UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS

Condições de qualificação para problemas de minimização

TRABALHO DE CONCLUSÃO DE CURSO

Renan Diogo Manfron

Florianópolis, 2012

Renan Diogo Manfron

Condições de qualificação para problemas de minimização

Trabalho de Conclusão de Curso apresentado ao Curso de Matemática do Departamento de Matemática do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina para obtenção do grau de Bacharel em Matemática e Computação Científica.

Orientador:
Juliano de Bem Francisco

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Florianópolis

2012

Esta monografia foi julgada adequada como TRABALHO DE CONCLUSÃO DE CURSO no Curso de Matemática - Habilitação Bacharelado e Computação Científica, e aprovada em sua forma final pela Banca Examinadora designada pela Portaria nº04/CCM/2012.

Prof. Nereu Estanislau Burin
Professor da disciplina

Banca Examinadora:

Prof. Juliano de Bem Francisco
Orientador

Profa. Melissa Weber Mendonça

Prof. Luciano Bedin

Agradecimentos

Agradeço à toda minha família pela força, apoio e carinho, tornando possível este importante passo em minha vida.

Ao meu orientador Prof.Juliano de Bem Francisco, pelas valiosas orientações que me guiaram durante o curso e a execução deste trabalho, pela dedicação e preocupação.

Aos professores membros da Banca Profa. Melissa Weber Mendonça e Prof. Luciano Bedin pela disposição em participar da banca deste trabalho.

Agradeço também a todos os meus professores, que me proporcionaram conhecimentos e incentivos para a realização deste trabalho e também ao meu futuro profissional e pessoal. Obrigado a todos.

Aos meus amigos e colegas que me ajudaram nos momentos difíceis e por tornarem a Universidade um lugar mais agradável. Vocês ficarão sempre na minha memória.

Sumário

Referências Bibliográficas **42**

Capítulo 1

Introdução

Em problemas da física, química, engenharia, economia e tantas outras áreas do conhecimento, faz-se necessária a diminuição de custos e/ou maximização de lucros. Nessa linha, surge o processo de otimização, que consiste em definir variáveis, quantificá-las através de relações, descrever seu comportamento e por fim determinar o objetivo a ser otimizado. Por este processo, chegamos a um modelo matemático para o problema, donde necessita-se de uma análise para verificar a efetividade - usando ferramentas matemáticas e observação do comportamento do modelo. Com isso, chega-se a uma conclusão a respeito do problema, a qual resultará numa inferência dada pelo resultado do problema. Nota-se, como num processo de planejamento, que se deve realizar uma avaliação do modelo, testando sua solução e fazendo as devidas alterações. Nesse sistema descrito, surge a Otimização Matemática, assunto abordado neste trabalho.

A Otimização Matemática é um conjunto de técnicas e métodos matemáticos para auxiliar a tomada de decisões nas atividades de organizações. Um problema de Otimização implica na busca de uma solução mais adequada, dentro de algum critério, entre diversas soluções alternativas, estabelecendo assim critério de identificação de solução, o qual chamamos função objetivo.

Trazendo para linguagem matemática, temos uma função f em espaços variados, chamada função objetivo, definida no conjunto de restrições, digamos Ω. Dependendo da estrutura de Ω, temos diferentes classes de problemas de otimização, para os quais uma variedade de métodos de solução tem sido desenvolvida.

Abordamos inicialmente problemas irrestritos, os quais podemos considerar $\Omega = \mathbb{R}^n$. Para tais problemas, obtemos resultados que minimizadores locais satisfazem ou que pontos devem satisfazer para que sejam candidatos a minimizadores locais. Vale destacar que tais problemas não ocorrem com tanta frequência na realidade, pois não se possui quantidades ilimitadas de produtos e de informações.

Em virtude do disposto acima, estudamos problemas com restrições, aqueles em que se têm certas condições a seguir e com maior aplicabilidade nas organizações. Em se tratando de restrições, decerto que focamos nas de igualdade, desigualdade e união destas. Enaltecendo esta teoria, cito o Teorema de Karush-Kuhn-Tucker, o qual coloca uma condição necessária para que um ponto seja um minimizador local. Verifica-se o teorema supra sempre que temos x^* um ponto minimizador local e mais uma condição de qualificação.

Estudamos, então, condições de qualificação, tais como Mangasarian-Fromovitz, posto constante, dependência linear positiva constante, pseudonormalidade, quasenormalidade. Discutimos implicações entre elas; demonstrando que a condição de qualificação de dependência linear positiva implica na de quasenormalidade; e exibindo contra-exemplos para mostrar que as respectivas recíprocas são falsas.

Por fim, apresentamos uma aplicação na análise de convergência de um algoritmo de Lagrangeano Aumentado usando a condição de qualificação de dependência linear positiva dada pela Definição 3.1.6.

Vale expor que fazemos o uso de técnicas iniciais de Análise Matemática, bem como Cálculo em várias variáveis, as quais são de extrema importância para chegarmos nos resultados.

Resumo

Otimização Matemática é um ramo interdisciplinar da Matemática Aplicada que faz uso de modelos matemáticos, estatísticos e de algoritmos na ajuda à tomada de decisões. Num problema de otimização, temos uma função objetivo e um conjunto de restrições, ambos relacionados às variáveis de decisão. No presente trabalho apresentamos a teoria relacionada a certos conjuntos de restrições e quais condições são necessárias e suficientes para que um ponto neste conjunto seja um minimizador local. O principal resultado nesse sentido é o Teorema de Karush-Kuhn-Tucker e provamo-lo sob as condições de qualificação para restrições lineares.

Capítulo 2

Condições de Otimalidade

Neste capítulo estudaremos métodos para minimizar funções diferenciáveis em conjuntos de $\mathbb{R}^n$. As condições de otimalidade são relações entre as derivadas da função objetivo e as derivadas das funções que definem as restrições. As condições necessárias devem ser obrigatoriamente satisfeitas por minimizadores, enquanto as condições suficientes, quando satisfeitas, asseguram que o ponto em consideração é um minimizador local.

Frequentemente, pontos limites de algoritmos são minimizadores, sobretudo quando o método trabalha ativamente diminuindo o valor da função objetivo em cada iteração. No entanto, garantir a condição de minimizador local costuma ser difícil.

A área de otimização desempenha um papel fundamental não só em matemática, mas também em economia, planejamento estratégico, análise de algoritmos, problemas de combinatória, criptografia, e muitos outros temas surpreendentemente distintos.

Comecemos pelo problema de otimização matemática com restrições arbitrárias, isto é, a solução deve estar num conjunto fechado $\Omega \subset \mathbb{R}^n$.

2.1 Minimização em conjuntos fechados arbitrários

Nesta classe de problemas sob conjuntos fechados arbitrários, consideraremos o seguinte problema

$$\text{minimizar } \ f(x), \ \text{ sujeito a } \ x \in \Omega, \tag{2.1}$$

em que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função continuamente diferenciável em $\mathbb{R}^n$ e $\Omega \subset \mathbb{R}^n$ é um conjunto fechado. Conhecendo a classe do problema a ser estudado, vamos à busca de um ponto x^* que quando aplicarmos à função f, esta seja menor do que quando aplicada em outros pontos em torno de x^*, o que nos leva à definição a seguir.

Definição 2.1.1. *Um ponto $x^* \in \Omega$ é dito minimizador local de f sobre Ω se $\exists\ \varepsilon > 0$ tal que $f(x) \geq f(x^*)$ sempre que $x \in \Omega$ e $|x - x^*| < \varepsilon$. Se $f(x) > f(x^*)$ sempre que $x \in \Omega$, $|x - x^*| < \varepsilon$ e $x \neq x^*$ para algum ε, então x^* é dito minimizador local estrito de f sobre Ω.*

Definição 2.1.2. *Um ponto $x^* \in \Omega$ é dito minimizador global de f sobre Ω se $f(x) \geq f(x^*)$, $\forall\ x \in \Omega$. Se $f(x) > f(x^*)$, $\forall\ x \in \Omega$ e $x \neq x^*$, então x^* é dito minimizador global estrito de f sobre Ω.*

Munidos da informação do que seria um ponto de minimizador local, podemos nos perguntar qual seria uma condição que todos os pontos minimizadores locais de uma função f têm em comum. Para responder esta pergunta, segue o teorema abaixo, o qual nos dá uma condição necessária para que um ponto x^* seja ponto minimizador local.

Definição 2.1.3. *Dizemos que $d \in \mathbb{R}^n$ é uma direção factível em relação ao conjunto Ω no ponto $\bar{x} \in \Omega$, quando existe $\varepsilon > 0$ tal que*

$$\bar{x} + td \in \Omega, \forall\ t \in [0, \varepsilon].$$

Teorema 2.1.4 (Condição necessária de primeira ordem). *Sejam Ω um subconjunto de $\mathbb{R}^n$ e $f \in C^1$ função em Ω. Se x^* é um minimizador local do Problema (2.1), então $\forall\ d \in \mathbb{R}^n$ direção factível em x^* dada pela Definição 2.1.3, temos que $\nabla f(x^*)^T d \geq 0$.*

Demonstração. Para todo α, em que $0 \leq \alpha \leq \bar{\alpha}$, e $\bar{\alpha}$ é o ε da Definição 2.1.3, o ponto $x(\alpha) = x^* + \alpha d \in \Omega$. Também, para $0 \leq \alpha \leq \bar{\alpha}$, defina a função $g(\alpha) = f(x(\alpha))$. Assim, g tem um mínimo local em $\alpha = 0$. Do cálculo, temos que

$$g(\alpha) - g(0) = g'(0)\alpha + o(\alpha),$$

em que $o(\alpha)$ denota um termo que tende a 0 mais rápido que α, isto é, $\lim\limits_{\alpha \to 0} \dfrac{o(\alpha)}{\alpha} = 0$. Agora, se $g'(0) < 0$, então para $\alpha > 0$ suficientemente pequeno, segue que $g'(0)\alpha + o(\alpha) < 0$ e assim $g(\alpha) - g(0) < 0$, ou seja, $g(\alpha) < g(0)$. O que contradiz o fato de $g(0)$ ser mínimo local.
Portanto, devemos ter $g'(0) = \nabla f(x^*)^T d \geq 0$, $\forall\ d \in \mathbb{R}^n$ direção factível.

■

Corolário 2.1.5. *Sejam $\Omega \subset \mathbb{R}^n$ e $f \in C^1$ função em Ω. Se x^* é um minimizador local do Problema (2.1) e se x^* é ponto interior de Ω, então $\nabla f(x^*) = 0$.*

Demonstração. Seja $x^* \in \Omega$ ponto minimizador local do Problema (2.1). Vamos supor que $\nabla f(x^*) \neq 0$ e considere $d = -\nabla f(x^*)$. Como x^* é ponto interior de Ω, para $d = -\nabla f(x^*)$, $\exists\, \lambda > 0$ suficientemente pequeno tal que $x^* + \lambda d \in \Omega$, logo d é direção factível. Dessa forma,

$$\nabla f(x^*)^T d = \nabla f(x^*)^T(-\nabla f(x^*)) = - \parallel \nabla f(x^*) \parallel^2 < 0.$$

O que contradiz a otimalidade de x^*. Portanto, $\nabla f(x^*) = 0$.

■

A seguir colocamos finalmente uma condição necessária para o problema irrestrito, isto é, $\Omega = \mathbb{R}^n$.

Corolário 2.1.6. *Seja x^* minimizador local de f em $\mathbb{R}^n$. Então $\nabla f(x^*) = 0$.*

Demonstração. Segue diretamente do Corolário 2.1.5.

Vimos, no Teorema 2.1.4, uma condição necessária sobre o gradiente da função f. Surge, então, o questionamento se há alguma relação entre pontos minimizadores locais e a hessiana da função f. O Teorema que se segue elucida essa questão.

