



TEXTO PARA DISCUSSÃO N°

***CRASH COURSE* EM MATEMÁTICA PARA ECONOMISTAS**

**André Braz Golgher
Luccas Assis Attílio
Helder Lara Ferreira Filho**

julho de 2016

Universidade Federal de Minas Gerais

Jaime Arturo Ramírez (Reitor)
Sandra Regina Goulart Almeida (Vice-reitora)

Faculdade de Ciências Econômicas

Paula Miranda-Ribeiro (Diretora)
Lizia de Figueirêdo (Vice-diretora)

Centro de Desenvolvimento e Planejamento Regional (Cedeplar)

Cássio Maldonado Turra (Diretor)

José Irineu Rangel Rigotti (Coordenador do
Programa de Pós-graduação em Demografia)

Marco Flávio da Cunha Resende (Coordenador do
Programa de Pós-graduação em Economia)

Laura Lúcia Rodríguez Wong (Chefe do
Departamento de Demografia)

Edson Paulo Domingues (Chefe do Departamento
de Ciências Econômicas)

Editores da série de Textos para Discussão

Aline Souza Magalhães (Economia)

Secretaria Geral do Cedeplar

Maristela Dória (Secretária-Geral)
Simone Basques Sette dos Reis (Editoração)

<http://www.cedeplar.ufmg.br>

Textos para Discussão

A série de Textos para Discussão divulga resultados preliminares de estudos desenvolvidos no âmbito do Cedeplar, com o objetivo de compartilhar ideias e obter comentários e críticas da comunidade científica antes de seu envio para publicação final. Os Textos para Discussão do Cedeplar começaram a ser publicados em 1974 e têm se destacado pela diversidade de temas e áreas de pesquisa.

G625c Golgher, André Braz.
2016 *Crash course* em matemática para economistas / André Braz Golgher, Luccas Assis Attílio, Helder Lara Ferreira Filho. - Belo Horizonte : UFMG/CEDEPLAR, 2016.
Inclui bibliografia (p. 26)

ISSN 2318-2377

1. Matemática. 2. Álgebra linear. I. Attílio, Luccas Assis. II. Ferreira Filho, Helder Lara. III. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. IV. Título. V. Série.

CDD: 510

Ficha catalográfica elaborada pela Biblioteca da
FACE/UFMG - JN063/2016

As opiniões contidas nesta publicação são de exclusiva responsabilidade do(s) autor(es), não exprimindo necessariamente o ponto de vista do Centro de Desenvolvimento e Planejamento Regional (Cedeplar), da Faculdade de Ciências Econômicas ou da Universidade Federal de Minas Gerais. É permitida a reprodução parcial deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções do texto completo ou para fins comerciais são expressamente proibidas.

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the publishers. The reproduction of parts of this paper of or data therein is allowed if properly cited. Commercial and full text reproductions are strictly forbidden.

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL**

***CRASH COURSE* EM MATEMÁTICA PARA ECONOMISTAS**

André Braz Golgher

Cedeplar/FACE/UFGM

Luccas Assis Atílio

Cedeplar/FACE/UFGM

Helder Lara Ferreira Filho

Cedeplar/FACE/UFGM

**CEDEPLAR/FACE/UFGM
BELO HORIZONTE
2016**

RESUMO

Este livro é composto de 10 capítulos que pretendem discutir de forma breve alguns dos pontos que normalmente são lecionados em cursos de Matemática para Economistas. Os capítulos são: 1 – Álgebra linear; 2 – Cálculo diferencial de várias variáveis; 3 – Cálculo Integral; 4 - Equações de diferença e equações diferenciais de 1ª ordem; 5 – Sistemas de equações de diferença; 6 – Sistemas de equações diferenciais; 7 – Elementos de Análise; 8 – Otimização não condicionada; 9 – Otimização condicionada; e 10 – Funções côncavas e quase côncavas. Assume-se que o leitor já esteja familiarizado com tópicos básicos da álgebra matricial, normalmente lecionados no Ensino Médio, e que também já tenha cursado uma disciplina introdutória de cálculo diferencial e integral. Além disso, um conhecimento elementar da teoria microeconômica facilita o entendimento dos capítulos.

Palavras-chave: Matemática para economistas.

ABSTRACT

This book is a set of ten chapters that briefly discuss topics normally presented in courses of mathematics for economists. These chapters are: 1 – Linear algebra; 2 – Differential calculus of many variables; 3 – Integral calculus; 4 - First order differential and difference equations; 5 – Systems of first order difference equations; 6 – Systems of first order differential equations; 7 – Elements of analysis; 8 – Unrestricted optimization; 9 – Restricted optimization; and 10 – Concave and quasi-concave functions. Knowledge of high school linear algebra and a previous elementary course of differential and integral calculus are required in order to better understand the concepts discussed in the book. Besides, familiarity with some of the basics concepts of microeconomics also helps.

Key-words: Mathematics for economists.

JEL: C00

INTRODUÇÃO

Este livro contém 10 capítulos que pretendem discutir de forma bastante breve alguns dos pontos que normalmente são lecionados em cursos de Matemática para Economistas tanto em nível de graduação como de pós-graduação. Como os capítulos são bastante resumidos e, conseqüentemente, incompletos, utilizou-se o termo em inglês *Crash Course*. O objetivo é discutir cada um dos 10 pontos de forma concisa e de modo mais abrangente possível. São eles: 1 – Álgebra linear; 2 – Cálculo diferencial de várias variáveis; 3 – Cálculo Integral; 4 - Equações de diferença e equações diferenciais de 1ª ordem; 5 – Sistemas de equações de diferença; 6 – Sistemas de equações diferenciais; 7 – Elementos de Análise; 8 – Otimização não condicionada; 9 – Otimização condicionada; e 10 – Funções côncavas e quasecôncavas.

Os capítulos dialogam entre si, sendo que os capítulos servem de base para os subsequentes e a complexidade das discussões aumenta paulatinamente. Todo o material apresentado aqui é amplamente discutido de forma muito mais abrangente e cuidadosa em diferentes livros-texto. A bibliografia utilizada na confecção desse texto inclui livros de Matemática para Economistas (Simon e Blume, 1994; Chiang e Wainwright, 2006; Golgher e Vidal, 2008), de Cálculo Diferencial e Integral (Stewart, 2006a, b), de Análise Matemática (Lima, 1995 e 2004), de Teoria Microeconômica (Mas-Colellet al, 1995; Kreps, 1990) e de sistemas dinâmicos (Banks, 1994; Gandolfo, 1997). Muitos dos problemas e demonstrações aqui apresentadas também estão presentes nesses livros-texto.

CAPÍTULO 1 – ÁLGEBRA LINEAR

A álgebra linear é amplamente utilizada em Economia em diversos problemas que aplicam matrizes e vetores. Por exemplo, muitos dentre os modelos econômicos mais estudados têm uma estrutura linear e são escritos como sistemas de equações lineares. Mesmo quando a relação entre as variáveis econômicas é não linear, as equações podem ser aproximadas por um sistema linear fazendo uso do cálculo diferencial. A álgebra linear fornece várias técnicas para resolver sistemas lineares, sendo uma poderosa ferramenta para o estudo de modelos econômicos, como veremos em diferentes capítulos que utilizam os conceitos discutidos nesse capítulo inicial.

Este capítulo pretende introduzir algumas das ferramentas da Álgebra Linear em um texto sintético, mas que objetiva ser abrangente, como todos os demais capítulos. Para tanto, o capítulo foi dividido nos seguintes tópicos: i) Sistema de equações lineares; ii) Posto de uma matriz; iii) Tipos especiais de matrizes; iv) A matriz inversa e sistemas lineares; v) Determinantes; vi) Utilização do determinante: invertendo matrizes; vii) Utilização do determinante: regra de Cramer; viii) Aplicações em economia: a matriz de insumo/produto; ix) Autovalores e autovetores; e x) Produto interno. Assume-se que o leitor já esteja familiarizado com tópicos básicos da álgebra matricial, normalmente lecionados no Ensino Médio, como soma e subtração de matrizes, multiplicação de matrizes por escalares, multiplicação de matrizes e operações elementares com linhas e colunas de matrizes.

1. SISTEMAS DE EQUAÇÕES LINEARES

Uma equação linear tem a seguinte forma geral $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$, onde a_i são parâmetros e x_i são variáveis.

Um sistema de equações lineares contém equações lineares e pode ser escrito de forma geral como:

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\&\dots \\a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m\end{aligned}$$

Sistemas de equações lineares podem ser expressos de forma eficaz em formato de matriz da seguinte forma:

$AX = b$, em que:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \text{ é a matriz dos coeficientes, } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ é o vetor das variáveis e } b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

é o vetor resposta.

Como forma de ilustração dos conceitos discutidos no capítulo apresentamos exemplos numéricos com sistemas com duas equações e duas incógnitas. Tome, por exemplo, o sistema de equações lineares:

$$x + 2y = 1$$

$$3x - y = 0$$

Esse sistema pode ser escrito em formato de matriz como:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}, X = \begin{bmatrix} x \\ y \end{bmatrix} \text{ e } b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Note que o sistema satisfaz a relação $AX = b$.

Outra forma de representação de sistemas lineares é por meio da matriz aumentada, \hat{A} . Essa matriz é obtida acrescentando à matriz dos coeficientes, A , uma última coluna com os elementos do vetor resposta, b :

$$\hat{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}$$

No caso do exemplo acima, teríamos:

$$\hat{A} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & 0 \end{bmatrix}.$$

Um sistema linear pode ser resolvido pela manipulação das linhas de matrizes de diferentes maneiras.

Uma prática bastante comum é escrever a matriz aumentada, \hat{A} , no formato escalonado por linhas.

Segue a definição de matriz escalonada por linhas.

Uma matriz escalonada por linhas é aquela em que cada linha tem mais zeros à esquerda do que a anterior.

Segue um exemplo para um sistema genérico de equações lineares com três equações e três incógnitas, que é representado pela matriz aumentada a seguir:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{bmatrix}.$$

Manipulando linhas, chega-se a matriz escalonada a seguir, onde a primeira linha foi mantida inalterada:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ 0 & c_{22} & c_{23} & d_2 \\ 0 & 0 & c_{33} & d_3 \end{bmatrix}.$$

A partir dessa matriz, obtém-se a solução do sistema:

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{b_1}{a_{11}} - a_{12}(y) - a_{13}(z) \\ \frac{d_2}{c_{22}} - \frac{c_{23}}{c_{22}}(z) \\ \frac{d_3}{c_{33}} \end{bmatrix}.$$

Segue um exemplo numérico com o mesmo sistema já descrito. Parte-se da matriz aumentada:

$$\hat{A} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & 0 \end{bmatrix}$$

Em seguida, essa matriz é escalonada. Para tanto, a linha 1 é mantida inalterada. Uma nova linha 2, l_2^* , é obtida a partir da manipulação das linhas 1, l_1 , e linha 2, l_2 , $l_2^* = l_2 - 3l_1$:

$$\hat{A}_{\text{escalonada}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -3 \end{bmatrix}$$

Obtêm-se, assim, os valores de $y = \frac{3}{7}$ e, em seguida, o de $x = \frac{1}{7}$.

2. POSTO DE UMA MATRIZ

Um sistema de equações lineares pode ter ou não solução. Caso tenha, um sistema pode ter uma única ou ∞ soluções. Como podemos determinar qual é o caso de um sistema específico? O critério fundamental para responder isso é o posto de uma matriz.

Posto - O posto de uma matriz é o número de linhas não nulas quando esta está escrita na sua forma escalonada por linhas. Uma linha é nula se todos os seus elementos são zero.

Note que, se uma linha da matriz puder ser escrita como combinação linear das demais, $l_n = a_1 l_1 + a_2 l_2 + \dots + a_{n-1} l_{n-1}$, necessariamente quando as linhas da matriz forem manipuladas e a matriz escalonada for obtida, uma das linhas será nula. Diz-se que as linhas não são independentes.

Antes de discutirmos a aplicabilidade do posto de uma matriz como critério definidor do número de soluções, devem-se diferenciar os sistemas homogêneos dos não homogêneos. Os sistemas homogêneos são aqueles em que todos os elementos do vetor resposta são nulos. Caso contrário, se pelo menos um elemento desse vetor for diferente de zero, o sistema é não homogêneo.

Os sistemas que têm maior aplicabilidade em economia são os que possuem o mesmo número de equações e incógnitas. Ou seja, o número de linhas é igual ao de colunas na matriz dos coeficientes. Por brevidade, trataremos apenas desse tipo de sistema.

Fato 1 - Um sistema tem solução se o posto da matriz dos coeficientes for igual ao posto da matriz aumentada.

Fato 2 – Para um sistema homogêneo, como a última coluna da matriz aumentada é nula, o posto da matriz dos coeficientes é sempre igual ao posto da matriz aumentada. Isto é, esse tipo de sistema sempre tem solução, que é a solução trivial.

Fato 3 – Se o posto da matriz for máximo, ou seja, se for igual ao número de linhas e colunas da matriz dos coeficientes, o sistema tem uma única solução. Dizemos que a matriz é **não singular**. Nesse caso, as linhas (e colunas) não podem ser escritas como combinação linear das demais, isto é, as linhas e, portanto, as equações do sistema, são independentes.

Fato 4 – Se o posto da matriz dos coeficientes não for máximo, existem duas possibilidades: o sistema pode ter ∞ soluções, se o posto dessa matriz for igual ao da matriz aumentada; ou zero soluções, caso isso não ocorra.

Seguem três exemplos de sistemas com duas equações e duas incógnitas, todos eles não homogêneos, que servirão de ilustração para esses conceitos discutidos. O primeiro exemplo é o apresentado na seção anterior:

$$\begin{aligned}x + 2y &= 1 \\ 3x - y &= 0\end{aligned}$$

Como vimos, a matriz aumentada desse sistema na forma escalonada é:

$$\hat{A}_{\text{escalonada}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & -7 & -3 \end{bmatrix}.$$

Portanto, o posto da matriz aumentada, isto é, o número de linhas não nulas, é igual ao posto da matriz dos coeficientes, e é máximo: $P(\hat{A}) = P(A) = 2$.

Pelo fato 1, o sistema tem solução e, pelo fato 3, ele tem uma única solução, como já foi obtido na seção anterior.

O segundo exemplo é composto da primeira equação do sistema anterior e dessa mesma equação multiplicada por dois:

$$\begin{aligned}x + 2y &= 1 \\ 2x + 4y &= 2\end{aligned}$$

Ou seja, as linhas da matriz não são independentes. Esse sistema tem como matriz aumentada na forma escalonada:

$$\hat{A}_{\text{escalonada}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

A segunda linha é toda nula, pois é um múltiplo da primeira. Portanto, o posto da matriz aumentada é igual ao posto da matriz dos coeficientes, porém não é máximo: $P(\hat{A}) = P(A) = 1$.

Pelo fato 1, o sistema tem solução e, pelo fato 4, ele tem infinitas soluções. Na verdade, qualquer relação entre x e y que satisfaz qualquer das equações.

O último exemplo com sistemas não homogêneos é similar ao anterior com uma mudança no vetor resposta:

$$\begin{aligned}x + 2y &= 1 \\ 2x + 4y &= 1\end{aligned}$$

Esse sistema tem como matriz aumentada na forma escalonada:

$$\hat{A}_{\text{escalonada}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

A segunda linha não é toda nula. O posto da matriz aumentada difere do posto da matriz dos coeficientes: $P(\hat{A}) = 2 \neq P(A) = 1$. Pelo fato 1, o sistema não tem solução.

Uma vez apresentados os exemplos para sistemas não homogêneos, seguem dois exemplos com sistemas homogêneos. O primeiro deles é similar ao primeiro dos exemplos não homogêneos:

$$\begin{aligned}x + 2y &= 0 \\ 3x - y &= 0\end{aligned}$$

A matriz aumentada na forma escalonada é:

$$\hat{A}_{\text{escalonada}} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -7 & 0 \end{bmatrix}.$$

Pelos fatos 2 e 3, o sistema tem uma única solução, que é a solução trivial: $x = y = 0$.

O segundo exemplo de sistema homogêneo é similar ao segundo (ou ao terceiro) dos exemplos não homogêneos:

$$\begin{aligned} x + 2y &= 0 \\ 2x + 4y &= 0 \end{aligned}$$

O sistema é representado na forma de matriz aumentada escalonada como:

$$\hat{A}_{\text{escalonada}} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Pelos fatos 2 e 4, o sistema tem infinitas soluções, além da solução trivial: quaisquer valores de x e de y que satisfaçam quaisquer das equações.

3. TIPOS ESPECIAIS DE MATRIZES

Alguns tipos de matrizes aparecem com certa regularidade em problemas econômicos ou servem de base para o desenvolvimento de conceitos em álgebra linear com aplicabilidade prática. Alguns desses são citados a seguir, sempre acompanhados de um exemplo 2 x 2.

Matriz quadrada – Uma matriz com o mesmo número de linhas e colunas:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$$

Matriz diagonal – Uma matriz quadrada com todos os elementos não diagonais iguais a zero:

$$B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Matriz identidade – Uma matriz identidade, escrita como **I**, é uma matriz diagonal com todos os elementos da diagonal principal iguais a 1:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Note que $AI = IA = A$:

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}.$$

Matriz triangular – Matriz normalmente quadrada com todos os elementos acima ou abaixo da diagonal principal iguais a zero.

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$$

Matriz nula – Uma matriz com todos os elementos iguais a zero. Escreve-se $A = 0$.

$$0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Matriz transposta – Uma matriz é transposta de outra quando a linha dessa última se transforma na coluna correspondente da primeira. Escreve-se a transposta de A como A^T ou A' .

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} \text{ e } A^T = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}$$

Matriz simétrica – Uma matriz A é simétrica quando $A = A^T$.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \text{ e } A^T = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$$

4. A MATRIZ INVERSA E SISTEMAS LINEARES

O foco da discussão desta seção são as matrizes quadradas. Essas representam sistema com o mesmo número de equações e de incógnitas. Caso o posto da matriz seja máximo, como vimos, existe somente uma solução. Seguem alguns pontos referentes à álgebra matricial de matrizes quadradas.

Matriz inversa – Seja A uma matriz quadrada, se a inversa de A , escrita como A^{-1} , existe, então: $AA^{-1} = A^{-1}A = I$.

A matriz inversa de A existirá se o seu posto for máximo, ou seja, se a matriz for não singular.

Vimos que um sistema linear não homogêneo pode se escrito como $AX = b$. Note que se a matriz inversa existe, temos uma única solução para o sistema, como demonstrado a seguir:

$$\begin{aligned} AX &= b \\ A^{-1}AX &= A^{-1}b \\ IX &= A^{-1}b \\ X &= A^{-1}b \end{aligned}$$

Para sistemas homogêneos, também temos uma única solução se A^{-1} existir:

$$\begin{aligned} AX &= 0 \\ A^{-1}AX &= 0 \\ IX &= 0 \\ X &= 0, \text{ que é a solução trivial.} \end{aligned}$$

Esta seção apresentou o conceito de matriz inversa. Essa matriz apresenta grande aplicabilidade na resolução de sistemas lineares, como mostrado nesta seção. As seções seguintes discutem como determinas se essa matriz existe e, caso ela exista, como obtê-la.

5. DETERMINANTES

As matrizes quadradas não singulares representam sistemas não homogêneos com uma única solução e tem grande aplicabilidade em problemas econômicos. Como determinar de forma direta se uma matriz tem inversa?

Para tanto, foi definido um número, conhecido como determinante, que quando diferente de zero, indica que a matriz é não singular. Representamos o determinante da matriz A como $\det(A)$ ou $|A|$.

Inicia-se a discussão com matrizes 1×1 . Uma matriz $A = (a_{11})$ terá inversa se existir uma matriz, também 1×1 , tal que $AA^{-1} = I = (1)$. Ou seja, $A^{-1} = (1/a_{11})$. Essa matriz existirá se $a_{11} \neq 0$, portanto, se define $|A| = a_{11}$.

Para matrizes maiores, segue uma definição formal de determinante:

Determinante – O determinante de uma matriz $A_{n \times n}$ é dado por $\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in}$, onde i é uma linha (ou coluna) qualquer, a_{ij} são os elementos da linha i e C_{ij} são os cofatores de i .

Cofator – O cofator C_{ij} é definido como $C_{ij} \equiv (-1)^{i+j} \det(A_{ij})$, onde A_{ij} é uma submatriz de A obtida pela eliminação da i -ésima linha e j -ésima coluna.

Segue um exemplo para uma matriz 2×2 genérica, seguido de um exemplo numérico. Dada uma matriz genérica:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Escolhe-se qualquer uma das linhas, por exemplo, a primeira. Então, pela definição de determinante:

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} = a_{11}(-1)^2 a_{22} + a_{12}(-1)^3 a_{21} = a_{11}a_{22} - a_{12}a_{21}.$$

Como exemplo numérico tome a matriz dos coeficientes do primeiro exemplo:

$$|A| = \begin{vmatrix} 1 & 2 \\ 3 & -1 \end{vmatrix} = (1)(-1) - (2)(3) = -7 \neq 0.$$

Como o determinante dessa última matriz é diferente de zero, A^{-1} existe.

O mesmo raciocínio utilizado para calcular o determinante de uma matriz 2×2 , a partir de determinantes de matrizes 1×1 , pode ser aplicado para matrizes maiores. Determinantes de matrizes 2×2 são utilizados para o cálculo de determinantes de matrizes 3×3 e assim sucessivamente.

6. UTILIZAÇÃO DO DETERMINANTE: INVERTENDO MATRIZES

Esta seção se baseia na discussão anterior e apresenta uma metodologia de como inverter matrizes. Como vimos, uma matriz tem inversa se seu determinante for diferente de zero. Por definição, o determinante de uma matriz $A_{n \times n}$ pode ser obtido com qualquer uma das linhas da matriz original. Assim, ele é dado por qualquer das relações a seguir:

$$\det(A) = a_{11}C_{11} + \dots + a_{1n}C_{1n}$$

$$\det(A) = a_{21}C_{21} + \dots + a_{2n}C_{2n}$$

.....

$$\det(A) = a_{n1}C_{n1} + \dots + a_{nn}C_{nn}$$

Ou seja, **(1)** $a_{j1}C_{i1} + \dots + a_{jn}C_{in} = \det(A)$ para todo $i = j$.

Designa-se uma matriz $B_{n \times n}$ cujas linhas são exatamente iguais às linhas da matriz A , com exceção de uma delas, a linha particular j . Essa linha de B é igual à outra qualquer da própria matriz. Assim, dado que duas linhas da matriz B são iguais, as linhas não são independentes, o posto não é máximo, e $\det(B) = 0$. Então, tomando essa mesma linha j como referência, temos:

$$\det(B) = b_{j1}C_{j1} + \dots + b_{jn}C_{jn} = 0.$$

Note que os elementos b_{ji} tiveram como origem uma linha i qualquer da matriz A , com $i \neq j$. Ou seja, pode ser qualquer uma das linhas desde que essa seja diferente de j . Além disso, os cofatores C_{jk} são os mesmos nas duas matrizes, uma vez que elas diferem apenas na linha j . Assim, para todo $i \neq j$, com i e j quaisquer, ficamos com **(2)** $a_{i1}C_{j1} + \dots + a_{in}C_{jn} = 0$.

De posse de **(1)** e **(2)**, podemos escrever a seguinte relação entre matrizes:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \dots & \dots & \dots & \dots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix} = \begin{pmatrix} \det(A) & 0 & \dots & 0 \\ 0 & \det(A) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \det(A) \end{pmatrix}$$

Essa relação pode ser escrita como:

$$A_{n \times n} C_{n \times n}^T = I \det(A)$$

$$A^{-1} A C^T = A^{-1} I \det(A)$$

$$C^T = A^{-1} \det(A)$$

$$A^{-1} = \left(\frac{1}{\det(A)} \right) C^T = \left(\frac{Adj(A)}{\det(A)} \right), \text{ onde } C^T = Adj(A), \text{ essa última denominada adjunta de } A.$$

Essa expressão é utilizada na obtenção da matriz inversa.

