

Universidade Estadual de Campinas

Instituto de Matemática, Estatística e

Computação Científica - IMECC

Departamento de Matemática

B C

Construção e análise de um algoritmo PQS globalmente convergente

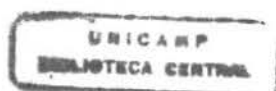
Dissertação de Mestrado

Roberto Carlos Antunes Thomé

Orientadora: Prof.^a Dr.^a Sandra Augusta Santos

Campinas

2001

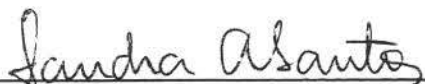


2001/15788

CONSTRUÇÃO E ANÁLISE DE UM ALGORITMO PQS GLOBALMENTE CONVERGENTE

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Roberto Carlos Antunes Thomé e aprovada pela comissão julgadora.

Campinas, 09 de março de 2001


Prof^a Dra. Sandra Augusta Santos
Orientadora

Banca Examinadora:

1. Sandra Augusta Santos
2. Roberto Andreani
3. Francisco de Assis Magalhães Gomes Neto

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

Dissertação de Mestrado defendida em 09 de março de 2001 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



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



Prof (a). Dr (a). ROBERTO ANDREANI



Prof (a). Dr (a). FRANCISCO DE ASSIS MAGALHÃES GOMES NETO

*À memória querida de meu avô
Antônio Antunes Filho*

Agradecimentos

A Deus, por estar tão presente nos momentos mais difíceis;

À minha noiva Jacqueline, uma companheira que soube compreender a distância com muito amor e carinho;

Aos meus pais Paulo e Vilma, por todo apoio e incentivo;

Aos meus irmãos Valéria, Paulo e Eduardo, pelo espírito de união;

Aos meus grandes amigos, pelos conselhos e pelos momentos de alegrias;

Finalmente, agradeço à minha orientadora Sandra não apenas pelo auxílio técnico, mas pelo encorajamento, paciência e por acreditar no meu potencial.

Summary

Sequential quadratic programming (SQP) methods are extensions of Newton's method for constrained nonlinear programming problems. In this work, an SQP-based algorithm is analysed for the general minimization problem in standard form. The merit function is of the augmented Lagrangian type, with a non-monotone updating scheme for the penalty parameter. Proofs that the algorithm is well defined and globally convergent are included. A strategy for dealing with the quadratic subproblems is introduced, based on bound-constrained minimization. Two approaches for choosing the Hessian of the quadratic model are suggested. Up-to-date bibliographical notes are included in the introduction.

◊ Key words: SQP Algorithms; global convergence; quadratic subproblems; augmented Lagrangian; bound-constrained minimization.

Sumário

Resumo	vi
Summary	vii
Introdução e notas bibliográficas	1
1 Motivação e elementos para a construção do algoritmo	3
1.1 Definições e condições de otimalidade	3
1.2 Ingredientes básicos do algoritmo	6
2 O Algoritmo PQS e a boa definição	13
2.1 O Algoritmo PQS	13
2.2 O algoritmo está bem definido	16
3 Propriedades de convergência global	20
3.1 Todo ponto limite é φ -estacionário	21
3.2 O algoritmo encontra um ponto crítico	29
4 Uma abordagem para os subproblemas	49
4.1 Resultado de equivalência	49
4.2 Os subproblemas do Algoritmo PQS	53
4.3 Escolhas para a matriz do modelo quadrático	55
Comentários finais	63
Referências bibliográficas	65

Introdução e notas bibliográficas

Dentre as estratégias para a minimização de problemas com restrições não-lineares gerais, podemos destacar três filosofias básicas: preservar a factibilidade, procurando lidar com a não linearidade das restrições (métodos de eliminação e direções factíveis), incorporar as restrições na função objetivo (métodos de penalização) ou linearizar o conjunto factível e simultaneamente construir um modelo quadrático para a função objetivo. É nesta última classe que se enquadram os algoritmos de programação quadrática sequencial (PQS).

O método PQS foi proposto pela primeira vez em 1963 por Wilson [69] em sua tese de doutorado e posteriormente desenvolvido por Garcia-Palomares e Mangasarian [30], Han [40, 41] e Powell [58, 59, 60] na década de 70. Para uma revisão bibliográfica, ver Boggs e Tolle [5]. O método de programação linear sequencial é descrito por Fletcher e Sainz de la Maza [22], e também pode ser encontrado resumidamente no final da seção 18.1 de [56]. A relação entre o método PQS e métodos de Lagrangeano aumentado é explorada por Tapia [65]. Procedimentos distintos para lidar com restrições lineares inconsistentes são descritos por Powell [58] e por Nocedal e Wright [56, seção 18.3].

Os métodos PQS com Hessianas reduzidas foram possivelmente propostos pela primeira vez por Murray e Wright e Coleman e Conn (ver [5]). Desde então, várias publicações nesse contexto têm surgido, e os algoritmos propostos têm se mostrado úteis em aplicações, tal como controle de processos químicos e otimização de trajetória.

Algumas análises mostram que várias, mas não todas, das boas propriedades de atualização BFGS são preservadas por atualizações BFGS com salvaguardas e controle do passo. Experimentos numéricos revelando as desvantagens das aproximações são relatados por Powell [61].

A maioria dos métodos PQS são infactíveis, significando que nem o ponto inicial nem qualquer outro ponto das iterações subsequentes precisam ser factíveis. Isso pode ser vantajoso quando o problema contém, de maneira significativa, restrições não-lineares, pois pode ser computacionalmente caro manter-se dentro da região factível. Entretanto, em algumas aplicações, as funções objetivo nem sempre estão definidas fora da região de factibilidade. Para lidar com essa situação, métodos PQS factíveis foram desenvolvidos; ver, por exemplo, Panier e Tits [57].

Vários programas PQS tratam as restrições lineares de maneira diferente das restrições não-lineares. Há um primeiro estágio onde os iterandos se tornam factíveis com relação a todas as restrições lineares, e todos iterandos subsequentes permanecem factíveis com relação a essas restrições. Esse é o caso de NPSOL [35] e SNOPT [33]. Para tratar de restrições inconsistentes ou quase dependentes, esses programas trabalham as restrições

do problema relaxadas.

O método de ponto interior para programação não-linear descrito por Byrd, Hribar e Nocedal [9] usa um algoritmo PQS com busca linear para resolver os subproblemas com restrições de igualdade que surgem nos métodos de barreira. Estratégias de correção de segunda ordem foram propostas por Coleman e Conn [13], Fletcher [20], Gabay [29] e Mayne e Polak [54]. A técnica *watchdog* foi proposta por Chamberlain et al. [11]. Procedimentos para calcular uma sequência de matrizes do espaço nulo, suavemente modificada, são descritos por Coleman e Sorensen [14] e Gill et al. [34]. Resultados de convergência mais realistas vem sendo provados para métodos quase-Newton com Hessiana reduzida; ver Byrd e Nocedal [10].

Trabalhos recentes apontam para um interesse especial para métodos PQS no contexto teórico [18, 28, 48, 49, 51, 55, 62, 70], computacional de grande porte [6, 7, 32, 39, 46, 64, 67] e aplicações em problemas de controle, otimização de sistemas dinâmicos, entre outras [19, 31, 42, 44, 47, 63, 68].

Nosso trabalho está fundamentado no desenvolvimento de [37], onde os autores apresentam um algoritmo baseado em programação quadrática sequencial para resolver o problema geral de otimização em sua forma padrão. Em [37], os autores estabelecem propriedades de convergência do algoritmo, que foram reproduzidas neste trabalho com pequenas modificações para tornar esta obra tão didática quanto possível. Também estamos nos apoiando fortemente no capítulo 12 de [52], que nos ajudou a fundamentar a motivação feita no capítulo 1. Neste capítulo, algumas definições necessárias são apresentadas. Outras referências utilizadas pontualmente são mencionadas ao longo do texto. No capítulo 2 está descrito com detalhes o algoritmo de programação quadrática sequencial proposto em [37], assim como a sua prova de boa definição. O capítulo 3 apresenta as propriedades de convergência global do algoritmo. À medida que se fazem necessárias, algumas hipóteses são adicionadas. No capítulo 4 é discutida a resolução dos subproblemas presentes no algoritmo com base na estratégia apresentada em [24], assim como possíveis escolhas para a matriz que compõe o modelo quadrático. Concluímos o trabalho com comentários finais e as referências mencionadas ao longo do texto.

Capítulo 1

Motivação e elementos para a construção do algoritmo

Um princípio básico em métodos numéricos é a conversão de problemas complexos numa sequência de subproblemas mais simples. O objetivo principal deste trabalho está em resolver o problema geral de programação não-linear em sua forma padrão:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & h(x) = 0, \\ & l \leq x \leq u, \end{array} \quad (1.1)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções *suaves* (C^1 ou C^2) e $m \leq n$. Os vetores l e u podem ter componentes $-\infty$ ou $+\infty$, respectivamente. Nesses casos, o símbolo $\leq$ deve ser interpretado como $<$. Sabemos que, de fato, qualquer problema de otimização com igualdades ou desigualdades pode ser levado à forma (1.1) pela introdução de variáveis de folga. Por exemplo, toda restrição do tipo $g(x) \leq 0$, onde $g : \mathbb{R}^n \rightarrow \mathbb{R}$ pode ser transformada em $g(x) + z = 0$, com $z \geq 0$. É claro que acrescentar uma variável z ao problema pode criar desvantagens. Por outro lado, o tratamento de restrições na forma padrão é geralmente mais simples e muitos algoritmos eficientes, com *software* bem desenvolvido, se baseiam na forma padrão (ver, por exemplo, [16]).

1.1 Definições e condições de otimalidade

Faremos a seguir algumas considerações teóricas importantes relativas ao problema geral de otimização (1.1) que queremos resolver.

Denotando o conjunto factível de (1.1) por $S = \{x \in \mathbb{R}^n \mid h(x) = 0, l \leq x \leq u\}$, dizemos que $x \in S$ é um *ponto factível*. Para $x \in S$ definimos um conjunto de índices das canalizações ativas em x por

$$\mathcal{I}_l(x) = \{i_1, i_2, \dots, i_{nl(x)}\}$$

e

$$\mathcal{I}_u(x) = \{i_1, i_2, \dots, i_{nu(x)}\}$$

de tal forma que $j \in \mathcal{I}_l(x) \Leftrightarrow x_j = l_j$ e $j \in \mathcal{I}_u(x) \Leftrightarrow x_j = u_j$.

Para o estabelecimento das condições necessárias de otimalidade do problema (1.1) precisamos da hipótese de regularidade, que definimos a seguir. A regularidade é uma caracterização do conjunto de restrições de forma a excluir a possibilidade de degenerações nos pontos de interesse. Utilizamos a notação e_i para o i -ésimo vetor elementar de $\mathbb{R}^n$, com 1 na i -ésima componente e 0 nas demais.

Definição (Ponto regular): O ponto $x \in S$ é regular se o conjunto

$$\{\nabla h_k\}_{1 \leq k \leq m} \cup \{-e_k\}_{k \in \mathcal{I}_l(x)} \cup \{e_k\}_{k \in \mathcal{I}_u(x)}$$

é linearmente independente (LI), ou, equivalentemente, se as colunas da matriz

$$\mathcal{M}(x) = \begin{pmatrix} h'(x)^T & E_l(x) & E_u(x) \end{pmatrix} \quad (1.2)$$

são LI, onde $h'(x)^T = [\nabla h_1(x) \ \cdots \ \nabla h_m(x)] \in \mathbb{R}^{n \times m}$ é a transposta da matriz Jacobiana,

$$E_l(x) = \begin{bmatrix} -e_{i_1} & \cdots & -e_{i_{nl(x)}} \end{bmatrix} \in \mathbb{R}^{n \times nl(x)} \text{ e } E_u(x) = \begin{bmatrix} -e_{i_1} & \cdots & -e_{i_{nu(x)}} \end{bmatrix} \in \mathbb{R}^{n \times nu(x)}.$$

Pelas características das restrições de canalização, para $x \in S$, os conjuntos $\mathcal{I}_l(x)$ e $\mathcal{I}_u(x)$ são tais que $\mathcal{I}_l(x) \cap \mathcal{I}_u(x) = \emptyset$, $\mathcal{I}_l(x) \subset \{1, \dots, n\}$ e $\mathcal{I}_u(x) \subset \{1, \dots, n\}$. Chamando de $a(x)$ o número total de restrições ativas em x temos $m \leq a(x) \equiv m + nl(x) + nu(x) \leq m + n$. Naturalmente, se $x \in S$ é regular, então $a(x) \leq n$.

As condições de otimalidade para o problema (1.1) são estabelecidas a seguir, e serão especialmente úteis no desenvolvimento deste trabalho.

Teorema 1 [Condições necessárias de primeira ordem (CN1)]

Consideramos o problema (1.1) com $f, h \in C^1$. Seja $x_* \in S$ um ponto regular. Se x_* é um minimizador local de (1.1), então existem $\lambda_* \in \mathbb{R}^m$, $\mu_* \in \mathbb{R}^{nl(x_*)}$ e $\nu_* \in \mathbb{R}^{nu(x_*)}$ tais que

$$\nabla f(x_*) + \sum_{i=1}^m (\lambda_*)_i \nabla h_i(x_*) - \sum_{k=1}^{nl(x_*)} (\mu_*)_k e_{i_k} + \sum_{k=1}^{nu(x_*)} (\nu_*)_k e_{i_k} = 0 \quad (1.3)$$

onde $i_k \in \{1, 2, \dots, n\}$ e

$$\mu_* \geq 0, \text{ e } \nu_* \geq 0. \quad (1.4)$$

Dem. Ver [23] ou [50].

- Obs.:** 1) No caso de $\mathcal{I}_l(x_*)$ ou $\mathcal{I}_u(x_*)$ ser um conjunto vazio, o multiplicador associado μ_* ou ν_* deixa de existir, bem como o termo correspondente na equação (1.3). Esta observação também se aplica aos próximos teoremas.
- 2) Um ponto que verifica CN1 é denominado *estacionário* ou *KKT* (Karush-Kuhn - Tucker).

Teorema 2 [Condições necessárias de segunda ordem (CN2)]

Consideramos o problema (1.1) com $f, h \in C^2$. Seja x_* um minimizador local de (1.1) e um ponto regular. Então existem $\lambda_* \in \mathbb{R}^m$, $\mu_* \in \mathbb{R}^{nl(x_*)}$ e $\nu_* \in \mathbb{R}^{nu(x_*)}$ tais que (1.3) e (1.4) se verificam. Além disso, definindo-se a matriz $E_p \in \mathbb{R}^{n \times n}$ para $p \in \{1, \dots, n\}$ tal que $(E_p)_{i,j} = 1$ se $i = j = p$ e $(E_p)_{i,j} = 0$ caso contrário, então a matriz

$$\nabla^2 \Upsilon(x_*, \lambda_*, \mu_*, \nu_*) \equiv \nabla^2 f(x_*) + \sum_{i=1}^m (\lambda_*)_i \nabla^2 h_i(x_*) - \sum_{k=1}^{nl(x_*)} (\mu_*)_k E_{i_k} + \sum_{k=1}^{nu(x_*)} (\nu_*)_k E_{i_k}, \quad (1.5)$$

onde $i_k \in \{1, 2, \dots, n\}$, verifica

$$y^T \nabla^2 \Upsilon(x_*, \lambda_*, \mu_*, \nu_*) y \geq 0$$

para todo $y \in T \equiv \{y \in \mathbb{R}^n \mid \mathcal{M}(x_*)^T y = 0\}$, com $\mathcal{M}(\cdot)$ definida em (1.2).

Dem. Ver [23] ou [50].

Teorema 3 [Condições suficientes de segunda ordem (CS2)]

Consideramos o problema (1.1) com $f, h \in C^2$. Seja $x_* \in S$ e tal que existem $\lambda_* \in \mathbb{R}^m$, $\mu_* \in \mathbb{R}^{nl(x_*)}$ e $\nu_* \in \mathbb{R}^{nu(x_*)}$ com (1.3) e (1.4) válidas. Suponhamos ainda que a matriz $\nabla^2 \Upsilon(x_*, \lambda_*, \mu_*, \nu_*)$ definida em (1.5) verifique $y^T \nabla^2 \Upsilon(x_*, \lambda_*, \mu_*, \nu_*) y > 0$ para todo $y \in \tilde{T} \equiv \{y \in \mathbb{R}^n \mid h'(x_*)y = 0 \text{ e } y_j = 0, j \in \mathcal{I}_l^o(x_*) \cup \mathcal{I}_u^o(x_*)\}$, onde $\mathcal{I}_l^o(x_*) = \{j \in \mathcal{I}_l(x_*) \mid (\mu_*)_j > 0\}$ e $\mathcal{I}_u^o(x_*) = \{j \in \mathcal{I}_u(x_*) \mid (\nu_*)_j > 0\}$. Então x_* é um minimizador local estrito de (1.1).

Dem. Ver [50].

Uma última definição antes de comentarmos os elementos básicos do algoritmo a ser analisado neste trabalho: seja $\Omega = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$. Dizemos que $z \in \Omega$ é

φ -estacionário se satisfaz as condições de otimalidade de primeira ordem para o problema

$$\begin{array}{ll} \text{minimizar} & \varphi(x) \\ \text{sujeito a} & x \in \Omega, \end{array}$$

onde $\varphi(x) = \frac{1}{2}\|h(x)\|_2^2$.

1.2 Ingredientes básicos do algoritmo

No próximo capítulo será apresentado um algoritmo para resolver o problema (1.1) baseado no método de Programação Quadrática Sequencial (PQS).

Este algoritmo foi proposto em [37], onde os autores estabelecem propriedades de convergência. Nosso trabalho está fundamentado no desenvolvimento de [37]. Estudamos com detalhes o algoritmo proposto e os resultados de convergência obtidos, reproduzindo-os com pequenas modificações nos próximos dois capítulos. Nossa contribuição nesses capítulos é apresentar o algoritmo e as propriedades teóricas de [37] da forma mais didática que conseguimos. Também estamos nos apoiando fortemente no capítulo 12 de [52], que nos ajudou a fundamentar a motivação feita a seguir. Outras referências utilizadas pontualmente são mencionadas ao longo do texto.

Os métodos PQS são as generalizações do método de Newton para o problema geral de otimização. A idéia consiste em substituir, em cada passo, a função objetivo por uma aproximação quadrática e as restrições por equações ou inequações lineares. Dessa maneira, o *subproblema* a ser resolvido em cada iteração do algoritmo é um problema de programação quadrática que, em comparação com o problema original, pode ser considerado *simples*.

As propriedades de convergência do algoritmo tornam-se mais fáceis de serem obtidas quando lidamos com Lagrangeanos aumentados, onde o *parâmetro de penalidade* θ pertence ao intervalo $[0, 1]$. O Lagrangeano aumentado clássico é dado por

$$\hat{\Phi}(x, \lambda, \rho) = \ell(x, \lambda) + \rho\varphi(x),$$

onde $\lambda \in \mathbb{R}^m$, $\rho > 0$, $\ell(x, \lambda)$ é a função Lagrangeana dada por

$$\ell(x, \lambda) = f(x) + h(x)^T \lambda$$

e

$$\varphi(x) = \frac{1}{2}\|h(x)\|_2^2.$$

Neste trabalho, seguindo o que foi feito em [37], também usaremos a função de mérito obtida pela combinação convexa de ℓ e φ . Dado $\theta \in [0, 1]$, definimos

$$\Phi(x, \lambda, \theta) = \theta\ell(x, \lambda) + (1 - \theta)\varphi(x). \quad (1.6)$$

A confiança que depositamos em Φ como função de mérito se baseia no fato de que quando θ está perto de zero, a função φ se torna dominante na combinação (1.6). Assim, as componentes de $h(x)$ serão obrigatoriamente empurradas para valores pequenos. Agora,

para valores pequenos de $\|h(x)\|_2$, se a aproximação dos multiplicadores de Lagrange é mantida limitada, o efeito redutor devido à variação destes será desprezível. Logo, a diminuição do primeiro termo da combinação (1.6) será devido à diminuição de f , como veremos adiante.

É claro que podemos relacionar o Lagrangeano aumentado $\hat{\Phi}$ com a função de mérito Φ , através da relação $\rho = \frac{1-\theta}{\theta}$. Esta relação é bastante coerente, pois quando ρ tende ao infinito, temos θ tendendo para zero. Por outro lado, quando ρ tende a zero, temos θ tendendo para um. Por esses motivos, nos sentimos à vontade para, como em [37], chamar a função de mérito Φ de Lagrangeano aumentado.

Seja $x_k \in \mathbb{R}^n$, $l \leq x_k \leq u$, uma aproximação da solução do problema (1.1) e seja $\lambda_k \in \mathbb{R}^m$ o multiplicador de Lagrange associado. Definimos a função $Q(B_k, x_k, \lambda_k, s)$ como a aproximação quadrática de $\ell(x_k + s, \lambda_k)$, ou seja,

$$Q_k(s) \equiv Q(B_k, x_k, \lambda_k, s) = \frac{1}{2}s^T B_k s + \nabla \ell(x_k, \lambda_k)^T s + \ell(x_k, \lambda_k), \quad (1.7)$$

onde $B_k \in \mathbb{R}^{n \times n}$ é uma matriz simétrica. É comum pensar, de início, que B_k é a matriz Hessiana $\nabla^2 f(x_k, \lambda_k)$. Veremos mais adiante que, contrariamente a intuição, esta escolha ingênua não é a mais adequada. Por ora, consideraremos B_k como uma aproximação de $\nabla^2 \ell(x_k, \lambda_k)$.

O princípio dos algoritmos PQS é a resolução do problema (1.1) por meio de uma sequência de problemas quadráticos do tipo

$$\begin{aligned} &\text{minimizar} && Q_k(s) \\ &\text{sujeito a} && h'(x_k)s + h(x_k) = 0, \\ &&& l \leq x_k + s \leq u, \end{aligned} \quad (1.8)$$

onde $h'(x_k) \in \mathbb{R}^{m \times n}$ é a matriz Jacobiana calculada em x_k , definida anteriormente. Resolvendo (1.8), esperamos encontrar uma melhor aproximação $x_k + s_c$ para a solução do problema (1.1), onde s_c é solução de (1.8). Infelizmente, o problema quadrático (1.8) pode não ter solução. Isto acontece em duas situações:

(i) Quando a região factível de (1.8) é vazia. Com efeito, a variedade afim $h'(x_k)s + h(x_k) = 0$ pode não interceptar a caixa $l \leq x \leq u$. Além disso, quando o posto de $h'(x_k)$ é menor que m e $h(x_k)$ não está no espaço coluna de $h'(x_k)$, a própria variedade afim é vazia;

(ii) Quando a função objetivo de (1.8) não é limitada inferiormente na região factível. Neste caso, pela continuidade da função quadrática, a região factível não pode ser compacta; em particular, alguma componente de l ou u deve ser infinita.

Um problema menor é que, mesmo quando (1.8) tem solução, ela pode não ser única.

A maioria dos métodos PQS para resolver o problema (1.1) são baseados em modificações de (1.8). Para solucionar o problema da não-limitação, apresentado em (ii), introduz-se em (1.8) a *região de confiança*

$$\|s\| \leq \Delta, \quad (1.9)$$

onde $\Delta > 0$ é o *raio* da região de confiança. Cabe observar que, se $\|\cdot\| = \|\cdot\|_\infty$, o problema (1.8) com a restrição adicional (1.9) continua sendo um problema de programação

quadrática.

Já o problema da infactibilidade, apresentado em (i), é mais complicado de se contornar. A idéia apresentada em [37] consiste em dividir a iteração em duas fases. Primeiro, considera-se o problema

$$\begin{aligned} &\text{minimizar} && \frac{1}{2} \|h'(x_k)s + h(x_k)\|_2^2 \\ &\text{sujeito a} && l \leq x_k + s \leq u, \\ &&& \|s\| \leq 0.8\Delta. \end{aligned} \quad (1.10)$$

Depois, se s_{nor} é uma solução de (1.10), o incremento s_c é obtido resolvendo o problema quadrático factível

$$\begin{aligned} &\text{minimizar} && Q_k(s) \\ &\text{sujeito a} && h'(x_k)s = h'(x_k)s_{nor}, \\ &&& l \leq x_k + s \leq u, \\ &&& \|s\| \leq \Delta. \end{aligned} \quad (1.11)$$

Claramente, a abordagem via (1.10)-(1.11) é consistente com (1.8)-(1.9), no sentido de que a solução de ambos coincide quando (1.8)-(1.9) tem solução, Δ é suficientemente grande e $h(x_k)$ está no espaço coluna de $h'(x_k)$.

A restrição $\|s\| \leq 0.8\Delta$ em (1.10) obriga a região factível de (1.11) a ser não-vazia. Isto também seria conseguido se, em vez dessa restrição tivesse sido colocado $\|s\| \leq r\Delta$ para qualquer $r \in [0, 1]$. A escolha $r = 0.8$ parece satisfazer simultaneamente os requisitos de que $\|h'(x_k)s + h(x_k)\|_2^2$ seja suficientemente pequeno, e que a região factível de (1.11) seja suficientemente ampla para permitir um decréscimo em sua função objetivo.

Todo esse processo de minimização desenvolvido em (1.10) e (1.11) pode ser computacionalmente muito caro, principalmente quando se trata de problemas de grande porte. Isto se deve ao fato desses problemas estarem sendo resolvidos de maneira *exata*. Portanto, no intuito de evitar esse incômodo, são desenvolvidos procedimentos capazes de resolver (1.10)-(1.11) de maneira aproximada. No próximo capítulo, onde apresentaremos o algoritmo, será dado o significado preciso de resolver (1.10)-(1.11) trocando a palavra *minimizar* por *minimizar aproximadamente*. A idéia é que um incremento s_{nor} será considerado como uma aproximação da solução de (1.10) se ele causar um decréscimo na função objetivo de (1.10) de mesma ordem do decréscimo produzido por um *passo de Cauchy generalizado* relacionado com tal função objetivo. Chamamos de passo de Cauchy generalizado a um passo no sentido oposto do gradiente projetado na caixa.