Teorema 2.1.7 (Condição necessária de segunda ordem). *Sejam $\Omega \subset \mathbb{R}^n$ e $f \in C^2$ função em Ω. Se x^* é minimizador local do Problema (2.1), então para qualquer $d \in \mathbb{R}^n$ direção factível em x^* dada pela Definição 2.1.3, temos:*

1. $\nabla f(x^*)^T d \geq 0$,

2. *Se* $\nabla f(x^*)^T d = 0$, *então* $d^T \nabla^2 f(x^*) d \geq 0$.

Demonstração. 1. Segue do Teorema 2.1.4.

2. Vale que $\nabla f(x^*)^T d = 0$ por hipótese. Denotando uma função $x(\alpha) = x^* + \alpha d$ e $g(\alpha) = f(x(\alpha))$, obtemos, sabendo que $g'(0) = \nabla f(x^*) = 0$ pelo Corolário 2.1.5, que

$$g(\alpha) - g(0) = \frac{1}{2} g''(0)\alpha^2 + o(\alpha^2)$$

conforme Teorema de Taylor, em que $o(\alpha^2)$ denota um termo que tende a 0 mais rápido que α^2. Agora, se $g''(0) < 0$, para α suficientemente pequeno, segue que

$$\frac{1}{2} g''(0)\alpha^2 + o(\alpha^2) < 0 \text{ e daí } g(\alpha) < g(0),$$

ou seja, $f(x^* + \alpha d) < f(x^*)$, o que contradiz o fato de x^* ser minimizador local de f em Ω. Portanto, $g''(0) = (f(x(\alpha)))'' = (\nabla^T f(x^*) d)' = d^T \nabla^2 f(x^*) d \geq 0$.

■

Para prosseguirmos no estudo do Problema (2.1), faz-se necessário um conhecimento breve a respeito de curvas viáveis, para isto:

Definição 2.1.8. *Dado $x \in \Omega$, dizemos que γ é uma curva viável em Ω de classe C^k passando por x se $\gamma : [-\varepsilon,\ \varepsilon] \to \Omega$, $\varepsilon > 0$; $\gamma(0) = x$, $\gamma'(0) \neq 0$ e $\gamma \in C^k$.*

Conseguimos condições necessárias de primeira e segunda ordens análogas as dos Teoremas 2.1.4 e 2.1.7, respectivamente, envolvendo curvas viáveis. Demonstrá-los-emos nos próximos dois resultados.

Lema 2.1.9. *Se $x^* \in int(\Omega)$ é um minimizador local do Problema (2.1) e γ é uma curva viável em Ω de classe C^1 passando por x^* em Ω de classe C^1 passando por x^*, então $\nabla f(x^*)^T \gamma'(0) = 0$.*

Demonstração. Definimos $\gamma_1 : [0,\varepsilon] \to \Omega$ por $\gamma_1(t) = \gamma(t)$ e $\gamma_2 : [0,\varepsilon] \to \Omega$ por $\gamma_2(t) = \gamma(-t)$. Pelo Teorema 2.1.4, sabemos que

$$\nabla f(x^*)^T \gamma_1'(0) \geq 0 \quad \text{e} \quad \nabla f(x^*)^T \gamma_2'(0) \geq 0$$

Mas $\gamma_1'(0) = \gamma'(0)$ e $\gamma_2'(0) = -\gamma'(0)$, assim

$$\nabla f(x^*)^T \gamma'(0) \geq 0 \ \text{ e } \ -\nabla f(x^*)^T \gamma'(0) \geq 0.$$

Logo, $\nabla f(x^*)^T \gamma'(0) = 0$.

■

A partir de agora, denotaremos por $\nabla^2 f(x) \geq 0$ a matriz hessiana da função f e diz que esta é simétrica semi-definida positiva. Também, $\nabla^2 f(x) > 0$ caracteriza a matriz hessiana da função f como simétrica definida positiva.

Corolário 2.1.10 (Condição necessária de segunda ordem para x^* no interior de Ω). *Seja x^* minimizador local do Problema (2.1), sendo também ponto interior de Ω. Se f tem derivadas segundas contínuas numa vizinhança de x^*, então $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*) \geq 0$.*

Demonstração. Sejam $d \in \mathbb{R}^n$, $d \neq 0$ arbitrário e $\gamma : [-\varepsilon,\varepsilon] \to \Omega$ curva viável definida por $\gamma(t) = x^* + td$. Pelo Corolário 2.1.5 e Lema 2.1.9, sgabemos que $\nabla f(x^*)^T d \equiv \nabla f(x^*)^T \gamma'(0) = 0$. Agora, como d é arbitrário, segue que $\nabla f(x^*) = 0$.

Definindo $\varphi : [-\varepsilon,\varepsilon] \to \mathbb{R}$ por $\varphi(t) = f(\gamma(t))$, temos que $\varphi'(0) = \nabla f(x^*)^T \gamma'(0) = 0$ e pelo Teorema 2.1.7, $0 \leq \varphi''(0) = \gamma'(0)^T \nabla^2 f(x^*) \gamma'(0) = d^T \nabla^2 f(x^*) d$. Da arbitrariedade de d, concluímos que $\nabla^2 f(x^*) \geq 0$.

■

Até agora, conseguimos condições que um ponto x^* deve cumprir a fim de que seja candidato a minimizador local. No próximo Teorema, daremos condições que garantem que um ponto $x^* \in \text{int}(\Omega)$ seja minimizador local, caso as cumpra.

Teorema 2.1.11 (Condição suficiente de segunda ordem para x^* no interior de Ω). *Sejam $f \in C^2(\Omega)$ e x^* ponto interior de Ω tal que $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*) > 0$. Então x^* é minimizador local estrito do Problema (2.1).*

Demonstração. Escrevendo a expansão de Taylor para f em torno de x^*,

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + o(\| x - x^* \|^2),$$

pois $\nabla f(x^*) = 0$, em que $\lim\limits_{x \to x^*} \dfrac{o(\|x - x^*\|^2)}{\|x - x^*\|^2} = 0$ e $\| \, . \, \|$ é qualquer norma em $\mathbb{R}^n$. Agora, da hipótese $\nabla^2 f(x^*) > 0$, sendo $\alpha > 0$ o menor dos autovalores de $\nabla^2 f(x^*)$, obtemos que $\forall \, x \neq x^*$,

$$(x - x^*)^T \nabla^2 f(x^*)(x - x^*) \geq (x - x^*)^T \alpha (x - x^*) = \alpha \|x - x^*\|^2.$$

Logo, $f(x) \geq f(x^*) + \dfrac{\alpha}{2}\|x - x^*\|^2 + o(\|x - x^*\|^2)$ e para $x \neq x^*$,

$$\frac{f(x) - f(x^*)}{\|x - x^*\|^2} \geq \frac{\alpha}{2} + \frac{o(\|x - x^*\|^2)}{\|x - x^*\|^2}.$$

Mas o termo $\dfrac{o(\|x - x^*\|^2)}{\|x - x^*\|^2}$ tende a 0 quando $x \to x^*$. Consequentemente, para x suficientemente próximo e diferente de x^*, $\left| \dfrac{o(\|x - x^*\|^2)}{\|x - x^*\|^2} \right| < \dfrac{\alpha}{4}$. Portanto, $\dfrac{f(x) - f(x^*)}{\|x - x^*\|^2} \geq \dfrac{\alpha}{2} + (-\dfrac{\alpha}{4}) = \dfrac{\alpha}{4} > 0$. Assim, $f(x) > f(x^*)$, $\forall \, x$ numa vizinhança de x^*, $x \neq x^*$.

■

2.2 Problemas com restrições de igualdade

Nos problemas que surgem de experimentos físicos ou de engenharia, economia ou outra ciência, em geral faz-se necessário o uso de restrições sobre os dados de um certo problema, em virtude de ter-se quantidade limitada de informação, bem como material, dinheiro, dados em geral. Desta forma, seguimos nosso estudo através de problemas dessa classe.

Nesta classe com restrições de igualdade, consideraremos o seguinte problema

$$\text{minimizar } \ f(x), \ \text{ sujeito a } \ h(x) = 0, \tag{2.2}$$

em que $h : \mathbb{R}^n \to \mathbb{R}^m$ é uma função continuamente diferenciável em $\mathbb{R}^n$. Chamamos Ω ao conjunto factível do problema. Em nosso caso, $\Omega = \{x \in \mathbb{R}^n | \ h(x) = 0\}$.

Nesta seção, derivaremos as caracterizações matemáticas das soluções do Problema (2.2) conforme fizemos na Seção 2.1, discutiremos as condições de otimalidade sob dois aspectos, a saber, condições necessárias e suficientes para um ponto viável ser minimizador local. Ainda da Seção 2.1, obtemos os seguintes resultados:

Condições necessárias: $\nabla f(x^*)^T d \geq 0$ e, caso $\nabla f(x^*)^T d = 0$, então $d^T \nabla^2 f(x^*) d \geq 0$, com $d \in \mathbb{R}^n$ direção factível em x^* dada pela Definição 2.1.3.
Condições suficientes: um ponto que cumpra simultaneamente $\nabla f(x^*) = 0$ e $d^T \nabla^2 f(x^*) d > 0$ é minimizador local de f, com $d \in \mathbb{R}^n$ direção factível em x^*.

Definição 2.2.1. *Se $x \in \Omega$, chamamos o conjunto tangente a Ω em x, aqui denotado por $M(x)$, ao conjunto de todos os vetores tangentes às curvas viáveis em Ω passando por x, ou seja: $M(x) = \{v \in \mathbb{R}^n | \ v = \ \gamma'(0) \ \text{para alguma curva viável } \ \gamma \ \text{passando por } x\ \}$.*

Denotamos

$$h'(x) = \begin{pmatrix} \dfrac{\partial h_1}{\partial x_1}(x) & \cdots & \dfrac{\partial h_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \dfrac{\partial h_m}{\partial x_1}(x) & \cdots & \dfrac{\partial h_m}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} h_1'(x) \\ \vdots \\ h_m'(x) \end{pmatrix} = \begin{pmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{pmatrix}$$

Relacionaremos o conjunto tangente $M(x)$ com o núcleo do Jacobiano de $h(x)$, denotado por $N(h'(x))$, pelo lema a seguir.

Lema 2.2.2. $\forall \, x \in \Omega$, $M(x) \subset N(h'(x))$.

Demonstração. Sejam $v \in M(x)$ e $\gamma : [-\varepsilon, \varepsilon] \to \Omega$ tal que $\gamma'(0) = v$, $\gamma(0) = x$. Defina $\phi(t) = h(\gamma(t))$, $\forall \ t \in [-\varepsilon, \varepsilon]$. Logo, $\phi'(t) \equiv ({\phi_1}'(t), \ldots, {\phi_m}'(t))^T = 0$, $\forall \ t \in (-\varepsilon, \varepsilon)$. Ademais, pela regra da cadeia, $\phi'(t) = h'(\gamma(t))\gamma'(t) = 0$, $\forall \, t \in (-\varepsilon, \varepsilon)$. Portanto, $0 = h'(x)\gamma'(0) = h'(x)v$. Isto é, $v \in N(h'(x))$.

■

Observe que a recíproca desse resultado é falsa. De fato, um contra-exemplo é $h(x_1, x_2) =$

x_1x_2, $x=(0,0)^T$ e assim $M(x)=\{v\in\mathbb{R}^2|\ v_1v_2=0\}$, mas $h'(x)=(0,0)$ e, claramente, $N(h'(x))=\mathbb{R}$.

Consideraremos agora uma particularidade de um ponto x^*, hipótese esta que é conhecida como condição de qualificação para restrições lineares. Com ela, conseguimos caracterizações importantes de minimizadores locais.

Definição 2.2.3. *Dizemos que* $x\in\Omega\equiv\{x\in\mathbb{R}^n|\ h(x)=0\}$ *satisfaz à condição de qualificação de independência linear se o posto de* $h'(x)$ *é igual a m, isto é,* $\{\nabla h_1(x),\dots,\nabla h_m(x)\}$ *é um conjunto linearmente independente. Chamaremos esta condição de LICQ.*

Faremos o uso do Teorema da Função Implícita, conhecido do estudo de Análise Matemática, o qual omitiremos a demonstração por não ser do objetivo deste trabalho. No entanto, ressalta-se sua importância no campo da otimização matemática.

