KKT à luz da Independência Linear das Restrições

André Luiz Zanin da Cruz - Universidade Estadual do Oeste do Paraná (UNIOESTE)
Paulo Domingos Conejo - Universidade Estadual do Oeste do Paraná (UNIOESTE)

(*Recebido em 14/11/2022. Aceito em 19/12/2022. Publicado em 22/12/2022*)

Resumo: As condições de Karush-Kuhn-Tucker são conhecidas por oferecerem uma condição necessária na busca de soluções para problemas da otimização mono objetivos, sujeitos a restrições de desigualdades. No entanto, essas condições sozinhas são insuficientes para garantir o status de condição de otimalidade. Para corrigir essa debilidade, são necessárias condições de qualificações que permitem a verificação de KKT. Neste trabalho utilizamos algumas das heranças de um vetor derivada dado pelo Teorema da Função Implícita, com objetivo de provar as condições de KKT à luz da qualificação de Independência Linear dos gradientes das restrições.

Palavras-chave: Otimização; Karush-Kuhn-Tucker; Condição de Qualificação.

1 Introdução

Otimização Matemática é a arte de, com base em algum critério, determinar o melhor elemento em um conjunto de alternativas dado. Mais especificamente, um problema de otimização consiste em minimizar ou maximizar uma determinada função f, denominada função objetivo, sujeita a determinadas condições Ω, denominadas restrições (RIBEIRO; KARAS, 2013). É usualmente escrito na forma

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & x \in \Omega, \end{array} \tag{1}$$

com o denominado conjunto viável $\Omega = \{x \in \mathbb{R}^n \,|\, h(x) = 0 \text{ e } g_j(x) \leq 0\}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g_j : \mathbb{R}^n \to \mathbb{R}$ para $j \in \{1, 2, \ldots, p\}$ e $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ funções diferenciáveis. Quando o conjunto Ω é o espaço $n-$dimensional, o problema de otimização é dito irrestrito. Se a função objetivo, ou uma das funções que descrevem as restrições é não linear, então o problema é dito de Programação Não Linear.

Na otimização irrestrita, as condições necessárias e suficientes de primeira e segunda ordem são formuladas a partir das derivadas da função objetivo, não necessitando de hipóteses adicionais em seus enunciados para torná-las condições de otimalidade (RIBEIRO; KARAS, 2013). Já na otimização restrita, segundo Marchand (2016), quando (1) tem somente restrições de igualdades, o matemático Joseph Louis Lagrange mostrou que se $x^0 \in \Omega$ é solução do problema, como consequência existirão vetores $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p$, conhecidos como multiplicadores de Lagrange, de modo que podemos escrever $\nabla f(x^0)$ como uma combinação linear dos vetores gradientes das restrições aplicados nesse ponto, isto é,

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

Esta condição é conhecida como condição de otimalidade de Lagrange, reiterando que vale apenas quando as restrições do Problema (1) são de igualdades. Para um problema com restrições gerais de igualdades e desigualdades, alguns elementos devem ser inseridos na otimalidade para obtermos as denominadas condições de Karush-Kuhn-Tucker (KKT) que é o objeto de estudo desse trabalho.

Sendo considerado por muitos estudiosos como o resultado central da Otimização, o teorema de KKT é constituído por testes de primeira derivadas, denominadas também de condições necessárias de primeira ordem para a solução de um problema geral de otimização não linear, sob alguma condição de qualificação. Entendemos aqui que condição de qualificação é qualquer hipótese que permite demonstrar as condições de KKT.

Ainda antes da década de 50, as condições de otimalidade foram estudadas para o Problema (1) com as restrições Ω gerais de desigualdades primeiro por Fritz John (JOHN, 1948) e, em seguida, por Harold W. Kuhn e Albert W. Tucker (KUHN; TUCKER, 1951). Mais tarde descobriu-se que estas condições de Kuhn-Tucker já haviam sido estabelecidas anteriormente por William Karush em sua dissertação de mestrado pelo Departamento de Matemática da Universidade de Chicago (HAESER, 2015). Hoje, as condições de otimalidade para restrições gerais são conhecidas como condições de KKT (Karush-Kuhn-Tucker).

Uma das hipóteses mais conhecidas, e a que será nosso foco, é a Condição de Qualificação de Independência Linear (LICQ). Essa condição se caracteriza por sua simplicidade, mas tem a desvantagem de ser uma hipótese muito forte para garantir KKT. Assim, existem diversos problemas que cumprem KKT, mas LICQ não é satisfeita, como mostrado ao final do artigo.

O objetivo neste texto é ir de Lagrange à KKT, utilizando propriedades de um vetor tangente dado pelo Teorema da Função Implícita, diferenciando pontualmente a demonstração apresentada neste trabalho, daquela apresentada em Luenberger (2008), quando a condição de qualificação é a de Independência Linear das restrições (LICQ). Não obstante, o texto deixa completo uma demonstração de KKT desde o Teorema dos Multiplicadores de Lagrange.

2 Introdução à Otimização

Nesta seção apresentamos alguns conceitos e resultados de Otimização e de Análise no $\mathbb{R}^n$ fundamentais para compreender e utilizar o teorema de KKT. Supomos que os resultados clássicos da Álgebra Linear sejam conhecidos, como por exemplo os teoremas de núcleo imagem e soma direta. Informações mais detalhadas podem ser encontradas em Elon (2004), Luenberger (2008), Ribeiro e Karas (2013) e Steinbruch (1987).

Definição 1. (Minimizador global/local) Considere uma função $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ e $x^0 \in \Omega$. Dizemos que x^0 é um minimizador local de f em Ω quando existe $\delta > 0$, tal que $f(x^0) \leq f(x)$, para todo $x \in B(x^0, \delta) \cap \Omega$. Caso $f(x^0) \leq f(x)$, para todo $x \in \Omega$, x^0 é dito minimizador global de f em Ω.

Observemos que um ponto de mínimo global é sempre um ponto de mínimo local, e dentre todos os pontos de mínimos locais, podemos ter um ou mais que são pontos de mínimo global.

De acordo com Haeser (2015), o termo *condição de otimalidade* é usado para se referir a qualquer condição satisfeita por uma solução de (1). Logo, por exemplo, a viabilidade ou a otimalidade local são condições de otimalidade. Identificar um ponto de mínimo local é muito difícil, por isso precisamos procurar por pontos que validem alguma propriedade que também é validada pelos mínimos locais. Isto é feito para o caso irrestrito na próxima definição, o que é também exemplo, portanto, de condição de otimalidade.

Definição 2. (Ponto estacionário) Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável, em $x^0 \in \mathbb{R}^n$ e satisfaz a condição de $\nabla f(x^0) = 0$, então x^0 é dito ser um ponto crítico (ou estacionário) de f.

Conjuntos compactos carregam propriedades relevantes na otimização.

Definição 3 (Conjunto Compacto)**.** Um conjunto $\Omega \subset \mathbb{R}^n$ é compacto quando Ω é fechado e limitado.

