
Mathematica

Revista eletrônica de divulgação matemática

Universidade Estadual do Oeste do Paraná - UNIOESTE
Centro de Ciências Exatas e Tecnológicas

Mathematica

Revista eletrônica de divulgação matemática

Conselho editorial:

Sandro Marcos Guzzo - UNIOESTE (Editor chefe)
Edson Carlos Licurgo Santos - UNIOESTE (Editor assistente)
Esdras Teixeira Costa - UFG (Editor assistente)
Fabiana Magda Garcia Papani - UNIOESTE (Editor assistente)
Karina Schiabel - UFSCar (Editor assistente)

Arte da Capa:

Diagramação:

Sandro Marcos Guzzo

Apresentação

A Revista Mathematica surgiu no final de 2019, com uma conversa informal entre alguns professores do curso de Matemática da Unioeste. Desejávamos inicialmente abrir um espaço para divulgação dos resultados obtidos em projetos de iniciação científica, monografias de conclusão de curso, dissertações de mestrado e pesquisas individuais de professores. Muitos destes trabalhos produzem belos textos de matemática que acabam por serem conhecidos por um número reduzido de pessoas. Surgiu então a ideia de criar uma revista que pudesse divulgar estes resultados na forma de textos curtos, didáticos e com conteúdo matemático que pudesse interessar aos alunos de graduação. Neste contexto, uma revista poderia também atrair a atenção de professores, pesquisadores e acadêmicos de instituições de todo o país.

A nossa intenção era publicar o primeiro número no ano de 2020. Contudo o ano de 2020 iniciou e com ele a pandemia causada pelo Coronavírus. Assim como todas as demais atividades, as atividades universitárias ficaram prejudicadas neste momento. As aulas foram suspensas e os projetos de pesquisa e de iniciação científica foram reestruturados. Não conseguimos no ano de 2020 divulgar a revista de forma satisfatória. Acreditamos ainda que mesmo chegando ao conhecimento de alguns professores e acadêmicos a existência da revista, havia outras preocupações maiores do que enviar um texto para uma revista nova e desconhecida.

Neste ano de 2022, a Revista Mathematica recebeu quatro textos que foram avaliados e considerados aptos para publicação. Temos assim nosso primeiro volume. O conselho editorial agradece aos autores pelo envio dos trabalhos e também à comissão científica pelas contribuições feitas durante o processo de avaliação e correção dos trabalhos.

O conselho editorial.

Índice de trabalhos

Soluções trigonométricas de equações polinomiais de grau 3	9
Otimização quadrática restrita e o método de programação quadrática sequencial	21
Representação de rotações no espaço com quatérnios	37
KKT à luz da Independência Linear das Restrições	51

Soluções trigonométricas de equações polinomiais de grau 3

Sandro Marcos Guzzo - Universidade Estadual do Oeste do Paraná (Campus de Cascavel)
Sandra Maria Tieppo - Universidade Federal do Paraná (Setor Palotina)

(Recebido em 16/09/2022. Aceito em 25/11/2022. Publicado em 22/12/2022)

Resumo: No início do século XVI, alguns matemáticos, fizeram esforços para encontrar uma fórmula para a resolução de equações polinomiais de grau 3. A ideia era obter uma fórmula que permitisse encontrar (pelo menos) uma raiz da equação manipulando apenas os coeficientes da própria equação. O objetivo deste texto é mostrar como podemos determinar (pelo menos) uma raiz de uma equação polinomial de grau 3 usando expressões trigonométricas.

Palavras-chave: Soluções trigonométricas; Equações polinomiais.

1 Introdução

Desde a descoberta da fórmula que resolve a equação do 2º grau, os matemáticos começaram a pensar em formas de resolver as equações de grau 3

$$ax^3 + bx^2 + cx + d = 0, \quad (1)$$

com a, b, c e d coeficientes reais (ou mesmo complexos) e $a \neq 0$.

No início do século XVI, matemáticos italianos, fizeram esforços para encontrar uma fórmula para a resolução de certas equações do terceiro grau. Eles consideraram a equação de grau 3 sem o termo quadrático,

$$x^3 + px + q = 0, \quad (2)$$

e encontraram uma raiz da forma

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}},$$

que ficou conhecida como fórmula de Cardano. O leitor interessado nos detalhes de como esta fórmula é obtida pode consultar Lima (2021).

Ludovico Ferrari, professor da Universidade de Bolonha e discípulo de Cardano, foi quem desenvolveu um método que permite eliminar o termo quadrático da equação original (1) transformando-a em uma equação sem o termo quadrático na forma (2). A técnica de Ferrari consiste em aplicar a mudança de variáveis $x = y - \frac{b}{3a}$ na equação (1), obtendo

$$a \left(y - \frac{b}{3a} \right)^3 + b \left(y - \frac{b}{3a} \right)^2 + c \left(y - \frac{b}{3a} \right) + d = 0,$$

e portanto

$$ay^3 + \left(c - \frac{b^2}{3a} \right) y + \left(\frac{2b^3}{27a^2} - \frac{bc}{3a} + d \right) = 0.$$

Dividindo agora toda a equação por $a \neq 0$, e designando $p = \frac{c}{a} - \frac{b^2}{3a^2}$ e $q = \frac{2b^3}{27a^3} - \frac{bc}{3a^2} + \frac{d}{a}$, recaímos em uma equação de grau 3 sem o termo quadrático, na forma de (2) (na variável y). A equação (2) é conhecida como equação reduzida da equação (1). Assim, se r é raiz da equação (2), então $x = r - \frac{b}{3a}$ é raiz de (1), desde que a, b, c, d, p e q satisfaçam as relações dadas.

Quando estudamos trigonometria, tanto a circular quanto a hiperbólica, nos deparamos com algumas expressões trigonométricas que envolvem potências de grau 3 das funções trigonométricas sem a presença de termos quadráticos. São as igualdades

$$4 \cos^3 u - 3 \cos u - \cos(3u) = 0, \quad (3)$$

$$4 \operatorname{sen}^3 u - 3 \operatorname{sen} u + \operatorname{sen}(3u) = 0, \quad (4)$$

$$4 \cosh^3 u - 3 \cosh u - \cosh(3u) = 0, \quad (5)$$

$$4 \operatorname{senh}^3 u + 3 \operatorname{senh} u - \operatorname{senh}(3u) = 0, \quad (6)$$

válidas para qualquer $u \in \mathbb{R}$. Estas identidades são facilmente obtidas a partir das chamadas fórmulas do arco triplo

$$\cos(3u) = 4 \cos^3 u - 3 \cos u,$$

$$\operatorname{sen}(3u) = -4 \operatorname{sen}^3 u + 3 \operatorname{sen} u,$$

$$\cosh(3u) = 4 \cosh^3 u - 3 \cosh u,$$

$$\operatorname{senh}(3u) = 4 \operatorname{senh}^3 u + 3 \operatorname{senh} u.$$

Para o leitor que deseja mais informações sobre trigonometria, em especial a trigonometria hiperbólica, recomendamos Guzzo (2021).

O objetivo deste trabalho é partir da equação (2), e obter (pelo menos) uma raiz desta equação, usando as expressões trigonométricas apresentadas. Com pelo menos uma raiz encontrada, podemos reduzir o grau do polinômio e usar a fórmula de Bháskara para determinar as outras duas raízes. Em resumo, o que pretendemos é obter (pelo menos) uma raiz da equação

$$x^3 + px + q = 0.$$

com $p, q \in \mathbb{R} - \{0\}$. Estamos considerando $p \neq 0$ e $q \neq 0$ pois caso $p = 0$ ou $q = 0$, a obtenção de uma raiz da equação não exigirá técnicas elaboradas. Se $p = 0$, a equação se reduz a $x^3 + q = 0$ e $x = \sqrt[3]{-q}$ é uma raiz. Se $q = 0$ então a equação se reduz a $x^3 + px = 0$ e $x = 0$ é uma raiz.

2 Discriminante de uma equação de grau 3

É conhecido que para uma equação de grau 2 na forma $ax^2 + bx + c = 0$ o discriminante $\Delta = b^2 - 4ac$ pode ser utilizado para decidir o comportamento das duas raízes antes mesmo de determinar estas raízes. A análise da nulidade ou do sinal de Δ nos indica a natureza das raízes. O objetivo desta seção é mostrar como esta ideia pode ser levada também para equações polinomiais de grau 3 que são o alvo do nosso estudo.

Consideremos uma equação polinomial geral de grau 3 na forma (1), isto é,

$$ax^3 + bx^2 + cx + d = 0,$$

em que $a, b, c, d \in \mathbb{R}$ e $a \neq 0$. De acordo com o Teorema Fundamental da Álgebra, sabemos que esta equação possui três raízes, reais ou complexas, distintas ou não. Designemos estas raízes por r_1, r_2 e r_3 . Podemos fazer uma análise do comportamento destas raízes analisando a expressão

$$\Delta = (r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2, \quad (7)$$

que é conhecida como discriminante da equação cúbica (1).

Observe que, de forma imediata, duas destas raízes coincidem se e somente se $\Delta = 0$.

Vamos verificar que uma destas raízes é um número complexo (não real), se e somente se, $\Delta < 0$. Suponha então que uma destas raízes é complexa não real. Como é de conhecimento da teoria geral das equações polinomiais com coeficientes reais, obrigatoriamente o complexo conjugado desta raiz é uma outra raiz da equação. Por este motivo, apenas duas destas raízes podem ser complexas não reais. Vamos supor então, sem perda de generalidade, que $r_1 \in \mathbb{R}$ e $r_3 = \bar{r}_2 \notin \mathbb{R}$. Então

$$\begin{aligned} \Delta &= (r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2 \\ &= (r_1 - r_2)^2(r_1 - \bar{r}_2)^2(r_2 - \bar{r}_2)^2 \\ &= ((r_1 - r_2)(r_1 - \bar{r}_2))^2 (r_2 - \bar{r}_2)^2 \\ &= (r_1^2 - r_1\bar{r}_2 - r_1r_2 + r_2\bar{r}_2)^2 (2Im(r_2)i)^2 \\ &= (r_1^2 - r_1(r_2 + \bar{r}_2) + |r_2|^2)^2 (2Im(r_2)i)^2 \\ &= (r_1^2 - r_1 2Re(r_2) + |r_2|^2)^2 (2Im(r_2)i)^2 \\ &= (r_1^2 - r_1 2Re(r_2) + (Re(r_2))^2 + (Im(r_2))^2)^2 (2Im(r_2)i)^2 \\ &= ((r_1 - Re(r_2))^2 + (Im(r_2))^2)^2 (2Im(r_2)i)^2, \end{aligned}$$

sendo que $Re(r_2)$ e $Im(r_2)$ referem-se respectivamente às partes real e imaginária do número complexo r_2 .

Basta ver agora que $Im(r_2) \neq 0$ pois r_2 é um número complexo não real e desta forma $(r_1 - Re(r_2))^2 + Im(r_2)^2 > 0$ e $(2Im(r_2)i)^2 < 0$ já que $i^2 = -1$. Segue que se alguma das raízes da equação (1) for complexa, então $\Delta < 0$. Claramente o recíproco é verdadeiro, isto é, se $\Delta < 0$ então obrigatoriamente algum dos fatores do produto é negativo. Mas como estes fatores são quadrados, obrigatoriamente um deles é um número complexo (mais ainda, um imaginário puro).

Por exclusão com os casos $\Delta = 0$ e $\Delta < 0$ temos então que as raízes da equação (1) são reais e distintas se e somente se $\Delta > 0$.

Esta análise do discriminante da equação é importante, mas o leitor deve ter percebido que esta forma de apresentação do discriminante torna-o sem muita utilidade. Isto porque para obter este discriminante precisamos conhecer explicitamente as três raízes da equação e

desta forma a análise do discriminante já seria desnecessária. Podemos reescrever a expressão do discriminante em termos dos coeficientes a , b , c e d da equação. Isto permitiria conhecer o comportamento das raízes antes de determiná-las. Para esta tarefa usamos as Relações de Girard, que estabelecem uma relação entre os coeficientes a , b , c e d e as raízes r_1 , r_2 , e r_3 da equação. Para a equação de grau 3 tais relações são dadas por

$$\begin{aligned} r_1 + r_2 + r_3 &= -\frac{b}{a} \\ r_1r_2 + r_1r_3 + r_2r_3 &= \frac{c}{a} \\ r_1r_2r_3 &= -\frac{d}{a}. \end{aligned}$$

Desta forma, temos que

$$\begin{aligned} \Delta &= ((r_1 - r_2)(r_1 - r_3)(r_2 - r_3))^2 \\ &= (r_1^2r_2 - r_1r_2^2 + r_2^2r_3 - r_1^2r_3 + r_1r_3^2 - r_2r_3^2)^2 \\ &= (r_1^4r_2^2 + r_1^2r_2^4 + r_2^4r_3^2 + r_1^4r_3^2 + r_1^2r_3^4 + r_2^2r_3^4) - 6(r_1r_2r_3)^2 \\ &\quad + 2(r_1^2r_2^3r_3 + r_1^3r_2r_3^2 + r_1^3r_2^2r_3 + r_1r_2^3r_3^2 + r_1r_2^2r_3^3 + r_1^2r_2r_3^3) \\ &\quad - 2(r_1^4r_2r_3 + r_1r_2^4r_3 + r_1r_2r_3^4) - 2(r_1^3r_2^3 + r_1^3r_3^3 + r_2^3r_3^3). \end{aligned} \quad (8)$$

Mas notemos que

$$\begin{aligned} (r_1 + r_2 + r_3)(r_1r_2 + r_1r_3 + r_2r_3) \\ = (r_1^2r_2 + r_1r_2^2 + r_1^2r_3 + r_1r_3^2 + r_2^2r_3 + r_2r_3^2) + 3(r_1r_2r_3) \end{aligned}$$

donde podemos escrever

$$(r_1^2r_2 + r_1r_2^2 + r_1^2r_3 + r_1r_3^2 + r_2^2r_3 + r_2r_3^2) = 3\frac{d}{a} - \frac{bc}{a} = \frac{3ad - bc}{a^2}. \quad (9)$$

Elevando ao quadrado os dois membros da identidade (9) podemos também obter

$$\begin{aligned} \left(\frac{3ad - bc}{a^2}\right)^2 &= (r_1^2r_2 + r_1r_2^2 + r_1^2r_3 + r_1r_3^2 + r_2^2r_3 + r_2r_3^2)^2 \\ &= (r_1^4r_2^2 + r_1^2r_2^4 + r_2^4r_3^2 + r_1^4r_3^2 + r_1^2r_3^4 + r_2^2r_3^4) + 6(r_1r_2r_3)^2 \\ &\quad + 2(r_1^3r_2r_3^2 + r_1^3r_2^2r_3 + r_1^2r_3^3r_2 + r_1r_2^3r_3^2 + r_1^2r_2r_3^3 + r_1r_2^2r_3^3) \\ &\quad + 2(r_1^4r_2r_3 + r_1r_2^4r_3 + r_1r_2r_3^4) + 2(r_1^3r_2^3 + r_1^3r_3^3 + r_2^3r_3^3). \end{aligned}$$

Desta última igualdade obtemos portanto

$$\begin{aligned} (r_1^4r_2^2 + r_1^2r_2^4 + r_2^4r_3^2 + r_1^4r_3^2 + r_1^2r_3^4 + r_2^2r_3^4) \\ + 2(r_1^3r_2r_3^2 + r_1^3r_2^2r_3 + r_1^2r_3^3r_2 + r_1r_2^3r_3^2 + r_1^2r_2r_3^3 + r_1r_2^2r_3^3) \\ = \left(\frac{3ad - bc}{a^2}\right)^2 - 6(r_1r_2r_3)^2 - 2(r_1^4r_2r_3 + r_1r_2^4r_3 + r_1r_2r_3^4) - 2(r_1^3r_2^3 + r_1^3r_3^3 + r_2^3r_3^3). \end{aligned} \quad (10)$$

Substituindo (10) em (8) obtemos

$$\Delta = \left(\frac{3ad - bc}{a^2}\right)^2 - 12(r_1 r_2 r_3)^2 - 4(r_1^4 r_2 r_3 + r_1 r_2^4 r_3 + r_1 r_2 r_3^4) - 4(r_1^3 r_2^3 + r_1^3 r_3^3 + r_2^3 r_3^3). \quad (11)$$