Teorema 2.2.4 (Teorema da Função Implícita). *Seja* $F:\mathbb{R}^k\times\mathbb{R}^m\to\mathbb{R}^m$ *uma função de classe* C^1. *Suponha que* $F(x_0,y_0)=0$ *e*

$$det\left[\frac{\partial F}{\partial y}(x_0,y_0)\right]\neq 0.$$

Então existem um aberto $W\subset\mathbb{R}^k$ *e* $\phi:W\to\mathbb{R}^k$ *função de classe* C^1 *tais que*

1. $x_0\in W$ *e* $\phi(x_0)=y_0$.

2. $F(x,\phi(x))=0,\ \forall x\in W$.

Demonstração. Encontra-se a demonstração no livro [3] ou outro texto de Análise.

■

Chegamos ao resultado que enunciamos anteriormente a respeito da relação entre o núcleo do Jacobiano de $h(x)$ e o conjunto tangente $M(x)$ no caso em que x é ponto que satisfaz LICQ.

Teorema 2.2.5. *Seja* $\Omega=\{x\in\mathbb{R}^n|\ h(x)=0\}$, $h\in C^k$, $x\in\Omega$ *um ponto que satisfaz LICQ. Então,* $\forall\ v\in N(h'(x))$, *existe uma curva viável de classe* C^k *passando por x tal que* $\gamma'(0)=v$. *Nesse caso,* $M(x)=N(h'(x))$.

Demonstração. Seja $v\in N(h'(x))$. Por definição temos que $h'(x)v=0$. Queremos encontrar uma curva viável γ em Ω passando por x tal que $\gamma'(0)=v$.

Consideramos o sistema de equações $h(x+tv+h'(x)^Tu)=0$, em que $u\in\mathbb{R}^m$, $t\in\mathbb{R}$. Para x e v fixos, esse é um sistema de m equações com $m+1$ variáveis e escolhendo $u=0$ e $t=0$, temos uma solução particular desse sistema. O Jacobiano de $h(x+tv+h'(x)^T u)$ em relação a u em $t=0$ é $h'(x)h'(x)^T \in \mathbb{R}^{m\times m}$ e é não singular, pois x cumpre LICQ, visto que $\{\nabla h_1(x),\ldots,\nabla h_m(x)\}$ é conjunto linearmente independente. Logo, pelo Teorema 2.2.4, existe $\bar{\gamma} \in C^k$ definida em $[-\varepsilon,\varepsilon]$, $\varepsilon > 0$, tal que $h(x+tv+h'(x)^T\bar{\gamma}(t)) = 0$.

Portanto, $h(x+tv+h'(x)^T\bar{\gamma}(t)) = 0$, $\forall\ t \in [-\varepsilon,\varepsilon]$. Derivando a expressão precedente em relação a t, para $t=0$, temos $h'(x)(v+h'(x)^T\bar{\gamma}'(0)) = 0$. Como $h'(x)v = 0$, segue que $h'(x)h'(x)^T\bar{\gamma}'(0) = 0$. Todavia, $h'(x)h'(x)^T$ é não singular e portanto $\bar{\gamma}'(0) = 0$. Consequentemente, definindo $\gamma : [-\varepsilon,\varepsilon] \to \Omega$ por $\gamma(t) = x+tv+h'(x)^T\bar{\gamma}(t)$, segue que

$$\gamma'(0) = v + h'(x)^T\bar{\gamma}'(0) = v.$$

Assim, γ é a curva viável procurada. Como v é arbitrário, temos que $N(h'(x)) \subset M(x)$. Do Lema 2.2.2, concluímos que $M(x) = N(h'(x))$.

■

O próximo resultado é importante no sentido de que o problema de minimização relaciona intimamente a função f e as restrições dadas por h.

Teorema 2.2.6. *Se x^* é minimizador local que cumpre LICQ colocada na Definição 2.2.3 com relação ao Problema (2.2), então $\nabla f(x^*) \perp N(h'(x^*))$.*

Demonstração. Seja $v \in N(h'(x^*))$, e então $v \in M(x^*)$ pois x^* satisfaz LICQ. Assim, existe γ em Ω passando por x^* tal que $\gamma'(0) = v$. Pelo Lema 2.1.9, $\nabla f(x^*)^T v = 0$, ou seja, $\nabla f(x^*) \perp N(h'(x^*))$.

■

Teorema 2.2.7 (Multiplicadores de Lagrange). *Se x^* é minimizador local que satisfaz LICQ com relação ao Problema (2.2), então existem únicos $\lambda_1,\ldots,\lambda_m$ reais tais que*

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

Neste caso, $\lambda_1,\ldots,\lambda_m$ são chamados multiplicadores de Lagrange do problema.

Demonstração. Pelo Teorema 2.2.6, $\nabla f(x^*) \perp N(h'(x^*))$. Logo, $\nabla f(x^*) \in R(h'(x^*)^T)$, isto é, existe $\lambda \in \mathbb{R}^m$ tal que $\nabla f(x^*) + h'(x^*)^T\lambda = 0$. Como x^* satisfaz LICQ, o Jacobiano $h'(x^*)$ tem posto completo e então esse vetor de multiplicadores é único.

■

Considerando os resultados obtidos para o Problema (2.2), os candidatos a minimizador local para esse problema serão pontos que cumprem LICQ e, ao mesmo tempo, sejam soluções do sistema não linear com $n+m$ equações e $n+m$ incógnitas:

$$\begin{cases} \nabla f(x) + h'(x)^T\lambda = 0 \\ h(x) = 0. \end{cases}$$

Esses pontos serão chamados de estacionários ou críticos. Naturalmente, os pontos que não cumprem LICQ com relação a Ω também seriam candidatos a minimizador local.

Definição 2.2.8. *Chamamos de lagrangeano do Problema (2.2) à função* $\ell(x,\lambda) = f(x) + h(x)^T\lambda$.

Até agora, conseguimos condições necessárias para que um ponto x^* seja minimizador local envolvendo apenas $\nabla f(x)$. A partir de agora, estamos interessados em que propriedade $\nabla^2 f(x)$ e $\nabla^2 h_i(x)$ devem cumprir neste caso, ou seja, analisar as informações de segunda ordem de f e cada $h_i(x)$.

Teorema 2.2.9 (Condições necessárias de segunda ordem para restrições de igualdade). *Vamos supor que* $f,\ h \in C^2$, x^* *é minimizador local que cumpre LICQ do Problema (2.2) e* λ *é o vetor de multiplicadores de Lagrange definido no Teorema 2.2.7. Então* $v^T\nabla^2{}_{xx}\ell(x^*,\lambda)v \geq 0,\ \forall\ v \in N(h'(x^*))$.

Demonstração. Do Teorema 2.2.7, sabemos que

$$\nabla f(x^*) + h'(x^*)^T\lambda = 0. \tag{2.3}$$

Seja $v \in N(h'(x^*))$ e então, pelo Teorema 2.2.5, existe uma curva viável γ em Ω de classe C^2 passando por x^*; satisfazendo $\gamma(0) = x^*$ e $v = \gamma'(0)$, também $\gamma'(0) \in N(h'(x^*))$.

Definindo $\varphi(t) = f(\gamma(t))$, pelo Lema 2.1.9, segue que $\varphi'(0) = \nabla f(x^*)^T\gamma'(0) = 0$ e usando o Teorema 2.1.7:

$$\varphi''(0) = \gamma'(0)^T\nabla^2 f(x^*)\gamma'(0) + \nabla f(x^*)^T\gamma''(0) \geq 0. \tag{2.4}$$

Agora, definindo $\phi_i(t) = \lambda_i h_i(\gamma(t))$, $i = 1,\ldots,m$:

$\phi_i'(t) = \lambda_i h_i'(\gamma(t))\gamma'(t) = 0$,

$\phi_i''(t) = \gamma'(t)^T\lambda_i\nabla^2 h_i(\gamma(t))\gamma'(t) + \lambda_i h_i'(\gamma(t))\gamma''(t) = 0$,

$\phi_i''(0) = \gamma'(0)^T\lambda_i\nabla^2 h_i(\gamma(0))\gamma'(0) + \lambda_i h_i'(\gamma(0))\gamma''(0) = 0$,

$\phi_i''(0) = \gamma'(0)^T\lambda_i\nabla^2 h_i(x^*)\gamma'(0) + \lambda_i h_i'(x^*)\gamma''(0) = 0$.

Logo,

$$\sum_{i=1}^{m} \phi_i''(0) = \gamma'(0)^T \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*) \gamma'(0) + \lambda^T h'(x^*) \gamma''(0) = 0. \tag{2.5}$$

Somando as Equações (2.4) e (2.5), obtemos que

$$\gamma'(0)^T [\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*)] \gamma'(0) + [\nabla f(x^*)^T + \lambda^T h'(x^*)] \gamma''(0) \geq 0.$$

Agora, da Equação (2.3), conseguimos $\nabla f(x^*)^T + \lambda^T h'(x^*) = 0$ e então

$$\gamma'(0)^T [\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*)] \gamma'(0) \geq 0.$$

De $\ell(x^*, \lambda) = f(x^*) + h(x^*)^T \lambda$ por definição e como $v = \gamma'(0)$ é qualquer, concluímos que

$$v^T \nabla^2{}_{xx} \ell(x^*, \lambda) v \geq 0, \ \forall \ v \in N(h'(x^*)).$$

■

Finalmente, podemos obter condições que garantem a minimalidade de um ponto x^*, envolvendo condições necessárias de primeira ordem, bem como os multiplicadores de Lagrange, o lagrangeano e núcleo do Jacobiano de h.

Corolário 2.2.10 (**Condições suficientes de segunda ordem para restrições de igualdade**). *Sejam f, $h \in C^2$, $x^* \in \Omega$ satisfazendo às condições necessárias de primeira ordem para o Problema (2.2), λ é o vetor de multiplicadores de Lagrange e $y^T \nabla^2{}_{xx} \ell(x, \lambda) y > 0$, $\forall \ y \in N(h'(x^*))$, $y \neq 0$. Então x^* é minimizador local estrito para o Problema (2.2).*

Demonstração. Suponha que x^* não é minimizador local do Problema (2.2). Assim, existe uma sequência $\{y_k\}_{k \in \mathbb{N}}$ com $\{y_k\} \to x^*$ e $f(y_k) < f(x^*)$, para cada $k \in \mathbb{N}$. Escreva $y_k = x^* + \delta_k s_k$, em que $\|s_k\| = 1$ e $\delta_k \to 0$.

Como $\{s_k\}$ é uma sequência num conjunto compacto $\bar{B}(0,1)$, segue que existe $K_1 \subseteq \mathbb{N}$ e $s^* \in \mathbb{R}^n$ com $\|s^*\| = 1$ tal que $\lim_{k \in K_1} s_k = s^*$.

Note que $0 = \lim_{k \in K_1} \dfrac{h(y_k) - h(x^*)}{\delta_k} = \lim_{k \in K_1} \dfrac{h(x^* + \delta_k s_k)}{\delta_k} = h'(x^*) s^*$, ou seja, $s^* \in N(h'(x^*))$.

