UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE CIÊNCIAS ECONÔMICAS

DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS

DANIEL DE SALES CASULA

TEORIA DE OTIMIZAÇÃO: TEOREMA DE KUHN-TUCKER – EXTENSÃO À PROBLEMAS MULTI-OBJETIVOS E APLICAÇÕES À ECONOMIA

Porto Alegre

2015

DANIEL DE SALES CASULA

TEORIA DE OTIMIZAÇÃO: TEOREMA DE KUHN-TUCKER – EXTENSÃO À PROBLEMAS MULTI-OBJETIVOS E APLICAÇÕES À ECONOMIA

Trabalho de conclusão submetido ao Curso de Graduação em Economia, da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Jorge Paulo de Araújo

Porto Alegre

2015

DANIEL DE SALES CASULA

TEORIA DE OTIMIZAÇÃO: TEOREMA DE KUHN-TUCKER – EXTENSÃO À PROBLEMAS MULTI-OBJETIVOS E APLICAÇÕES À ECONOMIA

Trabalho de conclusão submetido ao Curso de Graduação em Economia, da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Bacharel em Ciências Econômicas.

Aprovado em: Porto Alegre, ____ de __________ de 2015.

____

Prof. Dr. Jorge Paulo de Araújo – orientador – UFRGS

____

Prof. Dr. Fabricio Tourrucôo – UFRGS

____

Prof. Dr. Marcelo de Carvalho Griebeler – UFRGS

AGRADECIMENTOS

Em primeiro lugar, agradeço aos meus pais pelo incentivo aos estudos, apoio em todos os momentos e em todos os sentidos, pela confiança, carinho e por terem sido uma base sólida na minha vida. Agradeço de forma especial o meu melhor amigo, professor e irmão, Fabio, por estar sempre presente na minha vida tanto pessoal quanto acadêmica, por me servir de inspiração e por despertar em mim um gosto pela Matemática.

Agradeço, também, meu professor e amigo, Jorge Araújo, por ter sido um professor exemplar, por toda sua disponibilidade, sempre me auxiliando, ensinando e por sua dedicação contagiante à vida acadêmica.

Por fim, agradeço aos amigos que encontrei por esta jornada na UFRGS, principalmente os da FCE, os quais sempre estiverem presente tanto nos momentos de estudo quanto nos momentos de lazer, trocando conhecimentos e aprendizados que foram fundamentais para a conclusão do curso.

RESUMO

A Teoria de Otimização desempenha um papel muito importante em Ciências Econômicas, sendo ela aplicada em diversos problemas reais e cotidianos. Neste trabalho, procura-se apresentar os mais diversos métodos de resolução para problemas de otimização, assim como os principais teoremas e definições que os acompanham. Analisar-se-á problemas de otimização por diferenciabilidade com e sem restrições, introduzindo o método dos multiplicadores de Lagrange. Um dos objetivos do trabalho é analisar de forma precisa o Teorema de Kuhn-Tucker, tanto pela utilização do Teorema da Função Implícita quanto pelo uso do Lema de Farkas. A sequência do trabalho faz uma análise da Otimização Vetorial, via eficiência de Pareto, e, por fim, uma extensão à problemas multi-objetivos aplicados em Economia. Com algumas ressalvas, o trabalho pode ser visto como uma revisão bibliográfica referente à problemas de otimização estudados em um curso ordinário de graduação em Economia.

Palavras-chave: Otimização. Lagrange. Kuhn-Tucker. Multi-objetivo.

Classificação JEL: C02, C61.

ABSTRACT

Optimization Theory plays a very important role in Economics and it is applied in several real and daily problems. This paper aims to present the various methods for solving optimization problems, as well as the main theorems and definitions that follow them. It will be examined optimization problems by differentiability with and without constraints, introducing the method of Lagrange multipliers. One of the goals of this work is to analyze precisely the Kuhn-Tucker theorem by using the Implicit Function Theorem and also by using Farkas' Lemma. Afterwards, the work analyzes the Vector Optimization via Pareto efficiency, and finally an extension to multi-objective problems in applied Economics. With a few exceptions, this paper can be seen as a literature review related to optimization problems studied in an ordinary undergraduate degree in Economics.

Keywords: Optimization. Lagrange. Kuhn-Tucker. Multi-objective.

JEL Classification: C02, C61.

SUMÁRIO

ÍNDICE DE FIGURAS

1. CONTEXTUALIZAÇÃO

O artigo “*Nonlinear Programming*” publicado em 1950 por Harold Kuhn e Albert Tucker tem dois objetivos. O primeiro é generalizar o resultado estabelecido em Programação Linear que as soluções dos correspondentes problemas primal e dual são um ponto de sela do correspondente lagrangiano. O segundo objetivo dos autores é reduzir um problema com restrições e várias funções objetivos a um problema com uma função objetivo apenas.

Do trabalho destes autores, o que efetivamente resultou relevante em relação ao primeiro objetivo foi uma inovadora demonstração das condições necessárias num problema de otimização condicionada pelo uso de um resultado estabelecido pelo matemático húngaro Gyula Farkas (1847-1930) em 1901. Estas condições são hoje conhecidas como “condições de Kuhn-Tucker”. Em 1961, Kenneth Arrow e Alain Enthoven, mostraram que as condições de Kuhn-Tucker são suficientes, num problema de otimização condicionada, se exigimos adequadamente propriedades de quase-concavidade e quase-convexidade em relação às funções envolvidas.

Em relação ao segundo objetivo, os autores estabeleceram o primeiro resultado numa nova subdisciplina da Teoria da Otimização conhecida como Otimização Vetorial. Em 1972, ocorreu a 1ª conferência de Otimização Vetorial quando o problema é otimizar, via eficiência no sentido de Pareto, várias funções-objetivos simultaneamente. O primeiro livro de programação multi-objetivo foi publicado em 1978 (COHON, J. L. **Multiobjective Programming and Planning.** New York: Academic Press, 1978). Neste livro, o autor expõe as aplicações da Otimização Vetorial no uso de recursos hídricos, meio-ambiente, tributação, transporte urbano e relações internacionais, por exemplo.

1.1 Introdução

A Teoria de Otimização desempenha um papel muito importante em Ciências Econômicas, sendo ela aplicada em diversos problemas reais e cotidianos, tais como: retornos esperados em problemas de minimização de riscos em finanças, maximização do bem-estar para medidas governamentais, etc.

No capítulo 2, procura-se apresentar diversas ferramentas matemáticas clássicas para o auxílio da compreensão dos problemas de otimização. No capítulo 3, concomitantemente a apresentação de diferentes teoremas e definições fundamentais, analisar-se-á problemas de otimização sem restrições e, posteriormente, problemas condicionados à uma diversidade de restrições. Os problemas analisados envolverão otimização por diferenciabilidade.

No capítulo 4, tem-se como primeiro objetivo analisar o problema de maximização mais comum em Economia: maximizar uma função sujeita a restrições de desigualdade e restrições de não-negatividade, conhecida como o Teorema de Kuhn-Tucker. Apresentar-se-á a demonstração do Teorema da maneira como autores a elaboraram e, na sequência, uma demonstração completamente elementar a partir das simplificações que os próprios autores acabaram vislumbrando, como as simplificações introduzidas na demonstração do chamado “Lema de Farkas”.

Por fim, no capítulo 5, tem-se como segundo objetivo apresentar a Otimização Vetorial, assim como suas relações com o ótimo de Pareto e aplicações em problemas multi-objetivos. A parte final do capítulo está destinada a introduzir e analisar problemas multi-objetivos, buscando apresentar métodos para realizar comparações entre coisas incomensuráveis. Aplicações à Economia complementam o capítulo.

2. A MATEMÁTICA AUXILIAR PARA A COMPREENSÃO

Neste capítulo, busca-se apresentar algumas ferramentas matemáticas clássicas para o auxílio da compreensão dos problemas de otimização que serão estudados a *posteriori*.

2.1 Forma Quadrática

Depois das funções lineares, as funções quadráticas são as funções mais simples para envolver-se com as questões de otimização. Ambos os tipos de funções podem ser representados na forma matricial. Para tanto, analisar-se-á mediante definições, teoremas e exemplos das propriedades de uma forma quadrática, as quais reduzem os estudos das propriedades de uma matriz simétrica.

Muitos estudos relacionados à problemas de otimização econômicos são marcados pela presença de uma função objetivo quadrática, como por exemplo, questões relacionadas à minimização de risco em carteiras de investimentos, onde o risco é medido pela variância (quadrática) dos retornos dos investimentos.

Um modelo amplamente utilizado no estudo dos mercados financeiros é o CAPM (*Capital Asset Pricing Model*), o qual diz que o retorno esperado de qualquer ativo do mercado deve ser igual à soma do retorno sem risco e o ajuste pelo risco. Utiliza-se neste modelo o desvio padrão, o qual é a raiz quadrada da variância (mede a “dispersão” da distribuição). A partir dela utiliza-se funções quadráticas:

$$\sigma_w^2 = \sum_{s=1}^{S} \pi_s (w_s - \mu_w)^2 ,$$

sendo w uma variável aleatória assumindo os valores w_s (s = 1, 2, ..., S) com probabilidade π_s, e média μ_w. (VARIAN, 2006).

Definição 1: Uma forma quadrática em $\mathbb{R}^n$ é uma função real no formato

$$Q(x_1, \dots, x_n) = \sum_{ij=1}^{n} a_{ij} x_i x_j ,$$

onde cada termo é um monômio de grau dois.

Uma forma quadrática pode ser representada por uma matriz simétrica A, de tal forma que:

$$Q(x_1, \dots, x_n) = x^T A x.$$

Definição 2: Uma matriz é simétrica se, e somente se ela é igual a sua transposta. Ou seja, $A^T = A$. Necessariamente, a matriz é quadrada (A_{nxn}) e os elementos de sua diagonal são arbitrários (LAY, 2007). São exemplos:

$$\begin{bmatrix} 2 & 0 \\ 0 & -5 \end{bmatrix} \quad , \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

Observe que em uma matriz simétrica, a parte superior é uma “reflexão” da parte inferior, em relação à diagonal. (BOLDRINI, 1986).

2.2 Representação Matricial

Uma função linear $f(x_1, \dots x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ pode ser representada na forma matricial como:

$$f(x) = a \,.\, x = (a_1 \dots a_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Similarmente, pode-se escrever uma função quadrática geral em $\mathbb{R}^2$ como:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = (x_1 \; x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Tomando a forma quadrática geral:

$$Q(x_1, \dots, x_n) = \sum_{i \le j} a_{ij} x_i x_j$$

ou:

$$(x_1\ \ x_2\ \ \cdots\ \ x_n)\begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

De forma simplificada: $x^T A x$

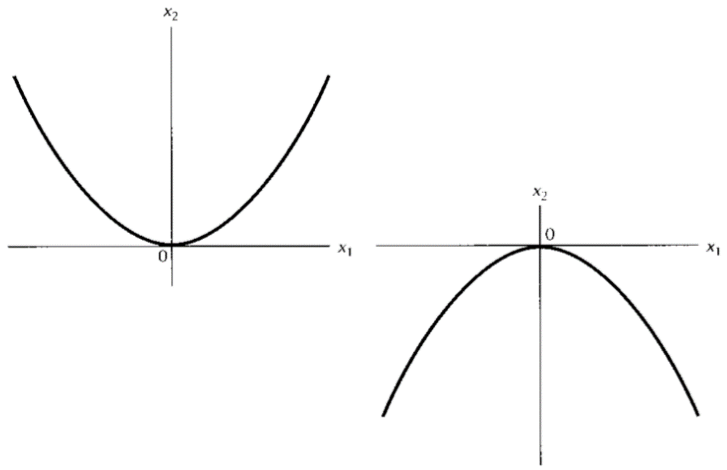
Sendo A uma matriz simétrica, a função a valores reais $Q(x) = x^T A x$ é uma forma quadrática. (SIMON; BLUME, 2008).

2.3 Formas Quadráticas Definidas

Uma forma quadrática assumirá o valor zero no ponto em que $x = 0$. Neste caso, procura-se saber se este ponto assume um máximo ou um mínimo das formas quadráticas já analisadas, ou nenhum dos dois extremos.

Analisando a forma quadrática de apenas uma variável, $y = ax^2$, tem-se que: se $a > 0$, então ax^2 será sempre ≥ 0, sendo igual a zero no ponto em que $x = 0$. Este é um exemplo de uma forma quadrática positiva onde $x = 0$ é o mínimo global. De maneira análoga, se $a < 0$, então ax^2 será sempre ≤ 0, sendo igual a zero no ponto em que $x = 0$. Tem-se aqui a forma quadrática negativa onde $x = 0$ é o máximo global.

Figura 1: Gráfico de funções quadráticas de uma variável



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 386).

Analisando agora a forma quadrática de duas variáveis.

Para $(x_1, x_2) \neq (0{,}0)$ tem-se:

$Q_1(x_1, x_2) = x_1^2 + x_2^2 > 0 \Longrightarrow Q_1$ é positiva.

$Q_2(x_1, x_2) = -x_1^2 - x_2^2 < 0 \Longrightarrow Q_2$ é negativa.

Formas quadráticas positivas ou negativas são chamadas de definidas.

No caso de: $Q_3(x_1, x_2) = x_1^2 - x_2^2$, a forma pode assumir tantos valores positivos quanto negativos, sendo, portanto, chamada de indefinida.

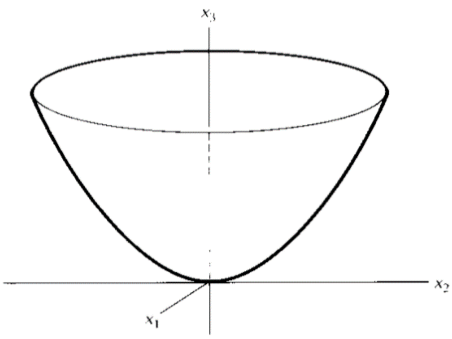
O caso intermediário se dá quando a forma quadrática assume valores ≤ 0 ou ≥ 0, sendo x não nulos. Observe:

$Q_4(x_1, x_2) = (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$, no qual a forma nunca assume valores negativos, mas pode ser zero em pontos fora da origem como (-1, 1). Essa forma é chamada de não-negativa.

Similarmente, tem-se: $Q_5(x_1, x_2) = -(x_1 + x_2)^2 = -x_1^2 - 2x_1x_2 - x_2^2$, onde a forma nunca assume valores positivos, mas pode ser zero em pontos não nulos. Forma chamada de não-positiva.

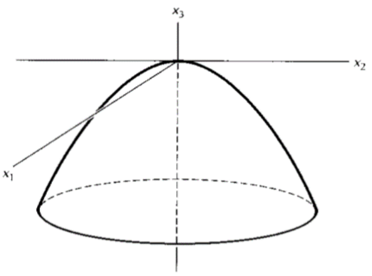
As formas não-negativa e não-positivas são chamadas de semidefinidas.

Figura 2: Gráfico da forma quadrática positiva definida



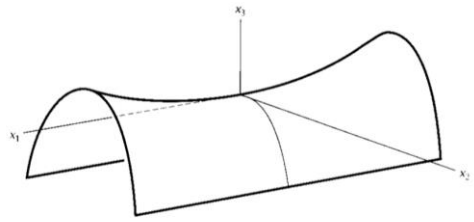
Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 387).

Figura 3: Gráfico da forma quadrática negativa definida



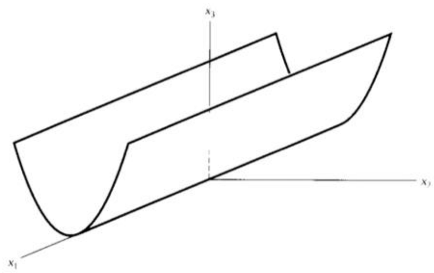
Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 388).

Figura 4: Gráfico da forma quadrática indefinida



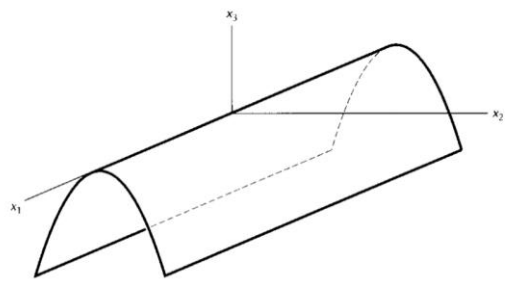
Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 388).

Figura 5: Gráfico da forma quadrática não-negativa



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 388).

Figura 6: Gráfico da forma quadrática não-positiva



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 389).

2.4 Matrizes Simétricas Definidas

Uma matriz simétrica recebe a mesma nomenclatura que a forma quadrática $Q(x) = x^T Ax$, ou seja: a matriz é dita positiva, negativa, não-positiva, etc., se a forma quadrática é positiva, negativa, não-positiva, etc., respectivamente. Além disso, assim como na forma quadrática, as matrizes simétricas positivas e negativas são chamadas definidas, e as matrizes simétricas não-negativas e não-positivas são chamadas semidefinidas.

Definição 3: Seja A_{nxn} uma matriz simétrica, então A é:

a) **positiva:** se $x^T Ax > 0, \ \forall x \neq 0$ em $\mathbb{R}^n$;
b) **não-negativa:** se $x^T Ax \geq 0$, $\forall x \neq 0$ em $\mathbb{R}^n$;
c) **negativa:** se $x^T Ax < 0$, $\forall x \neq 0$ em $\mathbb{R}^n$;
d) **não-positiva:** se $x^T Ax \leq 0$, $\forall x \neq 0$ em $\mathbb{R}^n$, e
e) **indefinida:** se $x^T Ax > 0$ para determinados x em $\mathbb{R}^n$ e $x^T Ax < 0$ para outros x em $\mathbb{R}^n$.