O próximo resultado garante que, se uma função for contínua em um conjunto compacto, então ela assume um valor máximo e um valor mínimo global nesse compacto. Esse teorema é utilizado na comprovação da existência de extremos globais para o Teorema dos multiplicadores de Lagrange e até de KKT.

Teorema 4 (Teorema de Weierstrass)**.** *Sejam $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ contínua e Ω compacto e não vazio. Então existe minimizador global de f em Ω.*

Prova. Em Ribeiro e Karas (2013). □

Na demonstração do Teorema de KKT apresentada neste texto, precisamos de resultados sobre planos tangentes a alguma superfície (hiperfície de nível zero) em um determinado ponto.

Definição 5. Um vetor que é ortogonal a todo vetor que seja tangente a alguma curva C da superfície S no ponto p é denominado de vetor normal a S no ponto p.

Teorema 6. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ e a equação de uma superfície S for dada por $f(x) = 0$, com todas as derivadas parciais contínuas e nem todas nulas no ponto $p \in \mathbb{R}^n$ em S, então o vetor normal a S no ponto p será $\nabla f(p)$.*

Prova. Fixado $p \in S$, devemos mostrar que $\nabla f(p)$ é ortogonal a S, isto é, mostrar que $\nabla f(p)$ é ortogonal a qualquer curva diferenciável sobre S que passe por p.

Considere $r(t) = (r_1(t), r_2(t), \ldots, r_n(t))$ uma curva diferenciável qualquer sobre S passando por p. Logo, por estar sobre S, a curva $r(t)$ satisfaz a $f(r_1(t), r_2(t), \ldots, r_n(t)) = 0$, e suponha $r(t_0) = p$. Derivando a equação $f(r_1(t), r_2(t), \ldots, r_n(t)) = 0$ em relação a t, teremos

$$\frac{\partial f(x)}{\partial r_1}\frac{dr_1(t)}{dt} + \frac{\partial f(x)}{\partial r_2}\frac{dr_2(t)}{dt} + \cdots + \frac{\partial f(x)}{\partial r_n}\frac{dr_n(t)}{dt} = 0.$$

Calculando em t_0, obtemos

$$\left\langle \left(\frac{\partial f(p)}{\partial r_1}, \frac{\partial f(p)}{\partial r_2}, \ldots, \frac{\partial f(p)}{\partial r_n}\right), \left(\frac{dr_1(t_0)}{dt}, \frac{dr_2(t_0)}{dt}, \ldots, \frac{dr_n(t_0)}{dt}\right) \right\rangle = 0,$$

que é $\langle \nabla f(p), r'(t_0) \rangle = 0$, portanto, $\nabla f(p) \perp r'(t_0)$. Assim, $\nabla f(p)$ é ortogonal a todos os vetores tangentes a S no ponto p. $\square$

Em seguida, introduzimos o conceito de direção de descida que é o vetor no qual a função estudada sempre decresce.

Definição 7. Considere uma função $f : \mathbb{R}^n \to \mathbb{R}$, um ponto $\overline{x} \in \mathbb{R}^n$ e um vetor direção $d \in \mathbb{R}^n - \{0\}$. Dizemos que d é direção de descida para f, a partir de $\overline{x}$, quando existe um $\theta > 0$ tal que $f(\overline{x} + td) < f(\overline{x}), \forall\, t \in (0, \theta)$.

O próximo resultado caracteriza direção de descida, a partir do vetor gradiente.

Teorema 8. *Se f é diferenciável em x e $\nabla f(\overline{x})^T d < 0$ então d é uma direção de descida para f, a partir de $\overline{x}$.*

Prova. Em Ribeiro e Karas (2013). $\square$

A seguir apresentamos o Teorema da Função Implícita que é utilizado na demonstração dos teoremas de multiplicadores de Lagrange em otimização de problemas com restrição de igualdade, e que é essencial para estabelecer o teorema de KKT com restrições gerais. Esse teorema caracteriza as condições locais para que de uma relação $G(x, y) = 0$, com $x \in \mathbb{R}^n$ e $y \in \mathbb{R}$, consigamos definir $y = y(x)$. A prova do Teorema da Função Implícita pode ser feita a partir do Teorema da Função Inversa, ou independente deste. Mais detalhes em Elon (2004).

Teorema 9 (Teorema da Função Implícita)**.** *Seja $G : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$ uma função de classe C^1. Suponha que $G(\overline{x}, \overline{y}) = 0$ e*

$$\left|\frac{\partial G(\overline{x}, \overline{y})}{\partial y}\right| \neq 0.$$

Então, existe um aberto $W \subset \mathbb{R}^k$ e $\varphi : W \to \mathbb{R}^m$ função de classe C^1 tais que

I) $\overline{x} \in W$ e $\varphi(\overline{x}) = \overline{y}$

II) $G(x, \varphi(x)) = 0, \forall\, x \in W$.

Prova. Em Elon (2004). $\square$

O próximo resultado é uma consequência do Teorema da Função Implícita, e garante a existência de uma curva diferenciável em uma superfície formada por $H_i(x) = 0$, $\forall i \in \{1, 2, \ldots, m\}$. Usaremos esse resultado mais a frente quando definimos plano tangente e provamos o Teorema de KKT.

Corolário 10. *Sejam $H : \mathbb{R}^n \to \mathbb{R}^m$, os vetores $\overline{x}$ e d pertencentes ao $\mathbb{R}^n$ e o conjunto de índices da função H sendo $\nu = \{1, 2, \ldots, m\}$ com $m < n$, tais que $H_\nu(\overline{x}) = 0$ e $\nabla H_i(\overline{x})^T d = 0$, para todo $i \in \nu$. Suponha que os gradientes $\nabla H_i(\overline{x})$, $i \in \nu$, sejam Linearmente Independentes. Então, existe uma curva diferenciável $\varphi : (-\epsilon, \epsilon) \to \mathbb{R}^n$ tal que $H_\nu(\varphi(t)) = 0$, para todo $t \in (-\epsilon, \epsilon)$, $\varphi(0) = \overline{x}$ e $\varphi'(0) = d$.*

Prova. Consideramos $\overline{x} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{n-m}, \ldots, \overline{x}_n)$. Temos que a matriz

$$M = (\nabla H_1(\overline{x}) \cdots \nabla H_m(\overline{x}))^T = \begin{pmatrix} \frac{\partial H_1(\overline{x})}{\partial x_1} & \frac{\partial H_2(\overline{x})}{\partial x_1} & \cdots & \frac{\partial H_m(\overline{x})}{\partial x_1} \\ \frac{\partial H_1(\overline{x})}{\partial x_2} & \frac{\partial H_2(\overline{x})}{\partial x_2} & \cdots & \frac{\partial H_m(\overline{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_1(\overline{x})}{\partial x_n} & \frac{\partial H_2(\overline{x})}{\partial x_n} & \cdots & \frac{\partial H_m(\overline{x})}{\partial x_n} \end{pmatrix} \in \mathbb{R}^{n \times m},$$

têm posto m, pois todos esses vetores gradientes são Linearmente Independentes.

