
Universidade Estadual de Campinas
Instituto de Matemática Estatística e Computação Científica
Departamento de Matemática Aplicada

Programação quadrática sequencial e condições de qualificação

Fernanda Téles Nunes*

Mestrado em Matemática Aplicada - Campinas - SP

Orientadora: Profa. Dra. Maria Aparecida Diniz Ehrhardt

* Este trabalho teve apoio financeiro do CNPq.

**PROGRAMAÇÃO QUADRÁTICA SEQUENCIAL
E CONDIÇÕES DE QUALIFICAÇÃO**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Fernanda Téles Nunes e aprovada pela comissão julgadora.

Campinas, 9 de março de 2009.



Profª. Dra. Maria Aparecida Diniz Ehrhardt
Orientadora

Banca Examinadora:

Profª. Dra. Maria Aparecida Diniz Ehrhardt (IMECC/UNICAMP)

Profª. Dra. Maria Laura Schuverdt (Universidad Nacional de La Plata/ Argentina)

Prof. Dr. Lúcio Tunes dos Santos (IMECC/UNICAMP)

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

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Júlia Milani Rodrigues – CRB8a 2116**

Nunes, Fernanda Téles

N922p Programação quadrática sequencial e condições de qualificação /
Fernanda Téles Nunes -- Campinas, [S.P. :s.n.], 2009.

Orientador : Maria Aparecida Diniz Ehrhardt

Dissertação (mestrado) - Universidade Estadual de Campinas,
Instituto de Matemática, Estatística e Computação Científica.

1. Programação quadrática sequencial. 2. Condições de
qualificação. 3. Busca não-monótona. I. Ehrhardt, Maria Aparecida
Diniz. II. Universidade Estadual de Campinas. Instituto de Matemática,
Estatística e Computação Científica. III. Título.

Título em inglês: Sequential quadratic programming and constraint qualification.

Palavras-chave em inglês (Keywords): 1. Sequential quadratic programming. 2. Constraint qualification. 3. Nonmonotone line search.

Área de concentração: Otimização não-linear

Título: Mestre em Matemática Aplicada

Banca examinadora:

Profa. Dra. Maria Aparecida Diniz Ehrhardt (IMECC-UNICAMP)

Profa. Dra. Maria Laura Schuverdt (Universidad de la Plata)

Prof. Dr. Lúcio Tunes dos Santos (IMECC-UNICAMP)

Data da defesa: 09/03/2009

Programa de pós-graduação: Mestrado em Matemática Aplicada

Dissertação de Mestrado defendida em 09 de março de 2009 e aprovada

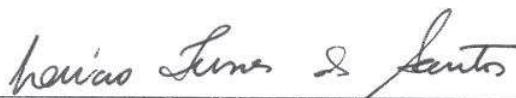
Pela Banca Examinadora composta pelos Profs. Drs.



Prof.(a). Dr(a). MARIA APARECIDA DINIZ EHRHARDT



Prof. (a). Dr (a). MARIA LAURA SCHUVERDT



Prof. (a). Dr (a). LÚCIO TUNES DOS SANTOS

*“ Que a inspiração chegue não depende de mim.
A única coisa que posso fazer é garantir que ela
me encontre trabalhando.”*

(Pablo Picasso)

Agradecimentos

Agradeço:

A Deus por iluminar minhas escolhas, pela força quando as dificuldades apareceram e pelas pessoas abençoadas que sempre colocou em meu caminho.

Aos meus pais, Nilce e Marcos, ao meu irmão, Marcos Paulo, e a meus avós, Margarida, Teodoro, Malvina e Lauro, que sempre me apoiaram incondicionalmente e, mesmo sem entender, torceram pelo meu sucesso neste trabalho.

À minha orientadora e grande amiga Cheti, que tão cuidadosa e carinhosamente me guiou durante todo o mestrado; para mim um exemplo de docente e ser humano.

Ao Prof. Lúcio Santos, pelo cuidado e dedicação na leitura deste trabalho e pelas valiosas sugestões.

A todos os professores que estiveram presentes em minha vida acadêmica, tanto àqueles com os quais convivi nas salas de aula, quanto àqueles que se tornaram professores (e alguns até grandes amigos) nos corredores e no saguão do IMECC. Agradeço, em especial, ao meu professor do colégio, Guilherme, que primeiro me mostrou quão encantadora a Matemática pode ser.

Ao meu namorado, Denis, por estar ao meu lado mesmo quando minhas horas eram gastas com os livros ou em frente ao computador.

A todos os amigos, por estarem sempre abertos às minhas perguntas e às discussões, por tornarem mais leves os momentos de estudo e compartilharem comigo tantos bons momentos durante o café. Também aos amigos de longe, que nunca deixaram de estar presentes.

Ao CNPq, pelo apoio financeiro.

Resumo

Abordando problemas de minimização de funções com restrições nos deparamos com as condições de otimalidade e, ainda, com condições de qualificação das restrições. Nosso interesse é o estudo detalhado de várias condições de qualificação, com destaque para a condição de dependência linear positiva constante, e sua influência na convergência de algoritmos de Programação Quadrática Sequencial. A relevância deste estudo está no fato de que resultados de convergência que têm, em suas hipóteses, condições de qualificação fracas são mais fortes que aqueles baseados em condições de qualificação fortes. Experimentos numéricos serão realizados tanto para investigar a eficiência destes métodos na resolução de problemas com diferentes condições de qualificação, quanto para comparar dois diferentes tipos de busca, monótona e não-monótona. Tentamos confirmar a hipótese de que algoritmos baseados em uma busca não-monótona atuam contra o Efeito Maratos, de comum ocorrência na resolução de problemas de minimização através de métodos de Programação Quadrática Sequencial.

Palavras-chave: minimização com restrições, programação quadrática sequencial, condições de qualificação, busca não-monótona.

Abstract

In the context of constrained optimization problems, we face the optimality conditions and also constraint qualification. Our aim is to study with details several constraint qualification, highlighting the constant positive linear dependence condition, and its influence in Sequential Quadratic Programming algorithms convergence. The relevance of this study is in the fact that convergence results having as hypothesis weak constraints qualification are stronger than those based on stronger constraints qualification. Numerical experiments will be done with the purpose of investigating the efficiency of these methods to solve problems with different constraints qualification and to compare two different kinds of line search, monotone and nonmonotone. We want to confirm the hypothesis that algorithms based on a nonmonotone line search act against the Maratos Effect, very common while solving minimization problems through Sequential Quadratic Programming methods.

Keywords: constrained minimization, sequential quadratic programming, constraint qualification, nonmonotone line search.

Sumário

1	Introdução	1
2	Conceitos Fundamentais	5
3	Condições de Qualificação	8
3.1	Regularidade, Mangasarian-Fromovitz e Posto Constante	9
3.2	A condição de dependência linear positiva constante	11
4	Programação quadrática sequencial	20
4.1	O subproblema quadrático	21
4.2	A Hessiana do Lagrangiano	23
4.3	Função de mérito e convergência global	32
4.4	Efeito Maratos e busca linear não-monótona	41
5	Métodos de PQS e condições de qualificação	44
5.1	Sequências de pontos KKT aproximados e a CPLD	45
5.2	Um método de PQS factível e MFCQ	51
6	Experimentos Numéricos	60
6.1	Testes preliminares	62
6.2	O problema da montanha	68
7	Conclusões	76

Notações utilizadas

$\mathbf{u}^T \in \mathbb{R}^{1 \times n}$ denota o transposto do vetor $\mathbf{u} \in \mathbb{R}^{n \times 1}$.

$\mathbf{u}^T \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$ é o produto escalar canônico em $\mathbb{R}^n$.

$\| \cdot \|$ representa a norma Euclidiana de vetores e matrizes.

$I \in \mathbb{R}^{n \times n}$ é a matriz identidade de ordem n .

Dada $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\nabla f \in \mathbb{R}^{n \times 1}$ é o vetor gradiente da função f .

Dada $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\nabla h \in \mathbb{R}^{n \times m}$ é a matriz Jacobiana de h com seus gradientes por colunas.

$\nabla^2 f \in \mathbb{R}^{n \times n}$ é a matriz hessiana da função $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

$|C|$ denota a cardinalidade do conjunto C .

Dado $v \in \mathbb{R}^n$, definimos $\text{supp}(v) = \{i \in \{1, \dots, n\} \mid v_i \neq 0\}$.

Capítulo 1

Introdução

Dentre os métodos computacionais de otimização mais estudados, utilizados na resolução de problemas de minimização com restrições envolvendo funções não-lineares, encontram-se os métodos de Programação Quadrática Sequencial (PQS) e os de Lagrangiano Aumentado. Resultados de convergência desses últimos foram apresentados recentemente por Schuverdt [36] utilizando a condição de dependência linear positiva constante, a qual provou-se, no mesmo trabalho, ser uma condição de qualificação das restrições. Estes estudos de Schuverdt tiveram início a partir de um trabalho de Qi e Wei [34], no qual os autores apenas conjecturaram a hipótese da CPLD ser uma condição de qualificação e utilizaram-na para analisar a convergência de métodos de PQS. Também é por este trabalho de Qi e Wei que se norteia o nosso.

Métodos de PQS baseiam-se na estratégia de encontrar a solução para um problema complicado através da resolução de uma sequência de problemas mais simples, cujas soluções convirjam para a procurada. Dado um problema que consiste na minimização de uma função sujeita a um conjunto de restrições, construímos uma sequência de subproblemas nos quais a função objetivo é substituída por uma aproximação quadrática e as restrições são substituídas por aproximações lineares. Daí o nome de Programação Quadrática Sequencial. Dada uma aproximação x^k para a solução, construímos o subproblema quadrático a partir deste iterando, que pelo menos numa vizinhança de x^k aproxima bem o problema original, e tomamos x^{k+1} como um ponto que resolve este subproblema, gerando assim uma sequência de iterandos que deve convergir para uma solução. A Programação Quadrática Sequencial se popularizou ao final da década de 70 e desde então vem sendo usada com sucesso para resolver problemas de otimização não-linear res-

trita. A primeira referência a algoritmos de PQS parece ter sido na tese de Wilson [38] na qual ele propôs o método que ficou conhecido como Newton-PQS. Veremos no Capítulo 4 que, sob certas condições, o método de PQS é equivalente ao método de Newton aplicado às condições KKT do problema de minimização. Tal equivalência foi primeiramente mencionada por Tapia [37].

Algumas propriedades dos métodos de PQS são relevantes. Primeiro, trata-se de um método que caminha por pontos infactíveis, não sendo necessário que a aproximação inicial pertença ao conjunto de restrições. Segundo, subproblemas quadráticos são fáceis de resolver e suas soluções podem resumir-se às de um sistema linear, quando consideramos apenas as restrições ativas no iterando atual. Terceiro, os métodos de PQS possuem boa velocidade de convergência, podendo ser quadrática ou superlinear conforme a construção dos subproblemas quadráticos. Garcia-Palomares e Mangasarian [15] investigaram em 1976 um algoritmo no qual toda a Hessiana do Lagrangiano, matriz das segundas derivadas com respeito a x e aos multiplicadores, era atualizada a cada iteração. Pouco depois, Han [17] apresentou teoremas de convergência global para algoritmos de PQS que utilizavam as atualizações PSB e BFGS, as quais serão apresentadas também no Capítulo 4. Para uma introdução à solução de problemas quadráticos usando técnicas clássicas ver [14].

Considerando que os métodos de PQS caminham por pontos infactíveis, devemos levar em conta não apenas o decréscimo no valor da função objetivo de uma iteração para a outra, mas também o quanto o iterando seguinte estará mais próximo do conjunto viável do que o iterando atual. O dever de manter o equilíbrio entre factibilidade e otimalidade nos métodos de PQS cabe a uma função de mérito. Além disso, as funções de mérito também desempenham um papel fundamental na análise de convergência global. Os resultados que apresentaremos neste contexto têm como função de mérito o Lagrangiano Aumentado, que possui as desejáveis características de ser uma função de penalidade exata e também diferenciável [6, 13, 4, 7]. Outra importante função de mérito, embora não diferenciável, é a função de penalidade exata ℓ_1 , a qual foi utilizada por Han [18] para provar a convergência global de um algoritmo de PQS no caso de problemas convexos.

O decréscimo no valor da função de mérito nos métodos de PQS pode significar ganho em otimalidade, ganho em factibilidade, ou ambos. No entanto, exigir que tal valor decresça sempre, de uma iteração para a outra, pode fazer com que o

algoritmo perca velocidade de convergência e, até mesmo, que pare em um ponto não estacionário para a função de mérito. Este problema, de comum ocorrência nos métodos de PQS, é conhecido como Efeito Maratos. Uma técnica utilizada para evitar tal efeito é a busca não-monótona. Trata-se de um tipo de busca que permite um aumento no valor da função objetivo de uma iteração para outra e favorece passos mais arrojados. A busca monótona pode impedir a aceitação do passo unitário mesmo perto de uma solução, o que desacelera o método ou, até mesmo, impede a convergência do algoritmo. Já a busca não-monótona favorece a aceitação do passo unitário, acelerando o método e melhorando sua performance. Este tipo de busca parece ter sido apresentado inicialmente por Grippo, Lampariello e Lucidi [16] juntamente com o método de Newton no contexto de minimização irrestrita. Sua atuação contra o Efeito Maratos na resolução de problemas com restrições de desigualdades via PQS aparece no trabalho de Bonnans et al. [8]. Os dois últimos autores estudaram o uso da busca linear não-monótona também para problemas do tipo minimax [41]. Um trabalho envolvendo otimização irrestrita sem o uso de derivadas (*derivative-free*) e busca linear não-monótona foi desenvolvido recentemente por Diniz-Ehrhardt et al. [12].

Embora os métodos de PQS tenham se popularizado como métodos infactíveis, abordagens factíveis estão presentes em diversos trabalhos. Se o conjunto viável do problema for composto por restrições de desigualdades, manter a factibilidade dos iterandos não custa caro e é de grande utilidade nas situações em que se trabalha com aplicações em tempo real e também nos casos em que a função objetivo não é definida fora da região viável. O uso de direções factíveis em um método de pontos interiores para otimização não-linear restrita foi apresentado por Herskovits [19, 20]. O mesmo autor desenvolveu também um trabalho em conjunto com Panier e Tits [32], no contexto de direções factíveis, onde foi apresentada uma técnica de curvar a direção de busca. Desta forma a busca deixa de ser linear e passa a ser realizada em um arco, o que também atua contra o Efeito Maratos e favorece a ocorrência, em conjunto, de passo completo e factibilidade. Uma combinação de direções de descida, factibilidade e convergência superlinear para problemas com restrições de desigualdades são encontrados em outros trabalhos de Panier e Tits [30, 31].

E as condições de qualificação? Onde entram em toda esta história? Tais como as condições de otimalidade dos problemas de minimização restrita, as condições de qualificação também estão presentes em todos os resultados de convergência dos métodos de PQS. A grande maioria destes resultados utiliza a regularidade

como hipótese. Porém, quanto mais fraca for uma condição de qualificação, maior é a quantidade de minimizadores que podem satisfazê-la e, portanto, maior pode ser o alcance do algoritmo que tem tal condição como hipótese. Isso quer dizer que condições de qualificação mais fracas implicam em resultados de convergência mais fortes. Nosso objetivo é o estudo de algoritmos de PQS e resultados baseados em diferentes condições de qualificação.

Alguns conceitos fundamentais da minimização restrita serão apresentados no Capítulo 2. No Capítulo 3 apresentaremos diferentes condições de qualificação e analisaremos algumas relações existentes entre elas. Uma idéia geral dos métodos clássicos de PQS, os inactíveis, será dada no Capítulo 4 e um resultado de convergência global será apresentado para a função de mérito Lagrangiano Aumentado. Ainda discutiremos brevemente a atuação da busca não-monótona contra o Efeito Maratos. No Capítulo 5 apresentaremos um método de PQS factível e resultados de convergência para diferentes condições de qualificação. Finalmente, no Capítulo 6, experimentos numéricos serão analisados.

Capítulo 2

Conceitos Fundamentais

Neste trabalho vamos considerar um problema de otimização não-linear com restrições de igualdade e desigualdade

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0 \\ & g(x) \leq 0, \end{aligned} \tag{2.1}$$

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

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

o qual chamaremos de *conjunto factível* ou *viável*. Em consequência, pontos pertencentes a D serão chamados *pontos factíveis* ou *viáveis*. A função a ser minimizada, f , é a função objetivo do problema.

A “melhor” solução que podemos encontrar para (2.1) é um ponto x^* tal que $f(x^*) \leq f(x)$ para todo $x \in D$, isto é, uma *solução global*. Mas sendo esta uma tarefa um tanto pretensiosa, torna-se bastante plausível contentar-nos com uma busca mais modesta, por *soluções locais*. Dizemos que $x^* \in D$ é um minimizador local de f em D se e somente se existe $\epsilon > 0$ tal que $f(x) \geq f(x^*)$ para todo $x \in D$ e $\|x - x^*\| < \epsilon$.

A grande maioria dos algoritmos de otimização buscam candidatos a minimizadores locais para o problema em questão através da análise de propriedades que sejam satisfeitas pelos pontos procurados. Considerando que um minimizador

do problema (2.1), seja ele global ou local, é um ponto ótimo, fica claro o uso da denominação *Condições Necessárias de Otimalidade* (CNO) para tais propriedades. Estas compõem as chamadas condições de *Karush-Kuhn-Tucker* e um ponto que as satisfazem é chamado *ponto KKT*.

Algebricamente, um ponto KKT do problema (2.1) é um ponto $x^* \in D$ para o qual existem vetores $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}_+^p$ tais que

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) &= 0, \\ h(x^*) &= 0, \quad g(x^*) \leq 0, \\ \mu_j^* g_j(x^*) &= 0, \quad j = 1, \dots, p. \end{aligned} \tag{2.2}$$

Note que o gradiente da função objetivo é uma combinação linear dos gradientes das restrições e os escalares que verificam tal combinação, λ_i^* e μ_j^* , são conhecidos como multiplicadores de Lagrange associados a x^* .

As condições KKT são fáceis de verificar. No entanto, elas não são, por si só, condições de otimalidade pois existem minimizadores (até mesmo globais) que não as satisfazem, ou seja, existem minimizadores que não são pontos estacionários. Vejamos o exemplo a seguir [26]:

Exemplo 2.1.

$$\begin{aligned} \min \quad & x_1 \\ \text{s.a} \quad & x_1^2 + x_2^2 = 0. \end{aligned}$$

Temos que $x^* = (0, 0)^T$ é o minimizador global do problema, já que é o único ponto factível, mas as condições KKT não se verificam neste ponto. Isso acontece porque $x^* = (0, 0)^T$ não possui uma propriedade que, quando satisfeita por um minimizador, torna obrigatório o cumprimento das condições KKT por tal ponto. Uma *condição de qualificação* é uma propriedade das restrições do problema (2.1) que quando satisfeita em um minimizador local x^* implica que x^* verifica (2.2).

Podemos então definir condições necessárias de primeira ordem para um minimizador local do problema (2.1).

Teorema 2.1 (CN1). *Seja x^* um minimizador local do problema (2.1) e suponha-se que x^* satisfaz uma dada condição de qualificação. Então existem $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}_+^p$ tais que*

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) &= 0, \\ h(x^*) &= 0, \quad g(x^*) \leq 0, \\ \mu_j^* g_j(x^*) &= 0, \quad j = 1, \dots, p. \end{aligned}$$

Demonstração: Ver Luenberger [25].

Antes de finalizarmos o capítulo, vamos à definição do conjunto de restrições ativas. Definimos por

$$I(x) = \{j \in \{1, \dots, p\} \mid g_j(x) = 0\},$$

o conjunto de índices das restrições ativas de desigualdade no ponto factível x . Nos casos em que $g_j(x) < 0$ temos que $j \notin I(x)$ e dizemos que a restrição está inativa em x . É importante deixar claro que as restrições de igualdade estão sempre ativas no conjunto viável, uma vez que são sempre satisfeitas na igualdade. Desta forma, quando falarmos em *restrições ativas* em um ponto factível x , estaremos nos referindo tanto às restrições de igualdade quanto às de desigualdade ativas em x .

Capítulo 3

Condições de Qualificação

O Exemplo 2.1 visto no Capítulo 2 mostra que, em alguns casos,

minimizador local $\nRightarrow$ condições KKT,

embora tal implicação fosse bastante desejável, uma vez que as condições KKT são de fácil verificação. Porém, se atreladas a uma condição adicional, podem ser consideradas boas CNO. É neste contexto que as condições de qualificação desempenham um importante papel.

Podemos afirmar que

minimizador local $\Rightarrow$ (KKT ou Não-CQ),

isto é, existem duas possibilidades para um minimizador x^* : ou x^* satisfaz as condições KKT, ou não satisfaz nenhuma condição de qualificação.

Há um trabalho adicional para determinar se um ponto satisfaz ou não uma CQ. Espera-se, então, que estas sejam propriedades de fácil verificação. Além disso, a implicação acima deve fornecer informações relevantes a respeito do ponto em questão para que o custo de verificar CQ e KKT não seja em vão. Ora,

minimizador local $\Rightarrow$ ponto factível,

mas esta é uma condição necessária de otimalidade tão fraca que talvez, por si só, nem mereça a denominação CNO. Daí, podemos intuir que boas condições de otimalidade devem ser fortes e boas condições de qualificação devem ser fracas [26].

3.1 Regularidade, Mangasarian-Fromovitz e Posto Constante

A condição de qualificação mais conhecida é a *regularidade* ou independência linear dos gradientes das restrições ativas.

Definição 3.1 (*LICQ - Linear Independence Constraint Qualification*). Dizemos que $x \in D$ é um ponto regular se os vetores

$$\nabla h_i(x), \quad i = 1, \dots, m, \quad \nabla g_j(x), \quad j \in I(x)$$

são linearmente independentes (LI).

A regularidade desempenha um papel importante em relação à obtenção de condições necessárias de otimalidade de problemas de programação não-linear, assim como em relação à análise de convergência de diversos algoritmos de otimização. No entanto, nem sempre é uma condição fraca o suficiente, pois deixa de ser satisfeita até mesmo para problemas simples. Vejamos:

Exemplo 3.1.

$$\begin{aligned} \min \quad & x_1 \\ \text{s.a} \quad & x_1^2 + x_2^2 = 0 \\ & -x_1 \leq 0. \end{aligned}$$

A solução global do exemplo acima é $x^* = (0, 0)^T$. Note que neste ponto todas as restrições são ativas e os gradientes destas em x^* , $\nabla h_1(x^*) = (0, 0)^T$ e $\nabla g_1(x^*) = (-1, 0)^T$, são linearmente dependentes. Portanto o minimizador global do exemplo não é um ponto regular.