De maneira análoga, um incremento s_c será considerado uma solução aproximada de (1.11) se o decréscimo $Q_k(s_c) - Q_k(s_{nor})$ é da mesma ordem do decréscimo produzido por um passo de Cauchy generalizado na região factível desse problema de programação quadrática.

Após calcularmos o incremento tentativo s_c e, conseqüentemente, o ponto tentativo $x_k + s_c$, precisaremos decidir se este ponto será aceito como uma nova estimativa da solução do problema (1.1). É nesse momento que faremos uso da função de mérito Φ definida em (1.6). O primeiro termo de (1.6) está relacionado com a otimalidade, enquanto o segundo termo está relacionado com a factibilidade. A função de mérito não envolve apenas o ponto tentativo $x_k + s_c$ e o parâmetro de penalidade θ , mas também uma nova estimativa $\lambda + \Delta\lambda$ dos multiplicadores de Lagrange. Neste trabalho, essa nova estimativa poderá ser

completamente arbitrária, sujeita apenas à restrição de ser limitada. Existem, no entanto, escolhas mais eficientes que outras. Por exemplo, uma boa idéia é escolher esse vetor de multiplicadores como o próprio vetor de multiplicadores associado à condição de otimalidade de (1.11). Para uma primeira leitura, e aproveitando-se dessa liberdade de escolha da nova estimativa dos multiplicadores de Lagrange, pode-se pensar que $\lambda + \Delta\lambda = 0$. De fato, essa é uma escolha admissível e a maior parte da teoria funciona com ela.

O nosso objetivo agora é mostrar que a função de mérito Φ se aproxima bem por uma quadrática nas variáveis x e λ . Da mesma forma que em [37], vamos supor que tanto f como h têm derivadas primeiras contínuas para todo $x \in \mathbb{R}^n$. Precisaremos escolher um parâmetro de penalidade adequado e que se relacione com x_k , s_c , λ_k e $\Delta\lambda$. Para tomar essa decisão, observamos que, se $\|s\|$ é suficientemente pequeno, então

$$h(x_k + s) \approx h'(x_k)s + h(x_k), \quad (1.12)$$

$$\varphi(x_k + s) = \frac{1}{2}\|h(x_k + s)\|_2^2 \approx \frac{1}{2}\|h'(x_k)s + h(x_k)\|_2^2 \quad (1.13)$$

e

$$\ell(x_k + s, \lambda_k) \approx Q(B_k, x_k, \lambda_k, s). \quad (1.14)$$

Por definição, sabemos que $\ell(x, \lambda) = f(x) + h(x)^T \lambda$. Logo

$$\ell(x_k + s, \lambda_k + \Delta\lambda) = f(x_k + s) + h(x_k + s)^T (\lambda_k + \Delta\lambda),$$

isto é,

$$\ell(x_k + s, \lambda_k + \Delta\lambda) = \ell(x_k + s, \lambda_k) + h(x_k + s)^T \Delta\lambda.$$

Portanto, usando (1.12) e (1.14), temos que

$$\ell(x_k + s, \lambda_k + \Delta\lambda) \approx Q(B_k, x_k, \lambda_k, s) + [h'(x_k)s + h(x_k)]^T \Delta\lambda, \quad (1.15)$$

para $\|s\|$ pequeno e $\|\Delta\lambda\|$ limitado.

Iremos definir a *redução verdadeira* da função de mérito Φ de (x_k, λ_k) até $(x_k + s, \lambda_k + \Delta\lambda)$, como

$$Ared(x_k, \lambda_k, s, \Delta\lambda, \theta) = \Phi(x_k, \lambda_k, \theta) - \Phi(x_k + s, \lambda_k + \Delta\lambda, \theta), \quad (1.16)$$

onde a notação *Ared* vem do inglês *actual reduction*. Se usarmos em (1.16) a definição para Φ dada por (1.6), então teremos que

$$Ared(x_k, \lambda_k, s, \Delta\lambda, \theta) = \theta[\ell(x_k, \lambda_k) - \ell(x_k + s, \lambda_k + \Delta\lambda)] + (1 - \theta)[\varphi(x_k) - \varphi(x_k + s)].$$

Logo, de (1.13) e (1.15), temos

$$\begin{aligned} Ared(x_k, \lambda_k, s, \Delta\lambda, \theta) &\approx \theta [Q_k(0) - Q_k(s) - [h'(x_k)s + h(x_k)]^T \Delta\lambda] + \\ &+ (1 - \theta) \left[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s + h(x_k)\|_2^2 \right], \end{aligned} \quad (1.17)$$

onde $Q_k(s)$ representa $Q(B_k, x_k, \lambda_k, s)$. Observamos que o membro do lado direito da expressão (1.17) pode ser considerado como uma boa aproximação da redução verdadeira

$Ared$ dada por (1.16). Isto nos leva a definir a *redução prevista* da função Φ de (x_k, λ_k) até $(x_k + s, \lambda_k + \Delta\lambda)$, como

$$\begin{aligned} Pred(x_k, \lambda_k, s, \Delta\lambda, B_k, \theta) &= \theta [Q_k(0) - Q_k(s) - [h'(x_k)s + h(x_k)]^T \Delta\lambda] + \\ &+ (1 - \theta) \left[\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s + h(x_k)\|_2^2 \right], \end{aligned} \quad (1.18)$$

onde a notação $Pred$ vem do inglês *predicted reduction*. Consequentemente, de (1.17) e (1.18), podemos escrever que

$$Ared(x_k, \lambda_k, s, \Delta\lambda, \theta) \approx Pred(x_k, \lambda_k, s, \Delta\lambda, B_k, \theta). \quad (1.19)$$

Baseados em (1.12), (1.13) e (1.14), reescrevemos (1.19) como

$$Ared(x_k, \lambda_k, s, \Delta\lambda, \theta) = Pred(x_k, \lambda_k, s, \Delta\lambda, B_k, \theta) + \mathcal{O}(\|s\|^2), \quad (1.20)$$

o que será útil na demonstração de alguns resultados. A notação $a = \mathcal{O}(b)$ significa que existe uma constante $M > 0$ tal que $a \leq Mb$. Mais adiante, na seção 3.1, veremos que a relação (1.20) é válida com a introdução de uma hipótese adequada.

Agora, queremos que a função de mérito calculada no novo ponto $x_k + s_c$ seja menor do que a calculada no ponto x_k da iteração em questão. De acordo com (1.16), isto significa que a redução verdadeira $Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)$ deve ser suficientemente maior que zero. Portanto, por (1.19), é bastante natural escolher o parâmetro de penalização θ de modo que a redução prevista $Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)$ também seja suficientemente maior que zero. No entanto, quando calculamos s_c , apenas o segundo termo de $Pred$ é maior ou igual a zero, isto é,

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2 \geq 0.$$

Em outras palavras, $Pred$ é uma combinação convexa do tipo $\theta a + (1 - \theta)b$, onde, necessariamente, $b \geq 0$. No entanto, o elemento a dessa combinação convexa não é necessariamente positivo. Sabemos apenas que se o segundo termo b se anula, então o primeiro termo a é não-negativo. Portanto, é possível escolher um parâmetro de penalização $\theta > 0$ pequeno o bastante para garantir que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2}{2},$$

ou seja,

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4} [\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2] \geq 0. \quad (1.21)$$

Por (1.19) e (1.21), $Ared$ será suficientemente positivo para Δ suficientemente pequeno. De fato, o ponto tentativo $x_k + s_c$ será aceito quando

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1 Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta). \quad (1.22)$$

Alem disso, a exigência (1.21) se verifica para $\theta = 0$. Isto nos leva a definir $\theta^{sup} \geq 0$ como

$$\theta^{sup} = \sup \left\{ \theta \in [0, 1] \mid Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4} \|h(x_k)\|_2^2 - \frac{1}{4} \|h'(x_k)s_c + h(x_k)\|_2^2 \right\}.$$

Para obter uma fórmula explícita para θ^{sup} usaremos a definição de $Pred$. Com efeito, temos que

$$\begin{aligned} Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) &= \theta [Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda] + \\ &\quad (1 - \theta) \left[\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2 \right], \end{aligned}$$

então temos que θ^{sup} deve satisfazer a desigualdade

$$\begin{aligned} &\theta [Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda] + \\ &\quad + (1 - \theta) \left[\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2 \right] \geq \frac{1}{4} \|h(x_k)\|_2^2 - \frac{1}{4} \|h'(x_k)s_c + h(x_k)\|_2^2. \end{aligned}$$

Colocando o parâmetro θ em evidência, teremos

$$\begin{aligned} &\theta [Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda - \frac{1}{2} \|h(x_k)\|_2^2 + \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2] \geq \\ &\quad \geq \frac{1}{4} \|h'(x_k)s_c + h(x_k)\|_2^2 - \frac{1}{4} \|h(x_k)\|_2^2. \end{aligned}$$

Isolando θ , obtemos

$$\theta \geq \frac{\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2}{2[Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta\lambda + \frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2]}.$$

Logo, a fórmula explícita para θ^{sup} é dada por

$$\theta^{sup} = \min \left\{ 1, \frac{\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2}{2[Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta\lambda + \frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2]} \right\}. \quad (1.23)$$

Se sempre escolhermos $\theta \leq \theta^{sup}$, então, por (1.21), a condição (1.22) implicará descida simples da função de mérito Φ , isto é, $\Phi(x_k + s_c, \lambda_k + \Delta\lambda, \theta) \leq \Phi(x_k, \lambda_k, \theta)$. Como valores maiores que θ^{sup} não satisfazem (1.21), parece bastante sensato impor a condição

$$\theta \leq \theta^{sup} \quad (1.24)$$

para a escolha do parâmetro de penalização θ . No entanto, esta condição ainda deixa bastante liberdade, quando $\theta^{sup} > 0$. Outras considerações serão necessárias para fazer uma eleição adequada, dentro das possíveis.

Um algoritmo baseado na função de mérito Φ poderia ser interpretado como um método destinado a minimizar Φ sujeita apenas às restrições de canalização $l \leq x \leq u$. Esta interpretação seria compatível com o conceito genérico do significado de uma função de mérito. No entanto, neste caso, tal interpretação não parece totalmente adequada, devido ao fato da função Φ mudar de uma iteração para outra, de acordo com a escolha de θ . Com efeito, θ estabelece pesos relativos para a factibilidade e a otimalidade no

algoritmo (com θ perto de 0 o método privilegia a factibilidade e com θ perto de 1 privilegia a otimalidade). Grandes variações de θ de uma iteração para outra pareceriam indicar que o método não consegue decidir qual é o peso adequado para cada um dos objetivos que são visados. Essa é uma motivação para limitar, pelo menos assintoticamente, as variações de θ . A maneira mais óbvia de forçar uma variação limitada de θ consiste em impor, além da condição (1.24), que

$$\theta \leq \theta_{k-1}, \quad \forall k \in \mathbb{N}, \quad (1.25)$$

onde θ_k é o valor de θ escolhido na k -ésima iteração do algoritmo. Juntando as condições (1.24) e (1.25), teremos que a sequência $\{\theta_k\}$ é monótona não-crescente e positiva, portanto convergente. Isso implicaria que, a longo prazo, a função de mérito seria, essencialmente, a mesma, e a interpretação criticada acima passaria a ser válida.

No entanto, a escolha monótona de θ também não é plenamente satisfatória. Lembrando que θ estabelece uma ponderação entre factibilidade e otimalidade, seria possível que, sobretudo nas primeiras iterações, valores muito pequenos de θ fossem impostos por (1.24) devido à necessidade de reforçar a factibilidade, e que esses valores muito pequenos fossem herdados por todas as iterações posteriores, onde valores maiores seriam toleráveis. Em outras palavras, a condição (1.25) carrega demasiadamente a história de dificuldades passadas do algoritmo, que podem não existir mais na iteração atual. Essas considerações levaram os autores de [37] à definição da seguinte estratégia *não-monótona* para θ : escolhe-se, independentemente de k , um número $N \geq 0$ que representará o *grau de não-monotonicidade* de $\{\theta_k\}$. A situação $N = 0$ corresponderá à escolha monótona, baseada em (1.25), e valores grandes de N aproximarão θ de θ^{sup} . Definimos

$$\theta_k^{min} = \min\{1, \theta_0, \dots, \theta_{k-1}\}, \quad (1.26)$$

$$\theta_k^{grande} = \left[1 + \frac{N}{(k+1)^{1.1}} \right] \theta_k^{min} \quad (1.27)$$

e

$$\theta = \min\{\theta_k^{grande}, \theta_k^{sup}\}. \quad (1.28)$$

Apesar de (1.26), (1.27) e (1.28) não implicarem na monotonia de $\{\theta_k\}$, essas escolhas implicam na convergência da sequência $\{\theta_k\}$, o que, do ponto de vista da interpretação da função de mérito, é igualmente satisfatório.

No próximo capítulo iremos descrever rigorosamente o algoritmo que acabou de ser esquematizado. Além disso, será provado que este algoritmo está bem definido. Completaremos os detalhes que até então foram omitidos nesta primeira apresentação.

Capítulo 2

O Algoritmo PQS e a boa definição

Neste capítulo será descrito em detalhes o algoritmo de programação quadrática sequencial proposto em [37]. A seguir, será provado que este algoritmo está bem definido. Assim como em [37], com o objetivo de facilitar a leitura do algoritmo, vários parâmetros foram trocados por números entre 0 e 1. O leitor deve ter em mente que, sempre que aparecer um número decimal, ele pode ser trocado por outro número no intervalo $(0, 1)$.

Antes de apresentarmos o algoritmo propriamente dito, comentaremos alguns de seus parâmetros. No começo de cada iteração será usado um limitante inferior $\Delta_{min} > 0$ para o raio da região de confiança. Portanto, a k -ésima iteração começará com um raio $\Delta \geq \Delta_{min}$. Para a estimativa dos multiplicadores de Lagrange e para as normas das matrizes que aproximam a Hessiana serão definidos limitantes superiores L_1 e L_2 , respectivamente. Finalmente, será definido um número $N > 0$ que reflete o grau de não-monotonicidade desejado para o parâmetro de penalidade.

Então, dadas a estimativa x_k para a solução do problema (1.1), a estimativa λ_k para o vetor dos multiplicadores de Lagrange e uma matriz simétrica B_k (que, geralmente, é uma aproximação para $\nabla^2 \ell(x_k, \lambda_k)$), o Algoritmo PQS a seguir descreve os passos para a obtenção de x_{k+1} e λ_{k+1} .

2.1 O Algoritmo PQS

Sejam $x_k \in \Omega$, $\lambda_k \in \mathbb{R}^m$, $\|\lambda_k\| \leq L_1$, $\Delta \geq \Delta_{min}$, $B_k = B_k^T \in \mathbb{R}^{n \times n}$, $\|B_k\| \leq L_2$ e $N \geq 0$.

Se x_k é um ponto estacionário de (1.1),
 x_k é φ -estacionário, mas não é factível, ou
 x_k é factível, mas não é regular,

então o algoritmo para.

Senão, x_{k+1} é obtido da seguinte maneira:

Passo 0: Cálculo de uma direção de descida factível para as restrições, ou, se o ponto é factível, para o Lagrangeano.

Se x_k não é φ -estacionário, então calcule $d_{nor} \in \mathbb{R}^n$ tal que:

$$l \leq x_k + d_{nor}(x_k) \leq u \quad (2.1)$$

$$d_{nor}(x_k)^T \nabla \varphi(x_k) < 0. \quad (2.2)$$

Se x_k é factível, regular e não-estacionário para o problema (1.1), então faça $d_{nor}(x_k) = 0$, $s_{nor} = s_{nor}(x_k, \Delta) = 0$ e calcule $d_{tan}(x_k) \in \mathbb{R}^n$ tal que:

$$l \leq x_k + d_{tan}(x_k) \leq u \quad (2.3)$$

$$h'(x_k)d_{tan}(x_k) = 0 \quad (2.4)$$

$$d_{tan}(x_k)^T \nabla \ell(x_k, \lambda_k) = d_{tan}(x_k)^T \nabla f(x_k) < 0. \quad (2.5)$$

Passo 1: *Cálculo do decréscimo nas restrições.*

Se x_k não é factível, calcule $s_{nor}^{dec} = s_{nor}^{dec}(x_k, \Delta) \in \mathbb{R}^n$ resolvendo:

$$\begin{aligned} &\text{minimizar} \quad \frac{1}{2} \|h'(x_k)s + h(x_k)\|_2^2 \\ &\text{sujeito a} \quad \|s\| \leq 0.8\Delta \\ &\quad \quad \quad l \leq x_k + s \leq u \\ &\quad \quad \quad s = td_{nor}(x_k), \quad t \geq 0. \end{aligned}$$

Passo 2: *Cálculo do passo normal.*

Se x_k não é factível, calcule $s_{nor} = s_{nor}(x_k, \Delta) \in \mathbb{R}^n$ tal que $l \leq x_k + s_{nor} \leq u$, $\|s_{nor}\| \leq 0.8\Delta$ e

$$\begin{aligned} &[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor} + h(x_k)\|_2^2] \geq \\ &0.9[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}^{dec} + h(x_k)\|_2^2]. \end{aligned} \quad (2.6)$$

Passo 3: *Cálculo do decréscimo do Lagrangiano no espaço tangente.*

Se x_k é factível, faça $d_{tan} = d_{tan}(B_k, x_k, \lambda_k, \Delta) = d_{tan}(x_k)$ (definido no Passo 0).

Caso contrário, calcule $d_{tan} = d_{tan}(B_k, x_k, \lambda_k, \Delta) \in \mathbb{R}^n$ tal que:

$$l \leq x_k + s_{nor} + d_{tan}(x_k) \leq u \quad (2.7)$$

$$h'(x_k)d_{tan} = 0 \quad (2.8)$$

$$d_{tan}^T \nabla Q(B_k, x_k, \lambda_k, s_{nor}) < 0, \quad (2.9)$$

onde $Q_k(s) \equiv Q(B_k, x_k, \lambda_k, s) = \frac{1}{2}s^T B_k s + \nabla \ell(x_k, \lambda_k)^T s + \ell(x_k, \lambda_k)$.

Se não existe d_{tan} satisfazendo as três condições acima, então faça $d_{tan} = 0$.

Calcule $s_{tan}^{dec} = s_{tan}^{dec}(B_k, x_k, \lambda_k, \Delta)$ resolvendo:

$$\begin{aligned} & \text{minimizar} && Q_k(s_{nor} + s) \\ & \text{sujeito a} && \|s_{nor} + s\| \leq \Delta \\ & && l \leq x_k + s_{nor} + s \leq u \\ & && s = td_{tan}, \quad t \geq 0, \end{aligned}$$

onde $Q_k(s_{nor} + s) = \frac{1}{2}(s_{nor} + s)^T B_k(s_{nor} + s) + \nabla \ell(x_k, \lambda_k)^T(s_{nor} + s) + \ell(x_k, \lambda_k)$.

Passo 4: *Cálculo do passo tangente.*

Calcule $s_{tan} = s_{tan}(B_k, x_k, \lambda_k, \Delta) \in \mathbb{R}^n$ tal que:

$$h'(x_k)s_{tan} = 0; \quad l \leq x_k + s_{nor} + s_{tan} \leq u; \quad \|s_{nor} + s_{tan}\| \leq \Delta \quad (2.10)$$

e

$$[Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan})] \geq 0.9[Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec})]. \quad (2.11)$$

Passo 5: *Cálculo da atualização dos multiplicadores e composição dos passos.*

Calcule $\Delta\lambda \in \mathbb{R}^m$ tal que $\|\lambda_k + \Delta\lambda\| \leq L_1$ e defina $s_c = s_c(B_k, x_k, \lambda_k, \Delta) = s_{nor} + s_{tan}$.

Passo 6: *Cálculo do parâmetro de penalização $\theta \in [0, 1]$.*

Calcule θ_k^{min} , θ_k^{grande} , θ_k^{sup} de tal maneira que:

Se $k = 0$, então $\theta_k^{min} = 1$, senão $\theta_k^{min} = \min\{1, \theta_0, \dots, \theta_{k-1}\}$;

$$\theta_k^{grande} = \left[1 + \frac{N}{(k+1)^{1.1}}\right] \theta_k^{min};$$

$$\theta_k^{sup} = \min \left\{ 1, \frac{\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2}{2[Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta\lambda + \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2]} \right\}.$$

Se Δ é o primeiro raio de confiança testado nesta iteração, então defina

$$\theta' = \theta'(x_k, \Delta) = \Delta_k^{grande}.$$

Caso contrário,

$$\theta' = \theta'(x_k, \Delta) = \theta(x_k, \Delta'),$$

onde Δ' é o raio de confiança testado antes de ser atualizado para Δ .

Faça $\theta = \min\{\theta_k^{sup}, \theta'\}$.

Passo 7: *Aceitação do passo s_c .*

Se $Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)$,
então defina $\begin{cases} x_{k+1} = x_k + s_c \\ \lambda_{k+1} = \lambda_k + \Delta\lambda \\ \theta_k = \theta. \end{cases}$
Senão escolha novo $\Delta \in [0.1\Delta, 0.9\Delta]$ e volte para o Passo 1.

Isto completa a descrição do algoritmo. □

Cabe observar que a equação (2.4) do Algoritmo PQS nos permite concluir que

$$[\nabla\ell(x_k, \lambda_k) - \nabla f(x_k)] \in \mathcal{Im}(h'(x_k)^T). \quad (2.12)$$

Com efeito, a equação (2.4) é dada por $h'(x_k)d_{tan}(x_k) = 0$. Logo,

$$d_{tan}(x_k) \in \mathcal{N}(h'(x_k)) = \mathcal{Im}(h'(x_k)^T)^\perp.$$

Temos que $\nabla\ell(x_k, \lambda_k) = \nabla f(x_k) + h'(x_k)^T \lambda_k$, pois $\ell(x, \lambda) = f(x) + \lambda^T h(x)$. Assim,

$$\begin{aligned} d_{tan}(x_k)^T \nabla\ell(x_k, \lambda_k) &= d_{tan}(x_k)^T [\nabla f(x_k) + h'(x_k)^T \lambda_k] = \\ &= d_{tan}(x_k)^T \nabla f(x_k) + [h'(x_k)d_{tan}(x_k)]^T \lambda_k = d_{tan}(x_k)^T \nabla f(x_k). \end{aligned}$$

Portanto, $d_{tan}(x_k)^T [\nabla\ell(x_k, \lambda_k) - \nabla f(x_k)] = 0$. Logo, podemos concluir que

$$[\nabla\ell(x_k, \lambda_k) - \nabla f(x_k)] \perp d_{tan}(x_k) \Rightarrow [\nabla\ell(x_k, \lambda_k) - \nabla f(x_k)] \in \mathcal{Im}(h'(x_k)^T).$$

Este fato será útil nas provas dos resultados de convergência do Algoritmo PQS.

2.2 O algoritmo está bem definido

Visando aplicação a problemas de grande porte, os subproblemas dos Passos 2 e 4 são resolvidos *aproximadamente*, com critérios convenientes para a precisão da resolução. Provaremos agora que o Algoritmo PQS está bem definido, isto é, que sob hipóteses adequadas, que incluem a possibilidade de x_k ainda não ser uma solução, pode-se encontrar x_{k+1} em tempo finito. Em outras palavras, mostraremos que o algoritmo está bem definido em duas situações:

- (i) x_k não é um ponto φ -estacionário;
- (ii) x_k é um ponto factível, regular e não-estacionário para o problema (1.1).

Observe que nos casos restantes, isto é, quando x_k é φ -estacionário, mas não é factível; quando x_k é factível e não-regular ou quando x_k é estacionário para o problema (1.1), o algoritmo termina. O caso (i) será tratado no Teorema 4, e o caso (ii) será abordado no Teorema 5.

É interessante observar que as condições suficientes que garantem o fato de x_{k+1} ser encontrado em tempo finito são bastante fracas. Em particular, as direções d_{nor} e d_{tan} precisam apenas satisfazer (2.2), (2.5), (2.6) e (2.9). Mais adiante, veremos que para obter

a convergência do Algoritmo PQS, precisaremos de condições mais fortes sobre d_{nor} e d_{tan} .

Teorema 4 (Boa definição em pontos não factíveis) *Seja x_k um ponto que não é φ -estacionário. Então, o Algoritmo PQS calcula um novo ponto x_{k+1} por meio de uma quantidade finita de passagens pelos Passos 1-7.*

Dem. Como x_k não é φ -estacionário, segue que existe uma direção $d_{nor} \neq 0$ calculada pelo Passo 0 do Algoritmo PQS. Considere $\beta(t) \equiv \frac{1}{2}\|h'(x_k)td_{nor} + h(x_k)\|_2^2$. Observe que $\beta(t)$ é uma função quadrática convexa e, por (2.2), $\beta'(0) = d_{nor}^T \nabla \varphi(x_k) < 0$. Se a quadrática é estritamente convexa (coeficiente de segunda ordem estritamente positivo), então $\beta(t)$ admite um minimizador irrestrito $\hat{t} > 0$. Propriedades elementares das quadráticas garantem, nesse caso, que

$$\beta(t) \leq \beta(0) + \frac{1}{2}\beta'(0)t, \quad (2.13)$$

para todo $t \in [0, \hat{t}]$. Se $\beta(t)$ não é estritamente convexa, então seu gráfico é uma reta, e (2.13) se satisfaz trivialmente para todo $t \geq 0$.