Pelo Teorema do núcleo e imagem, a dimensão do núcleo de M^T é igual a

$$dim\ (Nuc\ (M^T)) = n - m.$$

Então, podemos afirmar que existe uma matriz $Z \in \mathbb{R}^{n \times (n-m)}$, cujas colunas geram o núcleo da matriz M^T. Aplicando o complementar na igualdade $Nuc\ (M^T)^\perp = Im\ (M)$, obtemos

$$(Nuc\ (M^T)^\perp)^\perp = Im\ (M)^\perp \Rightarrow Nuc\ (M^T) = Im\ (M)^\perp \subseteq \mathbb{R}^n,$$

logo pelo teorema da soma direta, temos da equação anterior que

$$\mathbb{R}^n = Nuc\ (M^T) \oplus Nuc\ (M^T)^\perp = Nuc\ (M^T) \oplus Im\ (M).$$

Portanto, a matriz $(MZ)_{n \times n}$ gera o $\mathbb{R}^n$, já que $dim\ (Im\ (M)) = m$ e $dim\ (Z) = n - m = dim\ (Nuc\ (M^T))$.

Como (MZ) gera o $\mathbb{R}^n$, então ela tem colunas LI e, portanto, é inversível. Definimos a seguinte função auxiliar, $G : \mathbb{R}^{n+1} \to \mathbb{R}^n$ dada por

$$G\begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} H_\nu(x) \\ Z^T(x - \overline{x} - td) \end{pmatrix} = \begin{pmatrix} H_1(x) \\ H_2(x) \\ \vdots \\ H_m(x) \\ Z^T(x - (\overline{x} + td)) \end{pmatrix} \in \mathbb{R}^n.$$

Como $\dfrac{\partial G}{\partial x}\begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} \nabla H_1(x) \\ \vdots \\ \nabla H_m(x) \\ Z^T \end{pmatrix} = (MZ)$ é inversível e $G\begin{pmatrix} \overline{x} \\ 0 \end{pmatrix} = 0$, o Teorema 9 (Teorema da Função Implícita) garante a existência de uma curva diferenciável $\varphi : (-\epsilon, \epsilon) \to \mathbb{R}^n$

tal que $G\begin{pmatrix}\varphi(t)\\ t\end{pmatrix}=0$, para todo $t\in(-\epsilon,\epsilon)$. Assim,

$$H_\nu(\varphi(t))=0 \text{ e } Z^T(\varphi(t)-\overline{x}-td)=0.$$

Segue do Teorema da Função Implícita que $\varphi(0)=\overline{x}$. Derivando a equação $H_\nu(\varphi(t))=0$ em $t=0$, obtemos

$$\begin{pmatrix}\frac{\partial H_1(\overline{x})}{\partial x_1}*\frac{d\varphi_1(0)}{dt}+\frac{\partial H_1(\overline{x})}{\partial x_2}*\frac{d\varphi_2(0)}{dt}+\cdots+\frac{\partial H_1(\overline{x})}{\partial x_n}*\frac{d\varphi_n(0)}{dt}\\ \frac{\partial H_2(\overline{x})}{\partial x_1}*\frac{d\varphi_1(0)}{dt}+\frac{\partial H_2(\overline{x})}{\partial x_2}*\frac{d\varphi_2(0)}{dt}+\cdots+\frac{\partial H_2(\overline{x})}{\partial x_n}*\frac{d\varphi_n(0)}{dt}\\ \vdots\\ \frac{\partial H_m(\overline{x})}{\partial x_1}*\frac{d\varphi_1(0)}{dt}+\frac{\partial H_m(\overline{x})}{\partial x_2}*\frac{d\varphi_2(0)}{dt}+\cdots+\frac{\partial H_m(\overline{x})}{\partial x_n}*\frac{d\varphi_n(0)}{dt}\end{pmatrix}=\begin{pmatrix}\nabla H_1(\overline{x})\varphi'(0)\\ \nabla H_2(\overline{x})\varphi'(0)\\ \vdots\\ \nabla H_m(\overline{x})\varphi'(0)\end{pmatrix}=$$

$M^T*\varphi'(0)=0.$

Dividindo a segunda equação $Z^T(\varphi(t)-\overline{x}-td)=0$ por $t\neq 0$ e tomando o limite quando $t\to 0$, vamos ter $\lim_{t\to 0}Z^T(\frac{\varphi(t)}{t}-\frac{\varphi(0)}{t}-\frac{td}{t})=Z^T(\varphi'(0)-d)=0$. Como por hipótese $M^Td=0$, igualando as duas últimas equações, obtemos

$$\begin{pmatrix}M^T\\ Z^T\end{pmatrix}\varphi'(0)=\begin{pmatrix}M^T\\ Z^T\end{pmatrix}d,$$

donde segue que $\varphi'(0)=d$, completando a prova. $\square$

Desenvolvido pelo Matemático franco-italiano Joseph Louis Lagrange, o Teorema dos Multiplicadores de Lagrange permite transformar um problema de otimização com restrições de igualdades em um problema da otimização irrestrita, essa modificação é realizada pela inserção dos multiplicadores de Lagrange.

A partir desse teorema, um problema com n variáveis e m restrições de igualdades pode ser transformado em um problema irrestrito com $n+m$ variáveis (MARCHAND, 2016). O Teorema de KKT procura unir tanto a otimização irrestrita ao procurar por mínimos locais no interior de Ω quanto a otimização com restrições de igualdades ao buscar também por mínimos locais na fronteira desse conjunto, isto é, onde as restrições de igualdades são satisfeitas. Neste contexto, podemos dizer que KKT foi desenvolvido usando o Teorema dos Multiplicadores Lagrange como base, uma vez que ambos compartilham algumas propriedades, dentre elas, como veremos adiante, a mesma condição de regularidade.

Com os resultados apresentados até este momento podemos provar o Teorema dos Multiplicadores de Lagrange para o caso geral com m restrições de igualdades. A demonstração apresentada aqui foi retirada de William (2013), onde para simplificar notação, a condição de regularidade no Teorema 11 foi tomada considerando as m primeiras variáveis.