Isto sugere que a regularidade não é, em alguns casos, uma CQ tão fraca quanto gostaríamos. De fato, existem outras condições de qualificação mais fracas, de verificação mais fácil e que ainda assim desempenham um papel relevante relativo à convergência de algoritmos de programação não-linear. Seguem as definições de duas destas condições diretamente implicadas pela regularidade.

Definição 3.2 (*MFCQ - Mangasarian-Fromovitz Constraint Qualification*). Dizemos que $x \in D$ satisfaz a condição de Mangasarian-Fromovitz se os vetores $\nabla h_i(x)$, $i = 1, \dots, m$, são linearmente independentes e existe $d \in \mathbb{R}^n$ tal que

$$\nabla g_j(x)^T d < 0, \quad \forall j \in I(x)$$

e

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

Dado um certo conjunto de índices $I_0 \subset \{1, \dots, m\}$ e um ponto viável x , definimos a matriz $\nabla h_{I_0}(x)$ como aquela cujas colunas são formadas pelos gradientes das restrições que possuem índice pertencente a I_0 , isto é,

$$\nabla h_{I_0}(x) = \left(\nabla h_{i_1}(x) \quad \nabla h_{i_2}(x) \quad \dots \quad \nabla h_{i_{|I_0|}}(x) \right).$$

Definição 3.3 (*CRCQ - Constant Rank Constraint Qualification*). Dizemos que $x \in D$ satisfaz a condição de qualificação de posto constante se existem $\epsilon > 0$ e uma vizinhança $B(x, \epsilon)$ de x , tal que para todo $I_o \subset \{1, \dots, m\}$ e $J_o \subset I(x)$, a matriz

$$(\nabla h_{I_o}(y) \mid \nabla g_{J_o}(y))$$

tem posto constante para todo $y \in B(x, \epsilon)$.

Em outras palavras, x satisfaz CRCQ se o posto de qualquer matriz cujas colunas são formadas por algum subconjunto dos gradientes das restrições ativas não se altera numa vizinhança de x .

As implicações

$$\begin{aligned} \text{LICQ} &\Rightarrow \text{MFCQ}, \\ \text{LICQ} &\Rightarrow \text{CRCQ}, \end{aligned}$$

são verdadeiras e de fácil verificação. Já MFCQ e CRCQ não implicam uma na outra.

Considere um problema com uma única restrição, de igualdade, $h(x) = 0$. Se a quebrarmos em duas desigualdades, $h(x) \leq 0$ e $-h(x) \leq 0$, o problema não se altera; porém enquanto CRCQ é satisfeita por todo ponto viável, MFCQ não se satisfaz em nenhum ponto. Ou seja,

$$\text{CRCQ} \not\Rightarrow \text{MFCQ}.$$

Aqui fica claro que o fato de um ponto viável satisfazer ou não uma certa CQ está diretamente ligado à maneira como as restrições são escritas. Um ponto viável x que cumpre uma CQ pode deixar de cumpri-la se o conjunto factível for determinado por meio de diferentes restrições sem sofrer, no entanto, qualquer alteração. Por isso definimos condição de qualificação como uma propriedade das restrições que determinam a região viável do problema.

Considere agora um problema com as seguintes restrições de desigualdade:

$$\begin{aligned} g_1(x) &= x_1 + x_2^2 \leq 0 \\ g_2(x) &= x_1 \leq 0. \end{aligned}$$

No ponto $\bar{x} = (0, 0)^T$, $\nabla g_1(\bar{x}) = (1, 0)^T$ e $\nabla g_2(\bar{x}) = (1, 0)^T$, sendo impossível uma combinação linear não-negativa destes gradientes resultando no vetor nulo, a menos que os multiplicadores sejam ambos iguais a zero. Isto implica que MFCQ é satisfeita em $\bar{x}$. Já para todo ponto na forma $\tilde{x} = (0, x_2)^T$, $x_2 \neq 0$, numa vizinhança qualquer de $\bar{x}$ temos

$$\nabla g_1(\tilde{x}) = (1, 2x_2)^T \text{ e } \nabla g_2(\tilde{x}) = (1, 0)^T.$$

Assim, toda vizinhança de $\bar{x}$ tem pontos onde o posto da matriz cujas colunas são compostas pelos vetores gradientes das restrições ativas é igual a dois enquanto que, em $\bar{x}$, o posto é igual a um. Portanto CRCQ não se cumpre em $\bar{x}$ e podemos concluir que

$$\text{MFCQ} \not\Rightarrow \text{CRCQ}.$$

3.2 A condição de dependência linear positiva constante

Nesta seção trataremos de uma condição de qualificação relevante pelo papel que desempenha no estudo da convergência de importantes métodos de minimização como, por exemplo, métodos de Lagrangiano Aumentado [36] e métodos clássicos de Programação Quadrática Sequencial, conforme veremos no Capítulo 4. Antes de defini-la, faz-se necessária uma definição auxiliar, a de independência linear positiva.

Definição 3.4 (*PLI - Positive Linear Independence*). Dizemos que os vetores em $\mathbb{R}^n$

$a_1, \dots, a_m$ e $b_1, \dots, b_p$ são *positivo-linearmente dependentes (PLD)* [34] se existem $\alpha_i \in \mathbb{R}$, $i = 1, \dots, m$, $\beta_j \geq 0$, $j = 1, \dots, p$, não todos nulos, tais que

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

Caso contrário dizemos que tais vetores são *positivo-linearmente independentes (PLI)*.

A seguinte proposição foi apresentada na seção 1.8 de [33]:

Proposição 3.1. *Dado qualquer $x \in D$, assumamos que $\{1, \dots, m\} \cup I(x) \neq \emptyset$. Então MFCQ é satisfeita em x se e somente se*

$$\nabla h_i(x), \quad i = 1, \dots, m, \quad \nabla g_j(x), \quad j \in I(x)$$

são positivo-linearmente independentes.

Desta maneira podemos dizer, de maneira equivalente à Definição 3.2, que $x \in D$ satisfaz MFCQ se e somente se

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

com $\mu_j \geq 0$ para todo $j \in I(x)$, implica que

$$\lambda_1 = \dots = \lambda_m = \mu_j = 0,$$

para todo $j \in I(x)$.

Em [34], Qi e Wei apresentaram resultados de convergência relacionando métodos de programação quadrática sequencial com uma condição à qual chamaram condição de *dependência linear positiva constante* (CPLD - *Constant Positive Linear Dependence*). Sua definição, conforme apresentada em [34] segue abaixo.

Definição 3.5 (*CPLD-original*). Um ponto $x \in D$ satisfaz a condição CPLD-original [34] se para todo $I_o \subset \{1, \dots, m\}$ e $J_o \subset I(x)$, sempre que os vetores

$$\nabla h_i(x), i \in I_o, \nabla g_j(x), j \in J_o$$

são positivo-linearmente dependentes, existe uma vizinhança $N(x)$ de x tal que para todo $y \in N(x)$ os vetores $\nabla h_i(y), i \in I_o, \nabla g_j(y), j \in J_o$ são linearmente dependentes.

Em [34] a CPLD não foi usada como uma condição de qualificação, mas apenas como uma propriedade de pontos factíveis admitida para a prova de alguns resultados que serão apresentados no Capítulo 4. Em [36] a CPLD foi redefinida e provou-se que: (i) esta nova definição é equivalente à original; (ii) a CPLD é, de fato, uma condição de qualificação.

Definição 3.6 (*CPLD*). Dizemos que $x \in D$ satisfaz a condição CPLD [36] se para todo $I_o \subset \{1, \dots, m\}$ e $J_o \subset I(x)$, quando os gradientes

$$\nabla h_i(x), i \in I_o, \nabla g_j(x), j \in J_o$$

são positivo-linearmente dependentes, existe uma vizinhança $N(x)$ de x tal que os gradientes $\nabla h_i(y), i \in I_o, \nabla g_j(y), j \in J_o$, continuam positivo-linearmente dependentes para todo $y \in N(x)$.

O lema que segue será usado na prova da Proposição (3.2).

Lema 3.1 (*Lema de Caratheodory*). Sejam $u_1, \dots, u_m, v_1, \dots, v_p$ vetores em $\mathbb{R}^n$. Sejam $x \in \mathbb{R}^n$ e $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ escalares tais que $\mu_j \geq 0, \forall j = 1, \dots, p$ e

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

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

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

e os vetores $u_i, i \in I, v_j, j \in J$, são linearmente independentes.

Proposição 3.2. *CPLD-original* $\Leftrightarrow$ *CPLD*.

Demonstração: Se x^* satisfaz a condição CPLD, é óbvio que também satisfaz a condição CPLD-original, já que vetores positivo-linearmente dependentes são linearmente dependentes. Provemos então que CPLD-original $\Rightarrow$ CPLD. Para isso vamos assumir que $x^* \in D$ satisfaz a CPLD-original mas que não satisfaz a CPLD. Logo existem $I_0 \subset \{1, \dots, m\}$ e $J_0 \subset I(x^*)$ tais que os gradientes $\nabla h_i(x^*)$, $i \in I_0$, $\nabla g_j(x^*)$, $j \in J_0$, são positivo-linearmente dependentes e existe uma sequência (y_k) , $y_k \rightarrow x^*$ para a qual os gradientes $\nabla h_i(y_k)$, $i \in I_0$, $\nabla g_j(y_k)$, $j \in J_0$, são positivo-linearmente independentes.

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

$$\sum_{i \in I_0} \lambda_i \nabla h_i(x^*) + \sum_{j \in J_0} \mu_j \nabla g_j(x^*) = 0. \quad (3.1)$$

Definimos

$$\begin{aligned} I_+(x^*) &= \{i \in I_0 \mid \lambda_i > 0\}, \\ I_-(x^*) &= \{i \in I_0 \mid \lambda_i < 0\}, \\ I_0(x^*) &= \{j \in J_0 \mid \mu_j > 0\}. \end{aligned}$$

Temos que $I_+(x^*) \cup I_-(x^*) \cup I_0(x^*) \neq \emptyset$ e, por (3.1),

$$\sum_{i \in I_+(x^*)} \lambda_i \nabla h_i(x^*) + \sum_{i \in I_-(x^*)} \lambda_i \nabla h_i(x^*) + \sum_{j \in I_0(x^*)} \mu_j \nabla g_j(x^*) = 0. \quad (3.2)$$

Primeiramente, vamos assumir que $I_+(x^*) \neq \emptyset$. Seja $i_1 \in I_+(x^*)$. Sendo assim, por (3.2),

$$\lambda_{i_1} \nabla h_{i_1}(x^*) = - \sum_{i \in I_+(x^*) - \{i_1\}} \lambda_i \nabla h_i(x^*) - \sum_{i \in I_-(x^*)} \lambda_i \nabla h_i(x^*) - \sum_{j \in I_0(x^*)} \mu_j \nabla g_j(x^*). \quad (3.3)$$

Consideremos duas possibilidades:

- (a) $\nabla h_{i_1}(x^*) = 0$;
- (b) $\nabla h_{i_1}(x^*) \neq 0$.

Se (a) ocorre, o vetor $\nabla h_{i_1}(x^*)$ é positivo-linearmente dependente. Daí, pela condição CPLD-original, $\nabla h_{i_1}(y)$ é linearmente dependente para todo y numa vizinhança de x^* . Então, $\nabla h_{i_1}(y) = 0$ para todo y nessa vizinhança e segue que,

para k suficientemente grande, digamos $k \geq k_0$, temos que $\nabla h_{i_1}(y_k)$ é PLD, o que é uma contradição (já que qualquer subconjunto de um conjunto com vetores positivo-linearmente independentes possui apenas vetores positivo-linearmente independentes).

Vamos assumir agora que (b) se verifica. Então pelo Lema (3.1), existem

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

e escalares

$$\begin{aligned} \bar{\lambda}_i &> 0, \quad \forall i \in I_{++}, \\ \bar{\lambda}_i &< 0, \quad \forall i \in I_{--}, \\ \bar{\mu}_j &> 0, \quad \forall j \in I_{00}, \end{aligned}$$

tais que os vetores

$$\nabla h_i(x^*), i \in I_{++}, \quad \nabla h_i(x^*), i \in I_{--}, \quad \nabla g_j(x^*), j \in I_{00} \quad (3.4)$$

são linearmente independentes e

$$\lambda_{i_1} \nabla h_{i_1}(x^*) = - \sum_{i \in I_{++}} \bar{\lambda}_i \nabla h_i(x^*) - \sum_{i \in I_{--}} \bar{\lambda}_i \nabla h_i(x^*) - \sum_{j \in I_{00}} \bar{\mu}_j \nabla g_j(x^*). \quad (3.5)$$

Pela independência linear dos vetores em (3.4) e por argumentos de continuidade, os vetores

$$\nabla h_i(y), i \in I_{++}, \quad \nabla h_i(y), i \in I_{--}, \quad \nabla g_j(y), j \in I_{00} \quad (3.6)$$

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

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

são linearmente dependentes numa vizinhança de x^* . Portanto $\lambda_{i_1} \nabla h_{i_1}(y)$ deve ser uma combinação linear dos vetores (3.6) para todo y numa vizinhança de x^* .

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

$$\lambda_{i_1} \nabla h_{i_1}(y_k) = - \sum_{i \in I_{++}} [\alpha_k]_i \nabla h_i(y_k) - \sum_{i \in I_{--}} [\alpha_k]_i \nabla h_i(y_k) - \sum_{j \in I_{00}} [\beta_k]_j \nabla g_j(y_k). \quad (3.7)$$

Se a sequência $\{(\alpha_k, \beta_k)\}$ é não limitada, dividindo a igualdade (3.7) por $\|(\alpha_k, \beta_k)\|$, considerando uma subsequência adequada e fazendo k tender a infinito para esta subsequência obtemos:

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

com α_i, β_j não todos nulos. Logo os gradientes

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

são linearmente dependentes, o que contradiz (3.4).

Assim, a sequência $\{(\alpha_k, \beta_k)\}$ é limitada e existem α_i, β_j não todos nulos tais que $[\alpha_k]_i \rightarrow \alpha_i$ e $[\beta_k]_j \rightarrow \beta_j$. (O caso $\alpha = \beta = 0$ já foi analisado, pois resulta em $\nabla h_{i_1}(x^*) = 0$). Tomando o limite quando $k \rightarrow \infty$ em (3.7) temos que

$$\lambda_{i_1} \nabla h_{i_1}(x^*) = - \sum_{i \in I_{++}} \alpha_i \nabla h_i(x^*) - \sum_{i \in I_{--}} \alpha_i \nabla h_i(x^*) - \sum_{j \in I_{00}} \beta_j \nabla g_j(x^*) \quad (3.8)$$

e subtraindo (3.8) de (3.5) obtemos que

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

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

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

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

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

o que, por (3.7), implica que os gradientes

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

são PLD. Contradição.

As provas para os casos $I_-(x^*) \neq \emptyset$ e $I_0(x^*) \neq \emptyset$ são inteiramente análogas ao caso $I_+(x^*) \neq \emptyset$. Portanto, a condição CPLD-original implica na condição CPLD, como se queria demonstrar. ■

A prova de que a CPLD é uma condição de qualificação encontra-se em [36]. A importância destes resultados reside no fato de que a CPLD é uma condição ainda mais fraca que MFCQ e CRCQ. Em seu trabalho, Schuverdt [36] não somente provou que a CPLD é uma CQ, mas também posicionou esta condição em relação às demais. Implicações anteriores podem, assim, ser estendidas para:

$$\begin{aligned} \text{Regularidade} &\Rightarrow \text{MFCQ} \Rightarrow \text{CPLD}, \\ \text{Regularidade} &\Rightarrow \text{CRCQ} \Rightarrow \text{CPLD}. \end{aligned}$$

Consideremos o seguinte conjunto de restrições de desigualdade:

$$\begin{aligned} g_1(x) &= x_1 + x_2 \\ g_2(x) &= 2x_1 + 2x_2. \end{aligned}$$

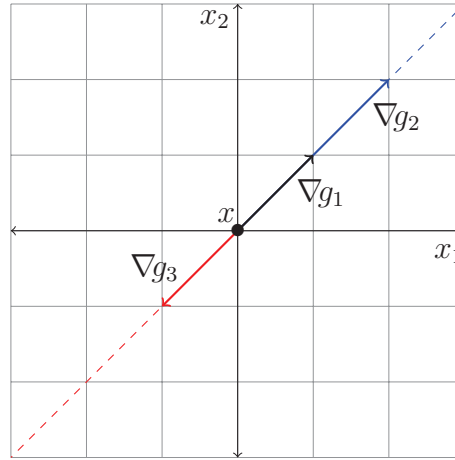


Figura 3.1: Uma análise geométrica de LICQ, MFCQ e CPLD.

Claramente os vetores $\nabla g_1(x)$ e $\nabla g_2(x)$ não satisfazem LICQ (assim como acontece para qualquer restrição $g_2(x)$ tal que $\nabla g_2(x)$ seja um vetor pertencente à reta pontilhada na Figura (3.2)). Mas satisfazem MFCQ já que $\mu_1 \nabla g_1(x) + \mu_2 \nabla g_2(x) = 0$, com $\mu_1, \mu_2 \geq 0$, só é possível quando $\mu_1 = \mu_2 = 0$.

Agora consideremos as restrições de desigualdade

$$\begin{aligned} g_1(x) &= x_1 + x_2 \\ g_3(x) &= -x_1 - x_2. \end{aligned}$$

Neste caso tanto LICQ quanto MFCQ não são satisfeitas pois os gradientes são linearmente dependentes e $\nabla g_1(x) + \nabla g_3(x) = 0$. No entanto $\nabla g_1(x)$ e $\nabla g_3(x)$ são vetores PLD qualquer que seja $x \in \mathbb{R}^2$ e, dessa forma, a CPLD é satisfeita por qualquer ponto viável. Note que apenas a CPLD será satisfeita na semi-reta que contém $\nabla g_3(x)$ na Figura (3.2) sempre que $g_3(x)$ for uma restrição do tipo $g_3(x) = -\epsilon x_1 - \epsilon x_2$, $\epsilon > 0$.

Duas últimas condições de qualificação, a quase-normalidade e a pseudo-normalidade, são apresentadas por [21] e [5], respectivamente.

Definição 3.7. Dizemos que $x \in D$ é quase-normal [21] se não existem escalares $\lambda_1, \dots, \lambda_m$, $\mu_1, \dots, \mu_p \in \mathbb{R}$ tais que:

1. $\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j=1}^p \mu_j \nabla g_j(x) = 0$.
2. $\mu_j \geq 0$, $j \in I(x)$.
3. $\lambda_1, \dots, \lambda_m, \{\mu_j\}_{j \in I(x)}$ não são todos nulos.
4. Em toda vizinhança $N(x)$ de x existe um ponto $y \in N(x)$ tal que $\lambda_i h_i(y) > 0$ para todo i com $\lambda_i \neq 0$ e $\mu_j g_j(y) \geq 0$ para todo j com $\mu_j \neq 0$.

Definição 3.8. Dizemos que $x \in D$ é pseudo-normal [5] se não existem escalares $\lambda_1, \dots, \lambda_m$, $\mu_1, \dots, \mu_p \in \mathbb{R}$ tais que:

1. $\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in I(x)} \mu_j \nabla g_j(x) = 0$.
2. $\mu_j \geq 0$, $j \in I(x)$.
3. Em toda vizinhança $N(x)$ de x existe um ponto $y \in N(x)$ tal que

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

A pseudo-normalidade, introduzida por Bertsekas e Ozdaglar em [5], é mais fácil de ser manipulada do que a quase-normalidade, porém esta última é mais fraca que a primeira. A pseudo-normalidade é utilizada para analisar a relação entre a existência de multiplicadores de Lagrange e a admissão de função de penalidade exata para problemas de programação matemática com conjunto abstrato de restrições [36].

Uma demonstração rigorosa de que a condição CPLD implica a quase-normalidade é encontrada em [1] e [36], o que significa, em outras palavras, que a quase-normalidade é uma CQ ainda mais fraca que a dependência linear positiva constante. No entanto, neste trabalho a CPLD mostra-se mais útil que a quase-normalidade. A condição de qualificação admitida em um certo método que busca minimizadores locais, testando as condições KKT, vai determinar quão grande é a proporção do conjunto viável que será “analisada” por este método. Quanto mais fraca for uma condição de qualificação, maior é a quantidade de pontos que devem, possivelmente, satisfazê-la. Então é bastante coerente imaginar que quanto mais fraca for uma CQ, melhor será o desempenho do método, já que corremos um risco menor de perder minimizadores “não-CQ”. No entanto, uma condição de qualificação fraca nem sempre é de fácil manipulação e utilizá-la numa prova de convergência pode ser inviável. A quase-normalidade é um exemplo de uma condição de qualificação fraca de difícil manipulação algébrica. Já a CPLD apresentou resultados muito significativos tanto teóricos, no trabalho de Qi e Wei em [34], quanto nos experimentos numéricos realizados em [36].

Capítulo 4

Programação quadrática sequencial

Problemas de programação não-linear com restrições de igualdade e desigualdade são, em sua grande maioria, de difícil resolução. Os métodos com o propósito de resolvê-los baseiam-se, freqüentemente, em estratégias que buscam simplificar estes problemas complicados. Uma dessas estratégias, por exemplo, é transformar o problema restrito em um irrestrito, de maneira que as soluções de ambos coincidam, e usar técnicas conhecidas de minimização irrestrita. Este é o funcionamento básico dos métodos de barreira e penalização.

Outra abordagem consiste em resolver uma sequência de problemas mais simples cujas soluções convirjam para a solução do problema original. Partindo de uma estimativa x^k , é de se esperar que a solução de um problema mais simples, que pelo menos localmente representa bem o problema original, forneça uma aproximação ainda melhor para a solução procurada. Esta é uma estratégia muitas vezes eficiente usada, por exemplo, pelo tão conhecido método de Newton no cálculo de zeros de funções. É exatamente nesta estratégia que se baseia a estrutura da *programação quadrática sequencial*. Dado um problema de programação não-linear com restrições, a cada iteração resolve-se um subproblema onde a função objetivo é substituída por uma aproximação quadrática e as restrições substituídas por restrições lineares.

Vamos supor nesta seção que nosso problema não-linear contém apenas restrições de igualdade a fim de simplificar as discussões e análises subsequentes. Pode-se pensar que estamos trabalhando em um caso em que somos capazes de

identificar as restrições ativas numa solução x^* . Sendo assim, e considerando a continuidade das funções envolvidas, é de se esperar que numa vizinhança de x^* o conjunto de restrições ativas não mude. Então abandonamos as desigualdades não-ativas e consideramos somente as ativas, trabalhando com um conjunto de restrições de igualdade que nos permitirá realizar uma boa análise local do problema.