Para a penúltima parcela da igualdade (11) usamos

$$(r_1 + r_2 + r_3)^3 = (r_1^3 + r_2^3 + r_3^3) + 3(r_1^2 r_2 + r_1 r_2^2 + r_1^2 r_3 + r_1 r_3^2 + r_2^2 r_3 + r_2 r_3^2) + 6r_1 r_2 r_3,$$

para escrever

$$\begin{aligned} & (r_1^4 r_2 r_3 + r_1 r_2^4 r_3 + r_1 r_2 r_3^4) \\ &= r_1 r_2 r_3 (r_1^3 + r_2^3 + r_3^3) \\ &= r_1 r_2 r_3 ((r_1 + r_2 + r_3)^3 - 3(r_1^2 r_2 + r_1 r_2^2 + r_1^2 r_3 + r_1 r_3^2 + r_2^2 r_3 + r_2 r_3^2) - 6r_1 r_2 r_3) \end{aligned}$$

e usando (9) conseguimos

$$(r_1^4 r_2 r_3 + r_1 r_2^4 r_3 + r_1 r_2 r_3^4) = -\frac{d}{a} \left(-\frac{b^3}{a^3} - 3\frac{3ad - bc}{a^2} + 6\frac{d}{a} \right) = \frac{b^3 d + 3a^2 d^2 - 3abcd}{a^4}. \quad (12)$$

Já para a última parcela da igualdade (11) podemos escrever

$$\begin{aligned} (r_1 r_2 + r_1 r_3 + r_2 r_3)^3 &= r_1^3 r_2^3 + r_1^3 r_3^3 + r_2^3 r_3^3 + 6r_1^2 r_2^2 r_3^2 \\ &\quad + 3(r_1^2 r_2^3 r_3 + r_1^3 r_2^2 r_3 + r_1 r_2^3 r_3^2 + r_1 r_2^2 r_3^3 + r_1^3 r_2 r_3^2 + r_1^2 r_2 r_3^3) \\ &= (r_1^3 r_2^3 + r_1^3 r_3^3 + r_2^3 r_3^3) + 6(r_1 r_2 r_3)^2 \\ &\quad + 3(r_1 r_2 r_3)(r_1 r_2^2 + r_1^2 r_2 + r_2^2 r_3 + r_2 r_3^2 + r_1^2 r_3 + r_1 r_3^2), \end{aligned}$$

e usando (9) obtemos

$$\begin{aligned} (r_1^3 r_2^3 + r_1^3 r_3^3 + r_2^3 r_3^3) &= (r_1 r_2 + r_1 r_3 + r_2 r_3)^3 - 6(r_1 r_2 r_3)^2 \\ &\quad - 3(r_1^2 r_2^3 r_3 + r_1^3 r_2^2 r_3 + r_1 r_2^3 r_3^2 + r_1 r_2^2 r_3^3 + r_1^3 r_2 r_3^2 + r_1^2 r_2 r_3^3) \\ &= \frac{c^3}{a^3} - 6\frac{d^2}{a^2} + \frac{3d}{a} \frac{3ad - bc}{a^2} = \frac{c^3 + 3ad^2 - 3bcd}{a^3}. \end{aligned} \quad (13)$$

Finalmente juntando as estimativas (12) e (13) em (11) obtemos

$$\begin{aligned} \Delta &= \left(\frac{3ad - bc}{a^2}\right)^2 - 12\frac{d^2}{a^2} - 4\frac{b^3 d + 3a^2 d^2 - 3abcd}{a^4} - 4\frac{c^3 + 3ad^2 - 3bcd}{a^3} \\ &= \frac{9a^2 d^2 - 6adbc + b^2 c^2}{a^4} - \frac{12d^2}{a^2} + \frac{12abcd - 4b^3 d - 12a^2 d^2}{a^4} + \frac{12bcd - 4c^3 - 12ad^2}{a^3} \\ &= \frac{9a^2 d^2 - 6adbc + b^2 c^2 - 12d^2 a^2 + 12abcd - 4b^3 d - 12a^2 d^2 + 12abcd - 4ac^3 - 12a^2 d^2}{a^4} \\ &= \frac{b^2 c^2 - 4ac^3 - 4b^3 d - 27a^2 d^2 + 18abcd}{a^4}. \end{aligned}$$

Já que a análise de sinal, bem como a nulidade de Δ , fica a cargo do numerador da fração, é comum apresentar o discriminante já na forma que interessa

$$\Delta = b^2 c^2 - 4ac^3 - 4b^3 d - 27a^2 d^2 + 18abcd.$$

Lembremos que o objetivo deste trabalho é obter (pelo menos) uma raiz da equação do terceiro grau reduzida na forma $x^3 + px + q = 0$, com $p \neq 0$ e $q \neq 0$. Desta forma o discriminante desta equação é $\Delta = -4p^3 - 27q^2$, e conforme os resultados desta seção, temos os seguintes casos:

- i) $\Delta = -4p^3 - 27q^2 > 0$, se e somente se, as três raízes da equação são reais e distintas;
- ii) $\Delta = -4p^3 - 27q^2 = 0$, se e somente se, a equação possui raízes repetidas. Neste caso, obrigatoriamente as raízes são reais, sendo pelo menos uma delas com multiplicidade;
- iii) $\Delta = -4p^3 - 27q^2 < 0$, se e somente se, a equação possui uma raiz real e duas raízes complexas conjugadas.

3 Raízes trigonométricas

Estamos agora prontos para o objetivo principal deste trabalho, a obtenção de soluções da equação polinomial de grau 3, usando expressões trigonométricas. Dividiremos o estudo nos casos $\Delta > 0$, $\Delta = 0$ e $\Delta < 0$.

Primeiro caso: (Três raízes reais e distintas) Vamos agora encontrar uma solução (raiz) da equação $x^3 + px + q = 0$, com $p \neq 0$ e $q \neq 0$, para o caso $\Delta = -4p^3 - 27q^2 > 0$. Neste caso, como visto anteriormente, sabemos que a equação cúbica possuirá 3 raízes reais e distintas. Mas notemos ainda que $\Delta = -4p^3 - 27q^2 > 0$ exigirá $-4p^3 > 27q^2$ e portanto $p < 0$.

Supondo então $\Delta = -4p^3 - 27q^2 > 0$ (consequentemente $p < 0$) e $q \neq 0$, começamos fazendo a substituição $x = 2\sqrt{\frac{-p}{3}} \cos u$ na equação (2), e com isso obtemos

$$\frac{-8p}{3} \sqrt{\frac{-p}{3}} \cos^3 u + 2p \sqrt{\frac{-p}{3}} \cos u + q = 0,$$

e reorganizando os termos

$$\frac{-2p}{3} \sqrt{\frac{-p}{3}} \left(4 \cos^3 u - 3 \cos u - \frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) = 0.$$

Como $p \neq 0$ então resta que

$$4 \cos^3 u - 3 \cos u - \frac{3q}{2p} \sqrt{\frac{3}{-p}} = 0$$

e comparando esta equação com (3), vemos que esta equação torna-se verdadeira para todos os valores de u que satisfazem $\cos(3u) = \frac{3q}{2p} \sqrt{\frac{3}{-p}}$.

Observe que como $-4p^3 - 27q^2 > 0$, então $27q^2 < -4p^3$, donde $9q^2 < 4p^2 \frac{-p}{3}$, e também $3|q| < 2|p| \sqrt{\frac{-p}{3}}$. Isto garante que $\left| \frac{3q}{2p} \sqrt{\frac{3}{-p}} \right| < 1$ e nos possibilita determinar o valor procurado para u .

Nestes termos, queremos encontrar u de forma que $\cos(3u) = \frac{3q}{2p}\sqrt{\frac{3}{-p}}$, e então, temos que

$$3u = \cos^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + 2k\pi, \quad \text{para } k \in \mathbb{Z},$$

ou ainda

$$u = \frac{1}{3} \cos^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3}, \quad \text{para } k \in \mathbb{Z}.$$

Voltando com este u em $x = 2\sqrt{\frac{-p}{3}} \cos u$, obtemos

$$x = 2\sqrt{\frac{-p}{3}} \cos \left(\frac{1}{3} \cos^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right), \quad \text{para } k \in \mathbb{Z}.$$

Notemos ainda que não há a necessidade de que k assuma todos os valores inteiros. Basta tomar $k \in \{0, 1, 2\}$ pois todos os demais valores de k inteiros ocasionarão a repetição do valor obtido com algum destes 3 valores de k . Segue que

$$x = 2\sqrt{\frac{-p}{3}} \cos \left(\frac{1}{3} \cos^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right), \quad \text{para } k \in \{0, 1, 2\}, \quad (14)$$

são as 3 raízes (reais e distintas) procuradas da equação $x^3 + px + q = 0$, quando $\Delta = -4p^3 - 27q^2 > 0$.

O leitor atento dirá que para este mesmo caso, também poderíamos considerar no início do processo a substituição $x = 2\sqrt{\frac{-p}{3}} \sin u$. De fato, esta substituição traria

$$\frac{-2p}{3} \sqrt{\frac{-p}{3}} \left(4 \sin^3 u - 3 \sin u - \frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) = 0.$$

e conseqüentemente

$$4 \sin^3 u - 3 \sin u - \frac{3q}{2p} \sqrt{\frac{3}{-p}} = 0.$$

Comparando agora esta equação com (4), ela torna-se verdadeira para todo $u \in \mathbb{R}$ que satisfaz $\sin(3u) = -\frac{3q}{2p}\sqrt{\frac{3}{-p}}$. Novamente precisamos da condição $\left| \frac{3q}{2p}\sqrt{\frac{3}{-p}} \right| < 1$ e como visto anteriormente, isto fica garantido da condição $-4p^3 - 27q^2 > 0$. Assim queremos determinar u de forma que $\sin(3u) = -\frac{3q}{2p}\sqrt{\frac{3}{-p}}$, o que nos trará de forma equivalente ao anterior

$$u = \frac{1}{3} \sin^{-1} \left(-\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3}, \quad \text{para } k \in \mathbb{Z}.$$

e

$$x = 2\sqrt{\frac{-p}{3}} \sin \left(\frac{1}{3} \sin^{-1} \left(-\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right), \quad \text{para } k \in \{0, 1, 2\}. \quad (15)$$

Não é difícil provar que os valores de x obtidos pelas expressões (14) e (15) coincidem, em virtude das identidades trigonométricas $\cos(\frac{\pi}{2} - u) = \sin u$, $\sin^{-1}(-u) = -\sin^{-1} u$ e $\cos^{-1} u = \frac{\pi}{2} - \sin^{-1} u$, válidas para todo $u \in \mathbb{R}$. Assim,

$$\cos \left(\frac{1}{3} \cos^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right) = \cos \left(\frac{1}{3} \frac{\pi}{2} - \frac{1}{3} \sin^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right)$$

$$\begin{aligned}
 &= \cos \left(-\frac{1}{3} \frac{\pi}{2} + \frac{1}{3} \operatorname{sen}^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) - \frac{2k\pi}{3} \right) \\
 &= \cos \left(\frac{\pi}{2} - \frac{4}{3} \frac{\pi}{2} + \frac{1}{3} \operatorname{sen}^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) - \frac{2k\pi}{3} \right) \\
 &= \cos \left(\frac{\pi}{2} - \frac{2\pi}{3} + \frac{1}{3} \operatorname{sen}^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) - \frac{2k\pi}{3} \right) \\
 &= \cos \left(\frac{\pi}{2} + \frac{1}{3} \operatorname{sen}^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) - \frac{2(k+1)\pi}{3} \right) \\
 &= \cos \left(\frac{\pi}{2} - \frac{1}{3} \operatorname{sen}^{-1} \left(-\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) - \frac{2(k+1)\pi}{3} \right) \\
 &= \operatorname{sen} \left(\frac{1}{3} \operatorname{sen}^{-1} \left(-\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2(k+1)\pi}{3} \right).
 \end{aligned}$$

Mas lembremos que $k \in \{0, 1, 2\}$ e portanto $(k+1) \in \{1, 2, 3\}$ e como o caso $(k+1) = 3$ coincide com o caso $k+1 = 0$ podemos claramente reorganizar $(k+1) \in \{0, 1, 2\}$ e não faz diferença escrever $k+1$ ou k .

Portanto as raízes de $x^3 + px + q = 0$, quando $-4p^3 - 27q^2 > 0$, são

$$\begin{aligned}
 x &= 2\sqrt{\frac{-p}{3}} \cos \left(\frac{1}{3} \cos^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right), \quad \text{para } k \in \{0, 1, 2\} \\
 &= 2\sqrt{\frac{-p}{3}} \operatorname{sen} \left(\frac{1}{3} \operatorname{sen}^{-1} \left(-\frac{3q}{2p} \sqrt{\frac{3}{-p}} \right) + \frac{2k\pi}{3} \right), \quad \text{para } k \in \{0, 1, 2\}.
 \end{aligned}$$

Segundo caso: (Três raízes reais com multiplicidade) Não trataremos extensivamente este caso, pois o caso anterior pode ser adaptado para contemplar também o caso em que a equação (2) admite raízes reais com multiplicidade.

Olharemos rapidamente para o caso anterior considerando que $\Delta = -4p^3 - 27q^2 = 0$, e portanto $-4p^3 = 27q^2$.

Se $p = q = 0$, então não há o que analisar pois neste caso a equação reduzida $x^3 + px + q = 0$ fica $x^3 = 0$ e a raiz $r = 0$ é a única raiz com multiplicidade 3.

Vamos então supor p e q não nulos satisfazendo $-4p^3 - 27q^2 = 0$ e portanto $-4p^3 = 27q^2$ o que obriga também $p < 0$. Neste caso temos que $\frac{27q^2}{-4p^3} = 1$, e portanto $\left| \frac{3q}{2p} \sqrt{\frac{3}{-p}} \right| = 1$.

A substituição $x = 2\sqrt{\frac{-p}{3}} \cos u$ ainda pode ser feita na equação (2), e nos conduz a

$$4 \cos^3 u - 3 \cos u - \frac{3q}{2p} \sqrt{\frac{3}{-p}} = 0$$

e da comparação com a identidade (3), obtemos $\cos(3u) = \frac{3q}{2p} \sqrt{\frac{3}{-p}} = \pm 1$.

Levando em conta que $p < 0$, se $q < 0$, então teremos $\cos(3u) = 1$ portanto $3u = 0 + 2k\pi$ com $k \in \mathbb{Z}$. Por outro lado, se $q > 0$ então teremos $\cos(3u) = -1$ e $3u = \pi + 2k\pi$ com $k \in \mathbb{Z}$. Lembremos que basta considerar em ambos os casos $k \in \{0, 1, 2\}$.

Sendo assim, as raízes da equação reduzida serão dadas por

$$x = 2\sqrt{\frac{-p}{3}} \cos\left(\theta_0 + \frac{2k\pi}{3}\right), \quad \text{para } k \in \{0, 1, 2\}$$

sendo que $\theta_0 = 0$ quando $q < 0$ e $\theta_0 = \frac{\pi}{3}$ quando $q > 0$. Podemos observar que quando $q < 0$ (ou quando $\theta_0 = 0$) duas raízes coincidirão para $k = 1$ e $k = 2$ já que $\cos(\frac{2\pi}{3}) = \cos(\frac{4\pi}{3})$. Também quando $q > 0$ (ou quando $\theta_0 = \frac{\pi}{3}$) duas raízes coincidirão para $k = 0$ e $k = 2$, já que, $\cos(\frac{\pi}{3}) = \cos(\frac{5\pi}{3})$.

De forma análoga, a substituição $x = 2\sqrt{\frac{-p}{3}} \operatorname{sen} u$ traz

$$4 \operatorname{sen}^3 u - 3 \operatorname{sen} u - \frac{3q}{2p} \sqrt{\frac{3}{-p}} = 0,$$

e pela comparação com a identidade (4), obtemos $\operatorname{sen}(3u) = -\frac{3q}{2p} \sqrt{\frac{3}{-p}} = \pm 1$.

O sinal de q novamente nos dirá o caminho a tomar. Se $q < 0$, então teremos $\operatorname{sen}(3u) = -1$ e assim $3u = \frac{3\pi}{2} + 2k\pi$ para $k \in \mathbb{Z}$. Se $q > 0$, então teremos $\operatorname{sen}(3u) = 1$ e assim $3u = \frac{\pi}{2} + 2k\pi$ para $k \in \mathbb{Z}$.