Por Taylor, sabemos que $\forall \ i \in \{1, \ldots, n\}$

$$0 = \lambda_i [h_i(y_k) = \lambda_i (h_i(x^*)) + \delta_k \nabla h_i(x^*)^T s_k + \frac{\delta_k^2}{2} s_k^T \nabla^2 f(x^*) s_k] + o(\delta_k^2), \tag{2.6}$$

o que implica em

$$0 > f(y_k) - f(x^*) = \delta_k s_k^T \nabla f(x^*) + \frac{\delta_k^2}{2} s_k^T \nabla^2 f(x^*) s_k + o(\delta_k^2). \tag{2.7}$$

Usando as Equações (2.6) e (2.7) resulta que

$$0 > \delta_k s_k^T \left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \right) + \frac{\delta_k^2}{2} s_k^T \left(\nabla^2 f(x^*) + \sum_{i=1}^{n} \lambda_i \nabla^2 h_i(x^*) \right) s_k + o(s_k^2),$$

$$0 > s_k^T \left(\nabla^2 f(x^*) + \sum_{i=1}^{n} \lambda_i \nabla^2 h_i(x^*) \right) s_k + o(s_k^2),$$

pois o primeiro termo da expressão acima é nulo por hipótese. Consequentemente,

$$0 > s_k^T \ell_{xx}(x^*, \lambda) s_k + \frac{o(\delta_k^2)}{\delta_k^2}.$$

Agora, tomando limite para $k \in K_1$, concluímos que ${s^*}^T \nabla_{xx} \ell(x^*, \lambda) s^* < 0$. O que é uma contradição, porque $s^* \in N(h'(x^*))$.

■

2.3 Problemas com restrições de desigualdade

Nota-se que nos problemas do cotidiano, não temos somente restrições de igualdade. Muitas vezes, temos uma quantidade máxima de um produto a ser consumido ou número de horas trabalhadas, o que envolve exatamente restrições de desigualdade. Digamos que a quantidade máxima de insumo para um certo produto seja x^1, então deveremos ter a combinação necessária que depende desse insumo limitada por x^1, isto é, $\sum_{i=1}^{m} a_i x_i \leq x^1$. Vale ressaltar que podemos transformar essas restrições de desigualdade em igualdade introduzindo variáveis de folga e assim tendo um sistema de restrições com mais variáveis. Todavia, não adentraremos ao assunto por não abranger o objetivo deste trabalho.

Analogamente ao que fizemos na Seção 2.2, nesta seção estenderemos os resultados das condições necessárias e suficientes para problemas com restrições gerais de desigualdade. Os conceitos e resultados são análogos aos feitos anteriormente e até agora, para problemas do tipo (2.2), temos:

Condições necessárias: $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$ e $y^T \nabla^2{}_{xx} \ell(x^*, \lambda) y \geq 0, \ \forall\, y \in N(h'(x^*))$.

Condições suficientes: $y^T \nabla^2{}_{xx} \ell(x, \lambda) y > 0, \forall\, y \in N(h'(x^*)), y \neq 0$.

Consideraremos agora o problema de minimização com restrições gerais de desigualdade

$$\text{minimizar } f(x), \text{ sujeito a } c(x) \leq 0, \tag{2.8}$$

em que $c : \mathbb{R}^n \to \mathbb{R}^p$ é continuamente diferenciável.

Para este problema, para cada $x \in \Omega = \{x \in \mathbb{R}^n | \, c(x) \leq 0\}$, chamamos de restrições ativas em x aquelas para as quais $c_i(x) = 0$ e denotaremos por $I(x)$ o conjunto das restrições ativas em x. Analogamente, chamamos restrições inativas em x aquelas para as quais $c_i(x) < 0$.

Definição 2.3.1. *Como na Definição 2.2.3, chamaremos ponto que satisfaz à condição de qualificação para restrições lineares um ponto de Ω onde os gradientes das restrições ativas são linearmente independentes e chamaremos de ponto que satisfaz LICQ para restrições de desigualdade.*

A questão que surge é se há relação entre o problema tratado nesta seção e o das restrições de igualdade. Para resolver esse impasse, ao Problema (2.8) relacionaremos o Problema (2.2) através do seguinte Lema:

Lema 2.3.2. *Se x^* é minimizador local do Problema (2.8) e $I(x^*) = \{i \in \{1, \ldots, p\} | \, c_i(x^*) = 0\}$, então x^* é minimizador local do Problema*

$$\textit{minimizar } f(x), \textit{ sujeito a } c_i(x) = 0, \ i \in I(x^*). \tag{2.9}$$

Demonstração. Suponha que x^* não é minimizador local do Problema (2.9). Então, $\forall \ \varepsilon > 0$, existe $x_\varepsilon \in B(x^*, \varepsilon)$ tal que $f(x_\varepsilon) < f(x^*)$ e $c_i(x_\varepsilon) = 0, \ \forall \ i \in I(x^*)$. Também, $c_i(x^*) = 0 \quad \forall \ i \in I(x^*)$ e $c_i(x^*) < 0 \quad \forall \ i \in \{1, \ldots, p\} \backslash I(x^*)$.

Como c_i é contínua $\forall \ i \in \{1, \ldots, p\}$, existe $\varepsilon_1 > 0$ satisfazendo $c_i(x) < 0, \ \forall x \in B(x^*, \varepsilon_1)$ e $i \in \{1, \ldots, p\} \backslash I(x^*)$. Assim, existe $\bar{x_{\varepsilon_1}} \in B(x^*, \varepsilon_1)$ tal que

$$f(\bar{x}_{\varepsilon_1}) < f(x^*) \text{ e } c_i(\bar{x}_{\varepsilon_1}) = 0, \ \ \forall \ i \in I(x^*) \ \text{ e } \ c_i(\bar{x}_{\varepsilon_1}) < 0, \ \ \forall \ i \in \{1, \ldots, p\} \backslash I(x^*).$$

Contradizendo o fato de x^* ser minimizador local.

■

Com base no Lema 2.3.2, ao Problema (2.8) podemos aplicar resultados já conhecidos para o problema de minimização com restrições de igualdade.

Lema 2.3.3. *Se x^* é minimizador local do Problema (2.8), $I(x^*) = \{i \in \{1,\dots,p\} |\ c_i(x^*) = 0\}$ e $\{\nabla c_i(x^*),\ i \in I(x^*)\}$ é um conjunto linearmente independente, então $\forall\ i \in I(x^*)$, existem $\mu_i \in \mathbb{R}$ tais que $\nabla f(x^*) + \sum_{i \in I(x^*)} \mu_i \nabla c_i(x^*) = 0$.*

Demonstração. Análoga a do Teorema 2.2.7.

■

Em outras palavras, o Lema 2.3.3 nos diz que o gradiente de f é combinação linear dos gradientes das restrições ativas num minimizador local que satisfaz LICQ para restrições de desigualdade do Problema (2.8). O Teorema a seguir mostra que sabemos algo sobre os sinais dos coeficientes dessa combinação linear, o resultado mais importante desta seção.

Teorema 2.3.4 (Condições de Karush-Kuhn-Tucker (KKT)). *Se x^* é minimizador local que satisfaz LICQ para restrições de desigualdade do Problema (2.8), então existem únicos $\mu_i \geq 0$, $i \in I(x^*)$ tais que*

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \mu_i \nabla c_i(x^*) = 0.$$

Demonstração. Do Lema 2.3.3, existem $\mu_i \in \mathbb{R}$, $i \in I(x^*)$ tais que

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \mu_i \nabla c_i(x^*) = 0. \tag{2.10}$$

Falta apenas mostrar que $\mu_i \geq 0,\ i \in I(x^*)$. Suponhamos que exista $k \in I(x^*)$ tal que $\mu_k < 0$ e chamemos

$$\Omega_I = \{x \in \mathbb{R}^n |\ c_i(x) = 0,\ i \in I(x^*)\},$$

$$\Omega_k = \{x \in \mathbb{R}^n |\ c_i(x) = 0,\ i \in I(x^*),\ i \neq k\} \ \text{ e}$$

$M_I(x^*)$ o conjunto tangente a Ω_I por x^* e $M_k(x^*)$ o conjunto tangente a Ω_k por x^*. Por x^* satisfazer LICQ para restrições de desigualdade, $\nabla c_k(x^*)$ não é combinação linear dos outros gradientes das restrições ativas em x^*. Portanto, existe $y \in M_k(x^*)$ tal que

$$\nabla c_k(x^*)^T y < 0. \tag{2.11}$$

Seja $\gamma(t)$ uma curva viável em Ω_k passando por x^* com $\gamma'(0) = y$. Então para $t \geq 0$ suficientemente pequeno, $\gamma(t) \in \{x \in \mathbb{R}^n |\ c(x) \leq 0\}$. Chamando $\varphi(t) = f(\gamma(t))$, temos que $\varphi'(0) = \nabla f(x^*)^T y$.

Logo, pelas Equações (2.10) e (2.11) e $\mu_k < 0$, segue que $\varphi'(0) < 0$. O que contradiz o fato de x^* ser minimizador local.

■

2.4 Problemas com restrições de Igualdade e Desigualdade

Nas duas seções precedentes, 2.1 e 2.2, estudamos problemas com restrições de igualdade e desigualdade, respectivamente. Agora, abrangeremo-las num problema com a união dessas restrições. Mais precisamente, trataremos do problema a seguir, que é o problema geral de programação não linear

$$\text{minimizar } \ f(x), \text{ sujeito a } \ h(x) = 0 \ \text{ e } \ c(x) \leq 0, \tag{2.12}$$

em que $h : \mathbb{R}^n \to \mathbb{R}^m$ e $c : \mathbb{R}^n \to \mathbb{R}^p$ são funções continuamente diferenciáveis.

Para cada $x \in \Omega = \{x \in \mathbb{R}^n |\ c(x) \leq 0\}$, chamamos de restrições ativas em x aquelas para as quais $c_i(x) = 0$ e denotaremos por $I(x)$ o conjunto das restrições ativas em x. Analogamente, chamamos restrições inativas em x aquelas para as quais $c_i(x) < 0$.

Definição 2.4.1. *De maneira similar aos casos anteriores, definimos ponto x que satisfaz à condição de qualificação das restrições lineares sobre o conjunto factível como um ponto onde os gradientes das restrições ativas são linearmente independentes, isto é,* $\{\nabla h_i(x)\}_{i=1}^m \cup \{\nabla c_i(x)\}_{i \in I(x)}$ *é um conjunto linearmente independente.*

Teorema 2.4.2 (Condições de Karush-Kuhn-Tucker gerais). *Sejam x^* minimizador local que cumpre a Definição 2.4.1 sobre o Problema (2.12) e $I(x^*) = \{i \in \{1, \dots, p\} |\ c_i(x^*) = 0\}$. Então existem únicos $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ e $\mu_i \geq 0$, $\forall\ i \in I(x^*)$ tais que*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i \in I(x^*)} \mu_i \nabla c_i(x^*) = 0.$$

Demonstração. Análoga a do Teorema 2.3.4.

■

Desta forma, se x^* é um ponto que satisfaz LICQ e é minimizador local para o Problema (2.12) , definindo $\mu_i = 0$ se $i \notin I(x^*)$, podemos reescrever as condições KKT da seguinte forma:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla c_i(x^*) = 0, \tag{2.13}$$

$$h(x^*) = 0, \tag{2.14}$$

$$\mu_i c_i(x^*) = 0,\ i = 1, \ldots, p, \tag{2.15}$$

$$\mu_i \geq 0,\ i = 1, \ldots, p, \tag{2.16}$$

$$c_i(x^*) \leq 0,\ i = \ 1, \ldots, p. \tag{2.17}$$

As $n+m+p$ Equações (2.13) - (2.15) formam um sistema não linear nas incógnitas $x^* \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$. As soluções desse sistema que satisfazem às Equações (2.16) e (2.17) são denominados pontos estacionários do Problema (2.12).

Teorema 2.4.3 (**Condições necessárias de segunda ordem para restrições de igualdade e desigualdade**). *Sejam x^* ponto que satisfaz à Definição 2.4.1 e é minimizador local do Problema (2.12) e A matriz cujas linhas são os gradientes das restrições ativas em x^*, excluindo os gradientes daquelas restrições de desigualdade cujo multiplicador é zero. Então, se λ e μ são os vetores dos multiplicadores de Lagrange dados no Teorema 2.4.2,*

$$y^T \nabla^2{}_{xx} \ell(x^*, \lambda, \mu) y \geq 0$$

para todo $y \in N(A)$, onde

$$\ell(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{p} \mu_i c_i(x) \tag{2.18}$$

é o lagrangeano do Problema (2.12).