Segue um exemplo para uma matriz 2 x 2 genérica:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Sabemos que $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

Além disso, temos:

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}.$$

Assim, obtemos:

$$A^{-1} = \left(\frac{1}{\det(A)} \right) C^T = \left(\frac{1}{a_{11}a_{22} - a_{12}a_{21}} \right) \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}.$$

Como exemplo numérico, tome sistema com $A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$, $X = \begin{bmatrix} x \\ y \end{bmatrix}$ e $b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Vimos que $\det(A) = -7$.

Assim, obtemos:

$$A^{-1} = \left(\frac{1}{-7} \right) \begin{pmatrix} -1 & -2 \\ -3 & 1 \end{pmatrix}.$$

$$\text{Note que } AA^{-1} = \left(\frac{1}{-7} \right) \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} -1 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Uma vez obtida a matriz inversa, A^{-1} , resolve-se facilmente um sistema de equações lineares não homogêneo, $AX = b$, a partir da relação $X = A^{-1}b$:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/7 & 2/7 \\ 3/7 & -1/7 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/7 \\ 3/7 \end{pmatrix},$$

como já sabíamos.

7. UTILIZAÇÃO DO DETERMINANTE: REGRA DE CRAMER

Outro método muito utilizado na resolução de sistemas lineares é a regra de Cramer. Como vimos na seção anterior, um sistema de equações lineares, cuja matriz dos coeficientes tem inversa, tem como solução:

$$X = A^{-1}b = \left(\frac{1}{\det(A)}\right)C^T b = \left(\frac{1}{\det(A)}\right) \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \dots & \dots & \dots & \dots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}.$$

Assim, o termo x_j do vetor das variáveis X pode ser obtido como:

$$x_j = \left(\frac{1}{\det(A)}\right) \begin{pmatrix} C_{1j} & C_{2j} & \dots & C_{nj} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} = \frac{b_1 C_{1j} + \dots + b_n C_{nj}}{\det(A)}.$$

Tome a definição de determinante de uma matriz qualquer, $\det(D) = d_{i1}C_{i1} + d_{i2}C_{i2} + \dots + d_{in}C_{in}$. O numerador da expressão acima representa o determinante de uma matriz cujos elementos da linha (ou coluna) escolhida são os mesmos que os elementos do vetor resposta. Além disso, os cofatores são exatamente iguais ao da matriz dos coeficientes.

Assim, essa matriz cujo determinante é calculado no numerador da expressão anterior difere da matriz A apenas na j -ésima coluna e cuja j -ésima coluna é o vetor resposta, b . Assim, utiliza-se a nomenclatura B_j para essa matriz e obtemos a seguinte expressão:

$$x_j = \frac{\det(B_j)}{\det(A)}.$$

Novamente, segue um exemplo genérico e um numérico para um sistema com duas equações e duas incógnitas. Dado o sistema,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ e } b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

pela regra de Cramer, temos:

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\det(A)} e$$

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\det(A)}.$$

Como exemplo numérico, utiliza-se o mesmo exemplo já usado:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}, X = \begin{bmatrix} x \\ y \end{bmatrix} e b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Pela regra de Cramer:

$$x = \frac{\begin{vmatrix} 1 & 2 \\ 0 & -1 \end{vmatrix}}{-7} = \frac{1}{7}$$

$$y = \frac{\begin{vmatrix} 1 & 1 \\ 3 & 0 \end{vmatrix}}{-7} = \frac{3}{7}.$$

8. APLICAÇÕES EM ECONOMIA: A MATRIZ DE INSUMO/PRODUTO

Como foi discutido na introdução desse capítulo, a álgebra linear pode ser aplicada em vários problemas econômicos. Será apresentada uma dessas aplicações, que é a matriz de insumo/produto, ou sistema de Leontief aberto. Posteriormente, em capítulos subsequentes, serão discutidas outras aplicações.

Por simplicidade, discute-se o sistema de Leontief aberto para dois insumos/produtos. Para que haja uma comparabilidade entre os produtos, as quantidades são expressas em unidades monetárias.

Existem dois produtos cujas quantidades totais produzidas são x_1 e x_2 . Parte da produção de cada um desses produtos é utilizada como insumo para a própria produção e também como insumo para a produção do outro produto. Tem-se, portanto, a_{ij} , a quantidade do bem i utilizada para se produzir uma unidade de j , onde $a_{ij} \geq 0$ e $\sum_{j=1}^2 a_{ij} < 1$, pois caso contrário o sistema não seria lucrativo. A parte não consumida endogenamente como insumo para a produção dos próprios bens é consumida exogenamente. A demanda exógena é representada por C_1 e C_2 .

O sistema abaixo representa essa aplicação. Por exemplo, a produção do primeiro produto serve como insumo para a produção própria, como insumo para a produção do segundo produto e para o consumo exógeno. O mesmo vale para o produto 2:

$$\begin{aligned}x_1 &= a_{11}x_1 + a_{12}x_2 + C_1 \\x_2 &= a_{21}x_1 + a_{22}x_2 + C_2\end{aligned}$$

Esse sistema pode ser escrito em formato de matriz como:

$$X = AX + C, \text{ onde } X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ e } C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}.$$

A matriz A é conhecida como matriz tecnológica e é considerada fixa para um determinado período. Rearranjando esse sistema com álgebra matricial, temos:

$$(I - A)X = C.$$

A solução é obtida a partir da inversa de $(I - A)$:

$$X = (I - A)^{-1}C.$$

Segue um exemplo numérico com o seguinte sistema:

$$\begin{aligned}x &= 0,5x + 0,2y + 100 \\y &= 0,3x + 0,4y + 200\end{aligned}$$

Note que todos os elementos são maiores ou iguais a zero e que a soma dos elementos em cada coluna é menor que um.

Assim, temos:

$$\begin{aligned}A &= \begin{pmatrix} 0,5 & 0,2 \\ 0,3 & 0,4 \end{pmatrix} \\I - A &= \begin{pmatrix} 0,5 & -0,2 \\ -0,3 & 0,6 \end{pmatrix} \\(I - A)^{-1} &= \begin{pmatrix} 1 \\ 0,24 \end{pmatrix} \begin{pmatrix} 0,6 & 0,2 \\ 0,3 & 0,5 \end{pmatrix} \\ \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 1 \\ 0,24 \end{pmatrix} \begin{pmatrix} 0,6 & 0,2 \\ 0,3 & 0,5 \end{pmatrix} \begin{pmatrix} 100 \\ 200 \end{pmatrix} = \begin{pmatrix} 1 \\ 0,24 \end{pmatrix} \begin{pmatrix} 100 \\ 130 \end{pmatrix} \approx \begin{pmatrix} 417 \\ 542 \end{pmatrix}.\end{aligned}$$

O total do produto x produzido é de aproximadamente 417. Destes, $(0,5)(417) = 208$ são utilizados como insumo para a produção do próprio bem. Outros $(0,2)(542) = 108$ são utilizados como insumo para a produção do bem y . O restante, 100, é o consumo exógeno do bem 1. A análise é semelhante para o bem 2.

9. AUTOVALORES E AUTOVETORES

Autovalores e autovetores são ferramentas extremamente poderosas no estudo de modelos dinâmicos, como será discutido nos capítulos cinco e seis, e em problemas de otimização, como na definição de matrizes, como apresentado no oitavo capítulo. Neste primeiro capítulo são apresentados esses conceitos e as aplicações desses são discutidas nesses capítulos subsequentes.

Segue a definição de autovalor.

Autovalores – Seja A uma matriz quadrada. O autovalor de A é um número, comumente representado por λ , que subtraído da diagonal principal zera o determinante. Ou seja: $|A - \lambda I| = 0$.

Seguem três exemplos para matrizes 2×2 .

Exemplo 9.1 - $A = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix}$

Então, utilizando a definição de autovalor, temos:

$$|A - \lambda I| = \begin{vmatrix} 4 - \lambda & 1 \\ -2 & 1 - \lambda \end{vmatrix} = 0$$

Daí resulta que:

$$(4 - \lambda)(1 - \lambda) + 2 = 0$$

$$\lambda^2 - 5\lambda + 6 = 0$$

$$\lambda_1 = 2$$

$$\lambda_2 = 3$$

Ou seja, foram obtidos dois autovalores reais distintos. Existem ainda duas outras possibilidades para matrizes 2×2 que são duas raízes complexas ou uma raiz dupla real.

Exemplo 9.2 - $B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$

Seguindo o mesmo procedimento, temos:

$$|B - \lambda I| = \begin{vmatrix} 1-\lambda & 2 \\ -2 & 1-\lambda \end{vmatrix} = 0$$

Daí resulta que:

$$(1-\lambda)(1-\lambda) + 4 = \lambda^2 - 2\lambda + 5 = 0$$

$$\lambda_1 = 1 + 2i$$

$$\lambda_2 = 1 - 2i$$

Nesse segundo exemplo foram obtidos dois autovalores complexos conjugados.

Exemplo 9.3 - $C = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix}$

Similarmente, temos:

$$|C - \lambda I| = \begin{vmatrix} 1-\lambda & 2 \\ -2 & 5-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)(5-\lambda) + 4 = \lambda^2 - 6\lambda + 9 = 0$$

$$\lambda_1 = \lambda_2 = 3$$

Nesse último exemplo foi obtido um autovalor real duplo.

Depois de apresentados esses três exemplos numéricos. Discute-se o conceito de autovetor.

Como a matriz $A - \lambda I$ tem o determinante nulo, ela é singular. Monta-se um sistema homogêneo tendo como matriz dos coeficientes essa matriz: $(A - \lambda I)X = 0$.

Na discussão sobre posto de matriz, vimos que um sistema homogêneo sempre tem solução. Vimos também que caso a matriz inversa exista, ou seja, se o determinante da matriz for diferente de zero, o sistema tem uma única solução, que é a trivial, caso contrário, existem ∞ soluções. Aqui, como o sistema é formado com uma matriz de coeficientes singular, pois $|A - \lambda I| = 0$, existem ∞ soluções não triviais.

Reescrevendo esse sistema, obtemos a relação com os autovetores:

$AX = \lambda X$, onde A é a matriz original, λ são os autovalores e X são os **autovetores**.

Para cada um dos autovalores, são infinitas as possibilidades de autovetores não triviais, todas múltiplas, e escolhe-se uma delas. Seguem os mesmos três exemplos numéricos.

Exemplo 9.1 - $A = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix}$ com $\lambda_1 = 2$ e $\lambda_2 = 3$.

Inicialmente, vejamos qual é o autovetor para o primeiro dos autovalores. Partindo da relação $AX = \lambda X$, temos:

$$\begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2 \begin{pmatrix} x \\ y \end{pmatrix}.$$

Daí decorre que:

$$4x + y = 2x, \text{ o que implica em } y = -2x; \text{ e}$$

$$-2x + y = 2y, \text{ o que também implica em } y = -2x$$

Assim, escolhe-se como autovetor qualquer múltiplo não nulo que respeite essa relação. Por exemplo:

$$V_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

Seguindo esse mesmo procedimento para o segundo dos autovalores, temos:

$$\begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 3 \begin{pmatrix} x \\ y \end{pmatrix}.$$

$$4x + y = 3x, \text{ o que implica em } y = -x.$$

$$-2x + y = 3y, \text{ o que também implica em } y = -x.$$

Daí, temos:

$$V_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Exemplo 9.2 - $B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$ com $\lambda_1 = 1 + 2i$ e $\lambda_2 = 1 - 2i$.

Para o primeiro autovetor, temos:

$$\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (1 + 2i) \begin{pmatrix} x \\ y \end{pmatrix};$$

$$x + 2y = (1 + 2i)x, \text{ o que implica em } y = ix; \text{ e}$$

$$-2x + y = (1 + 2i)y, \text{ o que também implica em } y = ix.$$

Assim:

$$V_1 = \begin{pmatrix} 1 \\ i \end{pmatrix}$$

Seguindo esse mesmo procedimento para o segundo dos autovalores, temos o complexo conjugado do primeiro:

$$V_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Exemplo 9.3 - $C = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix}$ com $\lambda_1 = \lambda_2 = 3$.

Neste exemplo, temos apenas um autovetor:

$$\begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 3 \begin{pmatrix} x \\ y \end{pmatrix};$$

$x + 2y = 3x$, o que implica em $y = x$; e

$-2x + 5y = 3y$, o que também implica em $y = x$.

Como único autovetor, temos:

$$V_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Os autovalores e autovetores apresentam inúmeras aplicações em Economia. Esses mesmos exemplos numéricos serão utilizados no estudo sobre sistemas de equações de diferença no quinto capítulo.

Dois outros conceitos associados aos autovalores, que também são utilizados em estudos sobre sistemas dinâmicos, são o traço e o determinante de matrizes, como será discutido no sexto capítulo com os sistemas de equações diferenciais. Segue a relação entre autovalores e traço e determinante de matrizes.

Traço – O traço de uma matriz A , representado como $tr(A)$, é a soma dos elementos da diagonal principal. O traço pode também ser calculado como a soma dos autovalores. Assim, temos: $tr(A) = a_{11} + a_{22} \dots + a_{kk} = \lambda_1 + \lambda_2 + \dots + \lambda_k$

O determinante de uma matriz pode também ser calculado a partir do produto dos autovalores: $\det(A) = \lambda_1 \lambda_2 \dots \lambda_k$.

Segue uma demonstração para ambas as relações para matrizes 2×2 :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ com autovalores } \lambda_1 \text{ e } \lambda_2.$$

Pela definição de autovalor, temos: $\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0$, o que implica em:

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0.$$

Dado que λ_1 e λ_2 são os autovalores da matriz, temos:

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2 = 0.$$

Comparando as duas expressões, obtemos:

$$\text{tr}(A) = a + d = \lambda_1 + \lambda_2; \text{ e}$$

$$\det(A) = (ad - bc) = \lambda_1\lambda_2.$$

Como exemplo numérico, utilizam-se as três matrizes descritas anteriormente.

$$\text{Para } A = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix} \text{ com } \lambda_1 = 2 \text{ e } \lambda_2 = 3: \text{tr}(A) = 5 \text{ e } |A| = 6.$$

$$\text{Para } B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \text{ com } \lambda_1 = 1 + 2i \text{ e } \lambda_2 = 1 - 2i: \text{tr}(B) = 2 \text{ e } |B| = 5.$$

$$\text{Para } C = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix} \text{ com } \lambda_1 = \lambda_2 = 3: \text{tr}(C) = 6 \text{ e } |C| = 9.$$

10. PRODUTO INTERNO

O produto interno entre dois vetores é um conceito utilizado no cálculo de várias variáveis, em particular no gradiente de funções, e também está presente em problemas de otimização com restrições de desigualdade, como veremos no segundo e nono capítulos.

Produto interno – Dados dois vetores no \mathfrak{R}^n , $U = (u_1, u_2, \dots, u_n)$ e $V = (v_1, v_2, \dots, v_n)$, o produto interno (ou produto escalar) de U e V , escrito como $U \bullet V$, é um número dado por $U \bullet V = u_1v_1 + u_2v_2 + \dots + u_nv_n$.

A norma, tamanho ou comprimento de um vetor podem ser obtidos através desse produto:

$$\|U\|^2 = U \bullet U \text{ ou } \|U\| = \sqrt{U \bullet U}.$$

O ângulo θ entre dois vetores U e V de norma positiva também pode ser obtido a partir desse produto:

$$\cos \theta = \frac{U \bullet V}{\|U\| \|V\|}.$$

Por exemplo, tome dois vetores no \Re^2 , $U = (1,1)$ e $V = (-1,2)$.

O produto escalar entre eles é:

$$U \bullet V = (1)(-1) + (1)(2) = 1.$$

Os tamanhos desses vetores são respectivamente:

$$\|U\| = \sqrt{(1,1) \bullet (1,1)} = \sqrt{2}$$

$$\|V\| = \sqrt{(-1,2) \bullet (-1,2)} = \sqrt{5}$$

O cosseno do ângulo entre eles é:

$$\cos \theta = \frac{U \bullet V}{\|U\| \|V\|} = \frac{1}{\sqrt{2} \sqrt{5}} \approx 0,32$$

$$\theta \approx \frac{\pi}{2,52}$$

CAPÍTULO 2 – FUNÇÕES DE VÁRIAS VARIÁVEIS

As funções de várias variáveis permeiam a análise econômica como em funções de produção, funções de custo, funções utilidade, etc. Essas funções são estudadas pelo cálculo de várias variáveis. Neste capítulo são discutidos pontos do cálculo de várias variáveis que, além de apresentarem inúmeras aplicações em economia, servem de base para vários outros conceitos, inclusive em tópicos normalmente abordados no cálculo de uma variável, como otimização e o estudo de concavidade e funções.

Os temas abordados aqui são: i) Funções de várias variáveis; ii) Cálculo de várias variáveis; iii) Derivadas de ordem superior e a matriz Hessiana; iv) Regra da cadeia; v) Derivadas direcionais e gradiente; vi) Funções e derivação implícita; vii) Curvas de nível e tangentes; e viii) Polinômios de Taylor. Assume-se que o leitor tenha conhecimento sobre conceitos básicos da álgebra linear, como discutido no primeiro capítulo, e também sobre o cálculo de uma variável, como normalmente apresentado em cursos introdutórios em nível universitário. Além disso, assume-se que o leitor também tenha noções de microeconomia, no formato discutido em cursos introdutórios em nível universitário.

1. FUNÇÕES DE VÁRIAS VARIÁVEIS

As funções de uma e de várias variáveis apresentam grande aplicabilidade em modelos econômicos. Este capítulo discute as funções e o cálculo de várias variáveis. Assume-se que o leitor já esteja familiarizado com alguns conceitos usualmente abordados em cursos introdutórios em nível universitário, como o conceito de função e como o cálculo de uma variável. Portanto, tópicos normalmente abordados em cursos introdutórios do cálculo diferencial não são discutidos aqui.

Inicia-se a apresentação com funções de várias variáveis. Seguem alguns exemplos dessas funções que são usualmente aplicadas em teoria microeconômica.

Uma função utilidade, $U = U(x_1, x_2, \dots, x_n)$, designa um valor real de utilidade a partir do consumo de n bens. Note que a função é definida como $f: X \rightarrow Y$, em que o domínio é definido a partir dos n bens em um espaço $X \subset \mathfrak{R}^n$ e o contradomínio é definido em $Y \subset \mathfrak{R}$.

Como exemplo numérico, assumamos que são dois bens, picanha (x) e cerveja (y), e que a função utilidade seja dada por $U(x, y) = x^2 y$. O indivíduo pode consumir diferentes quantidades de cada bem, por exemplo, os pares $(1, 0)$, $(1, 1)$ e $(2, 3)$, com as respectivas utilidades de $U(1,0) = 0$, $U(1,1) = 1$ e $U(2,3) = 12$.

Uma função de produção, $Q = Q(x_1, x_2, \dots, x_n)$, designa um valor real de produção a partir da utilização de n insumos. A função $Q: X \rightarrow Y$ também é definida com $X \subset \mathfrak{R}^n$ e $Y \subset \mathfrak{R}$.

Como exemplo, assuma que uma função de produção tenha três insumos, trabalho (L), capital (K) e tecnologia (A), em que foram dadas as nomenclaturas usuais, e que tem o seguinte formato:

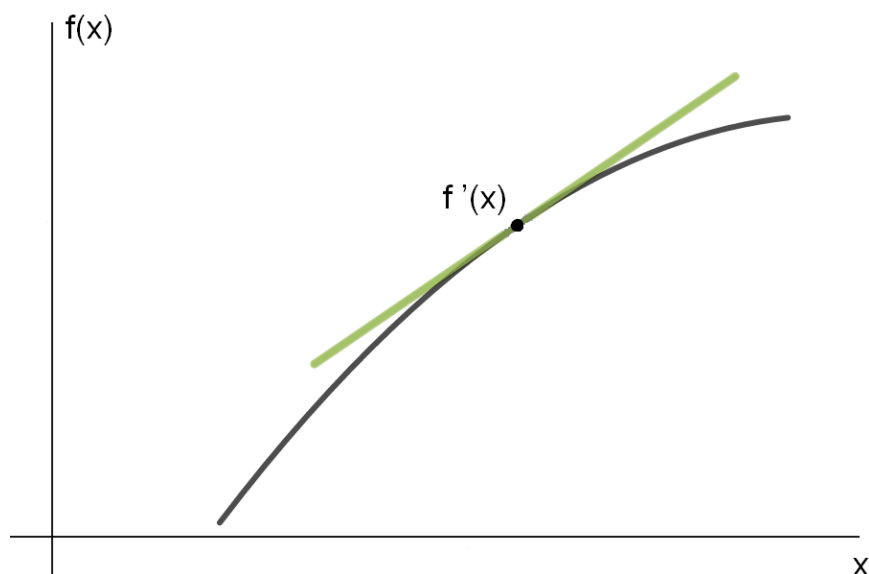
$Q = Q(L, K, A) = LK^2A^3$. A firma pode utilizar diferentes quantidades de cada insumo, por exemplo, os valores (10, 10, 10) e (10, 5, 20), com as respectivas produções de $Q(10,10,10) = 1000000$ e $Q(10,5,20) = 2000000$.

Note que o domínio e o contradomínio nos dois exemplos numéricos descritos foram respectivamente, $U: \mathbb{R}^2 \rightarrow \mathbb{R}$ e $Q: \mathbb{R}^3 \rightarrow \mathbb{R}$. Esses podem ter dimensões diversas, por exemplo, $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3$, $f: \mathbb{R}^5 \rightarrow \mathbb{R}^3$, $f: \mathbb{R} \rightarrow \mathbb{R}^n$, etc.

2. CÁLCULO DE VÁRIAS VARIÁVEIS

Um objetivo central da análise econômica é entender como que variáveis diversas afetam outras variáveis. Uma ferramenta essencial na obtenção desse entendimento é o cálculo de várias variáveis, que, como veremos, apresenta muitas semelhanças com o cálculo de uma variável. Os diagramas 1 e 2 comparam as duas abordagens, o primeiro com uma função de uma variável e o segundo com uma função de várias variáveis.

DIAGRAMA 1
A derivada em funções de uma variável

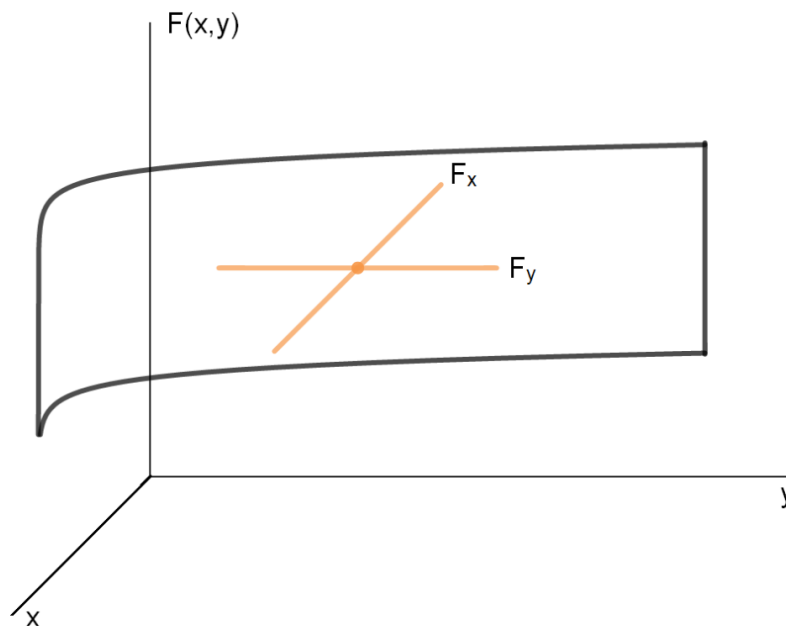


Para funções de uma variável, $f(x)$, a derivada, com a inclinação da reta tangente à curva, tem as seguintes nomenclaturas e definições:

$$\frac{df}{dx} = f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Para funções de várias variáveis, $F(x_1, x_2, \dots, x_n)$, o raciocínio é análogo. O diagrama 2 mostra uma função de duas variáveis, $F(x, y)$. Note que se calcula as inclinações das retas tangentes à superfície na direção do eixo x e na direção do eixo y . Como existe mais de uma possibilidade de derivada, diz-se que a derivada é parcial.