2.5 Menores Principais de uma Matriz

Definição 4: Seja $A = (a_{ij})$ uma matriz $n \ x \ n$. Para cada inteiro $k, 1 \leq k \leq n$, o $k - ésimo$ menor principal de A é obtido calculando o determinante da submatriz matriz principal de ordem k de A.

Definição 5: A submatriz principal de ordem k é obtida eliminando-se as últimas $n - k$ colunas e linhas de A.

$$A_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}.$$

Teorema: Seja A_{nxn} uma matriz simétrica, então:

a) A é positiva definida $\Leftrightarrow$ todos os n menores principais líderes de A são estritamente positivos. Ou seja: $|A_1| > 0, |A_2| > 0, \dots, |A_n| > 0$
b) A é negativa definida $\Leftrightarrow$ todos os n menores principais líderes de A alternam de sinal. Ou seja: $|A_1| < 0, |A_2| > 0, |A_3| < 0$, etc.

c) A é não-negativa $\Leftrightarrow$ todos os n menores principais de A são ≥ 0. Ou seja: $|A_1| \geq 0, |A_2| \geq 0, \dots, |A_n| \geq 0$

d) A é não-positiva $\Leftrightarrow$ todos os n menores principais de A de ordem ímpar são ≤ 0, e os de ordem par são ≥ 0. Ou seja: $|A_1| \leq 0, |A_2| \geq 0, |A_3| \leq 0, |A_4| \geq 0$, etc.

e) A é indefinida $\Leftrightarrow$ nenhuma das condições acima é satisfeita.

2.6 Teorema da Função Implícita

Muitas vezes nos modelos econômicos e, particularmente, nos modelos de otimização, as equações possuem variáveis exógenas e endógenas misturadas. Exemplo:

$$G(x_1, x_2, \dots, x_n, y) = 0.$$

A equação acima define a variável endógena y como uma função implícita das variáveis exógenas $x_1, x_2, \dots, x_n$.

A busca pela maximização ou minimização de uma função objetivo pode ser afetada pela variação de uma das variáveis endógenas sobre a variável exógena, por exemplo:

Exemplo: Dada uma função implícita $G(x, y) = c$ em torno do ponto (x_0, y_0). Além disso, sendo que exista uma função $y = y(x)$, C^1, e solução da equação $G(x, y) = c$, ou seja, $G\big(x, y(x)\big) = c$. Aplicando a Regra da Cadeia para derivar a equação em relação a x em x_0, tem-se:

$$\frac{\partial G}{\partial x}\big(x_0, y(x_0)\big).\frac{dx}{dx} + \frac{\partial G}{\partial y}\big(x_0, y(x_0)\big).\frac{dy}{dx}(x_0) = 0$$

ou

$$\frac{\partial G}{\partial x}(x_0, y_0) + \frac{\partial G}{\partial y}(x_0, y_0).y'(x_0) = 0.$$

Resolvendo para $y'(x_0)$,

$$y'(x_0) = -\frac{\frac{\partial G}{\partial x}(x_0, y_0)}{\frac{\partial G}{\partial y}(x_0, y_0)}.$$

Se a solução de $y(x)$ de $G(x, y) = c$ existe e é derivável, $\frac{\partial G}{\partial y}(x_0, y_0)$ deve ser não-nulo.

A mesma linha de raciocínio passa diretamente para a situação em que há várias variáveis exógenas e uma variável endógena. Neste caso, manter-se-iam constantes todas as variáveis exógenas, exceto uma. Ou seja, varia-se uma variável exógena de cada vez. Voltar-se-ia ao caso da situação bidimensional mencionada acima. Veja:

Definição 6: Teorema da Função Implícita: Seja $G(x_1, \dots, x_k, y)$ uma função C^1 numa bola em torno do ponto $(x_1^*, \dots, x_k^*, y^*)$. Suponha também que $(x_1^*, \dots, x_k^*, y^*)$ satisfaz

$$G(x_1^*, \dots, x_k^*, y^*) = c$$

e

$$\frac{\partial G}{\partial y}(x_1^*, \dots, x_k^*, y^*) \neq 0,$$

então existe uma função C^1, $y = (x_1, \dots, x_k)$, definida numa bola aberta B em torno de $(x_1^*, \dots, x_k^*)$ tal que:

a) $G(x_1, \dots, x_k, y(x_1, \dots, x_k)) = c$ para qualquer $(x_1, \dots, x_k) \in B$;
b) $y^* = y(x_1^*, \dots, x_k^*)$;
c) para cada índice i,

$$\frac{\partial y}{\partial x_i}(x_1^*, \dots, x_k^*) = -\frac{\frac{\partial G}{\partial x_i}(x_1^*, \dots, x_k^*, y^*)}{\frac{\partial G}{\partial y}(x_1^*, \dots, x_k^*, y^*)}.$$

(SIMON; BLUME, 2008).

Pode-se, também, usar diferenciais para aproximar a variação de um conjunto de nível $G(x, y) = c$ em torno do ponto (x_0, y_0). Seja:

$$G(x_0 + \Delta x, y_0 + \Delta y) - G(x_0, y_0) \approx \frac{\partial G}{\partial x}(x_0, y_0)\Delta x + \frac{\partial G}{\partial y}(x_0, y_0)\Delta y.$$

Ao encontrar a combinação de movimentos lineares Δx e Δy a partir do ponto (x_0, y_0) que levam a nenhuma variação de G, encontra-se a direção da reta tangente ao conjunto de nível $G(x, y)$ em (x_0, y_0). Para tanto, basta tomar o lado esquerdo da equação como zero, isto é, $\Delta G = 0$. Logo:

$$0 = \frac{\partial G}{\partial x}(x_0, y_0)\Delta x + \frac{\partial G}{\partial y}(x_0, y_0)\Delta y.$$

Resolvendo para $\Delta y/\Delta x$:

$$\frac{\Delta y}{\Delta x} = -\frac{\frac{\partial G}{\partial x}(x_0, y_0)}{\frac{\partial G}{\partial y}(x_0, y_0)}.$$

O Teorema da Função Implícita é muito aplicado na Microeconomia. Uma aplicação natural deste teorema ocorre quando se estuda uma curva de indiferença de uma função utilidade ou a curva de isoquanta de uma função de produção.

Perceba que a inclinação de uma curva de indiferença de uma função utilidade no ponto (x_0, y_0) é dada pela taxa marginal de substituição (TMgS) de U em (x_0, y_0), a qual mede, marginalmente, quanto do bem x o consumidor está disposto a abrir mão para obter uma unidade do bem y e manter o mesmo nível de satisfação. Resultado análogo pode ser visto nas isoquantas de uma função de produção, $Q = F(K, L)$, ao passo que a inclinação da isoquanta em (K_0, L_0), denominada Taxa Marginal de Substituição Técnica (TMgST), mede, marginalmente, a quantidade necessária de insumo para compensar a perda de outro mantendo-se o mesmo nível de produção.

3. OTIMIZAÇÃO POR DIFERENCIABILIDADE

Neste capítulo procura-se analisar, primeiramente, as condições de otimização não condicionada utilizando as condições de primeira e de segunda ordem sobre as derivadas as quais caracterizam os extremos de uma função diferenciável qualquer. Na sequência, analisar-se-á as condições de otimização para funções semelhantes, porém condicionadas a restrições.

3.1 Otimização Não Condicionada

Neste subcapítulo procura-se analisar as condições de primeira e segunda ordem para o problema matemático de otimizar funções diferenciáveis sendo estas não condicionadas a restrições.

3.1.1 Máximos e Mínimos para Funções de Duas Variáveis

A análise envolvendo funções[1] de apenas duas variáveis é muito comum em Ciências Econômicas. Um exemplo clássico é função Cobb-Douglas, a qual utiliza apenas capital (K) e trabalho (L) como variáveis dependentes. A *posteriori* ampliar-se-á a análise para as funções de várias variáveis.

Definição 7: Uma função de duas variáveis possui um máximo relativo no ponto (x_0, y_0) se há um círculo centrado em (x_0, y_0), de modo que $f(x_0, y_0) \geq f(x, y)$ para todos pontos (x, y) do domínio de f que estão dentro do círculo, e diz-se que f tem uma máximo absoluto em (x_0, y_0) se $f(x_0, y_0) \geq f(x, y)$ para todos os pontos (x, y) do domínio de f. (ANTON; BIVENS; DAVIS, 2007).

A mesma definição pode ser dada para os pontos de mínimo relativo e mínimo absoluto com a ressalva da troca de "≥" por "≤". Isto é: $f(x_0, y_0) \leq f(x, y)$ em ambos os casos.

Da mesma forma que uma função g de uma variável possui seu extremo relativo no ponto em que g é diferenciável, implicando que $g'(x_0) = 0$, análogo resultado deve ser obtido para uma função com duas variáveis. Logo, para que

[1] Quando não mencionado no decorrer do trabalho, o leitor deve observar que neste capítulo, assim como no capítulo 4, todas as funções mencionadas serão funções diferenciáveis.

$f(x, y)$ tenha um extremo relativo em (x_0, y_0) é necessário que as derivadas parciais de f existam em (x_0, y_0) e que os traços de uma determinada superfície $z = f(x, y)$ sobre os planos $x = x_0$ e $y = y_0$ tenham retas tangentes horizontais em (x_0, y_0). Portanto:

$$f_x(x_0, y_0) = 0 \quad \text{e} \quad f_y(x_0, y_0) = 0.$$

Definição 8: Dada uma função $f(x, y)$ e um ponto (x_0, y_0), se $f_x(x_0, y_0) = 0$ e $f_y(x_0, y_0) = 0$ ou se uma ou ambas derivadas parciais não existirem neste ponto, (x_0, y_0) é denominado ponto crítico da função.

A partir de uma função $y = f(x_1, x_2)$, busca-se descobrir quais condições devem ser impostas sobre a função que garantem que a derivada segunda da função é negativa para qualquer movimento a partir do ponto crítico. Dado que:

$$dy = f_1 dx_1 + f_2 dx_2$$

$$d^2y = (f_{11}dx_1 + f_{12}dx_2)dx_1 + (f_{21}dx_1 + f_{22}dx_2)dx_2$$

ou

$$d^2y = f_{11}dx_1^2 + f_{12}dx_2dx_1 + f_{21}dx_1dx_2 + f_{22}dx_2^2.$$

Aplicando o Teorema de Young[2], $(f_{12} = f_{21})$, tem-se:

$$d^2y = f_{11}dx_1^2 + 2f_{12}dx_1dx_2 + f_{22}dx_2^2.$$

A equação acima é representada na forma quadrática em dx_1 e dx_2. É natural combinar as derivadas parciais em uma matriz $n\,x\,n$ denominada matriz hessiana de F ou, simplesmente, a hessiana de F.

$$\text{Matriz Hessiana} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}.$$

(NICHOLSON; SNYDER, 2010).

[2] Teorema de Young: dada uma função $y = f(x_1, \dots, x_n)$, C^2, numa região aberta K de $\mathbb{R}^n$, então para cada x de K e para cada par de índices i e j, tem-se: $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$.

Teorema: Teste da Derivada Segunda: Seja $f(x_1, x_2)$ uma função com derivadas parciais de segunda ordem contínuas em algum círculo centrado em um ponto crítico (x_1, x_2) e seja:

$$D(x_1, x_2) = \frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} \cdot \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} - \left(\frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2}\right)^2,$$

então:

a) Se $D(x_1, x_2) > 0$ e $\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} > 0$, então (x_1, x_2) é ponto de mínimo relativo de f.
b) Se $D(x_1, x_2) > 0$ e $\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} < 0$, então (x_1, x_2) é ponto de máximo relativo de f.
c) Se $D(x_1, x_2) < 0$, então (x_1, x_2) é ponto de sela de f.
d) Se $D(x_1, x_2) = 0$, então o teste é inconclusivo.

(ANTON; BIVENS; DAVIS, 2007. *Adaptado*).

3.1.2 Máximos e Mínimos para Funções de Várias Variáveis

Definição 9: Seja $F: U \to \mathbb{R}^1$ uma função real de n variáveis, cujo domínio é um subconjunto de $\mathbb{R}^n$, tem-se:

a) Um ponto $x^* \in U$ é uma máximo de F em U se $F(x^*) \geq F(x)$, para cada $x \in U$.
b) $x^* \in U$ é um máximo estrito se x^* é um máximo e $F(x^*) \geq F(x)$, para cada $x \neq x^*$ em U.
c) $x^* \in U$ é um máximo local (ou relativo) de F se existe uma bola $B_r(x^*)$ em torno de x^* tal que $(x^*) \geq F(x)$, para cada $x \in B_r(x^*) \cap U$.
d) $x^* \in U$ é um máximo local estrito de F se existe uma bola $B_r(x^*)$ em torno de x^* tal que $(x^*) \geq F(x)$, para cada $x \neq x^*$ em $B_r(x^*) \cap U$.

(SIMON; BLUME, 2008).

Logo, x^* será um ponto de máximo local caso não existam pontos próximos em que F assume um valor maior. Ponto de máximo global será um ponto em que x^* é um máximo para F em todo o seu domínio U.

3.1.3 Condições de Primeira Ordem – CPO

Para que um determinado ponto x^* seja um ponto de máximo ou de mínimo da função F de n variáveis, as n derivadas parciais devem satisfazer $\frac{\partial F}{\partial x_i}(x^*) = 0$. Ou seja, x^* é ponto crítico de F.

Teorema: Seja $F: U \rightarrow \mathbb{R}^1$ uma função continuamente diferenciável (C^1) definida num subconjunto U de $\mathbb{R}^n$. Se x^* é um máximo ou mínimo local de F em U e se x^* é um ponto interior de U, então:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \text{ para } i = 1, \ldots, n.$$

(SIMON; BLUME, 2008).

3.1.4 Condições de Segunda Ordem – CSO

Sabe-se que para um ponto de máximo, a condição suficiente é que a derivada segunda de uma função f em $\mathbb{R}^1$ no ponto crítico seja negativa. Ou seja, sendo x^* um ponto crítico de f, $f''(x^*) < 0$. Correspondentemente, para uma função F de n variáveis, é necessário que $D^2F(x^*) < 0$, como uma matriz simétrica.

De maneira análoga, porém inversa, para um ponto de mínimo, $f''(x^*) > 0$ no caso de uma variável e $D^2F(x^*) > 0$ para um ponto crítico x^* n-dimensional.

Definição 10: Matriz Hessiana: Uma função de n variáveis tem n^2 derivadas parciais de segunda ordem. Arrumando as derivadas parciais numa matriz $n \; x \; n$, cuja (i,j)-ésima entrada é $\left(\frac{\partial^2 f}{\partial x_j \partial x_i}\right)(x^*)$, obtém-se a matriz hessiana de f, denotada por:

$$D^2f(x) \; ou \; D^2f_x \equiv \begin{bmatrix} \frac{d^2f}{\partial x_1^2} & \frac{d^2f}{\partial x_2 \partial x_1} & \cdots & \frac{d^2f}{\partial x_n \partial x_1} \\ \frac{d^2f}{\partial x_1 \partial x_2} & \frac{d^2f}{\partial x_2^2} & \cdots & \frac{d^2f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2f}{\partial x_1 \partial x_n} & \frac{d^2f}{\partial x_2 \partial x_n} & \cdots & \frac{d^2f}{\partial x_n^2} \end{bmatrix}.$$

(SIMON; BLUME, 2008).

Teorema: Seja $F: U \rightarrow \mathbb{R}^1$ uma função duas vezes continuamente diferenciável (C^2) cujo domínio é um conjunto aberto U em $\mathbb{R}^n$. Suponha que x^* é um ponto crítico de F, isto é:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \text{ para } i = 1, \dots, n.$$

Se tal condição é satisfeita e:

a) se a hessiana $D^2F(x^*)$ é uma matriz simétrica negativa. Ou seja: os n menores principais líderes de $D^2F(x^*)$ alternam de sinal em x^*:

$$\left|F_{x_1x_1}\right| < 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} \\ F_{x_1x_2} & F_{x_2x_2} \end{vmatrix} > 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} & F_{x_3x_1} \\ F_{x_1x_2} & F_{x_2x_2} & F_{x_3x_2} \\ F_{x_1x_3} & F_{x_2x_3} & F_{x_3x_3} \end{vmatrix} < 0,$$

então, x^* é um máximo local estrito de F.

b) se a hessiana $D^2F(x^*)$ é uma matriz simétrica positiva. Ou seja: os n menores principais líderes de $D^2F(x^*)$ são todos positivos em x^*:

$$\left|F_{x_1x_1}\right| > 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} \\ F_{x_1x_2} & F_{x_2x_2} \end{vmatrix} > 0, \quad \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} & F_{x_3x_1} \\ F_{x_1x_2} & F_{x_2x_2} & F_{x_3x_2} \\ F_{x_1x_3} & F_{x_2x_3} & F_{x_3x_3} \end{vmatrix} > 0,$$

então, x^* é um mínimo local estrito de F.

c) se os n menores principais líderes de $D^2F(x^*)$ violam o padrão de sinais das hipóteses acima, então x^* é um ponto sela de F. Ou seja, não é nem máximo nem mínimo local de F.

As condições mencionadas acima são chamadas de condições de segunda ordem suficiente para um máximo ou mínimo de uma função. Ao substituir as condições de definição positiva ou negativa sobre a hessiana de F para a exigência de que a hessiana seja não-negativa (para um mínimo local) e não-positiva (para um máximo local) obtém-se as denominadas condições necessárias. Perceba que as condições necessárias são mais fracas que as condições de segunda ordem suficiente. Veja um exemplo:

Exemplo: Suponha as funções $f_1(x) = x^3$ e $f_2(x) = -x^3$

Observe que para ambas as funções, $f_i'(0) = 0$ e $f_i''(0) = 0$, satisfazendo, assim, as exigências da condição necessária, a saber, que $f''(x^*) \leq 0$ para um máximo local e $f''(x^*) \geq 0$ para um mínimo local. Entretanto, para $x \geq 0$, f_1 tem um mínimo global estrito em $x = 0$ e f_2 tem um máximo global estrito em $x = 0$.

