2.10) não afeta os resultados de convergência apresentados no Capítulo 2.

Foi empregada a seguinte estratégia para a atualização do parâmetro de regularização:

- Usamos (3.4) com $R_{tol} = \max\{R_0, 1\}$ e $\gamma_1 = 0$.
- Se $R_k = \min\{R_{tol}, R_1, \dots, R_k\}$ definimos $\gamma_{k+1} = 0$. Senão, definimos

$$\gamma_{k+1} = \min\{\beta R_k, \gamma_k + 1\}. \quad (3.5)$$

Em problemas bem comportados Algencan satisfaz $R_{tol} \geq R_k \geq R_{k+1}$ para a maioria das iterações. Esta é a razão pela qual o desempenho de RAlgencan é quase idêntico ao desempenho de Algencan nestas situações. Como exemplo, consideremos a família de problemas definida por:

$$\min \sum_{i>j} \frac{1}{\|P_i - P_j\|}$$

sujeito a

$$P_k \in \mathbb{R}^{n_{dim}}, \|P_k\| \leq 1, k = 1, \dots, n_{pun}$$

para diferentes valores de n_{dim} , n_{pun} . Nós escrevemos $x = (P_1, \dots, P_{n_{pun}})$ e definimos o ponto inicial por

$$x_i^0 = \text{sen}(i), i = 1, \dots, n_{pun}.$$

Aplicamos Algencan e RAlgencan para os problemas definidos por $n_{dim} = 3$, $n_{pun} = 10, 20, \dots, 100$. Ambos os algoritmos obtiveram o mesmo resultado em 10 problemas. Em 6 problemas ($n_{pun} = 10, 20, 40, 60, 90, 100$) ambos usaram o mesmo número de iterações. Algencan convergiu em 27 iterações para $n_{pun} = 30$ e em 25 iterações para $n_{pun} = 50$. Nestes problemas RAlgencan gastou 28 iterações. Por outro lado, RAlgencan convergiu em 19 iterações para $n_{pun} = 70$ e em 18 iterações para $n_{pun} = 80$. Algencan precisou de 23 e 21 iterações, respectivamente, para esses problemas.

Consideremos os seguintes 6 problemas, nos quais Algencan apresentou voracidade. Em todos estes problemas o nível mais baixo do conjunto de restrições Ω foi a caixa (Artificial) $[-10^{20}, 10^{20}]^n$.

Problema 1

$$\min \sum_{i=1}^n x_i^3$$

sujeito a

$$g_i(x) \equiv -x_i \leq 0, i = 1, \dots, n.$$

Ponto inicial $(-7, \dots, -7)$, $n = 100$.

Resultados:

- Algencan parou na 9ª iteração, porque não conseguiu melhorar a viabilidade durante 9 iterações consecutivas. A norma final das restrições foi de $\approx 10^{20}$. A valor final da função objetivo de -10^{60} .

- RAlgencan detectou voracidade na 1ª iteração com $f(x^1) = -10^{62}$ e $R_1 \approx 10^{20}$. Então, x^0 foi utilizado como ponto de referência e $\gamma_2 = 1$. A mesma situação se repetiu nas 5 iterações seguintes, em que o parâmetro de regularização foi atualizado. Na iteração 7, utilizamos $\gamma_7 = 6$. A solução do sétimo subproblema foi finalmente aceitável, com $R_7 \approx 0.004$ e $f(x^7) \approx 0.98$. Como consequência, o algoritmo escolheu $\gamma_8 = 0$ e a convergência ocorreu na iteração 9.

Problema 2

$$\begin{aligned}
 \min \quad & -x_1 x_2 x_3 \\
 \text{s.a} \quad & h_1(x) \equiv x_1 - 4.2(\sin(x_4))^2 = 0 \\
 & h_2(x) \equiv x_2 - 4.2(\sin(x_5))^2 = 0 \\
 & h_3(x) \equiv x_3 - 4.2(\sin(x_6))^2 = 0 \\
 & h_4(x) \equiv x_1 + 2x_2 + 2x_3 - 7.2(\sin(x_7))^2 = 0,
 \end{aligned}$$

Ponto inicial (1, 2, 3, 4, 5, 6, 7).