Teorema 11 (Multiplicadores de Lagrange - m restrições)**.** *Sejam $f,g:D\subset\mathbb{R}^n\to\mathbb{R}$ funções diferenciáveis de classe C^1. Suponha que $n>m$. Se x^0 é ponto de extremo local de f sujeito as* *restrições* $g_1(x) = 0$, $g_2(x) = 0$, $g_3(x) = 0$, *...*, $g_m(x) = 0$, *mais a condição de regularidade*

$$\begin{vmatrix} \dfrac{\partial g_1(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_1(x^0)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_m(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{vmatrix} \neq 0. \tag{2}$$

Então existem as constantes reais $\lambda_1, \lambda_2, \cdots, \lambda_m$ *tais que para* $1 \leq i \leq n$,

$$\frac{\partial f(x^0)}{\partial x_i} - \lambda_1 \frac{\partial g_1(x^0)}{\partial x_i} - \lambda_2 \frac{\partial g_2(x^0)}{\partial x_i} - \cdots - \lambda_m \frac{\partial g_m(x^0)}{\partial x_i} = 0. \tag{3}$$

Prova. Denotamos $u = (x_{m+1}, x_{m+2}, \cdots, x_n)$ e $u^0 = (x^0_{m+1}, x^0_{m+2}, \ldots, x^0_n)$.

De (2), o Teorema 9 afirma que existem funções com derivadas contínuas $\varphi_\ell(u)$, $1 \leq \ell \leq m$, definidas numa vizinhança $W \subset \mathbb{R}^{n-m}$ do ponto u^0, de tal modo que

$$(\varphi_1(u), \varphi_2(u), \ldots, \varphi_m(u), u) \in D, \forall\ u \in W,$$

$$(\varphi_1(u^0), \varphi_2(u^0), \ldots, \varphi_m(u^0), u^0) = x^0, \tag{4}$$

e

$$g_\ell(\varphi_1(u), \varphi_2(u), \ldots, \varphi_m(u), u) = 0,\ u \in W, 1 \leq \ell \leq m. \tag{5}$$

Novamente de (2), temos o sistema

$$\begin{pmatrix} \dfrac{\partial g_1(x^0)}{\partial x_1} & \dfrac{\partial g_1(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_1(x^0)}{\partial x_m} \\ \dfrac{\partial g_2(x^0)}{\partial x_1} & \dfrac{\partial g_2(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_2(x^0)}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_m(x^0)}{\partial x_1} & \dfrac{\partial g_m(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} f_{x_1}(x^0) \\ f_{x_2}(x^0) \\ \vdots \\ f_{x_m}(x^0) \end{pmatrix} \tag{6}$$

que possui solução única. Isso implica em

$$\frac{\partial f(x^0)}{\partial x_i} - \lambda_1 \frac{\partial g_1(x^0)}{\partial x_i} - \lambda_2 \frac{\partial g_2(x^0)}{\partial x_i} - \cdots - \lambda_m \frac{\partial g_m(x^0)}{\partial x_i} = 0$$

para todo $1 \leq i \leq m$. Agora se $m + 1 \leq i \leq n$, derivamos (5) seguindo a regra da cadeia e utilizando (4) obtemos

$$\frac{\partial g_\ell(x^0)}{\partial x_i} + \sum_{j=1}^{m} \frac{\partial g_\ell(x^0)}{\partial x_j} \frac{\partial \varphi_j(x^0)}{\partial x_i} = 0, \quad 1 \leq \ell \leq m. \tag{7}$$

Se x^0 é um ponto de extremo local de f restrita aos pontos de $g_1(x) = 0$, $g_2(x) = 0$, ..., $g_m(x) = 0$, então u^0 é um ponto de extremo local irrestrito de

$$f(\varphi_1(u), \varphi_2(u), \ldots, \varphi_m(u), u).$$

Assim, temos que

$$\frac{\partial f(x^0)}{\partial x_i} + \sum_{j=1}^{m} \frac{\partial f(x^0)}{\partial x_j} \frac{\partial \varphi_j(x^0)}{\partial x_i} = 0, \quad 1 \leq \ell \leq m. \tag{8}$$

Das últimas equações (7) e (8), vamos ter a seguinte matriz com o seu determinante igual a zero

$$\begin{vmatrix} \dfrac{\partial f(x^0)}{\partial x_i} & \dfrac{\partial f(x^0)}{\partial x_1} & \cdots & \dfrac{\partial f(x^0)}{\partial x_m} \\ \dfrac{\partial g_1(x^0)}{\partial x_i} & \dfrac{\partial g_1(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_1(x^0)}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_m(x^0)}{\partial x_i} & \dfrac{\partial g_m(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{vmatrix} = 0.$$

Portanto, existem constantes reais $c_0, c_1, c_2, c_3, \ldots, c_m$, nem todas iguais a zero, tais que

$$\begin{pmatrix} \dfrac{\partial f(x^0)}{\partial x_i} & \dfrac{\partial g_1(x^0)}{\partial x_i} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_i} \\ \dfrac{\partial f(x^0)}{\partial x_1} & \dfrac{\partial g_1(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f(x^0)}{\partial x_m} & \dfrac{\partial g_1(x^0)}{\partial x_m} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{9}$$

Se $c_0 = 0$, então

$$\begin{pmatrix} \dfrac{\partial g_1(x^0)}{\partial x_1} & \dfrac{\partial g_1(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_1(x^0)}{\partial x_m} \\ \dfrac{\partial g_2(x^0)}{\partial x_1} & \dfrac{\partial g_2(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_2(x^0)}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_m(x^0)}{\partial x_1} & \dfrac{\partial g_m(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

e (2) implica que $c_1 = c_2 = c_3 = \cdots = c_m = 0$. Portanto, podemos supor que $c_0 = 1$ em uma solução não trivial de (9). Assim, obtemos

$$\begin{pmatrix} \dfrac{\partial f(x^0)}{\partial x_i} & \dfrac{\partial g_1(x^0)}{\partial x_i} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_i} \\ \dfrac{\partial f(x^0)}{\partial x_1} & \dfrac{\partial g_1(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f(x^0)}{\partial x_m} & \dfrac{\partial g_1(x^0)}{\partial x_m} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{pmatrix} \begin{pmatrix} 1 \\ c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{10}$$

que implica em

$$\begin{pmatrix} \dfrac{\partial g_1(x^0)}{\partial x_1} & \dfrac{\partial g_1(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_1(x^0)}{\partial x_m} \\ \dfrac{\partial g_2(x^0)}{\partial x_1} & \dfrac{\partial g_2(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_2(x^0)}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_m(x^0)}{\partial x_1} & \dfrac{\partial g_m(x^0)}{\partial x_2} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{pmatrix} \begin{pmatrix} -c_1 \\ -c_2 \\ \vdots \\ -c_m \end{pmatrix} = \begin{pmatrix} f_{x_1}(x^0) \\ f_{x_2}(x^0) \\ \vdots \\ f_{x_m}(x^0) \end{pmatrix},$$

E como (6) têm apenas uma solução, isso significa que $c_j = -\lambda_j$, $1 \leq j \leq m$, então de (10) teremos

$$\begin{pmatrix} \dfrac{\partial f(x^0)}{\partial x_i} & \dfrac{\partial g_1(x^0)}{\partial x_i} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_i} \\ \dfrac{\partial f(x^0)}{\partial x_1} & \dfrac{\partial g_1(x^0)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f(x^0)}{\partial x_m} & \dfrac{\partial g_1(x^0)}{\partial x_m} & \cdots & \dfrac{\partial g_m(x^0)}{\partial x_m} \end{pmatrix} \begin{pmatrix} 1 \\ -\lambda_1 \\ \vdots \\ -\lambda_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Calculando a multiplicação na primeira linha nesta última igualdade obtemos (3), o que completa a prova. $\square$

3 Teorema de Karush-Kuhn-Tucker

Nesta seção temos interesse no problema de otimização com restrições gerais de igualdades e desigualdades dado na forma

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & x \in \Omega \end{array} \tag{11}$$

com $\Omega = \{x \in \mathbb{R}^n \,|\, h(x) = 0 \text{ e } g(x) \leq 0\}$ tal que $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ e $h : \mathbb{R}^n \to \mathbb{R}^m$ são funções de classe C^1. As principais referências utilizadas nesta seção foram Luenberger (2008), Martínez (2006) e Ribeiro e Karas (2013).

