

Programação quadrática sequencial e condições de qualificação

Fernanda Teles Nunes *
Maria Aparecida Diniz-Ehrhardt
DMA - IMECC - Unicamp

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.

*Este autor foi financiado pelo CNPq

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 qualifications, 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 constraint qualifications are stronger than those based on stronger constraint qualifications. Numerical experiments will be done with the purpose of investigating the efficiency of these methods to solve problems with different constraint qualifications 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.

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 [19] utilizando a condição de dependência linear positiva constante, a qual provou-se, no mesmo trabalho, ser uma condição de qualificação das restrições. 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 principal objetivo é o estudo de algoritmos de PQS e resultados baseados em diferentes condições de qualificação.

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 converjam 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.

Algumas propriedades dos métodos de PQS são relevantes. Primeiro, trata-se de um método que tem a capacidade de caminhar 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 dependendo da construção dos subproblemas quadráticos.

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. O decréscimo no valor desta última 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.

As funções de mérito desempenham um papel fundamental na análise de convergência global dos algoritmos de PQS. 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 [4, 8, 3, 5]. Já a busca linear não-monótona parece ter sido apresentada inicialmente por Grippo, Lampariello e Lucidi [9] 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. [6].

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.

Na Seção 2 descreveremos brevemente alguns conceitos fundamentais da minimização restrita e apresentaremos diferentes condições de qualificação, assim como algumas relações existentes entre elas. Na terceira seção trataremos do método clássico de Programação Quadrática Sequencial e da atuação da busca não monótona contra o Efeito Maratos. Na Seção 4 apresentaremos um método de PQS factível e resultados de convergência para diferentes condições de qualificação. Por fim, na quinta e última Seção, seguem alguns testes numéricos realizados para comparar a eficiência de algoritmos de PQS utilizando busca linear monótona e não-monotona.

2 Condições de Qualificação

Iniciemos esta seção com alguns conceitos básicos da otimização não-linear. Seja

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0 \\ & g(x) \leq 0, \end{aligned} \tag{1}$$

um problema de otimização não-linear com restrições de igualdades e desigualdades, onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo do problema, e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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*.

A “melhor” solução que podemos encontrar para (1) é um ponto x^* tal que $f(x^*) \leq f(x)$ para todo $x \in D$, isto é, uma *solução global*. No entanto, a grande maioria dos algoritmos de minimização contenta-se com soluções mais modestas. 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 $\|x - x^*\| < \epsilon$.

Buscando propriedades que um minimizador local deve satisfazer, somos levados às condições de *Karush-Kuhn-Tucker*. Algebricamente, um ponto KKT do problema (1) é um ponto $x^* \in D$ para o qual existem vetores $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}_+^p$ tais que

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) &= 0, \\ h(x^*) &= 0, \quad g(x^*) \leq 0, \\ \mu_j^* g_j(x^*) &= 0, \quad j = 1, \dots, p. \end{aligned} \tag{2}$$

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 e gostaríamos que se constituíssem em *Condições Necessárias de Otimalidade*. 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 [14]:

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 (1) que quando satisfeita em um minimizador local x^* implica que x^* verifica (2).

Podemos então definir as condições necessárias de primeira ordem para um minimizador local do problema (1).

Teorema 2.1 (CN1) *Seja x^* um minimizador local do problema (1) e suponhamos 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 [12].

Por fim, denotamos 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 .

O Exemplo 2.1 mostra que, em alguns casos,

$$\text{minimizador local} \not\Rightarrow \text{condições KKT},$$

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

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

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,

$$\text{minimizador local} \Rightarrow \text{ponto factível},$$

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 [14].

2.1 Regularidade, Mangasarian-Fromovitz e Posto Constante

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

Definição 2.1 (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 2.2

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

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 2.2 (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,

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

Definição 2.3 (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

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

2.2 A condição de dependência linear positiva constante

Em [20], 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*). Na ocasião, 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 na próxima seção. Seguem algumas definições.

Definição 2.4 (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) [20] 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).

Definição 2.5 (CPLD) Dizemos que $x \in D$ satisfaz a condição CPLD [19] 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 $y \in N(x)$.