DIAGRAMA 2
As derivadas parciais em funções de duas variáveis



No cálculo de uma variável, utilizam-se os símbolos $\frac{df}{dx}$ ou $f'(x)$ para expressar a derivada. No exemplo descrito no diagrama 2 para o cálculo de várias variáveis, com a função $F(x, y)$, usa-se os símbolos $\frac{\partial F}{\partial x}$ ou F_x para designar a derivada parcial de F com relação a x e, similarmente, os símbolos $\frac{\partial F}{\partial y}$ ou F_y para designar a derivada parcial de F com relação a y . Note que o símbolo ∂ das derivadas parciais difere do símbolo d das derivadas de uma variável. Tampouco se representa a derivada parcial como uma linha, uma vez que se deve designar com relação a qual variável que a derivada parcial esta sendo feita.

Assim, similarmente à definição de derivadas de funções de uma variável, temos as seguintes expressões para as derivadas parciais:

$$\frac{\partial F}{\partial x} = F_x = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x, y) - F(x, y)}{\Delta x},$$

$$\frac{\partial F}{\partial y} = F_y = \lim_{\Delta y \rightarrow 0} \frac{F(x, y + \Delta y) - F(x, y)}{\Delta y}.$$

É importante ressaltar que quando a derivada parcial é com relação a x , o valor de y é constante e vice-versa.

Para uma função genérica $F(x_1, x_2, \dots, x_n)$, a definição de derivada parcial é similar. Para cada variável x_i , temos:

$$\frac{\partial F}{\partial x_i} = F_{x_i} = \lim_{\Delta x_i \rightarrow 0} \frac{F(x_1, \dots, x_i + \Delta x_i, \dots, x_n) - F(x_1, \dots, x_i, \dots, x_n)}{\Delta x_i}.$$

As condições necessárias para a existência das derivadas parciais são similares às condições para a existência da derivada de funções de uma variável. Seguem alguns exemplos. Inicialmente, deve-se ter em mente que as regras de derivação são as mesmas do cálculo de uma variável, levando em conta que quando a derivada é obtida com relação a uma variável específica, todas as demais são constantes.

Exemplo 1 - Obtenha as derivadas parciais da função $F(x, y) = x^2 + 5x + xy + y^3 + 4$.

Temos:

$$F_x = 2x + 5 + y$$

Note que quando a derivada parcial é feita com relação a x , y é constante. Portanto, y^3 é constante e a derivada do termo é zero.

A derivada parcial com relação a y é:

$$F_y = x + 3y^2$$

Exemplo 2 - Obtenha as derivadas parciais da função produção dada por $Q(L, K) = L^{1/4} K^{3/4}$.

As derivadas parciais de Q com relação a L e com relação a K podem ser representadas respectivamente por: $\frac{\partial Q}{\partial L}$ ou Q_L ; e $\frac{\partial Q}{\partial K}$ ou Q_K .

Vejamos:

$$Q_L = \frac{1}{4} L^{-3/4} K^{3/4} = \frac{1}{4} \left(\frac{K}{L} \right)^{3/4}$$

$$Q_K = \frac{3}{4} L^{1/4} K^{-1/4} = \frac{3}{4} \left(\frac{L}{K} \right)^{1/4}$$

As derivadas parciais têm uma ampla gama de aplicações em Economia. Por exemplo, essas duas últimas derivadas são respectivamente as produtividades marginais do trabalho e do capital.

Como exemplo numérico, determinam-se as produtividades marginais no ponto (16, 1).

$$Q_L(16, 1) = \frac{1}{4} \left(\frac{1}{16} \right)^{3/4} = \frac{1}{4} \left(\frac{1}{8} \right) = \frac{1}{32}.$$

$$Q_K(16, 1) = \frac{3}{4} \left(\frac{16}{1} \right)^{1/4} = \frac{3}{4} (2) = \frac{3}{2}.$$

Como veremos no oitavo e no nono capítulos, que, respectivamente, discutem problemas de otimização não condicionada e problemas de otimização condicionada, as derivadas parciais de primeira ordem são muito utilizadas na obtenção dos pontos ótimos. As derivadas parciais de ordem superior também têm grande aplicabilidade e esse é o tema da próxima seção.

3. DERIVADAS PARCIAIS DE ORDEM SUPERIOR E A MATRIZ HESSIANA

A seção anterior apresentou as derivadas parciais. Assim como no cálculo de uma variável, funções de várias variáveis também podem ter derivadas mais de uma vez, sendo obtidas as derivadas parciais de ordem superior.

Por exemplo, tome uma função polinomial de uma variável, $f(x) = 4x^2 + 5x + 6$. Pode-se derivar essa função diversas vezes, uma vez que as derivadas existam:

$$f'(x) = 8x + 5$$

$$f''(x) = 8$$

$$f'''(x) = 0$$

As funções de várias variáveis também podem ser derivadas diversas vezes, caso essas derivadas existam. Tome como exemplo uma função polinomial de duas variáveis, $F(x, y) = x^2 + 5x + xy + y^3 + 4$, que, como vimos, tem as seguintes derivadas de primeira ordem:

$$\frac{\partial F}{\partial x} = F_x = 2x + 5 + y$$

$$\frac{\partial F}{\partial y} = F_y = x + 3y^2$$

Assim, uma vez obtidas as derivadas parciais de primeira ordem, podem-se obter as derivadas parciais de segunda ordem. Porém, note que existem quatro possibilidades para o exemplo acima.

Pode-se derivar a derivada parcial com relação a x , $\frac{\partial F}{\partial x}$, uma segunda vez com relação a x ,

$$\frac{\partial}{\partial x} \left(\frac{\partial F}{\partial x} \right) = \frac{\partial^2 F}{\partial x^2} = F_{xx}, \text{ ou com relação a } y, \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial x} \right) = \frac{\partial^2 F}{\partial y \partial x} = F_{xy}.$$

O mesmo se aplica para a derivada parcial de primeira ordem com relação a y : $\frac{\partial^2 F}{\partial y^2} = F_{yy}$ e

$$\frac{\partial^2 F}{\partial x \partial y} = F_{yx}.$$

Nesse exemplo, temos:

$$F_{xx} = 2$$

$$F_{xy} = 1$$

$$F_{yx} = 1$$

$$F_{yy} = 6y$$

Note que as derivadas cruzadas são iguais, $\frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x}$. Esse fato não é uma coincidência. Isso

ocorrerá sempre que as derivadas de segunda ordem existirem e forem contínuas, como na grande maioria das funções utilizada em problemas econômicos.

Uma boa forma de organizar essa informação com as derivadas de segunda ordem é a partir da matriz Hessiana. Para uma função com duas variáveis como esse exemplo, ela tem o seguinte formato:

$$Hessiana = \begin{pmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 6y \end{pmatrix}.$$

Essa matriz é amplamente utilizada no estudo de concavidade de funções de várias variáveis, como veremos no oitavo capítulo. Em geral, a matriz é simétrica, pois as derivadas cruzadas são iguais. Essa matriz para uma função de n variáveis tem o seguinte formato:

$$Hessiana = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{pmatrix}$$

Segue um exemplo com a função de produção descrita na primeira seção do capítulo:

$$Q = Q(L, K, A) = LK^2 A^3.$$

Como derivadas parciais de primeira ordem, temos:

$$Q_L = K^2 A^3$$

$$Q_K = 2LKA^3$$

$$Q_A = 3LK^2 A^2$$

A partir dessas, obtém-se a matriz Hessiana da função:

$$Hessiana = \begin{pmatrix} \frac{\partial^2 Q}{\partial L^2} & \frac{\partial^2 F}{\partial L \partial K} & \frac{\partial^2 F}{\partial L \partial A} \\ \frac{\partial^2 F}{\partial K \partial L} & \frac{\partial^2 F}{\partial K^2} & \frac{\partial^2 F}{\partial K \partial A} \\ \frac{\partial^2 F}{\partial A \partial L} & \frac{\partial^2 F}{\partial A \partial K} & \frac{\partial^2 F}{\partial A^2} \end{pmatrix} = \begin{pmatrix} 0 & 2KA^3 & 3K^2 A^2 \\ 2KA^3 & 2LA^3 & 6LKA^2 \\ 3K^2 A^2 & 6LKA^2 & 6LK^2 A \end{pmatrix}.$$

4. REGRA DA CADEIA

Esta seção discute a regra da cadeia para funções de várias variáveis. No cálculo de uma variável, dada a função composta $g(t) = f[x(t)]$, obtemos a regra da cadeia em um ponto específico, t_0 , a partir da seguinte relação: $\frac{dg}{dt}(t_0) = \frac{df}{dx}(x(t_0)) \frac{dx}{dt}(t_0)$.

Estende-se esse raciocínio para as funções de várias variáveis. Inicialmente, define-se o que é uma curva e uma curva regular.

Curva – Uma curva no \mathbb{R}^n é representada por $X(t) = (x_1(t), x_2(t), \dots, x_n(t))$, em que variando o parâmetro $t \in \mathbb{R}$, obtém-se em cada uma das funções $x_i(t)$ um valor em \mathbb{R} .

Por exemplo, uma curva no \mathbb{R}^2 é uma função definida como $f: \mathbb{R} \rightarrow \mathbb{R}^2$, por exemplo, $X(t) = (x(t), y(t)) = (t+1, t^2)$. Variando $t \in \mathbb{R}$, temos valores de $x(t) \in \mathbb{R}$ e de $y(t) \in \mathbb{R}$, e esses formam uma curva no plano, $X(t) \in \mathbb{R}^2$.

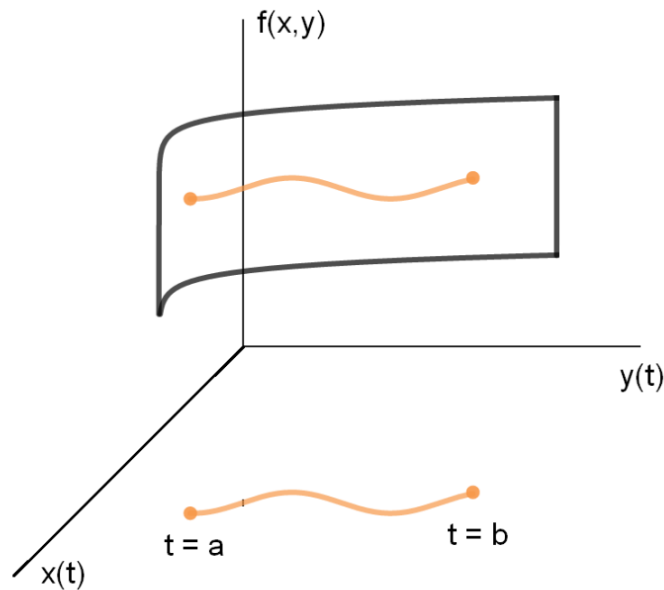
Curva regular – Uma curva no \mathbb{R}^n é regular se, e somente se, todo $x_i'(t)$ for contínuo em t e $(x_1'(t), x_2'(t), \dots, x_n'(t)) \neq (0, 0, \dots, 0)$ para todo t .

No exemplo acima, a curva é regular, pois $x(t)$ e $y(t)$ são contínuas e quando o primeiro for zero, o segundo não é nulo.

De posse dessas definições, segue a apresentação da regra da cadeia para funções de duas variáveis, sendo que a extensão para o caso com n variáveis é direto. O diagrama 3 mostra essa função.

Seja $(x(t), y(t))$ uma curva regular com $a \leq t \leq b$. O objetivo é determinar como que uma função no \mathbb{R}^3 , $f(x(t), y(t))$, se comporta no decorrer da curva. Como a função é parametrizada em t , temos: $g(t) = f(x(t), y(t))$.

DIAGRAMA 3
Regra da cadeia



No cálculo de várias variáveis o procedimento é similar ao que é feito no cálculo de uma variável. Necessita-se de mais uma definição. Uma função cujas derivadas de primeira ordem são contínuas é denominada C^1 .

Seja $(x(t), y(t))$ uma curva C^1 no intervalo em torno de t_0 . Seja $F(x(t), y(t))$ uma função C^1 no intervalo em torno de $(x(t_0), y(t_0))$. Então, $g(t) = F(x(t), y(t))$ também é uma função C^1 em t_0 . Além disso, temos:

$$\frac{dg}{dt}(t_0) = \frac{\partial F}{\partial x}(x(t_0), y(t_0)) \frac{dx}{dt}(t_0) + \frac{\partial F}{\partial y}(x(t_0), y(t_0)) \frac{dy}{dt}(t_0).$$

Ou seja, uma extensão do que é utilizado no cálculo de uma variável. Segue um exemplo numérico com a função de duas variáveis já utilizadas, $F(x, y) = x^2 + 5x + xy + y^3 + 4$, com as seguintes derivadas parciais de primeira ordem: $\frac{\partial F}{\partial x} = 2x + 5 + y$ e $\frac{\partial F}{\partial y} = x + 3y^2$. Assuma que

$$X(t) = (x(t), y(t)) = (t+1, t^2), \text{ como no exemplo anterior.}$$

A regra da cadeia toma o seguinte formato:

$$\frac{dg}{dt}(t_0) = (2x + 5 + y)|_{(x(t_0), y(t_0))} (1)|_{t_0} + (x + 3y^2)|_{(x(t_0), y(t_0))} (2t)|_{t_0}$$

Para $t_0 = 1$, temos:

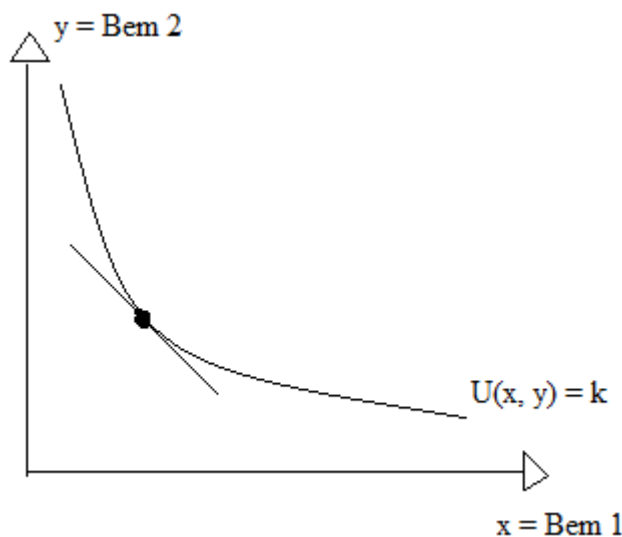
$$\frac{dg}{dt}(1) = (4 + 5 + 1)(1) + (2 + 3)(2) = 20.$$

5. DERIVADAS IMPLÍCITAS

Outro conceito amplamente utilizado na resolução de problemas econômicos é o de derivação implícita. Toda a discussão com relação às derivadas parciais até o momento foi feita com funções explícitas do tipo $y = F(x_1, \dots, x_n)$. Entretanto, é comum não podermos separar a variável y das demais variáveis x_i e, assim, temos uma função implícita, usualmente descrita como $G(x_1, \dots, x_n, y) = C$, em que C é uma constante.

Por exemplo, tome uma função utilidade com duas variáveis, $U(x, y)$, e defina uma curva de indiferença, $U(x, y) = k$, como representado no diagrama 4. Note que essa curva é uma curva de nível no \mathbb{R}^2 , que representa uma função no \mathbb{R}^3 .

DIAGRAMA 4
Curva de indiferença e a derivação implícita



Quer-se calcular a inclinação da curva de nível, porém, de forma implícita, tendo como base esse diagrama, com uma função com duas variáveis. O raciocínio para funções com mais variáveis é análogo.

Toda a dedução do teorema da função implícita assume que se trata de uma curva de nível, $U(x, y) = k$. Além disso, necessariamente pode-se representar uma variável como função da outra: $y = y(x)$. Assim, essa função implícita é reescrita como $U(x, y(x)) = k$.

Derivando essa expressão utilizando a regra da cadeia descrita na seção anterior:

$$\frac{dU}{dx} = \frac{\partial U}{\partial x} \frac{dx}{dx} + \frac{\partial U}{\partial y} \frac{dy}{dx} = 0.$$

Note que, como a derivada é feita em uma curva de nível, ou seja, $U(x, y) = k$, toda a derivada é nula.

Assim, em um ponto específico $(x_0, y(x_0))$, temos:

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

Rearranjando, ficamos com:

$$\frac{dy}{dx}(x_0) = - \frac{\frac{\partial U}{\partial x}(x_0, y_0)}{\frac{\partial U}{\partial y}(x_0, y_0)}.$$

Essa expressão permite obter a inclinação da curva de nível, $\frac{dy}{dx}$, a partir de duas derivadas parciais.

Note que a expressão da derivação implícita é válida se $\frac{\partial U}{\partial y}(x_0, y_0) \neq 0$. Se o denominador for nulo, não podemos assumir que $y = y(x)$, pois o conjunto dos pontos $U(x, y) = k$, seria vertical em pelo menos uma região e não representaria uma função de y com relação a x . Entretanto, talvez a derivação implícita possa ser representada como uma função $x = x(y)$.

Como ilustração, segue um exemplo numérico com a função utilidade $U(x, y) = xy$. Parte-se de uma curva de indiferença, $U(x, y) = k$.

As derivadas parciais são $U_x = y$ e $U_y = x$.

Utilizando a expressão da derivação implícita, temos:

$$\frac{dy}{dx}(x_0) = - \frac{\frac{\partial U}{\partial x}(x_0, y_0)}{\frac{\partial U}{\partial y}(x_0, y_0)} = - \frac{x}{y} \Big|_{(x_0, y_0)}$$

No ponto $(x_0, y_0) = (1, 1)$, temos:

$$\frac{dy}{dx}(1) = - \frac{1}{1} = -1.$$

As derivadas parciais de funções utilidades são denominadas utilidades marginais. A taxa marginal de substituição (TMS) é um exemplo de aplicação da derivação implícita. Por exemplo, como mostrado no diagrama4, essa taxa é o módulo da inclinação de uma curva de indiferença, sendo a divisão das

utilidades marginais: $TMS = \left| \frac{dy}{dx} \right| = \left| -\frac{U_x}{U_y} \right| = \frac{U_x}{U_y}$.

6. DERIVADAS DIRECIONAIS E GRADIENTES

A regra da cadeia também é a base para a obtenção de dois outros conceitos importantes no cálculo de várias variáveis e em aplicações em economia, que são as derivadas direcionais e o gradiente. O diagrama a seguir exemplifica o problema para uma função de duas variáveis e a discussão será feita para essa função. A extensão para o caso geral com n variáveis é direta.

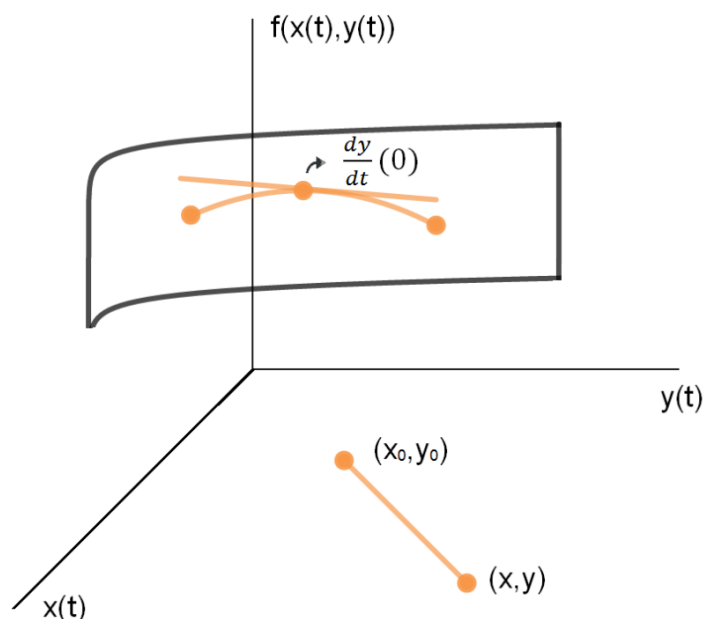
Inicialmente, parte-se de um ponto inicial dado, $(x_0(t), y_0(t))$. Também se define uma direção fixa, $V^T = (a, b)$, em que usualmente o módulo do vetor direção é unitário: $\|V\| = 1$. Assim, a curva $X(t) = (x(t), y(t))$ pode ser escrita como:

$$X(t) = (x_0 + at, y_0 + bt), \text{ pois } x(t) = x_0 + at \text{ e } y(t) = y_0 + bt.$$

Seguindo a definição da regra da cadeia, a função $g(t)$ é, portanto:

$$g(t) = f(x_0 + at, y_0 + bt).$$

DIAGRAMA 5
A derivada direcional



Note que o ponto inicial, (x_0, y_0) , e a direção, (a, b) , são dados. Utiliza-se a regra da cadeia para obter a derivada de $g(t)$ em $t = 0$:

$$\frac{dg}{dt}(0) = \frac{\partial f}{\partial x}(x(0), y(0)) \frac{dx}{dt}(0) + \frac{\partial f}{\partial y}(x(0), y(0)) \frac{dy}{dt}(0) = \frac{\partial f}{\partial x}(x_0, y_0)a + \frac{\partial f}{\partial y}(x_0, y_0)b.$$

Utiliza-se a nomenclatura $D_V f(x_0, y_0)$ para designar a derivada direcional na direção V no ponto

$$(x_0, y_0): D_V f(x_0, y_0) = \frac{dg}{dt}(0) = \frac{\partial f}{\partial x}(x_0, y_0)a + \frac{\partial f}{\partial y}(x_0, y_0)b.$$

No primeiro capítulo, discutiu-se o conceito de produto interno entre vetores. Essa relação representando a derivada direcional pode ser escrita em formato de vetores utilizando esse conceito:

$$D_V f(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) \\ \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} \bullet \begin{pmatrix} a \\ b \end{pmatrix} = \nabla F(x_0, y_0) \bullet V, \text{ em que } \nabla f(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) \\ \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} \text{ é o}$$

gradiente da função.

Segue um exemplo numérico com a mesma função de duas variáveis já utilizada:

$F(x, y) = x^2 + 5x + xy + y^3 + 4$, com as respectivas derivadas parciais de primeira ordem

$$\frac{\partial F}{\partial x} = 2x + 5 + y \text{ e } \frac{\partial F}{\partial y} = x + 3y^2.$$

Portanto, o gradiente dessa função é o seguinte:

$$\nabla F(x, y) = \begin{pmatrix} 2x + 5 + y \\ x + 3y^2 \end{pmatrix}.$$

Qual é a derivada direcional dessa função no ponto $(x_0, y_0) = (1, 1)$ e na direção $V = (3, 4)$?

Inicialmente, obtém-se o gradiente da função no ponto especificado:

$$\nabla F(1, 1) = \begin{pmatrix} 8 \\ 4 \end{pmatrix}.$$

Em seguida, obtém-se o vetor unitário na direção especificada. O vetor unitário é aquele de tamanho um. Como vimos no capítulo anterior, o tamanho de um vetor é dado por:

$$\|X\| = (X \bullet X)^{1/2}.$$

No caso específico do vetor $V = (3,4)$, temos:

$$\|V\| = ((3,4) \bullet (3,4))^{1/2} = ((3)(3) + (4)(4))^{1/2} = 5.$$

Uma vez obtido o comprimento do vetor que determina a direção da derivada direcional, obtém-se o vetor unitário na mesma direção:

$$U_V = \frac{V}{\|V\|} = \frac{(3,4)}{5} = \left(\frac{3}{5}, \frac{4}{5}\right).$$

Por fim, calcula-se a derivada direcional da função no ponto e na direção dadas:

$$D_V F(1,1) = \begin{pmatrix} 8 \\ 4 \end{pmatrix} \bullet \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix} = \frac{24 + 16}{5} = 8.$$

Também como discutido no capítulo anterior, sabe-se que o cosseno do ângulo entre dois vetores quaisquer é dado pela seguinte relação:

$$\cos \theta = \frac{X \bullet Y}{\|X\| \|Y\|}$$

Reescrevendo essa relação para a expressão da derivada direcional, temos:

$$\cos \theta = \frac{\nabla F \bullet U_V}{\|\nabla F\| \|U_V\|} = \frac{\nabla F \bullet U_V}{\|\nabla F\|}, \text{ pois } \|U_V\| = 1.$$

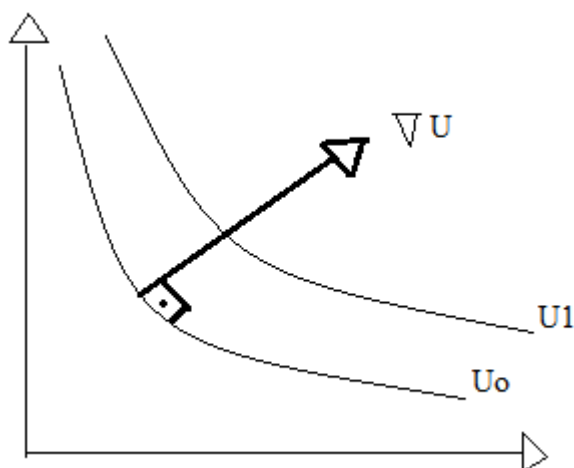
$$D_U F = \nabla F \bullet U_V = \|\nabla F\| \cos \theta.$$

Note que, para um dado ponto, o vetor gradiente é fixo. Assim, o único termo que pode variar na expressão acima é o $\cos \theta$. Isso implica que o maior valor possível para $D_U F$ é quando $\theta = 0$. Ou seja, o gradiente indica a direção onde a derivada direcional no ponto é a maior possível, isto é, a direção em que a função cresce o mais rápido possível no ponto especificado. Note ainda que, se $\theta = \pi/2$, $\cos \theta = 0$, tem-se $D_U F = 0$.