Exemplo: Suponha a função $F(x_1, x_2) = x_1^3 - x_2^2 + x_1x_2 + 5$

Para encontrar os pontos críticos da função, calcula-se as derivadas parciais de primeira ordem e igualem-nas à zero:

$$\frac{\partial F}{\partial x_1} = 3x_1^2 + x_2 = 0, \qquad \frac{\partial F}{\partial x_2} = -2x_2 + x_1 = 0.$$

Então: $2x_2 = x_1$ e $3x_1^2 + \frac{x_1}{2} = 0 \Rightarrow x_1\left(3x_1 + \frac{1}{2}\right) = 0$

Logo: $\begin{cases} x_1 = 0 \\ x_1 = \frac{-1}{6} \end{cases} \Rightarrow \begin{cases} x_2 = 0 \\ x_2 = \frac{-1}{12} \end{cases}$

Vê-se aqui que os únicos candidatos a máximo e mínimo local de F são os pontos (0,0) e $\left(\frac{-1}{6}, \frac{-1}{12}\right)$. Obtendo a hessiana de F:

$$D^2F(x) = \begin{vmatrix} F_{x_1x_1} & F_{x_2x_1} \\ F_{x_1x_2} & F_{x_2x_2} \end{vmatrix} = \begin{vmatrix} 6x_1 & 1 \\ 1 & -2 \end{vmatrix}.$$

O menor principal líder de primeira ordem é $F_{x_1x_1} = 6x_1$, e o de segunda ordem é $D^2F(x) = -12x_1 - 1$. Em (0,0), tem-se que o menor principal de segunda ordem é negativo, portanto, este ponto é uma sela de F. Em $\left(\frac{-1}{6}, \frac{-1}{12}\right)$, tem-se que o menor principal de primeira ordem é negativo $\left(F_{x_1x_1} = -1\right)$, e o menor principal de segunda ordem é positivo $\left(D^2F\left(\frac{-1}{6}, \frac{-1}{12}\right) = 1\right)$. Portanto, a hessiana é uma matriz simétrica negativa e $\left(\frac{-1}{6}, \frac{-1}{12}\right)$ é um máximo local estrito de F. Perceba que $\left(\frac{-1}{6}, \frac{-1}{12}\right)$ não é um máximo global, uma vez que no ponto $(n, 0)$, por exemplo, tem-se $F(n, 0) = n^3$, que vai para ∞ quando $n \to \infty$.

As condições suficientes de primeira e segunda ordem abordadas nos tópicos anteriores são satisfatórias para encontrar os pontos de máximos e mínimos locais de uma função diferenciável qualquer cujo domínio seja um conjunto aberto em $\mathbb{R}^n$. Entretanto, deve ficar claro que essas condições, assim como ilustrado no exemplo anterior, não são condições suficientes para determinar os pontos de máximos e mínimos globais de uma função real em $\mathbb{R}^n$.

Analisar-se-á as condições de primeira e segunda ordem necessárias para obter os extremos globais de uma função em $\mathbb{R}^n$ a partir da concavidade ou convexidade da função relacionada.

3.1.5 Funções Côncavas e Convexas

Definição 11: Uma função real f definida num subconjunto convexo U de $\mathbb{R}^n$, para quaisquer x, y em U, é:

a) **Estritamente convexa** $\Leftrightarrow f\big((1-\theta)x+\theta y\big) < (1-\theta)f(x)+\theta f(y)$, para $0 < \theta < 1$;

b) **Convexa** $\Leftrightarrow f\big((1-\theta)x+\theta y\big) \leq (1-\theta)f(x)+\theta f(y)$, para $0 \leq \theta \leq 1$;

c) **Estritamente côncava** $\Leftrightarrow f\big((1-\theta)x+\theta y\big) > (1-\theta)f(x)+\theta f(y)$, para $0 < \theta < 1$;

d) **Côncava** $\Leftrightarrow f\big((1-\theta)x+\theta y\big) \geq (1-\theta)f(x)+\theta f(y)$, para $0 \leq \theta \leq 1$.

Teorema: Seja uma função C^1 num subconjunto convexo U de $\mathbb{R}^n$. Então, para quaisquer x, y em U:

a) f é côncava se, e somente se: $f(y)-f(x) \leq Df(x)(y-x)$.

Ou seja:

$$f(y)-f(x) \leq \frac{\partial f}{\partial x_1}(x)(y_1-x_1)+\cdots+\frac{\partial f}{\partial x_n}(x)(y_n-x_n).$$

b) De modo análogo, f é convexa se, e somente se:

$$f(y)-f(x) \geq Df(x)(y-x).$$

Perceba que para uma função côncava, sendo $x_0 \in U$, então:

$Df(x_0)(y - x_0) \leq 0$ implica que $f(y) \leq f(x_0)$.

Se: $Df(x_0)(y - x_0) \leq 0, \forall y \in U \Longrightarrow x_0$ é um máximo global de f.

Pela forma geométrica, pode-se substituir a matriz das derivadas pelo vetor gradiente $\nabla f(x_0)$.

Se $\nabla f(x_0).(y - x_0) \leq 0$[3], ou seja, se o vetor $(y - x_0)$ forma um ângulo obtuso com o vetor gradiente, então $f(y) \leq f(x_0)$.

Em suma, se f é côncava e $Df(x_0) = 0$, então ter-se-á $f(y) \leq f(x_0), \forall y \in U$. Ou seja, qualquer ponto crítico de uma função côncava é automaticamente um ponto de máximo global.

(SIMON; BLUME, 2008).

Teorema: Seja uma função C^1 num subconjunto convexo U de $\mathbb{R}^n$. Se f é côncava e x_0, tal que $x_0 \in U,$ satisfaz $Df(x_0)(y - x_0) \leq 0, \forall y \in U$, então x_0 é um máximo global de f em U. Em contrapartida, se f é convexa e x_0, tal que $x_0 \in U,$ satisfaz $Df(x_0)(y - x_0) \geq 0, \forall y \in U$, então x_0 é um mínimo global de f em U.

Dentre algumas das propriedades das funções côncavas que as tornaram tão importantes no estudo das Ciências Econômicas, destaca-se: a primeira delas, como já mencionado, é que seus pontos críticos são automaticamente máximos globais. Outra propriedade continuamente presente em Economia é o fato de que os conjuntos de nível de uma função côncava têm o formato ideal para a teoria do consumidor e para a teoria da firma (produção).

[3] Ângulos em $\mathbb{R}^2$ e $\mathbb{R}^3$: se u e v são vetores não-nulos em $\mathbb{R}^2$ ou $\mathbb{R}^3$, existe uma conexão entre o produto interno e o ângulo ϑ entre os dois segmentos de reta da origem aos pontos identificados com u e v. A formula é: $u \,.\, v = \|u\| \; \|v\| \cos \vartheta$ (LAY, D. Álgebra Linear e suas aplicações. 2007).

3.2 Otimização Condicionada

O subcapítulo presente estará centrado em apresentar o problema matemático de maximizar ou minimizar uma função diferenciável de várias variáveis ao passo que a função está sujeita à algumas restrições. Normalmente, este problema é apresentado como segue:

Maximizar/minimizar $f(x_1, \ldots x_n)$, para $(x_1, \ldots x_n) \in \mathbb{R}^n$

Sujeito à $g_1(x_1, \ldots, x_n) \leq b_1$,..., $g_k(x_1, \ldots, x_n) \leq b_k$

$h_1(x_1, \ldots, x_n) \geq c_1$,..., $h_l(x_1, \ldots, x_n) \geq c_l$

$r_1(x_1, \ldots, x_n) = a_1$,..., $r_m(x_1, \ldots, x_n) = a_m$

Definição 12: A função $f(x_1, \ldots x_n)$ é chamada de função objetivo, enquanto que as funções $g_1(x_1, \ldots, x_n)$, ..., $g_k(x_1, \ldots, x_n)$, $h_1(x_1, \ldots, x_n)$, ..., $h_l(x_1, \ldots, x_n)$, $r_1(x_1, \ldots, x_n)$, ..., $r_m(x_1, \ldots, x_n)$ são denominadas funções restrição. Além disso, deve-se ressaltar que as funções g_i e h_j definem as restrições de desigualdade, enquanto que as funções r_p definem as restrições de igualdade.

3.2.1 Função Lagrangiana

Observe que, conforme visto nos tópicos anteriores, para resolver problemas de maximização ou minimização de uma função não condicionada, basta encontrar seus pontos críticos igualando suas respectivas derivadas parciais de primeira ordem a zero. Entretanto, é muito comum em problemas de otimização que a função objetivo esteja condicionada a funções de restrições.

Analisar-se-á um poderoso método para maximizar ou minimizar uma função sujeita a restrições sobre as variáveis desenvolvido pelo matemático e astrônomo franco-italiano Joseph Louis Lagrange (1736-1813).

A análise aqui presente iniciará através de um problema de maximização condicionada mais simples, envolvendo apenas uma função com duas variáveis e uma única restrição de igualdade. Em Microeconomia, um problema clássico é maximizar a função utilidade do consumidor condicionada aos preços dos produtos e a renda do consumidor. Seja:

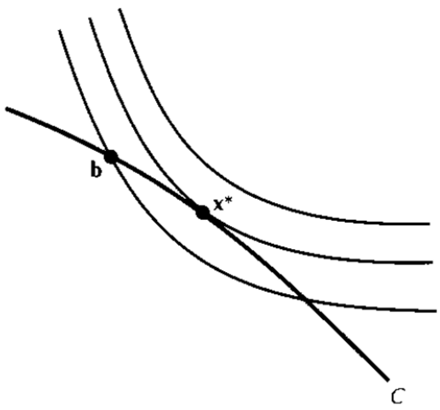
Maximizar $f(x_1, x_2)$

Sujeita a $p_1x_1 + p_2x_2 = I$

Para início de análise, ignora-se aqui a possibilidade de o consumidor não gastar toda sua renda (I) e a exigência de que os produtos $(x_1 \ e \ x_2)$ sejam não-negativos.

Da Microeconomia, sabe-se que: geometricamente, busca-se encontrar a curva de nível da função de maior valor que tangencia o conjunto restrição C. Esta situação ocorre no ponto x^* da figura 7 abaixo:

Figura 7: Ponto de tangencia entre a curva de nível e o conjunto restrição C



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 423).

Conforme ilustrado na figura, tem-se que quando a curva de nível de f tangencia o conjunto-restrição C, $\{h(x_1, x_2) = c\}$, no ponto de máximo condicionado x^*, a inclinação de ambas as “curvas” devem ser iguais. Aplicando o Teorema da Função Implícita, tem-se:

$$\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial f}{\partial x_2}(x^*)} = \frac{\frac{\partial h}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)},$$

arrumando:

$$\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_1}(x^*)} = \frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)}.$$

Para evitar denominadores possivelmente nulos, faça μ o valor comum dos quocientes acima. Logo:

$$\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_1}(x^*)} = \frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)} = \mu$$

ou

$$\frac{\partial f}{\partial x_1}(x^*) - \mu \frac{\partial h}{\partial x_1}(x^*) = 0,$$

$$\frac{\partial f}{\partial x_2}(x^*) - \mu \frac{\partial h}{\partial x_2}(x^*) = 0.$$

Tem-se aqui três incógnitas (x_1, x_2, μ). Para resolver o problema, utilizar-se-á as duas equações acima e a equação de restrição $h(x_1, x_2) = c$. Logo, ter-se-á um sistema com três incógnitas e três equações:

$$\frac{\partial f}{\partial x_1}(x^*) - \mu \frac{\partial h}{\partial x_1}(x^*) = 0,$$

$$\frac{\partial f}{\partial x_2}(x^*) - \mu \frac{\partial h}{\partial x_2}(x^*) = 0,$$

$$h(x_1, x_2) - c = 0.$$

Para resolver o sistema, construa a função conhecida como a função lagrangiana ou, simplesmente, o lagrangiano:

$$L(x_1, x_2, \mu) \equiv f(x_1, x_2) - \mu(h(x_1, x_2) - c).$$

Na sequência, encontrar-se-á os pontos críticos do lagrangiano calculando $\frac{\partial L}{\partial x_1}$, $\frac{\partial L}{\partial x_2}$ e $\frac{\partial L}{\partial \mu}$ e igualando a zero.

Observe que a introdução dessa nova variável μ, denominada multiplicador de Lagrange, reduzirá um problema de otimização com restrições em duas variáveis para um problema de otimização de três variáveis porém *sem* *restrições*. A penalidade deste método se resume à introdução de uma nova variável um tanto artificial. Porém, esta nova variável carrega significado econômico – entendida como a utilidade marginal da renda.

Deve-se mencionar que esta redução não teria êxito caso $\frac{\partial h}{\partial x_1}$ e $\frac{\partial h}{\partial x_2}$ fossem zero no máximo condicionado. Portanto, deve-se criar aqui a hipótese de que a parcial $\frac{\partial h}{\partial x_1}$ ou a parcial $\frac{\partial h}{\partial x_2}$ seja não-nula em x^* ou que ambas sejam nulas. Denomina-se esta imposição (fraca) no conjunto-restrição de qualificação de restrição, a qual pode ser entendida como:

$$\left(\frac{\partial h}{\partial x_1}(x^*), \frac{\partial h}{\partial x_2}(x^*)\right) \neq (0,0).$$

Para a melhor compreensão, veja um exemplo:

Exemplo: Sejam f e h funções C^1 de duas variáveis e suponha $x^* = (x_1^*, x_2^*)$.

Maximize $\quad f(x_1, x_2) = x_1^2 x_2$

Sujeita a $\quad h(x_1, x_2) \equiv 2x_1^2 + x_2^2 = 3$

Antes de escrever a função lagrangiana deve-se analisar se a qualificação de restrição é satisfeita. Para tanto, tome os pontos críticos de h:

$$\frac{\partial h}{\partial x_1}(x_1, x_2) = 4x_1 = 0 \iff x_1 = 0.$$

$$\frac{\partial h}{\partial x_2}(x_1, x_2) = 2x_2 = 0 \iff x_2 = 0.$$

Observe que o único ponto crítico da função ocorre em $(x_1, x_2) = (0,0)$, porém este ponto não satisfaz a restrição. Pode-se concluir, portanto, que a qualificação de restrição deste problema está satisfeita. Montando agora o lagrangiano:

$$L(x_1, x_2, \lambda) = x_1^2 x_2 - \lambda(2x_1^2 + x_2^2 - 3).$$

Calculando as derivadas parciais e igualando a zero, obtém-se:

$$\frac{\partial L}{\partial x_1} = 2x_1x_2 - 4\lambda x_1 = 2x_1(x_2 - 2\lambda) = 0$$

$$\frac{\partial L}{\partial x_2} = x_1^2 - 2\lambda x_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = -2x_1^2 - x_2^2 + 3 = 0$$

Para resolver o sistema, suponha diferentes casos:

Da primeira equação do sistema: $x_1 = 0$ ou $x_2 = 2\lambda$.

1º caso: da terceira equação: se $x_1 = 0 \Longrightarrow x_2 = \pm\sqrt{3}$ e, pela segunda equação, $\lambda = 0$. Portanto, tem-se $(x_1, x_2, \lambda) = (0, \sqrt{3}, 0)$ e $(0, -\sqrt{3}, 0)$.

2º caso: pela primeira equação, se $x_1 \neq 0 \Longrightarrow \lambda = x_2/2$. Substituindo na segunda equação obtém-se $x_1^2 = x_2^2$. Substituindo na terceira equação ter-se-á $3x_1^2 = 3$ ou $x_1 = \pm 1$, o que implica que $x_2 = \pm 1$. Pela primeira equação, perceba que: se $x_2 = 1 \Longrightarrow \lambda = 0{,}5$, se $x_2 = -1 \Longrightarrow \lambda = -0{,}5$. Ter-se-á, portanto, quatro soluções para (x_1, x_2, λ):

$(1{,}1{,}0{,}5)$, $(-1, -1, -0{,}5)$, $(1, -1, -0{,}5)$ e $(-1{,}1{,}05)$.

Já obtidos todos os candidatos, verifique na função qual ponto fornece o maior valor:

$$f(0, \sqrt{3}) = f(0, -\sqrt{3}) = 0$$

$$f(1{,}1) = f(-1{,}1) = 1$$

$$f(1, -1) = f(-1, -1) = -1$$

Observe que o máximo ocorre em $(1{,}1)$ e em $(-1{,}1)$. Pode-se observar também que $(1, -1)$ e $(-1, -1)$ minimizam a função condicionada.

Perceba do exemplo acima que o lagrangiano apenas “diz” que onde existe um mínimo ou um máximo, uma isoquanta da função objetivo deve tangenciar a restrição. Logo, as condições de primeira ordem do lagrangiano são necessárias mas não suficientes, pois pode existir a tangência sem que o ponto seja mínimo ou máximo.

Observe que os resultados anteriores surgiram da ideia de que as curvas de nível da função são tangentes ao conjunto restrição no ponto de máximo condicionado (x^*) e, portanto, possuem a mesma inclinação.

3.2.2 Derivadas Direcionais e Gradientes

Uma outra forma de apresentar os mesmos resultados obtidos no tópico anterior baseia-se nos vetores gradientes.

Antes de iniciar a próxima análise, observe que:

> “[...] as derivadas parciais de uma função dão as taxas de variação instantâneas dessa função nas direções paralelas aos eixos coordenados. As derivadas direcionais nos permitem calcular as taxas de variação de uma função em relação a qualquer direção”. (ANTON; BIVENS; DAVIS, 2007, p. 978).