Sabemos que basta considerar $k \in \{0, 1, 2\}$ e desta forma, as raízes da equação reduzida podem também ser dadas por

$$x = 2\sqrt{\frac{-p}{3}} \operatorname{sen}\left(\theta_0 + \frac{2k\pi}{3}\right), \quad \text{para } k \in \{0, 1, 2\},$$

sendo que $\theta_0 = \frac{\pi}{2}$ se $q < 0$ e $\theta_0 = \frac{\pi}{6}$ se $q > 0$. Naturalmente quando $q < 0$, duas raízes coincidirão quando $k = 1$ e $k = 2$ já que $\operatorname{sen}(\frac{7\pi}{6}) = \operatorname{sen}(\frac{11\pi}{6})$. Também quando $q > 0$ duas raízes coincidirão quando $k = 0$ e $k = 1$ já que $\operatorname{sen}(\frac{\pi}{6}) = \operatorname{sen}(\frac{5\pi}{6})$.

Terceiro caso: (Uma única raiz real) Para este caso queremos obter (pelo menos) uma solução da equação (2), com $p \neq 0$ e $q \neq 0$, considerando $-4p^3 - 27q^2 < 0$.

Neste caso teremos então que $-4p^3 < 27q^2$. Note que se $p > 0$ então esta condição sempre será cumprida, Mas podemos ainda ter valores de $p < 0$ satisfazendo $-4p^3 < 27q^2$. Por isso, vamos separar os casos em que $p > 0$ e $p < 0$. Como q está elevado ao quadrado, não conseguimos estimar se $q > 0$ ou se $q < 0$, e por isso, vamos utilizar $|q|$ nos cálculos.

Vamos considerar primeiro o caso um pouco mais restritivo, isto é, o caso em que $\Delta = -4p^3 - 27q^2 < 0$ com $p < 0$ e $q \neq 0$. Fazemos a substituição $x = -2\frac{q}{|q|} \sqrt{\frac{-p}{3}} \cosh u$, e obtemos

$$\frac{q^2}{|q|^2} \frac{8p}{3} \frac{q}{|q|} \sqrt{\frac{-p}{3}} \cosh^3 u - 2p \frac{q}{|q|} \sqrt{\frac{-p}{3}} \cosh u + q = 0.$$

Agora como $\frac{q^2}{|q|^2} = 1$ qualquer que seja o valor de $q \neq 0$, então reorganizando os termos obtemos

$$\frac{2p}{3} \frac{q}{|q|} \sqrt{\frac{-p}{3}} \left(4 \cosh^3 u - 3 \cosh u + \frac{3|q|}{2p} \sqrt{\frac{3}{-p}} \right) = 0.$$

Novamente, como $p \neq 0$ então

$$4 \cosh^3 u - 3 \cosh u + \frac{3|q|}{2p} \sqrt{\frac{3}{-p}} = 0$$

e comparando esta equação agora com a identidade (5), vemos que ela torna-se verdadeira se $\cosh(3u) = -\frac{3|q|}{2p} \sqrt{\frac{3}{-p}}$.

Notemos agora que como estamos considerando agora o caso $-4p^3 - 27q^2 < 0$, isto é $-4p^3 < 27q^2$, e como com $p < 0$, então $\frac{27q^2}{-4p^3} > 1$. Segue que $-\frac{3|q|}{2p} \sqrt{\frac{3}{-p}} > 1$ e portanto é possível determinar o valor procurado de u .

Nestes termos temos que

$$3u = \cosh^{-1} \left(-\frac{3|q|}{2p} \sqrt{\frac{3}{-p}} \right),$$

ou ainda

$$u = \frac{1}{3} \cosh^{-1} \left(-\frac{3|q|}{2p} \sqrt{\frac{3}{-p}} \right).$$

Voltando com este u em $x = 2\sqrt{\frac{-p}{3}} \cosh u$, obtemos

$$x = 2\sqrt{\frac{-p}{3}} \cosh \left(\frac{1}{3} \cosh^{-1} \left(-\frac{3|q|}{2p} \sqrt{\frac{3}{-p}} \right) \right),$$

sendo esta a única raiz real da equação (2) quando $-4p^3 - 27q^2 < 0$ com $p < 0$.

Resta agora considerar a equação $x^3 + px + q = 0$ com $\Delta = -4p^3 - 27q^2 < 0$ para o caso $p > 0$. Na verdade, basta considerar que $p > 0$ e a desigualdade $-4p^3 < 27q^2$ ocorrerá obrigatoriamente, garantindo que $\Delta = -4p^3 - 27q^2 < 0$. Isto ainda significa que a equação cúbica possui apenas uma raiz real.

Considerando então a equação cúbica $x^3 + px + q = 0$ com $p > 0$ e $q \neq 0$, usamos a substituição $x = -2\sqrt{\frac{p}{3}} \sinh u$, e obtemos a equação

$$-\frac{8p}{3} \sqrt{\frac{p}{3}} \sinh^3 u - 2p \sqrt{\frac{p}{3}} \sinh u + q = 0,$$

que após reorganização dos termos nos conduz a

$$\frac{2p}{3} \sqrt{\frac{p}{3}} \left(4 \sinh^3 u + 3 \sinh u - \frac{3q}{2p} \sqrt{\frac{3}{p}} \right) = 0.$$

Novamente, como $p \neq 0$ então

$$4 \sinh^3 u + 3 \sinh u - \frac{3q}{2p} \sqrt{\frac{3}{p}} = 0$$

e comparando com a identidade (6), queremos encontrar u de forma que $\sinh(3u) = \frac{3q}{2p} \sqrt{\frac{3}{p}}$.

Levando em conta a bijetividade da função seno hiperbólico de \mathbb{R} em \mathbb{R} , temos que

$$3u = \sinh^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{p}} \right),$$

donde

$$u = \frac{1}{3} \operatorname{senh}^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{p}} \right).$$

Voltando para a variável original x , levamos este u em $x = 2\sqrt{\frac{p}{3}} \operatorname{senh} u$, e obtemos

$$x = -2\sqrt{\frac{p}{3}} \operatorname{senh} \left(\frac{1}{3} \operatorname{senh}^{-1} \left(\frac{3q}{2p} \sqrt{\frac{3}{p}} \right) \right),$$

sendo esta a única raiz real da equação (2) quando $-4p^3 - 27q^2 < 0$ com $p > 0$.

4 Conclusões

Obtivemos neste trabalho soluções de equações polinomiais de grau 3, usando expressões trigonométricas. O objetivo dos autores não é o de tomar para si o crédito destas expressões que já estão divulgadas em textos matemáticos. O objetivo dos autores é o de dar uma demonstração para estas expressões.

É também conhecido que existe um desinteresse cada vez maior pela trigonometria e pelas funções trigonométricas por parte dos alunos de ensino médio e superior. Acreditamos que quanto mais pudermos falar a respeito das funções trigonométricas e mostrar suas relações com outros ramos da matemática, podemos despertar a curiosidade ou o interesse do aluno por esta classe de funções.

Referências

Biazzi, Ricardo Neves. *Polinômios irredutíveis: critérios e aplicações*. 74 f. Dissertação - (mestrado) - Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas, 2014. Disponível em: <<http://hdl.handle.net/11449/108811>>.

Lima, Elisangela D. de & Guzzo, Sandro M. *Solubilidade por meio de radicais de equações polinomiais de grau menor ou igual a quatro*. Mathematica - Revista Eletrônica de divulgação matemática. **V. 1 (2021)** pp 39-49.

Guzzo, Sandro M. *As funções trigonométricas circulares e hiperbólicas*. Editora Moan: Cascavel, 2021.

Otimização quadrática restrita e o método de programação quadrática sequencial

Alex Augusto Nunes Machado
Simone Aparecida Miloca - Unioeste

(Recebido em 21/10/2022. Aceito em 29/11/2022. Publicado em 22/12/2022)

Resumo: Este trabalho teve por intuito apresentar o algoritmo PQS para então mostrar, matematicamente, sua eficiência em resolver problemas de programação quadrática com restrições de igualdade em apenas uma iteração. Para tanto, foram enunciados alguns conceitos básicos da otimização restrita, bem como os conceitos mínimos associados ao método sequencial. Por fim, o algoritmo PQS foi implementado em linguagem Python, e três exemplos numéricos que mostram a capacidade de resolver problemas envolvendo restrições de igualdade com baixo custo temporal foram apresentados.

Palavras-chave: Otimização Quadrática Restrita, PQS, Python.

1 Introdução

De modo geral, o campo da Otimização consiste em

$$\begin{array}{ll} \text{Otimizar} & f(x) \\ \text{sujeito a} & x \in \Omega \subset \mathbb{R}^n \end{array} \quad (1)$$

sendo que f é denominada *função objetivo* e Ω representa, no caso da otimização restrita, um conjunto de restrições (outras funções) que podem ser de igualdade ou desigualdade, ou ainda, simplesmente representam uma restrição na qual o tipo de variável envolvida é explicitada, isto é, se as variáveis são números reais, não-negativos, inteiros, etc.

Em diversos campos de pesquisa nas quais os tomadores de decisões buscam obter os melhores resultados possíveis, dentro de certas condições podem surgir situações nas quais os problemas são modelados como um problema de otimização. Na construção desses modelos podem surgir funções com leis de formação das mais variadas e, com isso, conforme destacam Ribeiro e Karas (2013) e Goldberg e Luna (2005), é possível agrupar tais problemas de otimização em algumas categorias, cada qual possuindo técnicas específicas para sua resolução.

Dentre as diversas categorias possíveis, há os Problemas de Otimização Quadrática Restrita. Tal classe de problemas aparecem em vários campos do conhecimento humano, a exemplo da Engenharia Aeronáutica, Logística, Equações Diferenciais, Engenharia Estrutural, entre vários outros ¹.

¹A título de exemplo, ver ver Júnior (2007), Ojima e Yamakami (2006), Façanha, Carneiro e Filho (2013) e Barros et al. (2017).

Problemas de Otimização Quadrática Restrita ou, equivalentemente, de *Problemas de Programação Quadrática Restrita*, são caracterizados formalmente, e de modo matricial, por

$$\begin{aligned} \text{Minimizar} \quad & f(x) = \frac{1}{2}x^T Sx + v^T x + c \\ \text{sujeito a} \quad & Ax \leq b \end{aligned} \tag{2}$$

com $x, v \in \mathbb{R}^n$, $b \in \mathbb{R}^l$, $c \in \mathbb{R}$, $A \in M_{l \times n}(\mathbb{R})$ e $S \in M_{n \times n}(\mathbb{R})$ sendo que S é simétrica definida positiva ².

Diante disso, notemos que

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Sx + v^T x + c \\ &= \frac{1}{2}(x_{1j})(s_{ij})(x_{j1}) + (v_{1j})(x_{j1}) + c \\ &= \frac{1}{2}(x_{1j}) \left(\sum_{k=1}^n s_{ik}x_{k1} \right) + \left(\sum_{j=1}^n v_{1j}x_{j1} \right) + c \\ &= \frac{1}{2} \left(\sum_{p=1}^n x_{1p} \left(\sum_{k=1}^n s_{ik}x_{k1} \right) \right) + \left(\sum_{i=1}^n v_{1i}x_{i1} \right) + c \\ &= \frac{1}{2} \sum_{p=1}^n \sum_{k=1}^n x_{1p}s_{ik}x_{k1} + \left(\sum_{k=1}^n v_{1k}x_{k1} \right) + c \\ &= \frac{1}{2} \sum_{p=1}^n \sum_{k=1}^n x_p s_{pk} x_k + \sum_{k=1}^n v_k x_k + c \\ &= \frac{1}{2} \sum_{p=1}^n \sum_{k=1}^n s_{pk} x_p x_k + \sum_{k=1}^n v_k x_k + c \end{aligned}$$

Deste modo, podemos reformular (2) como

$$\begin{aligned} \text{Minimizar} \quad & f(x) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n s_{ij} x_i x_j + \sum_{i=1}^n v_i x_i + c \\ \text{sujeito a} \quad & Ax \leq b \end{aligned} \tag{3}$$

que consiste em outra forma de caracterizar os Problemas de Otimização Quadrática Restrita.

Diante disso, observemos que essa categoria se enquadra na classe de problemas não-lineares pois embora suas restrições sejam lineares, a função objetivo a ser otimizada possui termos quadráticos e/ou produto de duas ou mais variáveis.

Isto posto, visando o estudo de métodos de resolução de problemas nos moldes de (2), ou de forma equivalente (3), as próximas seções serão destinadas a apresentação de alguns conceitos e resultados de otimização restrita de modo geral.

²Tendo em mente os propósitos deste trabalho, vale comentar que a Definição 2 é a que aparece na literatura. Todavia, deve-se notar que do ponto de vista técnico a soma $\frac{1}{2}x^T Sx + v^T x$ gerará uma matriz de ordem um, e como c é uma constante real, o que seria somar uma matriz com um número? Um matemático resolve esse problema recordando que existe um isomorfismo entre o espaço das matrizes de ordem um e o conjunto dos números reais, deste modo, pode desconsiderar esse abuso de linguagem uma vez observado este detalhe. Todavia, embora esse detalhe possa parecer pequeno para um matemático, isso pode ser fonte de algumas mensagens de erro durante o processo de implementação computacional dependendo da linguagem utilizada (não é o caso para Python).

2 Alguns conceitos elementares sobre otimização restrita

Dada uma função $f: \mathbb{R}^n \rightarrow \mathbb{R}$, problemas de otimização restrita são geralmente caracterizados por

$$\begin{array}{ll} \text{Otimizar} & f(x) \\ \text{sujeito a} & \Omega = \{x \in \mathbb{R}^n; h(x) = 0 \text{ e } g(x) \leq 0\} \end{array} \quad (4)$$

em que $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$ e $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$. O conjunto Ω recebe o nome de **conjunto viável** do problema de otimização.

Vale destacar que é possível encontrar situações nas quais as restrições envolvidas são apenas de igualdade, nestes casos, (4) se resume a

$$\begin{array}{ll} \text{Otimizar} & f(x) \\ \text{sujeito a} & \Omega = \{x \in \mathbb{R}^n; h(x) = 0\} \end{array} \quad (5)$$

com $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$.

2.1 Condições de Lagrange

Consideremos a função diferenciável $f: A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $x \mapsto f(x) = z$ esteja sujeita à restrição $h: \mathbb{R}^n \rightarrow \mathbb{R}$ de forma que $x \mapsto h(x) = 0$, $x^{*T} = [x_1^*, \dots, x_n^*]$ seja minimizador local de f , $P^T = [x_1^*, \dots, x_n^*, f(x^*)] \in \mathbb{R}^{n+1}$ e seja S a superfície associada a f , isto é, $S = \{(x, f(x)) \in \mathbb{R}^{n+1}; x \in A\}$. Seja ainda a função vetorial $\gamma: X \subset \mathbb{R} \rightarrow A$ tal que $\gamma(t)^T = [x_1(t), \dots, x_n(t)]$ de modo que $\gamma(t) \in A$, $x^* \in gr(\gamma)$ e γ seja diferenciável, sendo $gr(\gamma)$ o gráfico de γ .

Com isso, tomemos t_0 como sendo o argumento de γ correspondente ao ponto x^* , assim $\gamma(t_0)^T = [x_1^*, \dots, x_n^*]$. Isto posto, a composta $f \circ \gamma: X \subset \mathbb{R} \rightarrow \mathbb{R}$ consiste nos valores de f restritos a curva γ , isto é, $(f \circ \gamma)(t) = f(\gamma(t)) = f(x)$ com $x \in A$.

Uma vez que x^* é minimizador local de f , segue que

$$f(x_1^*, \dots, x_n^*)^T = f(x^*) \leq f(x) = f(x_1, \dots, x_n)$$

para todo $x \in A$, assim segue da definição de γ que $x_1^* = x_1(t_0), \dots, x_n^* = x_n(t_0)$ e $x_1 = x_1(t), \dots, x_n = x_n(t)$ e assim

$$\begin{aligned} f(\gamma(t_0)) &= f(x_1(t_0), \dots, x_n(t_0)) \leq f(x_1(t), \dots, x_n(t)) = f(\gamma(t)) \\ &\iff (f \circ \gamma)(t_0) \leq (f \circ \gamma)(t) \end{aligned}$$

para todo $t \in X$. Portanto, segue que t_0 é minimizador local de $f \circ \gamma$, logo

$$\begin{aligned} 0 &= \nabla(f \circ \gamma)(t_0) = \nabla f(\gamma(t_0)) \\ &= \left(\frac{\partial}{\partial x_1} f(x^*) \cdot x_1'(t_0), \dots, \frac{\partial}{\partial x_n} f(x^*) \cdot x_n'(t_0) \right) \quad (\text{regra da cadeia}) \\ &= \left(\frac{\partial}{\partial x_1} f(x^*), \dots, \frac{\partial}{\partial x_n} f(x^*) \right) \cdot (x_1'(t_0), \dots, x_n'(t_0)) = \nabla f(x^*) \cdot \gamma'(t_0). \end{aligned}$$

Ou seja,

$$\nabla f(x^*) \cdot \gamma'(t_0) = 0.$$

Assim, segue por definição que $\nabla f(x^*)$ é perpendicular ao vetor da tangente $\gamma'(t_0)$ para todas as curvas C . Por outro lado, supondo³ $\nabla h(x^*)$ ortogonal a $\gamma'(t_0)$ para todas as curvas, isto é,

$$\nabla h(x^*) \cdot \gamma'(t_0) = 0,$$

obtemos que o $\nabla f(x^*)$ deve ser paralelo a $\nabla h(x^*)$, e conseqüentemente paralelo a $-\nabla h(x^*)$, ou seja, caso $\nabla h(x^*) \neq 0$ existe⁴ algum $\lambda \in \mathbb{R}$ de forma que

$$\nabla f(x^*) = -\nabla h(x^*)\lambda \quad \equiv \quad \nabla f(x^*) + \nabla h(x^*)\lambda = 0. \quad (6)$$

Neste caso, (6) é conhecido como condição de Lagrange e λ é denominado como **multiplicador de Lagrange**.