Resultados:

- No ponto inicial, temos $f(x^0) \approx -6$ e $R_0 \approx -7.9$. A primeira iteração de Algencan apresentou voracidade, obtendo $f(x^1) \approx -10^6$, $R_1 \approx 10^{20}$. Algencan não conseguiu melhorar a viabilidade nas iterações seguintes, apesar do crescimento dos parâmetros de penalidade. O algoritmo parou na iteração 25, devido a um parâmetro de penalidade inaceitavelmente grande.
- Para a primeira iteração, $\gamma_1 = 0$, RAlgencan obteve, obviamente, o mesmo iterando x^1 como Algencan fez. Assim, o algoritmo usou x^0 como ponto de referência e aumentou o parâmetro de regularização. Com $\gamma_2 = 1$, x^2 , era ainda próximo de x^1 , mas, com $\gamma_3 = 2$, era bastante razoável o iterando x^3 computado, com $f(x^3) \approx -2.76$ e $R_3 \approx 0.12$. Como consequência, RAlgencan escolheu $\gamma_4 = 0$. No entanto, devido à deterioração moderada

da viabilidade, após 4 iterações, foi necessário tomar $\gamma_8 = 1$ e $\gamma_9 = 2$. Isto foi suficiente para uma boa melhora de R_k , de modo que $\gamma_{10} = \gamma_{11} = 0$ e os critérios de convergência foram cumpridos em x^{11} .

Problema 3

$$\begin{aligned} \min \quad & -x_1 x_2^3 \\ \text{s.a} \quad & h_1(x) \equiv x_1 x_2 - 4 \operatorname{sen}^2(x_1) = 0, \end{aligned}$$

Ponto inicial $(1, 1)$.

Resultados:

- Algencan encontra voracidade, e computa uma grande perda de factibilidade na primeira iteração externa. Portanto, o parâmetro de penalidade é aumentado muitas vezes. Após cada aumento do parâmetro de penalidade, a viabilidade é ligeiramente melhorada. No entanto, após 40 iterações o método pára devido ao parâmetro de penalidade ser demasiadamente grande. A norma final das restrições é maior que 10^{20} .
- Após a primeira iteração com voracidade, RAlgencan utiliza como ponto de referência x^0 associado com $\gamma_2 = 1$ como parâmetro de regularização. A segunda iteração também apresentou voracidade e γ_k precisou ser aumentado mais uma vez. Com $\gamma_3 = 2$ a voracidade foi superada, o algoritmo obteve $f(x^3) \approx -43$ e $R_3 \approx 0.46$. A partir desta iteração, a precisão melhorou monotonicamente, e o parâmetro de regularização foi zero. Os critérios de convergência foram cumpridos na iteração 18, com $f(x^{18}) \approx -30$.

Problema 4

$$\begin{aligned} \min \quad & -x_1 e^{-x_1 x_2} \\ \text{s.a} \quad & h_1(x) \equiv -(x_1 + 1)^3 + 3(x_1 + 1)^2 - 1.5 + x_2 = 0. \end{aligned}$$

Ponto inicial $(1, -1.5)$.

Resultados:

- Algencan exibiu voracidade na primeira iteração, com $f(x_1) = -\infty$ e $R_1 \approx 10^{12}$. O parâmetro penalidade foi aumentado durante 9 iterações consecutivas sem modificação do presente estado dos fatos. Divergência foi declarada na iteração 21.
- A primeira iteração apresentou voracidade, por isso, RAlgencan aumentou o parâmetro de regularização. O ponto de referência para a segunda iteração foi o ponto inicial x^0 . Como consequência, a segunda iteração era menos infactível do que a primeira. No entanto, a sua infactibilidade é substancialmente maior do que a do ponto inicial, e o parâmetro regularização aumentou. O aumento de γ_k continuou até a iteração 7. Com $\gamma_7 = 6$, obtivemos $f(x^7) \approx -2.3$ e $R_7 \approx 10^{-4}$. A partir de então, o algoritmo utilizou $\gamma_k = 0$. A convergência ocorreu na iteração 10, com $f(x^{10}) \approx -2.28$.