Dada uma função $F(x_1, \dots, x_n)$ em um ponto x^* na direção de um vetor qualquer $v = (v_1, \dots, v_n)$, tome a equação paramétrica da reta por x^* na direção v:

$$x = x^* + tv.$$

A fim de obter a variação da função ao longo da reta, calcule:

$$g(t) \equiv F(x^* + tv) = F(x_1^* + tv_1, \dots, x_n^* + tv_n).$$

Aplicando a Regra da Cadeia para tomar a derivada de g em $t = 0$:

$$g'(0) = \frac{\partial F}{\partial x_1}(x^*)v_1 + \cdots + \frac{\partial F}{\partial x_n}(x^*)v_n$$

ou, em notação matricial:

$$\begin{pmatrix} \frac{\partial F}{\partial x_1}(x^*) & \dots & \frac{\partial F}{\partial x_n}(x^*) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = DF_{x^*}.v.$$

para: $DF_{x^*}.v$ = derivada direcional de F na direção e sentido de v em x^*.

Escrevendo a derivada de F em x^* como uma matriz-coluna, tem-se:

$$DF_{x^*} = \begin{pmatrix} \frac{\partial F}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial F}{\partial x_n}(x^*) \end{pmatrix}.$$

Esta matriz-coluna é interpretada como um vetor de $\mathbb{R}^n$ com cauda em x^*. Este vetor, denotado por $\nabla F(x^*)$, é conhecido como vetor gradiente de F em x^*. Deve-se lembrar que, como qualquer outro vetor, o vetor gradiente “leva” consigo as importantes características de um vetor qualquer: comprimento, direção e sentido.

Perceba que a derivada direcional de F em x^* na direção v, $DF_{x^*}.v$, pode ser representada pelo produto escalar dos vetores $\nabla F(x^*)$ e v:

$$\nabla F(x^*).v = \begin{pmatrix} \frac{\partial F}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial F}{\partial x_n}(x^*) \end{pmatrix}.(v_1 \dots v_n) = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(x^*)v_i.$$

Pela propriedade do produto escalar:

$$\nabla F(x^*).v = \|\nabla F(x^*)\|.\|v\|.\cos\theta.$$

Para: $\|v\| = 1$ e $\theta =$ ângulo entre os vetores $\nabla F(x^*)$ e v no ponto base x^*.

Observe que a função F cresce mais rapidamente quando $\cos\theta = 1$, ou seja, quando $\theta = 0°$. Portanto, pode-se concluir que, com v de tamanho unitário, $\nabla F(x^*).v$ será maior ao passo que v aponta para mesma direção e sentido que $\nabla F(x^*)$.

Teorema: Dada uma função $F: \mathbb{R}^n \to \mathbb{R}^1$, C^1, $\forall x \in Domínio\ de\ F$, em que $\nabla F(x) \neq 0$, o vetor gradiente de F em x aponta na direção em que F cresce mais rapidamente.

Ou, também:

a) se $\nabla F(x) = 0$ em x $\Longrightarrow$ todas as derivadas direcionas de F em x são nulas;
b) se $\nabla F(x) \neq 0$ em $x \Longrightarrow$ a derivada de F em x na direção e sentido de ∇F tem o maior valor;

c) se $\nabla F(x) \neq 0$ em $x \Longrightarrow$ a derivada de F em x no sentido oposto de ∇F tem o menor valor.

(ANTON; BIVENS; DAVIS, 2007. *Adaptado*).

Uma forma similar de resolver os problemas de otimização de uma função condicionada pode ser vista através dos vetores gradientes. Veja:

Exemplo: Sejam f e h funções C^1 de duas variáveis e suponha $x^* = (x_1^*, x_2^*)$.

Maximize $f(x_1, x_2)$

Sujeita a $h(x_1, x_2) = c$

Neste caso, os vetores gradientes são:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} \text{ e } \nabla h(x) = \begin{pmatrix} \frac{\partial h}{\partial x_1} \\ \frac{\partial h}{\partial x_2} \end{pmatrix}.$$

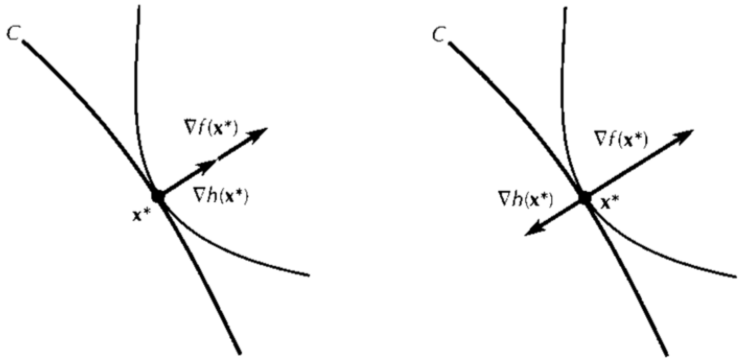
Sabe-se que os vetores gradientes devem estar alinhados em x^*, uma vez que os conjuntos de nível de f e h possuem a mesma inclinação no ponto x^*. Neste caso, os vetores gradientes devem apontar no mesmo sentido ou em sentidos opostos. Independente do caso, pode-se concluir que os vetores gradientes são múltiplos escalares um do outro, podendo ser escritos como:

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \mu^* \begin{pmatrix} \frac{\partial h}{\partial x_1} \\ \frac{\partial h}{\partial x_2} \end{pmatrix},$$

ou seja:

$$\nabla f(x^*) = \mu^* . \nabla h(x^*).$$

Figura 8: Vetores gradientes alinhados no máximo ou mínimo condicionado x*



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 427).

3.2.3 Diversas Restrições de Igualdade

O problema de maximização ou minimização de uma função sujeita a diversas restrições de igualdade pode ser entendido como uma “expansão” do problema condicionado à uma única restrição de igualdade. Veja:

Exemplo: Sejam $f, h_1, \dots, h_m$ funções C^1 de n variáveis. Busca-se maximizar ou minimizar $f(x)$ no conjunto-restrição:

$$C_h \equiv \{x = (x_1, \dots, x_n) | h_1(x) = a_1, \dots, h_m(x) = a_m\}.$$

Suponha que $x^* \in C_h$ e que x^* é um máximo ou mínimo (local) de f em C_h. Além disso, deve-se analisar se x^* satisfaz a qualificação de restrição, portanto:

$$\left(\frac{\partial h}{\partial x_1}(x^*), \frac{\partial h}{\partial x_2}(x^*), \dots, \frac{\partial h}{\partial x_n}(x^*)\right) \neq (0,0, \dots ,0).$$

A generalização da qualificação de restrição quando se está tratando com m funções de restrição, para $m > 1$, envolve a *derivada jacobiana* das funções de restrição:

$$Dh(x^*) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(x^*) & \dots & \frac{\partial h_1}{\partial x_n}(x^*) \\ \frac{\partial h_2}{\partial x_1}(x^*) & \cdots & \frac{\partial h_2}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(x^*) & \cdots & \frac{\partial h_m}{\partial x_n}(x^*) \end{pmatrix}.$$

O ponto x^* é denominado ponto crítico de $h = (h_1, \dots, h_m)$ se o posto[4] da matriz $Dh(x^*) < m$. Em outras palavras, $(h_1, \dots, h_m)$ satisfaz a qualificação de restrição não-degenerada (QRND) em x^* se o posto da matriz jacobiana $Dh(x^*)$ em x^* é m.

Assim, se x^* satisfaz a QRND, então existem $\mu_1^*, .., \mu_m^*$ tais que $(x_1^*, .., x_m^*, \mu_1^*, .., \mu_m^*) \equiv (x^*, \mu^*)$ é um ponto crítico do lagrangiano. Logo:

$$L(x,\mu) \equiv f(x) - \mu_1[h_1(x) - a_1] - \mu_2[h_2(x) - a_2] - \cdots - \mu_m[h_m(x) - a_m],$$

ou seja:

$$\frac{\partial L}{\partial x_1}(x^*, \mu^*) = 0, \dots, \frac{\partial L}{\partial x_n}(x^*, \mu^*),$$

$$\frac{\partial L}{\partial \mu_1}(x^*, \mu^*) = 0, \dots, \frac{\partial L}{\partial \mu_m}(x^*, \mu^*).$$

Conclua que para maximizar ou minimizar uma função condicionada à diversas restrições de igualdade deve-se simplesmente construir o lagrangiano, igualar as $(m + n)$ derivadas parciais de primeira ordem a zero e, por fim, resolver o sistema com $(m + n)$ equações e com $(m + n)$ incógnitas.

3.2.4 Restrições de Desigualdade

A maioria dos problemas de otimização em Economia envolvem restrições de desigualdade. Novamente, para simplificar a análise, iniciar-se-á analisando uma função com duas variáveis e apenas uma restrição de desigualdade. Ampliar-se-á para casos com várias restrições de desigualdade a *posteriori*.

Exemplo: Suponha que f e g são C^1 em $\mathbb{R}^2$, e que (x^*, y^*) maximiza f no conjunto restrição $g(x,y) \le b$. Busque:

Maximizar $\quad f(x,y)$

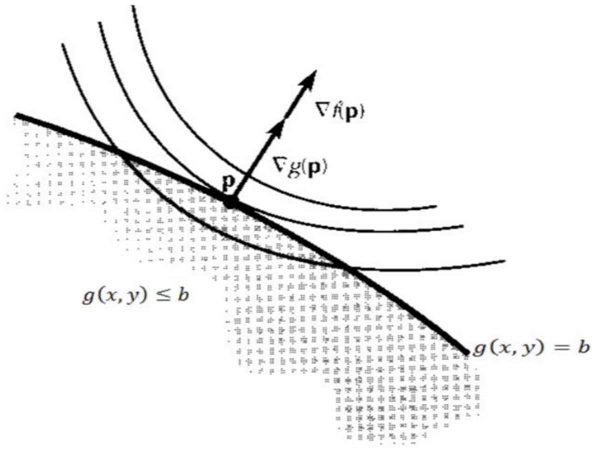
Sujeita a $\quad g(x,y) \le b$

---

[4] Definição: Seja uma matriz $A\ mxn$, o posto (ou a característica) de A é a dimensão do espaço das colunas de A. (LAY, D. Álgebra Linear e suas aplicações. 2007)

Antes de qualquer análise, observe a figura 9 abaixo:

Figura 9: Restrição de desigualdade e vetores gradientes



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 435).

A curva mais grossa representa $g(x, y) = b$, ao passo que a área hachurada abaixo representa o conjunto-restrição $g(x, y) \leq b$. As linhas mais finas representam as curvas de nível da função objetivo.

É sabido que o ponto de ótimo será dado onde o conjunto de nível de f e o conjunto de nível de g se tangenciam. Tal fato ocorre no ponto P, o qual está situado na fronteira do conjunto-restrição, onde $g(x, y) = b$. Deve-se dizer, portanto, que a restrição é ativa (também conhecida como vinculadora, eficaz ou justa).

Além disso, conforme também já mencionado, sabe-se que os vetores gradientes, $\nabla f(p)$ e $\nabla g(p)$, estão alinhados no ponto de ótimo, e, portanto, $\nabla f(p) = \lambda . \nabla g(p)$, ou seja, são múltiplos escalares. Entretanto, diferente do abordado nas restrições de igualdade, o sinal do multiplicador aqui é importante.

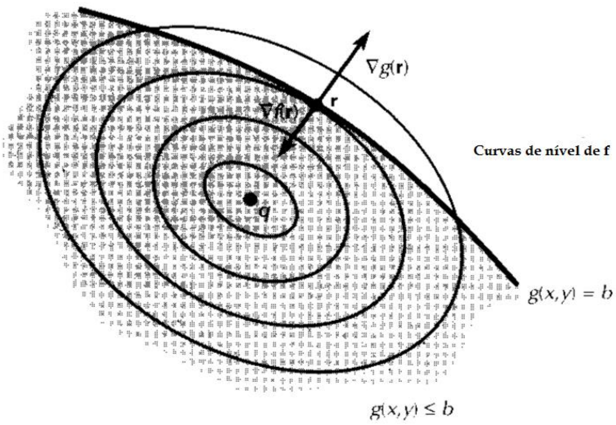
Sabe-se que o $\nabla f(p)$ aponta no sentido em que f cresce mais rapidamente. O $\nabla g(p)$, em particular, irá apontar sempre para o conjunto $g(x, y) \geq b$. Como o ponto P maximiza f no conjunto $g(x, y) \leq b$, o $\nabla f(p)$ deve apontar para a região onde $g(x, y) \geq b$. Portanto, conclui-se que para maximizar a função mencionada acima, $\nabla f(p)$ e $\nabla g(p)$ devem apontar no mesmo sentido.

A direção do correspondente multiplicador, λ (ou, alternativamente, μ), vai depender do conjunto-restrição e se procura-se maximizar ou minimizar a função objetivo. Isto é, quando se busca: maximizar $f(x^*)$ sujeita a $g(x^*) \leq c$, tal que x^*

é um máximo local de f, tem-se $\lambda^* \geq 0$. Quando busca-se maximizar $f(x^*)$ sujeita a $g(x^*) \geq c$, tal que x^* é um máximo local de f, tem-se $\mu^* \leq 0$. Em contrapartida, quando busca-se minimizar $f(x^*)$ sujeita a $g(x^*) \leq c$, tal que x^* é um mínimo local de f, tem-se $\lambda^* \leq 0$. E, por fim, quando busca-se minimizar $f(x^*)$ sujeita a $g(x^*) \geq c$, tal que x^* é um mínimo local de f, tem-se $\mu^* \geq 0$.

Uma outra possibilidade se dá quando o máximo da função f não ocorre em $g(x, y) = b$, mas sim em um ponto onde $g(x, y) < b$. Conforme pode-se observar na Figura 10 abaixo, suponha que o máximo ocorre no ponto q. Logo, apesar de existir um ponto r onde a função objetivo e a função restrição se tangenciam, o ponto r náo é um ponto de máximo, ao passo que o ponto q é um máximo local de f, ou seja, um máximo local não-condicionado. Neste caso, como q está no interior do conjunto-restrição, se diz que a restrição é inativa (também denominada não-vinculadora, ineficaz ou solta).

Figura 10: Situação de uma restrição inativa



Fonte: SIMON, C. P.; BLUME, L. Matemática para Economistas. (2008, p. 436).

Observe que no caso de um máximo condicionado na fronteira ativa, pode-se formar a função lagrangiana como segue:

$$L(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - b]$$

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0$$

$$\frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0$$

Com a ressalva da QRND, ou seja, que o máximo não seja um ponto crítico da função restrição g.

No caso da fronteira inativa, ainda se pode igualar as derivadas parciais do lagrangiano a zero, desde que o multiplicador seja igual a zero, isto é, $\lambda = 0$. Logo, se a restrição é ativa tem-se $g(x, y) - b = 0$, então: $\lambda \geq 0\ ou\ \lambda \leq 0$. Em contrapartida, se a restrição é inativa tem-se $g(x, y) - b \neq 0$, então $\lambda = 0$.

Deve-se concluir que se o ponto de máximo ou mínimo condicionado ocorre na fronteira, ou seja, a fronteira é ativa: $g(x^*, y^*) = b$. Logo, $g(x^*, y^*) - b = 0$. Portanto, $\lambda^* \geq 0\ ou\ \lambda^* \leq 0$. Caso o ponto de máximo ou mínimo condicionado ocorra no interior do conjunto-restrição, a fronteira é inativa. Logo, $g(x^*, y^*) - b \neq 0$, o que obriga que o correspondente multiplicador seja igual a zero, $\lambda^* = 0$.

Como não se sabe a *priori* se no ponto de máximo ou mínimo a restrição é ativa ou inativa, pode-se resumir o critério em:

$$\lambda\,.\,[g(x, y) - b] = 0,$$

ou seja:

$$g(x, y) - b = 0 \quad \text{ou} \quad \lambda = 0.$$

Vê-se, portanto, que a condição $\frac{\partial L}{\partial \mu} = 0$ para restrições de igualdade pode não servir para as restrições de desigualdade, uma vez que o ponto de ótimo pode se dar na fronteira ativa ou inativa. A condição é, portanto, substituída por duas condições:

$$\lambda\,.\,[g(x, y) - b] = 0,$$

e

$$\frac{\partial L}{\partial \lambda} = g(x, y) - b \leq 0.$$

4. ABORDAGEM KUHN-TUCKER

Em Economia, os problemas de otimização mais comuns referem-se à problemas de maximização condicionadas a restrições de desigualdade e restrições de não-negatividade. Como por exemplo:

Maximizar $f(x_1, \dots, x_n)$

Sujeita a $g_1(x_1, \dots, x_n) \leq a_1, \dots, g_k(x_1, \dots, x_n) \leq a_k$

$$h_1(x_1, \dots, x_n) \geq b_1, \dots, h_l(x_1, \dots, x_n) \geq b_l$$

$$x_1 \geq 0, \dots, x_n \geq 0$$

Essa situação é tão comum que um lagrangiano "especial" fora criado para tratar deste caso. Esta formulação especial é conhecida como Formulação de Kuhn-Tucker, a qual remonta ao trabalho dos dois matemáticos norte-americanos Albert W. Tucker e Harold W. Kuhn.

O objetivo dos tópicos a seguir é formular as condições necessárias e suficientes para um ponto sela de uma função diferenciável $\phi(x, u)$ com termos não negativos e aplicá-las, através do lagrangiano, para maximizar uma função $g(x)$ condicionada a funções diferenciáveis $f_h(x)$. Além disso, procura-se demonstrar que a equivalência entre a restrição para um máximo da função $g(x)$ e o ponto sela de um lagrangiano $\phi(x, u)$ se mantém, exigindo-se apenas que $g(x)$ e $f_h(x)$ sejam côncavas para x não negativo. Posteriormente, analisar-se-á o Teorema de Kuhn-Tucker e suas respectivas condições, envolvendo as demonstrações trazidas pelo Lema de Farkas.