Com (6) Joseph-Louis Lagrange desenvolveu o método de busca de máximos e mínimos de funções sujeitas a uma restrição $h(x) = 0$. Esse método afirma que caso f possua máximo e mínimo, e ainda, $\nabla h(x) \neq 0$ sobre a superfície $h(x) = 0$, então

Passo 1 Determine todos os valores $x \in A \subset \mathbb{R}^n$ tais que (6) é satisfeita;

Passo 2 Calcule f em cada um dos pontos obtidos no *Passo 1*, o maior dos valores obtidos será o máximo de f e o menor será o mínimo de f .

Isto posto, conforme destacam Izmailov e Solodov (2005), é comum reescrever⁵ (6) em termos da função Lagrangiana, como estabelece a Definição 1.

Definição 1. Consideremos o problema de otimização restrita (5). A função $\mathcal{L}: \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}$ tal que $x \mapsto \mathcal{L}(x, \lambda) = f(x) + h(x)\lambda$, com $h(x) = (h_1(x), \dots, h_l(x))^T$, é denominada **Lagrangiana** de (5).

Assim, quando $\nabla h(x) \neq 0$, a existência de um minimizador local de um problema de otimização restrita nos moldes de (5) está condicionada a existência e unicidade do multiplicador de Lagrange λ . Tal fato fica garantido pelo Teorema 3 conhecido como *Condição de otimalidade de Lagrange*. Diante disto, para enunciar tal teorema precisamos definir o que vem a ser uma condição de regularidade.

Definição 2. Seja o problema de otimização restrita (5). As condições $\{h'_i(x^*); i = 1, \dots, l\}$ ser um conjunto linearmente independente ou h ser uma função afim, isto é, $h(x) = Ax - a$ com $A \in M_{l \times n}$ e $a \in \mathbb{R}^l$, são denominadas **condições de regularidade das restrições** ou ainda condições de qualificação das restrições.

³Para mais detalhes, ver Stewart (2013, p. 845).

⁴Ver Guidorizzi (2000, p. 330) para mais detalhes.

⁵O leitor deve atentar-se ao fato de que a argumentação realizada anteriormente ao Teorema 3 considerou $h: B \subset \mathbb{R}^n \rightarrow \mathbb{R}$, porém, o problema de otimização estabelecido em (5) define $h: \mathbb{R}^n \rightarrow \mathbb{R}^l$. Dito isto, também é possível mostrar a existência de $\lambda \in \mathbb{R}^l$ para o caso em que a imagem de h ($Im(h)$) pertence a \mathbb{R}^l , contudo sua demonstração foge ao escopo deste trabalho. Ao leitor interessado, recomendamos a leitura do Capítulo 2 de Izmailov e Solodov (2005).

Teorema 3 (Condição de otimalidade de Lagrange). *Consideremos o problema de otimização restrita (5). Se $x^* \in \mathbb{R}^n$ é minimizador local de (5), f é diferenciável em x^* , h é diferenciável numa vizinhança de x^* , com derivada contínua em x^* , e uma das condições de regularidade das restrições (Definição 2) válidas, então existe $\lambda^* \in \mathbb{R}^l$ tal que*

$$\mathcal{L}'_x(x^*, \lambda^*) = 0.$$

Caso a primeira condição seja satisfeita, então $\lambda^ \in \mathbb{R}^l$ é único.*

Prova. Ver Izmailov e Solodov (2005, p. 46). □

Com isso, observemos que, para algum $\lambda \in \mathbb{R}^l$,

$$\mathcal{L}'_x(x, \lambda) = \nabla f(x) + (\nabla h(x))^T \lambda$$

enquanto que

$$\mathcal{L}'_\lambda(x, \lambda) = h(x),$$

deste modo, temos que

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

Ou seja, quando $\nabla \mathcal{L}(x, \lambda) = 0$ temos que

$$\nabla \mathcal{L}(x, \lambda) = \begin{bmatrix} \nabla f(x) + (\nabla h(x))^T \lambda \\ h(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (\text{Sistema de Lagrange}) \quad (7)$$

Conforme destacado por Izmailov e Solodov (2005) caso esta condição de Lagrange seja satisfeita significa que o gradiente da função objetivo ($f'(x^*)$) pode ser descrito como uma combinação linear das derivadas parciais das restrições, ou seja, combinação linear de $\{h'_i; i = 1, \dots, l\}$.

Isto posto, assim como mostrado para problemas de otimização restrita nos termos de (5), também é possível mostrar um resultado semelhante para problemas mais gerais, nos moldes de (4), tal resultado é conhecido como “*Condições de Karush-Kuhn-Tucker*”, ou mais economicamente, denominado por “*Condições de KKT*”. Para mais detalhes a respeito, recomendamos a leitura de Izmailov e Solodov (2005), Izmailov e Solodov (2007) e Ribeiro e Karas (2013).

3 Programação quadrática sequencial

Conforme apresentado até o momento existem resultados matemáticos que nos permitem assegurar a existência de solução para problemas de otimização, desde que estes satisfaçam certas condições, além de apresentar meios para sua determinação. Todavia, nem sempre a busca de tais soluções é viável (ou mesmo possível) analiticamente, haja vista a não linearidade das funções envolvidas Luenberger e Ye (2016, p. 06).

Neste sentido, ao longo dos anos foram desenvolvidos numerosos algoritmos para viabilizar a obtenção de solução para problemas de otimização, mesmo que de modo aproximado. Cabe salientar que no contexto de Otimização, a terminologia corrente utilizada como sinônimo de Algoritmo é *Método*.

Mas o que vem a ser um Algoritmo? Conforme definido em Burden, Faires e Burden (2015) um Algoritmo pode ser caracterizado como um “procedimento que descreve, sem ambiguidades, uma sequência finita de passos a serem feitos em uma ordem específica” tendo como objetivo “implementar um procedimento para resolver um problema ou aproximar uma solução do problema”.

Para ilustrar essas ideias, destacando a diferença entre algoritmos sequenciais e não-sequenciais, consideremos a seguinte situação. Imaginemos que estamos cansados de resolver equações de segundo grau na mão e desejamos automatizar isso por meio de um programa de computador, um algoritmo básico para essa finalidade é descrito no Algoritmo 1 a seguir.

```
Entrada  $a \leftarrow$  Valor do coeficiente a;  
 $b \leftarrow$  Valor do coeficiente b;  
 $c \leftarrow$  Valor do coeficiente c;  
if  $b^2 - 4ac \neq 0$  then  
     $x_1 \leftarrow \frac{-b + \sqrt{b^2 - 4ac}}{2a};$   
     $x_2 \leftarrow \frac{-b - \sqrt{b^2 - 4ac}}{2a};$   
else  
     $x_1 \leftarrow \frac{-b}{2a};$   
     $x_2 \leftarrow x_1;$   
end
```

Output: As raízes são x_1 e x_2

Algoritmo 1: Algoritmo básico para a resolução de equações segundo grau

Notemos que o Algoritmo 1 fornece uma receita, por assim dizer, de como obter as raízes de uma equação do segundo grau, independentemente de serem reais ou complexas ⁶. De toda forma, tal algoritmo ainda poderia ser melhorado no sentido de oferecer ao usuário a possibilidade de resolver mais de uma equação, conforme estabelece o Algoritmo 2. Assim, ao acrescentarmos o laço de repetição *while*, durante uma mesma execução do algoritmo teremos a possibilidade resolver varias equações de segundo grau, e mais que isso, cada resolução independe dos valores obtidos na execução do laço anterior, portanto, temos um algoritmo não-sequencial.

⁶Vale comentar que por definição um algoritmo consiste em uma sequência de passos que, ao serem implementados, resolvem um problema ou, ao menos, aproximam uma solução. Sendo assim, não há exigência de uma linguagem de programação em si, embora a utilização de uma seja necessária posteriormente.

De todo modo, a linguagem Python, que foi utilizada na implementação de PQS, tem suporte nativo aos números complexos, sendo assim o Algoritmo 1 tal como apresentado é passível de ser implementado sem qualquer problema quando $b^2 - 4ac < 0$.

```
B ← 1;
while B = 1 do
  a ← Valor do coeficiente a;
  b ← Valor do coeficiente b;
  c ← Valor do coeficiente c;
  if  $b^2 - 4ac \neq 0$  then
     $x_1 \leftarrow \frac{-b + \sqrt{b^2 - 4ac}}{2a}$ ;
     $x_2 \leftarrow \frac{-b - \sqrt{b^2 - 4ac}}{2a}$ ;
  else
     $x_1 \leftarrow \frac{-b}{2a}$ ;
     $x_2 \leftarrow x_1$ ;
  end
  Output: As raízes são  $x_1$  e  $x_2$ 
  Output: Deseja resolver outra equação? 1 para SIM. 0 para NÃO.
  B ← Entrada do usuário ;
end
```

Algoritmo 2: Algoritmo básico para a resolução de equações de segundo grau

Todavia, existem algoritmos que para cada nova repetição do laço (*while*, *for*, ...) utilizam-se dos resultados obtidos anteriormente. O método da bissecção descrito no Algoritmo 3 é um exemplo deste tipo de algoritmo.

```
ostrar um intervalo  $[a, b]$ ;
Escolher a condição inicial como sendo  $x_0 = \frac{b+a}{2}$ ;
k ← 0;
while  $f'(x_k) \neq 0$  do
  if  $f'(x_k) < 0$  then
    a ←  $x_k$ ;
  else
    b ←  $x_k$ ;
  end
   $x_k = \frac{b+a}{2}$ ;
end
```

Algoritmo 3: Exemplo genérico de implementação do método da bissecção

Observemos que a cada execução do laço *while* o método da bissecção reduz o intervalo de busca da solução do problema, que só é possível pois a informação do valor da derivada em x_k é levada em consideração para atualizar o valor de a , ou de b . Ou seja, enquanto o critério de parada não for satisfeito, enquanto $f'(x_k) = 0$ não ocorrer, cada execução do laço *while* atualiza o intervalo de busca para uma nova execução considerando a informação obtida na execução atual.

É intuitivo notar que o Algoritmo 3 gera portanto uma sequência de pontos de \mathbb{R} , deste modo, temos que tal algoritmo é caracterizado como um *Método Sequencial*. De modo geral, conforme formalizam Izmailov e Solodov (2007), temos um *Método Sequencial* “quando cada ponto se define de acordo com a informação obtida nos pontos anteriores”.

Definição 4. Dado um problema de otimização nos moldes de (4) as imagens $x^1, \dots, x^k, \dots \in \mathbb{R}^n$ da sequência de pontos geradas por um método sequencial são denominados *aproximações* à solução do problema, ou ainda, *iterandos* do método ⁷.

Convém mencionar que a sequência $(x^k)_{k \in \mathbb{N}}$ gerada por algum método sequencial também é conhecida como *trajetória*.

Definição 5. Seja $x: \mathbb{N} \rightarrow \mathbb{R}^n$ uma sequência gerada por algum método sequencial. A geração de uma nova aproximação x^{x+1} a partir de x^k é denominada *iteração* do método.

Isto posto, vale destacar que vários algoritmos foram desenvolvidos para a resolução computacional de problemas de Otimização Quadrática, dentre eles tem-se o Método de Pontos Interiores e o Simplex Modificado. Todavia, neste trabalho temos interesse em estudar o algoritmo conhecido como Programação Quadrática Sequencial (PQS) e vamos mostrar que no caso de problemas nos moldes de 2 (com restrições de igualdade) tal algoritmo os resolve em apenas uma iteração.

3.1 O algoritmo PQS

Conforme apresentado por Izmailov e Solodov (2007) e Ribeiro e Karas (2013) o método de Programação Quadrática Sequencial (PQS) consiste, em essência, “na resolução de uma sequência de problemas de otimização quadrática” em que

a ideia consiste em substituir, a cada iteração, a função objetivo por um modelo quadrático do Lagrangiano e as restrições por equações ou inequações lineares, aproximações de Taylor de primeira ordem em torno no ponto corrente (RIBEIRO; KARAS, 2013).

Convém destacar que esse modelo quadrático do Lagrangiano consiste na aproximação de Taylor de segunda ordem para a função objetivo.

Com isso, esta abordagem permite reduzir um problema inicialmente complexo, a resolução de uma sequência de problemas relativamente mais “simples”, “que muda a cada iteração de acordo com as informações disponíveis no ponto corrente” (RIBEIRO; KARAS, 2013). E conforme pontua Izmailov e Solodov (2007) quando um problema de programação quadrática é adequadamente construído pode ser uma boa aproximação do problema original, ao mesmo tempo em que é de resolução relativamente fácil.

⁷Notemos que a notação x_k , usada no contexto de \mathbb{R} , é substituída por x^k facilitando, desta forma, a associação com o \mathbb{R}^n .

Diante disso, considerando o problema (5), temos que (8) formaliza matematicamente estas ideias.

$$\begin{aligned} \text{Otimizar} \quad & \mathcal{L}(x^k, \lambda^k) + \nabla_x(x^k)(x - x^k) + \frac{1}{2}\nabla_{xx}^2\mathcal{L}(x^k, \lambda^k)(x - x^k) \\ \text{sujeito a} \quad & x \in \Omega = \{x \in \mathbb{R}^n; h(x^k) + h'(x - x^k) = 0\} \end{aligned} \quad (8)$$

Isto posto, o Algoritmo 4 apresenta o método PQS ⁸.

scolher o palpite inicial (x^0, λ^0) ;

$k \leftarrow 0$;

while $\nabla\mathcal{L}(x^k, \lambda^k) \neq 0$ **do**

Passo 1:

Resolva o problema $\nabla\mathcal{L}(x^*, \lambda) = 0$ obtendo um ponto estacionário do problema e o defina como x^{k+1} , bem como o multiplicador de Lagrange associado e o defina como λ^{k+1} .

Passo 2: $k \leftarrow k + 1$

end

Algoritmo 4: O Método PQS

Embora o Algoritmo 4 seja relativamente simples, e extremamente compacto, é natural o questionamento “*Como executamos o Passo 1?*”. Para isso, consideremos o problema de minimização conforme definido em (5). Com isso, de acordo com o apresentado na Subseção 2.1, em particular na igualdade (7), segue que para x^* ser minimizador local de f deverá ocorrer (para algum $\lambda \in \mathbb{R}^l$) que

$$\nabla\mathcal{L}(x^*, \lambda) = 0 \iff \begin{bmatrix} \nabla f(x^*) + (\nabla h(x^*))^T \lambda \\ h(x^*) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (9)$$

À luz disso, notemos que podemos fazer uso do método de Newton-Raphson⁹ ao considerarmos

$$\begin{aligned} \varphi: \mathbb{R}^n & \longrightarrow \mathbb{R}^{n+l} \\ x & \longmapsto \varphi(x) = \nabla\mathcal{L}(x, \lambda). \end{aligned}$$

Assim, reescrevendo (9) vem que

$$\varphi(x^*) = 0. \quad (10)$$

⁸Relembremos o conceito de ponto estacionário.

Definição 6. Sejam $f: \Omega \subset \mathbb{R}^m \rightarrow \mathbb{R}$ diferenciável em Ω e $x^* \in \Omega$. Definimos x^* como sendo *ponto crítico* ou (*estacionário*) de f se, e só se, $\nabla f(x^*) = 0$.

⁹Para detalhes a respeito deste método, ver Izmailov e Solodov (2007), Ribeiro e Karas (2013) ou <https://www.ufrgs.br/reatmat/CalculoNumerico/livro-py/livro-py.pdf>.