Definimos, para todo $\Delta > 0$,

$$\bar{t} = \max\{t > 0 \mid [x_k, x_k + td_{nor}] \subset \Omega, \text{ e } \|td_{nor}\| \leq 0.8\Delta\}$$

e $t(\Delta) = \min\{\hat{t}, \bar{t}\}$. Como $x_k + d_{nor} \in \Omega$, temos que $\|\bar{t}d_{nor}\| = 0.8\Delta$ se $\Delta \leq \|d_{nor}\|$. Naturalmente, (2.13) vale para todo $t \in [0, t(\Delta)]$. Como

$$\beta(0) = \frac{1}{2}\|h(x_k)\|_2^2 \text{ e } \beta'(0) = d_{nor}^T \nabla \varphi(x_k),$$

segue que existe $\bar{\Delta} \in (0, \|d_{nor}\|]$ tal que, de (2.13), vale a relação

$$\begin{aligned} \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)t(\Delta)d_{nor} + h(x_k)\|_2^2 &\geq -\frac{1}{2}d_{nor}^T \nabla \varphi(x_k)t(\Delta) \\ &= c\|d_{nor}\|t(\Delta) = 0.8c\Delta, \end{aligned} \quad (2.14)$$

para todo $\Delta \in (0, \bar{\Delta})$, onde $c = -\frac{d_{nor}^T \nabla \varphi(x_k)}{2\|d_{nor}\|} > 0$. Portanto, pelos Passos 1 e 2 do Algoritmo PQS e por (2.14), temos que

$$\begin{aligned} \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c(\Delta) + h(x_k)\|_2^2 &= \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}(\Delta) + h(x_k)\|_2^2 \geq \\ &\geq 0.9\left[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)t(\Delta)d_{nor} + h(x_k)\|_2^2\right] \geq 0.9[0.8c\Delta] = 0.72c\Delta. \end{aligned}$$

Como na escolha do parâmetro de penalização θ exigimos que (1.21) ocorra, temos que, para $\Delta \in (0, \bar{\Delta})$,

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq 0.36c\Delta. \quad (2.15)$$

Agora, pela diferenciabilidade de f e h , pela limitação de $\|\Delta\lambda\|$ e pelas definições de Q e $Pred$, temos que

$$\lim_{\Delta \rightarrow 0} \frac{|Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)|}{\Delta} = 0.$$

Então, de (2.15), temos

$$\lim_{\Delta \rightarrow 0} \left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} - 1 \right| = 0.$$

Isto significa que para Δ suficientemente pequeno, a desigualdade (1.22), dada por

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta),$$

valerá, completando nossa demonstração. $\square$

Teorema 5 (Boa definição em pontos factíveis) *Seja x_k um ponto factível, regular e não-estacionário para o problema (1.1). Então, o Algoritmo PQS calcula um novo ponto x_{k+1} por meio de uma quantidade finita de passagens pelos Passos 1-7.*

Dem. Neste caso, temos $s_{nor} = s_{nor}(\Delta) = 0$, para todo $\Delta > 0$. Como x_k é um ponto factível, regular e não-estacionário para o problema (1.1), existe uma direção $d_{tan}(x_k) \neq 0$, calculada no Passo 0 do Algoritmo PQS. Tomamos $\beta(t) \equiv Q_k(td_{tan})$, onde a função Q foi definida em (1.7). Usando o mesmo raciocínio do Teorema 4, temos que

$$\beta(t) \leq \beta(0) + \frac{1}{2}\beta'(0)t, \quad \forall t \in (0, \hat{t}], \quad (2.16)$$

onde $\hat{t} > 0$ é o minimizador irrestrito da função quadrática $\beta(t)$. Assim como no Teorema 4, definimos

$$\bar{t} = \max\{t > 0 \mid [x_k, x_k + td_{tan}] \subset \Omega, \text{ e } \|td_{tan}\| \leq \Delta\}$$

e $t(\Delta) = \min\{\hat{t}, \bar{t}\}$. Como $x_k + d_{tan} \in \Omega$, temos que $\|t(\Delta)d_{tan}\| = \Delta$, sempre que $\Delta \leq \|d_{tan}\|$. Naturalmente, (2.16) vale para todo $t \in [0, t(\Delta)]$. Logo, como $Q_k(s_{tan}^{dec}) \leq Q_k(t(\Delta)d_{tan})$, existe $\bar{\Delta} \in (0, \|d_{tan}\|]$ tal que, de (2.16), vale a relação

$$Q_k(0) - Q_k(s_{tan}^{dec}) \geq Q_k(0) - Q_k(t(\Delta)d_{tan}) \geq -\frac{1}{2}d_{tan}^T \nabla Q_k(0)t(\Delta) = c\Delta, \quad (2.17)$$

para todo $\Delta \in (0, \bar{\Delta})$, onde $c = -\frac{d_{tan}^T \nabla Q_k(0)}{2\|d_{tan}\|}$. Daí, pelos Passos 3 e 4 do Algoritmo PQS e por (2.17), temos que

$$Q_k(0) - Q_k(s_c(\Delta)) \geq 0.9[Q_k(0) - Q_k(s_{tan}^{dec})] \geq 0.9c\Delta,$$

para todo $\Delta \in (0, \bar{\Delta})$. Além disso, como, neste caso, $h(x_k) = 0$ e $h'(x_k)s_c = 0$, segue da definição de $Pred$ que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) = \theta[Q_k(0) - Q_k(s_c(\Delta))] \geq 0.9\theta c\Delta \geq 0. \quad (2.18)$$

Fica assim, evidente que a exigência (1.21) é válida para todo $\Delta \in (0, \bar{\Delta})$ e $\theta \in [0, 1]$. Isto significa que o parâmetro de penalização θ não é diminuído através de qualquer sequência de possíveis decréscimos de Δ no *loop* do algoritmo. Portanto, de (2.18), temos

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq 0.9\theta'c\Delta, \quad (2.19)$$

para todo $\Delta \in (0, \bar{\Delta})$, onde θ' é o valor de θ que satisfaz a exigência (1.21), no começo do *loop*. Daí, como no Teorema 4, temos que

$$\lim_{\Delta \rightarrow 0} \frac{|Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)|}{\Delta} = 0,$$

e portanto, de (2.19), vale

$$\lim_{\Delta \rightarrow 0} \left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} - 1 \right| = 0,$$

o que completa a demonstração. □

Capítulo 3

Propriedades de convergência global

Na seção 2.2, foi demonstrado que o Algoritmo PQS termina na k -ésima iteração apenas nas situações em que:

- (a) x_k é φ -estacionário, mas não é factível;
- (b) x_k é factível e não-regular;
- (c) x_k é factível e estacionário para o problema (1.1).

Neste capítulo, será investigado o comportamento das sequências infinitas geradas pelo Algoritmo PQS. É comum que a prova da convergência global de um algoritmo esteja muito relacionada com a prova de boa definição. Isto é bastante natural já que, na boa definição, provamos que os pontos onde o algoritmo deve *parar* têm determinadas características, e nos teoremas de convergência, geralmente, provamos que os pontos limite da sequência gerada têm essas mesmas características. Logo, os teoremas de convergência dizem sobre o limite o mesmo que os resultados de boa definição dizem sobre os iterandos. Muitas vezes, as provas de convergência global reproduzem, com variadas complicações analíticas, as idéias usadas para provar a boa definição.

A prova que apresentaremos está dividida em duas partes, que correspondem aos Teoremas 6 e 7. Nos dois casos usa-se como hipótese que a sequência gerada pelo Algoritmo PQS está totalmente contida em um compacto de $\mathbb{R}^n$. Evidentemente, quando as cotas l e u são finitas, esta é uma hipótese perfeitamente razoável. Na primeira parte prova-se que *todos* os pontos limites de uma sequência gerada pelo Algoritmo PQS são pontos φ -estacionários. Já na segunda parte demonstra-se a existência de *pelo menos um* ponto limite que é estacionário para o problema (1.1).

Conforme a necessidade, algumas hipóteses adicionais serão introduzidas. A primeira hipótese a ser considerada retrata a direção d_{nor} do Passo 0 do Algoritmo PQS. Uma escolha razoável para d_{nor} seria

$$d_{nor}(x_k) = \begin{cases} -\nabla\varphi(x_k), & \text{se } x_k \text{ está no interior de } \Omega \\ P_{\Omega}(-\nabla\varphi(x_k)), & \text{se } x_k \text{ está na fronteira de } \Omega, \end{cases}$$

onde P_{Ω} é a projeção ortogonal em Ω . Infelizmente, definida desta forma, a direção

$d_{nor}(x_k)$ não é contínua no ponto x_k da iteração em questão. Faremos, portanto, a seguinte hipótese:

(h.1) Hipótese da continuidade de d_{nor} : Para todo $x_k \in \Omega$, a aplicação d_{nor} , do Passo 0 do Algoritmo PQS, é contínua em x_k .

Uma possível escolha que satisfaz a hipótese (h.1) é dada por

$$d_{nor}(x_k) = P_\Omega(x_k - \alpha \nabla \varphi(x_k)) - x_k, \quad (3.1)$$

onde P_Ω é a projeção ortogonal em Ω e α é um parâmetro fixo. Nesse caso, a continuidade de d_{nor} deve-se ao fato de P_Ω e $\nabla \varphi$ serem aplicações contínuas. Cabe observar que, por (3.1), podemos interpretar a direção d_{nor} como sendo $y^* - x_k$, onde y^* é o minimizador de um modelo quadrático simples numa caixa, mais especificamente, do problema

$$\begin{aligned} &\text{minimizar} \quad \frac{1}{2} \|x_k - \alpha \nabla \varphi(x_k) - y\|_2^2 \\ &\text{sujeito a} \quad l \leq y \leq u. \end{aligned} \quad (3.2)$$

É interessante observar que a hipótese (h.1) é bastante fraca e que junto com a condição angular (2.2) (e, mais adiante, as hipóteses (h.2) e (h.3)), são suficientes para garantir a convergência para pontos φ -estacionários. Na seção 3.2, veremos que, para obter a convergência para pontos críticos do problema (1.1), serão necessárias mais hipóteses sobre d_{nor} e d_{tan} .

3.1 Todo ponto limite é φ -estacionário

Nesta seção, será feita a primeira parte da prova de convergência do Algoritmo PQS. Provaremos, no Teorema 6, que todos os pontos limites de uma sequência gerada pelo o Algoritmo PQS são φ -estacionários. Para isso, precisaremos fazer uso dos quatro lemas seguintes.

No Lema 3.1, será provado que, na vizinhança de um ponto que não é φ -estacionário, a redução prevista $Pred$ da função de mérito é proporcional ao raio Δ da região de confiança.

Lema 3.1 *Seja $x_* \in \Omega$ um ponto que não é φ -estacionário e considere válida a hipótese (h.1). Se o Algoritmo PQS é aplicado a x_k e*

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4} [\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2],$$

então existem $\epsilon, \Delta', c > 0$ tais que para todo $x_k \in \Omega$, $\|x_k - x_\| \leq \epsilon$, temos*

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq c \min\{\Delta, \Delta'\}.$$

Dem. Como $x_* \in \Omega$ não é φ -estacionário, então existe uma direção de descida contínua d_{nor} tal que

$$d_{nor}(x_*)^T \nabla \varphi(x_*) < 0. \quad (3.3)$$

Pela continuidade de $d_{nor}(x_k)$ e $\nabla \varphi(x_k)$, existe $\epsilon > 0$ tal que para todo $x_* \in \Omega$ e $\|x_k - x_*\| \leq \epsilon$ temos

$$2\|d_{nor}(x_*)\| \geq \|d_{nor}(x_k)\| \geq \frac{\|d_{nor}(x_*)\|}{2} > 0 \quad (3.4)$$

e

$$d_{nor}(x_k)^T \nabla \varphi(x_k) \leq \frac{1}{2} d_{nor}(x_*)^T \nabla \varphi(x_*) < 0. \quad (3.5)$$

De (3.4), obtemos

$$0 < \frac{1}{2\|d_{nor}(x_*)\|} \leq \frac{1}{\|d_{nor}(x_k)\|} \leq \frac{2}{\|d_{nor}(x_*)\|}. \quad (3.6)$$

Pela continuidade de $h'(x_k)$, temos que $|h'(x_k)^T h(x_k)|$ é limitado em $\|x_k - x_*\| \leq \epsilon$. Analogamente ao Teorema 4, temos que existe $t_1 \in (0, 1)$ tal que

$$\beta(t) \leq \beta(0) + \frac{1}{2} \beta'(0)t, \quad (3.7)$$

para todo $t \in (0, t_1]$, onde $\beta(t) \equiv \frac{1}{2} \|h'(x_k)t d_{nor}(x_k) + h(x_k)\|_2^2$. Observamos que

$$\beta(0) = \frac{1}{2} \|h(x_k)\|_2^2 \quad \text{e} \quad \beta'(0) = d_{nor}(x_k)^T \nabla \varphi(x_k).$$

Então, de (3.5) e (3.7), temos que

$$\begin{aligned} \frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)t d_{nor}(x_k) + h(x_k)\|_2^2 &\geq -\frac{t}{2} d_{nor}(x_k)^T \nabla \varphi(x_k) \geq \\ &\geq -\frac{t}{4} d_{nor}(x_*)^T \nabla \varphi(x_*) > 0. \end{aligned}$$

Portanto, obtemos:

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)t d_{nor}(x_k) + h(x_k)\|_2^2 \geq c_1 t d_{nor}(x_*)^T \nabla \varphi(x_*) > 0, \quad (3.8)$$

para todo $t \in (0, t_1]$ e $\|x_k - x_*\| \leq \epsilon$, com $c_1 = -\frac{1}{4} < 0$.

Considere $\Delta' = \frac{\|d_{nor}(x_*)\|}{2} t_1$.

Caso 1: Se $0 < \Delta \leq \Delta'$ e $\|x_k - x_*\| \leq \epsilon$.

Como $\Delta \leq \Delta'$ e, de (3.6), $\frac{1}{\|d_{nor}(x_k)\|} \leq \frac{2}{\|d_{nor}(x_*)\|}$, então $0 < \frac{\Delta}{2\|d_{nor}(x_k)\|} \leq \frac{\Delta'}{\|d_{nor}(x_*)\|} = \frac{t_1}{2} < t_1$, o que significa que $\frac{\Delta}{2\|d_{nor}(x_k)\|} \in (0, t_1]$. Logo (3.8) vale para $t = \frac{\Delta}{2\|d_{nor}(x_k)\|}$, ou seja,

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k) \frac{\Delta}{2\|d_{nor}(x_k)\|} d_{nor}(x_k) + h(x_k)\|_2^2 \geq c_1 d_{nor}(x_*)^T \nabla \varphi(x_*) \frac{\Delta}{2\|d_{nor}(x_k)\|}.$$

Mas, de (3.6), temos que $\frac{1}{\|d_{nor}(x_k)\|} \geq \frac{1}{2\|d_{nor}(x_*)\|}$. Segue daí que

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k) \frac{d_{nor}(x_k)}{2\|d_{nor}(x_k)\|} \Delta + h(x_k)\|_2^2 \geq \frac{c_1}{2} d_{nor}(x_*)^T \nabla \varphi(x_*) \frac{\Delta}{2\|d_{nor}(x_*)\|}.$$

Então

$$\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)\frac{d_{nor}(x_k)}{2\|d_{nor}(x_k)\|}\Delta + h(x_k)\|_2^2 \geq c_2 d_{nor}(x_*)^T \nabla \varphi(x_*) \frac{\Delta}{\|d_{nor}(x_*)\|}, \quad (3.9)$$

onde $c_2 = \frac{c_1}{4} = \frac{-\frac{1}{4}}{4} = -\frac{1}{16} < 0$.

Como $\left\| \frac{\Delta}{2\|d_{nor}(x_k)\|} d_{nor}(x_k) \right\| \leq 0.8\Delta$, então segue pela definição de s_{nor}^{dec} (Passo 1 do Algoritmo PQS) que

$$\begin{aligned} \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2 &\geq 0.9 \left[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}^{dec} + h(x_k)\|_2^2 \right] \geq \\ &\geq 0.9 \left[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)\frac{d_{nor}(x_k)}{2\|d_{nor}(x_k)\|}\Delta + h(x_k)\|_2^2 \right] \geq 0.9c_2 d_{nor}(x_*)^T \nabla \varphi(x_*) \frac{\Delta}{\|d_{nor}(x_*)\|}. \end{aligned}$$

Observamos que pela hipótese

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4}[\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2],$$

isto é,

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq c\Delta, \quad (3.10)$$

onde $c = \frac{0.9c_2 d_{nor}(x_*)^T \nabla \varphi(x_*)}{2\|d_{nor}(x_*)\|} = -\frac{0.9d_{nor}(x_*)^T \nabla \varphi(x_*)}{32\|d_{nor}(x_*)\|} > 0$.

Caso 2: Se $\Delta > \Delta'$.

Pela definição de s_{nor}^{dec} (Passo 1 do Algoritmo PQS), temos que

$$\frac{1}{2}\|h'(x_k)s_{nor}^{dec}(x_k, \Delta) + h(x_k)\|_2^2 \leq \frac{1}{2}\|h'(x_k)s_{nor}^{dec}(x_k, \Delta') + h(x_k)\|_2^2. \quad (3.11)$$

Grosseiramente falando, a desigualdade (3.11) nos diz que quando se aumenta Δ , é possível obter um minimizador melhor. Agora, multiplicando (3.11) por (-1) e somando com $\frac{1}{2}\|h(x_k)\|_2^2$, obtemos

$$\begin{aligned} \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}^{dec}(x_k, \Delta) + h(x_k)\|_2^2 &\geq \\ \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}^{dec}(x_k, \Delta') + h(x_k)\|_2^2. \end{aligned}$$

Como (3.10) vale para $\Delta = \Delta'$, segue pelo Passo 2 do Algoritmo PQS que

$$\begin{aligned} \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c(x_k, \Delta) + h(x_k)\|_2^2 &\geq \\ 0.9\left[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}^{dec}(x_k, \Delta) + h(x_k)\|_2^2\right] &\geq \\ 0.9\left[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor}^{dec}(x_k, \Delta') + h(x_k)\|_2^2\right] &\geq \\ \geq 0.9c_2 d_{nor}(x_*)^T \nabla \varphi(x_*) \frac{\Delta'}{\|d_{nor}(x_*)\|} = 2c\Delta', \end{aligned}$$

onde c é o mesmo do Caso 1. Analogamente ao desenvolvimento para a obtenção de (3.10), concluímos que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq c\Delta'. \quad (3.12)$$

Finalmente, de (3.10) e de (3.12), temos o resultado desejado, isto é,

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq c \min\{\Delta, \Delta'\}.$$

□

A próxima hipótese irá substituir as hipóteses sobre derivadas de segunda ordem que geralmente são usadas em regiões de confiança para algoritmos de programação quadrática sequencial. Esta será usada para garantir que a redução prevista $Pred$ da função de mérito é uma aproximação de segunda ordem para a redução verdadeira $Ared$.

(h.2) Hipótese da variação limitada: Para todo $x, y \in \Omega$, temos

$$\|\nabla f(x) - \nabla f(y)\| \leq \mathcal{O}(\|x - y\|) \quad (3.13)$$

e

$$\|h'(x) - h'(y)\| \leq \mathcal{O}(\|x - y\|). \quad (3.14)$$

Observe que as condições de Lipschitz dadas em (3.13) e (3.14) implicam que, para todo $x, y \in \Omega$,

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \mathcal{O}(\|y - x\|^2) \quad (3.15)$$

e

$$\|h(y) - h(x) - h'(x)(y - x)\| \leq \mathcal{O}(\|y - x\|^2). \quad (3.16)$$

Para uma demonstração de (3.15) e (3.16) indicamos [17, p.22 e 75]. A hipótese **(h.2)** serve para garantir a veracidade do fato mencionado em (1.20), isto é,

$$Ared(x_k, \lambda_k, s, \Delta\lambda, \theta) = Pred(x_k, \lambda_k, s, \Delta\lambda, B_k, \theta) + \mathcal{O}(\|s\|^2),$$

o que será bastante útil nas demonstrações dos próximos resultados.

O Lema 3.2 a seguir afirma que, numa vizinhança de um ponto que não é φ -estacionário, o passo tentativo s_c será aceito se o raio da região de confiança for menor do que uma quantidade fixa Δ'' .

Lema 3.2 *Seja $x_* \in \Omega$ um ponto que não é φ -estacionário e considere válidas as hipóteses (h.1) e (h.2). Se o Algoritmo PQS é aplicado a x_k e*

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4} [\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2],$$

então existem $\epsilon, \Delta'' > 0$ tais que para todo $x_k \in \Omega$ e $\|x_k - x_\| \leq \epsilon, \Delta \leq \Delta''$ temos*

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta).$$

Dem. Da definição de $Pred$, temos de (1.20) que

$$\begin{aligned} Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) &= Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) + \mathcal{O}(\|s_c\|^2) \Rightarrow \\ \Rightarrow |Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)| &= \mathcal{O}(\|s_c\|^2), \end{aligned}$$

ou seja, existe $\alpha > 0$ tal que

$$|Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)| \leq \alpha \|s_c\|^2.$$

Como $\|s_c\| \leq \Delta$, segue que

$$|Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)| \leq \alpha \Delta^2. \quad (3.17)$$

Pelo Caso 1 do Lema 3.1, temos que se $\Delta \leq \Delta''$, então $Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq c\Delta$, onde $c > 0$. Daí, obtemos que

$$\frac{1}{|Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)|} \leq \frac{1}{c\Delta}. \quad (3.18)$$

Portanto, de (3.17) e (3.18), obtemos:

$$\begin{aligned} \left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} - 1 \right| &= \left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} \right| \\ &= \frac{|Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)|}{|Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)|} \leq \frac{\alpha \Delta^2}{c\Delta} = \left(\frac{\alpha}{c} \right) \Delta, \end{aligned}$$

o que significa que $\left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} - 1 \right| \leq \mathcal{O}(\Delta)$. Logo

$$\lim_{\Delta \rightarrow 0} \left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} - 1 \right| = 0 \Rightarrow \lim_{\Delta \rightarrow 0} \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} = 1.$$

Então, para Δ suficientemente pequeno temos que $Ared$ e $Pred$ ficam da mesma ordem e satisfazem o teste

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1 Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta).$$

□

No Lema 3.3 abaixo será provado que, se a sequência gerada pelo Algoritmo PQS admite um ponto limite que não é φ -estacionário, então a redução verdadeira $Ared$ da função de mérito é suficientemente positiva. Este fato está relacionado com a não-limitação da função de mérito, essencial na argumentação por absurdo que está sendo construída.

Lema 3.3 *Seja $x_* \in \Omega$ um ponto que não é φ -estacionário e considere válidas as hipóteses (h.1) e (h.2). Seja K_1 um conjunto infinito de índices tal que*

$$\lim_{k \in K_1} x_k = x_*.$$

Então a subsequência $\Delta_k, k \in K_1$ é limitada inferiormente por um número estritamente positivo. Além disso, existe $c > 0$ tal que para $k \in K_1$ suficientemente grande, temos que

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq c.$$

Dem. Se $k \in K_1$ é suficientemente grande, então temos que $\|x_k - x_*\| \leq \epsilon$, onde ϵ é o mesmo do Lema 3.2. Pelo critério de aceitação de Δ no Passo 7 do Algoritmo PQS e pelo Lema 3.2, temos que se $\Delta_k \leq \Delta''$, então Δ_k será aceito, sempre que $\|x_k - x_*\| \leq \epsilon$. Isso significa que Δ_k não precisa ser exageradamente reduzido. Ou seja, existe $r > 0$ tal que $\Delta_k \geq r > 0$, para todo $k \in K_1$ suficientemente grande. Logo, pelo Lema 3.2, temos que

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta_k) \geq 0.1Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k). \quad (3.19)$$

Mas pelo Lema 3.1, temos que existe $c_1 > 0$ tal que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k) \geq c_1 \min\{\Delta_k, \Delta'\}. \quad (3.20)$$

Portanto, de (3.19) e (3.20), obtemos:

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq c,$$

onde $c = 0.1c_1 \min\{\Delta_k, \Delta'\} > 0$, pois $\Delta' > 0$ e $\Delta_k \geq r > 0$. □

A estabilidade assintótica de θ_k será provada no próximo lema, que garantirá a estabilidade da função de mérito para k suficientemente grande.

Lema 3.4 *Seja $\{x_k\}$ uma sequência infinita gerada pelo Algoritmo PQS. Então a sequência $\{\theta_k\}$ é convergente.*

Dem. Temos que a sequência $\{\theta_k^{min}\}$ é não-crescente, pois

$$\theta_k^{min} = \begin{cases} 1, & \text{se } k = 0, \\ \min\{1, \theta_0, \dots, \theta_{k-1}\}, & \text{se } k \geq 1. \end{cases} \quad (3.21)$$

Além disso, $\{\theta_k^{min}\}$ é limitada inferiormente, pois $\theta_k^{min} \geq 0$, para todo $k \in \mathbb{N}$. Portanto, temos que a sequência $\{\theta_k^{min}\}$ é convergente. Pelo Passo 6 do Algoritmo PQS, temos que

$$\theta_k^{grande} = \left[1 + \frac{N}{(k+1)^{1.1}}\right] \theta_k^{min}. \quad (3.22)$$

Fazendo $k \rightarrow \infty$ em (3.22), vemos que a sequência $\{\theta_k^{grande}\}$ converge para o mesmo limite da sequência $\{\theta_k^{min}\}$. Da definição de θ_k no Passo 6 do Algoritmo PQS, temos que $\theta_k \leq \theta_k^{grande}$. De (3.21), é fácil ver que $\theta_{k+1}^{min} \leq \theta_k$. Então

$$\theta_{k+1}^{min} \leq \theta_k \leq \theta_k^{grande}.$$

Portanto, pelo teorema do confronto, a sequência $\{\theta_k\}$ é convergente, e converge para o mesmo limite das sequências $\{\theta_k^{min}\}$ e $\{\theta_k^{grande}\}$. $\square$

A próxima hipótese faz com que a não limitação da função de mérito seja algo impossível.

(h.3) Hipótese de compacidade: A sequência $\{x_k\}$ gerada pelo Algoritmo PQS é limitada.

A hipótese (h.3) é bastante natural quando lidamos com problemas de canalização, pois a sequência $\{x_k\}$ será limitada, se os limites l e u são finitos (Ω é compacto). Pela limitação de $\|B_k\|$ e λ_k , temos que $\ell(x_k, \lambda_k)$, $\varphi(x_k)$ e $\Phi(x_k, \lambda_k, \theta)$ são limitadas, uma vez que ℓ , φ e Φ são funções contínuas.