A prova de que a CPLD é uma condição de qualificação encontra-se em [19]. 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 [19] 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:

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

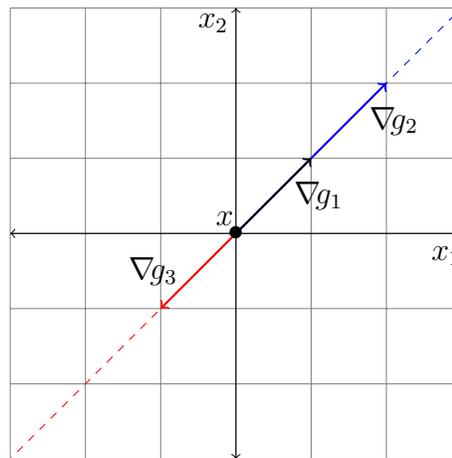


Figura 1: Uma análise geométrica de LICQ, MFCQ e CPLD.

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 1). 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 (2.2) sempre que $g_3(x)$ for uma restrição do tipo $g_3(x) = -\epsilon x_1 - \epsilon x_2$, $\epsilon > 0$.

3 O método de 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, frequentemente, em estratégias que buscam simplificar estes problemas complicados. O método de *Programação Quadrática Sequencial* consiste em resolver uma sequência de problemas mais simples cujas soluções converjam 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, estratégia esta usada pelo tão conhecido método de Newton no cálculo de zeros de funções. A fim de simplificar as discussões e análises subsequentes, vamos supor a partir de agora, nesta seção, que nosso problema não-linear contém apenas restrições de igualdade.

3.1 O subproblema quadrático

Dado um problema na forma

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & h(x) = 0, \end{aligned} \tag{3}$$

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:

$$\begin{aligned} \min \quad & \frac{1}{2}d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a.} \quad & \nabla h(x^k)^T d + h(x^k) = 0. \end{aligned} \tag{4}$$

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), 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

$$\begin{aligned} B_k d^k + \nabla f(x^k) + \nabla h(x^k) \lambda^k &= 0 \\ \nabla h(x^k)^T d^k + h(x^k) &= 0. \end{aligned} \quad (5)$$

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) em vez de (3). As condições KKT de (4) constituem um sistema linear, portanto de resolução mais barata que o sistema não-linear resultante das condições KKT do problema (3). O sistema linear reescrito fica:

$$\begin{aligned} B_k d^k + \nabla h(x^k) \lambda^k &= -\nabla f(x^k) \\ \nabla h(x^k)^T d^k &= -h(x^k). \end{aligned} \quad (6)$$

Diversos são os recursos que podem ser utilizados na resolução de (6). No entanto, não há garantia de que tal sistema tenha sempre solução e sem solução para (6) nossa idéia de tomar $x^{k+1} = x^k + d^k$ não pode seguir adiante.

Pensemos no problema (4). 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) como

$$\begin{aligned} \min \quad & \frac{1}{2} d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a} \quad & \nabla h(x^k)^T d + h(x^k) = \Psi(x_{nor}^k). \end{aligned} \quad (7)$$

Note que se (4) é factível, x_{nor}^k é tal que $\Psi(x_{nor}^k) = 0$ e, portanto, as soluções de (4) e (7) coincidem. Com isso garantimos a factibilidade dos subproblemas e, ainda mais, sem alterá-los nos casos em que já forem factíveis [13].

Uma vez determinada uma direção d^k que resolve (7), 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 (3), 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.

3.2 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 (3):

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

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

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

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 3.1 (CN2) *Seja x^* um minimizador local de (3) 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 [12].

Suponhamos que x^* seja um minimizador local do problema (3) 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 (3):

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

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 previamente, 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 (3) encontram-se no teorema a seguir:

Teorema 3.2 (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 (3).*

Demonstração: Ver Luenberger [12].

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 [3], 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,

$$\begin{bmatrix} \nabla^2 \ell(x^k, \lambda^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} d^k \\ p^k \end{bmatrix} = - \begin{bmatrix} \nabla \ell(x^k, \lambda^k) \\ h(x^k) \end{bmatrix}. \quad (8)$$

Daí, reescrevendo o sistema acima, ficamos com

$$\begin{aligned}\nabla^2 \ell(x^k, \lambda^k) d^k + \nabla h(x^k) p^k + \nabla \ell(x^k, \lambda^k) &= 0, \\ \nabla h(x^k)^T d^k + h(x^k) &= 0,\end{aligned} \quad (9)$$

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 (9), finalmente, da seguinte forma

$$\begin{aligned}\nabla^2 \ell(x^k, \lambda^k) d^k + \nabla f(x^k) + \nabla h(x^k) \lambda^{k+1} &= 0, \\ \nabla h(x^k)^T d^k + h(x^k) &= 0.\end{aligned} \quad (10)$$

Por outro lado, se considerarmos o subproblema (4) e o sistema (5) que descreve as condições KKT do mesmo, é possível perceber que (5) e (10) 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 $k \in \mathbb{N}$.