Tomando $x^k \in \mathbb{R}^n$ como sendo uma aproximação para a solução x^* e aplicando o teorema de Taylor de primeira ordem em torno de x^k (com k fixo), obtemos que

$$\varphi(x) = \varphi(x^k) + \nabla\varphi(x^k)(x - x^k) + r(x)$$

para $\lim_{x \rightarrow x^k} \frac{r(x)}{|x - x^k|} = 0$ e $x \in \text{Dom}(\varphi)$, isto é, para x^k suficientemente próximo à x^* , temos que

$$\varphi(x) = \varphi(x^k) + \nabla\varphi(x^k)(x - x^k), \quad (11)$$

assim, de (10) e (11) temos que

$$\varphi(x^k) + \nabla\varphi(x^k)(x^* - x^k) = 0,$$

dessa forma, a execução do Passo 1 no Algoritmo 4 depende da não-singularidade da Hessiana de \mathcal{L} em (x^k, λ^k) , além é claro que as somas e multiplicações matriciais sejam possíveis. Caso tais condições sejam satisfeitas, temos que

$$\begin{aligned} \varphi(x^k) + \nabla\varphi(x^k)(x^* - x^k) &= 0 \\ \implies \nabla\mathcal{L}(x^k, \lambda^k) + \nabla(\nabla\mathcal{L}(x^k, \lambda^k)) \left(\begin{bmatrix} x^* \\ \lambda \end{bmatrix} - \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} \right) &= 0 \\ \implies \nabla^2\mathcal{L}(x^k, \lambda^k) \left(\begin{bmatrix} x^* \\ \lambda \end{bmatrix} - \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} \right) &= -\nabla\mathcal{L}(x^k, \lambda^k) \end{aligned}$$

e para $(\nabla^2\mathcal{L}(x^k, \lambda^k))^{-1}$ não singular, temos que

$$\begin{aligned} \implies \begin{bmatrix} x^* - x^k \\ \lambda - \lambda^k \end{bmatrix} &= -(\nabla^2\mathcal{L}(x^k, \lambda^k))^{-1} \nabla\mathcal{L}(x^k, \lambda^k) \\ \implies \begin{bmatrix} x^* \\ \lambda \end{bmatrix} &= \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} - (\nabla^2\mathcal{L}(x^k, \lambda^k))^{-1} \nabla\mathcal{L}(x^k, \lambda^k). \end{aligned} \quad (12)$$

Portanto, a próxima aproximação x^{k+1} para x^* é tal que

$$\begin{bmatrix} x^{k+1} \\ \lambda^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} - (\nabla^2\mathcal{L}(x^k, \lambda^k))^{-1} \nabla\mathcal{L}(x^k, \lambda^k). \quad (13)$$

Esta abordagem é conhecida como *Método de Newton para o Sistema de Lagrange*.

Vale destacar que, em geral, como o custo computacional para inverter matrizes é maior que o custo da resolução de sistemas lineares¹⁰ e, visando um melhor desempenho computacional, (12) foi reescrita em termos do sistema linear a seguir

$$\nabla^2\mathcal{L}(x^k, \lambda^k) \begin{bmatrix} x \\ \lambda \end{bmatrix} = \nabla^2\mathcal{L}(x^k, \lambda^k) \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} - \nabla\mathcal{L}(x^k, \lambda^k) \quad (14)$$

deste modo, a próxima aproximação x^{k+1} para x^* consiste em nada mais que a solução do sistema linear (14), isto é, $x^{k+1} = x$ e $\lambda^{k+1} = \lambda$.

¹⁰Para mais detalhes ver <https://www.ufrgs.br/reamat/CalculoNumerico/livro-py/livro-py.pdf>.

Um caso particular

Isto posto, uma questão razoável que emerge é: “E caso a função objetivo do problema que estamos trabalhando já seja uma quadrática? Ou ainda, e se além da função objetivo ser quadrática, as restrições envolvidas nesse problema já fossem lineares? Como ficaria a aplicação deste método?”. Pois bem, visando responder a tais questionamentos, consideremos o problema quadrático similar ao definido em (2), porém de forma que as restrições envolvidas sejam de igualdade, isto é,

$$\begin{aligned} \text{Minimizar} \quad & f(x) = \frac{1}{2}x^T Sx + v^T x + c \\ \text{sujeito a} \quad & Ax - b = 0, \end{aligned} \tag{15}$$

com $x, v, b \in \mathbb{R}^n$, $c \in \mathbb{R}$, $S \in M_{n \times n}(\mathbb{R})$ e $A \in M_{l \times n}(\mathbb{R})$ sendo que S é simétrica definida positiva.

Neste caso uma vez que

$$\nabla^2 \mathcal{L}(x^k, \lambda^k)_{(n+l) \times (n+l)}, \begin{bmatrix} x \\ \lambda \end{bmatrix}_{(n+l) \times 1}, \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix}_{(n+l) \times 1} \quad \text{e} \quad \nabla \mathcal{L}(x^k, \lambda^k)_{(n+l) \times 1}$$

ou seja, os vetores e matrizes envolvidos cumprem os critérios exigidos pelas definições de soma e multiplicação matricial, temos que

$$\nabla \mathcal{L}(x^k, \lambda^k) = \begin{bmatrix} \nabla f(x^*) + (\nabla h(x))^T \lambda^k \\ h(x^*) \end{bmatrix} = \begin{bmatrix} Sx^* + v + A^T \lambda^k \\ Ax^* - b \end{bmatrix}, \tag{16}$$

bem como

$$\nabla^2 \mathcal{L}(x^k, \lambda^k) = \nabla \left(\nabla \mathcal{L}(x^k, \lambda^k) \right) = \nabla \left(\begin{bmatrix} Sx^* + v + A^T \lambda^k \\ Ax^* - b \end{bmatrix} \right) = \begin{bmatrix} S & A^T \\ A & 0 \end{bmatrix}. \tag{17}$$

Desta forma, levando (16) e (17) em (14) obtemos

$$\begin{bmatrix} S & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} S & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} - \begin{bmatrix} Sx^* + v + A^T \lambda^k \\ Ax^* - b \end{bmatrix}, \tag{18}$$

e com isso, a determinação de (x, λ) se resume a resolução do sistema Sistema Linear de Lagrange. Com isso, ao aplicar o PQS em (15), a solução é obtida em apenas uma iteração conforme (18) justifica, desde que, é claro, a Hessiana da Lagrangena em (x^k, λ^k) seja não-singular.

Vale frisar que o número de restrições é irrelevante para (14) do ponto de vista teórico, isto é, o número de linhas na matriz de restrições A pode ser qualquer número natural, não se limitando a ter a mesma ordem que S .

Além disso, convém pontuar que por mais que teoricamente o problema seja resolvido em apenas uma iteração, esta *uma iteração* pode demandar um alto custo temporal dependendo do número de variáveis, do número de restrições e do *hardware* envolvido.

Diante disso, a seção seguinte ilustra a resolução de problemas quadráticos com o PQS conforme apresentado nas linhas anteriores.

4 Exemplos

Com o objetivo de apresentar alguns exemplos numéricos da teoria exposta nas seções anteriores, bem como ter uma noção do desempenho computacional, o Algoritmo 4 foi traduzido para a linguagem Python considerando o Método de Newton para o Sistema de Lagrange nos termos de (14).

Com isso, um banco de problemas de Programação Quadrática Restrita foi criado por intermédio do algoritmo “`generating-input-data.py`” disponível em <https://github.com/AugustoAlex/Mathematica-V2-PQS>, sendo que a matriz S simétrica definida positiva foi criada com entradas pseudo-aleatórias ¹¹, por meio do módulo `random` ¹² da linguagem Python, conforme pode ser verificado no código-fonte supracitado.

Além disso, é pertinente mencionar as características físicas do *hardware* no qual os exemplos foram resolvidos, a saber, consistiu em um *notebook* pessoal com processador Intel(R) Core(TM) i7-5500U munido de dois núcleos e quatro *thread(s)* possuindo velocidade mínima de 500 MHz e máxima de 3000 MHz. Além de contar com dois chips de memória RAM SODIMM DDR3 de 4096MB cada, operando a 1600 MHz, bem como um SSD GIGABYTE GP-GSTFS31240GNTD.

Como características de software, utilizou-se o Python em sua versão 3.10.8 com seu interpretador padrão (CPython). Tal software foi executado dentro de um sistema operacional GNU/Linux, mais precisamente, Arch Linux.

Diante disso, seguem os exemplos.

Exemplo 1. Consideremos o problema de otimização

$$\begin{array}{l} \text{Otimizar} \\ \text{sujeito a} \end{array} \quad f(x) = x^T \begin{pmatrix} 142,71099 & -123,3046 \\ -123,30464 & 108,28907 \end{pmatrix} x + \begin{bmatrix} 47 & 59 \end{bmatrix} x + 4$$

$$\begin{bmatrix} -63 & -77 \\ 2 & -48 \\ -60 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 64 \\ 32 \\ 14 \end{bmatrix} = 0$$

Condição inicial x^1	$f(x^1) \approx$	Solução aproximada obtida (x^2)	$f(x^2) \approx$
$\begin{pmatrix} 59 & 36 \end{pmatrix}$	61.561,63686	$\begin{pmatrix} -0,07745 & -0,60868 \end{pmatrix}$	-20,87660

¹¹O algoritmo para obtenção dessa matriz S foi implementado segundo sugere Ribeiro e Karas (2013), enquanto que todos os demais elementos do problema (2) também foram gerados com entradas pseudo-aleatórias.

Ainda, convém também pontuar que quando falamos de programação computacional não existe aleatoriedade no sentido estrito, mas sim obtenção de números que *parecem* aleatórios, em suma na pseudo-aleatoriedade a saída do gerador *parece* aleatória para um observador externo, conforme conceituado em Gutterman, Pinkas e Reinman (2006).

¹²Para mais detalhes a respeito deste módulo recomenda-se a leitura de <https://docs.python.org/3/library/random.html?highlight=random#module-random>.

Ainda, convém destacar que a obtenção da solução para este exemplo (com apenas 3 variáveis) e três restrições levou cerca de 0,00345 segundos.

A Figura 1 ilustra o gráfico associado a este exemplo. Nele, o ponto em vermelho identifica o minimizador deste problema.

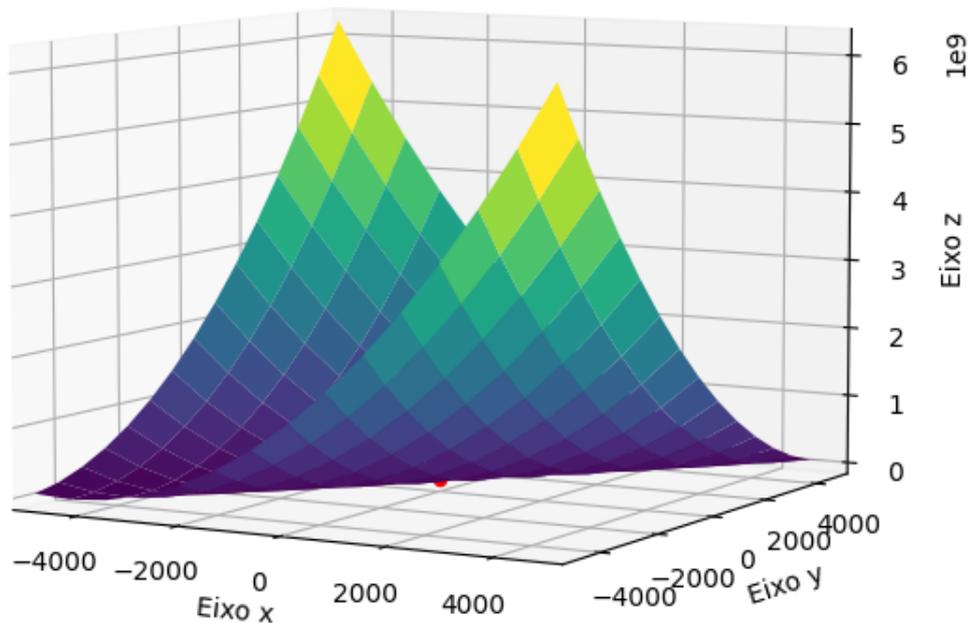


Figura 1: Gráfico associado ao problema de otimização do Exemplo 1

Na sequência, são apresentados os valores funcionais iniciais e finais, bem como o tempo de execução, de dois outros exemplos, um com 1000 variáveis e 1000 restrições. E, outro com 5000 variáveis e 5000 restrições.

Exemplo 2. Este exemplo trata-se de um problema contendo 1000 variáveis e 1000 restrições que, conforme descrito anteriormente, foi gerado pelo algoritmo “`generating-input-data.py`”. Os resultados obtidos são apresentados a seguir.

$f(x^1)$ aproximado	$f(x^2)$ aproximado	Tempo (s)
802.407.623,98335	1.798.430,50698	0,33481

Exemplo 3. Similarmente ao exemplo anterior, este consiste de um problema contendo 5000 variáveis e 5000 restrições que foi construído pelo algoritmo “`generating-input-data.py`”. Os resultados obtidos são dispostos a seguir.

$f(x^1)$ aproximado	$f(x^2)$ aproximado	Tempo (s)
442.862.516,44263	98.902,74119	18,48927

Conclusões

Neste trabalho escrevemos sobre o Método de Programação Quadrática Restrita, mostramos que no caso da função objetivo ser quadrática, e possuir restrições lineares (de igualdade), o problema é resolvido em apenas uma iteração. Por fim, visando exemplificar a teoria bem como ter noção do desempenho computacional, apresentamos três exemplos numéricos.

Referências

- BARROS, L. N. et al. Técnica de programação semidefinida por arco viável e aplicação à maximização da frequência natural de estruturas mecânicas. Universidade Federal da Paraíba, 2017.
- BUENO, H. P. *Álgebra linear: um segundo curso*. Rio de Janeiro: SBM, 2006. v. 1.
- BURDEN, R. L.; FAIRES, J. D.; BURDEN, A. M. *Análise numérica*. Tradução All Tasks. Revisão técnica Helena Maria Ávila de Castro. São Paulo: Cengage Learning, 2015. v. 3.
- FAÇANHA, T. S.; CARNEIRO, A. L.; FILHO, J. T. Filtro de kalman via programação quadrática. 2013.
- FARIAS, C. A. Modelos de otimização de portfólios: análise comparativa e aplicações ao mercado acionário brasileiro. Universidade Federal de Viçosa, 2003.
- GUIDORIZZI, H. L. *Um curso de cálculo*, vol. 2. [S.l.]: Grupo Gen-LTC, 2000.
- GUTTERMAN, Z.; PINKAS, B.; REINMAN, T. *Analysis of the Linux Random Number Generator*. 2006. Cryptology ePrint Archive, Paper 2006/086. Disponível em: <https://eprint.iacr.org/2006/086>.
- IZMAILOV, A.; SOLODOV, M. *Otimização - volume 1: Condições de otimalidade, elementos de análise convexa e de dualidade*. Rio de Janeiro: Impa, 2005.
- IZMAILOV, A.; SOLODOV, M. *Otimização - volume 2: Métodos computacionais*. Rio de Janeiro: Impa, 2007.
- JÚNIOR, F. de Souza. Simulação numérica de otimização de projeto de compressores axiais utilizando o método da programação sequencial quadrática. 2007.
- LUENBERGER, D. G.; YE, Y. *Linear and nonlinear programming*. [S.l.]: Springer, 2015. v. 4.
- OJIMA, A. L.; YAMAKAMI, A. Modelo de programação quadrática para análise da movimentação logística e comercialização da soja brasileira. *Engenharia Agrícola*, SciELO Brasil, v. 26, p. 552-560, 2006.
- RIBEIRO, A. A.; KARAS, E. W. *Otimização contínua: aspectos teóricos e computacionais*. São Paulo: Cengage Learning, 2013.

STEWART, J. *Cálculo*: volume 2. EZ2 Translate. 7. ed. São Paulo: Cengage Learning, 2013.

VANDERBEI, R. J. *Linear Programming*: Foundations and extensions. 4. ed. New York: Springer, 2014.

Alex Augusto Nunes Machado
augusttoalex@gmail.com

Simone Aparecida Miloca
smiloca@gmail.com

Representação de rotações no espaço com quatérnios

Flavio Roberto Dias Silva - Universidade Estadual do Oeste do Paraná (UNIOESTE)

(Recebido em 09/11/2022. Aceito em 05/12/2022. Publicado em 22/12/2022)

Resumo: É comum e intuitivo encontrar a representação de rotações no espaço em termos de ângulos e eixos, no entanto, existem outras representações para rotações com grande utilidade prática. Uma alternativa para a representação de rotações é a utilização de quatérnios. Os quatérnios, entre outras propriedades, podem representar uma rotação no espaço e ser operados, de maneira a representar novas rotações, apenas com operações básicas de soma e produto. Com isso em mente, esse artigo tem o objetivo de mostrar a representação de rotações através de quatérnios.