Durante esta seção, quando nos referirmos às funções f, h e g, estaremos supondo que sejam definidas como no problema (11).

Definição 12 (Restrições ativas e inativas)**.** Uma restrição de desigualdade é denominada ativa se $g_i(x) = 0$ e inativa se $g_i(x) < 0$, para todo $i \in \{1, 2, \ldots, p\}$.

Formalizamos agora a definição da condição de qualificação de Independência Linear (LICQ). Esta condição é muito utilizada, e dentre os motivos que podem explicar esta fama estão o fato de que pode ser facilmente enunciada e verificada, e também, por ser utilizada na análise de diversos algoritmos práticos como, por exemplo, o algoritmo de Lagrangeano aumentado apresentado em Martínez (2006).

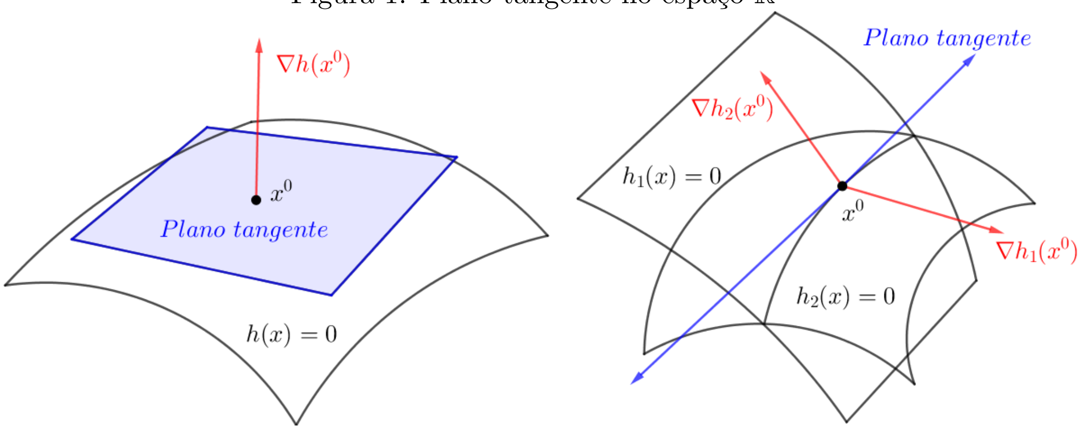
Definição 13. (Condição de Qualificação de Independência Linear - LICQ) Um ponto $\overline{x} \in \Omega$ é um ponto regular (LICQ) das restrições $h(x) = 0$ e $g(x) \leq 0$, se os vetores $\nabla h_i(\overline{x})$ para $i \in \{1, 2, \ldots, m\}$ e $\nabla g_j(\overline{x})$ para $j \in J(\overline{x})$ são Linearmente Independentes, onde $J(\overline{x}) = \{j \mid g_j(\overline{x}) = 0\}$ é o conjunto das restrições ativas de g em $\overline{x}$.

Notemos que, quando esta condição de regularidade (ou qualificação) LICQ é satisfeita, nenhum dos gradientes das restrições ativas podem ser nulo.

Vamos considerar agora uma hiperfície $S = \{x \in \mathbb{R}^n \,|\, h_i(x) = 0, \ \forall\ i \in \{1, 2, \ldots, m\}\}$, ela será dita suave quando todo $h_i(x)$ for de classe C^1. Seja $\overline{x}$ um ponto pertencente a essa hiperfície suave, podemos considerar um plano T passando por ele e tangenciando S. Na Figura 1 podemos observar essa situação em três dimensões. Formalizamos o conceito de plano T tangente a S no ponto $\overline{x}$, definindo curvas nessa superfície. Uma curva em uma superfície é uma família de pontos $\varphi(t) \subset S$ continuamente parametrizada por $t \in (-\delta, \delta)$. Essa curva é diferenciável se $\varphi'(t) = \dfrac{d\varphi(t)}{dt}$ existir, e duas vezes diferenciável se $\varphi''(t) = \dfrac{d^2\varphi(t)}{dt^2}$ existir. Vamos dizer que $\varphi(t)$ passe por $\overline{x}$ quando $t = 0$, com $-\delta \leq t \leq \delta$. Logo, a derivada da curva em $\overline{x}$ é $\varphi'(0)$. Tome agora todas as curvas diferenciáveis em S passando pelo ponto $\overline{x}$, a existência dessas curvas é garantida pelo Teorema 9.

Definição 14 (Plano tangente)**.** Definimos o plano T tangente a S no ponto $\overline{x}$ como a coleção das derivadas neste ponto, de todas as curvas diferenciáveis passando por este ponto.

Figura 1: Plano tangente no espaço $\mathbb{R}^3$



Fonte: Autor (2022)

Notemos que T é um subespaço do $\mathbb{R}^n$. Procuramos agora expressar o plano tangente com as derivadas das funções $h_i = 0$, $\forall\ i \in \{1, 2, \ldots, m\}$ no ponto $\overline{x}$. Introduzindo o subespaço $M = \{y \in \mathbb{R}^n \,|\, \nabla h(\overline{x})^T y = 0\}$, vamos investigar as condições em que M é igual ao plano tangente T, e a peça chave para essa igualdade é a regularidade de $\overline{x}$.

Teorema 15. *Em um ponto regular $\overline{x}$ da superfície S definida por $h(x) = 0$, o plano tangente T é igual ao conjunto $M = \{y \in \mathbb{R}^n \,|\, \nabla h(\overline{x})^T y = 0\}$.*

Prova. Devemos mostrar a princípio que $T \subseteq M$ e $M \subseteq T$. Dada uma curva diferenciável qualquer $\varphi(t) \subset S$, com $\varphi(0) = \overline{x}$ e $\varphi'(0) \in T$. Independente de $\overline{x}$ ser regular ou não, como $\nabla h(\overline{x})$ é vetor normal ao plano tangente (Teorema 6), então $\nabla h(\overline{x})^T \varphi'(0) = 0$ e $\varphi'(0) \in M$, ou seja, $T \subseteq M$.