O teorema a seguir estabelece o primeiro resultado de convergência global do Algoritmo PQS: todo ponto limite é φ -estacionário.

Teorema 6 *Seja $\{x_k\}$ uma sequência infinita gerada pelo Algoritmo PQS e considere válidas as hipóteses (h.1), (h.2) e (h.3). Então todo ponto limite de $\{x_k\}$ é φ -estacionário.*

Dem. Suponha que $x_* \in \Omega$ é um ponto limite de $\{x_k\}$ que não é φ -estacionário. Para simplificar a notação usaremos, para todo $k \in \mathbb{N}$,

$$\begin{cases} \ell_k &= \ell(x_k, \lambda_k) \\ \varphi_k &= \varphi(x_k) \\ \Phi_k &= \Phi(x_k, \lambda_k, \theta_k). \end{cases}$$

Então, para todo $k \in \mathbb{N}$, temos que

$$\begin{aligned} \Phi_{k+1} &= \theta_{k+1}\ell_{k+1} + (1 - \theta_{k+1})\varphi_{k+1} \\ &= \theta_{k+1}\ell_{k+1} + (1 - \theta_{k+1})\varphi_{k+1} - [\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}] + [\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}] \\ &= (\theta_{k+1} - \theta_k)\ell_{k+1} + (\theta_k - \theta_{k+1})\varphi_{k+1} + [\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}] \\ &= (\theta_k - \theta_{k+1})(\varphi_{k+1} - \ell_{k+1}) + [\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}]. \end{aligned}$$

Como $Ared(x_k, \lambda_k, s_k, \Delta\lambda_k, \theta_k) = \Phi(x_k, \lambda_k, \theta_k) - \Phi(x_{k+1}, \lambda_{k+1}, \theta_k)$, então

$$\Phi(x_{k+1}, \lambda_{k+1}, \theta_k) = \Phi(x_k, \lambda_k, \theta_k) - Ared(x_k, \lambda_k, s_k, \Delta\lambda_k, \theta_k).$$

Portanto,

$$[\theta_k \ell_{k+1} + (1 - \theta_k) \varphi_{k+1}] = \Phi_k - \beta_k,$$

onde $\beta_k = Ared(x_k, \lambda_k, s_k, \Delta \lambda_k, \theta_k)$. Segue daí que

$$\Phi_{k+1} = (\theta_k - \theta_{k+1})(\varphi_{k+1} - \ell_{k+1}) + \Phi_k - \beta_k. \quad (3.23)$$

Pelo Passo 6 do Algoritmo PQS, temos que

$$\theta_{k+1} \leq \theta_{k+1}^{grande} = \left[1 + \frac{N}{(k+2)^{1.1}} \right] \theta_{k+1}^{min} \leq \left[1 + \frac{N}{(k+1)^{1.1}} \right] \theta_k,$$

pois $\theta_{k+1}^{min} \leq \theta_k$ e $\frac{1}{(k+2)^{1.1}} \leq \frac{1}{(k+1)^{1.1}}$, para todo $k \in \mathbb{N}$. Logo

$$\begin{aligned} \theta_{k+1} &\leq \left[1 + \frac{N}{(k+1)^{1.1}} \right] \theta_k \Rightarrow \\ \Rightarrow \theta_k - \theta_{k+1} + \frac{\theta_k N}{(k+1)^{1.1}} &\geq 0, \quad \forall k \in \mathbb{N}. \end{aligned} \quad (3.24)$$

Como a sequência $\{x_k\}$ é limitada e $\ell(x, \lambda)$, $\varphi(x)$ e $\Phi(x, \lambda, \theta)$ são funções contínuas, então as sequências $\{\ell_k\}$, $\{\varphi_k\}$ e $\{\Phi_k\}$ são limitadas, uma vez que $\{\|B_k\|\}$ e $\{\lambda_k\}$ também o são. Isso significa que existem $m_1, m_2, m_3 > 0$ tais que $|\ell_k| \leq m_1$, $|\varphi_k| \leq m_2$ e $|\Phi_k| \leq m_3$, para todo $k \in \mathbb{N}$. Então podemos concluir que existe $m > 0$ tal que

$$\varphi_k - \ell_k \leq m, \quad \forall k \in \mathbb{N},$$

pois, $\varphi_k - \ell_k \leq |\varphi_k - \ell_k| \leq |\varphi_k| + |\ell_k| \leq m_1 + m_2 = m$.

Analogamente, temos que $\ell_k - \varphi_k \leq m$, para todo $k \in \mathbb{N}$. Portanto, de (3.23) e (3.24), temos

$$\begin{aligned} \Phi_{k+1} &= \left(\theta_k - \theta_{k+1} + \frac{\theta_k N}{(k+1)^{1.1}} \right) (\varphi_{k+1} - \ell_{k+1}) + \Phi_k - \beta_k - \frac{\theta_k N}{(k+1)^{1.1}} (\varphi_{k+1} - \ell_{k+1}) \\ &\leq \left(\theta_k - \theta_{k+1} + \frac{\theta_k N}{(k+1)^{1.1}} \right) m + \Phi_k - \beta_k + \frac{\theta_k N}{(k+1)^{1.1}} (\ell_{k+1} - \varphi_{k+1}) \\ &\leq \left(\theta_k - \theta_{k+1} + \frac{\theta_k N}{(k+1)^{1.1}} \right) m + \Phi_k - \beta_k + \frac{\theta_k N}{(k+1)^{1.1}} m \\ &= (\theta_k - \theta_{k+1})m + \Phi_k - \beta_k + \frac{2\theta_k N}{(k+1)^{1.1}} m, \end{aligned}$$

de onde podemos concluir que

$$\Phi_{k+1} - \Phi_k \leq (\theta_k - \theta_{k+1})m + \frac{2\theta_k N m}{(k+1)^{1.1}} - \beta_k. \quad (3.25)$$

Escrevendo (3.25) para $k = 0, 1, 2, \dots$, obtemos

$$\left\{ \begin{array}{l} \Phi_1 - \Phi_0 \leq (\theta_0 - \theta_1)m + \frac{2\theta_0 N m}{(0+1)^{1.1}} - \beta_0 \\ \Phi_2 - \Phi_1 \leq (\theta_1 - \theta_2)m + \frac{2\theta_1 N m}{(1+1)^{1.1}} - \beta_1 \\ \vdots \\ \Phi_k - \Phi_{k-1} \leq (\theta_{k-1} - \theta_k)m + \frac{2\theta_{k-1} N m}{[(k-1)+1]^{1.1}} - \beta_{k-1}. \end{array} \right.$$

Somando essas desigualdades, segue que

$$\Phi_k - \Phi_0 \leq (\theta_0 - \theta_k)m + 2Nm \sum_{j=0}^{k-1} \frac{\theta_j}{(j+1)^{1.1}} - \sum_{j=0}^{k-1} \beta_j. \quad (3.26)$$

Agora, $\theta_0 - \theta_k \leq 2$, pois $\theta_0 \leq 1$ e $-\theta_k \leq \theta_k \leq 1$. Além disso, $\sum_{j=0}^{k-1} \frac{\theta_j}{(j+1)^{1.1}} \leq \sum_{j=0}^{k-1} \frac{1}{(j+1)^{1.1}}$, pois $\theta_j \leq 1$, para todo $j \in \mathbb{N}$. Portanto, de (3.26), obtemos

$$\Phi_k \leq 2m + 2Nm \sum_{j=0}^{k-1} \frac{1}{(j+1)^{1.1}} - \sum_{j=0}^{k-1} \beta_j + \Phi_0, \quad \forall k \in \mathbb{N}. \quad (3.27)$$

São as séries $\sum_{j=0}^{\infty} \frac{1}{(j+1)^{1.1}}$ e $\sum_{j=0}^{\infty} \beta_j$ em (3.27) que irão determinar o comportamento de Φ_k , uma vez que N , m e Φ_0 são números fixos. É fácil ver que a série $\sum_{j=0}^{\infty} \frac{1}{(j+1)^{1.1}}$ converge, pois $1.1 > 1$. Pelo Lema 3.3, temos que

$$\beta_k = \text{Ared}(x_k, \lambda_k, s_k, \Delta\lambda_k, \theta_k) \geq c > 0, \quad \forall k \in K_1,$$

onde K_1 é um conjunto infinito de índices. Daí, temos que $\lim_{k \in K_1} \beta_k \neq 0$, o que implica que

a série $\sum_{j=0}^{\infty} \beta_j$ diverge, ou seja, $\sum_{j=0}^{\infty} \beta_j = \infty$, pois $\beta_k > 0$, para todo $k \in \mathbb{N}$.

Portanto, de (3.27), concluímos que Φ_k é ilimitado inferiormente. Isso é uma contradição, pois vimos que $|\Phi_k| \leq m_3$, para todo $k \in \mathbb{N}$. Esse absurdo vem da hipótese de x_* não ser φ -estacionário. $\square$

3.2 O algoritmo encontra um ponto crítico

Como foi mencionado anteriormente, a prova da convergência está dividida em duas partes. Acabamos de estabelecer a primeira, no Teorema 6, onde concluiu-se que todo ponto limite da sequência $\{x_k\}$ gerada pelo Algoritmo PQS é φ -estacionário.

Agora nesta seção trataremos da segunda parte da prova da convergência. Provaremos que, sob certas condições, existe pelo menos um ponto limite que é estacionário para o problema (1.1). Não existe ainda uma prova de que todos os pontos limites são estacionários.

Ao longo de toda a demonstração desta segunda parte usaremos como hipótese de absurdo que nenhum ponto limite da sequência $\{x_k\}$ é estacionário para o problema (1.1).

Outras hipóteses necessárias são:

- (i) Todos os pontos φ -estacionários são factíveis;
- (ii) Todos os pontos factíveis para o problema (1.1) são regulares.

Pela hipótese (i) podemos supor, ao longo da demonstração, que $\lim_{k \rightarrow \infty} \|h(x_k)\| = 0$. Isto significa que o ponto limite da sequência será factível para o problema (1.1).

Na primeira parte desta prova por absurdo, mostraremos que o modelo quadrático do Lagrangeano, dado por

$$Q_k(s) = \frac{1}{2} s^T B_k s + \nabla \ell(x_k, \lambda_k)^T s + \ell(x_k, \lambda_k),$$

tem um bom decréscimo (proporcional a Δ) desde s_{nor} até s_c . Chamaremos esta variação que vai de s_{nor} até s_c de *decréscimo tangencial*.

Na segunda parte da prova por absurdo, examinaremos a combinação convexa

$$\begin{aligned} Pred(x_k, \lambda_k, s, \Delta\lambda, B_k, \theta) &= \theta [Q_k(0) - Q_k(s) - [h'(x_k)s + h(x_k)]^T \Delta\lambda] + \\ &\quad (1 - \theta) \left[\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s + h(x_k)\|_2^2 \right], \end{aligned}$$

Assim como na prova da φ -estacionaridade dos pontos limites (Teorema 6), necessitaremos que $Pred$ seja positivo e proporcional a Δ .

Em outras palavras, precisamos mostrar inicialmente que o decréscimo tangencial de $Q_k(s)$ proporcional a Δ é um bom passo. A seguir, analisaremos a composição de $Pred$. O termo $[Q_k(0) - Q_k(s) - [h'(x_k)s + h(x_k)]^T \Delta\lambda]$ de $Pred$ que multiplica θ é composto, além do decréscimo tangencial, pela variação entre 0 e s_{nor} e pelo termo que envolve a variação dos multiplicadores de Lagrange. Esses dois termos atrapalharão o fato de $Pred$ ter que ser suficientemente positivo. Por outro lado, o termo $[\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s + h(x_k)\|_2^2]$ de $Pred$ que multiplica $(1 - \theta)$ é proporcional a $\|h(x_k)\|$, que tende a zero (lembramos que $\lim_{k \rightarrow \infty} \|h(x_k)\| = 0$). Daí, para conseguirmos que $Pred$ seja positivo e proporcional a Δ , precisaremos que θ não tenda para valores muito próximos de zero.

Uma escolha específica para a direção tangente d_{tan} será necessária. Assim como foi feito com d_{nor} , tomaremos d_{tan} como uma função contínua em $x_k \in \Omega$. Para isso, a escolha de d_{tan} nos Passos 0 e 3 do Algoritmo PQS será dada por

$$d_{tan} = P_{x_k}(-\eta \nabla Q_k(s_{nor})), \quad (3.28)$$

onde $\eta > 0$ é um parâmetro fixo e $P_{x_k}(z)$ é a projeção ortogonal de z no conjunto

$$\mathcal{T} = \{y \in \mathcal{N}(h'(x_k)) \mid l \leq x_k + s_{nor} + y \leq u\}.$$

Assim como a escolha feita para d_{nor} em (3.1), a direção (3.28) pode ser interpretada como o minimizador de um modelo quadrático simples para o Lagrangeano no espaço tangente $\mathcal{T}$. Além disso, a escolha de s_{nor} no Passo 2 do Algoritmo PQS deve ser feita de modo que o seu comprimento seja da mesma ordem que $\|h(x_k)\|$.

(h.4) Primeira hipótese algorítmica: A escolha de s_{nor} no Passo 2 do Algoritmo PQS é feita de tal maneira que

$$\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|). \quad (3.29)$$

Mais adiante veremos que essa condição para a escolha de s_{nor} é bastante natural.

Nos Lemas 3.5 e 3.6 veremos que a continuidade de d_{tan} e o tamanho controlado de s_{nor} nos fazem ganhar propriedades de decréscimo no modelo quadrático do Lagrangeano no espaço tangente, mas isso apenas para pontos que estão perto de se tornarem factíveis.

Para provar o Lema 3.5 necessitaremos das duas proposições seguintes.

Proposição 1 *Considere o conjunto $S = \{x \in \mathbb{R}^n \mid Ax = b, l \leq x \leq u\}$, onde $A \in \mathbb{R}^{m \times n}$. Seja $P(z)$ a projeção ortogonal de $z \in \mathbb{R}^n$ em S . Se $z_1 - z_2 \in \text{Im}(A^T)$, então $P(z_1) = P(z_2)$.*

Dem. Considere o problema

$$\begin{aligned} &\text{minimizar} \quad \frac{1}{2} \|z_1 - x\|_2^2 \\ &\text{sujeito a} \quad x \in S. \end{aligned}$$

Se $S \neq \emptyset$, então existem uma única projeção $P(z_1)$ de z_1 em S e $\lambda \in \mathbb{R}^m$, $\mu, \gamma \in \mathbb{R}^n$ tais que

$$\left\{ \begin{array}{l} P(z_1) - z_1 + A^T \lambda - \mu + \gamma = 0, \\ AP(z_1) = b, \\ (P(z_1) - l)^T \mu = 0, \\ (u - P(z_1))^T \gamma = 0, \\ l \leq P(z_1) \leq u, \\ \mu, \gamma \geq 0. \end{array} \right. \quad (3.30)$$

Agora dado que $z_1 - z_2 \in \text{Im}(A^T)$, então existe $\tilde{\lambda} \in \mathbb{R}^m$ tal que $z_1 - z_2 = A^T \tilde{\lambda}$, ou seja, $z_1 = z_2 + A^T \tilde{\lambda}$. Se chamarmos $\bar{\lambda} = \lambda - \tilde{\lambda}$, temos que (3.30) fica

$$\left\{ \begin{array}{l} P(z_1) - z_2 + A^T \bar{\lambda} - \mu + \gamma = 0, \\ AP(z_1) = b, \\ (P(z_1) - l)^T \mu = 0, \\ (u - P(z_1))^T \gamma = 0, \\ l \leq P(z_1) \leq u, \\ \mu, \gamma \geq 0. \end{array} \right. \quad (3.31)$$

Temos de (3.31) que $P(z_1)$ é a única solução do problema

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} \|z_2 - x\|_2^2 \\ & \text{sujeito a} \quad x \in S. \end{aligned}$$

o que resulta em $P(z_1) = P(z_2)$.

□

Proposição 2 Para todo $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ e $y \in \mathbb{R}^n$, seja $\psi(A, b, y)$ a projeção ortogonal de y no conjunto definido por $x \in \mathbb{R}^n$ tal que

$$\langle a_i, x \rangle \leq b_i, \quad i = 1, \dots, m,$$

onde $a_1^T, \dots, a_m^T$ são as linhas de A e $b = (b_1, \dots, b_m)^T$. Sejam $\bar{A} = (\bar{a}_1^T, \dots, \bar{a}_m^T)^T$, $\bar{b} = (\bar{b}_1, \dots, \bar{b}_m)^T$ e $\bar{y} \in \mathbb{R}^n$ tais que $\bar{x} = \psi(\bar{A}, \bar{b}, \bar{y})$,

$$\langle \bar{a}_i, \bar{x} \rangle = \bar{b}_i, \quad i \in \{1, \dots, p\},$$

$$\langle \bar{a}_i, \bar{x} \rangle < \bar{b}_i, \quad i \in \{p+1, \dots, m\}.$$

Suponha que os vetores $\{\bar{a}_i, i = 1, \dots, p\}$ são linearmente independentes. Então, ψ é contínua em $(\bar{A}, \bar{b}, \bar{y})$.

Dem. Seja $\bar{S} = \{x \in \mathbb{R}^n \mid \langle \bar{a}_i, x \rangle \leq \bar{b}_i, i = 1, \dots, m\}$. Sabemos que a projeção ortogonal $\bar{x} = \psi(\bar{A}, \bar{b}, \bar{y})$ do vetor $\bar{y}$ em $\bar{S}$ é o vetor de $\bar{S}$ mais próximo de $\bar{y}$, isto é,

$$\|\bar{y} - \bar{x}\| \leq \|\bar{y} - x\|, \quad \forall x \in \bar{S}.$$

Portanto, uma maneira de encontrar a projeção $\bar{x}$ é resolver o problema

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} \|\bar{y} - x\|_2^2 \\ & \text{sujeito a} \quad x \in \bar{S}. \end{aligned} \tag{3.32}$$

Escrevendo as condições KKT (ver, por exemplo, [50]) para o problema (3.32), temos que existem $\lambda_1, \lambda_2, \dots, \lambda_p \geq 0$ tais que

$$\begin{aligned} \bar{y} - \bar{x} &= \lambda_1 \bar{a}_1 + \dots + \lambda_p \bar{a}_p; \\ \langle \bar{a}_i, \bar{x} \rangle &= \bar{b}_i, \quad i = 1, \dots, p; \\ \langle \bar{a}_i, \bar{x} \rangle &< \bar{b}_i, \quad i = p+1, \dots, m. \end{aligned} \tag{3.33}$$

Segue daí que a solução $\bar{x}$ de (3.32) é também solução do problema

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} \|\bar{y} - x\|_2^2 \\ & \text{sujeito a} \quad x \in S, \end{aligned} \tag{3.34}$$

onde $S = \{x \in \mathbb{R}^n \mid \langle \bar{a}_i, x \rangle \leq \bar{b}_i, i = 1, \dots, p\}$.

Em outras palavras, estamos afirmando que $\bar{x}$ também é a projeção ortogonal de $\bar{y}$

em S .

O nosso objetivo é mostrar que a projeção ψ é contínua. Para isso, tomamos uma sequência (A^k, b^k, y^k) que converge para $(\bar{A}, \bar{b}, \bar{y})$ e, o nosso trabalho consistirá em provar que a sequência $\psi(A^k, b^k, y^k)$ converge para $\psi(\bar{A}, \bar{b}, \bar{y})$. Com efeito, para todo $k \in \mathbb{N}$, definimos os conjuntos

$$\bar{S}_k = \{x \in \mathbb{R}^n \mid \langle a_i^k, x \rangle \leq b_i^k, \quad i = 1, \dots, m\}$$

e

$$S_k = \{x \in \mathbb{R}^n \mid \langle a_i^k, x \rangle \leq b_i^k, \quad i = 1, \dots, p\}.$$

Seja $\bar{x}^k = \psi(A^k, b^k, y^k)$ a projeção de y^k em $\bar{S}_k$ e seja x^k a projeção de y^k em S_k . Queremos verificar que $\bar{x}^k \rightarrow \bar{x}$. Temos que para todo $k \in \mathbb{N}$, existem conjuntos $I^k \subset \{1, \dots, p\}$ e números $\lambda_1^k, \dots, \lambda_p^k \geq 0$ tais que

$$y^k - x^k = \lambda_1^k a_1^k + \dots + \lambda_p^k a_p^k,$$

$$\langle a_i^k, x^k \rangle \leq b_i^k, \quad i = 1, \dots, p,$$

$$\langle a_i^k, x^k \rangle = b_i^k, \quad \text{se } i \in I^k,$$

$$\lambda_i^k = 0, \quad \text{se } i \notin I^k.$$

Como $I^k \subset \{1, \dots, p\}$, então existe apenas um número finito de conjuntos distintos I^k (no máximo 2^p), para os infinitos índices $k \in \mathbb{N}$. Isto significa que existem infinitos conjuntos I^k que são iguais.

Sem perda de generalidade, consideraremos que $I^1, \dots, I^q$ são os diferentes possíveis subconjuntos de $\{1, \dots, p\}$. Seja $\{y^k, A^k, b^k, x^k\}_{k \in K_j}$ a subsequência correspondente ao subconjunto I^j , onde K_j é um conjunto infinito de índices, para todo $j = 1, \dots, q$. Se $I^j \in \{I^1, \dots, I^q\}$ e $K_j \in \{K_1, \dots, K_q\}$, então para todo $k \in K_j$ temos

$$y^k - x^k = \sum_{i \in I^j} \lambda_i^k a_i^k, \quad \langle a_i^k, x^k \rangle = b_i^k, \quad i \in I^j, \quad (3.35)$$

$$\lambda_i^k \geq 0, \quad \forall i \in I^j, \quad (3.36)$$

$$\langle a_i^k, x^k \rangle \leq b_i^k, \quad i = 1, \dots, p. \quad (3.37)$$

Novamente sem perda de generalidade, podemos reordenar os índices $i \in I^j$ de tal maneira que $I^j = \{1, \dots, r\}$. Seja A_r^k a matriz cujas linhas são as r primeiras linhas de A^k e seja $\bar{b}_r^k = (b_1^k, \dots, b_r^k)^T$. Por (3.35), temos que (x^k, λ^k) resolve o sistema linear $(n+r) \times (n+r)$ dado por

$$\begin{pmatrix} I & (A_r^k)^T \\ A_r^k & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} y^k \\ \bar{b}_r^k \end{pmatrix}.$$

Pela independência linear das p primeiras linhas de $\bar{A}$ e pela convergência de A^k , segue que, para k suficientemente grande, a matriz do sistema linear é não-singular e converge para uma matriz inversível. Portanto, existe (x, λ) tal que

$$\lim_{k \in K_j} x^k = x \quad \text{e} \quad \lim_{k \in K_j} \lambda^k = \lambda.$$

Tomando o limite em (3.35), (3.36) e (3.37), vemos que x satisfaz as condições KKT para a projeção de $\bar{y}$ em S ; logo $x = \bar{x}$. Usando o mesmo raciocínio para cada subsequência definida por $K_1, \dots, K_q$, temos que

$$\lim_{k \rightarrow \infty} x^k = \bar{x}. \quad (3.38)$$

Para k suficientemente grande, temos de (3.33) que

$$\langle a_i^k, x^k \rangle < b_i^k, \text{ se } i > p. \quad (3.39)$$

De (3.35), (3.36), (3.37) e (3.39), temos que x^k é a projeção de y^k em $\bar{S}_k$, isto é, $x^k = \bar{x}^k$. Portanto, de (3.38), temos que $\bar{x}^k \rightarrow \bar{x}$, o que completa a demonstração. $\square$

O nosso próximo resultado é o Lema 3.5, onde provaremos que em uma vizinhança de um ponto x_* factível, regular e não-estacionário para o problema (1.1), a derivada direcional do modelo quadrático do Lagrangeano na direção de d_{tan} está suficientemente longe de zero. Em outras palavras, se o ponto não é estacionário, então é possível sair dele com uma direção de descida.

Lema 3.5 *Seja $\{x_k\}$ a sequência infinita gerada pelo Algoritmo PQS. Considere a hipótese (h.4). Suponha que a subsequência $\{x_k\}_{k \in K_1}$ converge para o ponto $x_* \in \Omega$ que é factível e regular, mas não é um ponto estacionário para o problema (1.1). Então existem $k_1, c > 0$ tais que para todo x_k , com $k \in K_1$ e $k \geq k_1$ temos*

$$-\nabla Q_k(s_{nor}(x_k, \Delta))^T d_{tan}(B_k, x_k, \lambda_k, \Delta) \geq c.$$

Além disso, $\|d_{tan}(B_k, x_k, \lambda_k, \Delta)\|$ é limitada e é suficientemente maior que zero para x_k , $k \in K_1$, $k \geq k_1$.