4.1 Contextualização

O Teorema de Kuhn-Tucker articula-se com os problemas de Programação Linear, o qual, tendo como base os resultados obtidos por Dantzig-Neumann para problemas não lineares, destaca que os pontos ótimos em um problema de otimização são os pontos de sela do lagrangiano associado.

A Programação Linear lida, inclusive, com problemas como maximizar uma função linear $g(x) \equiv \sum c_i x_i$ com n variáveis reais $x_1, \dots, x_n$ formando um vetor x condicionado por $m + n$ restrições de desigualdade.

$f_h(x) \equiv b_h - \sum a_{hi} x_i \geq 0, x_i \geq 0,\ h = 1, \dots, m;\ m = i, \dots, n.$

Esse problema pode ser transformado em um problema de minimax (ponto de sela) caso sofra algumas adaptações nos métodos de cálculos usualmente aplicados.

Seja o lagrangiano:

$$\phi(x, u) \equiv g(x) + \sum u_h f_h(x).$$

Então um vetor x^0 maximizará a função $g(x)$ sujeita às $m + n$ restrições se, e somente se existir um vetor u^0 com componentes não negativos. Ou seja, se diz que (x^*, u^*) é ponto de sela de ϕ se:

$$\phi(x, u^0) \leq \phi(x^0, u^0) \leq \phi(x^0, u)\ ,\ \forall x \geq 0, \forall u \geq 0.$$

(KUHN; TUCKER, 1950).

4.2 Condições Necessárias e Suficientes para um ponto Sela

Seja $\phi(x, u)$ uma função diferenciável de um n-vetor x, com $x_i \geq 0$, e um m-vetor u, com $u_h \geq 0$. Sendo as derivadas parciais nos pontos x^0, u^0:

$$\phi_x^0 = \left[\frac{\partial \phi}{\partial x_i}\right]^0, \qquad \phi_u^0 = \left[\frac{\partial \phi}{\partial u_h}\right]^0.$$

Problema do ponto sela: procura-se encontrar os vetores x^0 e u^0 não negativos tais que:

$$\phi(x, u^0) \leq \phi(x^0, u^0) \leq \phi(x^0, u), \forall x \geq 0, u \geq 0.$$

Lema 1: As condições:

Condição 1: $\phi_x^0 \leq 0, \quad {\phi_x^0}' x^0 = 0, \quad x^0 \geq 0$ (1)

Condição 2: $\phi_u^0 \geq 0, \quad {\phi_u^0}' u^0 = 0, \quad u^0 \geq 0,$ (2)

são necessárias para que x^0 e u^0 forneçam a solução para o problema do ponto sela.

Lema 2: As condições (1), (2) e as condições:

Condição 3: $\phi(x, u^0) \leq \phi(x^0, u^0) + {\phi_x^0}'(x - x^0), \forall x \geq 0$ (3)

Condição 4: $\phi(x^0, u) \geq \phi(x^0, u^0) + {\phi_u^0}'(u - u^0), \forall u \geq 0,$ (4)

são suficientes para que x^0 e u^0 forneçam a solução para o problema do ponto sela.

As condições (3) e (4) são satisfeitas se $\phi(x, u^0)$ for uma função côncava de x e se $\phi(x^0, u)$ for uma função convexa de u.

(KUHN; TUCKER, 1950).

4.3 Multiplicadores de Lagrange para um máximo condicionado

Suponha $F(x)$ um m-vetor com componentes $f_1(x), \dots, f_m(x)$, os quais são funções diferenciáveis de $x, \forall x \geq 0$. Seja $g(x)$ uma função diferenciável de x para $x \geq 0$. Sejam as derivadas parciais, avaliadas em x^0:

$$F^0 = \left[\frac{\partial f_h}{\partial x_i}\right]^0, \qquad g^0 = \left[\frac{\partial g}{\partial x_i}\right]^0.$$

Problema de maximização: consiste em encontrar um x^0 tal que maximize $g(x)$ sujeito a $F(x) \geq 0, \forall x \geq 0$.

Supondo que u seja um conjunto de m múltiplos de Lagrange não negativos, forme a função:

$$\phi(x, u) \equiv g(x) + u'Fx.$$

Então:

$$\phi_x^0 = g^0 + F^{0'}u^0, \qquad \phi_u^0 = Fx^0.$$

Teorema 1: Para que x^0 seja solução do problema de maximização, é necessário que x^0 e que algum u^0 satisfaçam as condições (1) e (2) para $\phi(x,u) \equiv g(x) + u'F(x)$.

Teorema 2: Para que x^0 seja solução do problema de maximização, é suficiente que x^0 e que algum u^0 satisfaçam as condições $(1), (2)$ e (3) para $\phi(x,u) \equiv g(x) + u'F(x)$.

Prova: A partir das condições $(3), (1)$ e (2) tem-se:

$$g(x) + u^{0'}Fx = \phi(x,u^0) \leq \phi(x^0,u^0) + \phi_x^{0'}(x - x^0)$$
$$g(x) + u^{0'}Fx \leq \phi(x^0,u^0) = g(x^0) + u^{0'}Fx^0 = g(x^0), \forall x \geq 0,$$

mas $u^{0'}Fx^0 \geq 0, \forall x$ que satisfaça $Fx \geq 0$. Portanto, $g(x) \leq g(x^0), \forall x$ que satisfaça $Fx \geq 0, \forall x \geq 0$. Isto prova o Teorema 2.

Note que no Teorema 2, a condição (3) só é mantida para $Fx \geq 0, x \geq 0$.

(KUHN; TUCKER, 1950).

4.4 Concavidade-convexidade e o Teorema da Equivalência

Dada a definição de uma função côncava ou convexa já mencionada:

Lema 3: Se $f(x)$ é convexa e diferenciável, então:

$$f(x) \geq f(x^0) + f^{0'}(x - x^0),$$

sendo

$$f^0 = \left[\frac{\partial f}{\partial x_i}\right]^0,$$

para todo x^0 e x dentro da região de definição. Caso $f(x)$ seja côncava, a desigualdade é revertida.

Teorema 3: (***Equivalence Theorem)***. Sejam $f_1(x), \dots, f_m(x), g(x)$ funções côncavas e diferenciáveis para $x \geq 0$. Então, x^0 é uma solução para o problema de maximização se, e somente se x^0 e algum u^0 fornecer a solução do problema do ponto de sela para $\phi(x,u) \equiv g(x) + u'Fx$.

Prova: Pelo Lema 3 (para concavidade):

$$Fx \leq Fx^0 + F^{0'}(x - x^0),$$
$$g(x) \leq g(x^0) + g^{0'}(x - x^0),$$
$$\forall x^0 \geq 0, \forall x \geq 0.$$

Então, para qualquer $u^0 \geq 0$,

$$\phi(x, u^0) = g(x) + u^{0'}Fx$$

$$\phi(x, u^0) \leq g(x^0) + u^{0'}Fx^0 + \left(g^{0'} + u^{0'}F^{0'}\right).(x - x^0)$$

$$\phi(x, u^0) = \phi(x^0, u^0) + \phi_x^{0'}(x - x^0).$$

Isto é, a condição 3 se mantém para todo $x^0 \geq 0$ e $x \geq 0$. Sobre essas circunstâncias, o Teorema 1 e o Teorema 2 se combinam para fazer das condições (1) e (2) ambas necessárias e suficientes tal que x^0 forneça a solução para o problema de maximização.

A condição (4), por sua vez, se mantém automaticamente, desde que a linearidade de $\phi(x,u)$ com respeito a u implique que:

$$\phi(x^0, u) = \phi(x^0, u^0) + \phi_u^{0'}(u - u^0).$$

Conclui-se, portanto, que os Lemas 1 e 2 se combinam para fazer as condições (1) e (2) ambas necessárias e suficientes de tal modo que x^0 e u^0 forneçam a solução para o problema do ponto de sela. Completa-se aqui a prova do Teorema 3.

(KUHN; TUCKER, 1950).

4.5 Curvas em Trajetórias Admissíveis

Seja x^* um ponto na região

$$f_i(x) \geq 0, \dots, f_m(x) \geq 0.$$

Suponha que:

$$f_1(x^*) = 0, \dots, f_k(x^*) = 0,$$

e que

$$f_{k+1}(x^*) > 0, \dots, f_m(x^*) > 0,$$
$$x_{i_1}^* = 0, \dots, x_{i_l}^* = 0,$$

e que noutros casos $x_i^* > 0$.

Em outras palavras, as fronteiras $1, \dots, k, i_1, \dots, i_l$, são ativas e as demais inativas.

Denote por $\pi_1(x_1, \dots, x_n) = x_i$ as fronteiras dadas pelos hiperplanos $x_1 = 0, \dots, x_n = 0$, então: $\pi_i(x) \geq 0$.

Seja $x: [0, +\infty) \to \mathbb{R}^n$ uma curva, se diz que x é admissível em x^* se:

i. $x(0) = x^*$;

ii. $\frac{d}{dt} f_1(x(t))|_{t=0} = \nabla f_1(x(t)).x'(t)|_{t=0} \geq 0$

$$\vdots$$

$$\frac{d}{dt} f_k(x(t))|_{t=0} = \nabla f_k(x(t)).x'(t)|_{t=0} \geq 0$$

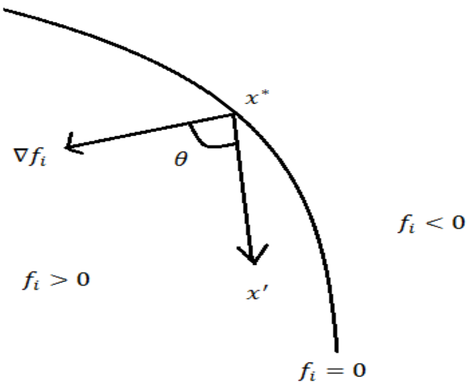
$$\frac{d}{dt} \pi_{i_1}(x(t))|_{t=0} = \nabla \pi_i(x(t)).x'(t)|_{t=0} = x_{i_1}'(0) \geq 0$$

$$\vdots$$

$$\frac{d}{dt} \pi_{i_l}(x(t))|_{t=0} = \nabla \pi_{\mathrm{i}}(x(t)).x'(t)|_{t=0} = x_{i_l}'(0) \geq 0.$$

Supondo que f_i seja ativa, então, se a curva é admissível, em $t = 0$ a curva tangencia a fronteira ou acaba "entrando" para o interior, onde $f_i > 0$.

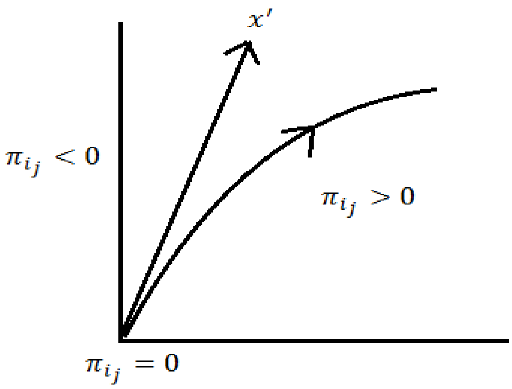
Figura 11: Curva admissível I



Fonte: Desenvolvida pelo autor.

O mesmo em relação a $x_{i_1}^* = 0, \ldots\ x_{i_l}^* = 0$.

Figura 12: Curva admissível II



Fonte: Desenvolvida pelo autor.

Observe que:

$$\nabla \pi_{i_j} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

4.6 O Teorema de Kuhn-Tucker

Teorema: Se x^* é máximo local de g sujeito a: $f_1 \geq 0, \dots, f_m \geq 0, x_1 \geq 0, \dots, x_n \geq 0,$ então as três seguintes condições, chamadas condições de Kuhn-Tucker, devem ser satisfeitas:

1ª condição:

$$\frac{\partial \phi}{\partial x_1} + u_1^* . \frac{\partial f_1^*}{\partial x_1} + \cdots + u_m^* . \frac{\partial f_m^*}{\partial x_1} \leq 0 \qquad (1)$$
$$\vdots$$
$$\frac{\partial \phi}{\partial x_n} + u_1^* . \frac{\partial f_1^*}{\partial x_n} + \cdots + u_m^* . \frac{\partial f_m^*}{\partial x_n} \leq 0$$

2ª condição:

$$u_1^* f_1^* = 0, \dots, u_m^* f_m^* = 0 \qquad (2)$$
$$\lambda_1^* x_1^* = 0, \dots, \lambda_n^* x_n^* = 0$$

3ª condição:

$$u_1^* \geq 0, \dots, u_m^* \geq 0, \lambda_1^* \geq 0, \dots, \lambda_n^* \geq 0 \qquad (3)$$

Teorema: Se valem a 1ª, 2ª e 3ª condição de Kuhn-Tucker (KT) e se $\phi(\,\cdot\,, u^*)$ é côncava $\Longrightarrow x^*$ é máximo global.

Teorema: Se valem a 1ª, 2ª e 3ª condição de Kuhn-Tucker (KT) e se $g, f_1, \dots, f_m$ são côncavas em $x \Longrightarrow x^*$ é máximo local.

Na prova do Teorema de Kuhn-Tucker vamos usar o Lema de Farkas.

Lema de Farkas: Seja A uma matriz $mxn, \vec{b} \in \mathbb{R}^n$.

$A'x = b, x \geq 0$, tem solução $\Leftrightarrow \forall y \in \mathbb{R}^n$, se $Ay \geq 0 \Longrightarrow b'y \geq 0$.

Prova do Teorema de Kuhn-Tucker:

No caso:

Se $x(t)$ é uma curva admissível, então:

$$\begin{pmatrix} \frac{\partial f_1^*}{\partial x_1} & \cdots & \frac{\partial f_1^*}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k^*}{\partial x_1} & \cdots & \frac{\partial f_k^*}{\partial x_n} \\ \delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_n \end{pmatrix} \begin{pmatrix} x_1'(0) \\ \vdots \\ x_n'(0) \end{pmatrix} \geq 0. \qquad (1)$$

Como x^* é máximo local para g, então:

$$-\nabla g^*.x'(0) \geq 0. \qquad (2)$$

Pelo Lema de Farkas existe solução $u^* \geq 0, \lambda^* \geq 0$:

$$\begin{pmatrix} \frac{\partial f_1^*}{\partial x_1} & \cdots & \frac{\partial f_k^*}{\partial x_1} & \delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1^*}{\partial x_n} & \cdots & \frac{\partial f_k^*}{\partial x_n} & 0 & \cdots & \delta_n \end{pmatrix} \begin{pmatrix} u_1^* \\ \vdots \\ u_k^* \\ \lambda_1^* \\ \vdots \\ \lambda_n^* \end{pmatrix} = \begin{pmatrix} \frac{-\partial g^*}{\partial x_1} \\ \vdots \\ \frac{-\partial g^*}{\partial x_n} \end{pmatrix}, \qquad (3)$$

ou seja, existe:

$$\begin{pmatrix} u^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} u_1^* \\ \vdots \\ u_k^* \\ \lambda_1^* \\ \vdots \\ \lambda_n^* \end{pmatrix} \geq 0, \qquad (4)$$

tal que, de (3):

$$u_1^*.\frac{\partial f_1^*}{\partial x_1} + \cdots + u_k^*.\frac{\partial f_k^*}{\partial x_1} + \delta_1 \lambda_1^* = -\frac{\partial g^*}{\partial x_1} \qquad (5)$$

$$\vdots$$

$$u_1^*.\frac{\partial f_1^*}{\partial x_n} + \cdots + u_k^*.\frac{\partial f_k^*}{\partial x_n} + \delta_n \lambda_n^* = -\frac{\partial g^*}{\partial x_n}$$

ou:

$$\frac{\partial g^*}{\partial x_1} + u_1^*.\frac{\partial f_1^*}{\partial x_1} + \cdots + u_k^*.\frac{\partial f_k^*}{\partial x_1} = -\delta_1 \lambda_1^* \leq 0 \qquad (6)$$

$$\vdots$$

$$\frac{\partial g^*}{\partial x_n} + u_1^*.\frac{\partial f_1^*}{\partial x_n} + \cdots + u_k^*.\frac{\partial f_k^*}{\partial x_n} = -\delta_n \lambda_n^* \leq 0.$$

Obs: para $k+1, \ldots, m, u_i^* = 0$.