O gradiente e a derivada direcional são utilizados em diferentes aplicações e conceitos em economia, como veremos na otimização condicionada. Segue um exemplo com curvas de indiferença para uma

função utilidade com dois bens, representada no diagrama a seguir. Em uma curva de indiferença, por definição, os valores de U são constantes. Assim, a derivada direcional na direção da curva de diferença é zero. Isso implica que o gradiente da função utilidade $U(x, y)$, ∇U , é perpendicular à curva de indiferença. Na verdade, esse resultado pode ser estendido para outras curvas de nível. O gradiente é perpendicular às curvas de nível em geral.

DIAGRAMA 5
A relação de gradiente e curva de nível



CAPÍTULO 3 - INTEGRAÇÃO

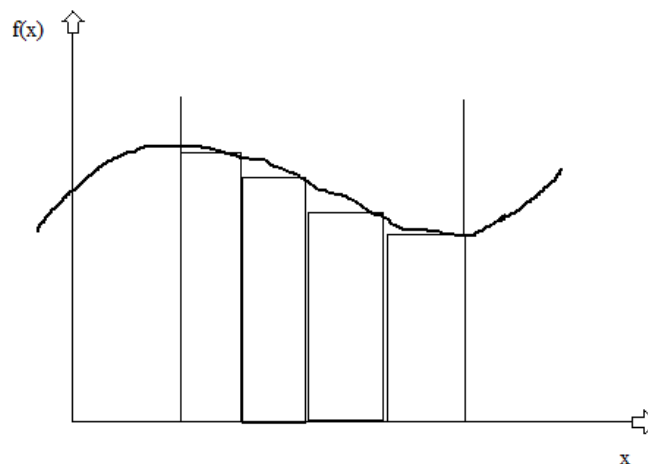
As integrais são amplamente utilizadas diretamente em aplicações em Economia, como será mostrado neste capítulo, e também servem para resolver diversos problemas associados a problemas econômicos, como será descrito em capítulos subsequentes, por exemplo, na resolução de equações diferenciais. Os tópicos abordados aqui são: i) Introdução à integração; ii) Expressões básicas de integração; iii) Substituição simples; iv) Integração por partes; e v) Aplicações da integral em Economia. Assume-se que o leitor tenha conhecimento sobre conceitos básicos do cálculo diferencial de uma variável, como normalmente apresentado em cursos introdutórios em nível universitário.

1. INTRODUÇÃO À INTEGRAÇÃO

O cálculo integral é diretamente relacionado ao cálculo diferencial, entretanto, enquanto utilizam-se as derivadas para estimar inclinações, as integrais são utilizadas para o cálculo de áreas, ou no caso das integrais múltiplas, tema não abordado aqui, são usadas no cálculo de volumes. Segue uma breve discussão sobre o que é e como se calcula uma integral.

A discussão do que é uma integral baseia-se no diagrama 1. Dada uma função contínua definida no intervalo $[a, b]$, tem-se como objetivo calcular a área entre a função e o eixo-x. Divide-se esse intervalo em n subintervalos de comprimento igual, $\Delta x = \frac{b-a}{n}$, no caso temos $n = 4$. Escolhe-se um ponto amostral $x_1^*, x_2^*, \dots, x_n^*$ em cada um dos subintervalos. Note que a área entre a função e o eixo-x pode ser aproximada por $A \cong \sum_{i=1}^4 f(x_i^*) \Delta x$, a soma aproximada das áreas de retângulos.

DIAGRAMA 1
A integral simples



Note que a medida que **n** aumenta, por exemplo, $n = 10$ com $A \cong \sum_{i=1}^{10} f(x_i^*) \Delta x$, a área estimada pelos retângulos tende a se aproximar do valor real. No limite, quando $n \rightarrow \infty$, a diferença entre as duas medidas tende a zero.

Temos então um somatório de infinitas áreas retangulares cuja soma é exatamente a área compreendida entre a função e o eixo-x. Dado que temos um limite com $n \rightarrow \infty$, usa-se como símbolo desse somatório infinito um S estendido que é o símbolo da integral:

$$A = \lim_{n \rightarrow \infty} \left[\sum_{i=1}^n f(x_i^*) \Delta x \right] = \int_a^b f(x) dx. \text{ Ou seja, a integral representa o somatório de infinitos}$$

retângulos cujas alturas são definidas a partir de uma função específica. Note que se a função for negativa, a área também será negativa.

Para calcular uma integral fazemos uso da antiderivada, como será mostrado a seguir com o auxílio do diagrama 2. Seja $A(x)$ a área compreendida entre uma função positiva e o eixo-x entre o ponto **a** e um ponto **x** qualquer. Varia-se esse ponto **x** para $x + h$ e tem-se um aumento de área. A diferença entre as duas áreas é dada por:

$$A(x+h) - A(x) = f^m(x)h, \text{ onde } f^m(x) \text{ é o valor médio da função entre } x \text{ e } x+h.$$

$$\text{Dividindo a expressão acima por } h \text{ obtemos, } \frac{A(x+h) - A(x)}{h} = f^m(x).$$

Tomando o limite quando $h \rightarrow 0$, observe que obtemos justamente a definição de derivada, sendo que a função que está sendo derivada é a área:

$$A'(x) = \lim_{h \rightarrow 0} \frac{A(x+h) - A(x)}{h} = f(x).$$

Assim, a função $f(x)$ representada no gráfico é a derivada de uma função $A(x)$, que é a função que define a área entre a função e o eixo-x. Portanto, sabe-se que $A'(x) = f(x)$ e deseja-se obter $A(x)$, a função cuja derivada é justamente $f(x)$. Isto é, o procedimento é justamente o inverso da derivada, onde temos uma função e obtém-se a derivada. Aqui temos a derivada e buscamos a função, temos o processo da antiderivada:

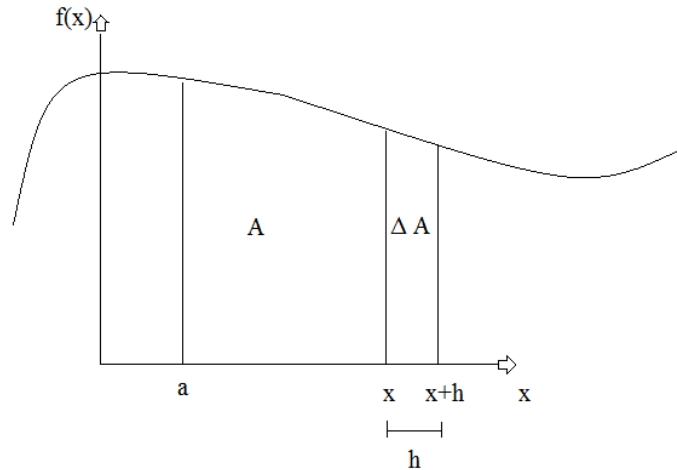
$$A = \int_a^b f(x) dx \text{ e } A'(x) = f(x).$$

Em uma apresentação mais formal:

Teorema fundamental do cálculo – Se f for contínua em $[a, b]$, então $\int_a^b f(x) dx = F(b) - F(a)$,

onde F é qualquer antiderivada de f , isto é, uma função tal que $F' = f$.

DIAGRAMA 2
A antiderivada



Como os limites a e b na integral são definidos, diz-se que a integral é definida:

$$\int_a^b f(x)dx = F(x)\Big|_a^b = F(b) - F(a). \text{ Caso a integral não tenha os limites definidos, diz-se que a integral}$$

é indefinida: $\int f(x)dx = F(x)$.

Essa seção introduziu o conceito de integral e o associou ao processo de antiderivada. Na próxima seção apresentam-se algumas das expressões básicas de integração.

2. EXPRESSÕES BÁSICAS DE INTEGRAÇÃO

Vimos na seção anterior que para calcular uma integral devemos obter a função, $F(x)$, que deu origem a uma derivada conhecida, $f(x)$. Grande parte das integrais presentes em problemas econômicos podem ser resolvidas direta ou indiretamente por uma das três expressões básicas de integração, que são descritas nesta seção. Existem outras, com funções seno e cosseno, que apresentam menor aplicabilidade em Economia e não são apresentadas aqui.

Segue a obtenção da primeira expressão básica de integração. Utiliza-se a variável genérica u na obtenção dessas expressões.

$$\text{Se } F(u) = \frac{u^{n+1}}{n+1} + C, \text{ onde } C \text{ é uma constante, então } F'(u) = n+1 \frac{u^n}{n+1} = u^n = f(u).$$

Então, sabendo que $\int f(u)du = F(u)$, obtém-se a primeira expressão básica de integração:

$$1) \int u^n du = \frac{u^{n+1}}{n+1} + C.$$

De forma análoga, obtém-se as outras duas expressões:

$$2) \int e^u du = e^u + C.$$

$$3) \int \frac{du}{u} = \ln(u) + C.$$

Lembrando que o logaritmo é definido apenas para valores positivos.

Seguem quatro exemplos de integrais que são resolvidas utilizando diretamente essas expressões.

Exemplo 2.1 - $\int 4x^2 dx$

O primeiro exemplo é de uma integral indefinida, pois os limites de integração não são definidos. Para resolvê-la, o objetivo inicial é reescrever a integral em um dos três formatos acima. O primeiro passo é passar a constante para fora da integral:

$$4 \int x^2 dx.$$

Depois se compara a integral que queremos resolver com as expressões básicas de integração. Verifica-se que essa integral é semelhante a primeira expressão, com $u = x$ e $n = 2$. Daí, temos:

$$4 \int x^2 dx = 4 \frac{x^3}{3} + C.$$

Exemplo 2.2 - $\int_0^2 e^x dx$.

Nesse segundo exemplo temos uma integral definida. Portanto, não é necessário o uso da constante C , uma vez que depois de resolvermos a integral, fazemos as substituições com os limites de integração:

$$\int_a^b f(x) dx = (F(x) + C) \Big|_a^b = (F(b) + C) - (F(a) + C) = F(b) - F(a)$$

Retornando ao exemplo, observa-se que a integral é semelhante a segunda expressão básica de integração, apenas com $u = x$. Então, obtemos:

$$\int_0^2 e^x dx = [e^x]_0^2 = e^2 - e^0 = e^2 - 1.$$

Exemplo 2.3 - $\int_1^2 \left(3t^4 + \frac{1}{t} \right) dt$

Novamente tenta-se reescrever a integral acima de forma similar às expressões básicas de integração. No caso, separamos a integral em duas:

$$3 \int_1^2 t^4 dt + \int_1^2 \frac{1}{t} dt.$$

Utilizando a primeira expressão de integração para a primeira das integrais e a terceira expressão para a seguinte, temos:

$$3 \int_1^2 t^4 dt = \left(3 \frac{t^5}{5} \right) \Big|_1^2 = \frac{3(32)}{5} = \frac{96}{5}.$$

$$\int_1^2 \frac{1}{t} dt = [\ln t]_1^2 = \ln 2 - \ln 1 = \ln 2 - 0 = \ln 2$$

Somando o resultado dessas duas integrais, temos:

$$\int_1^2 \left(3t^4 + \frac{1}{t} \right) dt = \frac{96}{5} + \ln 2.$$

Exemplo 2.4 - $\int_1^{\infty} \frac{1}{x^{1/2}} dx$

Note que na integral acima um dos limites de integração tende ao infinito. Integrais assim são denominadas integrais impróprias. Para resolvê-la, o processo inicial é exatamente o mesmo do que em outras integrais. Reescreve-se a integral no formato de uma das três expressões básicas de integração:

$$\int_1^{\infty} \frac{1}{x^{1/2}} dx = \int_1^{\infty} x^{-1/2} dx.$$

Utilizamos a primeira expressão básica de integração. Além disso, como a integral é imprópria, depois de obtida a expressão resultante da integração, inclui-se um limite:

$$\int_1^{\infty} x^{-1/2} dx = \lim_{b \rightarrow \infty} \left[\frac{x^{1/2}}{1/2} \right]_1^b.$$

Prosseguindo com o limite, temos:

$$\lim_{b \rightarrow \infty} \left[\frac{x^{1/2}}{1/2} \right]_1^b = 2 \lim_{x \rightarrow \infty} x^{1/2} - 2 = \infty - 2 = \infty.$$

As integrais discutidas nesta seção foram resolvidas a partir da utilização direta de uma das três expressões básicas de integração. As integrais que serão resolvidas nas próximas seções são também resolvidas utilizando uma dessas expressões, mas de forma indireta. Fazendo uso dos métodos de integração, o objetivo é transformar a integral que queremos calcular em uma integral que seja exatamente igual a uma dessas expressões. Esse capítulo apresenta dois métodos muito utilizados em problemas aplicados em Economia que são o de substituição simples, método discutido na próxima seção, e o de integração por partes, que é apresentado na seção 4. Existem diversos outros tipos de métodos que não são discutidos neste capítulo por apresentarem menor aplicabilidade em Economia.

3. SUBSTITUIÇÃO SIMPLES

No método de substituição simples deve-se fazer uma substituição de variável de forma a reescrever a integral que está sendo feita em um formato mais simples, se possível em um dos formatos descritos na seção anterior. Para fazer essa mudança de variável, busca-se uma parte ou a totalidade da função na integral que será a variável u , cuja derivada, du , também deve se encontrar na integral. Ou seja, buscamos uma função u na integral tal que du também esteja presente nesta. Por exemplo, este método é muito utilizado quando temos diferentes funções polinomiais dentro da integral, sendo que estas diferem em um no grau da função. O polinômio de maior grau será o u e o de menor grau será o du . Segue um exemplo.

Exemplo 3.1 - $\int_0^3 \frac{x}{1+x^2} dx$.

Note que na integral tem-se um polinômio de maior grau, grau 2, que será a escolha natural para u : $u = x^2 + 1$. Na integral também temos um polinômio de grau 1, por conseguinte, esse aparecerá na derivada de u : $du = 2x dx$. Esse último é reescrito como, $\frac{du}{2} = x dx$, pois é o formato que aparece na integral.

Uma vez proposta essa substituição simples, deve-se também mudar os limites de integração para a nova variável. Na integral do exemplo x varia entre 0 e 3. Na integral rescrita, a variável u não necessariamente varia entre esses mesmo valores. Note que, quando $x \rightarrow 0$, temos que $u \rightarrow 1$, e quando $x \rightarrow 3$, temos que $u \rightarrow 10$. Portanto, reescrevemos a integral desse exemplo como:

$$\int_0^3 \frac{x}{1+x^2} dx = \int_1^{10} \frac{du/2}{u} = \frac{1}{2} \int_1^{10} \frac{du}{u}.$$

Note que com essa substituição simples, a integral se tornou exatamente igual à terceira expressão básica de integração descrita na seção anterior. Assim, prosseguindo na resolução da integral, substituindo os limites de integração, obtemos:

$$\frac{1}{2} \int_1^{10} \frac{du}{u} = \frac{1}{2} [\ln(u)]_1^{10} = \frac{1}{2} (\ln(10) - \ln(1)) = \frac{1}{2} \ln(10).$$

Segue mais um exemplo de integral que é resolvido pelo método de substituição simples. Essa integral é a seguinte:

Exemplo 3.2 - $\int_0^{+\infty} \frac{4}{e^{3x}} dx.$

Inicialmente, reescreve-se essa integral imprópria para que ela se pareça com uma das três expressões básicas de integração. Assim, temos:

$$\int_0^{+\infty} \frac{4}{e^{3x}} dx = 4 \int_0^{+\infty} e^{-3x} dx.$$

Comparando essa integral com a segunda expressão básica, nota-se que devemos introduzir a substituição simples, $u = -3x$, o que implica em $du = -3dx$ ou $dx = du/-3$, para que elas se tornem iguais. Calculando os novos limites de integração, temos que, quando $x \rightarrow 0$, $u \rightarrow 0$, e quando $x \rightarrow \infty$, $u \rightarrow -\infty$. Fazendo essas substituições na integral acima, temos:

$$4 \int_0^{+\infty} e^{-3x} dx = -\frac{4}{3} \int_0^{-\infty} e^u du.$$

Apenas para reescrever a integral em um formato mais usual, com o limite superior de integração sendo maior que o inferior, trocam-se os limites de integração de posição e troca-se também o sinal da integral.

$$-\frac{4}{3} \int_0^{-\infty} e^u du = \frac{4}{3} \int_{-\infty}^0 e^u du = \frac{4}{3} \lim_{b \rightarrow -\infty} e^u \Big|_b^0 = \frac{4}{3} (1 - 0) = \frac{4}{3}.$$

Como veremos na seção 5, o método de substituição simples é bastante utilizado para resolver integrais que têm aplicabilidade em Economia. Outro método também bastante utilizado é o de integração por partes, discutido na próxima seção.

4. INTEGRAÇÃO POR PARTES

O método de integração por partes é realizado com uma expressão que se baseia em uma derivada de um produto de funções:

$$\frac{d}{dx}[u(x)v(x)] = v(x)\frac{d}{dx}u(x) + u(x)\frac{d}{dx}v(x).$$

Essa expressão é reescrita como:

$$u(x)dv(x) = d[u(x)v(x)] - v(x)du(x).$$

Em seguida, integra-se essa expressão tendo como limites genéricos **a** e **b**:

$$\int_a^b u dv = [uv]_a^b - \int_a^b v du.$$

Essa é a expressão que será utilizada na integração por partes. O princípio básico do método é transformar uma integral de difícil resolução, $\int u dv$, em outra mais simples, $\int v du$, que possa ser resolvida pelos métodos discutidos nas seções anteriores. Entretanto, é importante ressaltar que em muitos problemas deve-se aplicar mais de uma vez a integração por partes até que uma integral fácil seja obtida.

Seguem alguns exemplos.

Exemplo 4.1 - $\int_0^1 x e^{-x} dx$

Como descrito, devemos transformar uma integral difícil, no caso, $\int x e^{-x} dx$, em outra fácil que possa ser integrada diretamente ou por substituição simples. Note que nesse problema a integral a ser integrada por substituição simples é $\int e^{-x} dx$. Ou seja, queremos abaixar o grau do polinômio dentro da integral de x^1 para x^0 , sendo essa uma das principais aplicações do método.

Uma vez que a integral difícil é $\int u dv$, neste método de resolução de integrais devemos escolher uma parte, ou até mesmo a totalidade, da função que está sendo integrada como **u** e o restante será **dv**. Dada a integral do problema $\int_0^1 x e^{-x} dx$, escolhe-se como **u** o polinômio que queremos abaixar o grau, assim,

$u = x$. Consequentemente, o restante que está sendo integrado é $dv = e^{-x} dx$. Em seguida, devemos

obter os valores de du e de v para substituímos na expressão de integração por partes. Desta forma, é essencial que o termo escolhido para dv seja de fácil integração, como a escolha realizada aqui.

Vejam os:

se $u = x$, então $du = dx$; e

se $dv = e^{-x}dx$, então $v = \int e^{-x}dx$.

Essa integral é de fácil resolução e é feita por substituição simples, com $u = -x$: $v = \int e^{-x}dx = -e^{-x}$.

Depois de definidos $u = x$ e $dv = e^{-x}dx$, e obtidos $du = dx$ e $v = -e^{-x}$, fazemos as substituições na expressão de integração por partes:

$$\int_0^1 xe^{-x}dx = -xe^{-x}\Big|_0^1 + \int_0^1 e^{-x}dx.$$

Note que transformamos uma integral difícil, $\int_0^1 xe^{-x}dx$, em uma fácil, $\int_0^1 e^{-x}dx$, que será resolvida por substituição simples. Assim, ficamos com:

$$\int_0^1 xe^{-x}dx = -xe^{-x}\Big|_0^1 + \int_0^1 e^{-x}dx = -xe^{-x}\Big|_0^1 - e^{-x}\Big|_0^1 = (-e^{-1} - 0) - (e^{-1} - 1) = -2e^{-1} + 1.$$

A integração por partes, além de ser muito utilizada para abaixar o grau de polinômios dentro de integrais, é utilizada para simplificar integrais em geral. Segue um exemplo.

Exemplo 4.2 - $\int_1^e \ln x dx$

Note que essa integral não pode ser resolvida por uma das expressões básicas de integração, pois difere de todas elas. Além disso, o método de substituição simples também não pode ser usado, pois se

$u = \ln x$, $du = \frac{dx}{x}$, e não conseguiríamos utilizar esse método, pois esse du não aparece na integral.

Portanto, deve-se utilizar a integração por partes para simplificar a integral, para que esta seja resolvida por algum desses métodos.

Neste exemplo não há escolha para u , uma vez que não temos um produto de funções sendo integrado. Assim, temos $u = \ln x$, e o restante da integral é consequentemente dv : $dv = dx$. A partir desses,

obtemos, $du = \frac{dx}{x}$ e $v = x$, e fazemos as substituições na expressão de integração por partes:

$$\int_1^e \ln x dx = [x \ln x]_1^e - \int_1^e x \left(\frac{1}{x} \right) dx = [x \ln x]_1^e - \int_1^e dx = [x \ln x]_1^e - x \Big|_1^e = (e \ln e - 0) - (e - 1) = e \ln e - e + 1.$$

5. APLICAÇÕES DA INTEGRAL EM ECONOMIA

Existem muitas aplicações da integral em economia. Neste capítulo são brevemente descritas três delas: o cálculo do valor presente, o cálculo de probabilidades e a estimação do excedente do consumidor. Para uma discussão mais abrangente ver Chiang e Wainwright (2006).

5.1 - Valor presente – Imagine que $p(t)$ represente um fluxo de renda futuro entre os anos **a** e **b**. O valor presente desse fluxo, $VP(t)$, ou seja, quanto que todo o fluxo valeria em termos presentes, é

dado por: $VP(t) = \int_a^b e^{-rt} p(t) dt$, onde **r** é a taxa de desconto.

Como exemplo numérico, qual seria o valor presente de um fluxo contínuo de R\$1000 ao mês por 60 meses, sabendo que a taxa de desconto é 1% ao mês? Substituindo na integral acima, temos:

$$VP(60) = \int_0^{60} e^{-0,01t} 1000 dt = 1000 \int_0^{60} e^{-0,01t} dt.$$

Essa integral é feita por substituição simples com $u = -0,01t$ e, conseqüentemente, $du = -0,01dt$ ou

$$\frac{du}{-0,01} = dt.$$

Os limites de integração se tornam:

se $t \rightarrow 0$ temos que $u \rightarrow 0$ e quando $t \rightarrow 60$ temos que $u \rightarrow -0,6$.

Daí, temos:

$$\begin{aligned} VP(60) &= 1000 \int_0^{60} e^{-0,01t} dt = 1000 \int_0^{-0,6} e^u \frac{du}{-0,01} = -100000 \int_0^{-0,6} e^u du = 100000 \int_{-0,6}^0 e^u du \\ &= 100000 \left(e^u \Big|_{-0,6}^0 \right) = 100000(1 - e^{-0,6}) = 45118,84. \end{aligned}$$

Note que o fluxo de renda nominal foi de R\$60000, mas em valores atuais, ele vale menos por causa da taxa de desconto.