4.1 O subproblema quadrático

Dado um problema na forma

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0, \end{aligned} \tag{4.1}$$

e uma aproximação x^k para a solução do mesmo, construímos um subproblema quadrático tomando uma aproximação quadrática da função objetivo f e aproximações lineares das restrições em torno do ponto x^k . Obtemos, assim, o seguinte problema:

$$\begin{aligned} \min \quad & \frac{1}{2}(x - x^k)^T B_k (x - x^k) + \nabla f(x^k)^T (x - x^k) + f(x^k) \\ \text{s.a} \quad & h(x^k) + \nabla h(x^k)^T (x - x^k) = 0, \end{aligned}$$

onde $\nabla h(x^k) = [\nabla h_1(x^k), \dots, \nabla h_m(x^k)]$ e $B_k^T = B_k \in \mathbb{R}^{n \times n}$.

A solução deste problema consistirá numa melhor aproximação para a solução do problema original, passando a ser nosso ponto x^{k+1} . Desta forma, denotando $d^k = x^{k+1} - x^k$, podemos pensar no problema reescrito da seguinte maneira:

$$\begin{aligned} \min \quad & \frac{1}{2}d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a} \quad & \nabla h(x^k)^T d + h(x^k) = 0. \end{aligned} \tag{4.2}$$

Note que $f(x^k)$ é uma constante e por isso não integra a nova função objetivo. Como já sabemos, um minimizador do problema (4.2), se existe e satisfaz alguma condição de qualificação, deve também satisfazer as condições necessárias de otimalidade, ou condições KKT. Logo, existe $\lambda^k \in \mathbb{R}^m$ tal que

$$\begin{aligned} B_k d^k + \nabla f(x^k) + \nabla h(x^k) \lambda^k &= 0 \\ \nabla h(x^k)^T d^k + h(x^k) &= 0. \end{aligned} \tag{4.3}$$

É importante não esquecer que agora x^k é um ponto dado e d^k e λ^k são as variáveis a serem descobertas. Reescrevendo o sistema acima, torna-se claro o motivo pelo qual é preferível resolver (4.2) em vez de (4.1). As condições KKT de (4.2) recaem num sistema linear, portanto de resolução mais barata que o sistema não-linear resultante das condições KKT do problema (4.1). O sistema linear reescrito fica:

$$\begin{aligned} B_k d^k + \nabla h(x^k) \lambda^k &= -\nabla f(x^k) \\ \nabla h(x^k)^T d^k &= -h(x^k). \end{aligned} \tag{4.4}$$

Diversos são os recursos que podem ser utilizados na resolução de (4.4). No entanto, não há garantia de que tal sistema tenha sempre solução e sem solução para (4.4) nossa idéia de tomar $x^{k+1} = x^k + d^k$ não pode seguir adiante.

Pensemos no problema (4.2). Há duas possíveis causas para que ele não possua solução. A primeira é a inexistência de pontos factíveis, isto é, não existe $d \in \mathbb{R}^n$ tal que $\nabla h(x^k)^T d + h(x^k) = 0$. A segunda é o modelo quadrático ser ilimitado na região viável.

Para contornar a primeira causa, podemos encontrar a solução de quadrados mínimos para o conjunto de restrições resolvendo,

$$\min \|\Psi(x^k)\|_2^2$$

onde $\Psi(x^k) = \nabla h(x^k)^T (x - x^k) + h(x^k)$.

Denotando a solução encontrada por x_{nor}^k podemos redefinir (4.2) como

$$\begin{aligned} \min \quad & \frac{1}{2} d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a} \quad & \nabla h(x^k)^T d + h(x^k) = \Psi(x_{nor}^k). \end{aligned} \tag{4.5}$$

Note que se (4.2) é factível, x_{nor}^k é tal que $\Psi(x_{nor}^k) = 0$ e, portanto, as soluções de (4.2) e (4.5) coincidem. Com isso garantimos a factibilidade dos subproblemas

e, ainda mais, sem alterá-los nos casos em que já forem factíveis [27].

Uma vez determinada uma direção d^k que resolve (4.5), obtemos uma nova aproximação $x^{k+1} = x^k + d^k$ para a solução do problema original. A menos que x^k seja um minimizador de (4.1), muito provavelmente x^k será um ponto infactível para este problema. De fato, se x^k for factível, $h(x^k) = 0$ e portanto o sistema linear

$$\nabla h(x^k)^T d^k = -h(x^k) = 0$$

tem solução trivial $d^k = 0$ se x^k é um ponto regular. Logo $x^{k+1} = x^k + d^k = x^k$ e nenhum passo a mais na direção de um minimizador pode ser dado. Isto sugere que a solução ótima é atingida quando $d^k = 0$ resolve o subproblema na iteração k . Esta é a condição que determina o critério de parada de um algoritmo de programação quadrática sequencial.

Já para evitar que o modelo quadrático seja ilimitado no conjunto viável em questão, podemos pedir, por exemplo, que as matrizes B_k sejam definidas positivas; assim o modelo quadrático é convexo e, portanto, possui solução única. Ao formular tal modelo, uma escolha razoável e intuitiva para B_k parece ser a Hessiana da função objetivo $\nabla^2 f(x^k)$. No entanto veremos que isto não é verdade. A escolha das matrizes B_k , assunto de fundamental importância nos resultados relativos à convergência dos métodos de PQS, será discutida a seguir.

4.2 A Hessiana do Lagrangiano

Para iniciar a presente seção vamos definir uma função de fundamental importância em toda a teoria de otimização restrita, a *função Lagrangiana* associada ao problema (4.1):

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

para todo $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m$. Note que o gradiente desta função é dado por

$$\nabla L(x, \lambda) = \begin{bmatrix} \nabla f(x) + \nabla h(x)^T \lambda \\ h(x) \end{bmatrix}.$$

A partir desta seção vamos supor também a continuidade das derivadas de segunda ordem das funções $f(x)$ e $h_j(x) \forall j = 1, \dots, m$. Um minimizador local

regular x^* cumpre, adicionada às condições KKT, uma condição necessária de segunda ordem (CN2) exposta no teorema a seguir:

Teorema 4.1 (CN2). *Seja x^* um minimizador local de (4.1) e suponhamos que x^* seja um ponto regular. Então existe $\lambda^* \in \mathbb{R}^m$ tal que*

$$\begin{aligned}\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) &= 0, \\ h(x^*) &= 0, \\ y^T \nabla_{xx}^2 L(x^*, \lambda^*) y &\geq 0, \quad \forall y \in N(\nabla h(x^*)^T).\end{aligned}$$

Demonstração: Ver Luenberger [25].

Suponhamos que x^* seja um minimizador local do problema (4.1) com multiplicador de Lagrange associado λ^* e, por ora, consideremos que x^* seja um ponto regular. Então o par (x^*, λ^*) satisfaz as condições KKT de (4.1):

$$\begin{aligned}\nabla f(x^*) + \nabla h(x^*) \lambda^* &= 0 \\ h(x^*) &= 0,\end{aligned}$$

o que é equivalente a $\nabla L(x^*, \lambda^*) = 0$.

Além disso a matriz Jacobiana do sistema acima em x^* é dada por

$$J_* = \begin{bmatrix} \nabla^2 f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla^2 h_j(x^*) & \nabla h(x^*) \\ \nabla h(x^*)^T & 0 \end{bmatrix}$$

ou, de maneira simplificada,

$$J_* = \begin{bmatrix} \nabla_{xx}^2 L(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)^T & 0 \end{bmatrix}.$$

As condições necessárias, às quais chamamos CN1 no Capítulo 2, são condições que todo minimizador regular deve obrigatoriamente satisfazer. Existem também condições suficientes de otimalidade que, quando satisfeitas por um ponto, nos garantem que este é um minimizador local. As condições suficientes (CS) para o problema (4.1) encontram-se no teorema a seguir:

Teorema 4.2. *(CS) Se existe $\lambda^* \in \mathbb{R}^m$ tal que*

$$\nabla f(x^*) + \nabla h(x^*)\lambda^* = 0,$$

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

$$y^T \nabla_{xx}^2 L(x^*, \lambda^*) y > 0 \quad \forall y \in N(\nabla h(x^*)^T),$$

então x^ é um minimizador local estrito de (4.1).*

Demonstração: Ver Luenberger [25].

Aqui vamos pedir que as condições suficientes (CS) estejam satisfeitas no par (x^*, λ^*) e supor que x^* é um ponto regular (isto é importante para garantir a existência da solução do sistema que será apresentado a seguir). Para simplificar a notação, denotemos $\nabla_{xx}^2 L(x, \lambda)$ por $\nabla^2 \ell(x, \lambda)$ e $\nabla_x L(x, \lambda)$ por $\nabla \ell(x, \lambda)$. Com isso temos que $\nabla^2 \ell(x^*, \lambda^*)$ é definida positiva no núcleo de $\nabla h(x^*)^T$ e segue por argumentos de álgebra linear que a matriz Jacobiana J_* é não singular (ver [4], pp. 69 - Lema 1.27). Pela continuidade das funções envolvidas, para (x^k, λ^k) suficientemente próximo de (x^*, λ^*) , $\nabla^2 \ell(x^k, \lambda^k)$ permanece não singular e podemos, assim, aplicar o método de Newton ao sistema não-linear nas variáveis (x, λ)

$$\begin{aligned} \nabla f(x) + \nabla h(x)\lambda &= 0, \\ h(x) &= 0, \end{aligned}$$

obtendo, a partir de (x^k, λ^k) , o próximo iterando $(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k) + (d^k, p^k)$, onde (d^k, p^k) é a solução do sistema linear resultante ao aplicarmos o método de Newton,

$$\begin{bmatrix} \nabla^2 \ell(x^k, \lambda^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} d^k \\ p^k \end{bmatrix} = - \begin{bmatrix} \nabla \ell(x^k, \lambda^k) \\ h(x^k) \end{bmatrix}. \quad (4.6)$$

Daí, reescrevendo o sistema acima, ficamos com

$$\begin{aligned} \nabla^2 \ell(x^k, \lambda^k) d^k + \nabla h(x^k) p^k + \nabla \ell(x^k, \lambda^k) &= 0, \\ \nabla h(x^k)^T d^k + h(x^k) &= 0, \end{aligned} \quad (4.7)$$

e dado que

$$\begin{aligned}
\nabla h(x^k)p^k + \nabla \ell(x^k, \lambda^k) &= \nabla h(x^k)p^k + \nabla f(x^k) + \nabla h(x^k)\lambda^k \\
&= \nabla f(x^k) + \nabla h(x^k)(\lambda^k + p^k) \\
&= \nabla f(x^k) + \nabla h(x^k)\lambda^{k+1},
\end{aligned}$$

podemos escrever o sistema (4.7), finalmente, da seguinte forma

$$\begin{aligned}
\nabla^2 \ell(x^k, \lambda^k)d^k + \nabla f(x^k) + \nabla h(x^k)\lambda^{k+1} &= 0, \\
\nabla h(x^k)^T d^k + h(x^k) &= 0.
\end{aligned} \tag{4.8}$$

Por outro lado, se considerarmos o subproblema (4.2) e o sistema (4.3) que descreve as condições KKT do mesmo, é possível perceber que (4.3) e (4.8) serão exatamente o mesmo sistema (exceto pelo multiplicador de Lagrange que, neste último, refere-se ao multiplicador da próxima iteração, λ^{k+1}) se tomarmos B^k como sendo a Hessiana do Lagrangiano $\nabla^2 L(x^k, \lambda^k)$. E aqui fica claro porque a $\nabla^2 f(x^k)$ não é a melhor escolha para as matrizes B^k ; isso se explica pelo fato de que ao tomar $B^k = \nabla^2 f(x^k)$ estamos deixando de lado toda a informação referente à curvatura das restrições.

Com isso podemos concluir que, se $B^k = \nabla^2 \ell(x^k, \lambda^k)$, o método de programação quadrática sequencial é equivalente ao método de Newton aplicado às condições KKT do problema original. Logo, o método PQS apresentado possui as mesmas propriedades de convergência local que o método de Newton, ou seja, a sequência (x^k, λ^k) gerada por um algoritmo de PQS irá convergir quadraticamente para uma solução local (x^*, λ^*) se (x^0, λ^0) for tomado suficientemente próximo de (x^*, λ^*) , se x^* satisfizer a regularidade e as condições suficientes e B_k for escolhida como $\nabla^2 \ell(x^k, \lambda^k)$ para todo $k \in \mathbb{N}$.

Aplicando o método de programação quadrática sequencial ao problema do Exemplo 4.1, ao construir o subproblema quadrático com $x^k = (0.1, 1.5)$ e $B_k = \nabla^2 f(x^k)$, caímos em um subproblema ilimitado, onde é possível decrescer o valor da função objetivo tanto quanto se queira caminhando sobre a aproximação linear da restrição original, conforme Figura 4.1. Já escolhendo $B_k = \nabla^2 \ell(x^k, -0.5)$, o subproblema quadrático deixa de ser ilimitado e passa a ter uma única solução, denotada pela bolinha sobre a aproximação linear. Isso ocorre pois, ao considerar $B_k = \nabla^2 \ell(x^k, -0.5)$, levamos em conta informações referentes à curvatura da restrição, o que modifica a função objetivo do subproblema e permite a resolução do mesmo.

Exemplo 4.1.

$$\begin{aligned} \min \quad & (x_2 - 2)^2 - x_1^2 \\ \text{s.a} \quad & 4x_1^2 + x_2^2 - 1 = 0. \end{aligned}$$

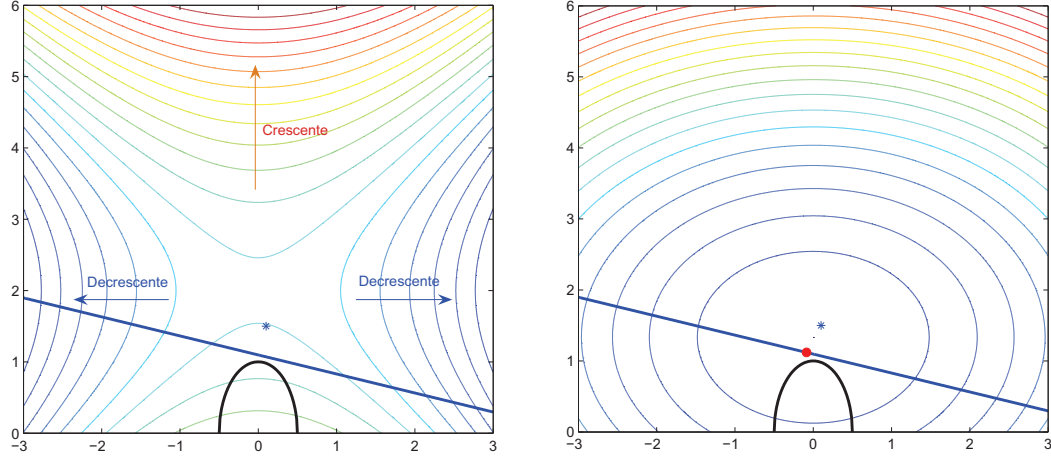


Figura 4.1: À esquerda, o subproblema quadrático com $B_k = \nabla^2 f(x^k)$ e à direita com $B_k = \nabla^2 \ell(x^k, -0.5)$, onde $x^k = (0.1, 1.5)$. A solução é $x^* = (0, 1)^T$ com multiplicador $\lambda^* = -1$.

Na Figura 4.2, segue o resultado de quatro iterações do método de programação quadrática sequencial clássico aplicado ao Exemplo 4.1, com $x^0 = (2, 4)^T$, $\lambda^0 = -0.5$ e $B_k = \nabla^2 \ell(x^k, \lambda^k)$ para todo $k = 0, 1, 2, 3$.

O resultado não poderia ser melhor se pudéssemos desconsiderar dois importantes e preocupantes fatores: escolher um ponto inicial suficientemente próximo de uma solução pode ser tão difícil quanto a própria resolução do problema em questão; calcular a Hessiana do Lagrangiano a cada iteração é um custo alto demais na maioria dos casos. Então faz-se necessário conhecer formas de manter as matrizes B_k próximas das verdadeiras Hessianas do Lagrangiano e atualizá-las a cada iteração sem grande custo computacional. Além disso precisamos de um algoritmo que não exija muito sobre o ponto inicial, mesmo que para isso tenhamos que perder velocidade de convergência.

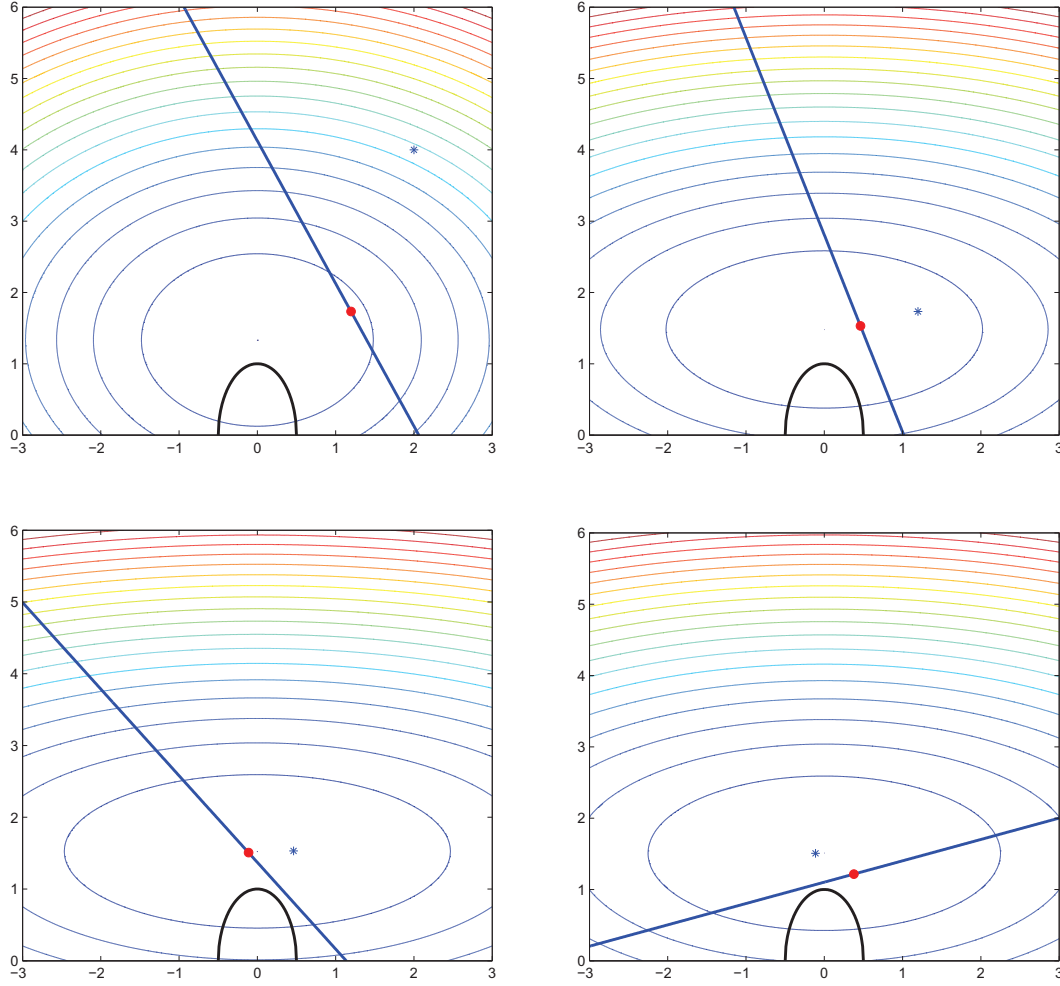


Figura 4.2: Quatro iterações do método de programação quadrática sequencial clássico aplicado ao Exemplo 4.1, com $x^0 = (2, 4)^T$, $\lambda^0 = -0.5$ e $B_k = \nabla^2 \ell(x^k, \lambda^k)$ para todo $k = 0, 1, 2, 3$. Acima, à esquerda, $x^0 = [2, 4]^T$ identificado pelo asterisco; à direita $x^1 = [1.1964, 1.7321]^T$; abaixo à esquerda $x^2 = [0.4621, 1.5308]^T$; e abaixo à direita $x^3 = [-0.1128, 1.5072]^T$. A bolinha vermelha indica, em todos os casos, a solução do subproblema quadrático. A solução global é $x^* = [0, 1]^T$ com multiplicador de Lagrange $\lambda^* = -1$.

Pelo Teorema 4.1 vemos que um fato importante a ser considerado é que $\nabla^2 \ell(x^*, \lambda^*)$ deve ser semi-definida positiva no núcleo de $\nabla h(x^*)^T$, o que não implica, no entanto, que tal matriz seja “totalmente” semi-definida positiva. É

neste contexto que surgem as aproximações *completas* e *reduzidas* para a Hessiana do Lagrangiano, matrizes que podem aproximar toda a $\nabla^2 \ell(x^*, \lambda^*)$ ou apenas aproximá-la no espaço nulo de $\nabla h(x^*)^T$.

Os principais exemplos de aproximações completas para a Hessiana estão entre as conhecidas aproximações *secantes*. Por argumentos de continuidade, temos que

$$\nabla \ell(x^{k+1}, \lambda^{k+1}) - \nabla \ell(x^k, \lambda^{k+1}) \approx \nabla^2 \ell(x^{k+1}, \lambda^{k+1})(x^{k+1} - x^k),$$

valendo a igualdade se a função Lagrangiana é quadrática em x . É razoável pedir então que B_{k+1} satisfaça a equação secante dada por

$$B_{k+1}(x^{k+1} - x^k) = \nabla \ell(x^{k+1}, \lambda^{k+1}) - \nabla \ell(x^k, \lambda^{k+1}).$$

Uma maneira comum de gerar aproximações satisfazendo a equação acima é computar B_{k+1} fazendo uso de fórmulas que atualizam B_k a partir de informações disponíveis até o momento; isto é, fazemos

$$B_{k+1} = B_k + U_k,$$

onde U_k é uma matriz de posto um ou posto dois que depende de B_k, x^k, x^{k+1} e λ^{k+1} .