Dem. De acordo com a escolha para d_{tan} feita em (3.28), temos que

$$d_{tan} = d_{tan}(B_k, x_k, \lambda_k, \Delta) = P_{x_k}(-\eta \nabla Q_k(s_{nor}(x_k, \Delta))).$$

Substituímos $\nabla Q_k(s_{nor}(x_k, \Delta))$ por $[B_k s_{nor}(x_k, \Delta) + \nabla \ell(x_k, \lambda_k)]$, pois

$$Q_k(s) = \frac{1}{2} s^T B_k s + \nabla \ell(x_k, \lambda_k)^T s + \ell(x_k, \lambda_k),$$

e obtemos

$$d_{tan}(B_k, x_k, \lambda_k, \Delta) = P_{x_k}(-\eta [B_k s_{nor}(x_k, \Delta) + \nabla \ell(x_k, \lambda_k)]). \quad (3.40)$$

Pela propriedade contrativa de projeção ortogonal (ver, por exemplo, [12, p.269]), temos que

$$\begin{aligned} \|P_{x_k}(-\eta [B_k s_{nor}(x_k, \Delta) + \nabla \ell(x_k, \lambda_k)]) - P_{x_k}(-\eta \nabla \ell(x_k, \lambda_k))\|_2 &\leq \\ &\leq \eta \|B_k\|_2 \|s_{nor}(x_k, \Delta)\|_2. \end{aligned} \quad (3.41)$$

Da hipótese de (h.4), da limitação de $\|B_k\|_2$ e usando (3.40), a desigualdade (3.41) pode ser reescrita como

$$\|d_{tan}(B_k, x_k, \lambda_k, \Delta) - P_{x_k}(-\eta \nabla \ell(x_k, \lambda_k))\|_2 \leq \mathcal{O}(\|h(x_k)\|). \quad (3.42)$$

Da Proposição 1, temos $P_{x_k}(-\eta \nabla \ell(x_k, \lambda_k)) = P_{x_k}(-\eta \nabla f(x_k))$, pois de (2.12) temos que $[\nabla \ell(x_k, \lambda_k) - \nabla f(x_k)] \in \mathcal{Im}(h'(x_k)^T)$. Portanto, de (3.42) temos

$$\|d_{tan}(B_k, x_k, \lambda_k, \Delta) - P_{x_k}(-\eta \nabla f(x_k))\|_2 \leq \mathcal{O}(\|h(x_k)\|). \quad (3.43)$$

Pela continuidade de $\nabla f(x_k)$ e pela convergência de $\{x_k\}_{k \in K_1}$, segue de (3.43) que

$$|\nabla f(x_k)^T P_{x_k}(-\eta \nabla f(x_k)) - \nabla f(x_k)^T d_{tan}(B_k, x_k, \lambda_k, \Delta)| \leq \mathcal{O}(\|h(x_k)\|), \quad (3.44)$$

para todo $k \in K_1$. Observe que, para todo $k \in K_1$, $P_{x_k}(-\eta \nabla f(x_k))$ é a solução do problema

$$\begin{aligned} & \text{minimizar} && \| -\eta \nabla f(x_k) - z \|_2^2 \\ & \text{sujeito a} && h'(x_k)z = 0 \\ & && l \leq x_k + s_{nor} + z \leq u. \end{aligned} \quad (3.45)$$

Definimos $P_{x_*}(-\eta \nabla f(x_*))$ como a solução do problema

$$\begin{aligned} & \text{minimizar} && \| -\eta \nabla f(x_*) - z \|_2^2 \\ & \text{sujeito a} && h'(x_*)z = 0 \\ & && l \leq x_* + s_{nor} + z \leq u. \end{aligned} \quad (3.46)$$

Como x_* é regular e não é um ponto estacionário para o problema (1.1), segue que $z = 0$ não é uma solução para o problema (3.46). Portanto, temos $P_{x_*}(-\eta \nabla f(x_*)) \neq 0$. Além disso, como $z = 0$ é um ponto factível do problema (3.46), então temos que

$$\| -\eta \nabla f(x_*) - P_{x_*}(-\eta \nabla f(x_*)) \|_2^2 < \| -\eta \nabla f(x_*) \|_2^2,$$

o que implica que $\nabla f(x_*)^T P_{x_*}(-\eta \nabla f(x_*)) < 0$.

Da Proposição 2, temos que $P_{x_k}(-\eta \nabla f(x_k))$ é uma função contínua em x_k , para todo x_k regular. Portanto, existem $c_1, c_2, c_3 > 0$ e $k' \in \mathbb{N}$ tais que, para todo x_k , $k \in K_1$, $k \geq k'$, temos

$$c_1 \leq \|P_{x_k}(-\eta \nabla f(x_k))\| \leq c_2 \quad \text{e} \quad \nabla f(x_k)^T P_{x_k}(-\eta \nabla f(x_k)) \leq -c_3. \quad (3.47)$$

Pela factibilidade do ponto x_* e pela continuidade de $h(x_k)$, temos de (3.43), (3.44) e (3.47) que existe $k'' \geq k'$ tal que

$$\frac{c_1}{2} \leq \|d_{tan}(B_k, x_k, \lambda_k, \Delta)\| \leq 2c_2 \quad \text{e} \quad \nabla f(x_k)^T d_{tan}(B_k, x_k, \lambda_k, \Delta) \leq -\frac{c_3}{2},$$

sempre que $k \in K_1$, $k \geq k''$. Temos portanto a limitação de $\|d_{tan}(B_k, x_k, \lambda_k, \Delta)\|$, e que este valor é suficientemente maior que zero. Além disso, como $d_{tan} \in \mathcal{N}(h'(x_k))$, $\|d_{tan}\|$ e $\|B_k\|_2$ são limitados e $\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|)$, segue que

$$\nabla Q_k(s_{nor})^T d_{tan}(B_k, x_k, \lambda_k, \Delta) = (B_k s_{nor} + \nabla \ell(x_k, \lambda_k))^T d_{tan}(B_k, x_k, \lambda_k, \Delta) =$$

$$= \nabla \ell(x_k, \lambda_k)^T d_{tan} + (B_k s_{nor})^T d_{tan} = \nabla f(x_k)^T d_{tan} + d_{tan}^T B_k s_{nor} \leq -\frac{c_3}{2} + \mathcal{O}(\|h(x_k)\|),$$

de onde obtemos

$$-\nabla Q_k(s_{nor}(x_k, \Delta))^T d_{tan}(B_k, x_k, \lambda_k, \Delta) \geq c,$$

apenas fazendo $c = \frac{c_3}{4}$ e $k_1 \geq k''$ tal que $\mathcal{O}(\|h(x_k)\|) \leq \frac{c_3}{4}$, sempre que $k \in K_1$, $k \geq k_1$. $\square$

O Lema 3.6 complementa o Lema 3.5, mostrando que numa vizinhança de um ponto factível, regular e não-estacionário para o problema (1.1), o decréscimo do modelo quadrático do Lagrangeano (de s_{nor} até s_c) é proporcional ao raio Δ da região de confiança.

Lema 3.6 *Seja $\{x_k\}$ uma sequência infinita gerada pelo Algoritmo PQS. Considere a hipótese (h.4). Suponha que a subsequência $\{x_k\}_{k \in K_1}$ converge para o ponto $x_* \in \Omega$ que é factível e regular, mas não é um ponto estacionário para o problema (1.1). Então existem $k_2, c > 0$, com $k_2 \geq k_1$, $\Delta' \in (0, \Delta_{min})$ tais que*

$$Q_k(s_{nor}(x_k, \Delta)) - Q_k(s_c(x_k, \Delta)) \geq c \min\{\Delta, \Delta'\},$$

sempre que $k \in K_1$, $k \geq k_2$.

Dem. Pelo Passo 4 do Algoritmo PQS, temos que

$$Q_k(s_{nor}) - Q_k(s_c) = Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}) \geq 0.9[Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec})]. \quad (3.48)$$

Escrevendo $s_{tan}^{dec} = t d_{tan}$ e usando que $Q_k(s) = \frac{1}{2} s^T B_k s + \nabla \ell(x_k, \lambda_k)^T s + \ell(x_k, \lambda_k)$, obtemos

$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) = -\frac{1}{2} t^2 d_{tan}^T B_k d_{tan} - t \nabla Q_k(s_{nor})^T d_{tan}. \quad (3.49)$$

Analisaremos o sinal de $d_{tan}^T B_k d_{tan}$ em dois casos. Primeiramente iremos supor que $d_{tan}^T B_k d_{tan} \leq 0$, o que implica que $-\frac{1}{2} t^2 d_{tan}^T B_k d_{tan} \geq 0$. Então, de (3.49) temos

$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) \geq -t \nabla Q_k(s_{nor})^T d_{tan}. \quad (3.50)$$

Pelo Lema 3.5, temos que existem $k_1, c_1 > 0$ tais que quando $k \in K_1$, $k \geq k_1$,

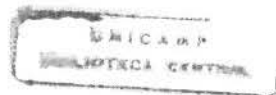
$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) \geq t_{sup}(-\nabla Q_k(s_{nor})^T d_{tan}) \geq t_{sup} c_1, \quad (3.51)$$

onde

$$t_{sup} = \sup\{\tau \geq 0 \mid l \leq x_k + s_{nor} + \tau d_{tan} \leq u, \text{ e } \|s_{nor} + \tau d_{tan}\| \leq \Delta\}. \quad (3.52)$$

Pela definição de d_{tan} no Algoritmo PQS, temos que $l \leq x_k + s_{nor} + d_{tan} \leq u$. Então podemos definir

$$t_1 = \sup\{\tau \in [0, 1] \mid \|s_{nor} + \tau d_{tan}\| \leq \Delta\},$$



de onde teremos $t_{sup} \geq t_1$. Portanto $t_1 = 1$ ou $\|s_{nor} + t_1 d_{tan}\| = \Delta$. No caso em que $\|s_{nor} + t_1 d_{tan}\| = \Delta$, teremos $t_1 \|d_{tan}\| \geq 0.2\Delta$, pois do algoritmo sabemos que $\|s_{nor}\| \leq 0.8\Delta$. Segue daí que $t_{sup} \geq 1$ ou $t_{sup} \geq \frac{0.2\Delta}{\|d_{tan}\|}$, ou seja,

$$t_{sup} \geq \min \left\{ 1, \frac{0.2\Delta}{\|d_{tan}\|} \right\}.$$

Pelo Lema 3.5, temos que existe $c_2 > 0$ tal que $\|d_{tan}\| \leq c_2$, sempre que $k \in K_1$ com $k \geq k_1$. Consequentemente, temos

$$t_{sup} \geq \min \left\{ 1, \frac{0.2\Delta}{\|d_{tan}\|} \right\} \geq \min \left\{ 1, \frac{0.2\Delta}{c_2} \right\} = \frac{0.2}{c_2} \min \left\{ \frac{c_2}{0.2}, \Delta \right\}.$$

Então, da desigualdade (3.51), escrevemos

$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) \geq t_{sup} c_1 \geq \frac{0.2c_1}{c_2} \min \left\{ \frac{c_2}{0.2}, \Delta \right\}, \quad (3.53)$$

sempre que $k \in K_1$, $k \geq k_1$.

Suponhamos agora que $d_{tan}^T B_k d_{tan} > 0$. Se escrevermos $s_{tan}^{dec} = t_{sup} d_{tan}$, onde t_{sup} é dado por (3.52), então, usando a mesma idéia utilizada nos Teoremas 4 e 5, e uma propriedade trivial de funções do segundo grau, obtemos

$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) \geq -\frac{t_{sup}}{2} \nabla Q_k(s_{nor})^T d_{tan}.$$

Usando os mesmos artifícios do caso em que $d_{tan}^T B_k d_{tan} \leq 0$, obtemos que

$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) \geq \frac{1}{2} \left(\frac{0.2c_1}{c_2} \right) \min \left\{ \frac{c_2}{0.2}, \Delta \right\}, \quad (3.54)$$

sempre que $k \in K_1$, $k \geq k_1$.

Finalmente, se escrevermos $s_{tan}^{dec} = t d_{tan}$ com $t < t_{sup}$, então t é o minimizador irrestrito de $Q_k(s_{nor} + t d_{tan})$. Logo,

$$t = -\frac{\nabla Q_k(s_{nor})^T d_{tan}}{d_{tan}^T B_k d_{tan}}$$

e

$$Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) = -\frac{t}{2} \nabla Q_k(s_{nor})^T d_{tan}.$$

Como, pelo Lema 3.5, $\|d_{tan}\|$ e $\|B_k\|$ são limitados para $k \in K_1$, $k \geq k_1$, segue que

$$t = -\frac{\nabla Q_k(s_{nor})^T d_{tan}}{d_{tan}^T B_k d_{tan}} \geq \frac{c_1}{c_3},$$

onde c_3 é um limitante uniforme para $d_{tan}^T B_k d_{tan}$. Portanto, teremos daí que

$$\begin{aligned} Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) &= -\frac{t}{2} \nabla Q_k(s_{nor})^T d_{tan} \\ &\geq \frac{1}{2} \left(\frac{c_1}{c_3} \right) (-\nabla Q_k(s_{nor})^T d_{tan}) \geq \frac{c_1^2}{2c_3}, \end{aligned} \quad (3.55)$$

sempre que $k \in K_1$, $k \geq k_1$. Então de (3.53), (3.54) e (3.55), temos

$$\begin{aligned} Q_k(s_{nor}) - Q_k(s_{nor} + s_{tan}^{dec}) &\geq 0.1 \frac{c_1}{c_2} \min \left\{ \frac{c_2}{0.2}, \frac{10c_1c_2}{2c_3}, \Delta \right\} = \\ &= 0.1 \frac{c_1}{c_2} \min \left\{ \min \left\{ \frac{c_2}{0.2}, \frac{5c_1c_2}{c_3} \right\}, \Delta \right\}. \end{aligned}$$

De (3.48), obtemos que

$$Q_k(s_{nor}) - Q_k(s_c) \geq c \min\{\Delta', \Delta\},$$

sempre que $k \in K_1$, $k \geq k_2$, onde $k_2 \geq k_1$, $c = \frac{0.09c_1}{c_2}$ e $\Delta' = \min \left\{ \Delta_{min}, \frac{c_2}{0.2}, \frac{5c_1c_2}{c_3} \right\}$. $\square$

O Lema 3.7 a seguir preocupa-se com o decréscimo do modelo quadrático do Lagrangeano de 0 até s_c , e não de s_{nor} até s_c , como ocorre no Lema 3.6. De fato, no Lema 3.6 foi provado que o decréscimo de s_{nor} até s_c é proporcional ao raio Δ . Agora, deve ser levado em consideração que pode acontecer um acréscimo no modelo quadrático do Lagrangeano quando se toma um passo de 0 até s_{nor} . A idéia é provar que, quando $\|h(x_k)\|$ é menor do que um múltiplo de Δ , tal acréscimo é insignificante e, então, o decréscimo no modelo quadrático do Lagrangeano é proporcional a Δ , mesmo considerando os termos que dependem da correção feita na aproximação para os multiplicadores $\Delta\lambda$. Como consequência, desde que o primeiro termo da combinação convexa que define $Pred$ seja suficientemente positivo, não será necessário diminuir o parâmetro de penalização θ dentro dessa região.

Lema 3.7 *Considere as hipóteses dos Lemas 3.5 e 3.6. Então, existem γ , c , $k_3 > 0$ tais que, quando $k \in K_1$, $k \geq k_3$ e $\|h(x_k)\| \leq \gamma\Delta$, temos*

$$Q_k(0) - Q_k(s_c(x_k, \Delta)) - [h'(x_k)s_c(x_k, \Delta) + h(x_k)]^T \Delta\lambda \geq c \min\{\Delta, \Delta'\}$$

e

$$\theta^{sup}(x_k, \Delta) = 1,$$

onde Δ' está definido no Lema 3.6.

Dem. Temos que

$$\begin{aligned} Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) &= Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda = \\ &= Q_k(s_{nor}) - Q_k(s_c) - [Q_k(s_{nor}) - Q_k(0)] - [h'(x_k)s_{nor} + h(x_k)]^T \Delta\lambda \geq \\ &\geq Q_k(s_{nor}) - Q_k(s_c) - |Q_k(s_{nor}) - Q_k(0)| - \|h'(x_k)s_{nor} + h(x_k)\| \|\Delta\lambda\|. \end{aligned}$$

Pela convergência de $\{x_k\}_{k \in K_1}$, pela continuidade de $h'(x_k)$ e $\nabla f(x_k, \lambda_k)$, pela limitação de λ_k , $\Delta\lambda_k$ e B_k e por $\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|)$, temos que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) \geq Q_k(s_{nor}) - Q_k(s_c) - \mathcal{O}(\|h(x_k)\|).$$

Então, pelo Lema 3.6, existem $c', k_2 > 0$, com $k_2 \geq k_1, \Delta' \in (0, \Delta_{min})$, tais que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) \geq c' \min\{\Delta, \Delta'\} - \mathcal{O}(\|h(x_k)\|), \quad (3.56)$$

sempre que $k \in K_1, k \geq k_2$. Ou seja, de (3.56), temos que existe $\beta > 0$ tal que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) \geq c' \min\{\Delta, \Delta'\} - \beta \|h(x_k)\|, \quad (3.57)$$

sempre que $k \in K_1, k \geq k_2$. E como por hipótese vale $\|h(x_k)\| \leq \gamma\Delta$, segue de (3.57) que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) \geq c' \min\{\Delta, \Delta'\} - \gamma\beta\Delta. \quad (3.58)$$

Tomando $c = \frac{c'}{2}, \gamma = \frac{c' \min\{\Delta, \Delta'\}}{2\beta\Delta}$ e $k_3 \geq k_2$, obtemos de (3.58) que

$$\begin{aligned} Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) &\geq c' \min\{\Delta, \Delta'\} - \frac{c' \min\{\Delta, \Delta'\}}{2\beta\Delta} \beta\Delta = \\ &= \left(c' - \frac{c'}{2}\right) \min\{\Delta, \Delta'\} = \frac{c'}{2} \min\{\Delta, \Delta'\} = c \min\{\Delta, \Delta'\}, \end{aligned}$$

sempre que $k \in K_1, k \geq k_3$.

Além disso, temos pela definição de θ^{sup} que

$$\theta^{sup} = \sup \left\{ \theta \in [0, 1] \mid Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4} \|h(x_k)\|_2^2 - \frac{1}{4} \|h'(x_k)s_c + h(x_k)\|_2^2 \right\}.$$

Como $\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|)$, temos que

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_c + h(x_k)\|_2^2 = \frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_{nor} + h(x_k)\|_2^2 \leq \mathcal{O}(\|h(x_k)\|^2).$$

Logo, temos

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) - \frac{1}{4} [\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2] \leq c \min\{\Delta, \Delta'\} - \mathcal{O}(\|h(x_k)\|^2).$$

Portanto, de maneira análoga ao desenvolvimento acima, basta escolher γ conveniente que obteremos

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, 1) \geq \frac{1}{4} [\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2],$$

ou seja,

$$\theta^{sup}(x_k, \Delta) = 1.$$

□

No Lema 3.8 é deduzida mais uma consequência da existência de um ponto limite da sequência $\{x_k\}$, que é factível, regular e não-estacionário para o problema (1.1). Basicamente, se θ_k é suficientemente maior que zero, então é possível construir uma sequência $\{Ared(s_k)\}$ de reduções verdadeiras que seja suficientemente maior que zero (ver Lema

3.7). Isso contradiz a hipótese de que $\{x_k\}$ é limitada. Com efeito, pelo Teorema de Bolzano - Weiertrass, temos que $\{x_k\}$ possui uma subsequência convergente. Suponha que

$$\lim_{k \in K_1} x_k = x_*.$$

Como a função de mérito Φ é contínua, segue que

$$\lim_{k \in K_1} \Phi(x_k) = \Phi(x_*).$$

Se θ_k é suficientemente maior que zero, então é possível construir uma sequência $Ared(s_k) \geq \beta > 0$. Como $Ared(s_k) = \Phi(x_k) - \Phi(x_{k+1})$, então temos que $\{\Phi(x_k)\}_{k \in K_1}$ diverge, pois nestas condições temos

$$\lim_{k \in K_1} Ared(s_k) \neq 0.$$

Logo, a subsequência $\{x_k\}_{k \in K_1}$ diverge, o que implica que a sequência $\{x_k\}$ é ilimitada. Portanto, sob as hipóteses dos Lemas 3.5, 3.6 e 3.7, pode ser concluído que $\theta_k \rightarrow 0$.

Vale a pena observar que já havíamos provado no Lema 3.4 que a sequência $\{\theta_k\}$ é convergente. Agora, no Lema 3.8, considerando algumas hipóteses adicionais, veremos que tal sequência converge para zero.

Lema 3.8 *Considere as hipóteses dos Lemas 3.5, 3.6 e 3.7. Considere também as hipóteses (h.1), (h.2) e (h.3). Então*

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

Dem. Suponha, por absurdo, que a sequência dos parâmetros penalizadores $\{\theta_k\}$ não converge para zero. Pelo Lema 3.4, temos que $\{\theta_k\}$ é convergente. Logo, $\{\theta_k\}$ é limitada. Como $\theta_k \in [0, 1]$ e $\theta_k \not\rightarrow 0$, então existem $k_4 \geq k_3$ e $\hat{\theta} > 0$ tais que

$$\theta_k \geq \hat{\theta}, \quad \forall k \geq k_4.$$

Pela definição de $Pred$, temos que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k) \geq \theta_k [Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda], \quad (3.59)$$

pois $\frac{1}{2} [\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2] \geq 0$. Pelo Lema 3.7 temos a existência de $c > 0$ tal que de (3.59) temos

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k) \geq \theta_k c \min\{\Delta, \Delta'\}, \quad (3.60)$$

sempre que $k \in K_1, k \geq k_3$. Segue de (3.60) que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k) \geq \hat{\theta} c \min\{\Delta, \Delta'\}, \quad (3.61)$$

sempre que $k \in K_1, k \geq k_4$. Agora, existe $\tilde{\Delta} \in (0, \Delta') \subset (0, \Delta_{\min})$ tal que quando $\Delta \in [\frac{\tilde{\Delta}}{10}, \tilde{\Delta})$ e $k \in K_1, k \geq k_4$, teremos de (3.61) que

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k) \geq \hat{\theta}c\Delta \geq \frac{\hat{\theta}c\tilde{\Delta}}{10}. \quad (3.62)$$

Assim como foi feito no Lema 3.2, temos por (3.17) que

$$|Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta_k) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k)| \leq \alpha\Delta^2 \leq \alpha\tilde{\Delta}^2, \quad (3.63)$$

onde $\alpha > 0$. Logo de (3.62) e (3.63), temos

$$\left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta_k) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k)} \right| \leq \mathcal{O}(\tilde{\Delta}).$$

Isto significa que quando $\Delta \in [\frac{\tilde{\Delta}}{10}, \tilde{\Delta})$ for suficientemente pequeno, teremos $Ared$ e $Pred$ satisfazendo o teste (1.22) dado por

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta_k) \geq 0.1Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta_k).$$

Então, pelo teste (1.22) e por (3.62), temos que

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta_k) \geq \beta > 0,$$

sempre que $k \in K_1, k \geq k_4$, onde $\beta = 0.01\hat{\theta}c\tilde{\Delta}$. Isto significa que a subsequência $\{\Phi(x_k)\}_{k \in K_1}$ diverge, o que implica que a subsequência $\{x_k\}_{k \in K_1}$ também diverge. Logo, a sequência $\{x_k\}$ é ilimitada, o que contraria a hipótese de compacidade. $\square$

Pensaremos agora no “plano” $(\|h(x)\|, \Delta)$ separado pela “reta” $\|h(x)\| = \gamma\Delta$. Nossos argumentos nos levam a considerar uma *zona boa* do plano, como o conjunto de pontos $(\|h(x)\|, \Delta)$ tais que $\|h(x)\| \leq \gamma\Delta$. Consequentemente, chamaremos de *zona ruim* a região na qual $\|h(x)\| > \gamma\Delta$.

No Lema 3.7 foi mostrado que dentro dessa zona boa, o fator que multiplica o parâmetro θ em $Pred$, isto é, $[Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda]$, é suficientemente grande e o fator que multiplica $(1-\theta)$ em $Pred$, ou seja, $[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2]$, é suficientemente pequeno de tal forma que a condição (1.21), dada por

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4}[\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2],$$

se satisfaz com valores grandes de θ . Portanto, sempre que o par $(\|h(x_k)\|, \Delta)$ está na zona boa, o parâmetro de penalização θ não precisará ser diminuído.

Agora, no Lema 3.9, provaremos que embora possa ser necessário diminuir θ na zona ruim, isto será feito de maneira controlada, isto é, o parâmetro de penalização continuará maior que um múltiplo de Δ . Para provarmos esta propriedade, será necessário fazermos uma hipótese adicional sobre o decréscimo de $[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor} + h(x_k)\|_2^2]$.

(h.5) Segunda hipótese algorítmica: Para cada ponto $x_* \in \Omega$ factível e regular, existem números $c, \epsilon > 0$ tais que, quando $\|h(x_k)\| \geq \gamma\Delta$,

$$\left[\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)s_{nor} + h(x_k)\|_2^2 \right] \geq c\Delta \|h(x_k)\|, \quad (3.64)$$

sempre que $\|x_k - x_*\| \leq \epsilon$.

Veremos agora que é possível escolher s_{nor} satisfazendo tanto a hipótese (h.4) dada por $\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|)$ como a hipótese (h.5) acima dada por (3.64). Com efeito, numa vizinhança de um ponto $x_* \in \Omega$ factível e regular, existe $v_{nor}(x_k) \in \mathbb{R}^n$ tal que

$$\begin{aligned} \|v_{nor}(x_k)\| &\leq \mathcal{O}(\|h(x_k)\|), \\ l &\leq x_k + v_{nor}(x_k) \leq u \end{aligned}$$

e

$$h'(x_k)v_{nor}(x_k) + h(x_k) = 0. \quad (3.65)$$

(Uma escolha possível para $v(x_k)$ pode ser obtida se selecionarmos m variáveis básicas $[x_k]_i$ tais que $l_i \leq [x_k]_i \leq u_i$ e as colunas correspondentes de $h'(x_k)$ formam uma matriz não-singular $B(x_k) \in \mathbb{R}^{m \times m}$. Essa escolha é possível em x_* pela hipótese de regularidade, e numa vizinhança de x_* pela continuidade. O vetor $v(x_k)$ é formado pelas componentes de $-B(x_k)^{-1}h(x_k)$ nas m primeiras coordenadas e por zeros nas $n - m$ coordenadas restantes. É claro que outras escolhas podem ser consideradas.)