Palavras-chave: quatérnios, rotações, Hamilton.

1 Álgebra dos quatérnios

Nessa seção faremos um levantamento geral da álgebra dos quatérnios, como o intuito do artigo não é o de detalhar e demonstrar todas as definições e propriedades dos quatérnios, sugerimos o artigo GUZZO & TOSTI(2017) que faz a exploração dos elementos da álgebra dos quatérnios.

Os quatérnios foram organizados para estender as propriedades dos números complexos para o espaço tridimensional e sua definição carrega alguns elementos que remetem aos complexos. A seguir damos a definição formal dos quatérnios.

Definição 1. O conjunto dos números quatérnios, denotado por \mathbb{H} , é formado por todos os números na forma $q = a + bi + cj + dk$, com $a, b, c, d \in \mathbb{R}$ e i, j, k satisfazendo $i^2 = j^2 = k^2 = -1$, $ij = k$ e $ji = -k$.

Da definição conseguimos retirar uma série de igualdades bastante usuais e recorrentes nas operações com quatérnios, entre elas:

$$ijk = (ij)k = kk = k^2 = -1$$

$$ik = i(ij) = i^2j = -j$$

$$kj = (ij)j = ij^2 = -i$$

$$ki = k(-j^2)i = -(kj)(ji) = -(-i)(-k) = -ik = j$$

$$jk = j(-i^2)k = -(ji)(ik) = -(-k)(-j) = -kj = i$$

Nos cálculos acima é assumida a associatividade da multiplicação nas componentes i, j e k .

Da definição 1, conseguimos observar várias semelhanças com os números complexos, em especial, gostaria de dar destaque aos elementos i, j e k que carregam a propriedade da unidade imaginárias dos números complexos, eles tem o quadrado igual a -1 .

Tal como nos complexos, para os números quatérnios, definimos parte real e parte imaginaria e é o objeto da definição a seguir.

Definição 2. Dado um quatérnio $q = a + bi + cj + dk$, definimos a parte real de q como o número real a e denotamos $a = Re(q)$. Também definimos a parte imaginaria de q como o vetor $\vec{u} = (b, c, d)$ e denotamos $Im(q) = (b, c, d)$.

Nos termos da definição acima identificamos o vetor $\vec{u} = (b, c, d) \in \mathbb{R}^3$ com o quatérnio $bi + cj + dk$, de onde podemos escrever $q = a + \vec{u}$.

Chamemos de \mathbb{H}^0 os quatérnios tais que $Re(q) = 0$ isso é,

$$\mathbb{H}^0 = \{q \in \mathbb{H} \mid Re(q) = 0\},$$

chamaremos também os elementos de \mathbb{H}^0 de quatérnios puros.

Definição 3. Dados dois números quatérnios, $q = a + bi + cj + dk$ e $r = e + fi + gj + hk$, dizemos que $q = r$ se, e somente se, $a = e, b = f, c = g$ e $d = h$. De outra forma $q = r$ se, e somente se, $Re(q) = Re(r)$ e $Im(q) = Im(r)$.

As considerações anteriores definem o conjunto dos números quatérnios. No que segue, apresentaremos as operações definidas nesse conjunto.

Definição 4. Definimos a soma entre dois números quatérnios $q = a + bi + cj + dk = a + \vec{u}$ e $r = e + fi + gj + hk = e + \vec{v}$ como a aplicação

$$\begin{aligned} + : \mathbb{H} \times \mathbb{H} &\longrightarrow \mathbb{H} \\ (q, r) &\longmapsto q + r \end{aligned}$$

onde a soma é definida por

$$q + r = (a + bi + cj + dk) + (e + fi + gj + hk) = (a + e) + (b + f)i + (c + g)j + (d + h)k$$

que é equivalente a

$$q + r = (a + \vec{u}) + (e + \vec{v}) = (a + e) + (\vec{u} + \vec{v})$$

onde $(\vec{u} + \vec{v})$ denota a soma usual de vetores em \mathbb{R}^3 .

É importante destacar que a soma de números quatérnios goza das propriedades usuais de soma, isso é, ela é associativa, comutativa, tem elemento neutro igual a $q = 0 + \vec{0}$ e um quatérnio $r = a + bi + cj + dk$ tem oposto em relação a soma dado por $w = -a - bi - cj - dk$. Não faremos a demonstração dessas propriedades aqui mas indicamos as bibliografias BUTZGE (2021), GUZZO & TOSTI(2017), HAMILTON(1844) onde podem ser encontradas.

Definição 5. Definimos a multiplicação de um número quaternião $q = a + bi + cj + dk$ por um número real α como a aplicação

$$\begin{aligned} \cdot : \mathbb{R} \times \mathbb{H} &\longrightarrow \mathbb{H} \\ (\alpha, q) &\longmapsto \alpha q \end{aligned}$$

onde a operação é definida por

$$\alpha q = \alpha(a + bi + cj + dk) = (\alpha a) + (\alpha b)i + (\alpha c)j + (\alpha d)k$$

que também pode ser visualizada na forma vetorial como

$$\alpha q = \alpha(a + \vec{u}) = (\alpha a) + (\alpha \vec{u}),$$

aqui $\alpha \vec{u}$ denota a operação usual de multiplicação por escalar de vetores de \mathbb{R}^3 .

Definição 6. Definimos a multiplicação entre dois números quaterniões $q = a + bi + cj + dk = a + \vec{u}$ e $r = e + fi + gj + hk = e + \vec{v}$ como a aplicação

$$\begin{aligned} \cdot : \mathbb{H} \times \mathbb{H} &\longrightarrow \mathbb{H} \\ (q, r) &\longmapsto qr \end{aligned}$$

onde a multiplicação é definida por

$$\begin{aligned} qr &= (a + bi + cj + dk)(e + fi + gj + hk) \\ &= (ae - bf - cg - dh) + (af + be + ch - dg)i + (ag - bh + ce + df)j + (ah - bg + cf + de)k \end{aligned}$$

que pode ser representado na forma vetorial como

$$qr = (a + \vec{u})(e + \vec{v}) = (ae - \langle \vec{u}, \vec{v} \rangle) + (a\vec{v} + e\vec{u} + \vec{u} \wedge \vec{v}),$$

onde $\langle \vec{u}, \vec{v} \rangle$ denota o produto interno (ou produto escalar) de vetores em \mathbb{R}^3 e $\vec{u} \wedge \vec{v}$ o produto vetorial entre vetores de \mathbb{R}^3 .

É importante destacar que a multiplicação de quaterniões tem a propriedade associativa, porém a multiplicação de quaterniões não é comutativa, basta tomar como exemplo $q = i$ e $r = j$ no conjunto dos quaterniões e verificar que

$$qr = ij = k \text{ e } rq = ji = -k.$$

A multiplicação de quaterniões tem elemento neutro igual a $1 + 0i + 0j + 0k = 1 + \vec{0}$, ao qual escreveremos simplesmente 1, ficando subentendido que $1 \in \mathbb{H}$. Isso é, dado $q = a + bi + cj + dk$, temos

$$q \cdot 1 = 1 \cdot q = q.$$

Para mais detalhes ver BUTZGE(2021), GUZZO & TOSTI(2017), HAMILTON(1844).

E finalmente, a multiplicação de quatérnios tem o elemento inverso. Dado um quatérnio $q = a + bi + cj + dk = a + \vec{u}$, tal que $a^2 + b^2 + c^2 + d^2 > 0$, este admite elemento inverso pela multiplicação o qual será denotado por q^{-1} e pode ser expresso por

$$q^{-1} = \left(\frac{a}{a^2 + |\vec{u}|^2} \right) + \left(\frac{-1}{a^2 + |\vec{u}|^2} \right) \vec{u} = \left(\frac{1}{a^2 + |\vec{u}|^2} \right) (a - \vec{u})$$

onde $|\vec{u}| = \sqrt{b^2 + c^2 + d^2}$ é a norma usual de um vetor de \mathbb{R}^3 . Além disso, esse q^{-1} é o único quatérnio que cumpre

$$qq^{-1} = q^{-1}q = 1$$

Definição 7. Seja $q = a + \vec{u} = a + bi + cj + dk \in \mathbb{H}$. O módulo de q é o número real definido por

$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2} = \sqrt{a^2 + |\vec{u}|^2}.$$

Definição 8. Seja $q = a + \vec{u} = a + bi + cj + dk \in \mathbb{H}$. O conjugado de q é o quatérnio definido por

$$\bar{q} = a - bi - cj - dk = a - \vec{u}.$$

Das duas ultimas definições podemos reescrever a expressão para o inverso de um quatérnio. Dado $q = a + bi + cj + dk = a + \vec{u}$ com $a^2 + b^2 + c^2 + d^2 > 0$, seu inverso pode ser escrito como

$$q^{-1} = \frac{1}{|q|^2} \bar{q}.$$

Observe que, da igualdade anterior, quando q é um quatérnio de módulo 1, vale que $\bar{q} = q^{-1}$.

2 Representação de rotações com quatérnios

Uma rotação no plano pode ser representada por um número complexo, podemos associar a matriz de rotação aplicada em um vetor com o produto de dois números complexos. A ideia para representar rotações com quatérnios segue o mesmo caminho, representar a matriz de rotação no espaço tridimensional aplicada a um vetor como o produto de dois quatérnios.

Considere o conjunto

$$S^3 = \{q \in \mathbb{H} \mid |q| = 1\}.$$

Sendo $q \in S^3$, ou seja, um quatérnio de módulo 1, é possível construir um operador de rotação com q de maneira que seu efeito sobre um vetor não nulo $v \in \mathbb{R}^3$, seja rotacioná-lo um angulo $\theta \in \mathbb{R}$ em torno de um eixo específico BUTZGE(2021), HANSON(2005).

Como já fizemos anteriormente, usaremos a identificação de $u \in \mathbb{R}^3$ com um quatérnio puro $q = 0 + \vec{u} \in \mathbb{H}^0$, identificando \mathbb{R}^3 e \mathbb{H}^0 podemos usar a multiplicação de quatérnios para representar as rotações no espaço. Nesse caminho, considere o operador definido abaixo:

Definição 9. Dado $q \in S^3$, o operador de rotação R_q obtido a partir de q é definido por

$$\begin{aligned} R_q : \mathbb{H}^0 &\longrightarrow \mathbb{H}^0 \\ v &\longmapsto qv\bar{q}. \end{aligned}$$

O operador definido envolve o produto entre três quatérnios, q e \bar{q} quatérnios de módulo unitário que fazem parte da definição do operador e o quatérnio puro v , o quatérnio no qual age o operador, que se identifica com um vetor de \mathbb{R}^3 .

Trabalhamos para que esse operador represente uma rotação no espaço, nesse caminho, uma verificação importante de se fazer nesse momento inicial é a boa definição do operador, isso é, se resultado dessa operação realmente é um quatérnio puro de \mathbb{H}^0 . De fato, seja $q = a + \vec{u}$ e $v = 0 + \vec{v}$, teremos

$$\begin{aligned} R_q(v) = qv\bar{q} &= (a + \vec{u})v(a - \vec{u}) \\ &= (-\langle \vec{u}, \vec{v} \rangle + a\vec{v} + \vec{u} \wedge \vec{v})(a - \vec{u}) \\ &= -a\langle \vec{u}, \vec{v} \rangle + \langle a\vec{v} + \vec{u} \wedge \vec{v}, \vec{u} \rangle + \langle \vec{u}, \vec{v} \rangle \vec{u} + a(a\vec{v} + \vec{u} \wedge \vec{v}) \\ &\quad + (a\vec{v} + \vec{u} \wedge \vec{v}) \wedge \vec{u}. \end{aligned} \tag{1}$$

Primeiramente olhemos para a parte real da expressão acima, esta é,

$$\begin{aligned} -a\langle \vec{u}, \vec{v} \rangle + \langle a\vec{v} + \vec{u} \wedge \vec{v}, \vec{u} \rangle &= -a\langle \vec{u}, \vec{v} \rangle + a\langle \vec{v}, \vec{u} \rangle + \langle \vec{u} \wedge \vec{v}, \vec{u} \rangle \\ &= \langle \vec{u} \wedge \vec{v}, \vec{u} \rangle. \end{aligned}$$

Lembrando que $\vec{u} \wedge \vec{v}$ é o produto vetorial da geometria analítica e representa um vetor simultaneamente ortogonal a \vec{u} e \vec{v} e, além disso, o produto interno entre dois vetores ortogonais é zero, de onde concluímos que a parte real $Re(R_q(v)) = 0$, então este é elemento de \mathbb{H}^0 , confirmando que o operador está bem definido.

Agora trabalhando com a parte vetorial de (1) temos

$$\begin{aligned} \langle \vec{u}, \vec{v} \rangle \vec{u} + a(a\vec{v} + \vec{u} \wedge \vec{v}) + (a\vec{v} + \vec{u} \wedge \vec{v}) \wedge \vec{u} &= \langle \vec{u}, \vec{v} \rangle \vec{u} + a^2\vec{v} + a(\vec{u} \wedge \vec{v}) \\ &\quad - a\vec{v} \wedge \vec{u} + (\vec{u} \wedge \vec{v}) \wedge (-\vec{u}) \\ &= \langle \vec{u}, \vec{v} \rangle \vec{u} + a^2\vec{v} + 2a(\vec{u} \wedge \vec{v}) \\ &\quad + (\vec{u} \wedge \vec{v}) \wedge (-\vec{u}). \end{aligned} \tag{2}$$

Observe que o último termo de (2) apresenta um duplo produto vetorial, conforme pode ser visto em BOULOS & CAMARGO(2007) e WINTERLE & STEINBRUCH(2000), esse termo pode ser escrito como

$$(\vec{u} \wedge \vec{v}) \wedge (-\vec{u}) = \langle \vec{u}, -\vec{u} \rangle \vec{v} - \langle \vec{v}, -\vec{u} \rangle \vec{u} = -|\vec{u}|^2 \vec{v} + \langle \vec{u}, \vec{v} \rangle \vec{u}$$

e substituindo em (2) teremos

$$\langle \vec{u}, \vec{v} \rangle \vec{u} + a(a\vec{v} + \vec{u} \wedge \vec{v}) + (a\vec{v} + \vec{u} \wedge \vec{v}) \wedge \vec{u} = (a^2 - |\vec{u}|^2) \vec{v} + 2\langle \vec{u}, \vec{v} \rangle \vec{u} + 2a(\vec{u} \wedge \vec{v}) \tag{3}$$

Desta ultima expressão, a equação (3), podemos escrever a atuação desse operador como uma operação matricial. Olhando separadamente para as tres parcelas da soma no lado direito de (3) e considerando que $\vec{v} = (x, y, z)$ além de $q = a + \vec{u} = a + bi + cj + dk$ teremos

1º termo

$$(a^2 - |\vec{u}|^2)\vec{v} = (a^2 - b^2 - c^2 - d^2) \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

2º termo

$$2\langle \vec{u}, \vec{v} \rangle \vec{u} = 2(bx + cy + dz) \begin{pmatrix} b \\ c \\ d \end{pmatrix} = 2 \begin{pmatrix} b^2 & bc & bd \\ bc & c^2 & cd \\ bd & cd & d^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

3º termo

$$\begin{aligned} 2a(\vec{u} \wedge \vec{v}) &= 2a \begin{pmatrix} i & j & k \\ b & c & d \\ x & y & z \end{pmatrix} \\ &= 2a[(-dy + cz)i + (dx - bz)j + (-cx + by)k] \\ &= 2a \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \end{aligned}$$

Somando os três termos teremos uma expressão para (3) que coloca o operador (1) em forma de operador linear

$$R_q(v) = \begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & a^2 - b^2 - c^2 + d^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (4)$$

Exemplo 1. Considerando os quatérnios unitários

$$q_x = \frac{\sqrt{3}}{2} + \frac{1}{2}i, \quad q_y = \frac{\sqrt{3}}{2} + \frac{1}{2}j \quad \text{e} \quad q_z = \frac{\sqrt{3}}{2} + \frac{1}{2}k$$

e um vetor $\vec{v} = (x, y, z)$ que pode ser associado ao quatérnio puro $v = 0 + xi + yj + zk$ temos

$$R_{q_x}(v) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

$$R_{q_y}(v) = \begin{pmatrix} \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

$$R_{q_z}(v) = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Observe que as matrizes acima são as conhecidas matrizes de rotação do vetor $\vec{v} = (x, y, z)$ um ângulo de $\frac{\pi}{3}$ em torno respectivamente dos eixos x, y e z .