Demonstração. Análoga a do Teorema 2.2.9.

∎

Teorema 2.4.4 (**Condições suficientes de segunda ordem para restrições de igualdade e desigualdade**). *Se x^* satisfaz à condição necessária de primeira ordem para o Problema (2.12) e além disso $y^T \nabla^2{}_{xx} \ell(x^*, \lambda, \mu) y > 0$ para todo $y \in N(A)$, $y \neq 0$, onde a matriz A e a função $\ell(x, \lambda, \mu)$ estão definidas no Teorema 2.4.3, então x^* é minimizador local estrito do Problema (2.12).*

Demonstração. Análoga a do Corolário 2.2.10.

∎

Capítulo 3

Condições de Qualificação

Abordando problemas de minimização de funções com restrições, conforme Seção 2.4, nos deparamos com as condições de otimalidade e, ainda, com condições de qualificação das restrições. A implicação do Teorema de Karush-Kuhn-Tucker é verificada sempre que temos x^* um minimizador local e mais uma condição de qualificação, que no capítulo anterior consistiu da Definição 2.4.1, LICQ para restrições de igualdade, desigualdade e igualdade e igualdade. Nosso interesse nesta seção é estudar as condições de qualificação e verificar que a condição de qualificação de dependência linear positiva implica na de quasenormalidade. Exporemos as relações de implicação entre as condições de qualificação estudadas e exibiremos contra-exemplos para verificar que as demais recíprocas não ocorrem.

3.1 Definição de algumas condições de qualificação e suas relações

Nesta seção, estudaremos o Artigo [4], mostrando que a condição de dependência linear positiva dos matemáticos QI e WEI implica em restrição de qualificação de quasenormalidade, mas a recíproca não é verdadeira. No artigo [7], foi definida pela primeira vez a condição de qualificação de dependência linear positiva e só recentemente, no artigo [6], foi provado que, de fato, é uma condição de qualificação. Em [7] esta condição foi essencial para analisar o algoritmo de Programação Quadrática Sequencial.

Definição 3.1.1. *Sejam $h : \mathbb{R}^n \to \mathbb{R}^m$ e $g : \mathbb{R}^n \to \mathbb{R}^p$ funções continuamente diferenciáveis. O conjunto viável Ω é definido como:*

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

Também, $\forall\, x \in \Omega$, defina o conjunto de índices de restrições ativas de desigualdade como:

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

Definição 3.1.2 (Condição de qualificação de independência linear - LICQ). *Um ponto factível $x \in \Omega$ verifica LICQ se os gradientes $\{\nabla g_i(x)\}_{i\in I(x)} \cup \{\nabla h_j(x)\}_{j\in J}$ são linearmente independentes.*

Definição 3.1.3 (Condição de qualificação de Mangasarian-Fromovitz - MFCQ). *Um ponto factível $x \in \Omega$ verifica MFCQ se os gradientes das restrições de igualdade, $\{\nabla h_j(x)\}_{j\in J}$, são linearmente independentes e existe uma direção $d \in \mathbb{R}^n$ tal que*

$$\nabla h_j(x)^T d = 0,\ j \in J \quad e$$

$$\nabla g_i(x)^T d < 0,\ i \in I(x).$$

Dizemos que $x \in \Omega$ é MF não regular se existem $\lambda_1,\dots,\lambda_m \in \mathbb{R},\ \mu_i \geq 0, \forall\, i \in I(x)$ com

$$\sum_{j\in J} |\lambda_j| + \sum_{i\in I(x)} \mu_i > 0, \quad satisfazendo$$

$$\sum_{j=1}^{m} \lambda_j \nabla h_j(x) + \sum_{i\in I(x)} \mu_i \nabla g_i(x) = 0.$$

Note que um ponto Mangasarian-Fromovitz não regular é exatamente um ponto que não satisfaz à condição de qualificação MF. Pontos factíveis que satisfazem MFCQ serão chamados Mangasarian-Fromovitz regulares.

Definição 3.1.4 (Condição de qualificação de posto constante - CRCQ). *Dizemos que $x \in \Omega$ satisfaz CRCQ se a independência linear do subconjunto dos gradientes ativos - restrições de igualdade e desigualdade - em x implica que esses gradientes são linearmente dependentes, $\forall\, y$ em alguma vizinhança de x.*

Definição 3.1.5 (Dependência Linear Positiva). *Sejam $x \in \Omega$, $I_0 \subset I(x)$, $J_0 \subset J$. Dizemos que o conjunto dos gradientes $\{\nabla g_i(x)\}_{i\in I_0} \cup \{\nabla h_j(x)\}_{j\in J_0}$ é linearmente dependente positivo se existem escalares $\{\alpha_j\}_{j\in J_0}, \{\beta_i\}_{i\in I_0}$ tais que*

$$\beta_i \geq 0,\ \forall\, i \in I_0,$$

$$\sum_{j\in J_0} |\alpha_j| + \sum_{i\in I_0} \beta_i > 0,$$

$$\sum_{j\in J_0} \alpha_j \nabla h_j(x) + \sum_{i\in I_0} \beta_i \nabla g_i(x) = 0.$$

Caso contrário, dizemos que $\{\nabla g_i(x)\}_{i\in I_0} \cup \{\nabla h_j(x)\}_{j\in J_0}$ é independente linear positivo.

Definição 3.1.6 (**Condição de qualificação de dependência linear positiva constante - CPLD**). *Um ponto factível $x \in \Omega$ é dito satisfazer CPLD se é MF regular ou, para alguns $I_0 \subset I(x)$, $J_0 \subset \{1,\ldots,m\}$ tal que o conjunto dos gradientes $\{\nabla g_i(x)\}_{i\in I_0} \cup \{\nabla h_j(x)\}_{j\in J_0}$ é linearmente dependente positivo, existe uma vizinhança $N(x)$ de x tal que o conjunto dos gradientes $\{\nabla g_i(y)\}_{i\in I_0} \cup \{\nabla h_j(y)\}_{j\in J_0}$ é linearmente dependente, $\forall\, y \in N(x)$.*

Claramente temos que MFCQ implica em CPLD. No entanto, a recíproca não é verdadeira. Basta considerar o seguinte conjunto: $h_1(x) = x_1,\ h_2(x) = x_1$.

Definição 3.1.7 (Propriedade B-sequencial). *Dizemos que $x \in \Omega$ com restrições ativas $I(x)$ satisfaz à propriedade B-sequencial relativa aos escalares $\lambda_1,\ldots,\lambda_m$ e $\mu_i,\ i \in I(x)$ se existe uma sequência de pontos, não necessariamente factíveis, $y_k \in \mathbb{R}^n$ tal que $\lim\limits_{k\to\infty} y_k = x$ e, $\forall\ j \in \{1,\ldots,m\}$ tal que $\lambda_j \neq 0$ e $\forall\ i \in I(x)$ tal que $\mu_i \neq 0$, tem-se*

$$\lambda_j h_j(y_k) > 0 \quad e \quad \mu_i g_i(y_k) > 0,\ \forall\, k \in \{0,1,\ldots\}.$$

Definição 3.1.8 (**Condição de qualificação de quasenormalidade**). *Dizemos que $x \in \Omega$ satisfaz à condição de qualificação de quasenormalidade se é MF regular ou, sendo MF não regular com escalares $\lambda_1,\ldots,\lambda_m$ e $\mu_i,\ i \in I(x)$, então não cumpre a propriedade B-sequencial relativa a esses escalares.*

Em outras palavras, a quasenormalidade diz que a propriedade B-sequencial não pode ser satisfeita se x não é MF regular.

Definição 3.1.9 (Propriedade P-sequencial). *Diz-se que $x \in \Omega$ com restrições ativas $I(x)$ obedece à propriedade P-sequencial relativamente aos escalares $\lambda_1,\ldots,\lambda_m$ e $\mu_i,\ i \in I(x)$ se existe uma sequência, não necessariamente de pontos factíveis, $y_k \in R^n$ tal que $\lim\limits_{k\to\infty} y_k = x$ implica em*

$$\sum_{j=1}^{m} \lambda_j h_j(y_k) + \sum_{i\in I(x)} \mu_i g_i(y_k) > 0.$$

Definição 3.1.10 (**Condição de qualificação de pseudonormalidade - PCQ**). *É dito que $x \in \Omega$ satisfaz PCQ se x é MF regular ou, sendo MF não regular com escalares $\lambda_1,\ldots,\lambda_m$ e $\mu_i,\ i \in I(x)$, então não cumpre a propriedade P-sequencial relacionada a estes escalares.*

Em outras palavras, PCQ traduz que a propriedade P-sequencial não pode ser satisfeita quando a qualificação MF não é cumprida em x.

A fim de provar que a CPLD implica na condição de quasenormalidade, precisaremos de 3 resultados do Cálculo de várias variáveis, os quais colocaremos nos Lemas 3.1.13, 3.1.14 e 3.1.15 e referenciaremos as demonstrações.

Definição 3.1.11. *Um conjunto $K \subset \mathbb{R}^n$ chama-se cone quando*

$$d \in K \;\; \Rightarrow \;\; td \in K, \, \forall \, t \in \mathbb{R}_+.$$

Definição 3.1.12. *Seja $S = \{x_1, \; x_2, \ldots, x_n\}$ um conjunto de vetores de um espaço vetorial. Então uma combinação linear cônica de elementos de S é formada através de*

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n,$$

em que $a_i \geq 0, \; i = 1, \ldots, n$ e $\displaystyle\sum_{i=1}^{n} a_i$ qualquer.

Lema 3.1.13. *Sejam $D \subset \mathbb{R}^n$ conjunto aberto, $x \in D, \; F : D \to \mathbb{R}^n$; $F \in C^1(D)$ e assuma $F'(x)$ não singular, denotando $F = (f_1, \ldots, f_n)$.*

Pelo Teorema 2.2.4, F é um homomorfismo entre uma vizinhança D_1 de x e $F(D_1)$; além disso, $F^{-1} : F(D_1) \to D_1$ está bem definida.

Assuma que $q \in \{1, \ldots, n-1\}, \; f : D \to \mathbb{R}, \; f \in C^1(D)$ é tal que $\nabla f(y)$ é uma combinação linear de $\{\nabla f_1(y), \ldots, \nabla f_q(y)\}, \; \forall \, y \in D$.

Defina $\varphi(u) = f[F^{-1}(u)], \; \forall \, u \in F(D_1)$, então

$$\frac{\partial \varphi}{\partial u_j}(u) = 0, \; \forall \, u \in F(D_1), j = q+1, \ldots, n.$$

Demonstração. Encontra-se a demonstração no livro [3] ou outro texto de Análise.

■

Lema 3.1.14. *Sejam $f, f_1, \ldots, f_q : D \subset \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciáveis, $x \in D$, sendo D conjunto aberto. Assuma que $\{\nabla f_1(x), \ldots \nabla f_q(x)\}$ é um conjunto linearmente independente e $\nabla f(y)$ é uma combinação linear de $\{\nabla f_1(y), \ldots, \nabla f_q(y)\}$,$\forall \; y \in D$. Em particular, $\nabla f(x) = \displaystyle\sum_{i=1}^{q} \lambda_i \nabla f_i(x)$. Então existem $D_2 \subset \mathbb{R}^q$ uma vizinhança aberta contendo $(f_1(x), \ldots, f_q(x))$ e uma função $\varphi : D_2 \to \mathbb{R}, \; \varphi \in C^1(D_2)$ tais que $\forall \; y \in D_1$ obtemos $(f_1(y), \ldots, f_q(y)) \in D_2$ e $f(y) = \varphi(f_1(y), \ldots, f_q(y))$. Além disso, $\lambda_i = \dfrac{\partial \varphi}{\partial u_i}(f_1(x), \ldots, f_q(x)), \; \forall \; i = 1, \ldots, q$.*

Demonstração. Encontra-se a demonstração no livro [3] ou outro texto de Análise.