Exemplo 3.1

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

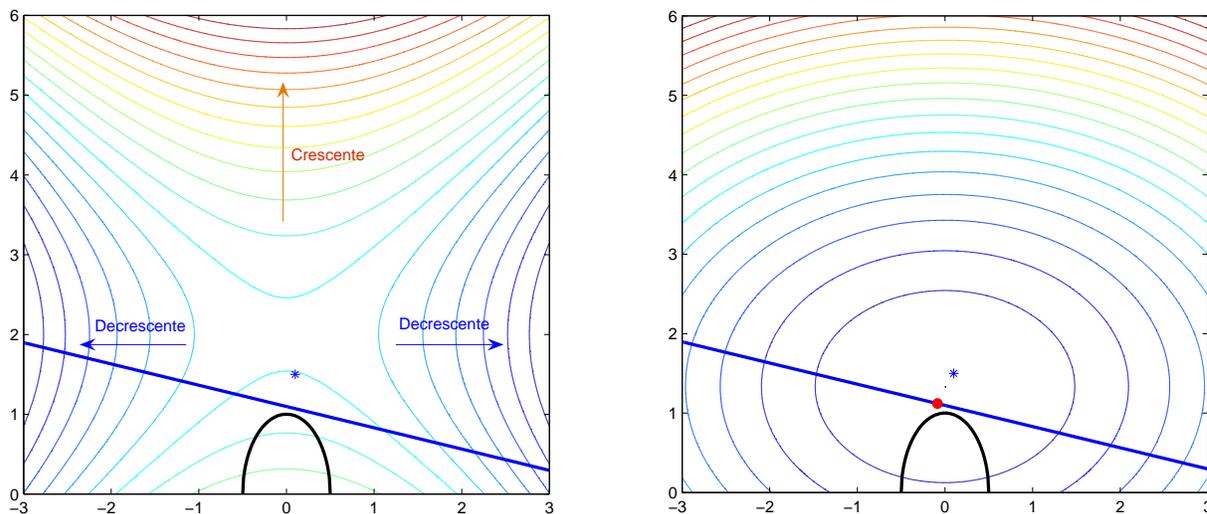


Figura 2: À esquerda, o subproblema quadrático com $B_k = \nabla^2 f(x^k)$ e à direita com $B_k = \nabla^2 \ell(x^k, -0.5)$, onde $x^k = (0.1, 1.5)$. A solução é $x^* = (0, 1)^T$ com multiplicador $\lambda^* = -1$.

Aplicando o método de programação quadrática sequencial ao problema do Exemplo 3.1, 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 2. 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 bolhinha 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.

3.3 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 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 [9], 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 [15] 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 [17] uniram estas duas estratégias, pedindo que o passo $t_k \in (0, 1)$ satisfaça

$$f(x^k + t_k d^k + t_k^2 \tilde{d}^k) \leq \max_{0 \leq i \leq M} \{f(x^{k-i})\} - \alpha t_k (d_0^k)^T B_k d_0^k, \quad (11)$$

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 (11), será realizado nas próximas seções. 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.

4 Um método factível de PQS e condições de qualificação

Definição 4.1 Dizemos que $\{x^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ é uma sequência de pontos KKT aproximados de (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:

$$\begin{aligned}
\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) &= \epsilon^k, \\
\|h(x^k)\| &\leq \eta_k, \\
g(x^k) &\leq \delta^k, \\
\mu^k &\geq 0, \\
(\mu^k)^T (g(x^k) - \delta^k) &= 0.
\end{aligned} \tag{12}$$

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 (1). Qi e Wei mostraram em [20] que se uma sequência deste tipo converge para um ponto que satisfaz a condição de dependência linear positiva constante introduzida na Seção 3, então este é um ponto KKT para o problema original (1). Vejamos o teorema a seguir:

Teorema 4.1 *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 (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 (1).