Exemplo 2. O quatérnio

$$q = \frac{\sqrt{2}}{2} + \sqrt{\frac{2}{5}}j + \frac{\sqrt{10}}{10}k$$

é um quatérnio unitário e sendo $\vec{v} = (x, y, z)$ que pode ser associado ao quatérnio puro $v = 0 + xi + yj + zk$ temos

$$R_q(v) = \begin{pmatrix} 0 & -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{\sqrt{5}}{5} & \frac{4}{5} & \frac{2}{5} \\ -\frac{2\sqrt{5}}{5} & \frac{2}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

A matriz anterior também representa uma rotação do vetor $\vec{v} = (x, y, z)$ no espaço tridimensional, mas precisamos de elementos adicionais para descrever com mais precisão essa rotação. Adiante retornaremos a esse exemplo para analisar qual rotação essa matriz representa.

Para mostrar que a imagem de v pelo operador $R_q(v)$ representa uma rotação no espaço para o vetor mostraremos a seguir que esse operador preserva a norma do vetor em \mathbb{R}^3 . Para isso demonstraremos a igualdade a seguir.

Proposição 10. *Dados $q, r \in \mathbb{H}$ vale*

$$|qr| = |q||r|.$$

Prova. Sejam $q = q_0 + \vec{v}$ e $r = r_0 + \vec{u}$. Temos

$$|qr|^2 = |(q_0 + \vec{v})(r_0 + \vec{u})|^2$$

$$= |(q_0 r_0 - \langle \vec{v}, \vec{u} \rangle) + (q_0 \vec{u} + r_0 \vec{v} + \vec{v} \wedge \vec{u})|^2.$$

A partir da igualdade anterior, usando a definição de módulo de quatérnio e módulo de vetor de \mathbb{R}^3 , podemos escrever

$$|qr|^2 = (q_0 r_0 - \langle \vec{v}, \vec{u} \rangle)^2 + |(q_0 \vec{u} + r_0 \vec{v} + \vec{v} \wedge \vec{u})|^2. \quad (5)$$

Na equação (5), lembramos mais uma vez que, $\vec{u} \wedge \vec{v}$ é simultaneamente ortogonal a \vec{u} e \vec{v} , e também o é para qualquer combinação linear dos dois vetores BOULOS & CARMARGO(2007) e WINTERLE & STEINBRUCH(2000), o que nos diz que

$$|q_0 \vec{u} + r_0 \vec{v} + \vec{v} \wedge \vec{u}|^2 = |q_0 \vec{u} + r_0 \vec{v}|^2 + |\vec{v} \wedge \vec{u}|^2.$$

Retornando essa igualdade a (5) e expandindo também o quadrado da diferença chegamos a

$$|qr|^2 = q_0^2 r_0^2 - 2q_0 r_0 \langle \vec{v}, \vec{u} \rangle + \langle \vec{v}, \vec{u} \rangle^2 + |q_0 \vec{u} + r_0 \vec{v}|^2 + |\vec{v} \wedge \vec{u}|^2. \quad (6)$$

Das definições de produto interno e produto vetorial de \mathbb{R}^3 temos

$$\langle \vec{v}, \vec{u} \rangle^2 + |\vec{v} \wedge \vec{u}|^2 = |\vec{v}|^2 |\vec{u}|^2$$

e, trabalhando com a norma da combinação linear colocando na forma de produto interno, conseguimos

$$\begin{aligned} |q_0 \vec{u} + r_0 \vec{v}|^2 &= \langle q_0 \vec{u} + r_0 \vec{v}, q_0 \vec{u} + r_0 \vec{v} \rangle \\ &= q_0^2 |\vec{u}|^2 + 2q_0 r_0 \langle \vec{u}, \vec{v} \rangle + r_0^2 |\vec{v}|^2. \end{aligned}$$

Substituindo as duas ultimas igualdades em (6) teremos

$$\begin{aligned} |qr|^2 &= q_0^2 r_0^2 + q_0^2 |\vec{u}|^2 + r_0^2 |\vec{v}|^2 + |\vec{v}|^2 |\vec{u}|^2 \\ &= q_0^2 (r_0^2 + |\vec{u}|^2) + |\vec{v}|^2 (r_0^2 + |\vec{u}|^2) \\ &= (q_0^2 + |\vec{v}|^2) (r_0^2 + |\vec{u}|^2) \\ &= |q|^2 |r|^2, \end{aligned}$$

de onde extraindo a raiz temos $|qr| = |q||r|$

□

Com a proposição 10, considerando que o módulo de um quatérnio puro coincide com a norma de vetores de \mathbb{R}^3 e que o módulo do quatérnio q é igual ao módulo do quatérnio conjugado \bar{q} , podemos verificar que o operador R_q preserva a norma de vetores. De fato

$$|R_q(v)| = |qv\bar{q}| = |q||v||\bar{q}| = |v|,$$

lembrando que q é um quatérnio unitário.

O próximo teorema mostra como é a atuação geométrica do operador R_q no vetor v .

Proposição 11. Dado $q = q_0 + \vec{u} \in S^3$, então, o vetor $\lambda \vec{u}$ é invariante por R_q para qualquer $\lambda \in \mathbb{R}$.

Prova. Considere a expressão de R_q dada por (3), reescrevendo para $R_q(\lambda \vec{u})$ teremos

$$R_q(\lambda \vec{u}) = (q_0^2 - |\vec{u}|^2) \lambda \vec{u} + 2 \langle \vec{u}, \lambda \vec{u} \rangle \vec{u} + 2q_0(\vec{u} \wedge \lambda \vec{u}).$$

Na expressão acima podemos observar que $\vec{u} \wedge \lambda \vec{u} = \vec{0}$ por ser o produto vetorial de dois vetores paralelos e além disso, $2 \langle \vec{u}, \lambda \vec{u} \rangle \vec{u} = 2|\vec{u}|^2 \lambda \vec{u}$. Com essas últimas identidades chegamos a

$$R_q(\lambda \vec{u}) = (q_0^2 + |\vec{u}|^2) \lambda \vec{u}.$$

Como $q \in S^3$ temos $(q_0^2 + |\vec{u}|^2) = |q|^2 = 1$ e portanto

$$R_q(\lambda \vec{u}) = \lambda \vec{u}$$

□

Teorema 12. Dado $q = q_0 + \vec{u} \in S^3$, tal que,

$$q = q_0 + \vec{u} = \cos\left(\frac{\theta}{2}\right) + \text{sen}\left(\frac{\theta}{2}\right) \frac{\vec{u}}{|\vec{u}|}$$

com $\vec{u} \neq \vec{0}$, a parte vetorial \vec{w} de $R_q(\vec{v}) = q \vec{v} \bar{q} = 0 + \vec{w}$ é a rotação de \vec{v} de um ângulo θ em torno do eixo gerado por \vec{u} , onde $\theta = 2 \cdot \arccos(q_0) \in [0, 2\pi)$

Prova. Chamemos de V o espaço vetorial gerado pelo vetor \vec{u} e V^\perp o complemento ortogonal de V e $\vec{v} \in \mathbb{R}^3$. Vamos escrever o vetor \vec{v} em duas componentes ortogonais, o vetor \vec{a} na direção de \vec{u} , isso nos diz que, $\vec{a} = \lambda \vec{u}$, para algum $\lambda \in \mathbb{R}$ e o vetor \vec{b} perpendicular ao vetor \vec{u} , logo \vec{v} pode ser escrito como

$$\vec{v} = \vec{a} + \vec{b} = \lambda \vec{u} + \vec{b} \text{ com } \lambda \in \mathbb{R}.$$

Como $\vec{v} \in \mathbb{R}^3$, \vec{v} está associado com um quatérnio puro e, da última igualdade podemos escrever

$$R_q(\vec{v}) = R_q(\vec{a} + \vec{b}) = R_q(\vec{a}) + R_q(\vec{b}),$$

onde a última igualdade vem do fato de R ser operador linear por (4).

Usando a proposição 11, teremos que $R_q(\vec{a}) = \vec{a}$, temos

$$R_q(\vec{v}) = \vec{a} + R_q(\vec{b}). \tag{7}$$

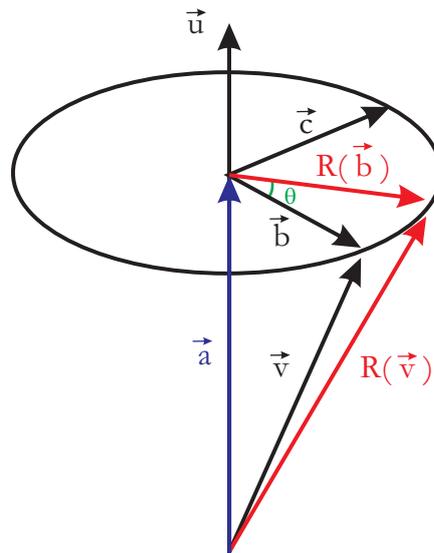


Figura 1: Ação do operador R sobre \vec{v}

Voltaremos nossa atenção para $R_q(\vec{b})$, usando para o operador a expressão (3) chegamos a

$$R_q(\vec{b}) = q \vec{b} \bar{q} = (q_0^2 - |\vec{u}|^2) \vec{b} + 2\vec{u} \langle \vec{u}, \vec{b} \rangle + 2q_0(\vec{u} \wedge \vec{b}),$$

onde podemos observar que $\vec{u} \perp \vec{b}$ o que nos dá que $\langle \vec{u}, \vec{b} \rangle = 0$ e essa expressão pode ser reescrita como

$$R_q(\vec{b}) = q \vec{b} \bar{q} = (q_0^2 - |\vec{u}|^2) \vec{b} + 2q_0(\vec{u} \wedge \vec{b}).$$

Vamos reescrever a segunda parcela da soma $2q_0(\vec{u} \wedge \vec{b}) = 2q_0|\vec{u}| \left(\frac{\vec{u}}{|\vec{u}|} \wedge \vec{b} \right)$ e chamemos $\vec{c} = \left(\frac{\vec{u}}{|\vec{u}|} \wedge \vec{b} \right)$, nesse caso, nossa igualdade se torna

$$R_q(\vec{b}) = (q_0^2 - |\vec{u}|^2) \vec{b} + 2q_0|\vec{u}| \vec{c}. \quad (8)$$

Observe que o vetor \vec{c} é elemento de V^\perp .

Como o quatérnio q é unitário, $|q| = q_0^2 + |\vec{u}|^2 = 1$, usando a identidade trigonométrica fundamental e o fato que existe um único $\theta \in [0, 2\pi)$ tal que

$$q_0 = \cos\left(\frac{\theta}{2}\right) \quad \text{e} \quad |\vec{u}| = \text{sen}\left(\frac{\theta}{2}\right)$$

Retornando essas informações para (8) temos

$$\begin{aligned} R_q(\vec{b}) &= \left(\cos^2\left(\frac{\theta}{2}\right) - \text{sen}^2\left(\frac{\theta}{2}\right) \right) \vec{b} + 2 \cos\left(\frac{\theta}{2}\right) \text{sen}\left(\frac{\theta}{2}\right) \vec{c} \\ &= \cos(\theta) \vec{b} + \text{sen}(\theta) \vec{c} \\ &= \vec{b} \cos \theta + \vec{c} \text{sen} \theta, \end{aligned}$$

que representa a rotação do vetor \vec{b} em torno do eixo gerado pelo vetor \vec{u} de um ângulo θ em V^\perp . Usando esse último resultado em conjunto com (7), podemos concluir que $R_q(\vec{v})$, como vetor, representa a rotação do vetor \vec{v} , um ângulo θ , em torno do eixo gerado por \vec{u} . \square

Exemplo 3. Para representar a rotação do vetor $\vec{v} = (1, 0, 1)$ em torno do eixo gerado por $\vec{u} = (0, 2, 1)$ de um ângulo de $\frac{\pi}{2}$, vamos construir o quatérnio segundo o enunciado do teorema 12.

Temos

$$|\vec{u}| = \sqrt{0^2 + 2^2 + 1^2} = \sqrt{5}$$

e

$$\frac{\vec{u}}{|\vec{u}|} = \left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right).$$

O que nos leva a

$$\begin{aligned} q &= q \cos\left(\frac{\theta}{2}\right) + \text{sen}\left(\frac{\theta}{2}\right) \frac{\vec{u}}{|\vec{u}|} \\ &= \cos\left(\frac{\pi}{4}\right) + \text{sen}\left(\frac{\pi}{4}\right) \left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) \\ &= \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} \left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) \\ &= \frac{\sqrt{2}}{2} + \left(0, \sqrt{\frac{2}{5}}, \frac{\sqrt{10}}{10}\right) \\ &= \frac{\sqrt{2}}{2} + \sqrt{\frac{2}{5}}j + \frac{\sqrt{10}}{10}k. \end{aligned} \tag{9}$$

A expressão (9) é o mesmo quatérnio mencionado no exemplo 2. Aqui damos o significado para a matriz do referido exemplo, essa matriz representa a rotação de um vetor em torno do eixo gerado por $(0, 2, 1)$ de um ângulo de $\frac{\pi}{2}$. O exemplo 2 nos diz que $qv\bar{q}$ pode ser calculado pelo produto de matrizes

$$\begin{pmatrix} 0 & -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{\sqrt{5}}{5} & 4 & \frac{2}{5} \\ -\frac{2\sqrt{5}}{5} & \frac{2}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{2\sqrt{5}}{5} \\ \frac{\sqrt{5}+2}{5} \\ \frac{1-2\sqrt{5}}{5} \end{pmatrix}.$$

Observe que $\vec{w} = R_q(v) = \left(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}+2}{5}, \frac{1-2\sqrt{5}}{5}\right)$ e $\vec{v} = (1, 0, 1)$ tem a mesma norma,

$$|\vec{v}| = \sqrt{1^2 + 1^2} = \sqrt{2},$$

$$\begin{aligned} |\vec{w}| &= \sqrt{\left(\frac{2\sqrt{5}}{5}\right)^2 + \left(\frac{\sqrt{5}+2}{5}\right)^2 + \left(\frac{1-2\sqrt{5}}{5}\right)^2} \\ &= \sqrt{\frac{20}{25} + \frac{5+4\sqrt{5}+4}{25} + \frac{1-4\sqrt{5}+20}{25}} \end{aligned}$$

$$= \sqrt{\frac{50}{25}} = \sqrt{2}$$

e para verificar que \vec{w} é a rotação de \vec{v} um ângulo de $\frac{\pi}{2}$ em torno do eixo gerado por \vec{u} , calculemos o ângulo entre as projeções de \vec{w} e \vec{v} no espaço ortogonal ao subespaço gerado por \vec{u} , sejam \vec{w}_0 e \vec{v}_0 essas projeções, temos

$$\begin{aligned} \vec{v}_0 &= \text{Proj}_{(0,2,1)}(1, 0, 1) - (1, 0, 1) \\ &= \frac{\langle (0, 2, 1), (1, 0, 1) \rangle}{5} (0, 2, 1) - (1, 0, 1) \\ &= \left(0, \frac{2}{5}, \frac{1}{5} \right) - (1, 0, 1) \\ &= \left(-1, \frac{2}{5}, -\frac{4}{5} \right) \end{aligned}$$

e

$$\begin{aligned} \vec{w}_0 &= \text{Proj}_{(0,2,1)} \left(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}+2}{5}, \frac{1-2\sqrt{5}}{5} \right) - \left(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}+2}{5}, \frac{1-2\sqrt{5}}{5} \right) \\ &= \frac{\left\langle (0, 2, 1), \left(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}+2}{5}, \frac{1-2\sqrt{5}}{5} \right) \right\rangle}{5} (0, 2, 1) - \left(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}+2}{5}, \frac{1-2\sqrt{5}}{5} \right) \\ &= \left(0, \frac{2}{5}, \frac{1}{5} \right) - \left(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}+2}{5}, \frac{1-2\sqrt{5}}{5} \right) \\ &= \left(-\frac{2\sqrt{5}}{5}, -\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5} \right). \end{aligned}$$

Finalmente, para concluir que o ângulo entre \vec{v}_0 e \vec{w}_0 é $\frac{\pi}{2}$, calculemos o produto escalar

$$\left\langle \left(-1, \frac{2}{5}, -\frac{4}{5} \right), \left(-\frac{2\sqrt{5}}{5}, -\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5} \right) \right\rangle = \frac{2\sqrt{5}}{5} - \frac{2\sqrt{5}}{25} - \frac{8\sqrt{5}}{25} = \frac{10\sqrt{5}}{25} - \frac{10\sqrt{5}}{25} = 0,$$

o que verifica o desejado.