■

Lema 3.1.15 (Caratheodory). *Sejam G um cone para algum $G \subset \mathbb{R}^n$ e $x \in G$. Então x pode ser escrito como uma combinação cônica de $m, \; m \leq n$ vetores linearmente independentes de G.*

Demonstração. Se x pertence ao cone G, então $x = \sum_{j=1}^{m} \lambda_j a_j$ com $a_1, \ldots, a_m \in G$ e $\lambda_1, \ldots, \lambda_m > 0$. Se os a_i's são linearmente independentes, então $m \leq n$ e o resultado fica provado. Se $a_1, \ldots, a_m$ são linearmente dependentes, então existe $(\mu_1, \ldots, \mu_m) \neq (0, \ldots, 0)$ tal que

$$\sum_{j=1}^{m} \mu_j a_j = 0.$$

Seja $\Delta^* = \max \{\Delta \geq 0 \mid \lambda_j - \Delta\mu_j \geq 0, \forall\, j = 1, \ldots, m\} = \min \left(\frac{\lambda_j}{\mu_j} \mid \mu_j > 0,\ j = 1, \ldots, m\right)$; como podemos assumir a existência de pelo menos um $\mu_i > 0$, temos que $0 < \Delta^* < \infty$.

Seja $\lambda_j{}' = \lambda_j - \Delta^*\mu_j \geq 0$. Então $x = \sum_{j=1}^{m} \lambda_j{}' a_j$ e existe um índice "i" tal que $\lambda_i' = 0$ e portanto podemos reduzir o número de $a_j's$.

Aplicando esse método recursivamente chega-se a um subconjunto de $\{a_1, \ldots, a_m\}$ linearmente independente que gera x.

■

Teorema 3.1.16. *Sejam $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ continuamente diferenciáveis, conjunto viável $X = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\}$ e $x \in X$ satisfazendo CPLD. Então, x obedece à condição de qualificação de quasenormalidade.*

Demonstração. Assuma que $x \in X$ satisfaz à condição CPLD. Assim, se x é MF regular temos a implicação. Suponha então que x é MF não regular. Logo, existem $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, $\mu_i \geq 0$, $\forall\, i \in I(x)$ tais que

$$\sum_{j=1}^{m} |\lambda_j| + \sum_{i \in I(x)} \mu_i > 0 \text{ e}$$

$$\sum_{j=1}^{m} \lambda_j \nabla h_j(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0. \tag{3.1}$$

Defina:

$$I_+(x) = \{j \in \{1, \ldots, m\} | \ \lambda_j > 0\},$$

$$I_-(x) = \{j \in \{1, \ldots, m\} | \ \lambda_j < 0\},$$

$$I_0(x) = \{i \in I(x) | \ \mu_i > 0\}.$$

Da Equação (3.1) ,

$$\sum_{j \in I_+(x)} \lambda_j \nabla h_j(x) + \sum_{j \in I_-(x)} \lambda_i \nabla h_j(x) + \sum_{i \in I_0(x)} \mu_i \nabla g_i(x) = 0. \tag{3.2}$$

Assuma, primeiramente, que $I_+(x) \neq \emptyset$. Seja $j_1 \in I_+(x)$. Então, da Equação (3.2),

$$\lambda_{j_1} \nabla h_{j_1}(x) = -\sum_{j \in I_+(x) \setminus \{j_1\}} \lambda_j \nabla h_j(x) - \sum_{j \in I_-(x)} \lambda_i \nabla h_j(x) - \sum_{i \in I_0(x)} \mu_i \nabla g_i(x).$$

Consideremos as duas seguintes possibilidades:

$$\nabla h_{j_1}(x) = 0, \tag{3.3}$$

$$\nabla h_{j_1}(x) \neq 0. \tag{3.4}$$

Se a Equação (3.3) acontece, o conjunto unitário $\{\nabla h_{j_1}(x)\}$ é linearmente dependente positivo. Então, pela condição CPLD, obtemos que o conjunto $\{\nabla h_{j_1}(y)\}$ é linearmente dependente para todo y em alguma vizinhança aberta de x. Ademais, $\nabla h_{j_1}(y) = 0$ para todo y em alguma vizinhança aberta de x . De $h_{j_1}(x) = 0$, concluímos que $h_{j_1}(y) = 0$ nessa vizinhança. Assim, a sequência $y_k \to x$ tal que $h_{j_1}(y_k) > 0$ não pode existir. Consequentemente, a propriedade B-sequencial relativa a $\lambda_1, \ldots, \lambda_m$ e $\mu_i,\ i \in I(x)$ não pode acontecer.

Agora, assuma que a Equação (3.4) ocorre. Então, pelo Lema 3.1.15 , existem

$$I_{++}(x) \subset I_+(x) \setminus \{j_1\},\ I_{--}(x) \subset I_-(x),\ I_{00}(x) \subset I_0(x)$$

tais que os vetores

$$\{\nabla h_j(x)\}_{j \in I_{++}(x)},\ \{\nabla h_j(x)\}_{j \in I_{--}(x)},\ \{\nabla g_i(x)\}_{i \in I_{00}(x)} \tag{3.5}$$

são linearmente independentes e

$$\lambda_{j_1} \nabla h_{j_1}(x) = -\sum_{j \in I_{++}(x)} \bar{\lambda}_j \nabla h_j(x) - \sum_{j \in I_{--}(x)} \bar{\lambda}_i \nabla h_j(x) - \sum_{i \in I_{00}(x)} \bar{\mu}_i \nabla g_i(x), \tag{3.6}$$

$$\text{com } \bar{\lambda}_j > 0,\ \forall\ j \in I_{++}(x),$$

$$\bar{\lambda}_j < 0,\ \forall\ j \in I_{--}(x),$$

$$\bar{\mu}_i > 0,\ \forall i\ \in I_{00}(x).$$

Da independência linear dos vetores na Equação (3.5) e da continuidade da derivada, os vetores

$$\{\nabla h_j(y)\}_{j \in I_{++}(x)},\ \{\nabla h_j(y)\}_{j \in I_{--}(x)},\ \{\nabla g_i(y)\}_{i \in I_{00}(x)} \tag{3.7}$$

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

$$\lambda_{j_1} \nabla h_{j_1}(y),\ \{\nabla h_j(y)\}_{j \in I_{++}(x)},\ \{\nabla h_j(y)\}_{j \in I_{--}(x)},\ \{\nabla g_i(y)\}_{i \in I_{00}(x)}$$

são linearmente dependentes para todo y numa vizinhança de x. Dessa forma, $\lambda_{j_1}\nabla h_{j_1}(y)$ deve ser uma combinação linear dos vetores na Equação (3.7) para todo y numa vizinhança de x.

Por simplicidade e sem perder generalidade, escrevemos:

$$I_{++}(x) = \{1,\ldots,m_1\},\ I_{--}(x) = \ \{m_1+1,\ldots,m_2\},\ I_{00}(x) = \{1,\ldots,p_1\},$$

com $m_2 + p_1 = q$.

Pelo Lema 3.1.14 , existe uma função suave φ definida numa vizinhança de $(0,\ldots,0) \in \mathbb{R}^q$ tal que, para todo y numa vizinhança de x,

$$\lambda_{j_1}h_{j_1}(y) = \varphi(h_1(y),\ldots,h_{m_2}(y),g_1(y),\ldots,g_{p_1}(y)). \tag{3.8}$$

Agora, suponha que $\{y_k\}$ é uma sequência que converge para x tal que

$$h_j(y_k) > 0,\ j = 1,\ldots,m_1,$$

$$h_j(y_k) < 0,\ j = m_1+1,\ldots,m_2,$$

$$g_i(y_k) > 0,\ i = 1,\ldots,p_1.$$

Das Equações (3.6) e (3.8) e do fato de que $\lambda_i = \dfrac{\partial\varphi}{\partial\mu_i}(h_1(y),\ldots,h_{m_2}(y),g_1(y),\ldots,g_{p_1}(y))$, para k suficientemente grande temos que $\lambda_{j_1}h_{j_1}(y_k) \leq 0$. Mas isso implica que a propriedade B-sequencial não acontece. As provas para os casos $I_-(x) \neq \emptyset$ e $I_0(x) \neq \emptyset$ são análogas a esses casos. Portanto, CPLD implica quasenormalidade.

■

3.2 Relações com outras condições de qualificação

Nesta seção, exibiremos alguns contra-exemplos mostrando que algumas implicações das condições de qualificação definidas na Seção 3.1 não ocorrem.

Exemplo 1: [Condição de qualificação de quasenormalidade não implica CPLD]

Caso $n = 2$, $m = 2$, $p = 0$

$$\begin{cases} h_1(x_1,x_2) = x_2 e^{x_1} \\ h_2(x_1,x_2) = x_2 \\ x^* = (0,0) \end{cases}$$

Temos que $\nabla h_1(x^*) = \nabla h_2(x^*)$. Da equação

$$\lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*) = 0,$$

seque que $\lambda_1 = -\lambda_2$. Consequentemente, a propriedade B-sequencial não ocorre porque o sinal de $h_1(x)$ é o mesmo que $h_2(x)$, $\forall\, x \in \mathbb{R}^2$. Logo, x^* satisfaz à condição de quasenormalidade e assim, para toda vizinhança de x^*, existem pontos em que ∇h_1 e ∇h_2 são linearmente independentes. Assim, CPLD não é satisfeita em x^*.

Exemplo 2: [CPLD não implica CRCQ ou MFCQ]

Tome $n = 2$, $m = 0$, $p = 4$, $x^* = (0,0)$

$$\begin{cases} g_1(x_1,x_2) = x_1 & \nabla g_1(x) = (1,0) \\ g_2(x_1,x_2) = x_1 + {x_2}^2 & \nabla g_2(x) = (1,2x_2) \\ g_3(x_1,x_2) = x_1 + x_2 & \nabla g_3(x) = (1,1) \\ g_4(x_1,x_2) = -x_1 - x_2 & \nabla g_4(x) = (-1,-1) \end{cases}$$

De $\nabla g_3(x^*) + \nabla g_4(x^*) = 0$, vemos que x^* não satisfaz MFCQ. Como $\nabla g_1(x^*)$ e $\nabla g_2(x^*)$ são linearmente independentes, pela continuidade da derivada, em toda vizinhança de x^* existem pontos em que $\nabla g_1(x)$ e $\nabla g_2(x)$ são linearmente independentes. Assim, x^* não satisfaz CRCQ.

Mostremos que vale CPLD. Assuma que $\mu_i \geq 0$, $i = 1, \ldots, 4$

$$\sum_{i=1}^{4} \mu_i > 0 \;\text{ e }\; \sum_{i=1}^{4} \mu_i \nabla g_i(x^*) = 0. \tag{3.9}$$

Se $\mu_2 = 0$, da Equação (3.9), temos que $\nabla g_1(x^*)$, $\nabla g_3(x^*)$ e $\nabla g_4(x^*)$ são linearmente dependentes. Agora, como g_1, g_3 e g_4 são lineares, $\nabla g_1(x)$, $\nabla g_3(x)$ e $\nabla g_4(x)$ são linearmente dependentes $\forall\, x \in \mathbb{R}^2$. Se $\mu_2 > 0$, temos que

$$\mu_1(1,0) + \mu_2(1,0) + \mu_3(1,1) + \mu_4(-1,-1) = (0,0),$$

o que nos fornece $\mu_3 = \mu_4$ e $\mu_1 = -\mu_2$ e assim $\mu_1 + \mu_2 + \mu_3 + \mu_4 = 2\mu_3 > 0$. Observe que para toda vizinhança $N(x)$ de $(0,0) = x^*$, temos que $\nabla g_3(x)$ e $\nabla g_4(x)$ são linearmente dependentes $\forall\, x \in N(x)$. Logo, cumpre CPLD.