Isto é, de (6) obtém-se:

$$\nabla_x \phi(x^*, u^*) \leq 0, \tag{7}$$

ou

$$\nabla_x \phi(x^*, u^*) \leq -\begin{pmatrix} \delta_1 \lambda_1^* \\ \vdots \\ \delta_n \lambda_n^* \end{pmatrix}. \tag{8}$$

Como $\delta_i \lambda_i^* \geq 0$ e $x_i \geq 0$,

$$\delta_i \lambda_i^* \,.\, x_i \leq 0 \Longrightarrow \nabla_x \phi(x^*, u^*).\, x \leq 0. \tag{9}$$

Pode-se ver que:

$$\nabla_x \phi(x^*, u^*).\, x^* = -\delta_1 \lambda_1^* x_1^* + \cdots + -\delta_n \lambda_n^* x_n^*. \tag{10}$$

Como:

$\delta_i \lambda_i^* x_i^* = 0$, se $x_i^* > 0$, pois neste caso $i \notin \{i_1, \dots, i_n\}$ e então $\delta_i = 0$ (11)

e

$\delta_i \lambda_i^* x_i^* = 0$, se $x_i^* = 0$. (12)

Assim, obtém-se:

$$\nabla_x \phi(x^*, u^*).\, x^* = 0. \tag{13}$$

De (9), observe que:

$$\nabla_x \phi(x^*, u^*).\, x \leq 0. \tag{14}$$

Ainda:

$$\nabla_u \phi(x^*, u^*) = \begin{pmatrix} f_1^* \\ \vdots \\ f_k^* \\ f_{k+1}^* \\ \vdots \\ f_m^* \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ f_{k+1}^* > 0 \\ \vdots \\ f_m^* > 0 \end{pmatrix} \geq 0, \tag{15}$$

então:

$$\nabla_u \phi(x^*, u^*) \geq 0 \quad \text{e} \quad u^* \geq 0. \tag{16}$$

Além disso:

$$\text{se } f_j\ (x^*) > 0 \Longrightarrow u_j^* = 0, \tag{17}$$
$$\text{se } f_j\ (x^*) = 0 \Longrightarrow u_j^*\ pode\ ser > 0.$$

De qualquer forma:

$$u_j^* f_j(x^*) = 0 \Longrightarrow \nabla_u \phi(x^*, u^*).\, u_j^* = 0. \tag{18}$$

Portanto, prova-se que: se x^* é máximo local, então existe:

$$0 \leq u^* \ \longrightarrow \text{3ª condição de Kuhn-Tucker;} \tag{19}$$

tal que:

$$\nabla_x \phi(x^*, u^*) \leq 0 \ \longrightarrow \text{1 ª condição de Kuhn-Tucker} \tag{20}$$

e

$$\nabla_x \phi(x^*, u^*).\, x^* = 0 \ \longrightarrow \text{2 ª condição de Kuhn-Tucker} \tag{21}$$
$$\nabla_x \phi(x^*, u^*).\, x \leq 0,$$
$$\nabla_u \phi(x^*, u^*).\, u^* = 0.$$

De fato:

$$\nabla_x \phi(x^*, u^*) \leq 0 \ \longrightarrow \text{1 ª condição de KT} \tag{22}$$
$$\nabla_x \phi(x^*, u^*).\, x^* = 0\ , \nabla_u \phi(x^*, u^*).\, u^* = 0 \longrightarrow \text{2 ª condição de KT}$$
$$u^* \geq 0\ ,\ \ \lambda^* \geq 0 \longrightarrow \text{3 ª condição de KT.}$$

Na verdade, o Lema de Farkas está afirmando que:

$$\nabla g^* = (-u_1^*)\nabla f_1^* + \cdots + (-u_m^*)\nabla f_m^* + (-\lambda_1^*)\overrightarrow{e_1} + \cdots + (-\lambda_n^*)\overrightarrow{e_n}. \tag{23}$$

Portanto, se x^* é máximo local de $g \longrightarrow \nabla g^*$ é combinação dos gradientes das restrições em:

$$-u_1^* \leq 0, \ldots, -u_m^* \leq 0, -\lambda_1^* \leq 0, \ldots, -\lambda_n^* \leq 0, \tag{24}$$
$$(-u_1^*)(f_1^*) = 0, \ldots, (-u_m^*)(f_m^*) = 0,$$
$$\lambda_1^* x_1^* = 0, \ldots, \lambda_n^* x_n^* = 0,$$

que são as condições necessárias de Kuhn-Tucker para um máximo.

Por outro lado, seja:

$$\phi(x,u) = g(x) + u'F(x) \tag{25}$$

e

$$\phi(x,u^*) = g(x) + u_1^* f_1(x) + \cdots + u_k^* f_k(x). \tag{26}$$

Supondo que ϕ seja côncava, por exemplo, tem-se:

$$\phi(x,u^*) \leq \phi(x^*,u^*) + \nabla\phi_x(x^*,u^*)(x - x^*) \tag{27}$$

$$\phi(x,u^*) \leq \phi(x^*,u^*) + \nabla\phi_x(x^*,u^*).x - \nabla\phi_x(x^*,u^*).x^*$$

Sabe-se que:

$$\nabla\phi_x(x^*,u^*).x \leq 0,\ \text{e} \tag{28}$$

$$\nabla\phi_x(x^*,u^*).x^* = 0,$$

então:

$$\phi(x,u^*) \leq \phi(x^*,u^*) + \nabla\phi_x(x^*,u^*).x. \tag{29}$$

Sabe-se, também, que:

$$\phi(x^*,u^*) = g(x^*) + u_1^* f_1(x^*) + \cdots + u_k^* f_k(x^*), \tag{30}$$

mas

$$u_1^* f_1(x^*) + \cdots + u_k^* f_k(x^*) = \phi_u(x^*,u^*).u^* = 0, \tag{31}$$

então:

$$\phi(x^*,u^*) = g(x^*). \tag{32}$$

Logo:

$$\phi(x,u^*) \leq g(x^*) + \nabla\phi_x(x^*,u^*).x \leq g(x^*) \tag{33}$$

e

$$\phi(x,u^*) = g(x) + u_1^* f_1(x) + \cdots + u_k^* f_k(x) \geq g(x), \tag{34}$$

sendo $u_1^* f_1(x) + \cdots + u_k^* f_k(x) \geq g(x) \geq 0$.

Por fim:

$$g(x) \leq \phi(x, u^*) \leq g(x^*) + \nabla\phi_x(x^*, u^*).\, x \leq g(x^*), \tag{35}$$

ou seja: x^* é máximo global.

Portanto, conclui-se que:

Se $u^* \geq 0$,

$$\begin{gathered} \nabla_x(x^*, u^*).\, x \leq 0 \longrightarrow \text{da 1ª condição de KT} \\ \nabla_x(x^*, u^*).\, x^* = 0 \longrightarrow \text{da 2ª condição de KT} \\ \nabla_u(x^*, u^*).\, u^* = 0 \longrightarrow \text{da 2ª condição de KT} \end{gathered} \tag{36}$$

e se $\phi(\,\cdot\,, u^*)$ é côncava, por exemplo, então x^* é um máximo global.

Ou seja, comprova-se aqui o teorema mencionado anteriormente:

Teorema: Se valem a 1ª, 2ª e 3ª condição de Kuhn-Tucker (KT) e se $\phi(\,\cdot\,, u^*)$ é côncava $\Longrightarrow x^*$ é máximo global.

Além disso, veja:

Se g é côncava e $f_1, \dots, f_m$ são côncavas, então:

$$\begin{gathered} g(x) \leq g(x^*) + \nabla g(x^*)(x - x^*), \\ u_1^* f_1(x) \leq u_1^* f_1(x^*) + u_1^* \nabla f_1(x^*)(x - x^*) \\ \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ u_m^* f_m(x) \leq u_m^* f_m(x^*) + u_m^* \nabla f_m(x^*)(x - x^*). \end{gathered} \tag{37}$$

Se $u^* \geq 0$,

$$\phi(\,\cdot\,, u^*) \leq \phi(x^*, u^*) + \nabla_{\mathrm{x}}\phi(x^*, u^*).\,(x - x^*), \tag{38}$$

portanto, $\phi(\,\cdot\,, u^*)$ é côncava em x.

Comprova-se aqui o outro teorema mencionado anteriormente:

Teorema: Se valem a 1ª, 2ª e 3ª condição de Kuhn-Tucker (KT) e se $g, f_1, \dots, f_m$ são côncavas em $x \Longrightarrow x^*$ é máximo local.

O chamado Teorema de Kuhn-Tucker resulta do grupo de matemáticos formados por Kuhn, Tucker e D. Gale, que durante anos se propuseram a desenvolver a Teoria das Inequações Lineares, cujos resultados acharam publicados nos *Annals of Mathematical* de Princeton durante quase duas décadas.

O Teorema de Kuhn-Tucker pode ser entendido como uma tentativa dos autores em aproximar a otimização não-linear à otimização linear. Neste sentido, o trabalho não teve maiores consequências, mas acabou favorecendo uma inovadora demonstração de resultados relativamente conhecidos (Teoria da Dualidade), além de uma extensão para a Otimização Vetorial.

Ao passo que estendem seu resultado para a Otimização Vetorial, os autores mostraram que as condições a serem satisfeitas por ótimos de Pareto podem ser expressas através das mesmas condições de Kuhn-Tucker com algumas modificações e que dadas as condições de concavidade estas condições são também suficientes.

Ainda no artigo “*Nonlinear Programming*”, o segundo objetivo dos autores era reduzir um problema com restrições e várias funções objetivos a um problema com uma função objetivo apenas. Uma nova subdisciplina da Teoria da Otimização, conhecida como Otimização Vetorial, ganhava espaço.

5. OTIMIZAÇÃO VETORIAL

5.1 Problemas de Programação Matemática

Uma forma geral de problemas de programação matemática pode ser dada como:

$$\max_{x} F(x) \text{ sujeito a } x \in X$$

Sendo x uma coluna de vetores com n variáveis

$$x = (x_1, x_2, \dots, x_n)'$$

(a apóstrofe denota a transposta do vetor linha)

Sendo $F(x)$ uma função dessas variáveis

$$F(x) = F(x_1, x_2, \dots, x_n),$$

e X um subconjunto Euclidiano n-dimensional.

$$x \subset E^n$$

Deve-se assumir que X não é vazio, isto é, existe algum vetor viável x, onde x é viável se, e somente se $x \in X$.

Existem diversos "tipos de soluções" para o problema mencionado acima. Um máximo global, por exemplo, é um vetor x^*, onde:

$$x^* \in X \quad \text{e} \quad F(X^*) \geq F(x), \quad \forall x \in X.$$

Um máximo global estrito seria dado por um vetor x^* que satisfaça:

$$x^* \in X \quad \text{e} \quad F(X^*) > F(x), \quad \forall x \in X, \ x \neq x^*.$$

(ARROW; INTRILIGATOR, 1981).

5.2 Ótimo Local-Global

Um vetor $x^* \in X$ será um máximo local para algum $\varepsilon > 0$, se:

$$F(x^*) \geq F(x), \qquad \forall x \in X \cap N_\varepsilon(x^*).$$

Sendo aqui:

$$N_\varepsilon(x^*) = \{x||x - x^*| \leq \varepsilon).$$

$N_\varepsilon(x^*)$ é uma ε vizinhança de x^*, isto é, para todos os pontos não mais distantes que ε de x^*.

Será um máximo local estrito se o vetor $x^* \in X$ satisfizer:

$$F(x^*) > F(x), \qquad \forall x \in X \cap N_\varepsilon(x^*), x \neq x^*.$$

Obviamente, um máximo global será um máximo local, mas o inverso não é verdadeiro. O mesmo é valido para os casos em que são estritos.

De acordo com o *Local-Global Theorem*, se a função objetivo $F(X)$ é côncava e conjunto do vetor X é um conjunto convexo, então todo máximo local será também um máximo global, o conjunto de todas as solução será convexo e a solução será única se $F(x)$ for uma função estritamente côncava.

A generalização do Teorema diz que se $F(x)$ é estritamente quase-côncava[5], então um máximo local é um máximo global único (ARROW; INTRILIGATOR, 1981).

Com a utilização deste Teorema é possível inferir, com as suposições adequadas de concavidade e convexidade, que uma solução de máximo local é também uma solução de máximo global.

5.3 Ótimo de Pareto

Dado $(x_1^*, x_2^*) \in R^n$,

$$x_1^* = \begin{pmatrix} x_{11}^* \\ \vdots \\ x_{n1}^* \end{pmatrix}, \quad x_2^* = \begin{pmatrix} x_{12}^* \\ \vdots \\ x_{n2}^* \end{pmatrix},$$

(x_1^*, x_2^*) é uma alocação ótima no sentido de Pareto, ou fortemente eficiente no sentido de Pareto, se não existe (x_1, x_2), tal que:

$$\begin{pmatrix} u_1(x_1) \\ u_2(x_2) \end{pmatrix} \geq \begin{pmatrix} u_1(x_1^*) \\ u_2(x_2^*) \end{pmatrix}.$$

[5] Uma função $F(x)$ é quase é estritamente quase-côncava se, e somente se dado $x^1, x^2 \in X, x^1 \neq x^2, F(\alpha x^1 + (1-\alpha)x^2) > F(x^2)$, onde $F(x^1) > F(x^2), \forall \alpha, 0 < \alpha < 1$.

Ou seja, não existe x_1, x_2 tais que $u_1(x_1) \geq u_1(x_1^*)$ e $u_2(x_2) \geq u_2(x_2^*)$ e alguma das desigualdades é estrita, isto é: $u_1(x_1) > u_1(x_1^*)$ ou $u_2(x_2) > u_2(x_2^*)$.

Se diz que (x_1^ϕ, x_2^ϕ) é fracamente eficiente se não existe (x_1, x_2) tais que $u_1(x_1) > u_1(x_1^\phi)$ e $u_2(x_2) > u_2(x_2^\phi)$, ou seja, quando não é possível melhorar ambos.

Exemplo: Dados dois bens: x_1, x_2

Maximizar $U_2\,(x_2, y_2) = U_2^*$

Sujeito a $x_1 + x_2 = \bar{x}$

$y_1 + y_2 = \bar{y}$

$u_1(x_1, y_1) \geq \overline{u_1}$

$x_1 \geq 0, y_1 \geq 0, x_2 \geq 0, y_2 \geq 0$

Supondo que maximizando este problema utilizando as condições do Teorema de Kuhn-Tucker encontrar-se-á $x_1^*, x_2^*, y_1^*, y_2^*$ como solução. Logo, conclui-se que:

a) No ponto de ótimo:
$u_2(x_2^*, y_2^*) \geq u_2(x_2, y_2)$
$u_1(x_1^*, y_1^*) \geq \overline{u_1}$

b) Pode existir:
$u_2(x_2', y_2') = u_2(x_2^*, y_2^*)$
$u_1(x_1', y_1') > u_1(x_1^*, y_1^*) \geq \overline{u_1}$
Logo: $(x_1^*, y_1^*), (x_2^*, y_2^*)$ não será fortemente eficiente, mas sim fracamente eficiente.

c) Não é possível:
$u_2(x_2', y_2') > u_2(x_2^*, y_2^*)$
$u_1(x_1', y_1') > u_1(x_1^*, y_1^*) \geq \overline{u_1}$

Montando o lagrangiano:

$$L(\cdot) \equiv u(x_2, y_2) - \lambda_1(x_1 + x_2 - \bar{x}) - \lambda_2(y_1 + y_2 - \bar{y}) - \mu(u_1(x_1, y_1) - \overline{u_1}).$$

Se as quantidades dos bens são positivas e as utilidades contínuas, as soluções dos problemas são fortemente eficientes no sentido de Pareto.

Teorema: O conjunto $x_1, \dots, x_n$ de n-uplas (x_i) é parcialmente ordenado da seguinte maneira: $(x_i') \geq (x_i)$ se e somente se $x_i' \geq x_i \; \forall i$.

A alocação (x_i^0, y_i^0) é ótimo de Pareto se:

- (x_i^0, y_i^0) é atingível
- Não existe alocação possível (x_i, y_i) tal que $(x_i) > (x_i^0)$. Isto é: é impossível fazer um consumidor melhor sem fazer o outro piorar.

(DEBREU, 1986. *Adaptado*).

5.4 Programação Linear

Um dos conceitos mais importantes em Programação Linear (PL) refere-se à Teoria da Dualidade. Qualquer problema de PL tem o problema original, chamado primal, e a ele associado um outro problema de PL, chamado dual. Veja um exemplo:

- **Problema primal:**

Maximizar $b_1x_1 + \cdots + b_nx_n$

Sujeito a $a_{11}x_1 + \cdots + a_{1n}x_n \leq c_1$

$$\vdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n \leq c_m$$

$$x_1 \geq 0, \dots, x_n \geq 0$$

Analogamente:

Maximizar $b'x$

Sujeito a $Ax \leq c$

$x \geq 0$

(a apóstrofe denota a transposta do vetor linha)

tal que:

$$b'x = (b_1 \ldots b_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}, \qquad c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

Montando o lagrangiano:

$$L = b_1x_1 + \cdots + b_nx_n - y_1(a_{11}x_1 + \cdots + a_{1n}x_n - c_1)$$

$$\vdots$$

$$- y_m(a_{m1}x_1 + \cdots + a_{mn}x_n - c_m)$$

ou seja:

$$L = b'x - y'(Ax - c).$$

- **Problema dual:**

Minimizar $c_1y_1 + \cdots + c_my_m$

Sujeito a $a_{11}y_1 + \cdots + a_{m1}y_m \geq b_1$

$\vdots$

$a_{m1}y_1 + \cdots + a_{mn}y_m \geq b_n$

$y_1 \geq 0, \ldots, y_m \geq 0$

Montando o lagrangiano: $L = c'y - x'(A'y - b)$

Observe que o lagrangiano do problema primal e do dual é o mesmo:

$$L = b'x - y'(Ax - c)$$

$$L = b'x - y'Ax + y'c$$

$$L' = x'b - x'A'y + c'y = c'y - x'(A'y - b)$$

Os multiplicadores de Lagrange do problema primal são as variáveis do problema dual e vice-versa.