Conclusões

O Teorema 12 nos diz que a rotação de um vetor \vec{v} um ângulo θ em torno de um eixo unitário \vec{u} pode ser representada pela multiplicação de quatérnios

$$q \vec{v} \bar{q}$$

onde

$$q = \cos\left(\frac{\theta}{2}\right) + \text{sen}\left(\frac{\theta}{2}\right) \vec{u}$$

que ainda, conforme a dedução em (4), quando representamos o quatérnio q na forma $q = a + bi + cj + dk$, a rotação pode ser representada pela multiplicação de matrizes

$$\begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & a^2 - b^2 - c^2 + d^2 \end{pmatrix} \vec{v}.$$

Referências

BOULOS, P; CAMARGO, I. Geometria Analítica - um tratamento vetorial. Prentice Hall, 3a edition, são Paulo, 2007.

BUTZGE, M. S.. Quatérnios e rotações no espaço.2021. 100f. Dissertação (Mestrado Profissional em Matemática em Rede Nacional - PROFMAT). Universidade Estadual do Oeste do Paraná - UNIOESTE. Cascavel - PR. Disponível

https://tede.unioeste.br/bitstream/tede/5748/5/Marta_Butzge2021.pdf. Acesso 15 de Ago. 2022.

GUZZO, S.M.; TOSTI, N. C. (2017, 14/08). Funções elementares no conjunto dos quatérnios: Abordagem por série de potências. *Jornal Eletrônico de Ensino e Pesquisa de Matemática*, v.1, n.1, p. 44-98, disponível em http://www.dma.uem.br/kit/jeepeema-1/art6_1701.pdf. Acesso em: 15 de ago. 2022.

HAMILTON, W. R.. On quaternions; or on a new system of imaginaries in algebra. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 25, n° 163 (1844): 2000.

HANSON, A. J.. Visualizing quaternions. In *ACM SIGGRAPH 2005 Courses*, 1-es. Canada: Association for Computing Machinery, New York, NY, United States, 2005. Disponível em <https://dl.acm.org/doi/proceedings/10.1145/1198555>. Acesso em: 15 de mar. 2021.

WINTERLE, P.; STEINBRUCH, A.. Geometria Analítica. Makron Books, São Paulo, 2000.

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

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

com o denominado conjunto viável $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ e } g_j(x) \leq 0\}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ para $j \in \{1, 2, \dots, p\}$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $r(t) = (r_1(t), r_2(t), \dots, r_n(t))$ uma curva diferenciável qualquer sobre S passando por p . Logo, por estar sobre S , a curva $r(t)$ satisfaz a $f(r_1(t), r_2(t), \dots, r_n(t)) = 0$, e

suponha $r(t_0) = p$. Derivando a equação $f(r_1(t), r_2(t), \dots, r_n(t)) = 0$ em relação a t , teremos

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

Calculando em t_0 , obtemos

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

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

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

Teorema 8. Se f é diferenciável em x e $\nabla f(\bar{x})^T d < 0$ então d é uma direção de descida para f , a partir de \bar{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 $G : \mathbb{R}^k \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ uma função de classe C^1 . Suponha que $G(\bar{x}, \bar{y}) = 0$ e*

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

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

- I) $\bar{x} \in W$ e $\varphi(\bar{x}) = \bar{y}$
- 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 $H_i(x) = 0, \forall i \in \{1, 2, \dots, m\}$. Usaremos esse resultado mais a frente quando definimos plano tangente e provamos o Teorema de KKT.

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

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

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

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, $G : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ dada por

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

$$\text{Como } \frac{\partial G}{\partial x} \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} \nabla H_1(x) \\ \vdots \\ \nabla H_m(x) \\ Z^T \end{pmatrix} = (MZ) \text{ é inversível e } G \begin{pmatrix} \bar{x} \\ 0 \end{pmatrix} = 0, \text{ o Teorema 9}$$

(Teorema da Função Implícita) garante a existência de uma curva diferenciável $\varphi : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n$

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

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

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

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

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

Dividindo a segunda equação $Z^T(\varphi(t) - \bar{x} - td) = 0$ por $t \neq 0$ e tomando o limite quando $t \rightarrow 0$, vamos ter $\lim_{t \rightarrow 0} Z^T \left(\frac{\varphi(t) - \varphi(0)}{t} - \frac{td}{t} \right) = Z^T(\varphi'(0) - d) = 0$. Como por hipótese $M^T d = 0$, 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. □

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 $f, g : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ funções diferenciáveis de classe C^1 . Suponha que $n > m$. Se x^0 é ponto de extremo local de f sujeito as*

restrições $g_1(x) = 0, g_2(x) = 0, g_3(x) = 0, \dots, g_m(x) = 0$, mais a condição de regularidade

$$\begin{vmatrix} \frac{\partial g_1(x^0)}{\partial x_1} & \dots & \frac{\partial g_1(x^0)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(x^0)}{\partial x_1} & \dots & \frac{\partial g_m(x^0)}{\partial x_m} \end{vmatrix} \neq 0. \quad (2)$$

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

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

Prova. Denotamos $u = (x_{m+1}, x_{m+2}, \dots, x_n)$ e $u^0 = (x_{m+1}^0, x_{m+2}^0, \dots, x_n^0)$.

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

$$\begin{aligned} (\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u), u) &\in D, \forall u \in W, \\ (\varphi_1(u^0), \varphi_2(u^0), \dots, \varphi_m(u^0), u^0) &= x^0, \end{aligned} \quad (4)$$

e

$$g_\ell(\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u), u) = 0, u \in W, 1 \leq \ell \leq m. \quad (5)$$

Novamente de (2), temos o sistema

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

que possui solução única. Isso implica em

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

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

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

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

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

Assim, temos que

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

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

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

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

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

Se $c_0 = 0$, então

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

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

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

que implica em

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

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

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

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

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

com $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ e } g(x) \leq 0\}$ tal que $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $i \in \{1, 2, \dots, p\}$.

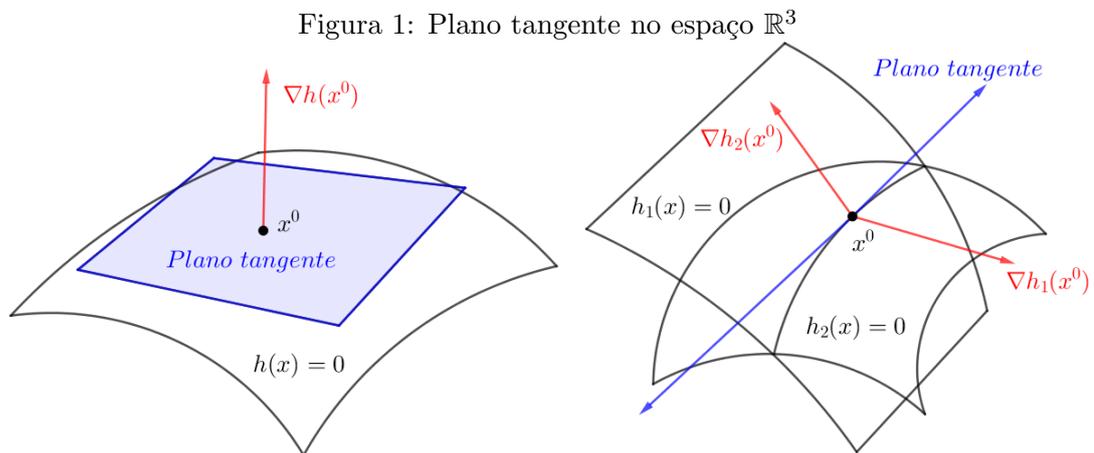
Formalizamos agora a definição da condição de qualificação de Independência Linear (LICQ). Esta condição é muito utilizada, e dentre os motivos que podem explicar esta fama estão o fato de que pode ser facilmente enunciada e verificada, e também, por ser utilizada na análise de diversos algoritmos práticos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

Definição 13. (Condição de Qualificação de Independência Linear - LICQ) Um ponto $\bar{x} \in \Omega$ é um ponto regular (LICQ) das restrições $h(x) = 0$ e $g(x) \leq 0$, se os vetores $\nabla h_i(\bar{x})$ para $i \in \{1, 2, \dots, m\}$ e $\nabla g_j(\bar{x})$ para $j \in J(\bar{x})$ são Linearmente Independentes, onde $J(\bar{x}) = \{j \mid g_j(\bar{x}) = 0\}$ é o conjunto das restrições ativas de g em \bar{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 $S = \{x \in \mathbb{R}^n \mid h_i(x) = 0, \forall i \in \{1, 2, \dots, m\}\}$, ela será dita suave quando todo $h_i(x)$ for de classe C^1 . Seja \bar{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 \bar{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 $\varphi'(t) = \frac{d\varphi(t)}{dt}$ existir, e duas vezes diferenciável se $\varphi''(t) = \frac{d^2\varphi(t)}{dt^2}$ existir. Vamos dizer que $\varphi(t)$ passe por \bar{x} quando $t = 0$, com $-\delta \leq t \leq \delta$. Logo, a derivada da curva em \bar{x} é $\varphi'(0)$. Tome agora todas as curvas diferenciáveis em S passando pelo ponto \bar{x} , a existência dessas curvas é garantida pelo Teorema 9.

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



Fonte: Autor (2022)

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

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

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

Através do Corolário 10, consideramos os vetores \bar{x} e $d \in \mathbb{R}^n$ com $\nabla h_i(\bar{x})^T d = 0$, $\forall i \in \{1, 2, \dots, m\}$ sendo que todo $\nabla h_i(\bar{x})$ é LI, ou seja, $d \in M$ e \bar{x} é um ponto regular, então vamos ter que $\varphi(0) = \bar{x}$ e $\varphi'(0) = d$, isto é, d vai ser um vetor tangente a alguma curva diferenciável $\varphi(t)$ passando por \bar{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 \bar{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 \bar{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 $h(\bar{x})$, 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 $S = \{(x, y) \in \mathbb{R}^2 \mid \gamma(x, y) = 0\} = \{(0, y) \in \mathbb{R}^2\}$. 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 $\bar{x} \in \Omega$ quando é nula ou existe uma sequência de pontos viáveis $(x^k) \subset \Omega$ tal que $x^k \rightarrow \bar{x}$ e

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \rightarrow \frac{d}{\|d\|}. \quad (12)$$

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 $\varphi : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n$ uma curva diferenciável tal que $\varphi(0) = \bar{x}$ e $\varphi'(0) = d$. Então existe uma sequência (x^k) tal que $x^k \rightarrow \bar{x}$ e

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \rightarrow \frac{d}{\|d\|}.$$

Prova. Temos que

$$\lim_{t \rightarrow 0} \frac{\varphi(t) - \bar{x}}{t} = \lim_{t \rightarrow 0} \frac{\varphi(t) - \varphi(0)}{t} = \varphi'(0) = d \neq 0.$$

Logo, $\varphi(t) \neq \bar{x}$ para todo $t \neq 0$ suficientemente pequeno. Tomamos uma sequência $(t_k) \subset \mathbb{R}_+$, com $t_k \rightarrow 0$ e definimos para todo $k \in \mathbb{N}$, $x^k = \varphi(t_k)$. Assim, $x^k \rightarrow \bar{x}$ e

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} = \frac{x^k - \bar{x}}{t_k} \frac{t_k}{\|x^k - \bar{x}\|} \rightarrow \frac{d}{\|d\|},$$

completando a prova. □

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

Lema 18. *Seja $\varphi : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n$ uma curva diferenciável tal que $\varphi(0) = \bar{x}$ 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, $x^k \rightarrow x^0$ e

$$\frac{x^k - x^0}{\|x^k - x^0\|} \rightarrow \frac{d}{\|d\|}.$$

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 Definição 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 $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ funções de classe C^1 , com $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ e } g(x) \leq 0\}$. 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 $i \in \{1, 2, \dots, p\}$.

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 $\mu_i \geq 0, \forall i \in \{1, 2, \dots, p\}$. 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 $S = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ e } g_j(x) = 0, \forall j \in J - \{k\}\}$ 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 $V = \{\nabla h_i(x^0) \text{ para todo } i \in \{1, \dots, m\} \text{ e } \nabla g_j(x^0) \text{ para } j \neq k, j = 1, \dots, p\}$ 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

$$M = \{y \in \mathbb{R}^n \mid Vy = 0\}. \quad (15)$$

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

$$\nabla g_k^T(x^0)^T d < 0. \quad (16)$$

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,

$$\nabla f(x^0)^T d \geq 0. \quad (17)$$

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

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

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, $\mu_i \geq 0, \forall i \in \{1, 2, \dots, p\}$. \square

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) $\nabla f(x^0) + \lambda^T \nabla h(x^0) + \mu^T \nabla g(x^0) = 0$, (ii) $\mu^T g(x^0) = 0$,
- (iii) $g(x^0) \leq 0$, (iv) $h(x^0) = 0$, (v) $\mu \geq 0$.

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

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

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

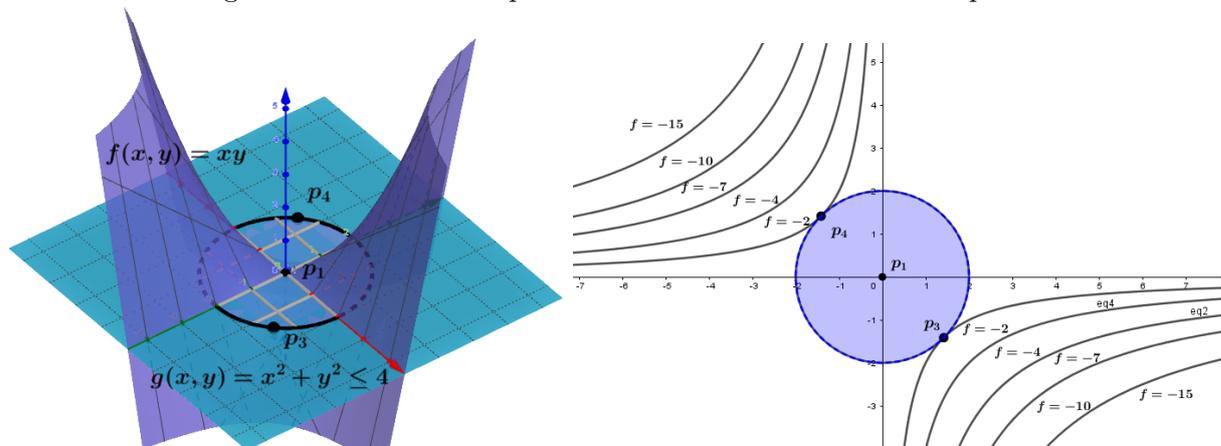
Da condição de complementariedade, $\mu g(x, y) = 0$ (iv), temos que a restrição vai estar inativa se $\mu = 0$ e ativa se $x^2 + y^2 - 4 = 0$ (v). Também devemos verificar se $\mu \geq 0$. Não devemos esquecer que no ponto estacionário x^0 , deve valer $g(x^0) \leq 0$ (vi). Assim, temos seis condições 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 $\mu = \frac{-x}{2y}$. Substituindo o valor de μ em (i), teremos a equação $x^2 - y^2 = 0$ (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

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

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, *Proceedings 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.

Créditos

Este volume foi gerado em código \LaTeX , editado no TeXstudio¹ versão 4.1.2 (64 bit, Qt5), e compilado pelo MiKTeX² versão 2.9.6000 (32 bit). Para esta edição foram utilizados os seguintes pacotes \LaTeX :

This issue was created in \LaTeX code, edited on TeXstudio¹ version 4.1.2 (64 bit, Qt5), and compiled by MiKTeX² version 2.9.6000 (32 bit). This issue uses the following \LaTeX packages:

<code>amssymb,</code>	<code>babel,</code>	<code>indentfirst,</code>
<code>amsfonts,</code>	<code>inputenc,</code>	<code>enumerate,</code>
<code>amsmath,</code>	<code>multicol,</code>	<code>fancyhdr,</code>
<code>latexsym,</code>	<code>color,</code>	<code>algorithm2e,</code>
<code>amsthm,</code>	<code>float,</code>	<code>tcolorbox,</code>
<code>makeidx,</code>	<code>graphicx,</code>	<code>hyperref.</code>

¹<https://www.texstudio.org/>

²<https://miktex.org/>

Índice de autores

Alex Augusto Nunes Machado, 21

André Luiz Zanin da Cruz, 51

Flavio Roberto Dias Silva, 37

Paulo Domingos Conejo, 51

Sandra Maria Tieppo, 9

Sandro Marcos Guzzo, 9

Simone Aparecida Miloca, 21