Então, podemos dividir nossa análise em duas possibilidades:

(a) Se $\|v_{nor}(x_k)\| \leq 0.8\Delta$, então fazemos $s_{nor}(x_k, \Delta) = v_{nor}(x_k)$. Logo, de (3.65), teremos (3.64) com $c = \frac{1}{\Delta}$;

(b) Se $\|v_{nor}(x_k)\| > 0.8\Delta$, então definimos

$$\tilde{v}_{nor} = \tilde{v}_{nor}(x_k, \Delta) = \frac{0.8\Delta}{\|v_{nor}(x_k)\|} v_{nor}(x_k)$$

e fazemos $s_{nor} = s_{nor}(x_k, \Delta)$ tal que

$$s_{nor} = \begin{cases} \tilde{v}_{nor}(x_k, \Delta), & \text{se } \|h'(x_k)\tilde{v}_{nor} + h(x_k)\|_2^2 \leq \|h'(x_k)s_{nor}^{dec} + h(x_k)\|_2^2, \\ s_{nor}^{dec}(x_k, \Delta), & \text{caso contrário.} \end{cases} \quad (3.66)$$

Escolhendo $d_{nor}(x_k) = P_\Omega(x_k - \alpha \nabla \varphi(x_k)) - x_k$, onde P_Ω é a projeção ortogonal em Ω e $\alpha > 0$ é um número fixo, então teremos $\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|)$. Observamos que

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)\tilde{v}_{nor}(x_k, \Delta) + h(x_k)\|_2^2 \geq \frac{0.8\Delta}{\|v_{nor}(x_k)\|} \frac{1}{2} \|h(x_k)\|_2^2,$$

e como $\|v_{nor}(x_k)\| \leq \mathcal{O}(\|h(x_k)\|)$, então segue que

$$\frac{1}{2} \|h(x_k)\|_2^2 - \frac{1}{2} \|h'(x_k)\tilde{v}_{nor}(x_k, \Delta) + h(x_k)\|_2^2 \geq 0.4\Delta \mathcal{O}(\|h(x_k)\|).$$

Logo, de (3.66), teremos a validade da condição (3.64) da hipótese (h.5).

Lema 3.9 *Considere as hipóteses dos Lemas 3.5, 3.6 e 3.7. Considere também a hipótese (h.5). Se $\|h(x_k)\| \geq \gamma\Delta, k \in K_1, k \geq k_3$, então $\frac{\Delta}{\theta^{sup}(x_k, \Delta)}$ é uniformemente limitado.*

Dem. Pela definição de θ^{sup} , se $\theta^{sup} = \theta^{sup}(x_k, \Delta) \neq 1$, então

$$\theta^{sup} = \frac{\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2}{2[Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta \lambda + \frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2]},$$

o que implica que

$$\frac{1}{2\theta^{sup}(x_k, \Delta)} = 1 + \frac{Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta \lambda}{\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_{nor} + h(x_k)\|_2^2}.$$

Pelas hipóteses, temos que se $k \in K_1, k \geq k_3$ e $\|h(x_k)\| \geq \gamma\Delta$, então

$$\frac{1}{2\theta^{sup}(x_k, \Delta)} \leq 1 + \frac{Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta \lambda}{c\Delta\|h(x_k)\|},$$

pois $\|x_k - x_*\| \leq \epsilon$. Além disso, pelo Lema 3.6, temos que $Q_k(s_c) - Q_k(s_{nor}) \leq 0$. Logo,

$$\begin{aligned} & Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta \lambda = \\ & = [Q_k(s_c) - Q_k(s_{nor})] + [Q_k(s_{nor}) - Q_k(0)] + [h'(x_k)s_{nor} + h(x_k)]^T \Delta \lambda \leq \\ & \leq |Q_k(s_{nor}) - Q_k(0)| + \|h'(x_k)s_{nor} + h(x_k)\| \|\Delta \lambda\|. \end{aligned}$$

Pela convergência de $\{x_k\}_{k \in K_1}$, pela continuidade de $h'(x_k)$ e $\nabla f(x_k, \lambda_k)$, pela limitação de $\|\lambda_k\|$, $\|\Delta \lambda_k\|$ e por $\|B_k\|$ e por $\|s_{nor}(x_k, \Delta)\| \leq \mathcal{O}(\|h(x_k)\|)$, temos que

$$|Q_k(s_{nor}) - Q_k(0)| + \|h'(x_k)s_{nor} + h(x_k)\| \|\Delta \lambda\| \leq \mathcal{O}(\|h(x_k)\|),$$

ou seja,

$$Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta \lambda \leq \mathcal{O}(\|h(x_k)\|).$$

Segue daí que

$$\frac{1}{2\theta^{sup}} \leq 1 + \frac{Q_k(s_c) - Q_k(0) + [h'(x_k)s_c + h(x_k)]^T \Delta \lambda}{c\Delta\|h(x_k)\|} \leq 1 + \frac{\mathcal{O}(\|h(x_k)\|)}{c\Delta\|h(x_k)\|} \leq 1 + \mathcal{O}\left(\frac{1}{\Delta}\right).$$

Portanto, $\frac{\Delta}{\theta^{sup}(x_k, \Delta)}$ é uniformemente limitado. $\square$

O Lema 3.10 é crucial. No Lema 3.7, provamos que o parâmetro de penalização θ não precisa ser diminuído na zona boa. Entretanto, através de sucessivos decréscimos de Δ , é possível entrarmos na indesejável zona ruim. É apenas nesta última situação que pode ser necessário diminuir o parâmetro θ . No Lema 3.10, prova-se que entraremos na

zona ruim não mais que uma vez em cada iteração. Isto significa que, para k suficientemente grande, o primeiro raio da região de confiança Δ será necessariamente aceito. Em outras palavras, para cada iteração k , pode haver apenas uma tentativa Δ dentro da zona ruim. Por essa razão, o parâmetro θ não será excessivamente diminuído. A prova do Lema 3.10 requer uma análise cuidadosa das aproximações de primeira ordem de f e h .

Lema 3.10 *Considere as hipóteses dos Lemas 3.5, 3.6 e 3.7. Considere também a hipótese da variação limitada (h.2) e a hipótese (h.5). Existem $k_6 \geq k_3$, $\tilde{\theta} \in (0, 1]$ tais que, quando $k \in K_1$, $k \geq k_6$, $\|h(x_k)\| \geq \gamma\Delta$ e $\theta \leq \tilde{\theta}$ satisfaz a condição (1.21) dada por*

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4}[\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2],$$

então teremos

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta).$$

Dem. Utilizando a hipótese da variação limitada (h.2), concluímos que

$$\|h(x_k + s_c) - h(x_k) - h'(x_k)s_c\| \leq \mathcal{O}(\|s_c\|^2). \quad (3.67)$$

Além disso, podemos escrever que

$$h(x_k + s_c) = h(x_k) + h'(x_k)s_c + \mathcal{E}(s_c), \quad (3.68)$$

onde a função vetorial $\mathcal{E}(s_c)$ denota o erro. Como $\|s_c\| \leq \Delta$, segue de (3.67) e (3.68) que

$$\|h(x_k + s_c) - h(x_k) - h'(x_k)s_c\| = \|\mathcal{E}(s_c)\| \leq \mathcal{O}(\|s_c\|^2) \leq \mathcal{O}(\Delta^2),$$

ou seja,

$$\|\mathcal{E}(s_c)\| \leq \mathcal{O}(\Delta^2). \quad (3.69)$$

Pela definição da redução verdadeira $Ared$ em (1.16), temos que

$$\begin{aligned} Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) &= \Phi(x_k, \lambda_k, \theta) - \Phi(x_k + s_c, \lambda_k + \Delta\lambda, \theta) \\ &= \theta[\ell(x_k, \lambda_k) - \ell(x_k + s_c, \lambda_k + \Delta\lambda)] + (1 - \theta)[\varphi(x_k) - \varphi(x_k + s_c)]. \end{aligned}$$

Novamente pela hipótese da variação limitada (h.2), temos que

$$\begin{aligned} Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) &= \theta[Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda + \mathcal{O}(\Delta^2)] + \\ &\quad + (1 - \theta)\frac{1}{2}[\|h(x_k)\|_2^2 - \|h(x_k + s_c)\|_2^2]. \end{aligned} \quad (3.70)$$

Por (3.68), temos que

$$\begin{aligned} \|h(x_k + s_c)\|_2^2 &= \|[h'(x_k)s_c + h(x_k)] + \mathcal{E}(s_c)\|_2^2 \\ &= \|h'(x_k)s_c + h(x_k)\|_2^2 + 2[h'(x_k)s_c + h(x_k)]^T \mathcal{E}(s_c) + \|\mathcal{E}(s_c)\|_2^2 \\ &= \|h'(x_k)s_c + h(x_k)\|_2^2 + \|\mathcal{E}(s_c)\|_2^2 + 2s_c^T h'(x_k)^T \mathcal{E}(s_c) + 2h(x_k)^T \mathcal{E}(s_c). \end{aligned}$$

Logo,

$$\begin{aligned} \|h(x_k)\|_2^2 - \|h(x_k + s_c)\|_2^2 &= \|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2 - \|\mathcal{E}(s_c)\|_2^2 \\ &\quad - 2s_c^T h'(x_k)^T \mathcal{E}(s_c) - 2h(x_k)^T \mathcal{E}(s_c). \end{aligned} \quad (3.71)$$

Como x_* é regular e $\|x_k - x_*\| \leq \epsilon$, então temos que $\|h'(x_k)\|$ é limitado, sempre que $k \in K_1, k \geq k_6$. Segue de (3.69) que

$$\begin{aligned} -2s_c^T h'(x_k)^T \mathcal{E}(s_c) &\leq 2\|s_c\| \|h'(x_k)\| \|\mathcal{E}(s_c)\| \leq \mathcal{O}(\Delta^3), \\ -2h(x_k)^T \mathcal{E}(s_c) &\leq 2\|h(x_k)\| \|\mathcal{E}(s_c)\| \leq \|h(x_k)\| \mathcal{O}(\Delta^2) \end{aligned}$$

e

$$-\|\mathcal{E}(s_c)\|_2^2 \leq \mathcal{O}(\Delta^4).$$

Então, temos de (3.71) que

$$\|h(x_k)\|_2^2 - \|h(x_k + s_c)\|_2^2 \leq \|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2 + \|h(x_k)\| \mathcal{O}(\Delta^2) + \mathcal{O}(\Delta^3). \quad (3.72)$$

Substituindo (3.72) em (3.70), obtemos

$$\begin{aligned} Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) &\leq \theta[Q_k(0) - Q_k(s_c) - [h'(x_k)s_c + h(x_k)]^T \Delta\lambda + \mathcal{O}(\Delta^2)] + \\ &\quad + (1 - \theta)[\frac{1}{2}\|h(x_k)\|_2^2 - \frac{1}{2}\|h'(x_k)s_c + h(x_k)\|_2^2 + \|h(x_k)\| \mathcal{O}(\Delta^2) + \mathcal{O}(\Delta^3)], \end{aligned}$$

ou seja,

$$\begin{aligned} Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) &\leq Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) + \theta \mathcal{O}(\Delta^2) + \\ &\quad + (1 - \theta)[\|h(x_k)\| \mathcal{O}(\Delta^2) + \mathcal{O}(\Delta^3)]. \end{aligned}$$

Assim, quando $\|h(x_k)\| \geq \gamma\Delta$, isto é, $\Delta \leq \frac{\|h(x_k)\|}{\gamma}$, usando propriedades de $\mathcal{O}(\cdot)$ (ver, por exemplo, [45, pp.104-107]) temos

$$\mathcal{O}(\Delta^2) = \Delta \mathcal{O}(\Delta) \leq \|h(x_k)\| \mathcal{O}(\Delta) \quad (3.73)$$

e

$$\mathcal{O}(\Delta^3) = \Delta \mathcal{O}(\Delta^2) \leq \|h(x_k)\| \mathcal{O}(\Delta^2). \quad (3.74)$$

Como $(1 - \theta) \leq 1$, segue de (3.73) e (3.74) que

$$\begin{aligned} |Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)| &\leq \theta \|h(x_k)\| \mathcal{O}(\Delta) \\ &\quad + \|h(x_k)\| \mathcal{O}(\Delta^2). \end{aligned} \quad (3.75)$$

Além disso, para $\theta \leq \tilde{\theta}$ e pela hipótese **(h.5)**, temos

$$Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta) \geq \frac{1}{4}[\|h(x_k)\|_2^2 - \|h'(x_k)s_c + h(x_k)\|_2^2] \geq \frac{c}{2}\Delta \|h(x_k)\|, \quad (3.76)$$

sempre que $k \in K_1, k \geq k_6$. Então de (3.75) e (3.76), temos

$$\left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} \right| \leq \frac{\theta \|h(x_k)\| \mathcal{O}(\Delta) + \|h(x_k)\| \mathcal{O}(\Delta^2)}{\frac{\varepsilon}{2} \Delta \|h(x_k)\|},$$

ou seja,

$$\left| \frac{Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) - Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)}{Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta)} \right| \leq \mathcal{O}(\theta) + \mathcal{O}(\Delta), \quad (3.77)$$

sempre que $k \in K_1, k \geq k_6, \|h(x_k)\| \geq \gamma\Delta$ e $\theta \leq \tilde{\theta}$. Logo, da desigualdade (3.77), teremos a validade do teste (1.22) dado por

$$Ared(x_k, \lambda_k, s_c, \Delta\lambda, \theta) \geq 0.1 Pred(x_k, \lambda_k, s_c, \Delta\lambda, B_k, \theta).$$

□

Finalmente, chegamos ao Lema 3.11 que é a segunda parte da prova da convergência do Algoritmo PQS. Nesse lema é provado que se todos os pontos limites da sequência gerada pelo algoritmo são factíveis e regulares, então um deles é necessariamente estacionário. Observe que os resultados anteriores dessa seção foram obtidos sob a hipótese de que existe uma subsequência que converge para um ponto factível, regular e não-estacionário para o problema (1.1). A técnica usada no Lema 3.11 consiste em mostrar que, quando as hipóteses são válidas, uma sequência adicional que necessariamente converge para um ponto estacionário pode ser construída.

Lema 3.11 *Seja $\{x_k\}$ uma sequência infinita gerada pelo Algoritmo PQS. Considere que todos os pontos limites de $\{x_k\}$ são factíveis e regulares. Considere também as hipóteses (h.1), (h.2), (h.3), (h.4) e (h.5). Então, existe um ponto limite da sequência $\{x_k\}$ que é estacionário para o problema (1.1).*

Dem. Pela hipótese de compacidade (h.3), existe uma subsequência $\{x_k\}_{k \in K_0}$ convergente, onde K_0 é um conjunto infinito de índices. Suponha que o ponto limite dessa subsequência não seja estacionário para o problema (1.1). Então, vale o Lema 3.8 para $K_0 = K_1$ e portanto, $\lim_{k \rightarrow \infty} \theta_k = 0$. Daí, novamente pela hipótese de compacidade (h.3), temos a existência de um número $k_7 \in \mathbb{N}$ tal que $\left[1 + \frac{N}{(k_7+1)^{1.1}}\right] \leq 2$ e K_1 , um subconjunto infinito de $\{k_7, k_7+1, k_7+2, \dots\}$ tal que $\{x_k\}_{k \in K_1}$ converge para um ponto $x_* \in \Omega$. Além disso, teremos que

$$\lim_{k \in K_1} \theta^{sup}(x_k, \bar{\Delta}_k) = 0 \quad (3.78)$$

e, pelo Passo 6 do Algoritmo PQS,

$$\theta^{sup}(x_k, \bar{\Delta}_k) \leq \theta'(x_k, \bar{\Delta}_k) \leq \theta_k^{grande}, \quad (3.79)$$

para todo $k \in K_1$, onde $\bar{\Delta}_k$ é um dos raios da região de confiança testados na iteração k . Temos, pelo Passo 6 do Algoritmo PQS, que a segunda desigualdade de (3.79) sempre vale, enquanto que a primeira desigualdade deve acontecer para um conjunto infinito de índices (caso contrário, θ^{sup} não tenderia para zero).

Por hipótese, temos a regularidade de x_* . Se x_* for estacionário para o problema (1.1), então o resultado está provado. Se não for estacionário, então valem as hipóteses dos Lemas 3.5 - 3.10 para $\{x_k\}_{k \in K_1}$. Sem perda de generalidade, podemos assumir que $k \geq k_6 \geq k_3$,

$$\theta^{sup}(x_k, \bar{\Delta}_k) \leq \frac{\bar{\theta}}{2} < 1 \quad \text{e} \quad \theta_k \leq \frac{\bar{\theta}}{2}, \quad (3.80)$$

para todo $k \in K_1$, onde k_3 é dado pelo Lema 3.7 e k_6 é dado pelo Lema 3.10. Pelo Lema 3.7, temos que $\theta^{sup}(x_k, \Delta) = 1$, sempre que $k \in K_1$ e $\|h(x_k)\| \leq \gamma\Delta$. Então, de (3.78) e (3.80), temos que

$$\|h(x_k)\| > \gamma\bar{\Delta}_k, \quad (3.81)$$

para todo $k \in K_1$. Portanto, como $\|h(x_k)\| \rightarrow 0$, segue que

$$\lim_{k \in K_1} \bar{\Delta}_k = 0. \quad (3.82)$$

Novamente sem perda de generalidade, assumiremos que

$$\bar{\Delta}_k \leq 0.1\Delta' < 0.1\Delta_{min}, \quad (3.83)$$

para todo $k \in K_1$, onde Δ' foi definido no Lema 3.6 e usado no Lema 3.7. Por (3.83), temos que o raio $\bar{\Delta}_k$ não pode ser o primeiro raio da região de confiança testado na iteração k . Seja $\hat{\Delta}_k$, então, o raio da região de confiança testado na iteração k imediatamente antes de $\bar{\Delta}_k$. Seja $\hat{\theta}_k$ o parâmetro de penalidade associado com o passo em que o raio $\hat{\Delta}_k$ foi rejeitado. Como $k_7 \geq N$, segue por (3.80) e pela escolha do parâmetro de penalização no Passo 6 do Algoritmo PQS que $\hat{\theta}_k \leq \bar{\theta}$, para todo $k \in K_1$. Portanto, podemos aplicar o Lema 3.10, de onde teremos

$$\|h(x_k)\| < \gamma\hat{\Delta}_k, \quad \forall k \in K_1. \quad (3.84)$$

Além disso, pelo Passo 7 do Algoritmo PQS, sabemos que $\bar{\Delta}_k \in [0.1\hat{\Delta}_k, 0.9\hat{\Delta}_k]$. Então, por (3.83), temos que

$$\hat{\Delta}_k \leq 10\bar{\Delta}_k \leq \Delta' < \Delta_{min}, \quad \forall k \in K_1. \quad (3.85)$$

Agora, pelo Passo 6 do Algoritmo PQS, por (3.78) e pelo Lema 3.7, temos

$$\begin{aligned} \hat{\theta}_k &= \min\{\theta'(x_k, \hat{\Delta}_k), \theta^{sup}(x_k, \hat{\Delta}_k)\} = \theta'(x_k, \hat{\Delta}_k) \\ &\geq \min\{\theta'(x_k, \hat{\Delta}_k), \theta^{sup}(x_k, \bar{\Delta}_k)\} = \theta^{sup}(x_k, \bar{\Delta}_k), \end{aligned} \quad (3.86)$$

para todo $k \in K_1$. Pelo Lema 3.7, por (3.85) e pela definição de $Pred$, temos que existe $c > 0$ tal que

$$Pred(x_k, \hat{\theta}_k, \hat{\Delta}_k) \geq \hat{\theta}_k c \hat{\Delta}_k. \quad (3.87)$$

Além disso, repetindo os mesmos argumentos do Lema 3.10, teremos que

$$\begin{aligned} & |Ared(x_k, \hat{\theta}_k, \hat{\Delta}_k) - Pred(x_k, \hat{\theta}_k, \hat{\Delta}_k)| \leq \\ & \leq \hat{\theta}_k \mathcal{O}(\hat{\Delta}_k^2) + (1 - \hat{\theta}_k) [\|h(x_k)\| \mathcal{O}(\hat{\Delta}_k^2) + \mathcal{O}(\hat{\Delta}_k^3)], \end{aligned} \quad (3.88)$$

para todo $k \in K_1$. Logo, de (3.87) e (3.88), temos

$$\begin{aligned} \left| \frac{Ared(x_k, \hat{\theta}_k, \hat{\Delta}_k) - Pred(x_k, \hat{\theta}_k, \hat{\Delta}_k)}{Pred(x_k, \hat{\theta}_k, \hat{\Delta}_k)} \right| & \leq \frac{\hat{\theta}_k \mathcal{O}(\hat{\Delta}_k^2) + (1 - \hat{\theta}_k) [\|h(x_k)\| \mathcal{O}(\hat{\Delta}_k^2) + \mathcal{O}(\hat{\Delta}_k^3)]}{\hat{\theta}_k c \hat{\Delta}_k} \\ & = \mathcal{O}(\hat{\Delta}_k) + (1 - \hat{\theta}_k) \frac{[\|h(x_k)\| \mathcal{O}(\hat{\Delta}_k) + \mathcal{O}(\hat{\Delta}_k^2)]}{\hat{\theta}_k}, \end{aligned} \quad (3.89)$$

para todo $k \in K_1$.

Agora, pelo Lema 3.9 e por (3.81), temos que $\frac{\hat{\Delta}_k}{\theta_{sup}(x_k, \hat{\Delta}_k)}$ é limitado para $k \in K_1$. Logo, de (3.85) e (3.86), temos que $\frac{\mathcal{O}(\hat{\Delta}_k)}{\hat{\theta}_k}$ também é limitado para $k \in K_1$. Finalmente, por (3.82), por (3.85) e pela factibilidade de x_* , temos que o termo do lado direito de (3.89) tende para zero, para $k \in K_1$. Isto implica que para k suficientemente grande, o critério de aceitação (1.22) irá valer com o raio de confiança $\hat{\Delta}_k$. Isto é uma contradição, pois assumimos que o raio $\hat{\Delta}_k$ foi rejeitado. $\square$

O Teorema 7 reúne os resultados provados nas seções 3.1 e 3.2, sistematizando a convergência global do Algoritmo PQS.

Teorema 7 *Seja $\{x_k\}$ a sequência infinita gerada pelo Algoritmo PQS. Considere as hipóteses (h.1), (h.2), (h.3), (h.4) e (h.5). Então, todos os pontos limites de $\{x_k\}$ são φ -estacionários. Além disso, se todos os pontos limites de $\{x_k\}$ são factíveis e regulares, então existe um ponto limite x_* que é estacionário para o problema (1.1). Em particular, se todos os pontos φ -estacionários são factíveis e regulares, então existe uma subsequência de $\{x_k\}$ que converge para um ponto factível, regular e estacionário para o problema (1.1).*

Dem. O resultado segue diretamente do Teorema 6 e do Lema 3.11. $\square$

Capítulo 4

Uma abordagem para os subproblemas

Conforme foi mencionado anteriormente, o objetivo principal deste trabalho está em resolver o problema geral de programação não-linear em sua forma padrão, apresentado em (1.1). No entanto, vimos que é possível gerar uma sequência de subproblemas quadráticos, dados em (1.10) e (1.11), para resolver (1.1) através de métodos de programação quadrática sequencial. Com base nestes subproblemas, neste capítulo discutiremos possibilidades de implementação dos Passos 0, 2, 3, e 4 do Algoritmo PQS, bem como escolhas convenientes para a matriz B_k , Hessiana do modelo quadrático.

4.1 Resultado de equivalência

Em [24], os autores propõem uma estratégia para minimizar uma função convexa num *politopo*¹ limitado, baseada na minimização de funções não-lineares em caixas. Tendo em vista a disponibilidade de algoritmos eficientes para a minimização de funções gerais em caixas (por exemplo, métodos de região de confiança [15, 27] ou métodos de gradiente espectral projetado [3]), analisaremos como explorar o resultado de equivalência de [24] na resolução dos nossos subproblemas. Apresentaremos a seguir o resultado de equivalência de [24] que apoiará nossas escolhas. A demonstração foi ligeiramente adaptada para permitir variáveis canalizadas no problema e será apresentada para deixar nosso trabalho completo.

Consideremos o problema

$$\begin{array}{ll} \text{minimizar} & g(x) \\ \text{sujeito a} & Ax = b \\ & c \leq x \leq d, \end{array} \quad (4.1)$$

¹ Um politopo é a interseção de um número finito de semi-espacos fechados [50, p.468].

onde $g \in \mathcal{C}^2(\mathbb{R}^n)$ é uma função convexa. Seja $\mathcal{W} = \{x \in \mathbb{R}^n \mid Ax = b, c \leq x \leq d\}$.

As condições (KKT) de otimalidade de primeira ordem de (4.1) são

$$\begin{cases} \nabla g(x) + A^T y - z + w = 0 \\ Ax - b = 0 \\ (x - c)^T z = 0 \\ (d - x)^T w = 0 \\ z \geq 0, w \geq 0 \\ c \leq x \leq d. \end{cases} \quad (4.2)$$

Seja

$$\Lambda(x, y, z, w) = \frac{1}{2} [\|\nabla g(x) + A^T y - z + w\|_2^2 + \|Ax - b\|_2^2 + ((x - c)^T z)^2 + ((d - x)^T w)^2],$$

e consideremos o problema

$$\begin{aligned} &\text{minimizar} && \Lambda(x, y, z, w) \\ &\text{sujeito a} && c \leq x \leq d, \\ &&& z \geq 0, w \geq 0. \end{aligned} \quad (4.3)$$

Como a função objetivo Λ não é convexa, ao resolvermos (4.3) esperaríamos apenas encontrar pontos estacionários, não necessariamente minimizadores globais. O resultado a seguir, no entanto, garante que todo ponto estacionário de (4.3) é um minimizador global deste problema que satisfaz as condições dadas em (4.2). Portanto, resolver (4.3) é equivalente a resolver (4.1). Para extensões e variações deste teorema veja, por exemplo, [1, 24, 25, 26].

Teorema 8 *Se $g \in \mathcal{C}^2(\mathbb{R}^n)$ é uma função convexa e o politopo $\mathcal{W}$ é não-vazio e limitado, então o problema (4.3) admite pelo menos um ponto estacionário (KKT) e todo ponto estacionário (x^*, y^*, z^*, w^*) de (4.3) é um minimizador global com $\Lambda(x^*, y^*, z^*, w^*) = 0$.*

Dem. A primeira parte é imediata. Como $\mathcal{W}$ é limitado e g é contínua, existe um minimizador global para o problema (4.1). Este minimizador tem que satisfazer (4.2) e, portanto, é um minimizador global de (4.3).