Problema 5

$$\begin{aligned} \min \quad & -\sum_{i=1}^n (x_i^8 + x_i) \\ \text{s.a} \quad & \sum_{i=1}^n x_i^2 \leq 1. \end{aligned}$$

Ponto inicial $(0.1, \dots, 0.1)$, $n = 50$.

Resultados:

- Para o ponto inicial, temos $f(x^0) \approx -0.5$ e $R_0 = 0$. A primeira iteração de Algencan leva para $f(x^1) \approx -10^{99}$, $R_1 \approx 10^{26}$. A situação não foi alterada nas 20 iterações seguintes e a divergência foi declarada na iteração 21.
- Após a primeira iteração com voracidade, Ralgencan tomou como ponto de referência x_0 e, para o segundo subproblema, $\gamma_2 = 1$. Como resultado, foi corrigida a voracidade, obtendo $f(x^2) \approx -8$ e $R_2 \approx 0.29$. Nas iterações seguintes do algoritmo utilizamos $\gamma_k = 0$. A convergência ocorreu na iteração 15 com $f(x^{15}) \approx -7.07$.

Problema 6

$$\begin{aligned} \min \quad & \sum_{i=1}^n \varphi(x_i) \\ \text{s.a} \quad & \sum_{i=1}^n x_i^2 \leq 1, \end{aligned}$$

onde $\varphi(t) = \log(\cos(t))$ se $\cos(t) > 0$; $\varphi(t) = -10^{30}$ caso contrário. Ponto inicial $(1/n, \dots, 1/n)$, $n = 100$.

Resultados:

A peculiaridade deste problema é que a função objetivo é definida arbitrariamente como $f(x) = -10^{30}$ nos pontos x onde, em princípio, $f(x)$ não seria definida. Todos estes pontos são infactíveis.

- Para o ponto inicial (factível) temos $f(x^0) \approx -0.005$ e $R_0 = 0$. A primeira iteração computada por Algencan forneceu $f(x^1) = -10^{30}$ e $R_1 \approx 250$. Em todas as subseqüentes iterações de Algencan o valor da função objectivo foi -10^{30} . A factibilidade-complementaridade medida por R_k oscilou em torno de 250 e divergência foi declarada na iteração 22.
- A segunda iteração do RAlgencan foi calculada com $\gamma_2 = 1$ e não foi capaz de corrigir a voracidade inicial. No entanto, com $\gamma_3 = 2$ e preservando x^0 como ponto de referência, foram obtidos $f(x^3) \approx 0.002$ e $R_3 = 0$. Com $\gamma_4 = 0$, a 4ª iteração apresentou voracidade novamente, obtendo $f(x^4) = -10^{30}$ e $R_4 \approx 246$. A voracidade foi corrigida na iteração 5, com $\gamma_5 = 1$, $f(x^5) = 0.5$ e $R_5 \approx 10^{-4}$. Uma vez que R_5 não foi o mínimo de $\{R_{tol}, R_1, \dots, R_5\}$ o algoritmo escolheu, de acordo com (3.5), $\gamma_6 = \beta R_5 \approx 0.102$. Esta ligeira regularização produziu novamente voracidade, na iteração $f(x^6) = -10^{-30}$, $R_6 \approx 246$. Então, o algoritmo definiu $\gamma_7 = 1.102$. A voracidade foi corrigida mais uma vez, permitindo a escolha do algoritmo para $\gamma_8 = 0.006$ devido à limitação de βR_k . Houve alternância entre voracidade e não voracidade até iteração 12. Com $\gamma_{12} = 1$ temos $f(x^{12}) \approx -0.5$ e $R_{12} \approx 10^{-8}$. Finalmente, a convergência foi decretada na iteração 14.