As Condições de Kuhn-Tucker para o problema primal são:

1) $\frac{\partial L^*}{\partial x_1} \leq 0, \ldots, \frac{\partial L^*}{\partial x_n} \leq 0$
2) $y^*(Ax^* - c) = 0$
3) $y^* \geq 0$

Percebe-se que: $L(x^*, y^*) = b'x^*$

Para o problema dual:

1) $\frac{\partial L^*}{\partial y_1} \geq 0, \ldots, \frac{\partial L^*}{\partial y_m} \geq 0$
2) $x^{*\prime}(A'y^* - b) = 0$
3) $x^* \geq 0$

Percebe-se que: $L(x^*, y^*) = c'y^*$

Então: $$L(x^*, y^*) = \underbrace{b'x^*}_{\substack{valor\ ótimo\ do \\ problema\ primal}} = \overbrace{c'y^*}^{\substack{valor\ ótimo\ do \\ problema\ dual}}$$

Observe que:

$$L(x^*, y) = b'x^* - y'(Ax^* - c) \geq b'x^* = L(x^*, y^*),$$

pois:

$$Ax^* \leq c \Longrightarrow y'(Ax^* - c) \leq 0, \quad y' \geq 0.$$

Portanto:

$$L(x^*, y) \geq L(x^*, y^*).$$

De maneira semelhante:

$$L(x, y^*) = c'y^* - x'(A'y^* - b) \leq c'y^* = L(x^*, y^*),$$

pois

$$A'y^* \geq b \Longrightarrow x'(A'y^* - b) \geq 0.$$

Portanto:

$$L(x, y^*) \leq L(x^*, y^*).$$

Por fim, observe que:

$$L(x, y^*) \leq L(x^*, y^*) \leq L(x^*, y).$$

Logo, como já previamente mencionado, conclui-se que $L(x^*, y^*)$ é uma sela do Lagrangiano. O resultado aqui obtido faz referência à Teoria da Dualidade.

É curioso observar que apesar da Teoria da Dualidade ser creditada à Kuhn, Tucker e Gale, o resultado obtido pelos autores já era anteriormente conhecido por um dos matemáticos mais influentes da época, o húngaro de origem judaica von Neumann.

No seu de artigo de 1981, *Reminiscences about the origins of linear programming*, o matemático estadunidense George B. Dantzig, considerado por muitos o pai da Programação Linear, comenta que ao passo em que vinha desenvolvendo o famoso Método Simplex, decidiu visitar von Neumann para ver o que ele poderia sugerir para técnicas de solução. A primeira visita ocorreu em 3 de outubro de 1947, no *Institute for Advanced Study* em Princeton.

Dantzig relata que ao passo que expõe as versões algébricas e geométricas no quadro-negro, von Neumann se levanta e inicia uma verdadeira aula sobre o assunto.

> *At one point seeing me sitting there with my eyes popping and my mouth open (after all I had searched the literature and found nothing), von Neumann said: "I don't want you to think I am pulling all this out of my sleeve on the spur of the moment like a magician. I have just recently completed a book with Oscar Morgenstern on the Theory of Games. What I am doing is conjecturing that the two problems are equivalent. The theory that I am outlining for your problem is an analogue to the one we have developed for games. Thus I learned about Farkas' Lemma, and about duality for the first time.* (DANTZIG, 1981, p. 5).

Em junho de 1948, outra visita de Dantzig em Princeton resultou no encontro com o professor A. Tucker e com seus estudantes H. Kuhn e D. Gale, os quais dedicavam-se à Teoria dos Jogos, Programação Não-linear e a Teoria da Dualidade. Doze anos depois, um novo encontro entre Dantzig e Tucker resultou numa conversa mencionado pelo autor:

> *Our conversation went like this: "Why", he asked, "do you ascribe duality to von Neumann and not to my group?" "Because he was the first to show it to me." He said, "that is strange for we have found nothing in writing about what von Neumann has done. What we have is his paper On a Maximizing Problem. "True," I said, "but let me send you a paper I wrote as a result of my first meeting with von Neumann." I sent him my report A Theorem on Linear Inequalities, dated 5 January 1948, which contained (as far as I know) the first rigorous proof of duality. Later Tucker asked me, "Why didn't you publish it?", to which I replied: "Because it was not my result -- it was von Neumann's. All I did was write up, for internal circulation, my own proof of what von Neumann outlined. It was my way of educating the people in my office In the Pentagon."* (DANTZIG, 1981, p. 6).

Atualmente, von Neumann é considerado o originador da Teoria da Dualidade, embora créditos são dados à Kuhn, Tucker e Gale como os primeiros publicadores de uma prova rigorosa sobre o tema.

Veja alguns exemplos cotidianos envolvendo Programação Linear:

Exemplo 1: Problema da dieta

Suponha a existência de $1, 2, \dots, n$ alimentos que contém $1,2, \dots, n$ nutrientes. Cada unidade do alimento j contém a_{ij} unidades do nutriente i. Sabe-se que são necessárias, no mínimo, b_i unidades do nutriente i em determinado período. Cada unidade do alimento j custa c_j dólares. O problema pode ser escrito como:

Minimizar $\quad c_1x_1 + \cdots + c_nx_n$

Sujeito a $\quad a_{11}x_1 + \cdots + a_{1n}x_n \geq b_1$

$$\vdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n \geq b_n$$

$$x_1 \geq 0, \dots, x_n \geq 0$$

Exemplo 2: Problema do Transporte

Quantidades $a_1, a_2, \dots, a_m$ devem ser transportadas em m locais e recebidos nas quantidades $b_1, b_2, \dots, b_n$ em n destinos. Para transportar 1 unidade do local i para o destino j, o custo é c_{ij}. Sabe-se que x_{ij} representa o número de unidades transportadas do local i para o destino j. O problema pode ser escrito como:

Minimizar $\quad \sum c_{ij}x_{ij}$

Sujeito a $\quad \sum_{j=1}^{n} x_{ij} = a_i \,, \quad i = 1, 2, \dots, m$

$$\sum_{i=1}^{n} x_{ij} = b_j \,, \quad j = 1, 2, \dots, n$$

$$x_{ij} \geq 0$$

5.5 Problemas Multi-objetivos e soluções ótimas de Pareto

Um problema de otimização único pode ser definido como:

Minimizar $f(x)$

Sujeito a $x \in S$

tal que f é a função escalar e S é o conjunto de restrição definido por:

$$S = \{x \in R^m : h(x) = 0, g(x) \geq 0\}.$$

No caso dos problemas multi-objetivos, deve-se encontrar, em termos matemáticos:

Minimizar $[f_1(x), f_2(x), \dots, f_n(x)]$

Sujeito a $x \in S$

onde $n > 1$ e S é o um conjunto de restrições definido acima.

O espaço em que o vetor objetivo pertence é chamado de espaço objetivo, e a imagem do conjunto factível é chamada de conjunto atingível, o qual pode ser representado por:

$$C = \{y \in R^n : y = f(x), x \in S\}.$$

O conceito escalar de ótimo não deve ser diretamente aqui aplicado, devendo, portanto, utilizar as noções do ótimo de Pareto. Essencialmente, o vetor $x^* \in S$ é visto como ótimo de Pareto para um problema multi-objetivo se todos os outros vetores $x \in S$ possuírem um valor mais alto para pelo menos uma função objetivo f_i, para $i = 1, 2, \dots, n$, ou possuírem o mesmo valor para todas as funções objetivos. Mais formalmente:

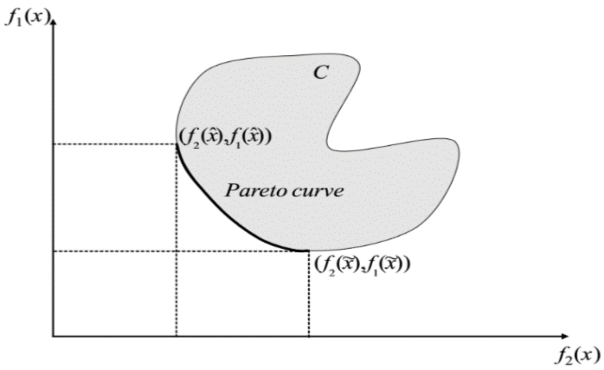
- O ponto x^* é dito fracamente eficiente no sentido de Pareto ou ótimo de Pareto fraco para uma função multi-objetivo se, e somente se não existir $x \in S$ tal que $f_i(x) < f_i(x^*), \forall\, i \in \{1, \dots, n\}$.
- O ponto x^* é dito fortemente (estritamente) eficiente no sentido de Pareto ou ótimo de Pareto estrito para uma função multi-objetivo se, e somente se não existir $x \in S$ tal que $f_i(x) \leq f_i(x^*), \forall\, i \in \{1, \dots, n\}$, com pelo menos uma desigualdade estrita.

(CARAMIA; DELL'OLMO. 2008. *Adaptado*).

Com relação as localizações de um ponto ótimo de Pareto deve-se ficar atento à vizinhança de x^*. Em outras palavras, se $B(x^*, \varepsilon)$ é um círculo de raio $\varepsilon > 0$ em torno do x^*, faz-se necessário que para algum $\varepsilon > 0$, não exista $x \in S \cap B(x^*, \varepsilon)$ tal que $f_i(x) \leq f_i(x^*), \forall\, i \in \{1, \dots, n\}$, com pelo menos uma desigualdade estrita.

A imagem com todos os pontos Pareto eficientes pode ser chamada de curva de Pareto. A forma da curva de Pareto indica a natureza do *trade-off* entre as diferentes funções objetivo. Um exemplo de uma curva de Pareto pode ser visto na Figura 13 abaixo, onde todos os pontos entre $(f_2(\hat{x}), f_1(\hat{x}))$ e $(f_2(\tilde{x}), f_1(\tilde{x}))$ definem a "fachada" da curva. Esses pontos são chamados de pontos não inferiores ou pontos não dominantes.

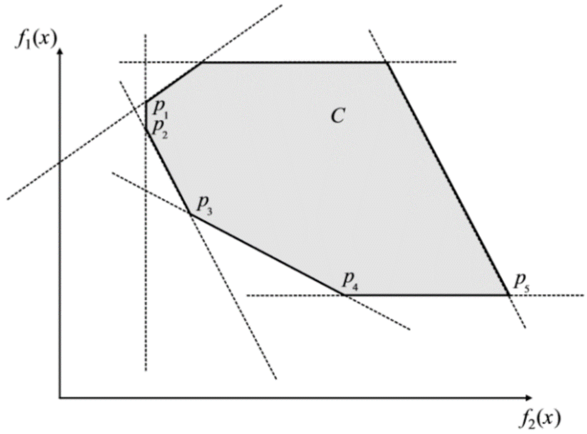
Figura 13: Curva de Pareto



Fonte: CARAMIA, M.; DELL'OLMO, P. Multi-objective management in freight logistics. (2008, p. 13).

Exemplos de pontos de ótimo de Pareto, fraco e estrito, podem ser vistos na Figura 14 a seguir. Os pontos P_1 e P_5 são pontos de ótimo fraco e os pontos P_2, P_3 e P_4 são ótimos estritos.

Figura 14: Pontos de Ótimo de Pareto: fraco e estrito



Fonte: CARAMIA, M.; DELL'OLMO, P. Multi-objective management in freight logistics. (2008, p. 13).

5.6 Programação Linear Multi-objetivo

A definição de Jared L. Cohon, 1978, clarifica ao leitor sobre do que se trata o estudo dos problemas multi-objetivos:

> *Multiobjective programming and planning is concerned with decision-making problems in which there are several conflicting objectives. Such problems are ubiquitous, particularly in the public sector, which must be concerned with society's objectives. (…) "tradeoff" and "conflict resolution" are now part of every public decision maker's vocabulary.* (COHON, 1978, p. 1).

Cohon destaca, também, a importância da utilização dos métodos multi-objetivos:

> *Perhaps the strongest support for the use of multiobjective methods comes from reality: Real-world problems are multiobjective. Public action generally impacts many different groups and social concerns. The imposition of a single-objective approach on such problems is overly restrictive and un-realistic. Multiobjective analysis allows several noncommensurable effects to be treated without artificially combining them. This is clearly a significant improvement in analytical capability.* (COHON, 1978, p. 4).

Problemas envolvendo otimização de diversas funções ao mesmo tempo são frequentes no dia a dia, tanto para o setor público quanto para o setor privado. Um exemplo clássico é o caso do regulador do governo o qual se depara com a produção de carros. O regulador busca um carro que maximiza a segurança, minimiza o consumo de combustível, minimiza a poluição do ar e minimiza o preço final. Obviamente, a união de todos esses requerimentos tende a ser impossível.

Embora possa ser impossível otimizar todas as “frentes”, decisões devem ser tomadas pelo regulador, e algumas decisões são melhores que outras. A tomada de decisão exige compromisso, provável *trade-off* entre as escolhas, e muitas vezes exige a comparação entre coisas incomensuráveis. A análise deste problema envolve o estudo da Programação Linear Multi-objetivo.

Mas afinal, quantos dólares vale uma vida humana? Se a produtora de carro aumentar em x seus custos na produção do carro visando aumentar a segurança, então, poder-se-ia salvar uma vida gastando x dólares. Logo, uma vida vale x dólares?

E se aumentarmos infinitamente a segurança do carro? O carro será extremamente seguro, mas o preço será maior que as pessoas estão dispostas ou até mesmo maior do que as pessoas podem pagar.

Algumas questões, apesar de estranhas num primeiro momento, são “respondidas” todos os dias. Como comparar laranjas e escovas de dentes? Para responder este tipo de questão utiliza-se das preferências próprias. Portanto, decisões são tomadas. É claro que algumas decisões podem parecer ou serem estúpidas. Os economistas separam as decisões eficientes das ineficientes. As eficientes são aquelas que não podem ser melhoradas em determinado aspecto sem piorar em um outro aspecto. Todas as outras são ineficientes.

Jared L. Cohon destaca com clareza em seu livro, 1978, como tais questões preenchem o cotidiano das pessoas:

> […] *Should I take the car or the bus? Well, the bus is cheaper (when the cost of gasoline, maintenance, and insurance are computed for the car), but the car is more convenient, particularly since I should stop at the store on my way home from work. The bus will save energy, but I can listen to the radio in the car. There are probably other attributes or objectives in addition to cost, convenience, energy consumption, and comfort that might be considered in choosing between the car and the bus. The point is that there is no single measure of what is best, like dollars. Instead, there are several measures or objectives of importance and making a decision requires the decision maker (the author, in this case) to*

articulate value judgments, at least implicitly, on the relative importance of the objectives. Problems like this abound. (COHON, 1978, p. 5).

Como já previamente mencionado, sabe-se que os problemas multi-objetivos levam em conta diversas funções objetivas. Relembre o caso do problema da dieta mencionado no problema primal e dual. A ideia era minimizar o custo dado um consumo mínimo de nutrientes, portanto, tem-se uma única função objetivo: minimizar o custo em dólares. Imagine agora a existência de um novo problema da dieta: minimizar o custo em dólares e, concomitantemente, minimizar o número de calorias consumidas (o indivíduo quer perder peso). Tem-se, portanto, duas funções objetivas.

É sabido que para ter uma alimentação mais saudável deve-se, por exemplo, comer mais filet mignon do que bife de hambúrguer. Entretanto, a primeira é mais cara. O consumidor pode consumir apenas filet mignon? Sua restrição orçamentária o inviabiliza? A otimização desse problema pode ser analisada através da Programação Linear Multi-objetivos. Veja um exemplo:

Exemplo[6]: Suponha que um determinado indivíduo tenha como um dos objetivos perder peso, portanto, necessita ter uma dieta balanceada, e como segundo objetivo, reduzir o custo da cesta de alimentação.

Seja a_{ij} a quantidade de nutrientes i no alimento j. Suponha b_i o requerimento diário mínimo do nutriente i. Seja $x_j \geq 0$ uma quantidade possível do alimento j na dieta. Então, exige-se

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i \qquad (1)$$

para satisfazer o requerimento diário mínimo do nutriente i.

Para cada alimento j, tem-se um custo em dólares c_{1j} e um "custo calórico" c_{2j}. Para uma determinada dieta x, ter-se-á um custo total em dólares igual a $\sum c_{1j}x_j$ e um custo calórico total de $\sum c_{2j}x_j$. Logo, o custo total terá dois

[6] Todos os exemplos a seguir e posteriores teoremas e demonstrações foram retirados de FRANKLIN, 2002, mediante adaptações.

componentes: dólares e calorias. Isto é, o custo total é um vetor com dois componentes. Se C é uma matriz com componentes c_{ij} $(i = 1, 2; j = 1, \dots, n)$, então o custo é o vetor Cx.

Para que uma determinada dieta x seja melhor que uma outra dieta x^0, a primeira deve ser ou mais barata sem ser mais calórica ou menos calórica sem ser mais cara. Em termos matemáticos:

$$\sum c_{1j} x_j < \sum c_{1j} x_j^0 \quad \text{e} \quad \sum c_{2j} x_j \leq \sum c_{2j} x_j^0, \tag{2}$$

ou

$$\sum c_{2j} x_j < \sum c_{2j} x_j^0 \quad \text{e} \quad \sum c_{1j} x_j \leq \sum c_{1j} x_j^0. \tag{3}$$

Usando a matriz de custos C, pode-se expressar tais inequações de forma compacta:

$$Cx \leq Cx^0 \text{ , } Cx \neq Cx^0. \tag{4}$$

Assim, x é melhor que x^0, ou seja, x^0 é ineficiente.

Caso nenhuma solução x satisfizer (4), x^0 será a solução eficiente e, portanto:

$$Cx^0 = mínimo. \tag{5}$$

As soluções eficientes de uma programação linear multi-objetivo podem ser expressas na forma canônica como:

$$Ax = b, \;\; x \geq 0, \;\; Cx = mínimo, \tag{6}$$

onde C é uma matriz custo com mais de uma linha.

Novamente, se nenhuma solução factível x satisfizer (4), x^0 será eficiente (ótima), portanto, x^0 resolve o problema convencional:

$$Ax = b, \;\; x \geq 0, \;\; (w^T C)x = min., \tag{7}$$

onde $w^T C$ é uma única linha, sendo todos os componentes do vetor w positivos.

Caso w seja um vetor positivo qualquer e x^0 seja ótimo para o problema convencional (7), então x^0 será uma solução eficiente de (6).