Através do Corolário 10, consideramos os vetores $\overline{x}$ e $d \in \mathbb{R}^n$ com $\nabla h_i(\overline{x})^T d = 0$, $\forall\ i \in \{1, 2, \dots, m\}$ sendo que todo $\nabla h_i(\overline{x})$ é LI, ou seja, $d \in M$ e $\overline{x}$ é um ponto regular, então vamos ter que $\varphi(0) = \overline{x}$ e $\varphi'(0) = d$, isto é, d vai ser um vetor tangente a alguma curva diferenciável $\varphi(t)$ passando por $\overline{x}$ em $t = 0$, indicando que pertence a T. Assim, $M \subseteq T$ e, portanto, $T = M$, provando o teorema.

□

É importante observar que a condição de $\overline{x}$ ser um ponto regular não é uma condição na própria superfície dada pelas restrições, mas sobre os vetores gradientes de sua representação em termos de $h(x)$. A formação do plano tangente T não depende de $\overline{x}$ ser regular ou não, apenas da derivada das curvas diferenciáveis na superfície que passam por aquele ponto, já M é formado a partir dos gradientes de $h(\overline{x})$, dependendo do ponto ser regular para ser igual a T.

O próximo exemplo retirado de Luenberger (2008) mostra que a relação entre M e T pode não ser de igualdade. Particularmente neste exemplo, escrevemos os vetores em linha para simplificar a notação.

Exemplo 1. Considere a função $\gamma(x, y) = x$ e a superfície $S = \{(x, y) \in \mathbb{R}^2 \,|\, \gamma(x, y) = 0\} = \{(0, y) \in \mathbb{R}^2\}$. Como S é o eixo das ordenadas no plano cartesiano, o plano tangente T coincide com essa superfície em qualquer ponto, isto é, $T = \mathbb{R}$. Agora vamos calcular o conjunto M. Como $\nabla\gamma(x, y) = (1, 0) \neq (0, 0)$, considere $k = (k_1, k_2) \in \mathbb{R}^2$ tal que $k \in M$, então $\nabla\gamma(x, y)k^T = 1 * k_1 + 0 * k_2 = 0$. Assim, $k_1 = 0$ e $k_2 \in \mathbb{R}$, ou seja, $M = \mathbb{R}$. Logo, o plano tangente $T = \mathbb{R} = M$ em qualquer ponto. No entanto, se considerarmos $\gamma(x, y) = x^2$, a superfície S continua sendo a mesma, mas $\nabla\gamma(x, y) = (2x, 0)$ e $\nabla\gamma(0, 0)k^T = 0$ implica em k_1 e k_2 poderem ser quaisquer reais, ou seja, $T \neq M = \mathbb{R}^2$ no ponto $(0, 0) \in M$. Claramente o ponto $(0, 0)$ não é regular, não contrariando o Teorema 15.

Definição 16. Uma direção $d \in \mathbb{R}^n$ é denominada tangente ao conjunto $\Omega \subset \mathbb{R}^n$ a partir de $\overline{x} \in \Omega$ quando é nula ou existe uma sequência de pontos viáveis $(x^k) \subset \Omega$ tal que $x^k \to \overline{x}$ e

$$\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} \to \frac{d}{\|d\|}. \tag{12}$$

O próximo lema garante que na hipótese de existência de uma curva diferenciável, podemos obter uma convergência do tipo (12), mas, ainda, não é direção tangente.

Lema 17. *Seja $\varphi : (-\epsilon, \epsilon) \to \mathbb{R}^n$ uma curva diferenciável tal que $\varphi(0) = \overline{x}$ e $\varphi'(0) = d$. Então existe uma sequência (x^k) tal que $x^k \to \overline{x}$ e*

$$\frac{x^k - \overline{x}}{\|x^k - \overline{x}\|} \to \frac{d}{\|d\|}.$$

Prova. Temos que
$$\lim_{t\to 0}\frac{\varphi(t)-\overline{x}}{t}=\lim_{t\to 0}\frac{\varphi(t)-\varphi(0)}{t}=\varphi'(0)=d\neq 0.$$
Logo, $\varphi(t)\neq\overline{x}$ para todo $t\neq 0$ suficientemente pequeno. Tomamos uma sequência $(t_k)\subset\mathbb{R}_+$, com $t_k\to 0$ e definimos para todo $k\in\mathbb{N}$, $x^k=\varphi(t_k)$. Assim, $x^k\to\overline{x}$ e
$$\frac{x^k-\overline{x}}{\|x^k-\overline{x}\|}=\frac{x^k-\overline{x}}{t_k}\frac{t_k}{\|x^k-\overline{x}\|}\to\frac{d}{\|d\|},$$
completando a prova. □

Como consequência deste lema, esta direção d não é de descida.

Lema 18. *Seja $\varphi:(-\epsilon,\epsilon)\to\mathbb{R}^n$ uma curva diferenciável tal que $\varphi(0)=\overline{x}$ e $\varphi'(0)=d$. Sejam Ω e f do Problema (11). Suponha $x^0\in\Omega$ um minimizador local da f em Ω. Então $\nabla f(x^0)^T d\geq 0$.*

Prova. Pelo Lema 17, $x^k\to x^0$ e
$$\frac{x^k-x^0}{\|x^k-x^0\|}\to\frac{d}{\|d\|}.$$

Por Taylor de primeira ordem, e pelo fato de x^0 ser minimizador local da f, temos
$$0\leq f(x^k)-f(x^0)=\nabla f(x^0)^T(x^k-x^0)+o(\|x^k-x^0\|),$$
para todo k suficientemente grande. Dividindo a expressão anterior por $\|x^k-x^0\|$ e aplicando o limite, obtemos que $\nabla f(x^0)^T d\geq 0$.

□

Podemos observar que, quando situamos esta curva diferenciável φ no conjunto viável Ω de (11), este vetor d é uma direção tangente no sentido da Definiçao 16, e mais, pelo Lema 18, esta direção não é de descida. Assim, temos todos os ingredientes para provar KKT sob LICQ. Notamos que, com a hipótese LICQ, temos o Problema (11) nas hipóteses do Lema 18, apesar que na demonstração de KKT iremos considerar um subconjunto de Ω para situar a curva φ.

Teorema 19 (Condições de Karush-Kuhn-Trucker)**.** *Sejam $f:\Omega\subset\mathbb{R}^n\to\mathbb{R}$, $g:\mathbb{R}^n\to\mathbb{R}^p$ e $h:\mathbb{R}^n\to\mathbb{R}^m$ funções de classe C^1, com $\Omega=\{x\in\mathbb{R}^n\,|\,h(x)=0 \text{ e } g(x)\leq 0\}$. Tome x^0 como ponto de mínimo local para o problema de minimizar f sujeito a $x\in\Omega$, e suponha que x^0 é um ponto regular para as restrições (LICQ). Então, existem vetores $\lambda\in\mathbb{R}^m$ e $\mu\in\mathbb{R}^p$ com $\mu\geq 0$ tais que*
$$\nabla f(x^0)+\lambda^T\nabla h(x^0)+\mu^T\nabla g(x^0)=0 \tag{13}$$
$$\mu^T g(x^0)=0. \tag{14}$$

Prova. Primeiramente, notamos que $\mu\geq 0$ e $g(x^0)\leq 0$, deixa (14) equivalente a afirmação de que uma componente de μ vai ser diferente de zero somente se a restrição correspondente for ativa. Essa é a chamada condição de complementariedade, se afirmado que $g_i(x^0)<0$ temos que $\mu_i=0$ e se $\mu_i>0$ vamos ter $g_i(x^0)=0$, com $i\in\{1,2,\ldots,p\}$.