A teoria de álgebra linear nos permite, para cada $k \in \mathbb{K}$, escolher multiplicadores (λ^k, μ^k) que satisfazem (12) e tais que os vetores

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

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 (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 seqüência $\{(\lambda^k, \mu^k) / \|(\lambda^k, \mu^k)\|\}_{k \in \mathbb{K}}$ 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 $\{(\lambda^k, \mu^k)\}_{k \in \mathbb{K}}$ assuma uma subsequência convergente, de onde segue que x^* é um ponto KKT para o problema (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 (1), e mesmo nos casos em que isto for verdade não há garantia de que ele seja um minimizador estrito.

Na Seção 3 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 [18] 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) 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 4.2 *Suponha que x seja um ponto KKT do problema (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 [18].*

Em [20] Qi e Wei demonstraram que se x^* é um ponto KKT de (1) que satisfaz a CPLD e a SSOSC, então x^* é um ponto KKT isolado de (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 [19] 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.

Antes do próximo teorema, vamos descrever um método de PQS geral. O algoritmo aparece na Seção 4 de [20].

Algoritmo 4.1 *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 subproblema 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.*

Assumindo que em toda iteração o subproblema quadrático possui solução, o bom desempenho do Algoritmo (4.1) 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, é fácil verificar que x^k é um ponto KKT do problema (1). Para os casos em que o Algoritmo (4.1) gera uma sequência infinita temos o seguinte resultado:

Teorema 4.2 *Assuma que o Algoritmo (4.1) 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*

$$\|B_k\| \leq C_1. \quad (13)$$

Se

$$\liminf_{k \in \mathbb{K}} \|d^k\| = 0, \quad (14)$$

então x^* é um ponto KKT de (1) [20].

Demonstração: Ver [16].

4.1 Um método de PQS factível e MFCQ

Até agora, nesta seção, consideramos problemas de programação não-linear no formato (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

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & g(x) \leq 0, \end{aligned} \quad (15)$$

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

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 (15), resolve-se o subproblema

$$\begin{aligned} \min \quad & \frac{1}{2} d^T B_k d + \nabla f(x^k)^T d \\ \text{s.a} \quad & \nabla g_j(x^k)^T d + g_j(x^k) \leq 0, \quad j \in \{1, \dots, p\}. \end{aligned} \quad (16)$$

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 (16), 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 $d^k = 0$.

Iremos expor, a partir de agora, o método factível de PQS proposto por Panier e Tits em 1993 [17], assim como os resultados de convergência apresentados por Qi e Wei em [20], com algumas modificações no conjunto de hipóteses do Teorema 4.3 e da Proposição 4.1.

Algoritmo 4.2 *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 $\theta_k = \frac{\|d_0^k\|^2}{1 + \|d_0^k\|^2}$, d_1^k a solução de

$$\min \left\{ \frac{1}{2} \|d\|^2 + \max\{\nabla f(x^k)^T d; \max\{g_j(x^k) + \nabla g_j(x^k)^T d \mid j \in \{1, \dots, p\}\}\} \right\} \quad (17)$$

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.

Assumiremos as hipóteses abaixo:

H1: O conjunto factível de (15) é não vazio.

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

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 4.3 *Assuma que as hipóteses **H1-H4** são satisfeitas. Então x^* é um ponto KKT de (15).*

O teorema a seguir foi provado por Qi e Wei em [20] 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 [17] por Panier e Tits trocando a hipótese **H4** pela regularidade em todos os pontos factíveis de (15). Mas, conforme já citado anteriormente nesta mesma seção, **H4** implica na existência de uma solução contínua de (17) para cada x^k , o que permitiu provar o Teorema 4.3 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 (17) para cada x^k , como requer o algoritmo 4.2.

Considerações e comentários mais detalhados a respeito do *passo a passo* do Algoritmo 4.2 podem ser encontrados em [16].

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 4.1 *Assuma que **H1-H6** são satisfeitas e $(x^k)_{k \in \mathbb{N}}$ é gerada pela Algoritmo 4.2. Então toda a sequência converge para x^* , isto é,*

$$\lim_{k \rightarrow \infty} x^k = x^*. \quad (18)$$

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 4.3 quanto a da Proposição 4.1 são, em essência, as mesmas apresentadas em [20]. Neste mesmo trabalho, os autores apresentaram uma versão modificada do Algoritmo 4.2 e provaram convergência superlinear a dois passos da mesma, sob hipóteses levemente mais fracas que a regularidade.