Vamos supor que (x, y, z, w) seja um ponto estacionário do problema (4.3). Então existem $\gamma, \mu, \nu, \xi \in \mathbb{R}^n$ tais que

$$\begin{aligned} &\nabla^2 g(x)(\nabla g(x) + A^T y - z + w) + A^T(Ax - b) + \\ &+ [(x - c)^T z]z - \nu - [(d - x)^T w]w + \xi = 0, \end{aligned} \quad (4.4)$$

$$A(\nabla g(x) + A^T y - z + w) = 0, \quad (4.5)$$

$$-(\nabla g(x) + A^T y - z + w) + [(x - c)^T z](x - c) - \mu = 0, \quad (4.6)$$

$$(\nabla g(x) + A^T y - z + w) + [(d - x)^T w](d - x) - \gamma = 0, \quad (4.7)$$

$$\nu^T(x - c) = 0, \quad \xi^T(d - x) = 0, \quad (4.8)$$

$$\mu^T z = 0, \quad \gamma^T w = 0, \quad (4.9)$$

$$c \leq x \leq d, \quad z \geq 0, \quad w \geq 0, \quad \gamma \geq 0, \quad \mu \geq 0, \quad \nu \geq 0, \quad \xi \geq 0. \quad (4.10)$$

Por (4.5), (4.6) e (4.7) temos que

$$[(x - c)^T z](x - c) - \mu \in \mathcal{N}(A) \quad (4.11)$$

e

$$-[(d - x)^T w](d - x) + \gamma \in \mathcal{N}(A) \quad (4.12)$$

onde $\mathcal{N}(A)$ é o núcleo da matriz A .

Assim, pré-multiplicando (4.4) por $[(x - c)^T z](x - c) - \mu$ e usando (4.5), (4.6) e (4.11) vem

$$\begin{aligned} & [[(x - c)^T z](x - c) - \mu]^T \nabla^2 g(x) [[(x - c)^T z](x - c) - \mu] + \\ & + [[(x - c)^T z](x - c) - \mu]^T [[(x - c)^T z]z - \nu - [(d - x)^T w]w + \xi] = 0. \end{aligned} \quad (4.13)$$

Como $\nabla^2 g$ é semi-definida positiva, segue de (4.13) que

$$[[(x - c)^T z](x - c) - \mu]^T [[(x - c)^T z]z - \nu - [(d - x)^T w]w + \xi] \leq 0. \quad (4.14)$$

Desenvolvendo (4.14) temos

$$\begin{aligned} & [(x - c)^T z]^3 - (x - c)^T z(x - c)^T \nu - (x - c)^T z(d - x)^T w(x - c)^T w + \\ & + (x - c)^T z(x - c)^T \xi - (x - c)^T z\mu^T z + \mu^T \nu + (d - x)^T w\mu^T w - \mu^T \xi \leq 0, \end{aligned}$$

e usando (4.8) e (4.9) nesta última desigualdade resulta em

$$\begin{aligned} & [(x - c)^T z]^3 + (x - c)^T z(x - c)^T \xi \\ & - (x - c)^T z(d - x)^T w(x - c)^T w + (d - x)^T w\mu^T w + \mu^T \nu - \mu^T \xi \leq 0. \end{aligned} \quad (4.15)$$

Analogamente, pré-multiplicando (4.4) por $-[(d - x)^T w](d - x) + \gamma$ e usando (4.5), (4.7) e (4.12) vem

$$\begin{aligned} & [-[(d - x)^T w](d - x) + \gamma]^T \nabla^2 g(x) [-[(d - x)^T w](d - x) + \gamma] + \\ & [-[(d - x)^T w](d - x) + \gamma]^T [[(x - c)^T z]z - \nu - [(d - x)^T w]w + \xi] = 0, \end{aligned} \quad (4.16)$$

e então, como $\nabla^2 g(x) \geq 0$, segue que

$$[-[(d - x)^T w](d - x) + \gamma]^T [[(x - c)^T z]z - \nu - [(d - x)^T w]w + \xi] \leq 0. \quad (4.17)$$

Desenvolvendo (4.17) temos

$$\begin{aligned} & -(d-x)^T w (d-x)^T z (x-c)^T z + (d-x)^T w (d-x)^T \nu + [(d-x)^T w]^3 \\ & -(d-x)^T w (d-x)^T \xi + (x-c)^T z \gamma^T z - \gamma^T \nu - (d-x)^T w \gamma^T w + \xi^T \gamma \leq 0, \end{aligned}$$

e usando (4.8) e (4.9) nesta última desigualdade implica em

$$\begin{aligned} & [(d-x)^T w]^3 + (d-x)^T w (d-x)^T \nu \\ & -(d-x)^T w (d-x)^T z (x-c)^T z - \gamma^T \nu + \gamma^T \xi + (x-c)^T z \gamma^T z \leq 0. \end{aligned} \quad (4.18)$$

Agora, somando (4.6) com (4.7) obtemos

$$[(x-c)^T z](x-c) - \mu + [(d-x)^T w](d-x) - \gamma = 0. \quad (4.19)$$

Então, pré-multiplicando (4.19) por ν^T , e utilizando (4.8) temos

$$\begin{aligned} & [(x-c)^T z][\nu^T(x-c)] - \nu^T \mu + [(d-x)^T w][\nu^T(d-x)] - \nu^T \gamma = 0 \Rightarrow \\ & \Rightarrow [(d-x)^T w][(d-x)^T \nu] = \nu^T \mu + \nu^T \gamma. \end{aligned} \quad (4.20)$$

Pré-multiplicando (4.19) por ξ^T , e utilizando (4.8) vem

$$\begin{aligned} & [(x-c)^T z][\xi^T(x-c)] - \xi^T \mu + [(d-x)^T w][\xi^T(d-x)] - \xi^T \gamma = 0 \Rightarrow \\ & \Rightarrow [(x-c)^T z][(x-c)^T \xi] = \xi^T \mu + \xi^T \gamma. \end{aligned} \quad (4.21)$$

Pré-multiplicando (4.19) por z^T , segue de (4.8) que

$$\begin{aligned} & [(x-c)^T z]^2 - \mu^T z + (d-x)^T w (d-x)^T z - \gamma^T z = 0 \Rightarrow \\ & \Rightarrow \gamma^T z = [(x-c)^T z]^2 + (d-x)^T w (d-x)^T z. \end{aligned} \quad (4.22)$$

Finalmente, pré-multiplicando (4.19) por w^T , e usando (4.8) fica

$$\begin{aligned} & (x-c)^T z (x-c)^T w - \mu^T w + [(d-x)^T w]^2 - \gamma^T w = 0 \Rightarrow \\ & \Rightarrow \mu^T w = [(d-x)^T w]^2 + (x-c)^T z (x-c)^T w. \end{aligned} \quad (4.23)$$

Agora, substituindo (4.21) e (4.23) em (4.15) vem

$$\begin{aligned} & [(x-c)^T z]^3 + \xi^T \mu + \xi^T \gamma - (x-c)^T z (d-x)^T w (x-c)^T w + \\ & + \mu^T \nu - \mu^T \xi + [(d-x)^T w]^3 + [(d-x)^T w](x-c)^T z (x-c)^T w \leq 0 \Rightarrow \\ & \Rightarrow [(x-c)^T z]^3 + [(d-x)^T w]^3 + \mu^T \nu + \xi^T \gamma \leq 0. \end{aligned} \quad (4.24)$$

Logo, de (4.10) e (4.24), temos

$$(x-c)^T z = 0, \quad (d-x)^T w = 0, \quad \nu^T \mu = 0, \quad \xi^T \gamma = 0. \quad (4.25)$$

Analogamente, substituindo (4.20) e (4.22) em (4.18) obtemos

$$[(d-x)^T w]^3 + \nu^T \mu + \nu^T \gamma - (d-x)^T w (d-x)^T z (x-c)^T z + \\ + \gamma^T \xi - \gamma^T \nu + [(x-c)^T z]^3 + (x-c)^T z [(d-x)^T w (d-x)^T z] \leq 0,$$

que é a desigualdade (4.24).

Por (4.6) e (4.25),

$$-(\nabla g(x) + A^T y - z + w) = \mu \geq 0. \quad (4.26)$$

Mas, por (4.5), $-(\nabla g(x) + A^T y - z + w) \in \mathcal{N}(A)$. Portanto, como $\mathcal{W}$ é limitado, a equação (4.26) implica necessariamente em

$$-(\nabla g(x) + A^T y - z + w) = 0. \quad (4.27)$$

Então, por (4.4), (4.25) e (4.27) temos

$$A^T(Ax - b) = \nu - \xi. \quad (4.28)$$

Agora, (4.8), (4.10) e (4.28) são as condições de otimalidade (necessárias e suficientes) do problema quadrático convexo

$$\begin{aligned} &\text{Minimizar} \quad \frac{1}{2} \|Ax - b\|^2 \\ &\text{sujeita a} \quad c \leq x \leq d. \end{aligned}$$

Como $\mathcal{W}$ é não vazio, temos que $Ax = b$. Esta igualdade, juntamente com (4.25) e (4.27) completam a prova. $\square$

4.2 Os subproblemas do Algoritmo PQS

Com base na idéia desenvolvida na seção 4.1 discutiremos a resolução aproximada de subproblemas em alguns dos passos que compõem o Algoritmo PQS.

No Passo 0 do Algoritmo PQS precisamos obter d_{nor} satisfazendo (2.1)-(2.2). Para isso, basta fazer $d_{nor} = y^* - x_k$, onde y^* é a solução do problema

$$\begin{aligned} &\text{minimizar} \quad \frac{1}{2} \|x_k - \alpha \nabla \varphi(x_k) - y\|_2^2 \\ &\text{sujeito a} \quad l \leq y \leq u, \end{aligned} \quad (4.29)$$

onde $\alpha > 0$ é um parâmetro fixo. O problema (4.29) consiste em minimizar uma função quadrática simples numa caixa e, portanto, basta projetar o ponto $x_k - \alpha \nabla \varphi(x_k)$ sobre a caixa.

Ainda no Passo 0 queremos encontrar d_{tan} satisfazendo (2.3)-(2.5). Assim sendo, fazemos $d_{tan} = z^*$, onde z^* é a solução do problema

$$\begin{aligned}
& \text{minimizar} && \frac{1}{2} \| -\eta \nabla f(x_k) - z \|_2^2 \\
& \text{sujeito a} && h'(x_k)z = 0 \\
& && l \leq x_k + z \leq u,
\end{aligned} \tag{4.30}$$

onde $\eta > 0$ é um parâmetro fixo. Observamos que a função objetivo do problema (4.30) é uma quadrática cuja Hessiana é a matriz identidade e, portanto, uma função convexa. Além disso, a região de factibilidade do problema (4.30), formada por restrições lineares de igualdade e uma caixa, é um politopo não-vazio, pois $0 \in \mathbb{R}^n$ é factível. Esse politopo será limitado sempre que a caixa original o for. Nesse caso, para encontrarmos a solução z^* de (4.30), aplicamos a estratégia apresentada no Teorema 8.

No Passo 2 precisamos calcular s_{nor} satisfazendo (2.6). Desta forma, fazemos $s_{nor} = \tilde{s}$, onde $\tilde{s}$ é uma solução aproximada do problema

$$\begin{aligned}
& \text{minimizar} && \frac{1}{2} \| h'(x_k)s + h(x_k) \|_2^2 \\
& \text{sujeito a} && \|s\|_\infty \leq 0.8\Delta \\
& && l \leq x_k + s \leq u
\end{aligned} \tag{4.31}$$

que satisfaz a desigualdade (2.6). Assim como no problema (4.29), temos em (4.31) um problema de minimização de uma quadrática canalizada e, podemos utilizar a mesma estratégia empregada em (4.29) na resolução de (4.31).

No Passo 3 queremos encontrar a direção d_{tan} que satisfaça (2.7)-(2.9). Assim sendo, fazemos $d_{tan} = \bar{z}$, onde $\bar{z}$ é a solução do problema

$$\begin{aligned}
& \text{minimizar} && \frac{1}{2} \| -\eta \nabla Q_k(s_{nor}) - z \|_2^2 \\
& \text{sujeito a} && h'(x_k)z = 0 \\
& && l \leq x_k + s_{nor} + z \leq u,
\end{aligned} \tag{4.32}$$

com $\eta > 0$ fixo. Assim como em (4.30), temos em (4.32) um problema de minimizar uma função convexa numa caixa e com restrições lineares de igualdade. Então, de maneira análoga, encontramos a solução $\bar{z}$ de (4.32) aplicando os mesmos argumentos apresentados para o problema (4.30). Na prática, nunca calculamos d_{tan} e esse vetor só é usado para obtermos resultados teóricos. Para efeito de implementação, costuma-se utilizar $d_{tan} = 0$.

Finalmente, iremos analisar o Passo 4. Consideremos o problema

$$\begin{aligned}
& \text{minimizar} && Q_k(s) \\
& \text{sujeito a} && h'(x_k)s = 0 \\
& && \|s_{nor} + s\|_\infty \leq \Delta \\
& && l \leq x_k + s_{nor} + s \leq u,
\end{aligned} \tag{4.33}$$

onde x_k é não-factível e $Q_k(s) = \frac{1}{2} s^T B_k s + \nabla \ell(x_k, \lambda_k)^T s + \ell(x_k, \lambda_k)$. Podemos calcular s_{tan} quando x_k é não-factível, apenas batizando $s_{tan} = \hat{s}$, onde $\hat{s}$ é uma solução aproximada

de (4.33) que verifica (2.11). Podemos dessa maneira identificar o problema (4.1) com o problema (4.33), apenas trocando a matriz A por $h'(x_k)$, x por s e b por 0. Observamos que a região factível do problema (4.33) é não-vazia ($0 \in \mathbb{R}^n$ é factível) e limitada (Δ é um número real). Além disso, essa mesma região é composta pela interseção de duas caixas, que é ainda uma caixa, e pode ser escrita como $c \leq s \leq d$, onde as componentes de c e d são respectivamente $c_i \equiv \max\{-\Delta - (s_{nor})_i, l_i - (x_k)_i - (s_{nor})_i\}$ e $d_i \equiv \min\{\Delta - (s_{nor})_i, u_i - (x_k)_i - (s_{nor})_i\}$, para $i = 1, \dots, n$. Portanto, para podermos aplicar o Teorema 8, precisaremos apenas que a função objetivo de (4.33) seja convexa. Para que isso aconteça, a matriz B_k deve ser semi-definida positiva, isto é, $s^T B_k s \geq 0$, para todo $s \in \mathbb{R}^n$. Encontramos assim, a primeira pista para escolher a matriz B_k do modelo quadrático, o que será discutido na próxima seção.

4.3 Escolhas para a matriz do modelo quadrático

Lembramos que até o presente momento, pensávamos na matriz B_k do modelo quadrático apenas como uma aproximação para a matriz Hessiana $\nabla^2 \ell(x_k, \lambda_k)$. Cabe mencionar que, quando a matriz B_k é a própria Hessiana da função objetivo do problema (1.1), perdemos informações sobre a curvatura das restrições. Vejamos isso graficamente na Figura 4.1. Imaginemos que a função objetivo f seja quadrática. Logo, a aproximação quadrática é ela própria.

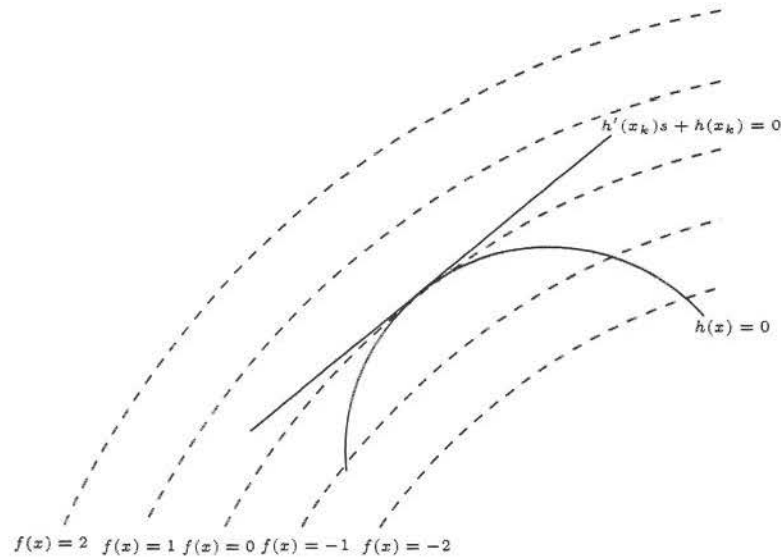


Figura 4.1: Ilustração para justificar a escolha da Hessiana no modelo quadrático

Observamos que “caminhando-se” por $h(x) = 0$ a partir de x_k , a função objetivo decresce, enquanto que “caminhando-se” por $h'(x_k)s + h(x_k) = 0$, a função objetivo cresce. Este incômodo pode ser contornado se incorporarmos em B_k informações sobre a segunda derivada de h . Estamos nos referindo a uma escolha para matriz B_k que geralmente está relacionada com o método de Newton. Para desenvolver esta idéia analiticamente,

consideremos o problema (1.1) sem a caixa, ou seja,

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && h(x) = 0. \end{aligned} \quad (4.34)$$

Se x_* é um ponto regular minimizador local do problema (4.34), então pelo Teorema 1 existe $\lambda_* \in \mathbb{R}^m$ tal que

$$\begin{cases} \nabla f(x_*) + h'(x_*)^T \lambda_* = 0 \\ h(x_*) = 0. \end{cases} \quad (4.35)$$

Olhando (4.35) como um sistema não-linear $F(x, \lambda) = 0$, o seu Jacobiano é dado por

$$F'(x, \lambda) = \begin{pmatrix} \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) & h'(x)^T \\ h'(x) & 0 \end{pmatrix}$$

e, portanto, aplicando o método de Newton a $F(x, \lambda) = 0$ temos

$$\left[\nabla^2 f(x_k) + \sum_{i=1}^m [\lambda_k]_i \nabla^2 h_i(x_k) \right] (x - x_k) + h'(x_k)^T (\lambda - \lambda_k) = -(\nabla f(x_k) + h'(x_k)^T \lambda_k)$$

e

$$h'(x_k)(x - x_k) = -h(x_k),$$

ou seja,

$$\begin{cases} [\nabla^2 f(x_k) + \sum_{i=1}^m [\lambda_k]_i \nabla^2 h_i(x_k)] s + \nabla \ell(x_k, \lambda) = 0 \\ h'(x_k)s + h(x_k) = 0, \end{cases} \quad (4.36)$$

onde $s = x - x_k$. Métodos baseados na iteração de Newton (4.36) para o problema (4.34) visando estabilidade e boas propriedades de convergência local foram propostos em [38] e [53].

Agora, as condições de otimalidade do problema quadrático (1.8), sem a caixa, são

$$\begin{cases} B_k s + \nabla \ell(x_k, \lambda) = 0 \\ h'(x_k)s + h(x_k) = 0, \end{cases} \quad (4.37)$$

onde $\lambda \in \mathbb{R}^n$. Se compararmos (4.36) com (4.37), o método de Newton nos sugere que

$$B_k = \nabla^2 f(x_k) + \sum_{i=1}^m [\lambda_k]_i \nabla^2 h_i(x_k), \quad (4.38)$$

onde λ_k é uma estimativa para os multiplicadores de Lagrange. Observamos que a escolha (4.38) para a matriz B_k engloba informações sobre a derivada segunda da restrição $h(x) = 0$, evitando, assim, o incômodo apresentado no exemplo gráfico anterior. Portanto, uma escolha possível para a matriz B_k seria a Hessiana do Lagrangeano, para a qual as propriedades de convergência local do método de Newton definido pelo subproblema (1.8) seriam as mesmas que as do método de Newton aplicado ao sistema não-linear dado em (4.35). Para outras aproximações secantes para B_k , a convergência local seria a mesma que a de um método quase-Newton. Boggs, Tolle e Wang [8] deram uma condição análoga à condição Dennis-Moré para a convergência superlinear de métodos quase-Newton aplicados a (4.35). Uma consequência dessa condição é que, supondo que a matriz Jacobiana do sistema (4.35) é não-singular, se as matrizes B_k convergem para a Hessiana do Lagrangeano na solução, então a convergência do par (x_k, λ_k) para (x_*, λ_*) é superlinear. Condições que garantem a convergência superlinear de x_k , λ_k e (x_k, λ_k) para métodos PQS podem ser encontradas em [66].

Embora o Algoritmo PQS utilize regiões de confiança para assegurar solubilidade dos subproblemas e garantir a convergência global, e em consequência seja capaz de lidar com Hessianas indefinidas, para empregarmos o resultado do Teorema 8 na solução dos subproblemas, é preciso que o modelo quadrático seja convexo. Neste sentido, cabe observar que a hipótese de modelo quadrático convexo, especialmente no caso de se trabalhar com uma aproximação para a matriz Hessiana do Lagrangeano, pode não ser realista. Combinando-se as condições de otimalidade de segunda ordem com a continuidade das funções envolvidas, o máximo que podemos esperar é que a matriz Hessiana do Lagrangeano seja semi-definida positiva no espaço tangente às restrições ativas para pontos nas vizinhanças de um minimizador local. Este fato nos motiva a considerar duas possíveis escolhas para as Hessianas e analisar as especificidades de cada uma: (a) aproximação quase-Newton para a Hessiana do Lagrangeano aumentado e (b) Hessiana reduzida.

(a) Aproximação quase-Newton para a Hessiana do Lagrangeano aumentado

Apesar do modelo (1.7) ser uma aproximação quadrática para o Lagrangeano e o desenvolvimento anterior ter sugerido que a aproximação $B_k \approx \nabla_{xx}^2 \ell(x_k, \lambda_k)$ pudesse ser uma alternativa, já que contempla informações sobre a curvatura das restrições, é possível que a Hessiana do Lagrangeano possua autovalores negativos, fazendo com que o modelo (1.7) não seja convexo. Uma alternativa que contorna esta dificuldade, ao mesmo tempo que respeita e tira proveito das condições de otimalidade de segunda ordem, é considerar a matriz Hessiana da função de mérito (1.6), utilizada pelo Algoritmo PQS:

$$\nabla_{xx}^2 \Phi(x, \lambda, \theta) = \theta \nabla_{xx}^2 \ell(x, \lambda) + (1 - \theta) \left(\sum_{i=1}^m h_i(x) \nabla_{xx}^2 h_i(x) + h'(x)^T h'(x) \right). \quad (4.39)$$

No desenvolvimento do capítulo 1, mostramos que a função de mérito (1.6) pode ser vista como o Lagrangeano aumentado tradicional, bastando definir o parâmetro de penalização usual de Lagrangeano aumentado como $\rho = \frac{1-\theta}{\theta}$. Neste sentido, a teoria clássica do Lagrangeano aumentado nos diz que em um minimizador (x_*, λ_*) de (1.1) satisfazendo as condições suficientes de segunda ordem, a matriz (4.39), que fica

$$\nabla_{xx}^2 \Phi(x_*, \lambda_*, \theta) = \theta \nabla_{xx}^2 \ell(x_*, \lambda_*) + (1 - \theta) h'(x_*)^T h'(x_*), \quad (4.40)$$

é definida positiva para todo $\theta \in (0, \theta_*)$ (ver, por exemplo, [56, cap.17] ou [50, cap.13]). Cabe observar que o último termo $h'(x)^T h'(x)$ de (4.39) acrescenta curvatura positiva ao Lagrangeano no espaço gerado pelas colunas de $h'(x)^T$, sem alterar a curvatura no espaço nulo de $h'(x)$. Podemos, portanto, trabalhar com $B_k \equiv \nabla_{xx}^2 \Phi(x_k, \lambda_k, \theta_k)$ ou alguma aproximação quase-Newton para esta matriz.

Uma abordagem quase-newtoniana bastante empregada é atualizar a matriz B_k com algo análogo à popular fórmula secante BFGS (Broyden, Fletcher, Goldfarb e Shanno) de minimização sem restrições dada por

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \quad (4.41)$$

onde $s_k = x_{k+1} - x_k$ e $y_k = \nabla_x \Phi(x_{k+1}, \lambda_{k+1}) - \nabla_x \Phi(x_k, \lambda_{k+1})$, conforme sugerido no capítulo 18 de [56]. Se B_k é definida positiva, como no caso de minimização sem restrições, a condição $s_k^T y_k > 0$ garante que B_{k+1} é definida positiva. No entanto, pode ser que s_k e y_k não satisfaçam essa desigualdade. Powell [60] propõe que y_k seja substituído por

$$\bar{y}_k = \theta y_k + (1 - \theta) B_k s_k,$$

onde

$$\theta = \begin{cases} 1, & s_k^T y_k \geq 0.2 s_k^T B_k s_k \\ \frac{0.8 s_k^T B_k s_k}{s_k^T B_k s_k - s_k^T y_k}, & s_k^T y_k < 0.2 s_k^T B_k s_k. \end{cases}$$

No entanto, o mesmo autor [61] observa que a substituição de y_k por $\bar{y}_k$ pode ser instável. Boggs e Tolle [4], por sua vez, propõem que $B_{k+1} = B_k$ quando $s_k^T y_k < 0$.

Nesta estratégia, no entanto, podem surgir algumas dificuldades. O limiar θ_* para

o qual (4.40) é definida positiva depende de limitantes para as segundas derivadas das funções envolvidas, em geral desconhecidos, de tal forma que pode ser difícil escolher valores apropriados para θ . Em geral, escolhas muito pequenas fazem com que o último termo de (4.39) domine, deteriorando o desempenho do algoritmo. Por outro lado, para escolhas muito grandes, a Hessiana (4.39) pode não ser definida positiva e a condição de curvatura $y_k^T s_k$ não seria satisfeita, sendo necessário ativar alguma salvaguarda, como as já mencionadas.