Exemplo 3: [CPLD não implica condição de pseudonormalidade]

Tome $n = 1$, $m = 0$, $p = 2$, $x^* = 0$

$$\begin{cases} g_1(x_1) = -x_1 & \qquad\qquad \nabla g_1(x_1) = -1 \\ g_2(x_1) = x_1 + {x_1}^2 & \qquad\qquad \nabla g_2(x_1) = 1 + 2x_1 \end{cases}$$

O ponto x^* não é MF regular pois $\nabla g_1(x^*) + \nabla g_2(x^*) = 0$. Mas, $g_1(x_1) + g_2(x_1) = {x_1}^2 > 0$, $\forall\, x_1 \neq x^*$, donde vale a propriedade P-sequencial. Agora, note que $\nabla g_1(x^*) \neq 0$ e $\nabla g_2(x^*) \neq 0$ e $\nabla g_1(x)$, $\nabla g_2(x)$ são linearmente dependentes, $\forall\, x \in \mathbb{R}$. Assim, x^* obedece CRCQ e, consequentemente, CPLD.

Exemplo 4:[**Pseudonormalidade não implica CPLD**]

Tome $n = 2$, $m = 0$, $p = 2$, $x^* = (0,0)$

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

Então, $\nabla g_1(x_1,x_2) = (-1,0)^T$ e $\nabla g_2(x_1,x_2) = (1 - 2x_1{x_2}^2, -2{x_1}^2x_2)^T$. Logo, $\forall\, u_1 = u_2 > 0$ temos:

$$\mu_1 \nabla g_1(0,0) + \mu_2 \nabla g_2(0,0) = (0,0)^T.$$

Também,

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

Portanto, x^* cumpre a condição de qualificação de pseudonormalidade. Por outro lado, $\nabla g_1(x)$ e $\nabla g_2(x)$ são linearmente independentes se $x_1 \neq 0$, $x_2 \neq 0$. O que nos diz que CPLD não ocorre.

Através dos conhecimentos supracitados, podemos compactá-los em uma tabela comparando as relações entre essas condições de qualificação, ou seja, quando uma implica ou não em outra e que algumas condições são mais fracas que outras, donde podemos verificar que pontos satisfaçam certas condições sem exigir muitas hipóteses.

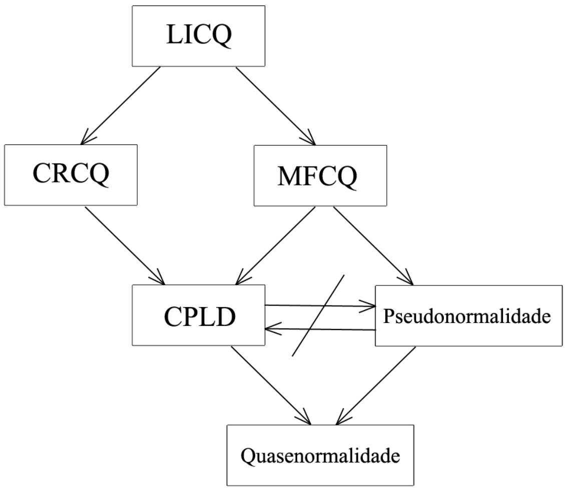

Figura 1: Tabela de implicações das condições de qualificação

Capítulo 4

Aplicação: Método de Lagrangeano Aumentado sob condição CPLD

Nesta seção estudaremos algoritmos de Lagrangeano Aumentado, como no Artigo [10] com destaque para a condição de dependência linear positiva constante dada na Definição 3.1.6, e sua influência na convergência desses algoritmos através do estudo do Artigo [9]. A relevância deste estudo esta no fato de que resultados de convergência que têm, em suas hipóteses, condições de qualificação fracas são mais fortes do que aqueles baseados em condições de qualificação fortes.

A ideia chave para desenvolver as condições necessárias e suficientes para um problema de otimização com restrições é transformá-lo em um problema de otimização sem restrições e aplicar as condições para este caso. Uma forma de fazer esta transformação é através da introdução de uma função auxiliar, chamada de função de Lagrange, definida como na Equação (2.18).

Sejam $F:\mathbb{R}^n\rightarrow\mathbb{R}^n,\ h:\mathbb{R}^n\rightarrow\mathbb{R}^m$ e $\Omega=\{x\in\mathbb{R}^n|\ l\leq x\leq u\}$. Assuma que h admita primeira derivada contínua num conjunto aberto contendo Ω.

Seja $\mathscr{P}_A$ o operador da projeção euclidiana num conjunto fechado e convexo A. Um ponto $x\in\mathbb{R}^n$ é dito satisfazer KKT com relação ao problema definido por $F,\ h$ e Ω se $\exists\ \lambda\in\mathbb{R}^m$ tal que

$$\mathscr{P}_\Omega[x-F(x)-\nabla h(x)\lambda]-x=0,\ h(x)=0,\ x\in\Omega. \tag{4.1}$$

Se $f:\mathbb{R}^n\rightarrow\mathbb{R}$ é diferenciável e $F=\nabla f$, a Equação (4.1) representa as condições de KKT do problema:

$$\text{minimizar } f(x),\ \text{sujeito a } h(x)=0,\ x\in\Omega. \tag{4.2}$$

Apresentaremos um algoritmo de Lagrangeano Aumentado para resolver o Problema (4.2). Por um lado, caracterizaremos situações em que pontos limites inviáveis existem, usando hipóteses mais fracas que a condição de qualificação LICQ dada pela Definição 3.1.2. Por outro lado, o fato de que pontos limites factíveis são pontos de KKT seguirá usando a condição de qualificação de CPLD dada na Definição 3.1.6.

4.1 Algoritmo do Lagrangeano Aumentado

Assumimos F contínua. Dados $x \in \Omega,\ \lambda \in \mathbb{R}^m,\ \rho \in \mathbb{R}^m_{++}$, definimos:

$$G(x,\lambda,\rho) = F(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{i=1}^{m} \rho_i h_i(x) \nabla h_i(x).$$

Se o sistema de KKT é originado de um problema de minimização suave, o mapeamento F é gradiente de alguma $f : \mathbb{R}^n \to \mathbb{R}$. Neste caso, definimos, para $\rho \in \mathbb{R}^m_{++}$:

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

Esta é a função do Lagrangeano Aumentado associado à Equação (4.2).

Algoritmo 4.1.1. *[Algoritmo A]*

Dados $x_0 \in \Omega,\ \tau \in [0,1], \gamma > 1, -\infty < \bar{\lambda}_{min} \le 0 \le \bar{\lambda}_{max} < \infty,\ \rho_1 \in \mathbb{R}^m_{++},\ \bar{\lambda}_1 \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m$. *Seja* $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *sequência que converge para* 0.

Passo 1. *Inicializa*

Faça $k \leftarrow 1$.

Passo 2. *Resolva o subproblema*

Compute $x_k \in \Omega$ *tal que*

$$\| P_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k \|_\infty \le \varepsilon_k. \tag{4.3}$$

Passo 3. *Estime os multiplicadores*

Defina, $\forall\ i = 1, \ldots, m$,

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k). \tag{4.4}$$

Se $h(x_k) = 0$ *e* $P_\Omega[x_k - G(x_k, \bar{\lambda}_k, \rho_k)] - x_k = 0$, *termine a execução do algoritmo. Neste caso,* x_k *é ponto KKT e* λ_{k+1} *é o vetor dos multiplicadores de Lagrange associado.*

Senão, compute

$$\bar{\lambda}_{k+1} \in [\bar{\lambda}_{min}, \bar{\lambda}_{max}]^m.$$

Passo 4. *Atualize os parâmetros de penalização*

Defina $\Gamma_k = \{i \in \{1,\ldots,m\} |\ h_i(x_k) > \tau \parallel h(x_{k-1}) \parallel_\infty\}$. *Se* $\Gamma_k = \emptyset$, *defina* $\rho_{k+1} = \rho_k$.

Senão, defina

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

Passo 5. *Comece nova iteração*

Faça $k \leftarrow k+1$ *e vá para Passo* 2.

Os Algoritmos de Lagrangeano Aumentado são baseados na resolução de iterações internas, onde calcula-se x_k satisfazendo à Equação (4.3). No caso de minimização ($F = \nabla f$), a maneira mais razoável para obter essas condições é resolver aproximadamente o problema de otimização:

$$\text{minimizar } \ L(x, \bar{\lambda}_k, \rho_k), \ \text{ sujeito a } \ x \in \Omega. \tag{4.5}$$

Agora investigaremos sobre que condições um ponto limite factível de uma sequência gerada pelo Algoritmo de Lagrangeano Aumentado gerado pelo Algoritmo 4.1.1 é um ponto KKT. Usaremos na prova que um ponto limite factível cumpre KKT se satisfaz à condição CPLD dada pela Definição 3.1.6. Recentemente, como no artigo [6], foi provado que, de fato, CPLD é condição de qualificação.

Considere o problema:

$$\textit{minimizar } f(x), \text{ sujeito a } \ h(x) = 0, \tag{4.6}$$

em que $h : \mathbb{R}^n \to \mathbb{R}^m$. Assumindo a condição CPLD conforme Definição 3.1.6, o resultado que se segue é mais forte do que se assumíssemos a condição LICQ da Definição 3.1.2.

Teorema 4.1.2. *Assuma que* $\{x_k\}$ *é sequência gerada pelo Algoritmo 4.1.1 e seja* x^* *ponto limite factível que satisfaz à condição CPLD da Definição 3.1.6. Então,* x^* *é ponto KKT do Problema 4.6.*

Demonstração. Escrevemos $G^k = G(x_k, \bar{\lambda}_k, \rho_k)$. Defina, $\forall\, k \in \mathbb{N}$, $v_k = \mathscr{P}_\Omega(x_k - G^k)$.

Logo, $v_k \in \mathbb{R}^n$ é solução de

$$\text{minimizar } \ \parallel v - (x_k - G^k) \|_2{}^2, \text{ sujeito a } \ l \leq v \leq u.$$

Pelas condições de KKT para este problema, $\exists\ {\mu_k}^u \in \mathbb{R}^n_+,\ {\mu_K}^l \in \mathbb{R}^n_+$ tais que $\forall\ k\ \in \mathbb{N}$

$$v_k - x_k + G^k + \sum_{i=1}^{n} [{\mu_k}^u]_i e_i - \sum_{i=1}^{n} [{\mu_k}^l]_i e_i = 0 \tag{4.7}$$

$$\text{e}\ \ [{\mu_k}^u]_i(u_i - [x_k]_i) = [{\mu_k}^l]_i(l_i - [x_k]_i) = 0,\ \forall\ i = 1, \ldots, n. \tag{4.8}$$

Pela Equação (4.4), $\lim\limits_{k\to\infty}(v_k - x_k) = 0$. Então, pela Equação (4.7),

$$\lim_{k\to\infty}\left(G^k + \sum_{i=1}^{n} [{\mu_k}^u]_i e_i - \sum_{i=1}^{n} [{\mu_k}^l]_i e_i\right) = 0.$$

Também, definindo λ_{k+1} como na Equação (4.4), ou seja, $[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i h_i(x_k)$,

$$\lim_{k\to\infty}\left(F(x_k) + \nabla h(x_k)\lambda_{k+1} + \sum_{i=1}^{n} [{\mu_k}^u]_i e_i - \sum_{i=1}^{n} [{\mu_k}^l]_i e_i\right) = 0. \tag{4.9}$$

Assuma agora que $K \underset{\infty}{\subseteq} \mathbb{N}$ tal que

$$\lim_{k\in K} x_k = x^*. \tag{4.10}$$

Defina:

$$I_l = \{i \in \{1, ..., n\} \mid [x^*]_i = l_i\},$$

$$I_u = \{i \in \{1, ..., n\} \mid [x^*]_i = u_i\},$$

$$I_0 = \{i \in \{1, ..., n\} \mid l_i < [x^*]_i < u_i\}.$$

Da equação 4.10, existe $k_0 \in \mathbb{N}$ tal que para todo $k \in K,\ k \geq k_0$,

$$i \in I_0 \Rightarrow l_i < [x_k]_i < u_i.$$

Agora, pela equação 4.8, $\forall\ k \in K, k \geq k_0$, temos que:

$$[{\mu_k}^u]_i = 0,\ \forall\ i \notin I_u$$

$$\text{e}\ \ [{\mu_k}^l]_i = 0,\ \forall\ i \notin I_l$$

Conforme equação 4.9,

$$\lim_{k\in K}\left(F(x_k) + \nabla h(x_k)\lambda_{k+1} + \sum_{i\in I_u} [{\mu_k}^u]_i e_i - \sum_{i\in I_l} [{\mu_k}^l]_i e_i\right) = 0.$$

Defina, para $k \in K,\ k \geq k_0$,

$$E_k = F(x_k) + \nabla h(x_k)\lambda_{k+1} + \sum_{i\in I_u} [{\mu_k}^u]_i e_i - \sum_{i\in I_l} [{\mu_k}^l]_i e_i.$$

Claramente,

$$\lim_{k\in K} E_k = 0$$

$$\text{e} \quad E_k - F(x_k) = \nabla h(x_k)\lambda_{k+1} + \sum_{i\in I_u} [{\mu_k}^u]_i e_i - \sum_{i\in I_l} [{\mu_k}^l]_i e_i$$

para todo $k \in K,\ k \geq k_0$.