5 Experimentos Numéricos

A relevância dos resultados apresentados na Seção 4 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). Estes problemas foram retirados do trabalho de Izmailov e Solodov [10] e a rotina utilizada nos testes foi a FFSQP, *Fortran Feasible Sequential Quadratic Programming* [21], que consiste na

implementação do método factível proposto por Panier e Tits em [17]. 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 4.2 apresentado na Seção 4, 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 Subseção 3.3, uma busca linear não-monótona atua contra a ocorrência do Efeito Maratos nos métodos de programação quadrática sequencial.

Para analisar o desempenho dos algoritmos na resolução de problemas com este tipo de solução, buscamos exemplos de dimensão pequena em [10] e [19], 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 [11]

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

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).

Neste exemplo o algoritmo FFSQP-AL chegou ao número máximo de iterações (500) antes de atingir a precisão pedida. Já FFSQP-NL alcançou o resultado em 28 iterações com norma do vetor KKT bem pequena.

Exemplo 2 [2]

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

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).

Nenhum dos dois algoritmos conseguiu encontrar a solução para o problema do Exemplo 2, 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.

Exemplo 3 [1]

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

onde

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

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).

O problema do Exemplo 3 possui as mesmas características daquele do Exemplo 2, mas neste caso FFSQP-NL conseguiu encontrar a solução enquanto que FFSQP-AL novamente atingiu o número máximo de iterações antes que o critério de parada fosse satisfeito.

Exemplo 4 [10]

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

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.

Neste quarto exemplo, 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.

Exemplo 5 [19]

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

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.

Na resolução deste último exemplo, ambos os algoritmos obtiveram sucesso.

Estes resultados sugerem que o método de programação quadrática sequencial factível, implementado nos algoritmos FFSQP-AL e FFSQP-NL, 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.

5.1 O problema da montanha

Com base nos resultados dos testes preliminares realizados, 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

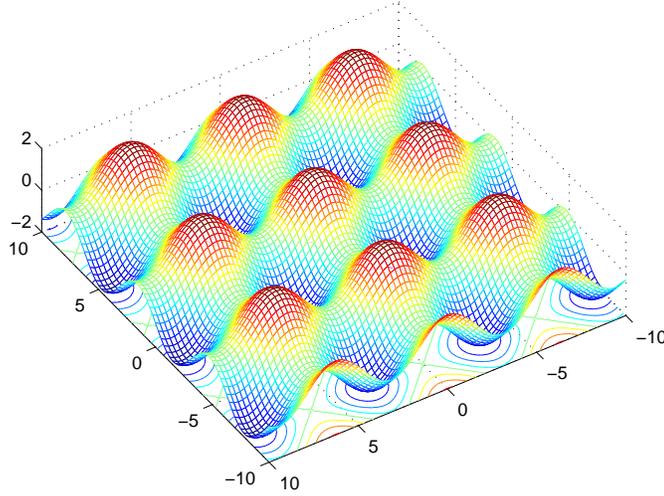


Figura 3: Superfície $S_2(x, y) = \text{sen}(x) + \text{cos}(y)$.

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 $S_1(x, y) = \text{sen}(x \cdot y) + \text{sen}(x + y)$ e $S_2(x, y) = \text{sen}(x) + \text{cos}(y)$, esta última representada na Figura 3.

Resolvemos o problema da montanha variando as dimensões do mesmo (entre 21 e 1001 variáveis) e comparamos a performance dos algoritmos FFSQP-AL e FFSQP-NL através da ferramenta “*performance profile*” introduzida por Dolan e Moré [7]. 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

$$r_{s,p} = \frac{\text{nit}_{s,p}}{\min\{\text{nit}_{s,p} \mid \forall s \in S\}},$$

com n_p igual ao número total de problemas e $\text{nit}_{s,p}$ 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.

Estabelecendo como critério de parada 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} e considerando o número de iterações gastas pelos algoritmos FFSQP-AL e FFSQP-NL como medida, obtivemos a comparação de desempenho apresentada na Figura 4. Esta nos mostra que FFSQP-AL não resolve nenhum problema (0%) com o número mínimo de iterações e resolve 10% deles com no mínimo duas vezes mais iterações que FFSQP-NL. 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 4. Na grande maioria dos casos FFSQP-NL apresentou um desempenho melhor que FFSQP-AL no que diz respeito ao número de iterações e avaliações de função.