5.2 - Cálculo de probabilidades – Dada uma variável aleatória contínua com função de probabilidade qualquer $f(x)$. A probabilidade que o valor da variável esteja entre **a** e **b** é dada por:

$$P(a \leq x \leq b) = \int_a^b f(x) dx.$$

Por exemplo, assuma que o tempo de espera em uma fila segue uma distribuição em formato

triangular, $f(t) = \frac{1}{15} - \frac{t}{450}$, onde o tempo máximo de espera é de 30 minutos.

Primeiro, note que:

$$P(0 \leq t \leq 30) = \int_0^{30} \left(\frac{1}{15} - \frac{t}{450} \right) dt = \left(\frac{t}{15} - \frac{t^2}{900} \right) \Big|_0^{30} = (2 - 1) = 1.$$

Ou seja, a probabilidade que um indivíduo espere de zero ao máximo de tempo é 1.
Qual é a probabilidade que o indivíduo irá esperar mais do que 10 minutos?

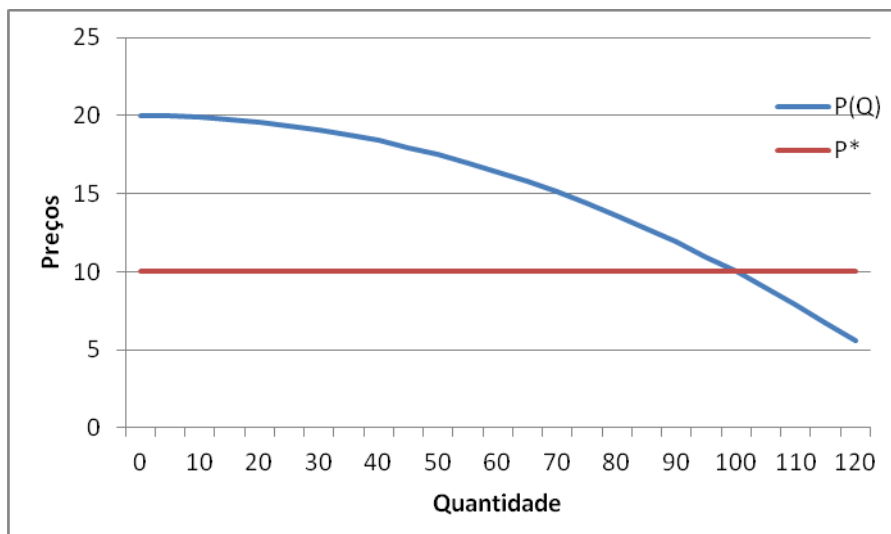
$$P(10 \leq t) = \int_{10}^{30} \left(\frac{1}{15} - \frac{t}{450} \right) dt = \left(\frac{t}{15} - \frac{t^2}{900} \right) \Big|_{10}^{30} = (2 - 1) - \left(\frac{10}{15} - \frac{100}{900} \right) = 1 - \frac{500}{900} = \frac{4}{9}.$$

5.3 - Excedente do consumidor – Em um mercado o preço de equilíbrio é representado por P^* e a quantidade de equilíbrio é Q^* . Dada uma função demanda $P(Q)$, o excedente do consumidor, indicando quanto que os consumidores estariam dispostos a pagar a mais do que o representado pelo

equilíbrio de mercado, é dado por: $E_C = \int_0^{Q^*} (P(Q) - P^*) dQ$.

Como exemplo numérico, assuma que $P^* = 10$, $Q^* = 100$ e que $P(Q) = 20 - \frac{Q^2}{1000}$, como mostra o diagrama a seguir. O excedente do consumidor é a área entre as curvas.

DIAGRAMA 3
Excedente do consumidor



Calculando o excedente do consumidor, temos:

$$E_C = \int_0^{100} \left(20 - \frac{Q^2}{1000} - 10 \right) dQ = \left(10Q - \frac{Q^3}{3000} \right) \Big|_0^{100} = \left(1000 - \frac{1000000}{3000} \right) = \frac{2000}{3}.$$

As integrais são também amplamente utilizadas na resolução de equações diferenciais, tema do quarto capítulo.

CAPÍTULO 4 - EQUAÇÕES DE DIFERENÇA DE 1ª ORDEM E EQUAÇÕES DIFERENCIAIS DE 1ª ORDEM

Existem muitos problemas econômicos que tratam de fenômenos dinâmicos, tais como evolução de taxas de desemprego, crescimento econômico, crescimento populacional, etc. Equações de diferença e equações diferenciais são poderosas ferramentas matemáticas que permitem estudar problemas econômicos desse tipo. Em um contexto onde o tempo é contínuo, capta-se essa variação por meio de derivadas e utilizam-se as equações diferenciais. Por exemplo, a variação populacional de uma região populosa pode ser considerada um fenômeno contínuo no tempo. Como veremos na seção 3, diferentes equações diferenciais podem ser utilizadas em estudos sobre variação populacional. Entretanto, em muitos outros problemas dinâmicos, como em análises sobre investimentos, em que se utilizam taxas de juros e depósitos definidos em períodos bem definidos, considera-se o tempo como discreto. As variáveis de interesse são medidas por períodos, como por exemplo, o montante aplicado no mês. Nestes casos, o uso de equações de diferença é uma poderosa ferramenta de análise como veremos nas seções 1 e 2.

Este capítulo discute as equações de diferença e diferenciais de 1ª ordem e, dentre essas, apenas algumas delas, que são bastante simples e têm aplicabilidade direta em Economia. Mais especificamente, os tópicos que serão abordados neste capítulo são: i) Equações de diferença de 1ª ordem; ii) Aplicações das equações de diferença de 1ª ordem; iii) Equações diferenciais ordinárias de 1ª ordem separáveis; iv) Equações diferenciais ordinárias lineares de 1ª ordem; v) Diagramas de fase e campos direcionais. Assume-se que o leitor tenha um relativo domínio das técnicas de integração, ponto discutido no capítulo.

1. EQUAÇÕES DE DIFERENÇA DE 1ª ORDEM

Nas equações de diferença, as variáveis são medidas em períodos, ou seja, assumem o mesmo valor por todo um período. Portanto, a variação das mesmas se dá entre períodos. Por exemplo, mede-se o PIB em um ano específico e compara-se com o PIB do ano anterior, obtendo a variação anual do PIB. Assim, temos a variável de interesse, y , representada em dois períodos diferentes, y_t e y_{t+1} . A diferença entre eles é $\Delta y = y_{t+1} - y_t$. Uma vez que as diferenças são calculadas entre períodos, não se utiliza o cálculo diferencial.

Este capítulo apresenta as equações de diferença de 1ª ordem. Assim, temos somente a diferença entre esses dois períodos apresentada acima. Como comparação, uma equação de 2ª ordem tem, além dos valores da variável de interesse nesses dois períodos, a informação em um terceiro período, y_{t+2} .

Dentre as equações de diferença de 1ª ordem, o capítulo discute as equações lineares com coeficientes constantes. Ou seja, um pequeno universo dentre as equações de diferença de 1ª ordem, que são de fácil resolução e são utilizadas em importantes aplicações em Economia.

Uma equação deste tipo pode ser escrita em diferentes formatos, sendo que em um, em particular, a interpretação da dinâmica da variável y se torna mais simples. Esse formato é o seguinte:

$$y_{t+1} = ay_t + b, \text{ onde } a \text{ e } b \text{ são constantes.}$$

Note que, o valor da variável no tempo subsequente, y_{t+1} , é expresso como uma constante vezes o valor da variável no tempo precedente, y_t , mais outra constante. Uma equação desse tipo é facilmente interpretável dependendo dos valores de a e b . Pode-se dividi-la em três tipos:

Tipo 1: $a = 1$ e $b \neq 0$;

Tipo 2: $a \neq 1$ e $b = 0$;

Tipo 3: $a \neq 1$ e $b \neq 0$.

O quarto tipo em potencial, com $a = 1$ e $b = 0$, não apresenta qualquer dinâmica, pois $y_{t+1} = y_t$. Ou seja, o valor da variável y não sofre qualquer mudança e não é um problema dinâmico, portanto, não é discutido aqui.

Cada um desses três tipos de equações de diferença de 1ª ordem lineares com coeficientes constantes tem um formato de resolução muito bem definido. Segue cada uma delas em separado. Para os dois primeiros tipos a obtenção da resolução é direta.

O primeiro tipo de equação de diferença pode ser escrito como:

$$y_{t+1} = y_t + b, \text{ com } b \neq 0.$$

Ou seja, o valor da variável em um período subsequente é o valor no período precedente acrescido de uma constante. Portanto, depois de um número determinado de períodos, o acréscimo é b vezes o

número de períodos. Assim, a obtenção da solução é direta, sendo que y_0 é o valor da variável no início.

$$y_t = y_0 + bt.$$

Como exemplo numérico, temos a seguinte equação do tipo 1:

$$y_{t+1} = y_t + 3, \text{ com } y_0 = 2.$$

Cuja solução é:

$$y_t = 2 + 3t$$

As equações do tipo 2 tem o seguinte formato:

$$y_{t+1} = ay_t, \quad \text{com } a \neq 1.$$

Elas apresentam o seguinte formato de resposta, também obtido de forma direta:

$$y_t = (a)^t y_0.$$

Como exemplo numérico, temos:

$$y_{t+1} = \left(\frac{1}{2}\right)y_t \text{ com } y_0 = 3.$$

Cuja solução é:

$$y_t = 3(1/2)^t$$

Para as equações do tipo 3, a solução não é tão direta como as anteriores, e é deduzida a seguir.

Partimos da equação de diferença, $y_{t+1} = ay_t + b$, sabendo que $a \neq 1$ e $b \neq 0$.

Escrevemos a relação:

$$\text{para } t = 0, \text{ temos } y_1 = ay_0 + b;$$

$$\text{para } t = 1, \text{ temos } y_2 = ay_1 + b; \text{ e}$$

$$\text{para } t = 2, y_3 = ay_2 + b.$$

Substituindo y_1 da primeira relação na segunda, temos:

$$y_2 = a(ay_0 + b) + b$$

Em seguida, substituímos esse resultado na terceira das relações:

$$y_3 = a[a(ay_0 + b) + b] + b$$

Reescrevemos essas duas últimas expressões no seguinte formato:

$$y_2 = a^2 y_0 + b(1 + a)$$

$$y_3 = a^3 y_0 + b(1 + a + a^2)$$

Assim, a partir da análise dessas expressões, deduzimos que para um t qualquer, temos:

$$y_t = a^t y_0 + b(1 + a + a^2 + \dots + a^{t-1})$$

Resta saber quanto vale o somatório entre parênteses:

$$S = 1 + a + a^2 + \dots + a^{t-1}$$

Multiplicamos toda a expressão por a e subtraímos o resultado da própria relação:

$$aS = a + a^2 + a^3 + \dots + a^t$$

$$aS - S = a^t - 1$$

$$(a - 1)S = a^t - 1$$

$$S = \frac{a^t - 1}{a - 1}.$$

Assim, chegamos a uma expressão que será utilizada sempre que a equação de diferença de 1ª ordem for do tipo 3:

$$y_t = a^t y_0 + b \left[\frac{a^t - 1}{a - 1} \right].$$

Como exemplo numérico, temos:

$$y_{t+1} = 2y_t + 3 \text{ com } y_0 = 4.$$

$$y_t = 2^t 4 + 3 \left[\frac{2^t - 1}{2 - 1} \right] = 2^t 7 - 3$$

2. APLICAÇÕES DAS EQUAÇÕES DE DIFERENÇA DE 1ª ORDEM

Nesta seção são apresentadas quatro aplicações das equações de diferença de 1ª ordem, sendo as três primeiras referentes às aplicações monetárias, cada uma abordando um dos três tipos de equação de diferença. A quarta aplicação é o modelo da teia de aranha.

2.1. Aplicações monetárias

A primeira das aplicações monetárias é quanto às **aplicações a juros simples**. A equação de diferença que representa uma aplicação desse tipo é:

$M_{t+1} = M_t + M_0 J$, onde M é o montante do investimento nos diferentes períodos, M_0 é o montante inicial aplicado e J é a taxa de juros simples.

Comparando essa equação com os três tipos discutidos na seção anterior, verifica-se que essa equação é do tipo 1 e tem como solução:

$$M_t = M_0 + JM_0 t.$$

Como exemplo numérico, assuma que o montante inicial é igual a R\$1000, que a taxa de juros simples é 1% ao mês, e que a aplicação é feita durante 12 meses. Então, ficamos com:

$$M_{12} = 1000 + (0,01)(1000)(12) = 1120.$$

O segundo exemplo de aplicação das equações de diferença de 1ª ordem é referente às **aplicações a juros compostos**. No caso, seguimos os mesmos conceitos descritos acima, porém, como a taxa de juros é composta, temos a seguinte equação de diferença:

$$M_{t+1} = (1 + J)M_t.$$

Esta equação é do tipo 2 e, como vimos, tem como solução:

$$M_t = (1 + J)^t M_0.$$

O exemplo numérico segue o mesmo montante inicial do problema anterior de R\$1000, a taxa de juros composta é de 0,5% ao mês e a aplicação é feita durante 15 meses:

$$M_{15} = (1 + 0,005)^{15} 1000 = 1077,68$$

Por fim, aborda-se o problema das **aplicações com juros compostos com depósitos ou retiradas fixos por período**. A equação de diferença que representa esse problema é dada por:

$M_{t+1} = (1+J)M_t + d$, onde d é o depósito ou retirada fixos em cada período e o restante já foi definido.

Essa equação é representada pelo tipo 3 e tem como solução:

$$M_t = (1+J)^t M_0 + d \left[\frac{(1+J)^t - 1}{1+J-1} \right] = (1+J)^t \left(M_0 + \frac{d}{J} \right) - \frac{d}{J}.$$

No exemplo numérico, temos o mesmo montante inicial de R\$1000, taxa de juros composta também de 0,5% ao mês, depósito mensal de R\$200 e o tempo de aplicação de 10 meses. Assim, temos:

$$M_t = (1+0,005)^{10} \left(1000 + \frac{200}{0,005} \right) - \frac{200}{0,005} = 3096,75$$

2.2. O modelo da teia de aranha

O modelo da teia de aranha é uma aplicação das equações de diferença que também é bastante simples (Chiang, 2006). Assuma que a demanda por um bem é linear com relação ao preço do próprio bem em um mesmo período:

$$Q_t = a + bP_t.$$

Assuma que a oferta do bem também é linear, mas que é definida a partir dos preços aplicados em um período precedente, pois o produtor tem a expectativa que os preços em um determinado período serão mantidos do período precedente:

$$Q_t = c + dP_{t-1}.$$

Igualando a oferta com a demanda, temos:

$$a + bP_t = c + dP_{t-1}.$$

Reescrevendo essa equação, ficamos com:

$$P_t = \frac{d}{b} P_{t-1} + \frac{c-a}{b}.$$

Para representar essa equação no formato discutido anteriormente, temos:

$$P_{t+1} = \frac{d}{b} P_t + \frac{c-a}{b}.$$

Assumindo que $d \neq b$ e que $c \neq a$, temos uma equação de diferença do terceiro tipo. Portanto, a solução desse problema é expressa como:

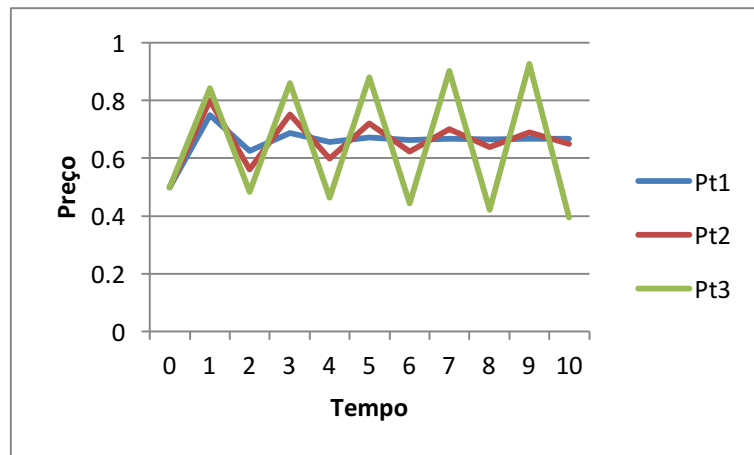
$$P_t = \left(\frac{d}{b}\right)^t P_0 + \left(\frac{c-a}{b}\right) \left[\frac{\left(\frac{d}{b}\right)^t - 1}{\left(\frac{d}{b}\right) - 1} \right] = \left(\frac{d}{b}\right)^t \left(P_0 - \frac{c-a}{d-b} \right) - \left(\frac{c-a}{d-b} \right).$$

Se $P_e = -\left(\frac{c-a}{d-b}\right)$, onde, como veremos, P_e é preço de equilíbrio, ficamos com:

$$P_t = \left(\frac{d}{b}\right)^t (P_0 - P_e) + P_e.$$

A dinâmica do preço e, conseqüentemente, também da quantidade ofertada/demandada desse modelo depende dos valores dos parâmetros a , b , c e d . O diagrama 1 mostra três dinâmicas diferentes obtidas com grupos distintos de parâmetros, definidos de forma que P_e é o mesmo em todos eles. P_0 também é o mesmo em todas as ilustrações. Além disso, como é mais usual, define-se $b < 0$ e $d > 0$. Em Pt1, observa-se uma convergência rápida nos preços, pois $\left|\frac{d}{b}\right| = 0,5$. Em Pt2, a convergência é mais lenta, pois $\left|\frac{d}{b}\right| = 0,8$, e em Pt3 ocorre uma divergência nos preços, pois $\left|\frac{d}{b}\right| > 1$.

DIAGRAMA 1
Modelo da teia de aranha



Esta seção e a anterior apresentam alguns conceitos introdutórios das equações de diferença de 1ª ordem. Na seção seguinte abordam-se também de forma introdutória alguns conceitos associados às equações diferenciais de 1ª ordem. Existem inúmeros tipos de equações diferenciais distintas e neste capítulo discutem-se algumas delas que, não só são bastante simples, como apresentam aplicabilidade direta em Economia.

3. EQUAÇÕES DIFERENCIAIS ORDINÁRIAS DE 1ª ORDEM

Como forma de ilustrar qual é o foco desse capítulo, apresenta-se uma breve classificação de equações diferenciais. O primeiro ponto a ser destacado é que existem as equações diferenciais ordinárias e as equações diferenciais parciais. Aqui, discutem-se apenas as equações diferenciais ordinárias. Nestas equações temos funções de uma variável, derivadas de uma variável e parâmetros. Exemplo disso são as equações $y' = ky$ e $y' + ay + by = 0$. Em contrapartida, as equações diferenciais parciais contêm

derivadas parciais, como a equação de transferência de calor em uma dimensão: $\frac{\partial T}{\partial t} = a \frac{\partial^2 T}{\partial x^2}$.

Assim, tratamos apenas das equações diferenciais ordinárias. Dentre essas, analisamos apenas aquelas de 1ª ordem, ou seja, a derivada de maior ordem é a derivada primeira, como em $y' = ky$. Equações diferenciais ordinárias de 2ª ordem, como $y'' + ay' + by = 0$, ou de ordem superior, não são discutidas. Portanto, neste capítulo são abordadas as equações diferenciais ordinárias de 1ª ordem, um pequeno universo dentre as equações diferenciais. Ainda assim, existem diversos tipos dentre essas equações, e apenas dois tipos distintos são analisados. A próxima subseção apresenta as equações separáveis e a seguinte as equações lineares. Ambas têm muitas aplicações em economia e são relativamente simples.

3.1. Equações separáveis

Muitas das equações diferenciais ordinárias de 1º ordem separáveis que são aplicadas em problemas na Economia podem ser resolvidas utilizando integrais não muito complicadas. Inicialmente, é discutido brevemente o método de resolução dessas equações. Depois, são apresentadas três aplicações destas em Economia.

As equações diferenciais ordinária de 1ª ordem separáveis podem ser representadas da seguinte forma:

$$\frac{dy}{dx} = f(x)g(y).$$

Assim, para resolver esse tipo de equação diferencial basta separar as variáveis e integrar:

$$\begin{aligned}\frac{dy}{g(y)} &= f(x)dx \\ \int \frac{dy}{g(y)} &= \int f(x)dx + C, \text{ onde } C \text{ é constante}\end{aligned}$$

Vejamos três exemplos aplicados.

3.1.1. Variáveis com variação a taxas constantes

Utilizam-se as equações separáveis em problemas em que a variável de interesse varia com taxas constantes, como preços em taxa de inflação constante, PIBs variando a taxas constantes, etc. Segue um exemplo em que a população de uma dada região varia com taxas constantes. A seguinte equação diferencial que descreve esse problema é:

$$\frac{dP}{dt} = kP, \text{ onde } P \text{ é população e } k \text{ é a taxa de crescimento geométrica constante.}$$

Note que a variação populacional, $\frac{dP}{dt}$, é igual ao produto da população instantânea e da taxa de variação constante. Para resolver esse problema, basta seguir os passos descritos anteriormente:

$$\begin{aligned}\frac{dP}{P} &= kdt \\ \int \frac{dP}{P} &= \int kdt + C \\ \ln P &= kt + C \\ P(t) &= e^{kt+C} = e^{kt} e^C = C_0 e^{kt}\end{aligned}$$

Para que o valor de C_0 seja definido necessita-se de uma condição inicial. Utilizando a informação sobre P_0 , a população inicial, ficamos com:

$$P(0) = C_0 = P_0$$

Então, a função $P(t) = P_0 e^{kt}$ é a solução dessa equação diferencial ordinária de 1ª ordem separável. Note como essa função satisfaz a equação diferencial.

3.1.2. Crescimento logístico

Outro exemplo de equação separável é a que trata da variação logística de uma variável. A equação diferencial que representa esse fenômeno pode ser escrita da seguinte forma:

$$\frac{dy}{dt} = (a - by)y.$$

Esse tipo de equação é muito utilizado em estudos com capacidade de carregamento limite, como em estudos ecológicos e trocas de tecnologia em aplicações diversas (Banks, 1994). Ela pode ser resolvida analiticamente de forma relativamente simples, com o uso de integração por frações parciais. Entretanto, essa equação é analisada na próxima seção com o uso de diagrama de fase e de campos direcionais.

A função logística pode ser representada pela seguinte equação:

$$y(t) = \frac{a}{b + \left(\frac{a}{y_0} - b \right) e^{-at}}.$$

3.1.3 - Análise de aversão ao risco

Uma terceira aplicação é quanto a medidas de aversão ao risco relativo de Arrow-Pratt.

Seja $u(x)$ a função de utilidade do indivíduo avesso ao risco, sendo que a medida de aversão ao risco é dada por:

$$b = -\frac{u''(x)x}{u'(x)}, \text{ onde } b > 0 \text{ é constante.}$$

Note que essa equação diferencial é de 2ª ordem, tema não abordado aqui. Entretanto, ela será transformada em uma equação de 1ª ordem, que será resolvida pelo método descrito anteriormente e, em seguida, outra equação diferencial de 1ª ordem será obtida e também resolvida.

O objetivo desse problema aplicado é obter uma função utilidade que apresente uma medida de aversão ao risco como definido anteriormente. Inicialmente define-se uma função $v(x)$, tal que $v(x) = u'(x)$. Dessa forma, a equação diferencial de 2ª ordem transforma-se em uma de 1ª ordem:

$$b = -\frac{v'(x)x}{v(x)}.$$

Resolvendo essa equação pelo método já descrito, temos:

$$b = -\frac{(dv/dx)x}{v}$$

$$\frac{dv}{v} = -b \frac{dx}{x}$$

$$\int \frac{dv}{v} = -b \int \frac{dx}{x} + C_1$$

$$\ln(v) = -b \ln x + C_1 = \ln x^{-b} + C_1.$$

$$v(x) = e^{\ln x^{-b} + C_1} = e^{\ln x^{-b}} e^{C_1} = C_2 x^{-b}, \text{ onde } C_2 \text{ é constante.}$$

Assim, a equação de 1ª ordem foi resolvida com a obtenção da função $v(x)$ que satisfaz a equação.

Continua-se a resolução do problema com uma nova equação de 1ª ordem: $v(x) = u'(x) = \frac{du}{dx}$.