Pelo Lema 3.1.15, sabemos que $\forall\ k \in K,\ k \geq k_0$ existem

$$\widetilde{I}_k \subseteq \{1,...,m\},\ \widetilde{I}_{uk} \subseteq I_u,\ \widetilde{I}_{lk} \subseteq I_l,$$

$$[\widetilde{\lambda}_k]_i \geq 0\ \forall\ i \in \widetilde{I}_k,\ [\widetilde{\mu}_k^u]_i \geq 0\ \forall\ i \in \widetilde{I}_{uk},\ [\widetilde{\mu}_k^l]_i \geq 0\ \forall\ i \in \widetilde{I}_{lk}$$

tais que os vetores

$$\{\nabla h(x_k)\}_{i\in\widetilde{I}_k},\ \{e_i\}_{i\in\widetilde{I}_{uk}},\ \{-e_i\}_{i\in\widetilde{I}_{lk}}$$

são linearmente independentes e

$$E_k - F(x_k) = \sum_{i\in\widetilde{I}_k} [\widetilde{\lambda}_k]_i \nabla h_i(x_k) + \sum_{i\in\widetilde{I}_{uk}} [\widetilde{\mu}_k^u]_i e_i - \sum_{i\in\widetilde{I}_{lk}} [\widetilde{\mu}_k{}^l]_i e_i. \tag{4.11}$$

Como só há finitos conjuntos possíveis para $\widetilde{I}_k,\ \widetilde{I}_{uk},\ \widetilde{I}_{lk}$, existe um conjunto infinito de índices

$$K_1 \subseteq \{k \in K \mid k \geq k_0\}$$

tal que

$$\widetilde{I}_k = \widetilde{I},\ \widetilde{I}_{uk} = \widetilde{I}_u,\ \widetilde{I}_{lk} = \widetilde{I}_l$$

$\forall\ k \in K_1$. Ainda, da equação 4.11,

$$E_k - F(x_k) = \sum_{i\in\widetilde{I}} [\widetilde{\lambda}_k]_i \nabla h_i(x_k) + \sum_{i\in\widetilde{I}_u} [\widetilde{\mu}_k^u]_i e_i - \sum_{i\in\widetilde{I}_l} [\widetilde{\mu}_k{}^l]_i e_i \tag{4.12}$$

e os vetores

$$\{\nabla h_i(x_k)\}_{i\in\widetilde{I}},\ \{e_i\}_{i\in\widetilde{I}_u},\ \{-e_i\}_{i\in\widetilde{I}_l} \tag{4.13}$$

são linearmente independentes $\forall\ k \in K_1$. Defina

$$S_k = \max\{\max\{|[\widetilde{\lambda}_k]_i|,\ i \in \widetilde{I}\},\ \max\{|[\widetilde{\mu}_k^u]_i|,\ i \in \widetilde{I}_u\},\ \max\{|[\widetilde{\mu}_k^l]_i|,\ i \in \widetilde{I}_l\}\}.$$

Consideremos as duas seguintes possibilidades:

1. $\{S_k\}_{k\in K_1}$ é limitada;

2. $\{S_k\}_{k\in K_1}$ é ilimitada.

No primeiro caso, exite K_2 um subconjunto infinito de K_1 tal que

$$\lim_{k\in K_2}[\widetilde{\lambda}_k]_i = \widetilde{\lambda}_i,$$

$$\lim_{k\in K_2}[\widetilde{\mu}_k^u]_i = \widetilde{\mu}_i^u \geq 0,\ \forall\ i \in \widetilde{I}_u,$$

$$\lim_{k\in K_2}[\widetilde{\mu}_k^l]_i = \widetilde{\mu}_i^l \geq 0,\ \forall\ i \in \widetilde{I}_l.$$

Então, tomando limite em 4.12 para $k \in K_2$, obtemos que x^* é ponto KKT.

Suponha agora que ocorra o segundo caso. Seja K_3 subconjunto infinito de K_1 tal que $\lim_{k\in K_3} S_k = \infty$ e $S_k > 1\ \ \forall\ k \in K_3$. Dividindo ambos os lados da equação 4.12 por S_k para todo $k \in K_3$, obtemos:

$$\frac{E_k - F(x_k)}{S_k} = \sum_{i\in\widetilde{I}} \frac{[\widetilde{\lambda}_k]_i}{S_k}\nabla h_i(x_k) + \sum_{i\in\widetilde{I}_u} \frac{[\widetilde{\mu}_k^u]_i}{S_k} e_i - \sum_{i\in\widetilde{I}_l} \frac{[\widetilde{\mu}_k{}^l]_i}{S_k} e_i. \tag{4.14}$$

Pela definição de S_k, essa quantidade é o módulo de um dos coeficientes $[\widetilde{\lambda}_k]_i$, $[\widetilde{\mu}_k^u]_i$, $[\widetilde{\mu}_k{}^l]_i$ que ocorre em (4.14). Portanto, podemos extrair um subconjunto infinito $K_4 \subseteq K_3$ tal que , para todo $k \in K_4$, S_k é o módulo do mesmo coeficiente. Mas, como

$$\left|\frac{[\widetilde{\lambda}_k]_i}{S_k}\right| \leq 1,\ \left|\frac{[\widetilde{\mu}_k^u]_i}{S_k}\right| \leq 1,\ \left|\frac{[\widetilde{\mu}_k{}^l]_i}{S_k}\right| \leq 1,$$

existe K_5, subconjunto infinito de K_4, tal que

$$\lim_{k\in K_5}\frac{[\widetilde{\lambda}_k]_i}{S_k} = \widetilde{\lambda}_i,\ \lim_{k\in K_5}\frac{[\widetilde{\mu}_k^u]_i}{S_k} = \widetilde{\mu}_i^u \geq 0,\ \lim_{k\in K_5}\frac{[\widetilde{\mu}_k{}^l]_i}{S_k} = \widetilde{\mu}_i{}^l \geq 0.$$

Tomando limite em ambos os lados da equação 4.14 para $k \in K_5$, obtemos que

$$\sum_{i\in\widetilde{I}} \widetilde{\lambda}_i \nabla h_i(x^*) + \sum_{i\in\widetilde{I}_u} \widetilde{\mu}_i^u e_i - \sum_{i\in\widetilde{I}_l} \widetilde{\mu}_i^l e_i = 0.$$

Além disso, pela escolha de K_4, o módulo de pelo menos um dos coeficientes $\widetilde{\lambda}_i$, $\widetilde{\mu}_i^u$, $\widetilde{\mu}_i^l$ é igual a 1. Logo, pela condição CPLD, os gradientes

$$\{\nabla h_i(x)\}_{i\in\widetilde{I}},\ \{e_i\}_{i\in\widetilde{I}_u},\ \{-e_i\}_{i\in\widetilde{I}_l}$$

devem ser linearmente dependentes numa vizinhança de x^*. Mas isso contradiz (4.13). Portanto, $\{S_k\}_{k\in K_1}$ é limitada e conforme prova do caso anterior, limitado, x^* é KKT.

$\blacksquare$

Este resultado é relevante pois, antes de ser apresentado em [9], a convergência do Algoritmo de Lagrangeano aumentado era obtida sob a condição MFCQ em relação ao ponto limite do algoritmo. Relembrando que, da Figura 1, temos que MFCQ é uma condição de qualificação mais forte do que a CPLD.

Capítulo 5

Conclusão

Neste trabalho, o objeto de estudo foi condições de otimalidade e condições de qualificação. Iniciamos a abordagem pelo problema de minimização sobre conjuntos irrestritos, concluindo condições a que um minimizador local deve satisfazer, bem como características que pontos devem cumprir a fim de serem candidatos a tais requeridos pontos. Do fato de que essa classe de problema não tem grande aplicabilidade na realidade, adentramos na minimização sobre conjuntos com restrições, passando por igualdade, desigualdade e conjunção destas. Note-se que as conclusões são análogas nessas hipóteses de restrições, com modificações pequenas e ideias parecidas.

Destaco os resultados de condições necessárias e suficientes, tratados no Teorema 2.2.9 e Corolário 2.2.10 para conjuntos de igualdade e posteriores modificações para as demais restrições. O Teorema de Karush-Kuhn-Tucker é importante computacional e teoricamente, e a demonstração é feita supondo a independência linear dos gradientes no ponto analisado. No entanto, aplicabilidade em algoritmos práticos é, por muitas vezes, difícil. Donde faz-se necessário o estudo de outras maneiras de determinar pontos minimizadores locais.

Sob esse aspecto, estudamos condições de qualificação, as quais quando aplicadas junto com a condição de um ponto ser minimizador local implica no Teorema de Karush-Kuhn-Tucker. Verificamos que a condição de qualificação de dependência linear positiva implica quasenormalidade, discutimos as implicações entre elas e verificado que algumas são realmente mais fortes que outras.

Com o objetivo de ilustrar uma aplicação, na Seção 4 estudamos um algoritmo de Lagrangeano Aumentado sob a condição de qualificação de dependência linear positiva constante.

No que concerne o estudo envolvido neste trabalho, ressalto que foi relevante para a cultura matemática, bem como desenvolvimento do gosto pelo saber, despertando interesse pela área de Otimização Matemática, a qual pretendo seguir.

Referências Bibliográficas

[1] Luenberger, David G. *Linear and nonlinear programming*. 3. ed. Boston: Ed. Spring, 2008.

[2] Martínez, M. M., Santos, S. A. *Métodos Computacionais de Otimização*. 2. ed. 247p. Departamento de Matemática Aplicada IMECC-UNICAMP, 1998.

[3] Lima, E. L. *Curso de Análise*. Volume 2. Rio de Janeiro: IMPA, 2001.

[4] Schuverdt, M. L., Martínez J. M., Andreani R. *The Constant Positive Linear Dependence condition of Qi and Wei implies the quasinormality constraint qualification*. Campinas, 2004.

[5] Izmailov, A., Solodov, M. *Otimização - Volume* 1. Volume 1. Rio de Janeiro: IMPA, 2005.

[6] Andreani, R., Martínez, J. M., Schuverdt, M. L. *On the relation between Constant Positive Linear Dependence condition and quasinormality constraint qualification*, Journal of Optimization Theory and Applications 125, p. 473–485, 2005.

[7] Qi ,L., Wei, Z. *On the constant positive linear dependence condition and its application to SQP methods*, SIAM Journal on Optimization 10, p. 963–981, 2000.

[8] Janin. R. *Directional derivative of the marginal function in nonlinear programming*, Mathematical Programming Study 21, p.110 – 126, 1984.

[9] Andreani, R., Birgin, E. G., Martínez, J. M., Schuverdt, M. L. *Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification*, Mathematical Programming manuscript - vol.111, p. 5 – 32, 2008.

[10] Conn, A. R., Gould, N. I. M., Toint, Ph. L. *A globally convergent Augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM Journal on Numerical Analysis 28, p. 545–572, 1991.