3.3 RAlgencan*

Nesta seção implementamos o Algoritmo que denominamos RAlgencan*, o qual foi baseado no Algoritmo 2.1, e sua implementação foi feita no quadro geral de Algencan, mantendo as opções-padrão, bem como as rotinas utilizadas na resolução dos subproblemas. Usamos a versão de Algencan disponível em março de 2007. As tolerâncias de viabilidade e otimalidade foram $\varepsilon_{feas} = 10^{-4}$ e $\varepsilon_{opt} = 10^{-4}$.

Em RAlgencan* o ponto inicial para o subproblema $k + 1$ é o iterando corrente x^k , exceto talvez em um número finito de iterações, nas quais poderemos utilizar o ponto de referência $\bar{x}^k$ como ponto inicial.

Seguem agora os detalhes de implementação de RAlgencan*:

Atualização para o ponto de referência:

- Se $R_k \leq \max\{\min\{R_0, \dots, R_k\}, 10^{-4}\}$

$$\bar{x}^k = x^k,$$

caso contrário:

$$\bar{x}^k = \bar{x}^{k-1}.$$

Isto garante que todo ponto obtido que é factível (dentro da precisão exigida), será o próximo ponto de referência.

A sequência γ_k fica definida da seguinte maneira:

Parâmetros: $\Theta_1 = 0$, $\gamma_1 = \rho_1$.

Atualização de γ_k .

$$\bar{R} = \min\{R_0, \dots, R_{k-1}\}$$

- Se $R_k \leq \max\{(\sqrt{0,1})\bar{R}, 10^{-4}\}$, por duas iterações consecutivas, definimos

$$\gamma_{k+1} = 0. \tag{3.6}$$

- Se

$$R_k \leq (1.5) \max\{\bar{R}, 10^{-4}\}, \text{ ou } \Theta_1 \geq 10 \quad (3.7)$$

escolhemos

$$\gamma_{k+1} = \frac{\gamma_k}{1 + \bar{R}} 10^{-1} \max\{R_k, 10^{-4}\}. \quad (3.8)$$

Neste caso o iterando corrente será o ponto inicial para o próximo subproblema.

- Caso contrário,

$$\Theta_1 = \Theta_1 + 1;$$

$$\gamma_{k+1} = \min \left\{ \max\{10^{-4}, \gamma_k\} \frac{R_k}{1 + \bar{R}}, \gamma_{max} \right\} \quad (3.9)$$

O ponto inicial para o próximo subproblema será o ponto de referência, também utilizaremos os multiplicadores do ponto de referência.

A decisão $\gamma_k = 0$ é tomada quando, em duas iterações consecutivas, a medida de in-factibilidade R^k melhorar uma ordem de grandeza. Julgamos que, quando isso acontece, a regularização não é mais necessária e, portanto, o de parâmetro regularização será nulo.

De acordo com a decisão (3.8) o parâmetro de regularização pode diminuir se houver uma pequena deterioração na viabilidade em relação à iteração anterior.

A escolha (3.9) é a única que permite aumentar o parâmetro de regularização γ_k . Fazemos esta escolha quando (3.7) não é satisfeita, indicando que estamos encontrado dificuldades para obter a viabilidade e, por isso, temos que dar mais peso ao parâmetro que penaliza a distância ao ponto de referência. No entanto γ_k é atualizado como em (3.9) no máximo dez vezes. A razão é que geralmente não vale a pena repetir uma correção que parece não estar funcionando.