A resolução é pelo mesmo método:

$$u(x) = \int v(x) dx + C_3 = \int C_2 x^{-b} dx + C_3.$$

Isso implica em:

a) Se $b \neq 1$, $u(x) = \frac{C_2}{1-b} x^{1-b} + C_3.$

b) Se $b = 1$, $u(x) = C_2 \ln x + C_3.$

Verifica-se que essas duas funções de utilidade satisfazem a medida de aversão ao risco relativo de Arrow-Pratt.

3.2. Equações lineares

Além das equações separáveis, outro tipo de equação diferencial de 1ª ordem muito utilizada em Economia é o de equações lineares. Essas equações podem ser representadas pela seguinte equação:

$$\frac{dy}{dt} + f(t)y = R(t).$$

As equações diferenciais de 1ª ordem que podem ser escritas nesse formato são lineares e são resolvidas por métodos bem definidos. Segue a demonstração de uma das metodologias empregadas na resolução desse tipo de equação.

Dada uma equação linear geral, como a descrita anteriormente, multiplica-se a equação por $e^{\int f(t)dt}$:

$$e^{\int f(t)dt} y' + e^{\int f(t)dt} f(t)y = e^{\int f(t)dt} R(t).$$

Note que, ao ser feita essa multiplicação, a parte a esquerda da igualdade representa a derivada de um produto de funções.

Assim, obtemos:

$$\frac{d}{dt} \left[e^{\int f(t)dt} y \right] = e^{\int f(t)dt} R(t).$$

Rearranjando e integrando, temos:

$$\int d \left[e^{\int f(t)dt} y \right] = \int e^{\int f(t)dt} R(t) dt + C$$

$$y(t) = \frac{1}{e^{\int f(t)dt}} \left[\int e^{\int f(t)dt} R(t) dt + C \right].$$

Apenas para simplificar, reescreve-se a expressão acima com $h = \int f(t)dt$:

$$y(t) = e^{-h} \left[\int e^h R(t) dt + C \right]$$

Essas duas expressões são utilizadas na resolução de equações diferenciais ordinárias lineares de 1ª ordem, como mostrado em um exemplo aplicado.

3.2.1. Variação populacional com migração constante

Assume-se que a população de uma região cresce vegetativamente a taxas constantes, como discutido na subseção 3.1.1. Além disso, a região recebe imigrantes, I , e perde emigrantes, E , de forma constante.

A equação diferencial a seguir representa esse problema:

$$\frac{dP}{dt} = kP + I - E, \text{ onde } I \text{ e } E \text{ são constantes.}$$

Reescreve-se essa equação no formato linear descrito anteriormente:

$$\frac{dP}{dt} - kP = I - E$$

Note que por comparação entre essa expressão e o formato padrão de equação linear, temos:

$$f(t) = -k \text{ e } R(t) = I - E.$$

Assim, uma vez definidos os termos $f(t)$ e $R(t)$, inicialmente, obtém-se h :

$$h = \int f(t)dt = \int -kdt = -kt.$$

Depois se substitui o valor de h e de $R(t)$ na outra expressão e integra-se:

$$\begin{aligned} P(t) &= e^{-h} \left[\int e^h R(t) dt + C \right] = e^{kt} \left[\int e^{-kt} (I - E) dt + C \right] \\ P(t) &= e^{kt} \left[(I - E) \frac{e^{-kt}}{-k} + C \right]. \end{aligned}$$

Utilizando a condição inicial $P(0) = P_0$, temos:

$$\begin{aligned} P(0) &= (I - E) \frac{1}{-k} + C = P_0 \\ C &= P_0 + \frac{(I - E)}{k} \end{aligned}$$

Substituindo, C na expressão de $P(t)$ obtemos a solução final do problema:

$$P(t) = e^{kt} \left[(I - E) \frac{e^{-kt}}{-k} + P_0 + \frac{(I - E)}{k} \right] = P_0 e^{kt} + \frac{(I - E)}{k} (e^{kt} - 1)$$

4. DIAGRAMA DE FASE E CAMPOS DIRECIONAIS

A equação diferencial que representa uma variação logística foi apresentada na subseção 3.1.2:

$$\frac{dP}{dt} = (a - bP)P.$$

Entretanto, essa equação não foi resolvida analiticamente, o que não requer um aparato matemático sofisticado, mas foi deixada para ser analisada a partir de um diagrama de fase e de campos direcionais. Os diagramas de fase e os campos direcionais permitem analisar fenômenos dinâmicos sem que uma solução analítica seja necessária.

Para facilitar a compreensão sobre o uso desses conceitos, assume-se que na equação anterior $a = 2$ e $b = 1$. Além disso, sabe-se que $t \geq 0$ e $P \geq 0$, pois tempo e população não podem ser negativos.

Portanto, temos a seguinte equação diferencial a ser analisada por um diagrama de fase na próxima subseção:

$$\frac{dP}{dt} = (2 - P)P.$$

4.1. Diagramas de fase

O objetivo dessa subseção é desenhar um diagrama $P \times \frac{dP}{dt}$ e, a partir deste, obter informações sobre a dinâmica do problema. Inicia-se a discussão com a obtenção dos estados estacionários ou de equilíbrios, que irão facilitar a obtenção do gráfico proposto acima. Nesses pontos, temos:

$$\frac{dP}{dt} = 0.$$

Portanto, os pontos de equilíbrio são:

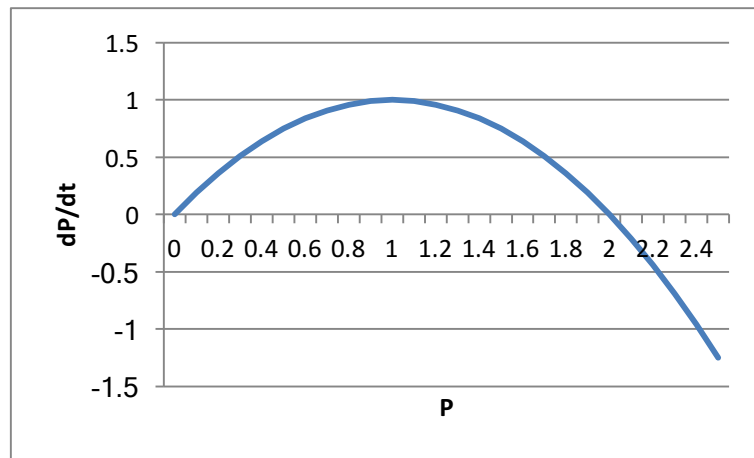
$$(2 - P)P = 0$$

$$P = 0 \text{ e } P = 2.$$

Uma vez determinados esses pontos, obtém-se o diagrama de fase, sabendo que a equação

$$\frac{dP}{dt} = (2 - P)P = 2P - P^2 \text{ é uma parábola com raízes } 0 \text{ e } 2.$$

DIAGRAMA 2
Diagrama de fase da equação logística

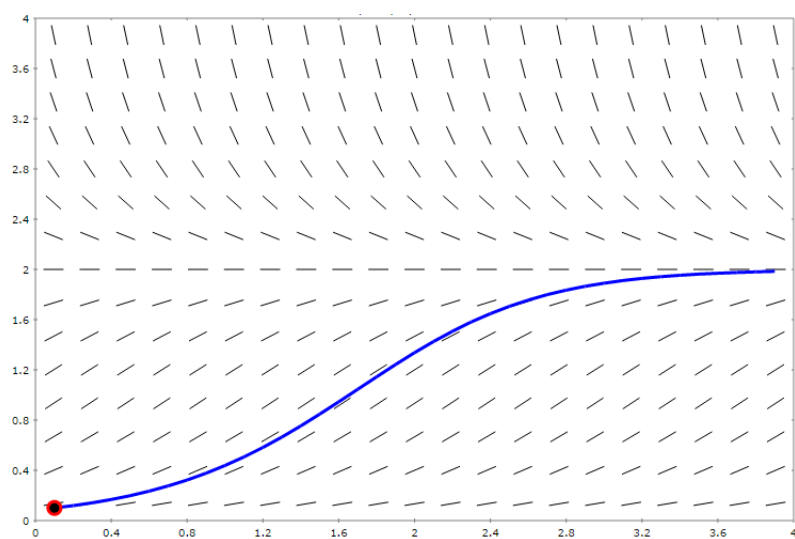


Para estudarmos os equilíbrios dos estados estacionários, fazemos uma pequena perturbação em torno dos mesmos. Note que o equilíbrio em $P = 0$ é instável. Uma vez que uma pequena perturbação com P pequeno e positivo, ou seja, a população era zero e se tornou pequena e positiva, faz com que o sistema se distancie desse ponto, pois $\frac{dP}{dt} > 0$ e P tende a crescer. Por outro lado, o equilíbrio em $P = 2$ é estável. Uma perturbação em torno do ponto, tanto para valores menores como maiores do que $P = 2$, dados os valores de $\frac{dP}{dt}$, respectivamente positivos e negativos, indicam uma tendência de retorno ao estado estacionário com $P = 2$.

4.2. Campos direcionais

Para a obtenção do campo direcional, baseia-se no diagrama acima, porém o diagrama a ser feito é de $t \times P$. Em valores de P próximos de 0,2, como pode ser visto no diagrama de fase, temos uma derivada positiva de pequena magnitude. Assim, desenhamos no diagrama do campo direcional um segmento de reta leve e positivamente inclinado, como mostrado no diagrama a seguir. Para valores de P próximos de 1, temos uma derivada positiva de maior magnitude e desenhamos um segmento de reta com inclinação positiva de maior magnitude. Seguimos esse mesmo procedimento para valores de P próximos de 1,8, temos um segmento de reta com inclinação levemente positiva, para valores de P próximos de 2,2, a inclinação é levemente negativa, e para valores próximos de 2,5, a inclinação é negativa com maior módulo. Como as inclinações independem de t , repetimos esses segmentos de reta para diferentes valores de t . Uma vez obtidos esses segmentos, representam-se possíveis trajetórias para a dinâmica do problema.

DIAGRAMA 3
Campos direcionais da equação logística



CAPÍTULO 5 - SISTEMAS DE EQUAÇÕES DE DIFERENÇA DE 1ª ORDEM

Existem muitos problemas econômicos que tratam de fenômenos dinâmicos, tais como evolução de taxas de desemprego, crescimento econômico, crescimento populacional, etc. Como vimos no capítulo anterior, as equações de diferença são poderosas ferramentas que permitem estudar problemas econômicos desse tipo, quando a dinâmica da variável de interesse é determinada pelos próprios valores desta. Entretanto, em muitos problemas econômicos, a dinâmica da variável envolve mais de uma variável, que se influenciam. O uso de sistemas de equações de diferença em muito facilita a resolução de problemas assim. Um tópico correlato, que é o uso dos sistemas de equações diferenciais de 1ª ordem, é abordado no sexto capítulo. Este quinto capítulo foi dividido nas seguintes seções: i) Introdução aos sistemas de equações de diferença lineares de 1ª ordem; ii) Sistemas de equações de diferença lineares de 1ª ordem homogêneos; iii) Autovalores reais aplicados à resolução de sistemas lineares; iv) Autovalores complexos aplicados à resolução de sistemas lineares; v) Autovalores repetidos aplicados à resolução de sistemas lineares; vi) Estabilidade de sistemas de equações de diferença e vii) Aplicações de sistema de equações de diferença. Assume-se que o leitor tenha conhecimento sobre conceitos básicos de álgebra linear, como já discutido no primeiro capítulo, e de equações de diferença de 1ª ordem, como apresentado no capítulo anterior.

1. INTRODUÇÃO AOS SISTEMAS DE EQUAÇÕES DE DIFERENÇA LINEARES DE 1ª ORDEM

No capítulo anterior foram apresentadas as equações de diferença de 1ª ordem. Com o intuito de facilitar a interpretação, elas foram escritas no seguinte formato:

$$y_{t+1} = ay_t + b.$$

Quando analisamos equações assim, a dinâmica da variável y depende somente do valor da própria variável e de parâmetros. Entretanto, em muitos problemas econômicos, a dinâmica de uma variável depende não só de seus próprios valores como também do valor de outras variáveis. Por exemplo, a dinâmica da variável y depende do valor da própria variável e também do valor da variável x . Similarmente, o mesmo ocorre para essa segunda variável, cuja dinâmica irá depender de ambas variáveis.

Em problemas em que dinâmicas de diferentes variáveis interagem entre si, podem-se representar essas dinâmicas com influência mútua com um sistema de equações de diferença. Existem diversos tipos de sistemas de equações de diferença. Toda a discussão apresentada aqui será feita para sistemas de equações de diferença lineares de 1ª ordem com duas equações e duas incógnitas. Esses sistemas, além de apresentarem muitas aplicações em economia, permitem uma apresentação mais breve de diversos conceitos, sendo que, em geral, esses podem ser generalizados para sistemas maiores.

Um sistema de equações de 1ª ordem linear genérico pode ser escrito como uma extensão da equação de diferença linear escrita acima, onde a s e b s são parâmetros:

$$\begin{aligned}x_{t+1} &= a_{11}x_t + a_{12}y_t + b_1 \\y_{t+1} &= a_{21}x_t + a_{22}y_t + b_2\end{aligned}$$

Note que a dinâmica da variável x depende dela própria e de y , além de parâmetros. Para y , idem. O sistema acima é um sistema de equações de diferença lineares de 1ª ordem não homogêneo. Note que esse sistema é muito bem representado fazendo uso da álgebra linear:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

Como veremos, o uso da álgebra linear irá facilitar em muito a resolução de problemas. Este capítulo analisa apenas os sistemas homogêneos, onde b_1 e b_2 são nulos:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}.$$

Para uma discussão sobre sistemas não homogêneos, ver Gandolfo (1997).

2. SISTEMAS DE EQUAÇÕES DE DIFERENÇA LINEARES DE 1ª ORDEM HOMOGÊNEOS

Este capítulo tem como foco os sistemas homogêneos, que são relativamente simples e apresentam grande aplicabilidade em Economia. Inicia-se a discussão com os sistemas desacoplados, em que a matriz com os coeficientes do sistema de equações é diagonal:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

Na verdade, o sistema conta com duas equações de diferença separadas e independentes: $x_{t+1} = a_{11}x_t$ e $y_{t+1} = a_{22}y_t$. Ou seja, elas são simplesmente equações de diferença de 1ª ordem do tipo 2, como discutido no capítulo anterior. O valor da variável x no período subsequente depende de uma constante a_{11} e de seu valor no período corrente. O mesmo ocorre com y . As soluções de equações assim já foram apresentadas no capítulo anterior: $x_t = a_{11}^t x_0$ e $y_t = a_{22}^t y_0$, onde x_0 e y_0 são os respectivos valores iniciais. Diz-se que as equações são desacopladas porque não existe uma interação entre as variáveis.

Entretanto, uma variável pode depender também de outra variável caso haja a interação entre elas, a matriz não é diagonal e deve-se desenvolver um método de resolução. A resolução de sistemas

desacoplados é bastante simples, como vimos. O método de resolução de sistemas acoplados faz uso de autovalores e autovetores, conceitos apresentados no primeiro capítulo.

Para facilitar a descrição do método, escreve-se o sistema, $\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}$, em um formato mais sucinto: $X_{t+1} = AX_t$.

Assim, o passo inicial é partir de um sistema acoplado, $X_{t+1} = AX_t$, e transformá-lo em um sistema desacoplado, $Z_{t+1} = DZ_t$, onde D é uma matriz diagonal. Em seguida, resolve-se esse sistema desacoplado, como já mostrado aqui. Por fim, transforma-se a resposta do sistema desacoplado em uma resposta associada ao sistema acoplado.

Para tanto, assume-se que existe uma matriz P invertível tal que se pode escrever a transformação de X_t em Z_t da seguinte forma: $X_t = PZ_t$ ou $Z_t = P^{-1}X_t$.

Então, partindo dessa última expressão, temos:

$$Z_t = P^{-1}X_t.$$

Em um período posterior:

$$Z_{t+1} = P^{-1}X_{t+1}.$$

Sabendo que $X_{t+1} = AX_t$, temos:

$$Z_{t+1} = P^{-1}AX_t.$$

Utilizando a relação $X_t = PZ_t$:

$$Z_{t+1} = P^{-1}APZ_t = (P^{-1}AP)Z_t$$

Sabendo que $Z_{t+1} = DZ_t$:

$$Z_{t+1} = (P^{-1}AP)Z_t = DZ_t$$

Assim, conclui-se que se as suposições feitas acima forem verdadeiras: $D = P^{-1}AP$.

Para definirmos quais são as matrizes D e P , partimos dessa última relação e multiplicamos os dois lados por P . Ficamos com:

$$AP = PD.$$

Como estamos analisando sistemas com matriz de coeficientes 2×2 , escreve-se $P = [V_1 \ V_2]$, onde V_1 e V_2 são as colunas da matriz P . Além disso, representamos a matriz diagonal como $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

Assim, a relação acima toma o seguinte formato:

$$A[V_1 \ V_2] = [V_1 \ V_2] \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Multiplicando essas matrizes, ficamos com:

$$[AV_1 \ AV_2] = [\lambda_1 V_1 \ \lambda_2 V_2]$$

Isso implica em $AV_1 = \lambda_1 V_1$ e $AV_2 = \lambda_2 V_2$, que representam a própria definição de autovalores e autovetores. Ou seja, a matriz P tem como colunas os autovetores da matriz A e a matriz D é formada pelos autovalores da mesma.

Assim, o sistema desacoplado, $Z_{t+1} = DZ_t$, pode ser escrito como:

$$\begin{pmatrix} z_{t+1} \\ w_{t+1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} z_t \\ w_t \end{pmatrix}.$$

Esse sistema tem como solução: $z_t = \lambda_1^t z_0 = A\lambda_1^t$ e $w_t = \lambda_2^t w_0 = B\lambda_2^t$, em que os valores iniciais de z_t e de w_t , z_0 e w_0 , foram substituídos para efeitos didáticos pelas constantes genéricas A e B .

Assim, temos a solução do sistema desacoplado, $Z_t = \begin{pmatrix} A\lambda_1^t \\ B\lambda_2^t \end{pmatrix}$. De posse desta, obtém-se a solução

final do sistema acoplado, fazendo uso da transformação, $X_t = PZ_t$:

$$X_t = PZ_t = (V_1 \ V_2) \begin{pmatrix} A\lambda_1^t \\ B\lambda_2^t \end{pmatrix} = A\lambda_1^t V_1 + B\lambda_2^t V_2.$$

Essa solução é obtida sempre que existirem as matrizes P e D . Isso ocorre quando temos dois autovalores reais ou dois autovalores complexos, como ilustrado numericamente nas seções 3 e 4, respectivamente. Entretanto, isso não necessariamente ocorrerá quando os autovalores forem repetidos, como veremos na seção 5.

No primeiro capítulo, onde foram introduzidos os conceitos de autovalores e autovetores, foram mostrados três exemplos numéricos. No primeiro foram obtidos dois autovalores reais, no segundo,

dois autovalores complexos conjugados e no terceiro uma raiz dupla real. Esses mesmos exemplos são utilizados nas três próximas seções a seguir como exemplo numérico.

3. AUTOVALORES REAIS APLICADOS À RESOLUÇÃO DE SISTEMAS LINEARES

Esta seção apresenta um exemplo numérico sobre como utilizar autovalores reais na resolução de sistemas de equações de diferença lineares de 1ª ordem. Seja um sistema representado por:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

O primeiro passo para resolver esse problema é obter os autovalores e autovetores da matriz

$$A = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix}.$$

Como vimos no primeiro capítulo, os autovalores e autovetores são:

$$\lambda_1 = 2$$

$$\lambda_2 = 3$$

$$V_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

$$V_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

De posse dos autovalores e autovetores, montamos a solução para o problema:

$$X_t = A\lambda_1^t V_1 + B\lambda_2^t V_2 = A2^t \begin{pmatrix} 1 \\ -2 \end{pmatrix} + B3^t \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

4. AUTOVALORES COMPLEXOS APLICADOS À RESOLUÇÃO DE SISTEMAS LINEARES

Como já discutido, todo o procedimento acima aplicado a duas raízes reais é válido se a matriz P tiver inversa, ou seja, sempre que tivermos dois autovetores distintos para formar essa matriz. Isso ocorre quando os autovalores são reais e distintos e também quando são complexos conjugados. Ou seja, a solução descrita nas seções anteriores é válida para ambos os casos. Entretanto, para valores complexos, deve-se reescrever essa solução para torná-la mais facilmente analisável.

Assim, partimos da solução obtida anteriormente:

$$X_t = A\lambda_1^t V_1 + B\lambda_2^t V_2.$$

Note que quando os autovalores são complexos, eles são conjugados:

$$\lambda_1 = \alpha + i\beta, \lambda_2 = \alpha - i\beta,$$

e o mesmo ocorre com os autovetores:

$$V_1 = u + iv, V_2 = u - iv.$$

Assim, a solução do sistema toma o seguinte formato:

$$X_t = A(\alpha + i\beta)^t (u + iv) + B(\alpha - i\beta)^t (u - iv).$$

Note que a soma de números complexos conjugados se torna um número real:

$(a + bi) + (a - bi) = 2a = 2\text{Re}(a + bi)$, onde Re indica que temos apenas a parte real do número complexo correspondente.

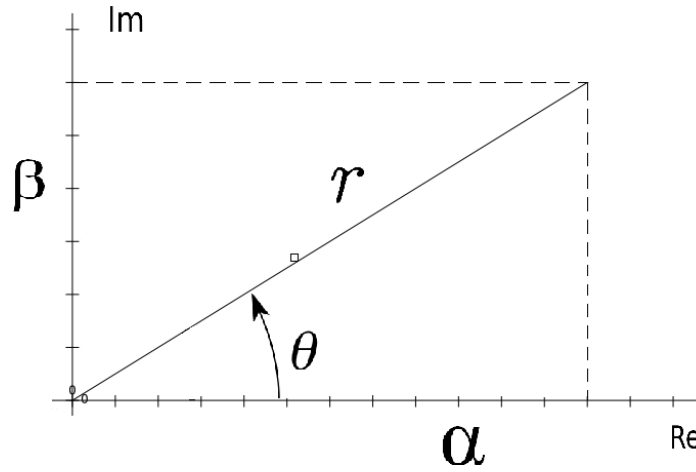
Usando desse artifício, se $A = C_1 + iC_2$ e $B = C_1 - iC_2$, em que C_1 e C_2 são constantes, a solução do sistema se torna uma soma de conjugados. Portanto, ela pode ser escrita como:

$$X_t = 2\text{Re}\{(C_1 + iC_2)(\alpha + i\beta)^t (u + iv)\}.$$

Para que essa solução seja mais facilmente analisável, deve-se simplificar o termo $(\alpha + i\beta)^t$.

O diagrama a seguir representa um número complexo $\alpha + i\beta$ em coordenadas polares:

DIAGRAMA 1
Número complexo em coordenadas polares



O diagrama permite obter as seguintes relações:

$$r = (\alpha^2 + \beta^2)^{1/2}, \quad \cos \theta = \frac{\alpha}{r} \text{ e } \sin \theta = \frac{\beta}{r}.$$

Assim, o número complexo $\alpha + i\beta$ é reescrito como:

$$\alpha + i\beta = r \left(\frac{\alpha}{r} + i \frac{\beta}{r} \right) = r(\cos \theta + i \sin \theta).$$

Isso implica, utilizando de Moivre, que: $(\alpha + i\beta)^t = r^t (\cos \theta + i \sin \theta)^t = r^t (\cos[t\theta] + i \sin[t\theta]).$

Substituindo essa última relação, obtém-se a seguinte expressão para a solução do sistema:

$$X_t = 2 \operatorname{Re}\{(C_1 + iC_2)r^t(\cos[t\theta] + i \sin[t\theta])(u + iv)\} = \operatorname{Re}\{(D_1 + iD_2)r^t(\cos[t\theta] + i \sin[t\theta])(u + iv)\},$$

em que D_1 e D_2 são constantes

Por fim, ficamos apenas com a parte real dessa expressão:

$$X_t = r^t \{D_1(\cos[t\theta]u - \sin[t\theta]v) - D_2(\sin[t\theta]u + \cos[t\theta]v)\}.$$

Esta é a expressão final que buscávamos. Segue um exemplo numérico para ilustrar o uso dessa expressão. Partimos do seguinte sistema:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

Inicialmente, obtêm-se os autovalores e autovetores da matriz $B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$, como já realizado no primeiro capítulo:

$$\lambda_1 = 1 + 2i$$

$$\lambda_2 = 1 - 2i$$

$$V_1 = \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + i \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$V_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - i \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

De posse dos autovalores e autovetores, montamos a solução para o problema:

$$r = (\alpha^2 + \beta^2)^{1/2} = (1^2 + 2^2)^{1/2} = 5^{1/2}.$$

$$\cos \theta = \frac{1}{5^{1/2}} \text{ e } \sin \theta = \frac{2}{5^{1/2}}, \text{ o que implica em } \theta \approx \frac{\pi}{2,84}.$$

Substituindo esses valores, bem como as partes reais e imaginárias dos autovetores, na expressão descrita acima, temos:

$$X_t = 5^{t/2} \left(D_1 \left(\cos \left[\frac{\pi}{2,84} t \right] \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \sin \left[\frac{\pi}{2,84} t \right] \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) - D_2 \left(\sin \left[\frac{\pi}{2,84} t \right] \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \cos \left[\frac{\pi}{2,84} t \right] \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \right)$$

5. AUTOVALORES REPETIDOS APLICADOS À RESOLUÇÃO DE SISTEMAS LINEARES

As duas seções anteriores apresentaram exemplos onde existiam dois autovetores distintos. Assim, pôde-se construir a matriz P , a partir desses autovetores, e obter a matriz diagonal. Diferentemente, quando os autovalores são repetidos, em geral, não se obtêm dois autovetores distintos. Isso ocorre apenas quando a matriz que deu origem ao autovalor repetido já for diagonal.