Uma dessas fórmulas é a atualização simétrica de posto dois PSB (Powell-symmetric-Broyden), dada por

$$B_{k+1} = B_k + \frac{1}{s^t s} \left[(y - B_k s) s^t + s (y - B_k s)^t \right] - \frac{(y - B_k s)^t}{(s^t s)^2} s s^t,$$

onde

$$s = x^{k+1} - x^k \quad e \quad y = \nabla \ell(x^{k+1}, \lambda^{k+1}) - \nabla \ell(x^k, \lambda^{k+1}).$$

Como estas são atualizações simétricas mas não definidas positivas, o sub-problema (4.2) pode não ter solução (ser ilimitado) e, além disso, as matrizes B_k não necessariamente irão convergir para uma matriz semi-definida positiva em $N(\nabla h(x^*)^T)$, como é o caso da verdadeira Hessiana do Lagrangiano. Desta forma, a menos que o ponto inicial seja escolhido próximo de uma solução, um algoritmo PQS usando a aproximação PSB para as matrizes B_k não possui boas propriedades de convergência. Porém, o custo para calcular estas atualizações é muito menor do que o custo de computar a hessiana verdadeira exigida pelo

método de Newton.

A atualização, também de posto dois, mais utilizada na prática é o método BFGS (Broyden-Fletcher-Goldfarb-Shanno) que possui a desejável característica de gerar matrizes definidas positivas sob certas condições. A fórmula de atualização é dada por

$$B_{k+1} = B_k + \frac{yy^t}{s^t y} - \frac{B_k s s^t B_k}{s^t B_k s},$$

onde s e y são os mesmos já explicitados acima. A melhor característica desta atualização é que se B_k é definida positiva e $y^T s > 0$, B_{k+1} herda a positividade de B_k . Este é um dos motivos pelos quais a atualização BFGS é largamente usada nas implementações algorítmicas dos métodos de PQS, pois o fato de B_k ser sempre definida positiva garante a existência de solução para o subproblema quadrático a cada iteração.

É claro que esta atualização também pode falhar. Se $\nabla^2 \ell(x^*, \lambda^*)$ é definida positiva somente no núcleo de $\nabla h(x^*)^T$, a sequência $\{B_k\}_{k \in \mathbb{N}}$ pode não se aproximar da Hessiana verdadeira, mesmo próximo da solução. Além disso é possível deparar-se com $y^T s \leq 0$, o que deixa de garantir a positividade de B_{k+1} . Uma saída para este último problema é tomar, por exemplo, $B_{k+1} = B_k$ na iteração $k + 1$. Sobre a atualização PSB ver [10] e sobre a atualização BFGS ver [29].

Uma abordagem diferente trata de aproximar apenas uma parte da Hessiana do Lagrangiano por uma matriz definida positiva, a parte referente ao subespaço nulo de $\nabla h(x^*)^T$. A vantagem dos métodos que utilizam aproximações reduzidas para a hessiana é que as atualizações definidas positivas, como BFGS, podem ser usadas sem qualquer empecilho e a dimensão do problema é reduzida, como veremos a seguir.

Vamos supor que $\nabla h(x^k)$ tem posto completo para todo $k \in \mathbb{N}$. Definamos por Z_k e Y_k as matrizes cujas colunas formam uma base, respectivamente, para $N(\nabla h(x^k)^T)$ e $\text{Im}(\nabla h(x^k))$. Sem perda de generalidade, vamos supor também que as colunas de Z_k são ortogonais.

Definição 4.1. *Seja x^k a solução aproximada na iteração k , com multiplicador associado λ^k . A matriz*

$$Z_k^T \nabla^2 \ell(x^k, \lambda^k) Z_k$$

é chamada a hessiana reduzida para a função Lagrangiana em (x^k, λ^k) [6].

A matriz Z_k depende da base escolhida para $N(\nabla h(x^k)^T)$ e, portanto, a hessiana reduzida não é única. Se o par (x^k, λ^k) estiver suficientemente próximo da solução verdadeira (x^*, λ^*) , teremos que a hessiana reduzida em (x^k, λ^k) é definida positiva.

Podemos decompor a direção d^k como

$$d^k = Z_k u + Y_k v, \quad u \in \mathbb{R}^{n-m}, \quad v \in \mathbb{R}^m \quad (4.9)$$

e substituindo na restrição linear do subproblema quadrático temos:

$$\begin{aligned} h(x^k) + \nabla h(x^k)^T d^k &= 0 \\ \nabla h(x^k)^T (Z_k u + Y_k v) &= -h(x^k) \\ \nabla h(x^k)^T Y_k v &= -h(x^k). \end{aligned}$$

Como, por hipótese, $\nabla h(x^k)^T$ tem posto completo, $\nabla h(x^k)^T Y_k$ é não-singular e podemos concluir que

$$v = -[\nabla h(x^k)^T Y_k]^{-1} h(x^k).$$

Agora, substituindo (4.9) na função objetivo do subproblema quadrático obtemos um problema irrestrito com $n - m$ variáveis

$$\min_u \frac{1}{2} u^T Z_k^T B_k Z_k u + (\nabla f(x^k)^T + v^T B_k) Z_k u.$$

A matriz $Z_k^T B_k Z_k$ é uma aproximação para a hessiana reduzida da função Lagrangiana e, em vez de atualizar B_k para depois computar $Z_k^T B_k Z_k$, a própria hessiana reduzida é atualizada. Denotemos $R_k = Z_k^T B_k Z_k$. Como R_k é definida positiva, esta matriz pode ser atualizada a cada iteração pela fórmula BFGS onde

$$s = Z_k^T (x^{k+1} - x^k) = Z_k^T Z_k u$$

e

$$y = Z_k^T [\nabla \ell(x^k + Z_k u, \lambda^k) - \nabla \ell(x^k, \lambda^k)],$$

o que se explica pelo fato de somente a hessiana reduzida estar sendo aproximada [6].

Os multiplicadores não são estimados por este método pois o problema resolvido é irrestrito. Na próxima seção veremos que uma estimativa possível é dada pela solução de quadrados mínimos para as condições de primeira ordem do problema (4.1). No entanto, é suficiente (ver [6], pp.24) que os multiplicadores sejam tais que satisfaçam

$$\|\lambda^k - \lambda^*\| = O(\|x^k - x^*\|).$$

4.3 Função de mérito e convergência global

Como vimos até aqui, um algoritmo de programação quadrática sequencial em cada iteração, partindo de uma aproximação x^k , resolve um subproblema quadrático que retorna uma direção d^k a partir da qual obtemos uma nova aproximação x^{k+1} . Se o método de PQS funcionar bem, é de se esperar que esta nova aproximação seja melhor que a anterior. Mas que características uma aproximação “melhor” deve possuir?

Podemos pensar, a princípio, que uma aproximação melhor é uma aproximação que dista menos de um minimizador do que a anterior; ou seja, está mais próxima de uma solução verdadeira. De uma iteração para outra a aproximação pode ser melhorada de duas maneiras: a primeira é chegar mais perto do conjunto viável, já que os métodos de PQS clássicos ca-minham por pontos infactíveis; a segunda é melhorar o valor da função objetivo.

A melhor situação possível é quando as duas coisas acontecem ao mesmo tempo: melhoramos a factibilidade e a otimalidade da aproximação. Mas se conseguirmos melhorar somente uma ou outra, ainda assim podemos ter em mãos uma aproximação melhor que a anterior.

Uma *função de mérito* Φ é introduzida em algoritmos de PQS justamente para desempenhar o papel de decidir se x^{k+1} é ou não melhor que x^k . Devemos construir uma função Φ que apresente um decréscimo considerável toda vez que nos aproximarmos de um minimizador do problema original. Desta forma, uma função de mérito deve ser construída de modo que os minimizadores do problema

com restrições sejam também minimizadores de Φ .

É fundamental também que a direção d^k que resolve o subproblema na iteração k seja de descida para Φ , pois desta maneira é sempre possível encontrar $t_k \in \mathbb{R}$ tal que $\Phi(x^k + t_k d^k) < \Phi(x^k)$. O cálculo desse tamanho de passo pode ser feito, por exemplo, por um procedimento de “*backtracking*”.

Se uma função de mérito busca equilibrar o ganho em factibilidade e otimalidade, é razoável pensar que, para pontos factíveis, apenas o ganho em otimalidade deve ser medido. Isto é, se estamos dentro do conjunto viável a única coisa que importa para decidir se uma aproximação melhorou ou não é o decréscimo na função objetivo do problema.

A função de mérito mais conhecida é o Lagrangiano aumentado:

$$\Phi_L(x, \rho) = f(x) + \lambda(x)^T h(x) + \frac{\rho}{2} \|h(x)\|_2^2, \quad (4.10)$$

onde ρ é uma constante a ser determinada e $\lambda(x)$ é uma estimativa dos multiplicadores de Lagrange associados a um minimizador x^* .

Outra importante função de mérito é a função de penalidade L_1

$$\Phi_1(x, \rho) = f(x) + \rho \|h(x)\|_1 = f(x) + \rho \sum_{i=1}^m |h_i(x)|, \quad (4.11)$$

onde ρ também é uma constante positiva a ser escolhida.

Ambas as funções têm a desejável característica de serem funções de penalização exata para escolhas adequadas dos multiplicadores de Lagrange. Em outras palavras, existe $\bar{\rho} > 0$ tal que toda solução do problema original é solução do problema

$$\min \Phi(x, \rho)$$

para todo $\rho > \bar{\rho}$, onde $\Phi(x, \rho)$ é uma das funções de mérito acima citadas.

A vantagem que o Lagrangiano aumentado possui sobre a penalização L_1 é que $\Phi_L(x, \rho)$ é diferenciável, ao contrário de $\Phi_1(x, \rho)$. Por outro lado, $\Phi_1(x, \rho)$ não requer nenhuma estimativa dos multiplicadores de Lagrange. É importante ressaltar que o parâmetro ρ é fixo nas funções de mérito e não funciona como um

parâmetro de penalização. Mas o fato de $\Phi_L(x, \rho)$ e $\Phi_1(x, \rho)$ serem funções de penalidade exata irá garantir a existência de um certo valor para ρ , necessário para a análise de convergência do método de PQS.

Nosso objetivo, a partir de agora, é fornecer uma prova de convergência global de um método de PQS clássico para um ponto estacionário da função de mérito, por ora representada pelo Lagrangiano aumentado. Faremos as seguintes hipóteses:

H1 : As matrizes B_k são uniformemente definidas positivas no núcleo de $\nabla h(x^k)^T$, isto é, existe uma constante $\beta_1 > 0$ tal que para todo $k \in \mathbb{N}$,

$$d^T B_k d \geq \beta_1 \|d\|_2^2$$

para toda d tal que $\nabla h(x^k)^T d = 0$.

H2 : A sequência (B_k) está uniformemente limitada, isto é, existe uma constante $\beta_2 > 0$ tal que para todo $k \in \mathbb{N}$,

$$\|B_k\|_2 \leq \beta_2.$$

H3 : As colunas de $\nabla h(x^k)$ são linearmente independentes para todo $k \in \mathbb{N}$.

Vamos definir

$$Q(x) = \nabla h(x) [\nabla h(x)^T \nabla h(x)]^{-1} \nabla h(x)^T \quad (4.12)$$

o projetor sobre $\text{Im}(\nabla h(x))$ e

$$P(x) = I - Q(x) \quad (4.13)$$

o projetor sobre $\text{N}(\nabla h(x)^T)$.

Considerando que o multiplicador λ^* associado a um minimizador x^* é tal que

$$\nabla f(x^*) + \nabla h(x^*) \lambda^* = 0, \text{ ou seja,}$$

$$\lambda^* = - [\nabla h(x^*)^T \nabla h(x^*)]^{-1} \nabla h(x^*)^T \nabla f(x^*),$$

fica fácil entender porque uma boa estimativa para os multiplicadores $\lambda(x)$ é dada por

$$\lambda(x) = - [\nabla h(x)^T \nabla h(x)]^{-1} \nabla h(x)^T \nabla f(x).$$

Calculando o gradiente da função de mérito $\Phi_L(x, \rho)$ obtemos,

$$\nabla \Phi_L(x, \rho) = \nabla f(x) + \nabla h(x) \lambda(x) + \nabla \lambda(x) h(x) + \rho \nabla h(x) h(x), \quad (4.14)$$

onde

$$\nabla \lambda(x) = -\nabla^2 L(x, \lambda(x)) \nabla h(x) [\nabla h(x)^T \nabla h(x)]^{-1} + R(x), \quad (4.15)$$

sendo $R(x)$ uma função tal que $R(x^*) = 0$ se x^* satisfaz as condições necessárias de primeira ordem do problema (4.1).

O lema abaixo apresenta um importante resultado que será usado posteriormente na prova de um teorema.

Lema 4.1. *Assuma que as hipóteses **H1-H3** estão satisfeitas. Então $d^k = 0$ resolve o subproblema quadrático (4.2) na iteração k se e somente se x^k é ponto estacionário do problema (4.1).*

Demonstração:

Suponhamos que $d^k = 0$ é solução de (4.2). Então, pelas condições necessárias do subproblema temos que existe $\hat{\lambda} \in \mathbb{R}^m$ tal que

$$\begin{aligned} B_k d^k + \nabla f(x^k) + \nabla h(x^k) \hat{\lambda} &= 0, \\ \nabla h(x^k)^T d^k + h(x^k) &= 0. \end{aligned}$$

Mas como supomos $d^k = 0$, as equações acima mostram que x^k é ponto estacionário do problema (4.1) com multiplicador $\hat{\lambda}$.

Suponhamos agora que x^k é ponto estacionário de (4.1) com multiplicador associado λ^k . Então,

$$\nabla f(x^k) + \nabla h(x^k) \lambda^k = 0, \quad (4.16)$$

$$h(x^k) = 0. \quad (4.17)$$

Seja $(d^k, \overline{\lambda^k})$ um par que resolve o subproblema quadrático. Então temos que

$$B_k d^k + \nabla f(x^k) + \nabla h(x^k) \bar{\lambda}^k = 0, \quad (4.18)$$

$$\nabla h(x^k)^T d^k + h(x^k) = 0. \quad (4.19)$$

Subtraindo (4.16) de (4.18) e substituindo (4.17) em (4.19) ficamos com:

$$\begin{aligned} B_k d^k + \nabla h(x^k)(\bar{\lambda}^k - \lambda^k) &= 0, \\ \nabla h(x^k)^T d^k &= 0. \end{aligned}$$

Como $(d^k)^T B_k d^k + \underbrace{(d^k)^T \nabla h(x^k)}_0 (\bar{\lambda}^k - \lambda^k) = 0$, concluímos que $d^k = 0$. Além disso, da igualdade $\nabla h(x^k)(\bar{\lambda}^k - \lambda^k) = 0$ e da hipótese **H3** concluímos também que $\bar{\lambda}^k = \lambda^k$. ■

Partindo das considerações e cálculos expostos até aqui, estamos aptos a dar início à prova do seguinte teorema:

Teorema 4.3. *Assuma que as hipóteses **H1-H3** são satisfeitas. Então, para algum valor finito do parâmetro ρ verifica-se que:*

(i) *Se x^* satisfaz CS para o problema (4.1), então x^* satisfaz CS para o problema irrestrito $\min \Phi_L(x, \rho)$;*

(ii) *Se na iteração k do método de PQS x^k não é um ponto KKT de (4.1), então d^k é uma direção de descida para Φ_L .*

Demonstração:

(i) Se x^* satisfaz CS para (4.1), em particular x^* satisfaz as condições necessárias de primeira ordem para este problema. Logo existe $\lambda^* \in \mathbb{R}^m$ tal que

$$\begin{aligned} \nabla f(x^*) + \nabla h(x^*) \lambda^* &= 0, \\ h(x^*) &= 0. \end{aligned}$$

Então temos $\lambda(x^*) = \lambda^*$ e para qualquer $\rho \in \mathbb{R}$

$$\nabla \Phi_L(x^*, \rho) = \underbrace{\nabla f(x^*) + \nabla h(x^*) \lambda^*}_0 + [\nabla \lambda(x^*) + \rho \nabla h(x^*)] \underbrace{h(x^*)}_0 = 0$$

Portanto x^* é ponto estacionário de $\Phi_L(x, \rho)$. Agora vamos mostrar que $\nabla^2 \Phi_L(x^*, \rho)$ é definida positiva para algum valor finito de ρ .

Temos que

$$\begin{aligned} \nabla^2 \Phi_L(x, \rho) &= \nabla^2 f(x) + \sum_{i=1}^m \nabla^2 h_i(x) \lambda_i(x) + \nabla h(x) \nabla \lambda(x)^T \\ &\quad + \sum_{i=1}^m \nabla^2 \lambda_i(x) h_i(x) + \nabla \lambda(x) \nabla h(x)^T \\ &\quad + \rho \left[\sum_{i=1}^m \nabla^2 h_i(x) h_i(x) + \nabla h(x) \nabla h(x)^T \right]. \end{aligned} \quad (4.20)$$

Computando (4.20) em x^* , considerando que $h(x^*) = 0$ e utilizando (4.12) e (4.15) ficamos com

$$\nabla^2 \Phi_L(x^*, \rho) = \nabla^2 \ell^* - Q^* \nabla^2 \ell^* - \nabla^2 \ell^* Q^* + \rho \nabla h(x^*) \nabla h(x^*)^T,$$

onde $\nabla^2 \ell^*$ e Q^* são, respectivamente, $\nabla^2 \ell(x, \rho)$ e $Q(x)$ computados em x^* .

Seja $0 \neq y \in \mathbb{R}^n$. Então,

$$y^T \nabla^2 \Phi_L(x^*, \rho) y = y^T \nabla^2 \ell^* y - 2y^T \nabla^2 \ell^* Q^* y + \rho \|\nabla h(x^*)^T y\|_2^2. \quad (4.21)$$

Vamos mostrar que existe $\bar{\rho} > 0$ tal que $\nabla^2 \Phi_L(x^*, \rho)$ é definida positiva para todo $\rho \geq \bar{\rho}$.

Suponhamos que tal afirmação não seja verdadeira. Então, para todo $k \in \mathbb{N}$, existe $0 < (\rho_k) \rightarrow \infty$ e $y^k \in \mathbb{R}^n$ com $\|y^k\|_2 = 1$ tal que

$$(y^k)^T \nabla^2 \Phi_L(x^*, \rho_k) y^k \leq 0. \quad (4.22)$$

De (4.22) e (4.21) segue que

$$(y^k)^T \nabla^2 \ell^* y^k - 2(y^k)^T \nabla^2 \ell^* Q^* y^k + \rho_k \|\nabla h(x^*)^T y^k\|_2^2 \leq 0 \quad \text{para todo } k \in \mathbb{N}.$$

Dado que a sequência $(y^k)_{k \in \mathbb{N}}$ pertence à bola unitária de $\mathbb{R}^n$, um conjunto fechado e limitado, garantimos a existência de uma subsequência $(y^k)_{k \in \mathbb{K}}$, $\mathbb{K} \subset \mathbb{N}$, convergindo para um vetor $\bar{y}$ com $\|\bar{y}\|_2^2 = 1$. Daí, tomando o limite em $\mathbb{K}$ temos

$$(\bar{y})^T \nabla^2 \ell^* \bar{y} - 2(\bar{y})^T \nabla^2 \ell^* Q^* \bar{y} + \lim_{k \in \mathbb{K}} \rho_k \|\nabla h(x^*)^T y^k\|_2^2 \leq 0. \quad (4.23)$$

Como $\|\nabla h(x^*)^T y^k\|_2^2 \geq 0$ para todo $k \in \mathbb{K}$ e $\rho_k \rightarrow \infty$, (4.23) implica que

$$\lim_{k \rightarrow \infty} \|\nabla h(x^*)^T y^k\|_2^2 = 0.$$

Logo $\nabla h(x^*)^T \bar{y} = 0$, isto é, $\bar{y} \in N(\nabla h(x^*)^T)$. Então temos que $Q^* \bar{y} = 0$, resultando em

$$(\bar{y})^T \nabla^2 \ell^* \bar{y} \leq 0,$$

o que contraria a hipótese de que x^* satisfaz CS para o problema (4.1). Portanto existe $\bar{\rho} > 0$ finito tal que $\nabla^2 \Phi_L(x^*, \rho) > 0$ para todo $\rho \geq \bar{\rho}$ e com isto provamos (i).

(ii) Suponhamos que na iteração k do método PQS, x^k não é um ponto estacionário de (4.1). Então pelo Lema (4.1) temos que $d^k \neq 0$ e existe $\bar{\lambda}^k \in \mathbb{R}^m$ tal que

$$\begin{aligned} B_k d^k + \nabla f(x^k) + \nabla h(x^k) \bar{\lambda}^k &= 0, \\ \nabla h(x^k)^T d^k + h(x^k) &= 0. \end{aligned}$$

De (4.14) segue que

$$\begin{aligned} (d^k)^T \nabla \Phi_L(x^k, \rho) &= (d^k)^T \nabla f(x^k) + (d^k)^T \nabla h(x^k) \lambda(x^k) \\ &\quad + (d^k)^T \nabla \lambda(x^k) h(x^k) + \rho (d^k)^T \nabla h(x^k) h(x^k). \end{aligned} \quad (4.24)$$

Como $\nabla h(x^k)^T d^k = -h(x^k)$, substituindo em (4.24) ficamos com

$$\begin{aligned} (d^k)^T \nabla \Phi_L(x^k, \rho) &= (d^k)^T \nabla f(x^k) - h(x^k) \lambda(x^k) \\ &\quad + (d^k)^T \nabla \lambda(x^k) h(x^k) - \rho \|h(x^k)\|_2^2. \end{aligned} \quad (4.25)$$

Analisemos dois casos:

Caso 1: Se x^k não é factível para o problema (4.1), $h(x^k) \neq 0$ e temos que para ρ suficientemente grande $(d^k)^T \nabla \Phi_L(x^k, \rho) < 0$.

Caso 2: Se $h(x^k) = 0$ temos que

$$(d^k)^T \nabla \Phi_F(x^k, \rho) = (d^k)^T \nabla f(x^k).$$

Como $d = 0$ é uma direção factível para o subproblema com valor nulo de função objetivo, então a solução d^k é tal que

$$\frac{1}{2}(d^k)^T B_k d^k + \nabla f(x^k)^T d^k \leq 0.$$

Mas $h(x^k) = 0$ implica que $d^k \in N(\nabla h(x^k)^T)$. Daí, por hipótese, $(d^k)^T B_k d^k > 0$ e, obrigatoriamente, devemos ter $\nabla f(x^k)^T d^k < 0$. ■

Em [6] Boggs fez uso de uma hipótese adicional para provar o mesmo resultado. Ele supôs, juntamente com as três hipóteses que também consideramos aqui, que a seqüência $(x^k)_{k \in \mathbb{N}}$ gerada por um algoritmo baseado no método de PQS está contida em um conjunto compacto. Sua demonstração está baseada fortemente na idéia de decompor um vetor qualquer como soma de projeções em subespaços ortogonais. Adotamos, no entanto, uma estratégia diferente que segue a mesma linha de raciocínio presente na prova de limitação do parâmetro de penalização do Lagrangiano aumentado encontrada em [4], uma vez que esta foi a função de mérito considerada, e provamos o mesmo resultado sem a hipótese de que os iterandos estão contidos em um conjunto compacto.