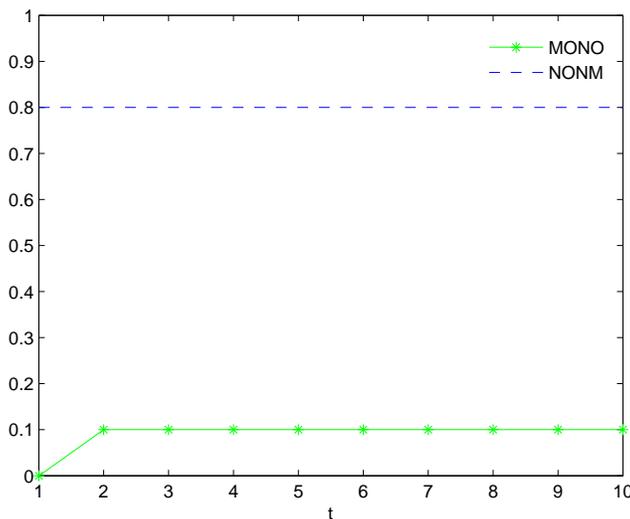


Figura 4: Perfil de desempenho dos algoritmos FFSQP-AL e FFSQP-NL com relação ao número de iterações gastas na resolução dos problemas com superfície $S(x, y) = \sin(x) + \cos(y)$.

Os testes também evidenciaram a ocorrência do Efeito Maratos e como FFSQP-NL parece agir contra o mesmo. Em alguns casos FFSQP-AL atinge o número máximo de iterações antes que algum dos critérios de parada sejam satisfeitos. Em outros casos, detecta-se que o novo iterando é equivalente ao anterior, no entanto também não se satisfazem os dois critérios de parada.

Especificamente para a superfície $S_2(x, y) = \text{sen}(x) + \text{cos}(y)$, nos problemas resolvidos 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-NL 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.

6 Conclusão

Neste trabalho apresentamos 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. 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. A resolução de problemas com poucas variáveis, nos quais conseguimos identificar as condições de qualificação satisfeitas na solução, confirmou a boa performance do método em aplicações envolvendo conjuntos viáveis formados por restrições de desigualdade. Já com problemas de dimensões maiores comparamos o comportamento de dois diferentes tipos de busca, monótona e não-monótona, e concluímos 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 nos métodos de PQS.

Referências

- [1] 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.
- [2] 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.
- [3] D. P. Bertsekas “Constrained Optimization and Lagrange Multiplier Methods”, Academic Press, 1982.
- [4] P. T. Boggs, J. W. Tolle, “Sequential quadratic programming”, *Acta Numerica*, pp. 1–51, 1995.

-
- [5] 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.
- [6] 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.
- [7] E. D. Dolan, J. J. Moré, “Benchmarking Optimization Software with Performance Profiles”, *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] D. G. Luenberger, “Linear and Nonlinear Programming”, 2.ed., Springer, 2005.
- [13] J. M. Martínez, S. A. Santos, “Métodos Computacionais de Otimização”, Sociedade Brasileira de Matemática, 1995.
- [14] J. M. Martínez, “Otimização Prática Usando o Lagrangiano Aumentado”, Opúsculo, Departamento de Matemática Aplicada, Universidade Estadual de Campinas, Brasil, 2006.
- [15] D. Q. Mayne, E. Polak, “A superlinearly convergent algorithm for constrained optimization problems”, *Mathematical Programming Study*, vol. 16, pp. 45–61, 1982.
- [16] F. T. Nunes, “Programação Quadrática Sequencial e Condições de Qualificação”, Dissertação de Mestrado, Universidade Estadual de Campinas, 2009.
- [17] E. R. Panier, A. L. Tits “On combining feasibility, descent and superlinear convergence in inequality constrained optimization”, *Mathematical Programming*, vol. 59, no. 2, pp. 133–278, 1993.
- [18] S. M. Robinson, “Generalized equations and their solutions, part II: Applications to nonlinear programming”, *Mathematical Programming Study*, vol. 19, pp. 200–221, 1982.

-
- [19] 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.
- [20] 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.
- [21] 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.