Por exemplo:

$$C = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

Essa matriz tem como autovalor duplo real, $\lambda_1 = \lambda_2 = 3$. Como a matriz já é diagonal e, portanto, representa um sistema desacoplado, é possível obter dois autovetores distintos e não múltiplos, por exemplo, $V_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $V_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Para uma matriz genérica em que os autovalores são iguais, isso não ocorrerá.

Como vimos, para a matriz $D = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix}$, os autovalores são $\lambda_1 = \lambda_2 = 3$, e temos apenas um autovetor: $V_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Como temos apenas um autovetor, não existe a matriz P , tal que $P^{-1}AP = D$, e também não existe a matriz diagonal com os autovalores. Portanto, como resolver um sistema como:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}?$$

Uma vez que a matriz diagonal não existe, partimos de uma matriz semelhante a esta, que é a quase diagonal (QD): $QD = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.

Propõem-se ainda uma matriz P tal que a relação $P^{-1}AP = QD$ exista.

Assim, seguindo passos similares aos utilizados para quando a matriz diagonal existia, temos:

$$AP = P \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

$$A[V_1 \ V_2] = [V_1 \ V_2] \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

$$[AV_1 \ AV_2] = [\lambda V_1 \ V_1 + \lambda V_2].$$

A primeira relação, $AV_1 = \lambda V_1$, indica que λ é o autovalor real duplo e V_1 é o autovetor da matriz A , ambos já obtidos numericamente anteriormente.

Da segunda relação, temos $AV_2 = V_1 + \lambda V_2$, que é rescrita como:

$$(A - I\lambda)V_2 = V_1, \text{ onde } V_2 \text{ é o autovetor generalizado.}$$

Note que, nos exemplos anteriores, a matriz diagonal existe, $P^{-1}AP = D$, e a matriz P é formada pelos autovetores da matriz A . Nesse caso com autovalores repetidos e um único autovetor, em que a matriz diagonal não existe, a matriz P é formada tendo como colunas o único autovetor da matriz A em conjunto com esse autovetor generalizado, $P = [V_1 \ V_2]$.

Seguindo o exemplo numérico, qual seria esse vetor?

$$(A - I\lambda)V_2 = V_1$$

$$\begin{pmatrix} 1-3 & 2 \\ -2 & 5-3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Temos então em ambas as relações que $-2x + 2y = 1$. Qualquer vetor que respeite essa relação pode ser o escolhido, por exemplo:

$$V_2 = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

O próximo passo é a obtenção da solução do sistema quase desacoplado:

$$Z_{t+1} = (QD)Z_t$$

$$\begin{pmatrix} z_{t+1} \\ w_{t+1} \end{pmatrix} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} z_t \\ w_t \end{pmatrix}.$$

No caso da variável w , tem-se uma equação desacoplada, $w_{t+1} = \lambda w_t$, cuja solução já conhecemos:

$$w_t = \lambda^t w_0.$$

Para z , temos uma equação acoplada: $z_{t+1} = \lambda z_t + w_t$. Para que a solução dessa equação seja inferida, escreve-se a mesma com diferentes valores de t :

$$z_1 = \lambda z_0 + w_0,$$

$$z_2 = \lambda z_1 + w_1 = \lambda(\lambda z_0 + w_0) + \lambda w_0 = \lambda^2 z_0 + (2\lambda)w_0,$$

$$z_3 = \lambda z_2 + w_2 = \lambda(\lambda^2 z_0 + (2\lambda)w_0) + \lambda^2 w_0 = \lambda^3 z_0 + (3\lambda^2)w_0$$

Por inspeção, infere-se que:

$$z_t = \lambda^t z_0 + t\lambda^{t-1}w_0.$$

De posse das soluções de w e z ,

$$Z_t = \begin{pmatrix} z_t \\ w_t \end{pmatrix} = \begin{pmatrix} \lambda^t z_0 + t\lambda^{t-1}w_0 \\ \lambda^t w_0 \end{pmatrix} = \begin{pmatrix} A\lambda^t + Bt\lambda^{t-1} \\ B\lambda^t \end{pmatrix},$$

em que as constantes foram trocadas por outras genéricas para fins didáticos, obtemos a solução do sistema acoplado por meio da relação $X_t = PZ_t$:

$$X_t = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} z_t \\ w_t \end{pmatrix} = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} A\lambda^t + Bt\lambda^{t-1} \\ B\lambda^t \end{pmatrix} = (A\lambda^t + Bt\lambda^{t-1})V_1 + B\lambda^t V_2.$$

Continuando o exemplo numérico:

$$X_t = (A3^t + Bt3^{t-1}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + B3^t \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

Em todos os exemplos numéricos, não se definiu os valores das constantes A e B . Esse procedimento será realizado na seção 7, quando serão mostradas duas aplicações de sistema de equações de diferença. A próxima seção discute a estabilidade dos sistemas tendo como base as expressões obtidas nesta e nas duas seções anteriores.

6. ESTABILIDADE DE SISTEMAS DE EQUAÇÕES DE DIFERENÇA

Um sistema de equações de diferença pode ser convergente, oscilante ou divergente. No primeiro caso, a solução de longo prazo assume um único valor bem definido, ou seja, $\lim_{t \rightarrow \infty} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$, em

que k_1 e k_2 são constantes. Um sistema diverge quando $\lim_{t \rightarrow \infty} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$, em que k_1 e/ou k_2

tendem para o infinito. Um sistema pode ainda apresentar uma solução oscilante, quando esta for finita, mas não assumir um único valor bem definido e, sim, um conjunto deles.

Segue uma análise sobre a convergência das soluções para cada uma das possibilidades de resposta discutidas nas seções anteriores. Assume-se que ambas constantes, A e B , são diferentes de zero.

6.1. Dois autovalores reais distintos

Como vimos, a solução quando temos dois autovalores reais distintos é dada por

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = A\lambda_1^t V_1 + B\lambda_2^t V_2. \text{ Note que:}$$

i) se $\lambda_1 \in (-1,1)$ e $\lambda_2 \in (-1,1)$, $\lim_{t \rightarrow \infty} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$;

ii) se $\lambda_1 \in (-1,1)$ e $\lambda_2 = 1$, $\lim_{t \rightarrow \infty} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = BV_2$;

iii) se $\lambda_1 \in (-1,1]$ e $\lambda_2 = -1$, $\lim_{t \rightarrow \infty} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = (-1)^t BV_2$, o sistema oscila entre dois valores; e

iv) $\lambda_1 \in (-\infty, -1) \cup (1, \infty)$ e /ou $\lambda_2 \in (-\infty, -1) \cup (1, \infty)$ o sistema diverge, com oscilações ou não.

6.2. Dois autovalores complexos

Partimos da solução para dois autovalores complexos que é dada por:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = r^t \{ D_1 (\cos[t\theta]u - \text{sen}[t\theta]v) - D_2 (\text{sen}[t\theta]u + \cos[t\theta]v) \}.$$

Note que:

i) se $r \in [0,1)$, ou seja, $\alpha^2 + \beta^2 < 1$, o sistema converge oscilando;

ii) se $r = 1$, o sistema oscila; e

iii) se $r > 1$, o sistema diverge oscilando.

6.3. Autovalores repetidos para matrizes não diagonais

Como vimos, a solução para autovalores repetidos para matrizes não diagonais é dada por:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = (A\lambda^t + Bt\lambda^{t-1})V_1 + B\lambda^t V_2.$$

Note que:

- i) se $\lambda \in (-1, 1)$, o sistema converge; e
- ii) se $|\lambda| \geq 1$, o sistema diverge.

7. APLICAÇÕES DE SISTEMA DE EQUAÇÕES DE DIFERENÇA

Os sistemas de equações de diferença são aplicados em diversos modelos econômicos. Serão apresentadas duas dessas aplicações: o modelo populacional de Leslie e o processo de Markov. Ambos os exemplos são facilmente resolvíveis com o ferramental discutido aqui.

7.1. Modelo populacional de Leslie

Considere uma espécie que pode viver no máximo durante dois anos. Sejam x os indivíduos no primeiro ano de vida e y os indivíduos no segundo. Sejam α_1 e α_2 as respectivas taxas de reprodução e β_1 a taxa de mortalidade dos indivíduos no primeiro ano de vida. O sistema de equações que traduz a dinâmica populacional dessa espécie é dado por:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 - \beta_1 & 0 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}.$$

Como exemplo numérico, assuma que $\alpha_1 = 1$ e que $\alpha_2 = 4$, isto é, cada indivíduo no primeiro ano de vida tem um filhote e cada indivíduo no segundo ano de vida tem quatro filhotes por ano. Além disso, assuma que $\beta_1 = 1/2$, ou seja, metade dos indivíduos morre entre o primeiro e o segundo ano de vida. Substituindo esses valores no sistema acima, temos:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 1/2 & 0 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

Os autovalores e autovetores da matriz $A = \begin{pmatrix} 1 & 4 \\ 1/2 & 0 \end{pmatrix}$ são: $\lambda_1 = 2$, $\lambda_2 = -1$, $V_1 = \begin{pmatrix} 1 \\ 1/4 \end{pmatrix}$ e

$$V_2 = \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}.$$

Assim, montamos a resposta do sistema:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = A2^t \begin{pmatrix} 1 \\ 1/4 \end{pmatrix} + B(-1)^t \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}$$

Note que as constantes A e B aparecem nas soluções de sistemas de equações de diferença apresentadas anteriormente nas seções 3, 4 e 5. Entretanto, até este ponto no capítulo não se discutiu como que estas constantes são estimadas. Isso será realizado nesta aplicação e também na próxima.

Obtêm-se os valores dessas constantes a partir de alguma informação complementar sobre o sistema. Uma informação que é comumente utilizada é a condição inicial do sistema. Por exemplo, assumamos que a população dessa espécie em estudo era, inicialmente, igualmente distribuída entre indivíduos no primeiro e no segundo ano de vida, em um total de 100 indivíduos:

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 50 \\ 50 \end{pmatrix}.$$

Utiliza-se essa informação na obtenção de A e B :

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = A \begin{pmatrix} 1 \\ 1/4 \end{pmatrix} + B \begin{pmatrix} 1 \\ -1/2 \end{pmatrix} = \begin{pmatrix} 50 \\ 50 \end{pmatrix}$$

Daí, tem-se que: $A = 100$ e $B = -50$.

Assim, após a obtenção do valor dessas constantes, a solução do problema se torna:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = 2^t \begin{pmatrix} 100 \\ 25 \end{pmatrix} + (-1)^t \begin{pmatrix} -50 \\ 25 \end{pmatrix}$$

A partir dessa expressão, podemos obter as populações para qualquer t . Por exemplo, em $t = 1$, temos:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = 2 \begin{pmatrix} 100 \\ 25 \end{pmatrix} + (-1) \begin{pmatrix} -50 \\ 25 \end{pmatrix} = \begin{pmatrix} 250 \\ 25 \end{pmatrix}$$

Ou seja, dentre os 50 indivíduos no primeiro ano de vida em $t = 0$, 25 sobreviveram para estar no segundo ano de vida em $t = 1$. Além disso, os 50 indivíduos no primeiro ano de vida em $t = 0$ tiveram 50 filhotes e os 50 indivíduos no segundo ano de vida em $t = 0$ tiveram 200 filhotes, totalizando 250 indivíduos no primeiro ano de vida em $t = 1$.

7.2. Processo de Markov

O processo de Markov é muito utilizado em estudos dinâmicos que envolvam probabilidades de transição. No exemplo apresentado aqui, temos as probabilidades de transição entre os estados de empregado e desempregado, como mostra a tabela 1. Dentre os empregados no tempo t , a probabilidade que o indivíduo continue empregado em $t + 1$ é q e a probabilidade que ele se torne um desempregado é $1 - q$. Similarmente para os desempregados em t , as probabilidades são respectivamente p e $1 - p$ de se tornar empregado ou continuar desempregado em $t + 1$.

TABELA 1
Processo de Markov

		Empregado no tempo t	
		Sim	Não
Empregado no tempo $t + 1$	Sim	q	p
	Não	$1 - q$	$1 - p$

Representando como x os empregados e y os desempregados, essa tabela representa o seguinte sistema de equações de diferença:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} q & p \\ 1 - q & 1 - p \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}.$$

Essa matriz com as probabilidades de transição, a **matriz de Markov**, é uma matriz com elementos não negativos cuja soma em cada coluna em separado é igual a um.

$$A = \begin{pmatrix} q & p \\ 1 - q & 1 - p \end{pmatrix}, \text{ com } p, q \in [0, 1].$$

O procedimento para resolver esse sistema de equações de diferença é o mesmo já realizado, e inicia-se com a obtenção dos autovalores e autovetores:

$$\begin{aligned} \begin{vmatrix} q - \lambda & p \\ 1 - q & 1 - p - \lambda \end{vmatrix} &= 0 \\ (q - \lambda)(1 - p - \lambda) - (p)(1 - q) &= 0 \\ \lambda^2 + (p - q - 1)\lambda + (q - p) &= 0 \\ \lambda_1 &= 1 \\ \lambda_2 &= q - p \end{aligned}$$

Para $\lambda_1 = 1$, o autovetor é $X_1 = \begin{pmatrix} 1 \\ (1-q)/p \end{pmatrix}$.

Para $\lambda_2 = q - p$, temos: $X_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

De posse dessa informação, montamos a resposta:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = A \begin{pmatrix} 1 \\ (1-q)/p \end{pmatrix} + B(q-p)^t \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Como exemplo numérico, assumamos que $q = 0,8$ e $p = 0,5$, então temos:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = A \begin{pmatrix} 1 \\ 2/5 \end{pmatrix} + B(0,3)^t \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Para determinar os valores das constantes, assumamos que a taxa de desemprego hoje é de 40%:

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 60 \\ 40 \end{pmatrix}$$

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = A \begin{pmatrix} 1 \\ 2/5 \end{pmatrix} + B \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 60 \\ 40 \end{pmatrix}.$$

Dessa relação, obtêm-se os valores de $A = 500/7$ e de $B = -80/7$. Substituindo na expressão acima, temos:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 500/7 \\ 200/7 \end{pmatrix} + (0,3)^t \begin{pmatrix} -80/7 \\ 80/7 \end{pmatrix}$$

Qual é a taxa de desemprego em $t = 1$?

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 500/7 \\ 200/7 \end{pmatrix} + (0,3) \begin{pmatrix} -80/7 \\ 80/7 \end{pmatrix} = \begin{pmatrix} 476/7 \\ 224/7 \end{pmatrix} = \begin{pmatrix} 68 \\ 32 \end{pmatrix}.$$

No longo prazo, $t \rightarrow \infty$, temos:

$$\begin{pmatrix} x_\infty \\ y_\infty \end{pmatrix} = \begin{pmatrix} 500/7 \\ 200/7 \end{pmatrix} \approx \begin{pmatrix} 71,4 \\ 28,6 \end{pmatrix}.$$

Este capítulo discutiu alguns pontos relacionados aos sistemas de equações de diferença de 1ª ordem. O próximo capítulo apresenta um tópico similar, que são os sistemas de equações diferenciais de 1ª ordem.

CAPÍTULO 6 - SISTEMAS DE EQUAÇÕES DE DIFERENCIAIS DE 1ª ORDEM

A dinâmica unidimensional da evolução de uma variável contínua pode ser analisada a partir de equações diferenciais, como foi discutido no quarto capítulo. Entretanto, de forma geral, muitas variáveis tendem a evoluir por meio de interações com outras variáveis, caracterizando dinâmicas que são mutuamente afetadas por um conjunto de variáveis. Para problemas assim, como discutido no capítulo anterior, quando o tempo é discreto, podem ser utilizados os sistemas de equações de diferença. Todavia, quando o tempo é contínuo, uma das ferramentas utilizadas no estudo da dinâmica de múltiplas variáveis que interagem é o sistema de equações diferenciais.

Existem muitos tipos distintos de sistemas de equações diferenciais. Muitas aplicações em economia utilizam sistemas de equação de 1ª ordem e a discussão aqui se restringe a esses sistemas. Serão apresentados os seguintes pontos sobre o tema: i) Introdução aos sistemas de equações diferenciais de 1ª ordem; ii) Resolvendo sistemas com autovalores reais distintos; iii) Resolvendo sistemas com autovalores complexos conjugados; iv) Resolvendo sistemas com autovalores repetidos; v) Estabilidade de sistemas lineares; vi) Estabilidade de sistemas não lineares; vii) Aplicação de sistema de equações diferenciais lineares de 1ª ordem; e viii) Diagrama de fase: aplicação de sistema de equações diferenciais não lineares de 1ª ordem. Assume-se que o leitor tenha conhecimento sobre conceitos básicos de álgebra linear, de integração, e de equações diferenciais de 1ª ordem, respectivamente como apresentado no primeiro, terceiro e quarto capítulos. Além disso, esse sexto capítulo apresenta similaridades com o capítulo anterior, pois os formatos de sistemas de equações de diferença de 1ª ordem e sistemas de equações diferenciais de 1ª ordem são similares, e os métodos de resolução também são semelhantes.

1. INTRODUÇÃO AOS SISTEMAS DE EQUAÇÕES DIFERENCIAIS DE 1ª ORDEM

No quarto capítulo foram apresentadas as equações diferenciais de 1ª ordem. Por exemplo, a equação separável $\frac{dy}{dt} = ky$, que representa a dinâmica de uma variável que varia com taxas constantes. Outro

exemplo foram as equações lineares com coeficientes constantes, $\frac{dy}{dt} = ay + b$.

Quando analisamos equações assim, a dinâmica da variável y depende somente do valor da própria variável e de parâmetros. Entretanto, como já discutido no capítulo anterior, quando foram apresentados os sistemas de equações de diferença, em muitos problemas econômicos a dinâmica de uma variável depende não só de seus próprios valores como também do valor de outras variáveis. Essa mesma perceptiva está presente nesse sexto capítulo.

Portanto, em problemas com o tempo contínuo, diferentemente das equações de diferença que tratam do tempo discreto, onde dinâmicas de diferentes variáveis interagem entre si, pode-se representar essas dinâmicas com um sistema de equações diferenciais. Entre os diversos tipos de sistemas, a apresentação aqui tem como foco os sistemas de equações diferenciais de 1ª ordem com coeficientes constantes. Inicialmente, discutem-se os sistemas de equações lineares com duas equações e duas

incógnitas. Esses sistemas, além de apresentarem muitas aplicações em economia, permitem uma apresentação mais breve de diversos conceitos, muitos destes semelhantes aos já apresentados na discussão anterior sobre sistema de equações de diferença.

Um sistema de equações linear de 1ª ordem é uma extensão natural da última das equações diferenciais apresentadas anteriormente e tem o seguinte formato:

$$\begin{aligned}\frac{dx}{dt} &= a_{11}x + a_{12}y + b_1 \\ \frac{dy}{dt} &= a_{21}x + a_{22}y + b_2\end{aligned}$$

Note que a dinâmica da variável x , $\frac{dx}{dt}$, depende do valor da própria variável, da variável y e dos parâmetros. O mesmo ocorre com a variável y .

De forma similar ao realizado no capítulo anterior, aqui também se analisam os sistema homogêneos, onde $b_1 = b_2 = 0$.

Um sistema de equações diferenciais lineares de 1ª ordem homogêneo com coeficientes constantes pode ser escrito fazendo uso de matrizes:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \text{ onde } \dot{x} = \frac{dx}{dt} \text{ e } \dot{y} = \frac{dy}{dt}.$$

Ou de forma mais sucinta, $\dot{X} = AX$, onde os elementos da matriz A são todos constantes.

A metodologia de resolução analítica de sistemas assim é apresentada nas próximas três seções. Faz-se uso de autovalores e autovetores, tema apresentado no primeiro capítulo. Na seção cinco discute-se a estabilidade das soluções analíticas apresentadas nessas seções. Na seção 7 apresenta-se um exemplo aplicado. Essa apresentação é, em muitos aspectos, similar a discutida para os sistemas de equações de diferença discutidos anteriormente.

Esse capítulo também apresenta os sistemas não lineares. Segue um exemplo ilustrativo que será discutido em maior profundidade na seção 8.

Imagine duas firmas e assuma que cada uma delas ocupe um determinado mercado a partir de uma variação logística das vendas. Como vimos no quarto capítulo, a equação logística pode ser escrita como:

$$\frac{dy}{dt} = (a - by)y, \text{ onde } a \text{ e } b \text{ são constantes positivas.}$$

Assim, sejam x e y as respectivas vendas das firmas 1 e 2, as equações de crescimento das vendas são respectivamente representadas por:

$$\frac{dx}{dt} = (a_1 - b_1 x)x$$

$$\frac{dy}{dt} = (a_2 - b_2 y)y, \text{ onde } a_s \text{ e } b_s \text{ são constantes positivas.}$$

Note que as equações diferenciais nesse formato ainda são desacopladas, ou seja, a dinâmica de vendas de cada uma das firmas é determinada unicamente pelas vendas da própria empresa.

Agora assumamos que existe uma competição entre as firmas pelo mesmo mercado, isto é, as vendas de uma firma influenciam as vendas da outra de forma negativa. Inclui-se, assim, um termo negativo que dependa das vendas de ambas as firmas. Assim, obtemos:

$$\frac{dx}{dt} = (a_1 - b_1 x - c_1 y)x$$

$$\frac{dy}{dt} = (a_2 - b_2 x - c_2 y)y, \text{ onde } a_s, b_s \text{ e } c_s \text{ são constantes positivas.}$$

Desta forma, obtém-se um sistema de equações diferenciais não lineares de 1ª ordem. Esses problemas podem apresentar uma resolução analítica muito sofisticada, o que não é realizado nesse capítulo. Entretanto, a estabilidade desses sistemas é discutida na seção 7 e utilizam-se diagramas de fase para que uma ideia da dinâmica do problema seja obtida, como discutido na seção 9.

2. RESOLVENDO SISTEMAS COM AUTOVALORES REAIS DISTINTOS

No capítulo anterior os sistemas de equações de diferença lineares de 1ª ordem foram resolvidos fazendo uso de autovalores e autovetores. Os sistemas de equações diferenciais lineares de 1ª ordem são resolvidos de forma similar. Essa seção descreve como resolver esses sistemas também tendo como foco sistemas com duas equações e duas incógnitas, em uma discussão semelhante a já apresentada para equações de diferença.