Assim exceto em um número finito de vezes, γ_k se atualiza como em (3.8). Logo $\gamma_{k+1} \leq \gamma_k$ para todo k suficientemente grande, assim se existir $K \subset \infty \mathbb{N}$ tal que

$$\lim_{k \in K} R_k = 0, \text{ então } \lim_{k \rightarrow \infty} \gamma_k = 0.$$

Portanto os resultados teóricos sobre o Algoritmo 2.1 vistos no Capítulo 2 são validos para o Algoritmo RAlgencan*.

Testamos os mesmos 6 problemas da seção anterior, nos quais Algencan apresentou voracidade, e RAlgencan obteve sucesso. Em todos estes problemas o nível mais baixo do conjunto de restrições Ω foi a caixa (Artificial) $[-10^{20}, 10^{20}]^n$.

Problema 1

$$\min \sum_{i=1}^n x_i^3$$

sujeito a

$$g_i(x) \equiv -x_i \leq 0, \quad i = 1, \dots, n.$$

Ponto inicial $(-7, \dots, -7)$, $n = 100$.

Resultado:

- Para a primeira iteração $\gamma_1 = 10$, mesmo assim RAlgencan* detectou voracidade com $f(x^1) = -10^{62}$ e $R_1 \approx 10^{20}$. Então, x^0 foi utilizado como ponto de referência e $\gamma_2 = 10^4$. O método não apresentou mais voracidade e γ_k decresceu até zero. Como consequência a convergência ocorreu na iteração 9.

Sequência γ_k :

$$\{10, 10^4, 8759.66, 7665.6, 6621.4, 4945.59, 1077.14, 0\}$$

Problema 2

$$\min \quad -x_1 x_2 x_3$$

$$s.a \quad h_1(x) \equiv x_1 - 4.2(\sin(x_4))^2 = 0$$

$$h_2(x) \equiv x_2 - 4.2(\sin(x_5))^2 = 0$$

$$h_3(x) \equiv x_3 - 4.2(\sin(x_6))^2 = 0$$

$$h_4(x) \equiv x_1 + 2x_2 + 2x_3 - 7.2(\sin(x_7))^2 = 0,$$

Ponto inicial (1, 2, 3, 4, 5, 6, 7).

Resultado:

- Para primeira iteração, $\gamma_1 \approx 1.6$, RAlgencan* não apresentou voracidade até a iteração 2, na iteração 3 RAlgencan* apresentou voracidade, então $\gamma_8 = 10^4$, e o ponto de referência x^2 foi utilizado, γ_k decresceu até atingir zero na iteração 9, os critérios de convergência foram cumpridos em x^{10} .

Sequência γ_k :

$$\{1.60, 0.23, 0.05, 10^4, 3141, 906.8, 74.83, 12.04, 0.22, 0\}$$

Problema 3

$$\begin{aligned} \min \quad & -x_1 x_2^3 \\ \text{s.a} \quad & h_1(x) \equiv x_1 x_2 - 4 \operatorname{sen}^2(x_1) = 0, \end{aligned}$$

Ponto inicial (1, 1).

Resultado:

- Após a primeira iteração com voracidade, RAlgencan* utiliza como ponto de referência x^0 associado com $\gamma_2 = 10^4$ como parâmetro de regularização. O método não apresenta voracidade até a iteração 6, com $\gamma_7 = 0$ o método apresenta voracidade, e então $\gamma_8 = 10^4$ e utilizamos o ponto de referência x^6 . O método não apresentou mais voracidade e os critérios de convergência foram cumpridos na iteração 13, com $f(x^{18}) \approx -30$.

Sequência γ_k :

$$\{5.95, 10^4, 6438.6, 4096.5, 2275.34, 379.17, 0, 10^4, 823.92, 19.46, 0.242, 0.001, 0\}$$

Problema 4

$$\begin{aligned} \min \quad & -x_1 e^{-x_1 x_2} \\ \text{s.a} \quad & h_1(x) \equiv -(x_1 + 1)^3 + 3(x_1 + 1)^2 - 1.5 + x_2 = 0. \end{aligned}$$

Ponto inicial $(1, -1.5)$.