Como x^0 é um ponto de mínimo local sobre o conjunto das restrições Ω, então também o é sobre o subconjunto de Ω, definindo as restrições ativas como zero. Assim, como x^0 é regular, pelo Teorema 11 para o problema com restrições de igualdade resultante definido em uma vizinhança de x^0, existem multiplicadores de Lagrange $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p$ satisfazendo (13). Se $g_j(x^0) < 0$, necessariamente $\mu_j = 0$ para não atrapalhar a validade de (13). Como consequência, (14) também vale.

Resta mostrar que o multiplicador μ satisfaz $\mu_i \geq 0,\ \forall\, i \in \{1, 2, \ldots, p\}$. Suponhamos por absurdo que exista uma coordenada $\mu_k < 0$ de μ para algum $k \in J(x^0)$, sendo $J(x^0)$ o conjunto de todas as restrições ativas de g em x^0. Considere a superfície $S = \{x \in \mathbb{R}^n \,|\, h(x) = 0 \text{ e } g_j(x) = 0\,, \forall\, j \in J - \{k\}\}$ definida por todas as restrições ativas em x^0, exceto por $g_k(x) = 0$.

Notamos que por x^0 ser ponto regular das restrições, temos que o conjunto definido por $V = \{\nabla h_i(x^0) \text{ para todo } i \in \{1, \ldots, m\} \text{ e } \nabla g_j(x^0) \text{ para } j \neq k,\ j = 1, \ldots, p\}$ é Linearmente Independentes, ou seja, $\nabla g_k(x^0)$ não é combinação linear dos vetores de V.

Logo, pelo Teorema 15 o plano tangente a S no ponto x^0 é dado por

$$M = \{y \in \mathbb{R}^n \,|\, Vy = 0\}. \tag{15}$$

De (15), existe $d \in M$ tal que

$$\nabla g_k^T(x^0)^T d < 0. \tag{16}$$

De V ser Linearmente Independente e pelo Corolário 10, existe uma curva diferenciável $\varphi(t) \subset S$ para $t \in (-\delta, \delta)$, com $\varphi(0) = x^0$ e $\varphi'(0) = d$. Como x^0 é um ponto de mínimo local, segue do Lema 18 que d não é direção de descida, ou seja,

$$\nabla f(x^0)^T d \geq 0. \tag{17}$$

Mas, de (13), se multiplicarmos em ambos os lados da igualdade por d vamos ter

$$\nabla f(x^0)^T d + \lambda^T \nabla h(x^0) d + \mu_k \nabla g_k(x^0)^T d + \sum_{i=1, i\neq k}^{p} \mu_i \nabla g_i(x^0)^T d = 0. \tag{18}$$

De (16) $\nabla g_k(x^0)^T d < 0$, e pela hipótese de absurdo $\mu_k < 0$, logo, temos que $\nabla f(x^0)^T d < 0$, o que contradiz (17). Logo, não existe $k \in J(x^0)$ tal que $\mu_k < 0$. Portanto, $\mu_i \geq 0,\ \forall\, i \in \{1, 2, \ldots, p\}$. □

Observação 1. Podemos dizer que um ponto de extremo local x^0 vai satisfazer as seguintes condições para se enquadrar no caso de KKT quando

(i) $\nabla f(x^0) + \lambda^T \nabla h(x^0) + \mu^T \nabla g(x^0) = 0$, $\quad (ii)$ $\mu^T g(x^0) = 0$,

(iii) $g(x^0) \leq 0$, $\quad (iv)$ $h(x^0) = 0$, $\quad (v)$ $\mu \geq 0$.

A condição de complementariedade $\mu^T g(x^0) = 0$, também pode ser vista como $\mu_i g_i(x^0) = 0$, $\forall\, i \in \{1, 2, \ldots, p\}$.

Exemplo 2. Minimizar a função $f(x,y) = xy$ sujeita à restrição de desigualdade $g(x,y) = x^2 + y^2 \leq 4$.

Primeiramente escrevemos a função lagrangena $l(x,y,\mu) = xy + \mu(x^2 + y^2 - 4)$ e suas derivadas parciais. Em seguida, precisamos identificar as condições necessárias que cada ponto encontrado deve satisfazer.

$$(i)\ \frac{\partial l}{\partial x} = y + 2x\mu = 0;\ (ii)\ \frac{\partial l}{\partial y} = x + 2y\mu = 0;\ (iii)\ \frac{\partial l}{\partial \mu} = x^2 + y^2 - 4 = 0.$$

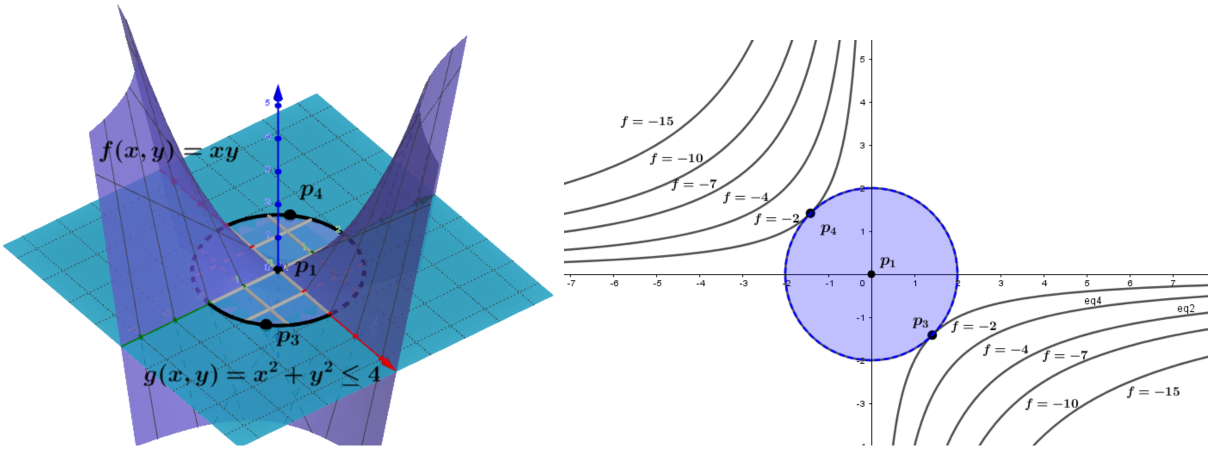
Da condição de complementariedade, $\mu g(x,y) = 0 \quad (iv)$, temos que a restrição vai estar inativa se $\mu = 0$ e ativa se $x^2 + y^2 - 4 = 0 \quad (v)$. Também devemos verificar se $\mu \geq 0$. Não devemos esquecer que no ponto estacionário x^0, deve valer $g(x^0) \leq 0 \quad (vi)$. Assim, temos seis condições que o ponto estacionário deve satisfazer antes de verificarmos se ele é um ponto extremo. Vamos começar pela restrição inativa. Como $\mu = 0$, obtemos de (ii) que $x = 0$ e de (i) que $y = 0$. Logo, o primeiro ponto crítico é $p_1^T = (0,0)$. Partindo agora para a restrição ativa, de (ii) obtemos $\mu = \dfrac{-x}{2y}$. Substituindo o valor de μ em (i), teremos a equação $x^2 - y^2 = 0 \quad (vii)$. Após resolver o sistema formado por (iii) e (vii), determinamos os seguintes pontos críticos com seus respectivos multiplicadores.