Na verdade, quando $d^k \in N(\nabla h(x^k)^T)$, temos pela hipótese **H1** que

$$\frac{(d^k)^T B_k d^k}{\|d^k\|_2} \geq \beta_1 > 0.$$

Logo,

$$\frac{(d^k)^T \nabla \Phi_L(x^k, \rho)}{\|d^k\|_2} = \frac{(d^k)^T \nabla f(x^k)}{\|d^k\|_2} \leq -\beta_1 < 0.$$

Nos casos em que $d^k \notin N(\nabla h(x^k)^T)$, temos $h(x^k) \neq 0$ e por (4.25) é possível escolher ρ finito de maneira que

$$\frac{(d^k)^T \nabla \Phi_L(x^k, \rho)}{\|d^k\|_2} \leq -\beta_3 < 0$$

para alguma constante positiva β_3 .

Com isto garantimos a existência de uma constante $\beta_4 > 0$ tal que

$$\cos \theta_k = \frac{-(d^k)^T \nabla \Phi_L(x^k, \rho)}{\|d^k\|_2 \|\nabla \Phi_L(x^k, \rho)\|_2} \geq \beta_4 > 0, \quad (4.26)$$

onde θ_k é o ângulo entre d^k e $-\nabla \Phi_L(x^k, \rho)$.

Para garantir decréscimo suficiente da função de mérito vamos pedir que o tamanho de passo t_k dado na direção d^k satisfaça as condições de Wolfe [6]:

$$\Phi_L(x^k + t_k d^k, \rho) \leq \Phi_L(x^k, \rho) + \sigma_1 t_k \nabla \Phi_L(x^k, \rho)^T d^k \quad (4.27)$$

$$\nabla \Phi_L(x^k + t_k d^k, \rho)^T d^k \geq \sigma_2 \nabla \Phi_L(x^k, \rho)^T d^k, \quad (4.28)$$

onde $0 < \sigma_1 < \sigma_2 < 1$. A desigualdade (4.28) tem o papel de garantir que o passo t_k não seja muito pequeno.

Teorema 4.4. *Assuma que as hipóteses **H1-H3** estão satisfeitas e que a cada iteração ρ é escolhido de maneira que (4.26) se cumpra. Se a seqüência $(x^k)_{k \in \mathbb{N}}$ gerada pelo método de PQS com tamanho de passo satisfazendo as condições de Wolfe convergir para x^* , então este é um ponto estacionário de Φ_L .*

Demonstração:

De (4.26) e (4.27) temos que a desigualdade

$$\Phi_L(x^{k+1}, \rho) \leq \Phi_L(x^k, \rho) - \beta_5 \cos^2 \theta_k \|\nabla \Phi_L(x^k, \rho)\|_2^2$$

é satisfeita para alguma constante positiva β_5 . Dado um valor finito $n \in \mathbb{N}$ temos

$$\Phi_L(x^{n+1}, \rho) - \Phi_L(x^0, \rho) \leq -\beta_5 \sum_{k=0}^n \cos^2 \theta_k \|\nabla \Phi_L(x^k, \rho)\|_2^2.$$

Agora, fazendo n tender ao infinito ficamos com

$$\Phi_L(x^0, \rho) - \Phi_L(x^*, \rho) \geq \beta_5 \sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla \Phi_L(x^k, \rho)\|_2^2,$$

o que implica na convergência da série em questão. Logo, o termo geral da mesma converge para zero, ou seja

$$\cos^2 \theta_k \|\nabla \Phi_L(x^k, \rho)\|_2^2 \rightarrow 0.$$

Como $\cos \theta_k$ é mantido longe do zero por hipótese, então

$$\lim_{k \rightarrow \infty} \nabla \Phi_L(x^k, \rho) = 0.$$

Pela continuidade das funções envolvidas concluímos que x^* é um ponto estacionário de Φ_L . ■

Os resultados acima mostram que nos casos em que os iterandos x^k do método de PQS geram uma sequência convergente, o ponto limite de tal sequência é ponto estacionário da função de mérito $\Phi_L(x, \rho)$ para algum valor finito do parâmetro ρ . Uma boa questão a ser analisada é se o mesmo resultado do Teorema 4.4 vale para qualquer ponto de acumulação da sequência $(x^k)_{k \in \mathbb{N}}$ nos casos em que x^k não for convergente. A garantia de que a convergência da sequência de fato ocorre é uma questão que depende do problema de minimização; se temos uma região viável onde a função objetivo é ilimitada, não podemos esperar que $(x^k)_{k \in \mathbb{N}}$ convirja para algum ponto. Muitos autores admitem a hipótese de que os iterandos estão contidos em um conjunto compacto, o que resulta na existência de pelo menos uma subsequência convergente. Nos dois teoremas diretamente anteriores não trabalhamos com esta hipótese justamente para não desconsiderar os casos em que o minimizador procurado pode não existir. É claro que quando a convergência ocorre, os iterandos estão contidos em um compacto e todos os resultados que têm a compacidade como hipótese continuam válidos.

4.4 Efeito Maratos e busca linear não-monótona

Na resolução de problemas de otimização com restrições via programação quadrática sequencial é comum a ocorrência do conhecido *efeito Maratos*. Este caracteriza-se pela impossibilidade de dar um passo completo, isto é, passo unitário, mesmo estando próximo a uma solução. Quando tal efeito ocorre, o método perde velocidade de convergência podendo, até mesmo, estacionar em um ponto que não é KKT. Vale ressaltar que a necessidade de truncar o passo ou, equivalentemente,

realizar uma busca linear, dá-se pela exigência de decrescer o valor da função objetivo (ou da função de mérito) a cada iteração.

Uma técnica bem sucedida na prevenção da ocorrência do efeito Maratos é a realização de uma busca linear não-monótona, o que significa que o valor da função objetivo não precisa, necessariamente, decrescer a cada iteração. Tal técnica foi primeiramente proposta por Grippo, Lampariello e Lucidi em [16], trabalho no qual os autores mostraram que uma busca linear não-monótona contorna várias dificuldades com as quais podemos nos deparar quando utilizamos o método de Newton e ainda mostraram que a convergência global deste método se mantém com esse tipo de busca. Mais especificamente, uma busca linear não-monótona impõe que o valor da função objetivo no novo iterando deve satisfazer o critério de Armijo com relação ao máximo valor de função nos últimos M iterandos, onde M é um número pré-fixado. Matematicamente falando, devemos determinar $t_k \in (0, 1)$ tal que

$$f(x^k + t_k d^k) \leq \max_{0 \leq i \leq M} f(x_{k-i}).$$

Mayne e Polak em [28] propuseram, para contornar o efeito Maratos, curvar a direção de busca. Uma direção de correção $\tilde{d}^k$ é determinada e a busca, agora não mais linear, é realizada no arco

$$x^k + t_k d^k + t_k^2 \tilde{d}^k,$$

onde t_k é um valor pertencente ao intervalo $(0, 1)$ tal que

$$f(x^k + t_k d^k + t_k^2 \tilde{d}^k) < f(x^k) + t_k \nabla f(x^k)^T d^k.$$

Esta é uma técnica que favorece a aceitação do passo unitário com maior frequência do que na busca linear monótona, porém requer avaliações de função adicionais em cada iteração.

Sendo assim, Panier e Tits em [31] uniram estas duas estratégias, pedindo que o passo $t_k \in (0, 1)$ satisfaça

$$f(x^k + t_k d^k + t_k^2 \tilde{d}^k) \leq \max_{0 \leq i \leq M} \{f(x^{k-i})\} - \alpha t_k (d_0^k)^T B_k d_0^k, \quad (4.29)$$

onde $\alpha \in (0, 1)$, B_k é uma estimativa para a Hessiana do Lagrangiano em x^k e d_0^k é a direção que resolve o subproblema quadrático na iteração k . Um estudo visando analisar o comportamento de algoritmos que utilizam busca linear monótona e

não-monótona, conforme (4.29), será realizado nos próximos capítulos. Os dois tipos de busca serão comparadas em algoritmos de programação quadrática sequencial e a ocorrência ou não do efeito Maratos poderá ser detectada e analisada.

Uma outra técnica conhecida na literatura que também combate o efeito Maratos é a chamada técnica *watchdog* [9], proposta juntamente com métodos de programação quadrática sequencial para problemas com restrições. Seu propósito também é forçar que o passo unitário seja aceito, mas há uma limitação que pode resultar em repetidos *backtrackings* nas primeiras iterações. A busca linear não-monótona tem ainda aplicação contra o efeito Maratos no contexto de problemas minimax restritos e irrestritos, conforme trabalho de Zhou e Tits [41].

Capítulo 5

Métodos de PQS e condições de qualificação

Retomando o que foi exposto no terceiro capítulo, o método de PQS resolve a cada iteração um subproblema quadrático que aproxima bem o problema original buscando a solução de um sistema linear que consiste nas condições KKT do subproblema. À medida que conseguimos iterandos melhores é intuitivo que os pontos x^k se tornem “quase KKT” para o problema original. Sob que condições uma seqüência de pontos “quase KKT” converge para um ponto KKT de fato? Esta é uma das questões que vamos responder ao longo deste quinto capítulo.

Outro ponto relevante a ser considerado é o uso da hipótese, nos resultados do capítulo anterior, de que os gradientes das restrições ativas são linearmente independentes, o que garante a unicidade dos multiplicadores de Lagrange. Seria possível alcançar os mesmos resultados de convergência relaxando esta hipótese? Em outras palavras, as boas propriedades do método PQS continuam válidas se substituirmos a LICQ por uma condição de qualificação mais fraca? Os resultados a seguir têm por base os estudos realizados por Qi e Wei em [34] e algumas modificações serão apresentadas e comentadas ao final da Seção 5.2.

5.1 Sequências de pontos KKT aproximados e a CPLD

Vamos começar definindo o que vem a ser, rigorosamente, uma sequência de pontos “quase KKT”. Agora vamos voltar a pensar num problema de minimização com restrições de igualdade e desigualdade, no formato (2.1).

Definição 5.1. Dizemos que $(x^k)_{k \in \mathbb{N}} \subset \mathbb{R}^n$ é uma sequência de pontos KKT aproximados de (2.1) se existe uma sequência $\{(\lambda^k, \mu^k, \epsilon^k, \delta^k, \eta^k)\} \subset \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}$, $\delta^k \geq 0$, $\eta^k \geq 0$, tal que para cada k é satisfeito:

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) &= \epsilon^k, \\ \|h(x^k)\| &\leq \eta^k, \\ g(x^k) &\leq \delta^k, \\ \mu^k &\geq 0, \\ (\mu^k)^T (g(x^k) - \delta^k) &= 0. \end{aligned} \tag{5.1}$$

e $\{(\epsilon^k, \delta^k, \eta^k)\}_{k \in \mathbb{N}}$ converge para zero.

Sequências de pontos KKT aproximados são geradas pelos métodos de PQS, métodos de equações KKT e alguns outros métodos destinados a resolver o problema (2.1). Qi e Wei mostraram em [34] que se uma sequência deste tipo converge para um ponto que satisfaz a condição de dependência linear positiva constante introduzida no Capítulo 3, então este é um ponto KKT para o problema original (2.1). Vejamos o teorema a seguir:

Teorema 5.1. Se uma sequência de pontos KKT aproximados $(x^k)_{k \in \mathbb{N}}$ possui um ponto limite x^* no qual a CPLD é satisfeita, então x^* é um ponto KKT de (2.1).

Demonstração:

Se x^* é um ponto limite de $(x^k)_{k \in \mathbb{N}}$ existe um subconjunto infinito dos naturais $\mathbb{K}$ tal que $x^k \rightarrow x^*$ para $k \in \mathbb{K}$. Passando o limite em $\|h(x^k)\| \leq \eta^k$ e $g(x^k) \leq \delta^k$ verifica-se que $h(x^*) = 0$ e $g(x^*) \leq 0$, o que garante a factibilidade de x^* para o problema (2.1).

A teoria de álgebra linear nos permite, para cada $k \in \mathbb{K}$, escolher multiplicadores (λ^k, μ^k) que satisfazem (5.1) e tais que os vetores

$$\nabla h_i(x^k), i \in \text{supp}(\lambda^k) \text{ e } \nabla g_j(x^k), j \in \text{supp}(\mu^k)$$

sejam linearmente independentes. Esta será a escolha feita para todo $k \in \mathbb{K}$. Denotemos $I_k = \text{supp}(\lambda^k)$ e $J_k = \text{supp}(\mu^k)$.

Vamos considerar dois casos:

Caso 1: Se $\{(\lambda^k, \mu^k)\}_{k \in \mathbb{K}}$ admite uma subsequência convergente, digamos para $k \in \mathbb{K}_1 \subset \mathbb{K}$, então existem $\lambda^* \in \mathbb{R}^m$ e $\mu^* \in \mathbb{R}_+^p$ tais que $\lambda^k \rightarrow \lambda^*$ e $\mu^k \rightarrow \mu^*$ quando $k \rightarrow \infty$.

Passando o limite na equação

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) = \epsilon^k$$

ficamos com

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

Além disso, fazendo $k \rightarrow \infty$ também em

$$(\mu^k)^T (g(x^k) - \delta^k) = 0,$$

verificamos que se cumpre a complementaridade $(\mu^*)^T g(x^*) = 0$.

Portanto (x^*, λ^*, μ^*) satisfaz as condições KKT do problema (2.1).

Caso 2: Se $\{(\lambda^k, \mu^k)\}_{k \in \mathbb{K}}$ não admite uma subsequência convergente então

$$\lim_{k \rightarrow \infty} \|(\lambda^k, \mu^k)\| = +\infty.$$

Mas a sequência $\{(\lambda^k, \mu^k) / \|(\lambda^k, \mu^k)\|\}_{k \in \mathbb{K}}$ está contida em um conjunto compacto e por isso admite uma subsequência convergente, digamos para $k \in \mathbb{K}_2 \subset \mathbb{K}$. Sendo assim, existem $\alpha \in \mathbb{R}^m$ e $\beta \in \mathbb{R}_+^p$, com $\|(\alpha, \beta)\| = 1$, tais que

$$\lim_{k \in \mathbb{K}_2} \frac{(\lambda^k, \mu^k)}{\|(\lambda^k, \mu^k)\|} = (\alpha, \beta).$$

Além disso, existe $k_1 \in \mathbb{K}_2$ tal que para todo $k \geq k_1$ verifica-se que $\text{supp}(\alpha) \subset I_k$ e $\text{supp}(\beta) \subset J_k$.

Dividindo ambos os lados da equação

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) = \epsilon^k$$

por $\|(\lambda^k, \mu^k)\|$ e fazendo k tender ao infinito com $k \in \mathbb{K}_2$ obtemos

$$\sum_{i \in \text{supp}(\alpha)} \alpha_i \nabla h_i(x^*) + \sum_{j \in \text{supp}(\beta)} \beta_j \nabla g_j(x^*) = 0$$

o que equivale a dizer que

$$\sum_{i \in I_k} \alpha_i \nabla h_i(x^*) + \sum_{j \in J_k} \beta_j \nabla g_j(x^*) = 0 \quad , \quad \forall k \geq k_1, k \in \mathbb{K}_2.$$

Isto nos mostra que os vetores $\nabla h_i(x^*)$, $i \in I_k$ e $\nabla g_j(x^*)$, $j \in J_k$, são positivo linearmente dependentes $\forall k \geq k_1$, $k \in \mathbb{K}_2$. Como x^* satisfaz a CPLD por hipótese, existe uma vizinhança de x^* na qual os gradientes continuam linearmente dependentes. Isto é, existe $\epsilon > 0$ tal que para todo $k \in \mathbb{K}_2$, $k \geq k_1$ e satisfazendo $\|x^* - x^k\| < \epsilon$, os gradientes acima são linearmente dependentes. Isto, no entanto, contradiz o que foi considerado no início da demonstração deste teorema, a independência linear de tais gradientes para todo $k \in \mathbb{K}$.

Portanto a única possibilidade é que $\{(\lambda^k, \mu^k)\}_{k \in \mathbb{K}}$ assumam uma subsequência convergente, de onde segue que x^* é um ponto KKT para o problema (2.1). ■

Como o método de PQS gera uma sequência de pontos KKT aproximados, com este resultado temos então que qualquer ponto limite desta sequência, no qual a CPLD se cumpra, é um ponto KKT do problema original. Não podemos garantir, entretanto, que tal ponto limite seja um minimizador de (2.1), e mesmo nos casos em que isto for verdade não há garantia de que ele seja um minimizador estrito.

No Capítulo 4 definimos as condições suficientes considerando que a regularidade é satisfeita pelo minimizador. Neste caso, quando o ponto é regular,

os multiplicadores de Lagrange são unicamente determinados. A partir do momento em que deixamos de exigir que a regularidade seja satisfeita e passamos a trabalhar com condições de qualificação mais fracas, os pontos estacionários satisfazem as condições KKT com infinitos multiplicadores. Esta é a motivação para a condição suficiente que vamos definir agora. A SSOSC - *strong second-order sufficiency conditions* - foi introduzida por Robinson [35] em 1982.

A função *Lagrangiana* associada a um problema com restrições de igualdades e desigualdades é dada por

$$L(x, \lambda, \mu) = f(x) + h(x)^T \lambda + g(x)^T \mu.$$

Dizemos que uma tripla (x, λ, μ) satisfaz a SSOSC se satisfaz as condições KKT (2.2) e a matriz $\nabla_{xx}^2 L(x, \lambda, \mu)$ é definida positiva no subespaço

$$G(x, \lambda, \mu) = \{d \in \mathbb{R}^n \mid \nabla h_i(x)^T d = 0 \text{ para } i \in \{1, \dots, m\}, \\ \nabla g_j(x)^T d = 0 \text{ para } j \in \text{supp}(\mu)\}.$$

Definição 5.2. *Suponha que x seja um ponto KKT do problema (2.1). Se para todos os multiplicadores de Lagrange (λ, μ) de x a tripla (x, λ, μ) satisfaz a SSOSC, então dizemos que x satisfaz a SSOSC [35].*

Em [34] Qi e Wei demonstraram que se x^* é um ponto KKT de (2.1) que satisfaz a CPLD e a SSOSC, então x^* é um ponto KKT isolado de (2.1). Neste trabalho, os autores introduziram a condição CPLD e conjecturaram que ela poderia ser uma condição de qualificação, deixando tal questão em aberto. Em [36] Schuverdt explorou esta questão e provou que a CPLD é de fato uma CQ. Desta maneira o resultado citado no início deste parágrafo segue como consequência imediata das condições suficientes para um minimizador. De fato, se um ponto satisfaz SSOSC juntamente com alguma condição de qualificação, ele não somente é um ponto KKT isolado, mas um minimizador estrito do problema de otimização.

Schuverdt, no mesmo trabalho, também posicionou a CPLD em relação às demais condições de qualificação, mostrando ser essa uma CQ mais fraca que a regularidade, Mangasarian-Fromovitz e posto constante; ou seja, dado um certo problema de minimização é possível que existam minimizadores satisfazendo a CPLD mas não a regularidade, por exemplo. Isto nos capacita a buscar pontos estacionários em um universo maior, já que todo ponto estacionário “qualificado” deve obrigatoriamente satisfazer KKT.

Antes do próximo teorema, vamos descrever um método de PQS geral. O algoritmo aparece na Seção 4 de [34].

Algoritmo 5.1. *Seja $C > 0$.*

Dados: $x^0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n \times n}$ simétrica definida positiva.

Passo 0 (*Inicialização*): $k = 0$.

Passo 1 (*Cálculo de uma direção de busca*): *Compute d^k que resolve o sub-problema quadrático (QP)*

$$\begin{aligned} \min \quad & \frac{1}{2} d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a} \quad & \nabla h_i(x^k)^T d + h_i(x^k) = 0, \quad i \in \{1, \dots, m\}, \\ & \nabla g_j(x^k)^T d + g_j(x^k) \leq 0, \quad j \in \{1, \dots, p\}. \end{aligned}$$

Se $d^k = 0$, parar: x^k é uma solução.

Passo 2 (*Busca linear e correção adicional*): *Determine o tamanho de passo $t_k \in (0, 1)$ e uma direção de correção $\tilde{d}^k$ tal que*

$$\|\tilde{d}^k\| \leq C \|d^k\|.$$

Passo 3 (*Atualizações*): *Compute uma nova aproximação simétrica definida positiva B_{k+1} para a hessiana do Lagrangiano. Faça $x^{k+1} = x^k + t_k d^k + \tilde{d}^k$ e $k = k + 1$. Volte ao Passo 1.*

Cada método de PQS deve especificar as regras que irão determinar o cálculo de t_k , $\tilde{d}^k$ e B_{k+1} . O tamanho de passo deve ser tal que garanta decréscimo suficiente de uma dada função de mérito e a matriz B_{k+1} pode ser obtida, por exemplo, pela atualizações BFGS, conforme vimos no Capítulo 4. Nos métodos clássicos de PQS temos sempre $\tilde{d}^k = 0$.

Assumindo que em toda iteração o subproblema quadrático possui solução, o bom desempenho do Algoritmo (5.1) requer a análise de dois casos: quando ele pára no Passo 1 e quando gera uma sequência infinita.

Se o Algoritmo pára no Passo 1, seguindo um raciocínio análogo àquele realizado na primeira parte da prova do Lema 4.1 e tomando o cuidado de considerar as restrições de desigualdade ao escrever as condições KKT do subproblema quadrático acima, é fácil verificar que x^k é um ponto KKT do problema (2.1).

Para os casos em que o Algoritmo (5.1) gera uma sequência infinita temos o seguinte resultado:

Teorema 5.2. *Assuma que o Algoritmo (5.1) gera uma sequência infinita $(x^k)_{k \in \mathbb{N}}$ e que esta sequência possui um ponto de acumulação x^* . Digamos que $x^k \rightarrow x^*$ para $k \in \mathbb{K} \subset \mathbb{N}$. Suponha que x^* satisfaz a CPLD e que as estimativas para a Hessiana do Lagrangiano $(B_k)_{k \in \mathbb{N}}$ são limitadas, isto é, existe um escalar $C_1 > 0$ tal que para todo $k \in \mathbb{N}$ temos que*

$$\|B_k\| \leq C_1. \quad (5.2)$$

Se

$$\liminf_{k \in \mathbb{K}} \|d^k\| = 0, \quad (5.3)$$

então x^* é um ponto KKT de (2.1) [34].