Resultado:

- A primeira iteração apresentou voracidade, por isso, RAlgencan* aumentou o parâmetro de regularização $\gamma_2 = 10^4$ o ponto de referência para a segunda iteração foi o ponto inicial x^0 . Como consequência, o método superou o problema da voracidade, a convergência ocorreu na iteração 10, com $f(x^{10}) \approx -2.28$.

Sequência γ_k :

$$\{10, 10^4, 4992.74, 2482.22, 1176.78, 17.01, 10.44, 0.172, 5.14 \times 10^{-5}, 0\}$$

Problema 5

$$\begin{aligned} \min \quad & -\sum_{i=1}^n (x_i^8 + x_i) \\ \text{s.a} \quad & \sum_{i=1}^n x_i^2 \leq 1. \end{aligned}$$

Ponto inicial $(0.1, \dots, 0.1)$, $n = 50$.

Resultado:

- RAlgencan* apresentou voracidade na terceira iteração e obtemos para $f(x^3) \approx -10^{99}$, $R_3 \approx 10^{26}$. Ralgencan* tomou como ponto de referência x^2 e para o quarto subproblema $\gamma_4 = 10^4$, nas iterações seguintes γ_k decresceu até $\gamma_7 = 0.003$, na iteração 8, encontramos novamente voracidade, assim $\gamma_8 = 10^4$, o problema da voracidade foi superado. A convergência ocorreu na iteração 12 com $f(x^{12}) \approx -7.07$.

Sequência γ_k :

$$\{10, 1573.01, 0.1573, 10^4, 1, 33.28, 0.0033, 10^4, 1, 0.35, 3.5 \times 10^{-5}, 3.5 \times 10^{-9}\}$$

Problema 6

$$\begin{aligned} \min \quad & \sum_{i=1}^n \varphi(x_i) \\ \text{s.a} \quad & \sum_{i=1}^n x_i^2 \leq 1, \end{aligned}$$

onde $\varphi(t) = \log(\cos(t))$ se $\cos(t) > 0$; $\varphi(t) = -10^{30}$ caso contrário. Ponto inicial $(1/n, \dots, 1/n)$, $n = 100$.

Resultado:

- Para o ponto inicial (factível) temos $f(x^0) \approx -0.005$ e $R_0 = 0$. RAlgencan* apresentou voracidade na segunda iteração. A factibilidade-complementaridade medida por R_k variou entre 0 e 250, e a sequência γ_k variou entre 10^{-4} e 10^4 até a iteração 20, a partir de então γ_k não aumentou mais, e a divergência foi declarada na iteração 30.

A seguir apresentamos uma tabela que resume o desempenho de Algencan, RAlgencan e RAlgencan* nos problemas com voracidade:

Problema	n	m	Método	Resultado	Iterações externas
Problema 1	100	100	Algencan	Divergiu	-
			RAlgencan	Convergiu	9
			RAlgencan*	Convergiu	9
Problema 2	7	4	Algencan	Divergiu	-
			RAlgencan	Convergiu	11
			RAlgencan*	Convergiu	10
Problema 3	2	1	Algencan	Divergiu	-
			RAlgencan	Convergiu	18
			RAlgencan*	Convergiu	18
Problema 4	2	1	Algencan	Divergiu	-
			RAlgencan	Convergiu	10
			RAlgencan*	Convergiu	10
Problema 5	50	1	Algencan	Divergiu	-
			RAlgencan	Convergiu	15
			RAlgencan*	Convergiu	12
Problema 6	100	1	Algencan	Divergiu	-
			RAlgencan	Convergiu	14
			RAlgencan*	Divergiu	-

Tabela 3.1: Comparação entre Algencan, RAlgencan e RAlgencan*.

Na Tabela 3.1 n e m são respectivamente o número de variáveis e de restrições para cada problema.

Com exceção do problema 6, RAlgencan* resolveu todos os problemas que apresentavam voracidade. Portanto mesmo com as alterações o método continuou robusto para problemas com voracidade.