$$\begin{cases} p_1^T = (0,0) \text{ com } \mu = 0 \\ p_2^T = (\sqrt{2},\sqrt{2}) \text{ com } \mu = -\frac{1}{2} < 0 \\ p_3^T = (\sqrt{2},-\sqrt{2}) \text{ com } \mu = \frac{1}{2} > 0 \\ p_4^T = (-\sqrt{2},\sqrt{2}) \text{ com } \mu = \frac{1}{2} > 0 \\ p_5^T = (-\sqrt{2},-\sqrt{2}) \text{ com } \mu = -\frac{1}{2} < 0 \end{cases}$$

Podemos notar que apenas os pontos p_1, p_3, p_4 satisfazem todas as condições iniciais dadas na Observação 1. Como $g(x,y) = x^2 + y^2 = 4$ é um círculo e os pontos em seu interior também pertencem ao conjunto viável Ω, e esse conjunto também é limitado, segue que Ω é compacto. Pelo Teorema 4, f possui pelo menos um ponto de mínimo em Ω. Sobre o ponto p_1 que está na restrição inativa, podemos usar as condições de otimalidade para problemas irrestritos para verificar se ele pode ser um candidato a extremo local ou um ponto de sela. Como a matriz $\nabla^2 f(p_1)^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ tem autovalores -1 e 1, logo, p_1 é um ponto de sela. Como $f(p_3) = -2$ e $f(p_4) = -2$, concluímos que p_3 e p_4 são os pontos de mínimo global de f restrita aos pontos de Ω.

Na Figura 2 estão, respectivamente, os gráficos do exemplo em uma visão no $\mathbb{R}^3$ e no $\mathbb{R}^2$ com as curvas de nível tangenciando o conjunto viável onde f tem valor de mínimo -2.

Figura 2: Gráfico das superfícies e curvas de níveis do Exemplo 2



Fonte: Autor (2022)

Finalmente, mostramos um exemplo de quando valem as condições do Teorema de KKT, mas a Condição de Qualificação de Independência Linear (LICQ) não é satisfeita.

Exemplo 3. Considere o problema

$$\begin{array}{cc} \text{minimizar} & f(x,y) = x \\ \text{sujeito a} & g_1(x,y) = x^2 + 3x - y \leq 0 \\ & g_2(x,y) = x^2 + y^2 + 2x + 2y \leq 0 \\ & g_3(x,y) = -x \leq 0 \end{array}$$

O ponto $p^T = (0, 0) \in \Omega$ é o único ponto que pertence as três restrições, mais especificamente, quando elas são ativas e satisfaz as condições de KKT para ser o mínimo local de f. No entanto, os vetores gradiente $\nabla g_1(0, 0) = (3\ -1)^T$, $\nabla g_2(0, 0) = (2\ 2)^T$ e $\nabla g_3(0, 0) = (-1\ 0)^T$ são Linearmente Dependentes, ou seja, p não é um ponto regular para as restrições. Observe que a equação $\nabla f(0, 0) = -\lambda \nabla g_1(0, 0) - \mu \nabla g_2(0, 0) - \theta \nabla g_3(0, 0)$ implica no sistema

$$\begin{cases} -3\lambda - 2\mu + \theta = 1 \\ \lambda - 2\mu = 0 \end{cases}$$

em que $\lambda = 2\mu$ e $\theta = 1 + 8\mu$. Logo, existem infinitas combinações com multiplicadores de Lagrange positivos que satisfazem as equações anteriores.

Conclusões

Apesar de que a demonstração das condições de KKT sob a qualificação de independência linear dos gradientes das restrições esteja consagrada na literatura, neste trabalho provamos KKT diferenciando pontualmente das provas usuais, utilizando resultados que herdam propriedades de um vetor tangente dado pelo Teorema da Função Implícita. Como complemento, estudamos a relação que KKT tem com o Teorema dos Multiplicadores de Lagrange.

Referências

HAESER, G. *Condições de otimalidade de primeira e segunda ordem em otimização não linear*. Instituto de Matemática e Estatística, Universidade de São Paulo - USP. 2015.

JOHN, F. *Extremum problems with inequalities as subsidiary conditions*. In O. E. Neugebauer K. O. Friedrichs and J. J. Stoker, editors, Studies and Essays: Courant Anniversary,Wiley-Interscience, New York, pages 187–204, 1948.

KUHN, H. W.; TUCKER, A. W. *Nonlinear programming*. 1951. In J. Neyman, editor, Proceendings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, pages 481–492.

LIMA, E. L. *Análise Real*. Rio de Janeiro: IMPA, 2004. v. 2.

LUENBERGER, D. G. *Introduction to Linear and Nonlinear Programming*. 3. ed. Cambridge: Addison-Wesley, 2008.

MARCHAND, L. *Multiplicadores de Lagrange*: uma aplicação em problemas de otimização global restrita. 2016. Trabalho de Conclusão de Curso (Bacharelado em Matemática Aplicada) - Instituto de Matemática, Estatística e Física, Universidade Federal do Rio Grande - FURG, Rio Grande, 2016.

MARTÍNEZ, J. M. *Otimização prática usando o Lagrangiano aumentado*. Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas - UNICAMP. 2006.

RIBEIRO, A. A.; KARAS, E. W. *Otimização Contínua Aspectos Teóricos e Computacionais*. São Paulo: Cengage Learning, 2013.

STEINBRUCH, A. *Álgebra Linear*. São Paulo: McGraw-Hill, 1987.

WILLIAM, F. T. *The Method of Lagrange Multipliers*. São Antônio: Person Education, 2013. Disponível em: http://ramanujan.math.trinity.edu/wtrench/texts/TRENCH_LAGRANGE_METHOD.PDF. Acesso em: 09 set. 2022.