Demonstração:

De (5.3) temos que existe uma subsequência $\mathbb{K}' \subset \mathbb{K}$ tal que

$$\lim_{k \in \mathbb{K}'} \|d^k\| = 0. \quad (5.4)$$

Das condições KKT do problema quadrático (QP) seguem as seguintes igualdades e desigualdades:

$$\begin{aligned} B_k d^k + \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) &= 0, \\ h(x^k) + \nabla h(x^k)^T d^k &= 0, \\ g(x^k) + \nabla g(x^k)^T d^k &\leq 0, \\ \mu^k &\geq 0, \\ (\mu^k)^T (g(x^k) - \nabla g(x^k)^T d^k) &= 0. \end{aligned}$$

Tomando o limite para $k \in \mathbb{K}'$, pela hipótese (5.2) e por (5.4), temos que

$$\epsilon_k \equiv -B_k d^k \rightarrow 0,$$

$$\delta_k \equiv -\nabla g(x^k)^T d^k \rightarrow 0,$$

e

$$\eta_k \equiv \|\nabla h(x^k)^T d^k\| \rightarrow 0.$$

Temos então uma sequência de pontos KKT aproximados e portanto, pelo Teorema (5.1), segue que x^* é um ponto KKT de (2.1). ■

Estes resultados mostram que a CPLD desempenha um importante papel na análise de convergência de métodos de programação quadrática sequencial de maneira geral. No entanto, não podemos estender a este capítulo o resultado de convergência global provado no Capítulo 4 para o método de PQS clássico, onde utilizamos o Lagrangiano Aumentado como função de mérito, pois lá precisávamos da hipótese de que a regularidade se cumpria em todos os iterandos. Na próxima seção introduziremos a estrutura básica de um método factível de programação quadrática sequencial e estabeceremos a convergência global de um método factível específico tomando como hipóteses a SSOSC e a condição de qualificação de Mangasarian-Fromovitz.

5.2 Um método de PQS factível e MFCQ

Até agora, neste capítulo, consideramos problemas de programação não-linear no formato (2.1) com restrições de igualdade e desigualdade. Neste momento, iremos alterar esta consideração inicial e pensar em problemas com restrições de desigualdade apenas. Assim nosso problema original passa a ser

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & g(x) \leq 0, \end{aligned} \tag{5.5}$$

onde $g(x) = [g_1(x), \dots, g_p(x)]^T$.

Um método factível de PQS baseia-se na mesma estratégia que os métodos clássicos, não factíveis. Dada uma aproximação factível x^k para a solução de (5.5), resolve-se o subproblema

$$\begin{aligned}
\min \quad & \frac{1}{2}d^T B_k d + \nabla f(x^k)^T d \\
\text{s.a} \quad & \nabla g_j(x^k)^T d + g_j(x^k) \leq 0, \quad j \in \{1, \dots, p\}.
\end{aligned} \tag{5.6}$$

onde B_k é uma matriz simétrica definida positiva.

A diferença essencial entre os métodos factíveis e não-factíveis aparece no momento de determinar x^{k+1} . Obtida a direção d^k que resolve (5.6), e considerando que a aproximação inicial x^0 é factível, o próximo ponto é dado por

$$x^{k+1} = x^k + t_k d^k$$

onde t_k é tal que $g(x^k + t_k d^k) \leq 0$. Ou seja, é escolhido um tamanho de passo t_k que torne a aproximação x^{k+1} factível para o problema original. Desta forma andamos sempre dentro da região viável e paramos em um ponto ótimo quando o subproblema na iteração k retorna como solução $d^k = 0$.

Claramente, há um custo para determinar o tamanho de passo que mantém a factibilidade dos iterandos e isto, juntamente com a condição de dar um ponto inicial factível, parece ser uma desvantagem em relação aos métodos clássicos de PQS. Mas conforme vimos na seção 4.3, os resultados existentes de convergência global também precisam calcular um tamanho de passo que garanta decréscimo suficiente da função de mérito. Além disso, se estamos sempre nos mantendo factíveis ao longo das iterações, não há necessidade de medir factibilidade. Isto nos leva a pensar que uma função de mérito para um método PQS factível deve levar em conta apenas a otimalidade do problema original, consistindo assim na própria função objetivo do problema. Com isso contornamos a dificuldade de selecionar um critério adequado no qual a busca linear na direção d^k deve basear-se.

Um outro problema com o qual também podemos nos deparar ao resolver o subproblema quadrático (5.6) é a inconsistência das restrições. Mas tendo em mãos um algoritmo que, de alguma maneira, mantém os iterandos factíveis para o problema (5.5), as restrições de (5.6) serão sempre consistentes, pois $d = 0$ é factível a cada iteração.

Direções factíveis foram utilizadas por Herskovits [19] na resolução de problemas de otimização com restrições de igualdade e desigualdade. O autor propôs um algoritmo de primeira ordem usando direções factíveis e também um algoritmo com convergência superlinear que inclui estimativas quase-Newton para a

Hessiana da função Lagrangiana. Ambos os algoritmos definem a cada iteração uma direção de descida por meio da resolução de um sistema linear nas variáveis primais e duais. Em essência, trata-se do sistema linear resultante ao se aplicar o método de Newton às condições KKT do problema usando, em vez da Hessiana do Lagrangiano, uma estimativa para a mesma. Em um segundo estágio, o sistema linear é perturbado de maneira a se obter uma direção de descida e factível. Realiza-se então uma busca linear a fim de se conseguir um novo ponto no interior do conjunto viável e assegurar a convergência global.

Quando são consideradas apenas restrições de desigualdade a sequência de pontos gerada reduz o valor da função objetivo a cada iteração. Já nos casos em que se incluem restrições de igualdade, para que estas sejam satisfeitas pode ser necessário que o valor da função objetivo aumente de uma iteração para outra. Trata-se de um método simples, que não requer a solução de problemas quadráticos e não se assemelha aos métodos de penalização ou barreira. Apenas dois sistemas lineares com a mesma matriz são resolvidos a cada iteração, mas para isso requer-se fortemente a hipótese de que a regularidade é satisfeita por todos os iterandos.

Métodos factíveis têm grande importância em muitos contextos como, por exemplo, aplicações em tempo real e também nos casos em que a função objetivo não está definida fora da região viável [8].

Nosso objetivo, a partir de agora, é expor com detalhes o método factível de PQS proposto por Panier e Tits em 1993 [31], assim como os resultados de convergência apresentados por Qi e Wei em [34], com algumas modificações no conjunto de hipóteses do Teorema 5.3 e da Proposição 5.3.

O algoritmo em questão requer que, para todo $k \in \mathbb{N}$, os iterandos satisfaçam

$$g(x^k) \leq 0 \tag{5.7}$$

e

$$f(x^{k+1}) \leq f(x^k). \tag{5.8}$$

Seja d^0 a solução de (5.6) na iteração k e suponhamos $d^0 \neq 0$. Como $d = 0$ é uma direção factível com valor nulo de função objetivo, então d^0 satisfaz

$$\frac{1}{2}(d^0)^T B_k d^0 + \nabla f(x^k)^T d^0 \leq 0.$$

Dado que B_k é definida positiva por hipótese, temos $\nabla f(x^k)^T d^0 < 0$ e segue daí que d^0 é sempre uma direção de descida para $f(x)$ em x^k . No entanto esta não é necessariamente uma direção factível. De fato, se $g_j(x^k) = 0$ e $\nabla g_j(x^k)^T d^0 = 0$, qualquer movimento nesta direção implicará em perda de factibilidade. Mesmo nos casos em que d^0 for factível, uma busca linear que force as desigualdades (5.7) e (5.8) a serem satisfeitas, pode não permitir que o passo unitário seja sempre aceito numa vizinhança de uma solução x^* e com isso bons resultados de convergência não podem ser obtidos. O algoritmo proposto contorna estas dificuldades através de dois mecanismos simples que correspondem a “inclinar” e “curvar”, sucessivamente, a direção de busca.

Em primeiro lugar, d^0 será substituída por uma combinação convexa

$$d = (1 - \theta)d^0 + \theta d^1 \quad (5.9)$$

entre d^0 e alguma direção de descida factível d^1 , com $\theta \in (0, 1)$. A idéia é que θ seja escolhido perto de 1 quando d^0 não for factível, mas que vá se aproximando de 0 à medida que os iterandos se aproximam de uma solução x^* , pois assim a característica quase-Newton de d^0 , ao menos localmente, não será perdida.

Em segundo lugar, a busca, agora não mais linear, se dará ao longo do arco $x^k + td + t^2 \tilde{d}$, onde $\tilde{d}$ é uma direção de correção escolhida de maneira que

- (i) numa vizinhança de x^* , $x^k + d + \tilde{d}$ seja factível e satisfaça $f(x^k + d + \tilde{d}) < f(x^k)$;
- (ii) $d + \tilde{d}$ convirja para d perto de x^* .

Mesmo perto da solução, $x^k + d$ pode violar tanto a factibilidade quanto o descréscimo da função. Realizar uma busca ao longo do arco $x^k + td + t^2 \tilde{d}$ ajuda a contornar esta situação [31].

Iremos assumir as seguintes hipóteses:

H1: O conjunto factível de (5.5) é não vazio.

H2: As funções f, g_j , $j = 1, \dots, p$, são continuamente diferenciáveis.

A direção factível e de descida d^1 será obtida a partir do iterando x^k por uma aplicação contínua $d^1 : \mathbb{R}^n \rightarrow \mathbb{R}^n$ satisfazendo

$$d^1(x) = 0 \quad \text{se } x \text{ é um ponto KKT}; \quad (5.10)$$

$$\nabla f(x)^T d^1(x) < 0 \quad \text{se } x \text{ não é um ponto KKT}; \quad (5.11)$$

$$\nabla g_j(x)^T d^1(x) < 0 \quad \text{se } x \text{ não é um ponto KKT e } j \in I(x). \quad (5.12)$$

Em [31] Panier e Tits indicaram que se x^k é um ponto regular, a aplicação contínua $d^1(x^k)$ satisfazendo (5.10), (5.11) e (5.12) pode ser tomada, por exemplo, como a solução de

$$\min \left\{ \frac{1}{2} \|d\|^2 + \max\{\nabla f(x^k)^T d; \max\{g_j(x^k) + \nabla g_j(x^k)^T d \mid j \in \{1, \dots, p\}\}\} \right\}. \quad (5.13)$$

Considerando (5.12) é fácil verificar que a existência da aplicação $d^1(x^k)$ implica que a MFCQ é satisfeita por todos os pontos não-KKT. Por outro lado, se todos os pontos não-KKT satisfazem MFCQ, ainda podemos garantir a existência de tal aplicação contínua (ver Seção 1.8 de [33]).

O coeficiente θ da combinação convexa também deve ser determinado por uma aplicação, $\theta : \mathbb{R}^n \rightarrow [0, 1]$. Quando $\|d^0\|$ for grande, maior do que uma dada constante positiva, devemos ter $d = d^1$ e portanto $\theta(x)$ deve ser tal que satisfaça $\theta(d^0) \approx 1$. Quando $\|d^0\|$ for pequena, devemos ter

$$\theta(d^0) = O(\|d^0\|^2) \quad (5.14)$$

para que não somente θ vá a zero à medida que $\|d^0\|$ for, mas também para que a busca ao longo de $x^k + td + t^2 \tilde{d}$ forneça assintoticamente um passo completo, isto é, passo igual a um [31]. Uma aplicação que satisfaz os requisitos especificados é dada, por exemplo, por

$$\theta(d) = \frac{\|d\|^2}{1 + \|d\|^2}. \quad (5.15)$$

Finalmente, a direção de correção $\tilde{d}$ pode ser escolhida como a solução do problema quadrático na variável $\tilde{d}$

$$\begin{aligned}
\min \quad & \frac{1}{2}(d + \tilde{d})^T B_k(d + \tilde{d}) + \nabla f(x^k)^T(d + \tilde{d}) \\
\text{s.a} \quad & \nabla g_j(x^k)^T \tilde{d} + g_j(x^k + d) \leq -\|d\|^\tau, \quad j \in \{1, \dots, p\},
\end{aligned}$$

caso esta solução possua norma menor ou igual ao $\min\{\|d\|, C\}$, onde C é uma constante positiva grande e $\tau \in (2, 3)$ é predeterminada. Caso contrário, tomamos $\tilde{d} = 0$. O objetivo desta correção é fazer com que $x^k + d + \tilde{d}$ seja um novo iterando satisfatório quando x^k encontra-se próximo a uma solução. Desta maneira, $\tilde{d}$ deve ser pequena quando comparada à d e deve, ao mesmo tempo, garantir que $g(x^k + d + \tilde{d}) < 0$ e $f(x^k + d + \tilde{d}) < f(x^k)$.

Com base nas considerações acima, Panier e Tits [31] propuseram o seguinte método factível de programação quadrática sequencial.

Algoritmo 5.2. *Seja $C > 0$, $\tau_1 \in (0, \frac{1}{2})$, $\tau_2 \in (0, 1)$, $\tau_3 \in (2, 3)$.
Dados: x^0 factível, $B_0 \in \mathbb{R}^{n \times n}$ simétrica definida positiva.*

Passo 0 (Inicialização): $k=0$.

Passo 1 (Calculando um arco de busca):

(i) Compute d_0^k que resolve o subproblema quadrático

$$\begin{aligned}
\min \quad & \frac{1}{2}d^T B_k d + \nabla f(x^k)^T d \\
\text{s.a} \quad & \nabla g_j(x^k)^T d + g_j(x^k) \leq 0, \quad j \in \{1, \dots, p\}.
\end{aligned}$$

Se $d_0^k = 0$, parar: x^k é uma solução.

(ii) Seja d_1^k a solução de (5.13), $\theta_k = \theta(d_0^k)$, e $d^k = (1 - \theta_k)d_0^k + \theta_k d_1^k$.

(iii) Tome uma direção de correção $\tilde{d}^k$ como sendo a solução do problema

$$\begin{aligned}
\min \quad & \frac{1}{2}(d^k + d)^T B_k(d^k + d) + \nabla f(x^k)^T(d^k + d) \\
\text{s.a} \quad & \nabla g_j(x^k)^T d + g_j(x^k + d^k) \leq -\|d^k\|^{\tau_3}, \quad j \in \{1, \dots, p\},
\end{aligned}$$

se tal direção existir e tiver norma menor que $\min\{\|d^k\|, C\}$. Caso contrário faça $\tilde{d}^k = 0$.

Passo 2 (Busca no arco): Tome t_k como o primeiro número t da sequência $\{1, \tau_2, \tau_2^2, \dots\}$ que satisfaz

$$f(x^k + td^k + t^2\tilde{d}^k) \leq f(x^k) + \tau_1 t \nabla f(x^k)^T d^k$$

e

$$g_j(x^k + td^k + t^2\tilde{d}^k) \leq 0, \quad \forall j \in \{1, \dots, p\}.$$

Passo 3 (Atualizações): Compute uma nova aproximação simétrica definida positiva B_{k+1} para a Hessiana do Lagrangiano. Faça $x^{k+1} = x^k + t_k d^k + t_k^2 \tilde{d}^k$ e $k = k + 1$. Volte ao Passo 1.

As seguintes proposições [31] mostram que o Algoritmo 5.2 está bem definido e, ou pára em um ponto KKT de (5.5), ou gera uma sequência infinita de iterandos $(x^k)_{k \in \mathbb{N}}$.

Proposição 5.1. A direção d_0^k está sempre bem definida e é igual a zero se, e somente se, x^k é um ponto KKT do problema original (5.5). Desta forma, se o Algoritmo 5.2 pára no Passo 1(i), então x^k é um ponto KKT de (5.5). Caso contrário, d_0^k é tal que

$$\nabla f(x^k)^T d_0^k < 0$$

e

$$\nabla g_j(x^k)^T d_0^k \leq 0, \quad \text{para todo } j \in I(x^k).$$

Proposição 5.2. No Passo 2, a busca no arco $x^k + td^k + t^2\tilde{d}^k$ fornece um passo $t_k = \tau^i$ para algum valor finito $i = i(k)$.

As proposições acima nos dizem, portanto, que enquanto um ponto KKT do problema (5.5) não for alcançado, a direção d_0 obtida a partir do subproblema quadrático a cada i-teração é sempre uma direção de descida para $f(x)$ e também uma direção factível. Além disso, é possível encontrar um tamanho de passo que mantenha a factibilidade dos iterandos e ainda forneça decréscimo da função de mérito, que neste caso nada mais é do que a própria função objetivo do problema.

Considerando que uma sequência infinita $(x^k)_{k \in \mathbb{N}}$ é gerada pelo Algoritmo 5.2, vamos supor que x^* é um ponto de acumulação de tal sequência. Para o resultado de convergência global do algoritmo, faremos as seguintes hipóteses adicionais:

H3: As matrizes B_k são uniformemente limitadas para todo $k \in \mathbb{N}$, isto é, existe $C_1 > 0$ tal que $\|B_k\| \leq C_1$ para todo k .

H4: Todos os iterandos satisfazem MFCQ.

Teorema 5.3. *Assuma que as hipóteses **H1-H4** são satisfeitas. Então x^* é um ponto KKT de (5.5).*

O teorema a seguir foi provado por Qi e Wei em [34] supondo, em vez de **H4**, que

- (i) todos os pontos não-KKT satisfazem MFCQ
- (ii) o ponto limite x^* satisfaz a CPLD.

No entanto, estas duas hipóteses não podem ocorrer ao mesmo tempo sem que os pontos KKT também satisfaçam MFCQ. De fato, suponha que x^* seja um minimizador local que satisfaz a CPLD com os gradientes das restrições sendo PLD. Então existe uma vizinhança $N(x^*)$ de x^* tal que para todo $x \in N(x^*)$ os gradientes das restrições ativas em x continuam PLD. Mas em toda vizinhança existem pontos não-KKT, que por hipótese satisfazem MFCQ e, portanto, seus vetores gradientes são PLI, o que é absurdo. Desta forma, a única opção é que x^* satisfaça CPLD sendo PLI, o que resulta na hipótese **H4**. Uma questão que fica em aberto é a seguinte: seria possível substituir **H4** por “Todos os iterandos satisfazem a CPLD?” Em caso afirmativo, teríamos um resultado teórico muito mais forte, dado que as hipóteses consideradas seriam mais fracas.

O mesmo teorema foi provado em [31] por Panier e Tits trocando a hipótese **H4** pela regularidade em todos os pontos factíveis de (5.5). Mas, conforme já citado anteriormente nesta mesma seção, **H4** implica na existência de uma solução contínua de (5.13) para cada x^k , o que permitiu provar o Teorema 5.3 supondo MFCQ em vez de LICQ.

Porém, trocar MFCQ por uma condição de qualificação mais fraca neste resultado parece não ser possível. Pensemos, por exemplo, em um problema que

contém as desigualdades $g(x) \leq 0$ e $-g(x) \leq 0$. Para tais restrições se satisfazem CRCQ e CPLD enquanto que MFCQ nunca é satisfeita. É fácil perceber, porém, que não existe $d_1(x)$ tal que $\nabla g(x)^T d_1(x) < 0$ e $-\nabla g(x)^T d_1(x) < 0$, já que estas duas últimas desigualdades não podem ser ambas satisfeitas ao mesmo tempo. Isso implica na inexistência de uma solução contínua de (5.13) para cada x^k , como requer o algoritmo 5.2.

Substituindo **H2** pela hipótese de que f e g são duas vezes continuamente diferenciáveis e acrescentando

H5: Existe um escalar $C_2 > 0$ tal que, para todo $k \in \mathbb{N}$, as estimativas da Hessiana satisfazem

$$d^T B_k d \geq C_2 \|d\|^2 \quad \text{para todo } d \in \mathbb{R}^n;$$

H6: x^* satisfaz a SSOSC,

também podemos mostrar a seguinte proposição:

Proposição 5.3. *Assuma que **H1-H6** são satisfeitas e $(x^k)_{k \in \mathbb{N}}$ é gerada pela Algoritmo 5.2. Então toda a seqüência converge para x^* , isto é,*

$$\lim_{k \rightarrow \infty} x^k = x^*. \quad (5.16)$$

Também neste caso, a proposição foi mostrada sob hipóteses mais fracas que aquelas utilizadas anteriormente por Panier e Tits e, apesar da inconsistência das hipóteses supostas por Qi e Wei, tanto a prova do Teorema 5.3 quanto a da Proposição 5.3 são, em essência, as mesmas apresentadas em [34]. Neste mesmo trabalho, os autores apresentaram uma versão modificada do Algoritmo 5.2 e provaram convergência superlinear a dois passos da mesma, sob hipóteses levemente mais fracas que a regularidade.

Capítulo 6

Experimentos Numéricos

A relevância dos resultados apresentados no Capítulo 5, para métodos de programação quadrática sequencial, está nas hipóteses que pedem condições de qualificação mais fracas que a regularidade. Desta maneira, temos garantia teórica de que os algoritmos apresentados podem identificar uma quantidade maior de minimizadores locais de um dado problema de programação não-linear. Visando verificar a veracidade dessa teoria, buscamos problemas com soluções degeneradas, isto é, soluções cujos multiplicadores de Lagrange não são únicos (uma vez que os gradientes das restrições ativas são vetores linearmente dependentes). A rotina utilizada nos testes foi a FFSQP, *Fortran Feasible Sequential Quadratic Programming* [42], que consiste na implementação do método factível proposto por Panier e Tits em [31]. FFSQP é composta por dois algoritmos factíveis de programação quadrática sequencial, os quais diferem pela busca linear realizada em um e em outro.

No primeiro algoritmo, identificado pelos autores como FFSQP-AL, e que de maneira simplificada consiste no Algoritmo 5.2 apresentado no Capítulo 5, realiza-se uma busca linear baseada no critério de Armijo e gera-se uma sequência de iterandos nos quais o valor da função objetivo decresce de uma iteração para a outra monotonicamente. No segundo, identificado como FFSQP-NL, realiza-se uma busca linear não-monótona em que o valor da função objetivo no próximo ponto deve ser menor que o máximo valor de função objetivo nos quatro últimos iterandos. Desta forma o critério de Armijo é relaxado, permite-se um possível aumento no valor da função objetivo de uma iteração para a outra e passos maiores são favorecidos. Conforme citado na Seção 4.4 do Capítulo 4, uma busca linear não-monótona atua contra a ocorrência do Efeito Maratos nos métodos de

programação quadrática sequencial.

De maneira geral, FFSQP-NL consiste dos seguintes passos:

Algoritmo 6.1. *Seja $C > 0$, $\tau_1 \in (0, \frac{1}{2})$, $\tau_2 \in (0, 1)$, $\tau_3 \in (2, 3)$.
Dados: x^0 factível, $B_0 \in \mathbb{R}^{n \times n}$ simétrica definida positiva.*

Passo 0 (Inicialização): $k=0$.