RAlgencan* utiliza desde o início γ_k diferente de zero, e decreta $\gamma_k = 0$ se, e somente se, γ_k atualiza-se como em (3.6). Assim, devido à regularização, os subproblemas para RAlgencan* são melhor condicionados, se comparados aos subproblemas de Algencan, geralmente esse tipo de melhoria é acompanhada de um ganho na velocidade de convergência do método. Testamos 15 problemas da coleção LaCumparsita (www.ime.usp.br/~egbirgin/collection) onde Algencan não apresentou voracidade.

Confirmamos que a introdução da regularização não prejudicou o comportamento de Algencan. Em todos os problemas encontramos as mesmas soluções, com quase o mesmo número de iterações externas. Observamos também que RAlgencan* apresentou uma melhora significativa para 7 problemas em relação ao número de iterações internas e de avaliações de função, em 2 problemas RAlgencan* consumiu significativamente mais iterações internas e avaliações de função que Algencan, nos demais problemas os métodos apresentaram praticamente o mesmo desempenho, como podemos observar a comparação de Algencan versus RAlgencan* na tabela a seguir:

Problema	n	m	Método	It ext	It int	Aval.	f
America	264	17	Algencan	9	36	55	3.25
			RAlgencan*	9	37	56	3.25
Cache	384	258	Algencan	5	535	1534	-1.0
			RAlgencan*	8	1836	5894	-1.0
Hard-spheres nd=3, np=18	55	171	Algencan	7	114	1269	0.649
			RAlgencan*	7	117	1130	0.649
Hard-spheres nd=3, np=200	601	20100	Algencan	11	1794	18101	0.966
			RAlgencan*	10	1312	13366	0.966
Hard-spheres nd=4, np=250	1001	31375	Algencan	7	2490	25489	0.895
			RAlgencan*	7	1312	13366	0.895
Balls nd=3, np=20	61	380	Algencan	8	226	2400	5.87
			RAlgencan*	6	583	7520	5.85
Balls nd=5, np=50	251	2450	Algencan	6	1084	12389	4.06
			RAlgencan*	6	838	9609	4.1
Balls nd=6, np=100	601	9900	Algencan	7	1630	16120	4.40
			RAlgencan*	7	1184	11840	4.44
Contour I np = 100	100	98	Algencan	9	119	3086	1.10
			RAlgencan*	10	44	712	1.10
Contour I np = 2000	2000	1998	Algencan	13	32	59	1.88
			RAlgencan*	13	32	59	1.88
Condor nm=100, pm=50, qm=10	100	98	Algencan	14	98	133	6.15
			RAlgencan*	14	96	131	6.15
Condor nm = 500, pm=300, qm=10	500	498	Algencan	15	1734	1766	0.038
			RAlgencan*	15	1800	1831	0.038
vi8EA	23	30	Algencan	11	3650	11004	17.9
			RAlgencan*	11	3085	12023	18.0
simfock kd = 3, nd=2	9	19	Algencan	8	54	137	7.92
			RAlgencan*	6	46	83	7.92

Problema	n	m	Método	It ext	It int	Aval.	f
simfoc	289	579	Algencan	14	81	139	19.3
kd = 17, nd=8			RAlgencan*	14	78	136	19.3

Tabela 3.2: Comparação de Algencan versus RAlgencan*.

Na Tabela 3.2 n e m são respectivamente o número de variáveis e de restrições para cada problema, It ext e It int são respectivamente o número de iterações externas e internas que os métodos consumiram, para Algencan Aval. é o número de avaliações da função Lagrangiano aumentado, para RAlgencan* é o número de avaliações da função Lagrangiano aumentado regularizado, e f é o valor da função objetivo na solução encontrada por cada método.