(b) Hessiana reduzida

Contrariamente à abordagem anterior, que lida com matrizes completas $B_k \in \mathbb{R}^{n \times n}$, consideraremos agora aproximações para a Hessiana reduzida do Lagrangeano,

$$Z_k^T \nabla_{xx}^2 \ell(x_k, \lambda_k) Z_k, \quad (4.42)$$

onde a matriz Z_k é tal que suas colunas geram o espaço nulo de $h'(x_k)$. A Hessiana (4.42) tem dimensão inferior à de (4.39) e sob hipóteses clássicas é definida positiva numa vizinhança da solução. Para utilizarmos a matriz (4.42) no subproblema (4.33) podemos proceder da seguinte maneira: supondo que $h'(x_k)$ tem posto completo m , uma vez determinada a matriz $Z_k \in \mathbb{R}^{n \times (n-m)}$, a mudança de variáveis $s = Z_k v$, com $v \in \mathbb{R}^{(n-m)}$, absorve a restrição de igualdade e (4.33) fica

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} v^T Z_k^T \nabla_{xx}^2 \ell(x_k, \lambda_k) Z_k v + v^T Z_k^T \nabla_x \ell(x_k, \lambda_k) + \ell(x_k, \lambda_k) \\ & \text{sujeito a} \quad \begin{aligned} & \|s_{nor} + Z_k v\|_\infty \leq \Delta \\ & l \leq x_k + s_{nor} + Z_k v \leq u. \end{aligned} \end{aligned} \quad (4.43)$$

O conjunto factível de (4.43) pode ser reescrito na forma

$$a \leq Z_k v \leq b, \quad (4.44)$$

onde as componentes de a e b são respectivamente $a_i = \max\{-\Delta - (s_{nor})_i, l_i - (x_k)_i - (s_{nor})_i\}$ e $b_i = \min\{\Delta - (s_{nor})_i, u_i - (x_k)_i - (s_{nor})_i\}$, para $i = 1, \dots, n$, ou seja, temos um problema com restrições lineares de desigualdade. Completando-se com folgas, (4.44) fica

$$\begin{cases} Z_k v \leq b \\ Z_k v \geq a \end{cases} \Rightarrow \begin{cases} Z_k v + p = b \\ 0 \leq p \leq b - a, \end{cases}$$

onde $p \in \mathbb{R}^n$, e então, reescrevemos (4.43) como

$$\begin{aligned}
& \text{minimizar} \quad \frac{1}{2} v^T Z_k^T \nabla_{xx}^2 \ell(x_k, \lambda_k) Z_k v + v^T Z_k^T \nabla_x \ell(x_k, \lambda_k) + \ell(x_k, \lambda_k) \\
& \text{sujeito a} \quad \begin{pmatrix} Z_k & I \end{pmatrix} \begin{pmatrix} v \\ p \end{pmatrix} = b \\
& \quad \quad \quad 0 \leq p \leq b - a.
\end{aligned} \tag{4.45}$$

O problema (4.45) possui $2n - m$ variáveis, n restrições lineares de igualdade e n canalizações. Além disso, (4.45) é um problema convexo na vizinhança de um minimizador e seu conjunto factível é um politopo não-vazio, pela maneira com que foi construído, já que $Z_k v = 0$ é um ponto viável do problema (4.43). Pela equivalência entre (4.43) e (4.45), este politopo também é limitado. Portanto, o resultado do Teorema 8 pode ser aplicado a (4.45), uma formulação equivalente a (4.33).

Assim como fizemos na abordagem anterior, vamos apresentar as expressões que permitem aproximar a Hessiana (4.42) no contexto quase-Newton. Para tanto, consideremos o sistema (4.37) que estabelece as condições de otimalidade de primeira ordem (Teorema 1) para o problema

$$\begin{aligned}
& \text{minimizar} \quad Q_k(s) \\
& \text{sujeito a} \quad h'(x_k)s = -h(x_k),
\end{aligned}$$

reescrito sob a forma

$$\begin{cases} B_k s + \lambda^T h'(x_k) &= -\nabla f(x_k) \\ h'(x_k)s &= -h(x_k) \end{cases} \quad \text{ou} \quad \begin{pmatrix} B_k & h'(x_k)^T \\ h'(x_k) & 0 \end{pmatrix} \begin{pmatrix} s \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) \\ -h(x_k) \end{pmatrix}. \tag{4.46}$$

Com as hipóteses de que $h'(x_k)$ tem posto completo m e B_k é tal que $y^T B_k y > 0$ para todo $y \neq 0$ tal que $h'(x_k)y = 0$, então o sistema (4.46) tem solução única, que no contexto de programação quadrática sequencial pode ser tomada como $s = s_k$ e $\lambda = \lambda_{k+1}$.

Suponhamos agora conhecidas as matrizes Y_k e Z_k , cujas colunas geram o espaço coluna de $h'(x_k)^T$ e o espaço nulo de $h'(x_k)$, respectivamente (ver, por exemplo, [56, cap.15]). Tomar $Y_k \equiv h'(x_k)^T$ é uma escolha possível sempre que $h'(x_k)$ tenha posto completo m . Escrevendo

$$s_k = Y_k s_y + Z_k s_z, \tag{4.47}$$

podemos substituir (4.47) em (4.46) e obter o sistema desacoplado, a ser resolvido para s_y e s_z :

$$\begin{cases} (h'(x_k)Y_k) s_y = -h(x_k) \\ (Z_k^T B_k Z_k) s_z = -Z_k^T B_k Y_k s_y - Z_k^T \nabla f(x_k). \end{cases} \quad (4.48)$$

Os multiplicadores de Lagrange λ_{k+1} podem então ser obtidos resolvendo-se:

$$(h'(x_k)Y_k)^T \lambda_{k+1} = Y_k^T (\nabla f(x_k) + B_k s_k). \quad (4.49)$$

A única hipótese a ser feita sobre B_k é que a Hessiana reduzida $Z_k^T B_k Z_k$ seja definida positiva. Chamando de M_k a aproximação quase-Newton para a matriz (4.42), e supondo que o passo $\alpha_k s_k$ foi tomado de (x_k, λ_k) para (x_{k+1}, λ_{k+1}) , pelo Teorema de Taylor temos

$$\nabla_{xx}^2 \ell(x_{k+1}, \lambda_{k+1}) \alpha_k s_k \approx \nabla_x \ell(x_k + \alpha_k s_k, \lambda_{k+1}) - \nabla_x \ell(x_k, \lambda_{k+1}), \quad (4.50)$$

onde $s_k = x_{k+1} - x_k = Z_k s_z + Y_k s_y$. Pré-multiplicando (4.50) por Z_k^T e usando (4.47), vem

$$\begin{aligned} Z_k^T \nabla_{xx}^2 \ell(x_{k+1}, \lambda_{k+1}) Z_k \alpha_k s_z &\approx -Z_k^T \nabla_{xx}^2 \ell(x_{k+1}, \lambda_{k+1}) Y_k \alpha_k s_y + \\ &+ Z_k^T [\nabla_x \ell(x_k + \alpha_k s_k, \lambda_{k+1}) - \nabla_x \ell(x_k, \lambda_{k+1})]. \end{aligned} \quad (4.51)$$

A equação secante para M_k é obtida eliminando-se o termo $Z_k^T \nabla_{xx}^2 \ell(x_{k+1}, \lambda_{k+1}) Y_k \alpha_k s_y$ em (4.51). Esta simplificação é inspirada na eliminação análoga do termo em s_k no lado direito de (4.49), motivada pelo desacoplamento na computação de s_k e λ_{k+1} , e justificada pelo fato de que s_k vai para zero à medida que a sequência se aproxima da solução, mas $\nabla f(x_k)$ em geral não converge para zero. Completando o raciocínio, de (4.49) vemos que os novos multiplicadores serão boas estimativas para os multiplicadores do problema quadrático na vizinhança da solução.

Assim, a equação secante é

$$M_{k+1} \tilde{s}_k = \tilde{y}_k,$$

onde

$$\begin{cases} \tilde{s}_k &= \alpha_k s_z \\ \tilde{y}_k &= Z_k^T [\nabla_x \ell(x_{k+1}, \lambda_{k+1}) - \nabla_x \ell(x_k, \lambda_{k+1})]. \end{cases} \quad (4.52)$$

A mesma fórmula de correção (4.41) da abordagem anterior se aplica para obter M_{k+1} a partir de M_k e (4.52), bem como as salvaguardas mencionadas para a manutenção de matrizes definidas positivas.

Cabe observar que os passos tangencial e normal do desenvolvimento acima, na nomenclatura do Algoritmo PQS, ficam $Z_k s_z = s_{tan}$ e $Y_k s_y = s_{nor}$, de tal forma que $s_k = x_{k+1} - x_k = s_c$. Em termos práticos, se Z_k foi obtida pela fatoração QR de $h'(x_k)^T$ (ver algoritmo em [14] e [36] para implementação que preserva continuidade), então $s_z = Z_k^T s_{tan}$, ou seja, em (4.52) temos $\tilde{s}_k = \alpha_k Z_k^T s_{tan}$. Uma outra alternativa seria buscar uma “base” de $h'(x_k)^T$, de modo que se

$$h'(x_k) = [B \mid N]$$

temos $Y = \begin{bmatrix} B^{-1} \\ 0 \end{bmatrix}$ e $Z = \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix}$ (ver [9, 21]).

Comentários finais

O objetivo principal deste trabalho é a resolução do problema geral de programação não-linear na forma padrão. Para tanto optamos pelo Algoritmo PQS proposto em [37]. Estudamos e reproduzimos aqui as provas de boa definição e convergência global deste algoritmo, que utiliza como função de mérito o Lagrangeano aumentado (diferenciável, simples e estável) e uma estratégia não-monótona para atualizar o parâmetro de penalização. Cabe destacar que a abordagem PQS considerada utiliza regiões de confiança para assegurar solubilidade dos subproblemas, o que por sua vez aumenta a estabilidade destes quando se precisa reduzir o raio. Além disso, os subproblemas não precisam ser resolvidos exatamente, o que é relevante para problemas de grande porte.

Embora tenhamos interagido bastante com o Algoritmo PQS de [37] e suas propriedades teóricas, o elemento original deste trabalho concentra-se no capítulo 4, onde reduzimos os subproblemas do algoritmo PQS ao contexto de otimização em caixas, quer pela própria característica do subproblema em questão (originalmente quadráticas canalizadas), quer pelo intermédio de um resultado de equivalência. Tal resultado permite que a minimização de uma função convexa num politopo não-vazio e limitado seja equivalentemente formulada como minimização numa caixa. Visando a manutenção da convexidade do modelo quadrático, propomos duas escolhas para a sua matriz Hessiana, uma aproximação quase-Newton para a Hessiana do Lagrangeano aumentado, e a outra baseada na formulação via Hessiana reduzida.

Por tratar-se de um trabalho essencialmente teórico, os aspectos práticos permanecem abertos e sugerem interessantes questões a serem investigadas. Por exemplo:

- Em que medida a hipótese de limitação do politopo no Teorema 8 é importante na prática? O que ocorre se a caixa do problema (1.1) é ilimitada?
- Como se comparam as escolhas propostas para as matrizes hessianas do modelo quadrático? Como se comporta o compromisso custo versus desempenho em cada escolha?

Com relação à primeira questão, na defesa desta dissertação o professor R. Andreani mencionou um resultado teórico mais geral, no contexto de problemas não-lineares com restrições de equilíbrio (Teorema 2, [2]), que assegura a validade do Teorema 8 sem precisarmos da limitação para o politopo $\mathcal{W}$. As únicas hipóteses necessárias são que $\mathcal{W}$ seja não-vazio e que a função g seja quadrática, que se enquadram perfeitamente ao nosso caso.

Como o Algoritmo PQS já foi testado computacionalmente em [37], para investi-

garmos a validade e eficiência de nossa abordagem para os subproblemas, gostaríamos de desenvolver uma implementação computacional utilizando o algoritmo Box-Quacan [27] para a minimização em caixas, já que todos os subproblemas podem ser colocados neste contexto. Nossa idéia é fazer os primeiros ajustes com problemas-teste de [43] e posteriormente comparar com o desempenho do algoritmo originalmente implementado em [37].

Referências bibliográficas

- [1] ANDREANI, R. & MARTÍNEZ, J. M. - On the solution of bounded and unbounded mixed complementarity problems, por aparecer em *Optimization*, (1999).
- [2] ANDREANI, R. & MARTÍNEZ, J. M. - On the solution of mathematical programming problems with equilibrium constraints using nonlinear programming algorithms, em preparação, (2000).
- [3] BIRGIN, J. M., MARTÍNEZ, J. M. & RAYDAN, M. - Nonmonotone spectral projected gradient methods on convex sets, TR-IMECC, Universidade Estadual de Campinas, Brazil, por aparecer em *SIAM Journal on Optimization*, (1999).
- [4] BOGGS, P. T., TOLLE, J. W. - A family of descent funtions for constrained optimization. *SIAM Journal on Numerical Analysis*, 21 (1984), pp.1146-1161.
- [5] BOGGS, P. T., TOLLE, J. W. - Sequential quadratic programming. *Acta Numerica*, 4 (1996), pp.1-51.
- [6] BOGGS, P. T., KEARSLEY, A. J. & TOLLE, J. W. - A Pratical Algorithm for General Large-Scale Nonlinear Optimization Problems. *SIAM Journal on Optimization*, 9 (1999), pp.755-778.
- [7] BOGGS, P. T., KEARSLEY, A. J. & TOLLE, J. W. - A Global Convergence Analysis of an Algorithm for Large-Scale Nonlinear Optimization Problems. *SIAM Journal on Optimization*, 9 (1999), pp.833-862.
- [8] BOGGS, P. T., TOLLE, J. W. & WANG, P. - On the local convergence of quasi-Newton methods for constrained optimization. *SIAM Journal on Control and Optimization*, 20 (1982), pp.161-171.
- [9] BYRD, R. H., HRIBAR, M. E. & NOCEDAL, J. - An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM Journal on Optimization*, 9 (1999), pp.877-900.
- [10] BYRD, R. H. & NOCEDAL, J. - An analysis of reduced Hessian methods for constrained optimization. *Mathematical Programming*, 49 (1991), pp.285-323.
- [11] CHAMBERLAIN, R., LEMARÉCHAL, C., PEDERSEN, H. C. & POWELL, M. J. D. - The watchdog technique for forcing convergence in algorithms for constrained optimization. *Mathematical Programming*, 16 (1982), pp.1-17.

- [12] CIARLET, P. G. - *Introduction to numerical linear algebra and optimisation*. Cambridge University Press, 1989.
- [13] COLEMAN, T. F. & CONN, A. R. - Non-linear programming via an exact penalty-function: Asymptotic analysis. *Mathematical Programming*, 24 (1982), pp.123-136.
- [14] COLEMAN, T. F. & SORENSEN, D. C. - A note on the computation of an orthonormal basis for the null space of a matrix. *Mathematical Programming*, 29 (1984), pp.234-242.
- [15] CONN, A. R., GOULD, N. & TOINT, P. - Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis*, 25 (1988), pp.433-460. Ver também *SIAM Journal on Numerical Analysis*, 26 (1989), pp.764-767.
- [16] CONN, A. R., GOULD, N. & TOINT, P. - *LANCELOT: a Fortran package for large-scale nonlinear optimization (release A)*. Springer Verlag, Berlin and New York, (1988).
- [17] DENNIS, J. E. & SCHNABEL, R. B. - *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. New Jersey: Prentice-Hall, Inc., 1983.
- [18] DENNIS, J. E., HEINKENSCHLOSS, M. & VICENTE, L. N. - Trust-Region Interior-Point SQP Algorithms for a class of Nonlinear Programming Problems. *SIAM Journal on Control and Optimization*, 36 (1998), pp.1750-1794.
- [19] DONTCHEV, A. L., HAGER, W. W. & VELIOV, V. M. - Uniform Convergence and Mesh Independence of Newton's Method for Discretized Variational Problems. *SIAM Journal on Control and Optimization*, 39 (2000), pp.961-980.
- [20] FLETCHER, R. - Second order corrections for non-differentiable optimization. Em *Numerical Analysis*, D. Griffiths, ed., Springer Verlag, 1982, pp.85-114. Proceedings Dundee 1981.
- [21] FLETCHER, R. - *Practical Methods of Optimization*, 2^a ed., John Wiley & Sons, Chichester, 1987.
- [22] FLETCHER, R. & MAZA, E. S. de la - Nonlinear programming and nonsmooth optimization by successive linear programming. *Mathematical Programming*, (1989), pp.235-256.
- [23] FRIEDLANDER, A. - *Elementos de Programação Não-Linear*. Campinas, SP: Editora da Unicamp, 1994.
- [24] FRIEDLANDER, A., MARTÍNEZ, J. M. & SANTOS, S. A. - On the resolution of linearly constrained convex minimization problems. *SIAM Journal on Optimization*, 4 (1994), pp.331-339.

- [25] FRIEDLANDER, A., MARTÍNEZ, J. M. & SANTOS, S. A. - Solution of linear complementarity problems using minimization with simple bounds. *Journal of Global Optimization*, 6 (1995), pp.253-267.
- [26] FRIEDLANDER, A., MARTÍNEZ, J. M. & SANTOS, S. A. - A new strategy for solving variational inequalities in bounded polytopes. *Numerical functional analysis and optimization*, 16 (1995), pp.653-668.
- [27] FRIEDLANDER, A., MARTÍNEZ, J. M. & SANTOS, S. A. - A New Trust Region Algorithm for Bound constrained minimization. *Applied Mathematics & Optimization*, 30 (1994), pp.235-266.
- [28] FUKUSHIMA, M. & PANG, J. - Some Feasibility Issues in Mathematical Programs with Equilibrium Constraints. *SIAM Journal on Optimization*, 8 (1998), pp.673-681.
- [29] GABAY, D. - Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization. *Mathematical Programming Studies*, 16 (1982), pp.18-44.
- [30] GARCIA-PALOMARES, U. M. & MANGASARIAN, O. L. - Superlinearly convergent quasi-Newton methods for nonlinearly constrained optimization problems. *Mathematical Programming*, 11 (1976), pp.1-13.
- [31] GILL, P. E. - An SQP Method for the Optimization of Dynamical Systems, *Presentation at 16th International Symposium on Mathematical Programming*, Lausanne, Switzerland, Abstracts, p.106, August, 1997.
- [32] GILL, P. E. - SQP Method for Large-Scale Optimization, *Contributed talk at 19th IFIP TC7 Conference on System Modelling and Optimization*, Cambridge, England, Book of Abstracts, p.22, July, 1999.
- [33] GILL, P. E., MURRAY, W. & SAUNDERS, M. A. - User's guide for SNOPT (Version 5.3): A FORTRAN package for large-scale nonlinear programming. *Technical Report NA 97-4*, Department of Mathematics, University of California, San Diego, 1997.
- [34] GILL, P. E., MURRAY, W., SAUNDERS, M. A., STEWART, G. W. & WRIGHT, M. H. - Properties of a representation of a basis for the null space. *Mathematical Programming*, 33 (1985), pp.172-186.
- [35] GILL, P. E., MURRAY, W., SAUNDERS, M. A. & WRIGHT, M. H. - User's guide for NPSOL (Version 4.0): A FORTRAN package for nonlinear programming. *Technical Report SOL 86-2*, Department of Operations Research, Stanford University, Stanford, CA, 1986.
- [36] GOLUB, G. H. & VAN LOAN, C. F. - *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.

- [37] GOMES, F. A. M., MACIEL, M. C. & MARTÍNEZ, J. M. - Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters. *Mathematical Programming*, 84 (1) (1999), pp.161-200.
- [38] GOULD, N. I. M. - On the convergence of a sequential penalty function method for constrained minimization. *SIAM Journal on Numerical Analysis*, 26 (1989), pp.107-128.
- [39] GOULD, N. & TOINT, P. - SQP Methods for Large-Scale Nonlinear Programming, *Invited talk at 19th IFIP TC7 Conference on System Modelling and Optimization*, Cambridge, England, Book of Abstracts, p.4, July, 1999.
- [40] HAN, S. P. - Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Mathematical Programming*, 11 (1976), pp.263-282.
- [41] HAN, S. P. - A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications*, 22 (1977), pp.297-309.
- [42] HEIKENSCHLOSS, M. - SQP-Based Optimization Methods for Flow Control, *Contributed presentation at Optimization 98*, Coimbra, Portugal, Abstracts, p.42, July, 1998.
- [43] HOCK, W. & SCHITTKOWSKI, K. - *Test examples for nonlinear programming codes*. Lecture Notes in Economics and Mathematical Systems 187, Springer Verlag, Berlin, 1981.
- [44] KLEIS, D. & SACHS, E. W. - Optimal Control of the Sterilization of Prepackaged Food. *SIAM Journal on Optimization*, 10 (2000), pp.1180-1195.
- [45] KNUTH, D. E. - *The art of computer programming*. 2^a ed., Addison-Wesley Publishing Company, 1981.
- [46] LAWRENCE, C. T. & TITS, A. - A Computationally Efficient Feasible SQP Algorithm, *Contributed talk at 19th IFIP TC7 Conference on System Modelling and Optimization*, Cambridge, England, Book of Abstracts, p.55, July, 1999.
- [47] LEIBFRTZ, F. & SACHS, E. W. - Inexact SQP Interior Point Methods and Large Scale Optimal Control Problems. *SIAM Journal on Control and Optimization*, 38 (1999), pp.272-293.
- [48] LEYFFER, S. - Nonlinear Programming without a penalty function, *Presentation at 16th International Symposium on Mathematical Programming*, Lausanne, Switzerland, Abstracts, p.165, August, 1997.
- [49] LIU, X. & YUAN, Y. - A Robust Algorithm for Optimization with General Equality and Inequality Constraints. *SIAM Journal on Scientific Computing*, 22 (2000), pp.517-534.
- [50] LUENBERGER, D. G. - *Linear and nonlinear programming*. 2^a ed., New York: Addison-Wesley Publishing Company, 1986.

- [51] LUO, Z-Q., FUKUSHIMA, M. & PANG, J-S. - A Globally Convergent Sequential Quadratic Programming Algorithm for Mathematical Programming with Linear Complementarity Constraints, *Presentation at 16th International Symposium on Mathematical Programming*, Lausanne, Switzerland, Abstracts, p.170, August, 1997.
- [52] MARTÍNEZ, J. M. & SANTOS, S. A. - *Métodos Computacionais de Otimização*. 20^o Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, 1995.
- [53] MARTÍNEZ, J. M. & SANTOS, L. T. - Some new theoretical results on recursive quadratic programming algorithms. *Journal of Optimization Theory and Applications* 97, 1998, pp.435-454.
- [54] MAYNE, D. Q. & POLAK, E. - A Superlinearly convergent algorithm for constrained optimization problems. *Mathematical Programming Studies*, 16 (1982), pp.45-61.
- [55] MITCHELL, J. E. - An Interior Point SQP Method for Solving Relaxations of Mixed Integer Nonlinear Programming Problems, *Presentation at 16th International Symposium on Mathematical Programming*, Lausanne, Switzerland, Abstracts, p.185, August, 1997.
- [56] NOCEDAL, J. & WRIGHT, S. J. - *Numerical Optimization*. Springer-Verlag New York, Inc., 1999.
- [57] PANIER, E. R. & TITS, A. L. - On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Mathematical Programming*, 59 (1993), pp.261-276.
- [58] POWELL, M. J. D. - A fast algorithm for nonlinearly constrained optimization calculations. Em *Numerical Analysis*, Dundee 1977, G. A. Watson, ed., Springer Verlag, Berlin, 1977, pp.144-157.
- [59] POWELL, M. J. D. - Algorithms for nonlinear constraints that use Lagrangian functions. *Mathematical Programming*, 14 (1978), pp.224-248.
- [60] POWELL, M. J. D. - The convergence of variable metric methods for nonlinearly constrained optimization calculations, in *Nonlinear Programming 3* (Mangasarian, O. L., Meyer, R. R. & Robinson, S. M., eds), Academic Press, London and New York, 1978, pp.27-63.
- [61] POWELL, M. J. D. - The performance of two subroutines for constrained optimization on some difficult test problems, in *Numerical Optimization 1984* (Boggs, P. T., Byrd, R. H. & Schnabel, R. B., eds), SIAM, Philadelphia, 1985, pp.160-177.
- [62] QI, L. & WEI, Z. - On the Constant Positive Linear Dependence Condition and Its Application to SQP Methods. *SIAM Journal on Optimization*, 10 (2000), pp.963-981.

- [63] SACHS, E. W. - Optimization Methods in Food Processing, *Presentation at 16th International Symposium on Mathematical Programming*, Lausanne, Switzerland, Abstracts, p.237, August, 1997.
- [64] SARGENT, R. W. H. & DING, M. - A New SQP Algorithm for Large-Scale Non-linear Programming. *SIAM Journal on Optimization*, 11 (2000), pp.716-747.
- [65] TAPIA, R. A. - Quasi-Newton methods for equality constrained optimization: Equivalence of existing methods and a new implementation. In *Nonlinear Programming 3* (O. Mangasarian, R. Meyer & S. Robinson, eds), Academic Press, New York, NY, 1978, pp.125-164.
- [66] TAPIA, R. & CHIANG, J. - Convergence Rates for the Variable, the Multiplier, and the Pair in SQP Methods. *Technical Report 89-2*, Rice University, March 1989.
- [67] TOMLIN, J. - Implementation of a Filter Algorithm for Regularized SQP Solution of Nonlinear Programming, *Mini-Symposium at sixth SIAM Conference on Optimization*, Atlanta, GA, Final Program and Abstracts, p.115, May, 1999.
- [68] TRÖLTZSCH, F. - On the Lagrange-Newton-SQP Method for the Optimal Control of Semilinear Parabolic Equations. *SIAM Journal on Control and Optimization*, 38 (1999), pp.294-312.
- [69] WILSON, R. B. - *A simplicial Algorithm for Concave Programming*, Ph.D. thesis, Graduate School of Business Administration, Harvard University, 1963.
- [70] YUAN, Y. & CHEN, X. - Optimality Conditions for the CDT Subproblem, in *Numerical Linear Algebra and Optimization*, Yuan, Y. ed, Science Press, Beijing, New York, 1999, pp.111-121.