Passo 1 (Resolvendo o subproblema quadrático):

(i) Compute d_0^k que resolve o subproblema quadrático

$$\begin{aligned} \min \quad & \frac{1}{2} d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a} \quad & \nabla g_j(x^k)^T d + g_j(x^k) \leq 0, \quad j \in \{1, \dots, p\}. \end{aligned}$$

Se $d_0^k = 0$, parar: x^k é uma solução.

(ii) Seja d_1^k a solução de (5.13), $\theta_k = \theta(d_0^k)$, e $d^k = (1 - \theta_k)d_0^k + \theta_k d_1^k$.

Passo 2 (Calculando um arco de busca):

(i) Se $f(x^k + d^k) \leq \max_{0 \leq k \leq 4} \{f(x^k)\} + \tau_1 \nabla f(x^k)^T d^k$

e

$$g_j(x^k + d^k) \leq 0, \quad \forall j \in \{1, \dots, p\},$$

tome $t_k = 1$, $\tilde{d}^k = 0$ e vá para o Passo 3. Senão,

(ii) Tome uma direção de correção $\tilde{d}^k$ como sendo a solução do problema

$$\begin{aligned} \min \quad & \frac{1}{2} (d^k + d)^T B_k (d^k + d) + \nabla f(x^k)^T (d^k + d) \\ \text{s.a} \quad & \nabla g_j(x^k)^T d + g_j(x^k + d^k) \leq - \|d^k\|^{\tau_3}, \quad j \in \{1, \dots, p\}, \end{aligned}$$

se tal direção existir e tiver norma menor que $\min\{\|d^k\|, C\}$. Caso contrário faça $\tilde{d}^k = 0$.

Tome t_k como o primeiro número t da sequência $\{1, \tau_2, \tau_2^2, \dots\}$ que satisfaz

$$f(x^k + td^k + t^2\tilde{d}^k) \leq \max_{0 \leq k \leq 4} \{f(x^k)\} + \tau_1 t \nabla f(x^k)^T d^k$$

e

$$g_j(x^k + td^k + t^2\tilde{d}^k) \leq 0, \quad \forall j \in \{1, \dots, p\}.$$

Passo 3 (Atualizações): Compute uma nova aproximação simétrica definida positiva B_{k+1} para a Hessiana do Lagrangiano. Faça $x^{k+1} = x^k + t_k d^k + t_k^2 \tilde{d}^k$ e $k = k + 1$. Volte ao Passo 1.

Todos os detalhes da implementação de FFSQP-AL e FFSQP-NL, assim como a abordagem para problemas do tipo minimax utilizando estes algoritmos, encontram-se em [42].

6.1 Testes preliminares

A fim de facilitar a identificação dos algoritmos, vamos denotar FFSQP-AL por FFSQP-MONO, fazendo referência à busca linear monótona, e FFSQP-NL por FFSQP-NONM, fazendo referência à busca linear não-monótona. No contexto de condições de qualificação mais fracas que a regularidade, como MFCQ, CRCQ e CPLD, estão os problemas com infinitos multiplicadores de Lagrange associados a uma única solução primal, uma solução degenerada. Para analisar o desempenho dos algoritmos na resolução de problemas com este tipo de solução, buscamos exemplos de dimensão pequena, nos quais é possível identificar algebricamente que se trata, de fato, de uma solução degenerada e também verificar quais condições de qualificação são satisfeitas pelo conjunto de restrições ativas nesta solução.

Denotemos por $M(x^*)$ o conjunto dos multiplicadores associados à solução x^* .

• Exemplo 1

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.a} \quad & (x_1 + 2)^2 \leq 4 \\ & -x_1 \leq 0. \end{aligned}$$

Solução $x^* = (0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*) \in \mathbb{R}_+^2 \mid 4\mu_1^* - \mu_2^* = 0\}$. Esta solução satisfaz CRCQ, mas não satisfaz MFCQ e, conseqüentemente, não satisfaz LICQ.

- Exemplo 2

$$\begin{array}{ll} \min & x_1^2 \\ \text{s.a} & x_2 = 0 \\ & -x_2 \leq 0. \end{array}$$

Solução $x^* = (0, 0)^T$ e $M(x^*) = \{(\lambda_1^*, \mu_1^*) \in \mathbb{R}_+^2 \mid \lambda_1^* - \mu_1^* = 0\}$. Esta solução viola MFCQ mas satisfaz a SSOSC para todo $(\lambda_1^*, \mu_1^*) \in M(x^*)$.

- Exemplo 3 [39]

$$\begin{array}{ll} \min & x_1 \\ \text{s.a} & (x_1 - 2)^2 + x_2^2 \leq 4 \\ & (x_1 - 4)^2 + x_2^2 \leq 16. \end{array}$$

Solução $x^* = (0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*) \in \mathbb{R}_+^2 \mid \mu_1^* + 2\mu_2^* = 1/4, 0 \leq \mu_2^* \leq 1/8\}$. Esta solução satisfaz MFCQ (mas não LICQ).

- Exemplo 4 [24]

$$\begin{array}{ll} \min & -x_1 \\ \text{s.a} & x_1 + -x_2^2 \leq 0 \\ & x_1 + x_2^2 \leq 0. \end{array}$$

Solução $x^* = (0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*) \in \mathbb{R}_+^2 \mid \mu_1^* + \mu_2^* = 1\}$. Esta solução satisfaz MFCQ (mas não LICQ).

- Exemplo 5 [40]

$$\begin{array}{ll} \min & x_1 \\ \text{s.a} & (x_1 - 2)^2 + x_2^2 \leq 4 \\ & -x_1 \leq 0. \end{array}$$

Solução $x^* = (0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*) \in \mathbb{R}_+^2 \mid \mu_2^* = 1 - 4\mu_1^*, 0 \leq \mu_1^* \leq 1/4\}$. Esta solução satisfaz MFCQ (mas não LICQ).

- Exemplo 6 [36]

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.a} & x_1 + x_2^2 \leq 0 \\ & x_1 + x_2 \leq 0 \\ & -x_1 - x_2 \leq 0 \\ & x_1 \leq 0. \end{array}$$

Solução $x^* = (0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*, \mu_3^*, \mu_4^*) \in \mathbb{R}_+^4 \mid \mu_1^* + \mu_4^* = 0, \mu_2^* - \mu_3^* = 0\}$. Esta solução satisfaz CPLD mas não satisfaz LICQ, MFCQ e CRCQ.

• **Exemplo 7** [40]

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\ \text{s.a} \quad & x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_1 - x_2 + x_3 - x_4 \leq 8 \\ & x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 \leq 10 \\ & 2x_1^2 + x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_2 - x_4 \leq 5 \\ & -x_2^3 + -2x_1^2 - x_4^2 - x_1 + 3x_2 + x_3 - 4x_4 \leq 7. \end{aligned}$$

Solução $x^* = (0, 1, 2, -1)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*, \mu_3^*, \mu_4^*) \in \mathbb{R}_+^4 \mid \mu_1^* = 3 - \mu_3^*, \mu_2^* = 0, 2 \leq \mu_3^* \leq 3, \mu_4^* = \mu_3^* - 2\}$. Esta solução satisfaz MFCQ (mas não LICQ).

• **Exemplo 8** [3]

$$\begin{aligned} \min \quad & x_3 \\ \text{s.a} \quad & \langle Q_i(x_1, x_2), (x_1, x_2) \rangle - x_3 \leq 0, \quad i = 1, 2, 3 \end{aligned}$$

onde

$$Q_1 = \begin{pmatrix} 0 & \sqrt{3} \\ \sqrt{3} & -2 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & -\sqrt{3} \\ -\sqrt{3} & -2 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} -3 & 0 \\ 0 & 1 \end{pmatrix}.$$

Solução $x^* = (0, 0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*, \mu_3^*) \in \mathbb{R}_+^3 \mid \sum_{i=1}^3 \mu_i^* = 1\}$. Esta solução satisfaz MFCQ (mas não LICQ).

• **Exemplo 9** [2]

$$\begin{aligned} \min \quad & x_3 \\ \text{s.a} \quad & \langle Q_i(x_1, x_2), (x_1, x_2) \rangle - x_3 \leq 0, \quad i = 1, 2, 3, 4, \end{aligned}$$

onde

$$Q_i = U_i^T Q U_i, \quad U_i = \begin{pmatrix} \cos[\pi(i-1)/4] & \sin[\pi(i-1)/4] \\ -\sin[\pi(i-1)/4] & \cos[\pi(i-1)/4] \end{pmatrix} \text{ e } Q = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}.$$

Solução $x^* = (0, 0, 0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*, \mu_3^*, \mu_4^*) \in \mathbb{R}_+^4 \mid \sum_{i=1}^4 \mu_i^* = 1\}$. Esta solução satisfaz MFCQ (mas não LICQ).

• **Exemplo 10** [40]

$$\begin{aligned} \min \quad & x_1 \\ \text{s.a} \quad & (x_1 - 2)^2 + x_2^2 \leq 4 \\ & (x_1 - 4)^2 + x_2^2 \leq 16 \\ & x_1^2 + (x_2 - 2)^2 \leq 4. \end{aligned}$$

Solução $x^* = (0, 0, 0)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*, \mu_3^*) \in \mathbb{R}_+^3 \mid \mu_1^* + 2\mu_2^* = 1/4, 0 \leq \mu_2^* \leq 1/8, \mu_3^* = 0\}$. Esta solução satisfaz MFCQ (mas não LICQ).

• **Exemplo 11** [22]

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 + x_3^2 \\ \text{s.a} \quad & \sin x_1 + \sin x_2 + \sin x_3 = 0 \\ & x_1 + x_2 + x_3 + x_1^2 + \sin x_1 x_3 = 0. \end{aligned}$$

Solução $x^* = (0, 0, 0)^T$ e $M(x^*) = \{(\lambda_1^*, \lambda_2^*) \in \mathbb{R}^2 \mid \lambda_1^* + \lambda_2^* = 0\}$. Esta solução viola LICQ, MFCQ, CRCQ e CPLD.

• **Exemplo 12** [23]

$$\begin{aligned} \min \quad & x_1 + x_2 + x_1^* + x_2^2 + (x_3 - 1)^2 \\ \text{s.a} \quad & -x_1 - x_2^2 \leq 0 \\ & -x_1^2 + (x_3 - 1)^2 \leq 0 \\ & (x_1 + x_2)x_2 + (x_1^2 - (x_3 - 1)^2)x_3 \leq 0. \end{aligned}$$

Solução $x^* = (0, 0, 1)^T$ e $M(x^*) = \{(\mu_1^*, \mu_2^*, \mu_3^*) \in \mathbb{R}_+^3 \mid \mu_1^* = 1, \mu_2^* \geq 0, \mu_3^* \geq 0\}$. Esta solução viola LICQ, MFCQ, CRCQ e CPLD.

Estes problemas foram resolvidos por ambos os algoritmos, FFSQP-MONO e FFSQP-NONM, com a finalidade de mensurar a eficácia dos mesmos na identificação de soluções degeneradas. Os resultados estão apresentados na Tabela 6.1. Nesta, “n” representa o número de variáveis, “Iter” o número de iterações, “Avalf” o número de avaliações da função objetivo, “Avalg” o número de avaliações das restrições e “Norma KKT” a norma do vetor KKT na solução encontrada. Os problemas marcados com um “OK” no campo “Solução” foram resolvidos com sucesso. Os demais, marcados com um “X”, serão discutidos com maiores detalhes. Os algoritmos param quando a norma do vetor KKT for menor ou igual a 10^{-6} ou quando a norma da direção d_0 for menor ou igual a 10^{-8} . No entanto o sucesso da solução encontrada só é acusado quando ambas as condições são satisfeitas ao mesmo tempo. Se apenas uma ou outra condição for satisfeita, uma mensagem de erro é devolvida juntamente com a solução.

Começando pelo problema do Exemplo 4, verificamos que o algoritmo FFSQP-MONO chegou ao número máximo de iterações (500) antes de atingir a precisão pedida. Já FFSQP-NONM alcançou o resultado em 28 iterações com norma do

Problema	Algoritmo	n	Iter	Avalf	Avalg	Norma KKT	Solução
1	MONO	2	2	3	4	0.0000+00	OK
	NONM		2	3	4	0.0000+00	OK
2	MONO	2	2	3	4	0.0000+00	OK
	NONM		2	3	4	0.0000+00	OK
3	MONO	2	16	16	70	0.8615E-16	OK
	NONM		10	10	28	0.3243E-10	OK
4	MONO	2	500	501	5057	0.1726E-05	X
	NONM		28	28	33	0.5326E-15	OK
5	MONO	2	108	109	637	0.9354E-10	OK
	NONM		28	28	33	0.5326E-15	OK
6	MONO	2	2	2	2	0.0000+00	OK
	NONM		2	2	2	0.0000+00	OK
7	MONO	4	12	12	83	0.3904E-07	OK
	NONM		14	14	73	0.9420E-19	OK
8	MONO	3	-	-	-	-	X
	NONM		11	11	353	-	X
9	MONO	3	500	501	7883	0.6903E-05	X
	NONM		29	29	125	0.2001E-10	OK
10	MONO	2	15	15	97	0.4803E-10	OK
	NONM		8	8	42	0.9252E-12	OK
11	MONO	3	65	34	163	0.1452E-05	X
	NONM		72	78	175	0.4327E-05	X
12	MONO	3	10	10	33	0.5587E+01	X
	NONM		37	37	216	0.5007E+01	X

Tabela 6.1: Resultados apresentados pelos algoritmos FFSQP-MONO e FFSQP-NONM para doze problemas com soluções degeneradas.

vetor KKT bem pequena.

Nenhum dos dois algoritmos conseguiu encontrar a solução para o problema do Exemplo 8, mas este insucesso já era esperado para problemas desse tipo ao se aplicar métodos de programação quadrática sequencial. Como a função objetivo é linear, os subproblemas quadráticos acabam reduzidos a problemas de programação linear. Ao deparar-se com regiões viáveis ilimitadas, o *solver* responsável pela resolução dos subproblemas quadráticos fica impossibilitado de determinar a direção d_0 , uma vez que o valor da função objetivo pode decrescer tanto quanto se queira caminhando pelo conjunto dos pontos factíveis. Isto é exatamente o que acontece neste exemplo.

O problema do Exemplo 9 possui as mesmas características daquele do Exemplo 8, mas neste caso FFSQP-NONM conseguiu encontrar a solução enquanto que FFSQP-MONO novamente atingiu o número máximo de iterações antes que o critério de parada fosse satisfeito.

No caso do Exemplo 11, por tratar-se de um conjunto de restrições apenas com igualdades, os métodos factíveis são incapazes de satisfazer a factibilidade dos iterandos e, mesmo gerando uma sequência de pontos que caminham para a solução, a norma da direção torna-se pequena demais antes que um ponto KKT seja alcançado, fazendo com que o algoritmo pare por não conseguir sair do lugar.

Finalmente, o motivo de nenhum dos dois algoritmos ter resolvido o problema do Exemplo 12 parece ter forte relação com a função Lagrangiana. O método estaciona em um ponto que não é KKT do problema original (note que a norma desse vetor é grande), mas é ponto de sela da função Lagrangiana. Como este ponto de sela mora no conjunto viável do problema, tanto FFSQP-MONO quanto FFSQP-NONM são atraídos por ele e não conseguem sair daí. Os algoritmos param pois a norma da direção d_0 novamente torna-se muito pequena.

Estes resultados sugerem que o método de programação quadrática sequencial factível, implementado nos algoritmos FFSQP-MONO e FFSQP-NONM, são eficientes na resolução de problemas que possuem apenas restrições de desigualdades, nos quais o conjunto viável é “amplo” o suficiente para que os algoritmos consigam caminhar pelo seu interior em direção à solução. Podemos dizer que o método de programação quadrática sequencial factível compara-se, de maneira geral, a um método de pontos interiores para problemas não-lineares.

6.2 O problema da montanha

A partir das considerações expostas na última seção, buscamos problemas de dimensões maiores, com restrições de desigualdade apenas, para continuar testando os algoritmos, agora com relação à quantidade de variáveis. Dados uma superfície $S(x, y)$, um ponto inicial p_i e um ponto final p_f , ambos em $\mathbb{R}^2$, o problema escolhido para os testes consiste em encontrar um caminho $p_i, p_1, p_2, \dots, p_N, p_f$ de p_i até p_f tal que $\max_{1 \leq k \leq N} \{S(p_k)\}$ seja o menor possível, onde N é o número de pontos intermediários. Além disso, a distância entre dois pontos consecutivos no caminho não deve ser maior que uma tolerância previamente estabelecida. Devido ao formato das superfícies, tal problema é conhecido como *problema da montanha*.

Este problema pode ter uma formulação do tipo minimax com N funções objetivos e $N + 1$ restrições de desigualdade, ou então com $2N + 1$ restrições de desigualdade e uma única função objetivo linear. Nos testes realizados, utilizou-se a seguinte formulação:

$$\begin{aligned} \min \quad & z \\ \text{s.a} \quad & d(p_i, p_1)^2 \leq d_{max}^2 \\ & d(p_{k-1}, p_k)^2 \leq d_{max}^2, \quad k = 2, \dots, N \\ & d(p_N, p_f)^2 \leq d_{max}^2 \\ & S(p_k) \leq z, \quad k = 1, \dots, N, \end{aligned}$$

onde $d(.,.)$ é a distância Euclidiana e d_{max} é a distância máxima definida previamente pelo usuário e que deve satisfazer $d_{max} \geq d(p_i, p_f)/(N + 1)$. Temos então um problema com $2N + 1$ variáveis e $2N + 1$ restrições de desigualdade.

Os valores iniciais para $p_1, p_2, \dots, p_N$ correspondem a uma perturbação dos pontos igualmente espaçados entre p_i e p_f e o valor inicial de z é

$$z = \max\{S(p_i), S(p_1), \dots, S(p_N), S(p_f)\}.$$

Os testes foram realizados utilizando como pontos iniciais $p_i = (-10, -10)$ e $p_f = (10, 10)$ e as superfícies $S_1(x, y) = \sin(x \cdot y) + \sin(x + y)$ e $S_2(x, y) = \sin(x) + \cos(y)$, esta última representada na figura 6.1.

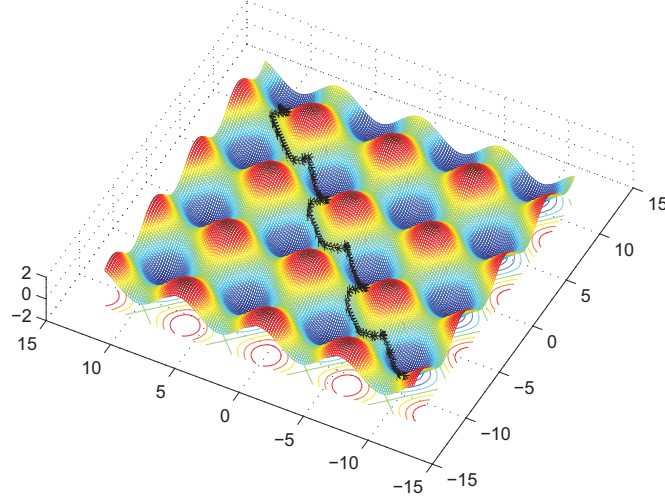


Figura 6.1: Superfície $S_2(x, y) = \text{sen}(x) + \cos(y)$ com um caminho ligando o ponto inicial $p_i = (-10, -10)$ ao ponto final $p_f = (10, 10)$.

As Tabelas 6.2 e 6.3 trazem os resultados obtidos pelos algoritmos FFSQP-MONO e FFSQP-NONM para as superfícies $S_1(x, y) = \text{sen}(x \cdot y) + \text{sen}(x + y)$ e $S_2(x, y) = \text{sen}(x) + \cos(y)$, respectivamente, e para diferentes quantidades de pontos intermediários.

Assim como na Tabela 6.1, “n” representa o número de variáveis, “Iter” o número de iterações, “Avalf” o número de avaliações da função objetivo, “Avalg” o número de avaliações das restrições e “Norma KKT” a norma do vetor KKT na solução encontrada. Os problemas marcados com um “OK” no campo “Solução” foram resolvidos com sucesso. Os demais, marcados com (*) e (**), devolveram mensagens de erro que constam nas tabelas. O critério de parada dos algoritmos estabelece que a norma da direção d_0 deve ser menor ou igual a 10^{-8} ou que a norma do vetor KKT deve ser menor ou igual a 10^{-6} .

É possível perceber, tanto na tabela 6.2 quanto na 6.3 que FFSQP-NONM apresenta um desempenho melhor que FFSQP-MONO na grande maioria dos casos no que diz respeito ao número de iterações e avaliações de função. Em particular na Tabela 6.3, fica claro a ocorrência do Efeito Maratos e como FFSQP-NONM parece agir contra o mesmo. Em alguns casos FFSQP-MONO atinge o número máximo de iterações antes algum dos critérios de parada sejam satisfeitos. Em outros casos, detecta-se que o novo iterando é equivalente ao anterior, no entanto

Pontos intermediários	n	Algoritmo	Iter	Avalf	Avalg	Norma KKT	Solução
10	21	MONO	20	23	596	0.2942E-07	OK
		NONM	27	29	675	0.2662E-08	OK
20	41	MONO	37	107	2582	0.5459E-11	OK
		NONM	23	30	1291	0.2531E-10	OK
30	61	MONO	33	57	3187	0.3666E-09	OK
		NONM	31	35	2359	0.8940E-09	OK
40	81	MONO	49	122	6599	0.1577E-10	OK
		NONM	35	36	3005	0.5207E-09	OK
50	101	MONO	35	44	4913	0.5875E-09	OK
		NONM	49	51	5358	0.3597E-10	OK
100	201	MONO	412	457	148390	0.1483E-08	OK
		NONM	136	232	41789	0.1161E-10	OK
200	401	MONO	85	402	64178	0.3736E-08	OK
		NONM	29	57	18887	0.2946E-09	OK
300	601	MONO	39	126	48801	0.1769E-14	OK
		NONM	14	14	17097	0.3444E-08	OK
400	801	MONO	41	105	77118	0.1625E-09	OK
		NONM	17	17	30305	0.1380E-07	OK
500	1001	MONO	500	3619	987822	0.4517E-07	(*)
		NONM	20	23	52235	0.1771E-07	OK

Tabela 6.2: Resultados apresentados pelos algoritmos FFSQP-MONO e FFSQP-NONM para $S(x, y) = \sin(x \cdot y) + \sin(x + y)$. * Número máximo de iterações atingido antes que uma solução seja encontrada.