CAPÍTULO 4

Conclusões

O fenômeno de voracidade é definido neste trabalho como a tendência de alguns métodos de programação não linear de buscar pontos demasiadamente infactíveis com valores de função muito pequenos, em geral, nas primeiras iterações. Este fenômeno pode ocorrer mesmo quando temos boas aproximações iniciais para a solução do problema.

A solução proposta neste trabalho para diminuir a voracidade, foi substituir o Lagrangiano aumentado pelo Lagrangiano aumentado regularizado, utilizando o ponto de referência. Mostra-

mos que o método regularizado preserva as propriedades de convergência global do método Lagrangiano aumentado. Do ponto de vista prático, mostramos que as implementações para o método regularizado RAlgencan e RAlgencan* parecem não prejudicar o comportamento de Algencan, quando este algoritmo funciona bem. Os testes também forneceram indícios de que, em situações onde Algencan falhar devido à voracidade, a regularização pode ajudar o algoritmo a encontrar a trajetória correta, para escapar da voracidade.

Outra consequência da estratégia de regularização é o condicionamento dos subproblemas,

que é melhorado. Normalmente, este tipo de melhoria é acompanhada por uma diminuição da velocidade de convergência. No entanto, este fenômeno não foi observado nos experimentos numéricos de RAlgencan utilizando problemas sem voracidade. Isto é devido à particular estratégia de atualização do parâmetro de regularização, que inicia $\gamma_1 = 0$, e γ_k tende a zero rapidamente quando a viabilidade-complementaridade melhora. Já os experimentos com RAlgencan* mostraram que o método parece apresentar sim uma tendência a consumir menos iterações internas, se comparado a Algencan, uma vez que nesta implementação o parâmetro de regularização é diferente de zero na primeira iteração, e sua particular atualização faz com que a regularização aja efetivamente em mais iterações, mesmo em problemas que não apresentam o fenômeno da voracidade.

BIBLIOGRAFIA

- [1] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. *Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification*. Mathematical Programming 112, pp 5-32, 2008.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt. *On Augmented Lagrangian Methods with general lower-level constraints*, SIAM Journal on Optimization 18 (2007), pp. 1286-1309.
- [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] M. S. Bazaraa, H. D. Sherali, C. M. Shetty. *Nonlinear programming: theory and algorithms*. 2.ed. Nova Iorque, Wiley, 1993.
- [5] D. P. Bertsekas. *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.

- [6] E. V. Castelani, A. L. M. Martinez, J. M. Martínez, B. F. Svaiter, *Addressing the greediness phenomenon in nonlinear programming by means of proximal augmented Lagrangians*. Computational Optimization and Applications, 2010 (por aparecer).
- [7] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.
- [8] A. V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, 1968.
- [9] R. Fletcher, *Practical Methods of Optimization*, Academic Press, London, 1987.
- [10] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications 4 (1969), pp. 303–320.
- [11] A. N. Iusem, *Augmented Lagrangian methods and proximal point methods for convex optimization*, Investigación Operativa 8 (1999), pp. 11-49.
- [12] E. L. Lima. *Curso de Análise*, Projeto Euclides. Rio de Janeiro, IMPA, 1994.
- [13] E. L. Lima. *Análise no Espaço $\mathbb{R}^n$* , Edgar Blücher Ltda.
- [14] J. R. Munkres. *Topology*, Prentice Hall, Upper Saddle River, NJ 2000.
- [15] D. G. Luenberger. *Linear and nonlinear programming*. 2.ed. Reading, Addison-Wesley, 1984.
- [16] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
- [17] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher (ed.), Academic Press, New York, NY, pp. 283–298, 1969.
- [18] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM Journal on Optimization 10 (2000), pp. 963–981.

- [19] R. T. Rockafellar, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Mathematical Programming 5 (1973) pp. 354–373.
- [20] R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM Journal on Control and Optimization 12 (1974), pp. 268–285.
- [21] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research 1 (1976), 97–116.
- [22] R. T. Rockafellar, *Lagrange multipliers and optimality*, SIAM Review 35 (1993), pp. 183–238.