Teorema: O vetor x^0 será uma solução eficiente da *multiobjective program* (6) se, e somente se existir um vetor $w > 0$ para cada x^0 como solução ótima da *single-objective program* (7). (FRANKLIN, 2002. *Adaptado*).

Demonstração: Observe a prova do teorema acima:

Suponha que x^0 seja um ótimo para (7), onde $w > 0$. Suponha $Cx \leq Cx^0, Cx \neq Cx^0$. Então, $w^T(Cx - Cx^0) < 0$, contradizendo que x^0 seja ótimo.

De outra forma, supondo que x^0 seja eficiente para (6), deve-se construir um vetor $w > 0$ para x^0 sendo ótimo para (7). Dado x^0, deve-se considerar Cx^0 como um vetor constante e, então, pode-se definir a programação linear convencional abaixo:

$$\begin{gathered} Ax = b \\ Cx + z = Cx^0 \\ x \geq 0, \qquad z \geq 0 \\ \sum z_i = max. \end{gathered} \tag{8}$$

Uma solução factível para (8) é: $x = x^0, z = 0$. Caso contrário, deve-se ter $\sum z_i > 0$, então x irá satisfazer:

$$Ax = b, \quad x \geq 0, \quad Cx \leq Cx^0, \quad Cx \neq Cx^0, \tag{9}$$

o que é impossível caso x^0 seja eficiente para (7).

Se A tem m linhas e C tem k linhas, é possível particionar o vetor dual como segue:

$$[-y_1, \dots, -y_m, w_1, \dots, w_k] = [-y^T, w^T].$$

Então y e w satisfazem:

$$-y^T A + w^T C \geq 0, \tag{10}$$

$$w^T I \geq [1, \dots, 1], \tag{11}$$

$$-y^T b + w^T(Cx^0) = min. = 0. \tag{12}$$

O mínimo dual iguala zero pois o máximo primal iguala zero em (8).

Das equações $(8), (10)$ e (12), obtém-se:

$$Ax^0 = b,\;\; x^0 \geq 0;\quad y^T A \leq (w^T C), \tag{13}$$
$$(w^T C)x^0 = y^T b.$$

Portanto, x^0 é ótimo para o problema primal.

$$Ax = b,\;\; x \geq 0,\;\; (w^T C)x^0 = min., \tag{14}$$

enquanto y é ótimo para o problema dual:

$$y^T A \leq (w^T C),\;\; y^T b = max. \tag{15}$$

De (14), x^0 resolve o problema convencional (7); por (11), os pesos w_i são todos positivos.

(FRANKLIN, 2002. *Adaptado*).

O significado dos pesos: de acordo com o teorema anterior, uma dieta eficiente deve ser factível e que minimiza uma combinação linear dos custos dos componentes:

$$w_1 \,.\, (total\ de\ dólares) + w_2 \,.\, (total\ de\ calorias) = min. \tag{16}$$

com coeficientes positivos w_1 e w_2. Isto é: $w^T(Cx) = min.,\;\; w > 0$.

Os coeficientes na equação (16) não podem ser adimensionais, uma vez que não se pode adicionar dólares à calorias.

Exemplo: Suponha valores hipotéticos para os coeficientes:

$$w_1 = \frac{300}{dólar}\quad,\quad w_2 = \frac{2}{caloria}\,. \tag{17}$$

Busca-se minimizar a combinação linear adimensional descrita na equação (16):

$$\left(\frac{300}{dólar}\right).(total\ de\ dólares) + \left(\frac{2}{caloria}\right).(total\ de\ calorias) = min. \tag{18}$$

Observe que a adição de \$2 dólares representa o mesmo que adicionar 300 calorias na dieta, logo, 300 calorias valem \$2 dólares.

De modo geral, os valores do pesos representam o custo adimensional do seu relativo. No exemplo acima, $w_1 = 300/dólar$ diz que 300 é o custo adimensional de um dólar.

O custo adimensional relativo pode ser visto como:

$$\frac{valor\ de\ uma\ caloria}{valor\ de\ um\ dólar} = \frac{2}{300} = 0{,}067. \tag{19}$$

Ou seja: 300 calorias valem \$2 dólares ou, analogamente, 1 caloria vale \$0,067 dólar.

O leque de possíveis pesos para os coeficientes: Supondo que os pesos $w_i > 0$ para os componentes do custo $\sum c_{ij}x_j$ e que se resolva o problema convencional de programação linear:

$$Ax = b,\ \ x \geq 0,\ \ (w^T C)x = min., \tag{20}$$

a solução ótima x será uma solução eficiente da programação linear multi-objetivo:

$$Ax = b,\ \ x \geq 0,\ \ Cx = min. \tag{21}$$

De forma análoga, supondo que x^0 resolva $Ax = b, x \geq 0$, e supondo que x^0 seja uma solução eficiente de (21), então sabe-se que x^0 é uma solução eficiente de (20) para algum $w > 0$.

Geralmente, x^0 determina o leque de possibilidades dos possíveis vetores de ponderação (vetores com respectivos pesos). Portanto, a solução eficiente x^0 determina um conjunto Ω o qual para cada $w \in \Omega$, x^0 será uma solução eficiente do problema convencional (20). Logo, o valor do *ratio* w_i/w_j não será fixo, devendo variar dentro de intervalos específicos. Apenas quando se sabe toda a gama de possíveis valores para os pesos do vetor $w \in \Omega$ é que se pode inferir os possíveis preços relativos (*ratio* w_i/w_j factíveis).

O conjunto Ω representa o leque de possibilidades para o peso dos vetores w. Portanto, cada solução eficiente x^0 determina um conjunto de vetores de ponderação.

Suponha que x^0 seja ótimo para o problema convencional (20). Logo, existe um vetor dual y que satisfaz:

$$y^T A \leq w^T C, \quad y^T b \geq w^T C x^0. \tag{22}$$

Reciprocamente, as inequações em (22) implicam que x^0 é ótimo para (20).

Exemplo: Considere o problema multi-objetivo:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} x = \begin{bmatrix} 3 \\ 9 \end{bmatrix}, \qquad x \geq 0,$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 3 & 2 & 1 \\ 4 & 2 & 2 \end{bmatrix} x = min.$$

Observe que os custos possuem quatro componentes, logo, a matriz custo possui quatro linhas. Além disso, observe que uma solução factível pode ser dada por: $x^0 = [1,1,0]^T$.

Busca-se responder duas questões:

1. x^0 é uma solução eficiente?
2. Se sim, determine o conjunto Ω possível para os vetores de ponderação.

Sabe-se que as inequações $y^T A \leq w^T C, y^T b \geq w^T C x^0$, implicam que x^0 é ótimo para $Ax = b, x \geq 0, (w^T C)x = min$.

Aplicando na questão presente, e sabendo que a solução factível deve ser $x^0 = [1,1,0]^T$, obtém-se:

$$y^T A \leq w^T C, \tag{23}$$

$$[y_1, y_2]A = [w^T c^1, w^T c^2],$$

$$[y_1, y_2] \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = [w^T c^1, w^T c^2],$$

onde c^1, c^2 e c^3 são as colunas da matriz custo.

No equilíbrio, x^0 é ótimo para (20) se a solução de equilíbrio y^T satisfaz:

$$[y_1, y_2]\begin{bmatrix}3\\6\end{bmatrix} \leq [w^T c^3]. \tag{24}$$

Eliminando-se y através de (23):

$$[y_1, y_2]AA^{-1} = [w^T c^1, w^T c^2]A^{-1} \tag{25}$$
$$[y_1, y_2] = w^T[c^1, c^2].\, A^{-1}$$
$$[y_1, y_2] = w^T[c^1, c^2].\frac{1}{3}\begin{bmatrix}-5 & 2\\ 4 & -1\end{bmatrix},$$

sendo A^{-1} a inversa de A.

Logo, a inequação é dada por:

$$w^T[c^1, c^2].\frac{1}{3}\begin{bmatrix}-5 & 2\\ 4 & -1\end{bmatrix}.\begin{bmatrix}3\\6\end{bmatrix} \leq w^T c^3 \tag{26}$$

ou

$$w^T[c^1, c^2]\begin{bmatrix}-1\\2\end{bmatrix} \leq w^T c^3 \tag{27}$$

ou ainda

$$w^T[-1c^1 + 2c^2] \leq w^T c^3, \tag{28}$$

onde:

$$-c^1 = \begin{bmatrix}-1\\-1\\-3\\-4\end{bmatrix}, \qquad 2c^2 = \begin{bmatrix}4\\2\\4\\4\end{bmatrix}, \qquad c^3 = \begin{bmatrix}3\\0\\1\\2\end{bmatrix}.$$

A partir da matriz custo, obtém-se:

$$3w_1 + w_2 + w_3 \leq 3w_1 + w_3 + 2w_4, \tag{29}$$

ou, por fim:

$$w_2 \leq 2w_4. \tag{30}$$

Desde que a inequação $w_2 \leq 2w_4$ tenha solução $w > 0$, a solução factível x^0 será eficiente.

O conjunto Ω consiste em todos os quatro-dimensionais vetores w que satisfazem a normalização $w_i \geq 1, i = (1,2,3,4)$, e solucionam a inequação $w_2 \leq 2w_4$.

Portanto, conclui-se que os possíveis vetores de ponderação adimensionais são simplesmente vetores que satisfazem:

$$w_2 \leq 2w_4, \quad w_i > 0, \quad i = (1,2,3,4).$$

Conclui-se que a solução eficiente x^0 nada diz sobre o primeiro e o terceiro componente do custo. Com relação ao segundo e ao quarto componente, pode-se afirmar apenas que: uma unidade do segundo componente vale menos ou o mesmo que duas unidades do quarto componente. Assim, através destes passos, pode-se realizar comparação entre coisas incomensuráveis – dólares e calorias, vida humana e dólares, laranjas e escovas de dentes.

CONCLUSÃO

A Teoria de Otimização possui aplicações em diversas áreas, embora seja especialmente relevante para a Economia. Sob forte contribuição dos matemáticos, os avanços nas áreas de otimização geram avanços no estudo das Ciências Econômicas.

Nesta monografia, procedeu-se a análise dos principais teoremas e demonstrações referentes à Teoria de Otimização em Ciências Econômicas à nível de graduação.

Assim como em qualquer estudo, os problemas relacionados à otimização requerem um *background*, neste caso em Matemática. Para tanto, no capítulo 2, fora apresentado algumas ferramentas clássicas para a melhor compreensão dos capítulos posteriores.

No que diz respeito às condições de otimização, tanto para um máximo quanto para um mínimo, buscou-se analisar, no capítulo 3, as condições de otimização de funções não condicionadas utilizando as condições de primeira e de segunda ordem sobre as derivadas as quais caracterizam os extremos de uma função diferenciável qualquer. Na sequência do capítulo, buscou-se analisar as condições de otimização para funções semelhantes, porém condicionadas a restrições. Considerou-se, portanto, a análise de otimização por diferenciabilidade.

Na área da Microeconomia, um problema clássico é maximizar a função utilidade do consumidor condicionada aos preços dos produtos e a renda do consumidor ou, de forma parecida, minimizar uma função dispêndio condicionada a obter um determinado nível de produção. A fim de solucionar problemas análogos de otimização, o poderoso método – de multiplicadores de Lagrange – para maximizar ou minimizar uma função sujeita a restrições fora desenvolvido pelo matemático e astrônomo franco-italiano Joseph Louis Lagrange (1736-1813).

Em Economia, os problemas de otimização mais comuns referem-se à problemas de maximização condicionadas a restrições de desigualdade e restrições de não-negatividade. Essa situação é tão comum que um lagrangiano “especial” fora criado para tratar deste caso. Esta formulação especial é conhecida como Formulação de Kuhn-Tucker, a qual remonta ao trabalho dos dois matemáticos norte-americanos Albert W. Tucker e Harold W. Kuhn.

Devido ao indubitável avanço que a Formulação de Kuhn-Tucker trouxe para a Teoria de Otimização e, particularmente, para o estudo das Ciências Econômicas, buscou-se, no capítulo 4, apresentar de forma detalhada os resultados obtidos pelos autores.

O presente trabalho teve por objetivo apresentar duas versões de demonstração para o Teorema de Kuhn-Tucker. A primeira demonstração, que é a mais usual nos livros, utiliza o Teorema da Função Implícita. A segunda demonstração, através do uso do Lema de Farkas, é a demonstração original do Kuhn e do Tucker.

Em 1950, o artigo "*Nonlinear Programming*", publicado por Harold Kuhn e Albert Tucker, teve como primeiro objetivo generalizar o resultado estabelecido em Programação Linear – com base nos resultados obtidos por Dantzig-Neumann para problemas não-lineares – de que as soluções dos correspondentes problemas primal e dual são um ponto de sela do lagrangiano associado.

Resultados semelhantes ao Teorema de Kuhn-Tucker já eram conhecidos antes do artigo de Kuhn-Tucker. A novidade do artigo deles está no uso do Lema de Farkas e no fato que apresentar as condições necessárias e suficientes para um máximo global, dada as restrições, pela existência de um ponto de sela para o lagrangiano. Resultado, obviamente, inspirado pelo resultado equivalente obtido em Teoria dos Jogos por George Dantzig sob orientação de von Neumann. Portanto, o artigo de Kuhn e Tucker se insere dentro do grande programa de pesquisa sobre desigualdades lineares orientado por von Neumann na década de 1940 e 1950. Nos anos posteriores a 1951, Kuhn e Tucker continuaram a obter resultados sobre otimização linear, todos apoiados de uma maneira ou outra em alguma proposição derivada do Lema de Farkas. A grande vantagem da demonstração de Kuhn e Tucker é esta apresentar os resultados da Programação Linear e não linear de uma maneira unificada.

Por outro lado, Kuhn e Tucker estendem seu resultado para a Otimização Vetorial, mostrando que as condições a serem satisfeitas por ótimos de Pareto podem ser expressas através das mesmas condições de Kuhn-Tucker com algumas modificações e que dadas as condições de concavidade estas condições são também suficientes.

Ainda no artigo "*Nonlinear Programming*", o segundo objetivo dos autores era reduzir um problema com restrições e várias funções objetivos a um problema com uma função objetivo apenas. Uma nova subdisciplina da Teoria da Otimização, conhecida como Otimização Vetorial, ganhava espaço.

Sob essa influência, e como objetivo final deste trabalho, buscou-se apresentar a Otimização Vetorial, a qual consiste no problema de otimizar, via eficiência no sentido de Pareto, várias funções-objetivos simultaneamente.

Os problemas multi-objetivos estão presentes tanto no setor público quanto no setor privado, sendo de particular interesse quando relacionado às questões da sociedade como um todo. Os conflitos para tomar decisões e o *trade-off* presentes nos problemas multi-objetivos os aproximam da realidade enfrentada no cotidiano das pessoas, o que serve de um forte suporte para o desenvolvimento do estudo da área. Baseado no livro de Jared L. Cohon, 1978, primeiro livro de programação multi-objetivo, buscou-se nesta monografia apresentar de maneira concisa alguns dos problemas hipotéticos, porém recorrentes em nosso dia a dia, e relacioná-los de forma simples com a Otimização Vetorial, a fim de obter uma solução dita eficiente no sentido de Pareto.

O estudo de problemas multi-objetivos busca resolver a abordagem demasiadamente restritiva e não realista das *single-objective*. O curioso deste método é observar a sua capacidade em realizar comparação entre coisas incomensuráveis – dólares e calorias, vida humana e dólares, laranjas e escovas de dentes. A análise multi-objetivo passa a tratá-los sem combiná-los artificialmente, representando uma melhoria significativa na capacidade analítica dos problemas de otimização. A Teoria da Otimização mostra-se aplicável em qualquer área de estudo.

REFERÊNCIAS

ANTON, H.; BIVENS, I.; DAVIS, S. **Cálculo.** 8 ed. Porto Alegre: Bookman, 2007. Reimpressão 2010, v. 1 e v. 2.

ARROW, K. J.; ENTHOVEN, A. Quase-Concave Programming. **Econometrica**, v. 29, n. 4, 1961.

ARROW, K. J.; INTRILIGATOR, M. D. **Handbook of mathematical economics.** Amsterdam: North-Holland, c1981. v. 1.

BOLDRINI, J.; COSTA, S; FIGUEIREDO, V.; WETZLER, H. **Álgebra Linear.** 3 ed. São Paulo: Harbra, 1986.

CARAMIA, M.; DELL'OLMO, P. **Multi-objective management in freight logistics.** Increasing capacity, service level and safety with optimizations algorithms. London: Springer-Verlag London, 2008.

COHON, J. L. **Multiobjective programming and planning.** New York: Academic Press, 1978.

DANTZIG, G. B. Reminiscences about the origins of linear programming. **Technical Report SOL 81-5**. Standford, California, 1981.

DEBREU, G. **Mathematical economics:** twenty papers of Gerard Debreu. Cambrige: Cambridge University Press, 1995, 1986.

FRANKLIN, J. N. **Methods of mathematical economics:** Linear and nonlinear programming, Fixed-Points Theorems. Philadelphia, PA: SIAM, 2002.

KUHN, H. W; TUCKER, A.W. **Nonlinear programming.** Proceedings of the second Berkeley symposium on mathematical statistics and probability: p. 481-492, Berkeley: University of California Press, 1950.

LAY, D. C. **Álgebra Linear e suas aplicações.** 2 ed. Rio de Janeiro: LTC, 2007.

NICHOLSON, W.; SNYDER, C. **Microeconomic theory:** basic principles and extensions. 11 ed. Mason, OH: Cengage Learning, 2010.

SIMON, C. P.; BLUME, L. **Matemática para economistas.** Porto Alegre: Bookman, 2004. Reimpressão 2008.

VARIAN, H. R. **Microeconomia:** princípios básicos. 7 ed. Rio de Janeiro: Elsevier, 2006.