Pontos intermediários	n	Algoritmo	Iter	Avalf	Avalg	Norma KKT	Solução
10	21	MONO	100	158	5344	0.7137E-09	(**)
		NONM	50	50	1075	0.6625E-09	OK
20	41	MONO	24	24	1201	0.2333E-08	OK
		NONM	20	20	840	0.7730E-09	OK
30	61	MONO	131	156	12453	0.1536E-09	(**)
		NONM	39	40	2512	0.6087E-10	OK
40	81	MONO	412	433	52288	0.5039E-08	(**)
		NONM	28	28	2436	0.7309E-10	OK
50	101	MONO	109	204	17353	0.2346E-08	(**)
		NONM	53	54	5576	0.2553E-09	OK
100	201	MONO	143	295	44199	0.1450E-07	(**)
		NONM	109	135	29927	0.8910E-09	OK
200	401	MONO	500	3659	456291	0.1172E-08	(*)
		NONM	153	205	88934	0.1978E-08	(**)
300	601	MONO	500	3321	571461	0.7994E-08	(*)
		NONM	201	217	159833	0.2333E-13	OK
400	801	MONO	500	3118	877588	0.7084E-08	(*)
		NONM	180	187	194839	0.1534E-09	OK
500	1001	MONO	500	501	842890	0.1245E-02	(*)
		NONM	500	501	710810	0.1847E-03	(*)

Tabela 6.3: Resultados apresentados pelos algoritmos FFSQP-MONO e FFSQP-NONM para $S(x, y) = \sin(x) + \cos(y)$.

(*) Número máximo de iterações atingido antes que uma solução seja encontrada. (**) O novo iterando é numericamente equivalente ao iterando atual, embora o critério de parada não tenha sido satisfeito.

também não se satisfazem os dois critérios de parada.

Ainda analisando a Tabela 6.3, com exceção do último teste que possui 500 pontos intermediários, a norma do vetor KKT já é da ordem de 10^{-6} , mas a norma da direção d_0 ainda não é pequena o suficiente. Como a busca linear (ou no arco) realizada é monótona, são dados passos cada vez menores e o algoritmo não consegue detectar que já encontrou uma solução. Poderíamos dizer que ele “patina” perto da solução e nunca consegue chegar até ela, o que resulta em perda na velocidade de convergência. Por outro lado, FFSQP-NONM contorna esta dificuldade ao realizar uma busca não-monótona, permitindo que o valor da função objetivo aumente de uma iteração para a outra e favorecendo a aceitação do passo um. Assim, a busca não-monótona encontra solução para problemas que a busca monótona não é capaz de resolver.

Seguem algumas análises comparativas entre o desempenho de FFSQP-MONO e FFSQP-NONM com relação ao número de iterações, número de avaliações de função objetivo e avaliações das funções que compõem o conjunto de restrições dos problemas. As comparações foram realizadas apenas para a superfície $S_1(x, y) = \sin(x \cdot y) + \sin(x + y)$, cujos resultados constam na Tabela 6.2, pois os dois algoritmos tiveram sucesso na resolução dos problemas, a não ser para o último deles que possui 500 pontos intermediários. No caso da superfície $S_2(x, y) = \sin(x) + \cos(y)$, como observamos na Tabela 6.3, apenas FFSQP-NONM teve sucesso na resolução da maioria dos problemas e por isso não faz sentido uma comparação de valores.

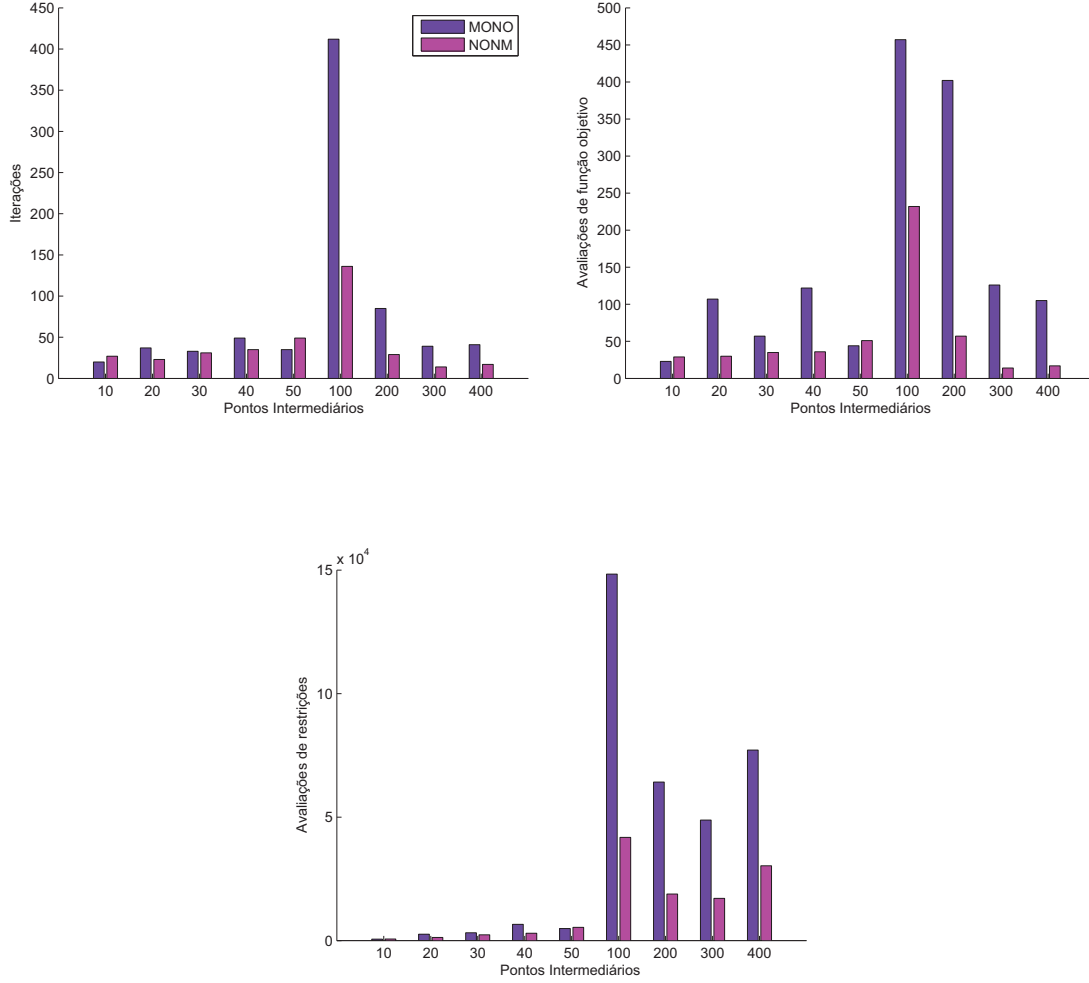


Figura 6.2: Análise comparativa entre o desempenho dos algoritmos FFSQP-MONO e FFSQP-NONM com relação ao número de iterações, avaliações de função objetivo e avaliações das restrições.

Achamos conveniente, entretanto, comparar a performance dos algoritmos FFSQP-MONO e FFSQP-NONM através da ferramenta “*performance profile*” introduzida por Dolan e Moré [11]. Trata-se de uma função de distribuição acumulada $\rho_s : \mathbb{R} \rightarrow [0, 1]$, onde s faz referência ao algoritmo considerado.

Seja S o conjunto dos algoritmos e P o conjunto dos problemas a serem resolvidos. Se o problema p é resolvido pelo algoritmo s então

$$\rho_s(t) = \frac{1}{n_p} |p \in P \mid r_{s,p} \leq t|,$$

onde

$$r_{s,p} = \frac{nit_{s,p}}{\min\{nit_{s,p} \mid \forall s \in S\}},$$

com n_p igual ao número total de problemas e $nit_{s,p}$ o número de iterações que o algoritmo s realizou para resolver o problema p . Se o problema p não é resolvido pelo algoritmo s então $r_{s,p} = r_M$, um número grande qualquer. Além do número de iterações, a medida de desempenho do *performance profile* pode ser o número de avaliações de função objetivo, tempo gasto na resolução do problema, etc.

Considerando o número de iterações gastas pelos algoritmos FFSQP-MONO e FFSQP-NONM, obtivemos a seguinte comparação de desempenho:

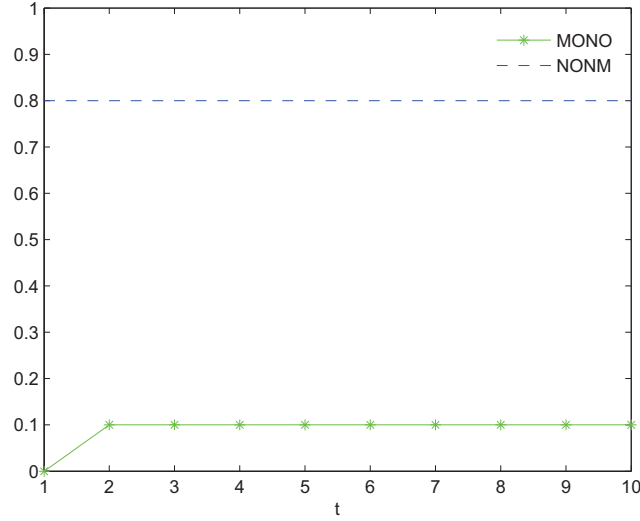


Figura 6.3: Perfil de desempenho dos algoritmos FFSQP-MONO e FFSQP-NONM com relação ao número de iterações gastas na resolução dos problemas com superfície $S(x, y) = \sin(x) + \cos(y)$.

A Figura 6.3 mostra que FFSQP-MONO não resolve nenhum problema (0%) com o número mínimo de iterações e resolve 10% deles com no mínimo duas vezes

mais iterações que FFSQP-NONM. Este, por sua vez, resolve 80% dos problemas, todos eles com o número mínimo de iterações. Os gráficos referentes ao *performance profile* baseado no número de avaliações de função objetivo e ao baseado no número de avaliações de restrições apresentaram o mesmo comportamento do gráfico da Figura 6.3. Em todos os casos FFSQP-NONM apresentou melhor performance que FFSQP-MONO.

Capítulo 7

Conclusões

O propósito deste trabalho foi o estudo de métodos de programação quadrática sequencial na resolução de problemas de programação não-linear e das propriedades de um método de PQS em particular, factível, quando diferentes condições de qualificação são consideradas. Apresentamos uma análise das relações existentes entre quatro condições de qualificação em especial, a regularidade (LICQ), Mangasarian-Fromovitz (MFCQ), posto constante (CRCQ) e a condição de dependência linear positiva constante (CLPD), a mais fraca dentre estas. Resultados de convergência que têm, em suas hipóteses, condições de qualificação fracas são mais fortes do que aqueles que estão baseados em condições de qualificação fortes. Quanto mais fraca uma condição de qualificação, mais pontos viáveis a satisfazem e, com isso, maior é o alcance do método ao buscar pontos KKT. Daí a importância deste estudo.

Vimos que, sob a hipótese da independência linear dos gradientes das restrições ativas nos iterandos gerados pelo método clássico de PQS, é possível garantir a limitação do parâmetro de penalização da função de mérito e a convergência para um ponto estacionário desta nos casos em que a sequência de iterandos não diverge. Este é um fato importante uma vez que, sob a mesma garantia de limitação do parâmetro de penalização, temos que todo ponto KKT do problema original é também um ponto estacionário da função de mérito. Estes resultados foram mostrados neste trabalho sem supor que os iterandos estão contidos em um compacto, hipótese esta utilizada por Boggs em [6] na prova dos mesmos resultados. Um questionamento que surgiu ao longo do trabalho, podendo ser analisado futuramente, é a possibilidade de obtenção de um ponto estacionário da função de mérito nos casos em que a sequência possui vários pontos de acumulação e não

somente quando esta possui um único ponto limite.

Com base no trabalho de Qi e Wei [34] apresentamos também resultados de convergência de métodos de PQS atrelados a condições de qualificação mais fracas que a regularidade, tais como Mangasarian-Fromovitz e a condição de dependência linear positiva constante. Em particular, exploramos as propriedades de um método de PQS factível e verificamos que duas das hipóteses utilizadas por Qi e Wei na prova de convergência deste método poderiam ser sintetizadas em uma única. Para que o ponto limite x^* satisfaça a CPLD e todos os pontos não-KKT satisfaçam MFCQ a única possibilidade é que x^* também satisfaça MFCQ. Com isso, substituímos as duas hipóteses previamente citadas pela hipótese de que MFCQ é satisfeita por todos os iterandos gerados pelo método de PQS factível. Uma questão surge a partir destas considerações: seria possível provar o mesmo resultado de convergência para pontos KKT sob uma condição de qualificação mais fraca que a de Mangasarian-Fromovitz?

Finalmente, com os testes numéricos validamos a eficiência do método de PQS factível na resolução de problemas com soluções degeneradas, caso em que estão envolvidas condições de qualificação mais fracas que LICQ. Neste caso, selecionamos problemas com poucas variáveis, para os quais era possível analisar quais condições de qualificação eram satisfeitas na solução. Com a resolução destes, percebemos que a performance do método de PQS factível depende fortemente do conjunto de restrições, que deve ser composto apenas por desigualdades. Frequentemente, quando há uma ou mais restrições de igualdade o método não consegue manter a factibilidade, pois é difícil caminhar em um conjunto viável muito “estrito”. Estes problemas prévios nos deram uma idéia das características que os problemas candidatos a serem bem resolvidos pelo método de PQS factível devem possuir. A partir daí, selecionamos problemas com dimensões maiores, contendo características semelhantes às que observamos previamente, e comparamos o comportamento de dois diferentes tipos de busca: monótona e não-monótona. Os resultados obtidos levaram à conclusão de que a busca não-monótona tem melhor performance quanto ao número de iterações e avaliações de função realizadas, além de atuar contra a ocorrência do Efeito Maratos, frequente quando métodos de PQS são utilizados. A busca não-monótona favorece a aceitação do passo um e com isso, quando comparada à busca monótona, ganhamos não apenas no número de iterações e avaliações de função mas também em velocidade de convergência. Assim, em muitos casos, há maior eficiência computacionalmente quando somos menos conservadores e permitimos a aceitação de passos maiores,

mesmo se o valor da função objetivo (ou função de mérito) sofrer um acréscimo temporário.

Métodos de programação quadrática sequencial, condições de qualificação e resultados de convergência envolvendo ambos foram apresentados, estudados e analisados neste trabalho, o qual mostrou quão úteis podem ser na prática, em aplicações envolvendo conjuntos viáveis formados por restrições de desigualdade, e quão vasto e rico é o campo teórico abordado por estes tópicos na área da otimização não-linear.

Referências Bibliográficas

- [1] R. Andreani, J. M. Martínez, M. L. Schuverdt, “On the relation between Constant Positive Linear Dependence Condition and Quasi-normality Constraint Qualification”, *Journal of Optimization Theory and Applications*, vol. 125, no. 2, pp. 473–485, 2005.
- [2] A. V. Arutyunov, “Perturbations of extremal problems with constraints and necessary optimality conditions”, *Journal of Mathematical Sciences*, vol. 54, no. 6, pp. 1342–1400, 1991.
- [3] A. Baccari, A. Trad, “On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints”, *SIAM Journal on Optimization*, vol. 15, no. 2, pp. 394–408, 2004.
- [4] D. P. Bertsekas “Constrained optimization and Lagrange Multiplier Methods”, Academic Press, 1982.
- [5] D. P. Bertsekas, A. E. Ozdaglar, “Pseudonormality and a Lagrange multiplier theory for constrained optimization”, *Journal of Optimization Theory and Applications*, vol. 114, no. 2, pp. 287–343, 2002.
- [6] P. T. Boggs, J. W. Tolle, “Sequential Quadratic Programming”, *Acta Numerica*, pp. 1–51, 1995.
- [7] P. T. Boggs, J. W. Tolle, “Augmented Lagrangians which are quadratic in the multiplier”, *Journal of Optimization Theory and Applications*, vol. 31, no. 1, pp. 17–26, 1980.
- [8] J. F. Bonnans, E. R. Panier, A. L. Tits, J. L. Zhou, “Avoiding the Maratos effect by means of a nonmonotone line search. II - Inequality constrained problems - Feasible iterates”, *SIAM Journal on Numerical Analysis*, vol. 29, no. 4, pp. 1187–1202, 1992.

- [9] R. M. Chamberlain, J. D. Powell, C. Lemarechal, H. C. Pedersen, “The watchdog technique for forcing convergence in algorithms for constrained optimization”, *Mathematical Programming Study*, vol. 16, pp. 1–17, 1982.
- [10] J. E. Dennis Jr., R. B. Schnabel, “Numerical Methods for unconstrained Optimization and Nonlinear Equations”, *Society for Industrial and Applied Mathematics*, 1996.
- [11] E. D. Dolan, J. J. Moré, “Benchmarking Optimization Software with Performance Profiles”, *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.
- [12] M. A. Diniz-Ehrhardt, J. M. Martínez, M. Raydan, “A derivative-free non-monotone line-search technique for unconstrained optimization”, *Journal of Computational and Applied Mathematics*, vol. 219, no.2, pp. 383–397, 2008.
- [13] R. Fletcher, “A class of methods for nonlinear programming, III: Rates of convergence”, *Numerical Methods for Nonlinear Optimization* (F. A. Lootsma, ed.), Academic Press, pp. 371–382, 1972.
- [14] R. Fletcher, “Practical Methods of Optimization”, Wiley, 2.ed., 1987.
- [15] U. M. Garcia-Palomares, O. L. Mangasarian, “Superlinearly convergent quasi-Newton methods for nonlinearly constrained optimization problems”, *Mathematical Programming*, vol. 11, no. 1, pp. 1–13, 1976.
- [16] L. Grippo, F. Lampariello, S. Lucidi, “A nonmonotone line search technique for Newton’s method”, *SIAM Journal on Numerical Analysis*, vol. 23, no. 4, pp. 707–716, 1986.
- [17] S. P. Han, “Superlinearly convergent variable metric algorithms for general nonlinear programming problems”, *Mathematical Programming*, vol. 11, no. 1, pp. 263–282, 1976.
- [18] S. P. Han, “A globally convergent method for nonlinear programming”, *Journal of Optimization Theory and Applications*, vol. 22, no. 3, pp. 297–309, 1997.
- [19] J. Herskovits, “A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization”, *Mathematical Programming*, vol. 36, no. 1, pp. 19–38, 1986.

- [20] J. Herskovits, “Feasible Directions Interior Point Technique For Nonlinear Optimization”, *Journal for Optimization Theory and Applications*, vol. 99, no. 1, pp. 121–146, 1998.
- [21] M. R. Hestenes, “Optimization Theory, The finite dimensional case”, John Wiley, 1975.
- [22] A. F. Izmailov, M. V. Solodov, “Examples of Dual-Behavior of Newton-Type Methods on Optimization Problems with Degenerated Constraints”, *Computational Optimization and Applications*, vol. 42, pp. 231–264, 2009.
- [23] A.F. Izmailov, M.V. Solodov, “Newton-type methods for optimization problems without constraint qualifications”, *SIAM Journal on Optimization*, vol. 15, no. 1, pp. 210–228, 2004.
- [24] A.F. Izmailov, M.V. Solodov, “On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions”, *Mathematical Programming*, vol. 117, no. 1, pp. 271–304, 2009.
- [25] D. G. Luenberger, “Linear and Nonlinear Programming”, 2.ed., Springer, 2005.
- [26] J. M. Martínez, “Otimização Prática usando o Lagrangiano Aumentado”, Opúsculo, Departamento de Matemática Aplicada, Universidade Estadual de Campinas, Brasil, 2006.
- [27] J. M. Martínez, S. A. Santos, “Métodos Computacionais de Otimização”, Sociedade Brasileira de Matemática, 1995.
- [28] D. Q. Mayne, E. Polak, “A superlinearly convergente algorithm for constrained optimization problems”, *Mathematical Programming Study*, vol. 16, pp. 45–61, 1982.
- [29] J. Nocedal, S. J. Wright, “Numerical Optimization”, Springer-Verlag, 1999.
- [30] E. R. Panier, A. L. Tits, “A superlinearly convergent feasible method for the solution for inequality constrained optimization problems”, *SIAM Journal on Control and Optimization*, vol. 25, no. 4, pp. 934–950, 1987.
- [31] E. R. Panier, A. L. Tits, “On combining feasibility, descent and super-linear convergence in inequality constrained optimization”, *Mathematical Programming*, vol. 59, no. 2, pp. 133–278, 1993.

- [32] E. R. Panier, A. L. Tits, J. Herskovits, “A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization”, *SIAM Journal on Control and Optimization*, vol. 26, no. 4, pp. 788–811, 1988.
- [33] E. Polak, “Optimization: Algorithms and Consistent Approximations”, *Applied Mathematical Sciences*, vol. 124, 1997.
- [34] L. Qi, Z. Wei, “On the constant positive linear dependence condition and its applications to SQP methods”, *SIAM Journal on Optimization*, vol. 10, no. 4, pp. 963–981, 2000.
- [35] S. M. Robinson, “Generalized equations and their solutions, part II: Applications to nonlinear programming,”, *Mathematical Programming Study*, vol. 19, pp. 200–221, 1982.
- [36] M. L. Schuverdt, “Métodos de Lagrangiano Aumentado com Convergência Utilizando a Condição de Dependência Linear Positiva Constante”, Tese de Doutorado, Universidade Estadual de Campinas, 2006.
- [37] R. A. Tapia, “Quasi-Newton methods for equality constrained optimization: Equivalence of existing methods and a new implementation”, *Nonlinear Programming 3* (O. Mangasarian, R. Meyer e S. Robinson, eds.), pp. 125–164, 1978.
- [38] R. B. Wilson, “A simplicial algorithm for concave programming”, Thesis PhD, Universidade de Harvard, 1963.
- [39] S. J. Wright, “Superlinear convergence of a stabilized SQP method to a degenerate solution”, *Computational Optimization and Applications*, vol. 11, no. 3, pp. 253–275, 1998.
- [40] S. J. Wright, “Constraint identification and algorithm stabilization for degenerate nonlinear programs”, *Mathematical Programming*, vol. 95, no. 1, pp. 137–160, 2003.
- [41] J. L. Zhou, A. L. Tits, “Nonmonotone Line Search for Minimax Problems”, *Journal of Optimization Theory and Applications*, vol. 76, no. 3, pp. 455–476, 1993.

-
- [42] J. L. Zhou, A. L. Tits, C. T. Lawrence, “ User’s Guide for FFSQP Version 3.7: A FORTRAN Code for Solving Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality and Linear Constraints”, *Institute for Systems Research*, University of Maryland, College Park, MD 20742, 1997.