Parte-se de um sistema genérico, $\dot{X} = AX$, como o representado a seguir:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

O objetivo é transformá-lo em um sistema desacoplado, cuja matriz dos coeficientes é diagonal, $\dot{Z} = DZ$. Como vimos, essa matriz diagonal é formada pelos autovalores da matriz A :

$$\begin{pmatrix} \dot{z} \\ \dot{w} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix}.$$

Esse sistema é, na verdade, composto de 2 equações diferenciais independentes de 1ª ordem separáveis:

$$\dot{z} = \lambda_1 z$$

$$\dot{w} = \lambda_2 w$$

Como vimos no quarto capítulo, essas equações diferenciais de 1ª ordem separáveis são facilmente resolvidas pelo método de separação de variáveis. As soluções são respectivamente:

$$z = Be^{\lambda_1 t}$$

$$w = Ce^{\lambda_2 t}, \text{ onde } B \text{ e } C \text{ são constantes.}$$

Assim como foi feito para o sistema de equações de diferença, assume-se que existe uma matriz P , cuja inversa P^{-1} existe, tal que $X = PZ$. Novamente essa matriz P será formada pelos autovetores da matriz A .

Daí, temos:

$$X = PZ = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} Be^{\lambda_1 t} \\ Ce^{\lambda_2 t} \end{bmatrix}$$

Por fim:

$$X(t) = Be^{\lambda_1 t} V_1 + Ce^{\lambda_2 t} V_2.$$

Note que essa solução é válida quando as matrizes P e P^{-1} existirem, o que ocorrerá sempre que os autovalores forem reais e distintos.

A discussão prossegue com um exemplo numérico, utilizando a mesma matriz apresentada no 1º capítulo e também utilizada no capítulo anterior.

Seja um sistema representado por:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Como vimos, os autovalores e autovetores da matriz $A = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix}$ são:

$$\lambda_1 = 2 \text{ e } \lambda_2 = 3.$$

$$V_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \text{ e } V_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

De posse dos autovalores e autovetores, temos como solução do problema:

$$X(t) = Be^{2t} \begin{pmatrix} 1 \\ -2 \end{pmatrix} + Ce^{3t} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

3. RESOLVENDO SISTEMAS COM AUTOVALORES COMPLEXOS CONJUGADOS

Como já discutido para as equações de diferença, todo o procedimento descrito acima para dois autovalores reais distintos também é válido para autovalores complexos conjugados. Ou seja, a solução descrita na seção anterior também é válida para autovalores complexos. Entretanto, reescreve-se a solução para torná-la mais facilmente analisável.

Partimos da solução obtida anteriormente:

$$X(t) = Be^{\lambda_1 t} V_1 + Ce^{\lambda_2 t} V_2.$$

Substituem-se na expressão os autovalores e autovetores complexos conjugados:

$$X(t) = Be^{(\alpha + \beta i)t} (u + iv) + Ce^{(\alpha - \beta i)t} (u - iv).$$

Seguindo procedimento análogo ao realizado para equações de diferença, fazendo as substituições $B = A_1 + iA_2$ e $C = A_1 - iA_2$, e colocando $e^{\alpha t}$ em evidência, temos:

$$X(t) = e^{\alpha t} 2 \operatorname{Re} \left[(A_1 + iA_2) e^{i\beta t} (u + iv) \right] \text{ onde Re indica a parte real do número complexo.}$$

Utiliza-se a relação de Euler, $e^{i\beta t} = \cos(\beta t) + i \operatorname{sen}(\beta t)$, para simplificar a expressão:

$$X(t) = e^{\alpha t} 2 \operatorname{Re} [(A_1 + iA_2)(\cos(\beta t) + i \operatorname{sen}(\beta t))(u + iv)],$$

Por fim, temos o formato final para a resolução de sistemas lineares com autovalores complexos:

$$X(t) = e^{\alpha t} [(B_1 \cos(\beta t) - B_2 \operatorname{sen}(\beta t))u - (B_1 \operatorname{sen}(\beta t) + B_2 \cos(\beta t))v].$$

Como ilustração, segue um exemplo numérico, também com a mesma matriz utilizada no capítulo anterior.

Partimos do seguinte sistema:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Como vimos, os autovalores e autovetores da matriz $A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$ são:

$$\lambda_1 = 1 + 2i \text{ e } \lambda_2 = 1 - 2i$$

$$V_1 = \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + i \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ e } V_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - i \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Substituindo esses valores na expressão descrita acima, temos:

$$X(t) = e^t \left[(B_1 \cos(2t) - B_2 \sin(2t)) \begin{pmatrix} 1 \\ 0 \end{pmatrix} u - (B_1 \sin(2t) + B_2 \cos(2t)) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right].$$

4. RESOLVENDO SISTEMAS COM AUTOVALORES REPETIDOS

Toda a discussão anterior é válida para problemas onde as matrizes P , P^{-1} e D existem. Como vimos na discussão com os sistemas de equações de diferença, em problemas com autovalores repetidos em

matrizes não diagonais isso não ocorre, e utiliza-se uma matriz quase diagonal (QD): $QD = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.

Também como visto, a matriz $P = (V_1 \ V_2)$ é formada por V_1 , o autovetor da matriz dos coeficientes, e V_2 , o autovetor generalizado, obtido a partir da relação $(A - I\lambda)V_2 = V_1$.

Inicialmente, determinamos a solução do sistema quase desacoplado, $\dot{Z} = (QD)Z$:

$$\begin{pmatrix} \dot{z} \\ \dot{w} \end{pmatrix} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix}.$$

Assim como foi realizado na seção 2 deste capítulo, para w obtém-se uma equação de 1ª ordem separável, que tem como solução:

$$w = Be^{\lambda t}.$$

Para z , temos a seguinte equação diferencial:

$$\dot{z} = \lambda z + w = \lambda z + Be^{\lambda t}.$$

Reescrevemos essa equação no formato linear de 1ª ordem, como:

$$\dot{z} - \lambda z = Be^{\lambda t}.$$

O método de resolução de equações diferenciais lineares de 1ª ordem foi apresentado no quarto capítulo. Comparando essa expressão com o formato padrão desse tipo de equação,

$\frac{dz}{dt} + f(t)z = R(t)$, observa-se que $f(t) = -\lambda$ e que $R(t) = Be^{\lambda t}$. Utilizando as expressões associadas ao método de resolução desse tipo de equação, temos:

$$h = \int f(t)dt = \int -\lambda dt = -\lambda t$$

$$z(t) = e^{-h} \left[\int e^h R(t)dt + C \right] = e^{\lambda t} \left[\int e^{-\lambda t} B e^{\lambda t} dt + C \right] = e^{\lambda t} [Bt + C]$$

Uma vez obtida essa solução do sistema quase desacoplado, obtém-se a solução do sistema acoplado:

$$X = PZ = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} e^{\lambda t} (Bt + C) \\ B e^{\lambda t} \end{pmatrix} = (Bt + C)e^{\lambda t} V_1 + B e^{\lambda t} V_2.$$

Segue um exemplo ilustrativo.

Como vimos, a matriz $A = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix}$ tem como autovalores, $\lambda_1 = \lambda_2 = 3$, como autovetor, $V_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, e

como autovetor generalizado, $V_2 = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$.

Portanto, o sistema $\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ tem como solução:

$$X = (Bt + C)e^{3t} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + B e^{3t} \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}.$$

Assim como no capítulo anterior, em todos os exemplos numéricos ilustrados, não se definiu os valores das constantes das soluções. Esse procedimento será realizado na seção 7, quando será mostrada uma aplicação de sistema de equações diferenciais lineares.

5. ESTABILIDADE DE SISTEMAS LINEARES

Assim como foi feito anteriormente na discussão sobre sistemas de equações de diferença, a estabilidade dos estados estacionários de sistemas de equações diferenciais lineares de 1ª ordem pode ser analisada a partir dos autovalores das soluções descritas nas três seções anteriores.

Em um estado estacionário, as derivadas com relação ao tempo de todas as variáveis do sistema são nulas. Em sistemas com duas variáveis, como os descritos neste capítulo, temos que no estado estacionário:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

O estado estacionário, x^* , pode ser estável ou instável. Caso seja estável, este pode ser:

i) **assintoticamente estável**, ou seja, se o sistema estiver próximo de x^* , ele convergirá para esse ponto;

- ii) **globalmente estável**, isto é, não importa onde esteja o sistema, ele convergirá para x^* ; ou
- iii) o ponto pode ter **estabilidade neutra**, quando o sistema estiver próximo de x^* , ele permanecerá próximo desse ponto.

A partir dessas definições segue uma discussão para sistemas lineares com duas variáveis, inicialmente para sistemas resolvidos com dois autovalores reais, depois com dois autovalores complexos e por fim com uma raiz dupla real. Assume-se que as constantes são diferentes de zero.

Para **autovalores reais e distintos**, vimos que a solução do sistema é dada por:

$$X(t) = Be^{\lambda_1 t} V_1 + Ce^{\lambda_2 t} V_2. \text{ Então, temos:}$$

- i) Se $\lambda_1 < 0, \lambda_2 < 0$, temos: $\lim_{t \rightarrow \infty} X(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$;
- ii) Se $\lambda_1 < 0, \lambda_2 = 0$, temos: $\lim_{t \rightarrow \infty} X(t) = CV_2$;
- iii) Se $\lambda_1 > 0$ e/ou $\lambda_2 > 0$, o sistema é instável e divergirá, pois $\lim_{t \rightarrow \infty} X(t)$ não é finito.

Nos dois primeiros casos, o sistema converge e o estado estacionário é globalmente estável.

Para **autovalores complexos conjugados**, sabe-se que a solução é dada por:

$$X(t) = e^{\alpha t} [(B_1 \cos(\beta t) - B_2 \sin(\beta t))u - (B_1 \sin(\beta t) + B_2 \cos(\beta t))v].$$

O que determina a estabilidade do equilíbrio é o valor da parte real do autovalor α , pois o restante da solução é necessariamente limitado. Então, temos:

- i) Se $\alpha < 0$, $\lim_{t \rightarrow \infty} X(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, e o estado estacionário é globalmente estável;
- ii) Se $\alpha = 0$, temos uma solução somente com senos e cossenos, a solução oscila em torno de um ponto de equilíbrio, e a estabilidade é neutra;
- iii) Se $\alpha > 0$, o termo $e^{\alpha t} \rightarrow \infty$ quando $t \rightarrow \infty$, e o sistema é instável.

Para **autovalores repetidos**, a solução do sistema é $X(t) = (Bt + C)e^{\lambda t} V_1 + Ae^{\lambda t} V_2$.

O sistema é globalmente estável para $\lambda < 0$, e instável para $\lambda > 0$. Se $\lambda = 0$, o equilíbrio é instável, por causa do termo BtV_1 .

Em vez de autovalores como discutido aqui, em muitas aplicações em Economia usa-se o traço e o determinante da matriz dos coeficientes para analisar a estabilidade de estados estacionários. Esses dois conceitos foram descritos no primeiro capítulo.

Segue uma discussão sobre estabilidade de estados estacionários utilizando o traço e o determinante.

Seja um sistema cuja matriz dos coeficientes é representada por $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Portanto, a equação

característica para a obtenção dos autovalores é:

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0.$$

Assim, temos como autovalores: $\lambda = \frac{(a + d) \pm ((a + d)^2 - 4(ad - bc))^{1/2}}{2}.$

Como vimos acima, os autovalores podem ser reais e distintos, complexos conjugados, ou reais duplos. Segue uma discussão para cada um deles a partir dos valores de $\Delta = (a + d)^2 - 4(ad - bc).$

Se $\Delta = 0$, temos uma raiz dupla real, $\lambda = \frac{(a + d)}{2}$. Como vimos acima, para que o equilíbrio seja globalmente estável, necessita-se que $\lambda < 0$. Isso implica em $a + d = \text{tr}(A) < 0$.

Se $\Delta < 0$, as raízes são complexas conjugadas e para estabilidade assintótica global, a parte real dos autovalores deve ser negativa. Ou seja, como no caso anterior, $a + d = \text{tr}(A) < 0$.

Se $\Delta > 0$, os autovalores são reais e distintos e se ambos forem negativos, o equilíbrio será assintoticamente estável. Se $\text{tr}(A) < 0$, então, necessariamente:

$$\lambda_1 = \frac{(a + d) - ((a + d)^2 - 4(ad - bc))^{1/2}}{2} < 0.$$

O outro autovalor, $\lambda_2 = \frac{(a + d) + ((a + d)^2 - 4(ad - bc))^{1/2}}{2} < 0$, será negativo se:

$$(a + d) + ((a + d)^2 - 4(ad - bc))^{1/2} < 0.$$

Isso somente irá ocorrer se $\text{tr}(A) < 0$ e, além disso:

$$(a + d) < -((a + d)^2 - 4(ad - bc))^{1/2}.$$

Isso implica em:

$$(a + d)^2 > (a + d)^2 - 4(ad - bc)$$

$$0 > -4(ad - bc).$$

Por fim, $(ad - bc) = \det(A) > 0$.

Assim, se a matriz A tiver traço negativo e determinante positivo, o equilíbrio é estável, uma vez que isso ocorrerá para as três possibilidades de resposta.

6. ESTABILIDADE DE SISTEMAS NÃO LINEARES

Toda a discussão anterior sobre estabilidade foi feita para sistemas lineares. Na primeira seção desse capítulo foi apresentado um sistema não linear, que será discutido em maiores detalhes na oitava seção desse capítulo:

$$\begin{aligned}\frac{dx}{dt} &= (a_1 - b_1x - c_1y)x \\ \frac{dy}{dt} &= (a_2 - b_2x - c_2y)y \text{ onde } a_s, b_s \text{ e } c_s \text{ são constantes positivas.}\end{aligned}$$

Segue uma discussão sobre a estabilidade de sistemas não lineares, que se baseia na linearização do mesmo. Vejamos.

Um sistema não linear como esse pode ser escrito sinteticamente como:

$$\dot{Y} = F(Y)$$

Por definição, no estado estacionário, Y^* , temos: $\dot{Y} = F(Y^*) = 0$.

Defina $H(t) = Y(t) - Y^*$, com $H(t)$ pequeno, ou seja, $Y(t)$ é próximo do estado estacionário.

Seja $Y(t) = Y^* + H(t)$ solução de $\dot{Y} = F(Y)$. Então o sistema de equações, $\dot{Y} = F(Y)$, próximo ao estado estacionário, pode ser escrito como:

$$\frac{d}{dt}(Y^* + H(t)) = F(Y^* + H(t)).$$

Obtemos a expansão de Taylor de $F(Y^* + H(t))$ em torno do estado estacionário, Y^* :

$$F(Y^* + H(t)) = F(Y^*) + DF(Y^*)H(t) + R(H(t)), \text{ onde } R \text{ é de restante do polinômio.}$$

Por definição, $F(Y^*) = 0$. Além disso, como $H(t)$ é pequeno, $R(H(t))$ tende a zero, e ficamos com:

$$F(Y^* + H(t)) \approx DF(Y^*)H(t).$$

Daí resulta que:

$$\frac{d}{dt}(Y^* + H(t)) = DF(Y^*)H(t).$$

Note que $\frac{d}{dt}(Y^* + H(t)) = \frac{d}{dt}H(t)$, e então ficamos com:

$$\frac{d}{dt}H(t) = DF(Y^*)H(t).$$

Note que esse sistema é exatamente igual ao sistema linear $\dot{X} = AX$. Ou seja, é a linearização do sistema não linear quando esse estiver próximo de um estado estacionário. Assim, em torno desse ponto podemos tratá-lo como um sistema linear, como discutido anteriormente, mas com a matriz Jacobiana, $DF(Y^*)$, no lugar da matriz original. Assim como o equilíbrio do sistema linear é discutido a partir dos autovalores da matriz A , o mesmo é feito para o sistema não linear com a matriz Jacobiana.

Ou seja, se os autovalores da matriz Jacobiana $DF(Y^*)$ forem negativos ou tiverem uma parte real negativa, então Y^* é assintoticamente estável. Se um dos autovalores da matriz Jacobiana for positivo ou se a parte real do autovalor for positiva, o equilíbrio é instável. Ou, alternativamente, se $\text{tr}(DF(Y^*)) < 0$ e $\det(DF(Y^*)) > 0$, Y^* é estável.

Um exemplo numérico é mostrado na seção 8.

7. APLICAÇÃO COM SISTEMA DE EQUAÇÕES DIFERENCIAIS LINEARES DE 1ª ORDEM

Essa seção apresenta uma aplicação com um sistema linear em Economia. Estuda-se a interação entre inflação e desemprego. Inicialmente são definidas quatro equações, com as quais será montado um sistema de equações diferenciais de 1ª ordem.

Assuma que a relação entre a taxa de variação salarial e taxa de desemprego é linear e negativamente inclinada:

$s = a - bD$, onde s é taxa de variação salarial, D é taxa de desemprego, e a e b são constantes positivas.

A inflação é uma função de s , do aumento de produtividade, T , e da taxa de inflação esperada, π :

$$p = a - bD - T + g\pi, \text{ onde } g \in (0,1].$$

A inflação esperada é ajustada a partir da seguinte relação adaptativa:

$$\dot{\pi} = j(p - \pi), \text{ onde } j \in (0,1].$$

A taxa de desemprego varia como função da taxa de variação nominal de moeda disponível, μ , e da inflação, p , a partir da seguinte relação:

$$\dot{D} = k(p - \mu), \text{ onde } k > 0.$$

O objetivo é escrever um sistema de equações diferenciais lineares a partir dessas três últimas equações. Substituindo p da segunda das equações nas duas últimas e reescrevendo, temos:

$$\dot{\pi} = j(a - bD - T + g\pi - \pi) = j(g - 1)\pi - jbD + j(a - T)$$

$$\dot{D} = k(a - bD - T + g\pi - \mu) = kg\pi - kbD + k(a - T - \mu)$$

Esse sistema é não homogêneo, assunto não abordado aqui. Apenas de forma ilustrativa, para torná-lo homogêneo, assuma que $a = T$ e que $\mu = 0$:

$$\dot{\pi} = j(a - bD - T + g\pi - \pi) = j(g - 1)\pi - jbD$$

$$\dot{D} = k(a - bD - T + g\pi - \mu) = kg\pi - kbD$$

Em formato de matriz:

$$\begin{pmatrix} \dot{\pi} \\ \dot{D} \end{pmatrix} = \begin{pmatrix} j(g-1) & -jb \\ kg & -kb \end{pmatrix} \begin{pmatrix} \pi \\ D \end{pmatrix}.$$

Para ilustrar o problema numericamente defina as constantes como: $j = 3/16$, $g = 1$, $b = 1$ e $k = 1$:

$$\begin{pmatrix} \dot{\pi} \\ \dot{D} \end{pmatrix} = \begin{pmatrix} 0 & -3/16 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \pi \\ D \end{pmatrix}.$$

Os autovalores e autovetores da matriz $A = \begin{pmatrix} 0 & -3/16 \\ 1 & -1 \end{pmatrix}$ são:

$$\lambda_1 = 3/4 \text{ e } \lambda_2 = 1/4$$

$$X_1 = \begin{pmatrix} 1 \\ -4 \end{pmatrix} \text{ e } X_2 = \begin{pmatrix} 1 \\ -4/3 \end{pmatrix}$$

A solução é:

$$\begin{pmatrix} \pi \\ D \end{pmatrix} = B e^{3t/4} \begin{pmatrix} 1 \\ -4 \end{pmatrix} + C e^{t/4} \begin{pmatrix} 1 \\ -4/3 \end{pmatrix}.$$

8. DIAGRAMA DE FASE: APLICAÇÃO DE SISTEMA DE EQUAÇÕES DIFERENCIAIS NÃO LINEARES DE 1ª ORDEM

Vimos anteriormente como obter a solução analítica de sistemas lineares. Entretanto, em muitos casos, principalmente quando a solução analítica é de difícil obtenção, podemos analisar sistemas sem que seja necessária a solução analítica do problema. No caso específico dos sistemas não lineares, a obtenção de uma solução analítica pode ser um processo bastante complicado, e o uso dos diagramas de fase pode ser particularmente conveniente.

Segue um exemplo completo de obtenção de um diagrama de fase para um problema não linear com equações que representam a competição entre duas firmas, como já descrito no capítulo. Como vimos, as equações são:

$$\frac{dx}{dt} = (a_1 - b_1x - c_1y)x$$

$$\frac{dy}{dt} = (a_2 - b_2x - c_2y)y, \text{ onde } x \text{ e } y \text{ são, respectivamente, as vendas da firma 1 e da firma 2.}$$

Note que se c_1 e c_2 forem negativos, a presença de uma firma é benéfica à outra, ou seja, há sinergia entre elas.

Para facilitar o entendimento, usam-se como constantes os seguintes valores: $a_1 = 4$, $b_1 = 1$, $c_1 = 1$, $a_2 = 6$, $b_2 = 3$ e $c_2 = 1$. Com a introdução dessas constantes obtemos as seguintes equações designadas respectivamente de $F(x, y)$ e $G(x, y)$:

$$F(x, y) = \frac{dx}{dt} = (4 - x - y)x,$$

$$G(x, y) = \frac{dy}{dt} = (6 - 3x - y)y.$$

Toda a discussão será feita em cinco passos.

Passo 1 – Obtenção dos estados estacionários

Como vimos, nos estados estacionários, temos:

$$\frac{dx}{dt} = \frac{dy}{dt} = 0.$$

Então, ficamos com:

$$(4 - x - y)x = 0$$

$$(6 - 3x - y)y = 0$$

Esse sistema de duas equações fornece quatro pontos como estados estacionários:

$$(0, 0), (0, 6), (4, 0) \text{ e } (1, 3).$$

No primeiro deles, as firmas tem venda zero, ou seja, ainda não existem ou já existiram e não existem mais. No segundo ponto, a firma 2 domina todo o mercado e vende 6 unidades. No terceiro, a firma 1 domina todo o mercado e vende 4. No último, as firmas coexistem, sendo que a primeira vende uma unidade e a segunda 3.

Passo 2 – Análise de estabilidade dos estados estacionários

Ilustra-se nesse passo a discussão da seção 7. Uma vez obtidos os estados estacionários do sistema não-linear, também como vimos, a estabilidade desses é analisada via matriz Jacobiana:

$$DF(x,y) = \begin{pmatrix} F_x & F_y \\ G_x & G_y \end{pmatrix} = \begin{pmatrix} 4-2x-y & -x \\ -3y & 6-2y-3x \end{pmatrix}$$

Como vimos, assume-se que nas proximidades dos estados estacionários, o sistema pode ser linearizado. Analisam-se os autovalores da matriz Jacobiana para cada um dos estados estacionários em separado.

Ponto (0, 0):

$$DF(0,0) = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}, \text{ com autovalores } \lambda_1 = 4 > 0 \text{ e } \lambda_2 = 6 > 0.$$

Como pelo menos um dos autovalores é positivo, o equilíbrio é instável. Note também, de forma alternativa, que o traço é positivo.

Ou seja, seguindo a discussão sobre estabilidade de estados estacionários de sistema não lineares, qualquer perturbação $H(t)$ em torno desse equilíbrio, com a entrada de uma ou mesmo de duas firmas pequenas simultaneamente no mercado promove um afastamento do ponto. Isto é, as vendas tenderão a crescer a partir dessa pequena perturbação. Seguindo uma função logística, uma vez que a competição é insignificante porque as firmas são pequenas.

Ponto (0, 6):

$$DF(0,6) = \begin{pmatrix} -2 & 0 \\ -18 & -6 \end{pmatrix}, \text{ com autovalores } \lambda_1 = -2 < 0 \text{ e } \lambda_2 = -6 < 0.$$

Ambos os autovalores são negativos, ou, alternativamente, o traço é negativo e o determinante é positivo. Isso implica que o equilíbrio é estável. Ou seja, qualquer perturbação em torno desse equilíbrio, com a entrada de uma firma pequena em um mercado dominado pela segunda firma, promove uma competição entre elas, com vantagem comparativa para a segunda, e a retomada de todo o mercado pela firma dois.

Ponto (4, 0):

$$DF(4,0) = \begin{pmatrix} -4 & -4 \\ 0 & -6 \end{pmatrix}, \text{ com autovalores } \lambda_1 = -4 < 0 \text{ e } \lambda_2 = -6 < 0.$$

O equilíbrio é estável e a situação é análoga à anterior, mas com domínio de mercado pela firma 1.

Ponto (1, 3):

$$DF(1,3) = \begin{pmatrix} -1 & -1 \\ -9 & -3 \end{pmatrix}, \text{ com autovalores } \lambda_1 < 0 \text{ e } \lambda_2 > 0 \text{ ou, alternativamente, com traço e o}$$

determinante negativos. O equilíbrio é instável. Quase qualquer perturbação no equilíbrio de coexistência das firmas leva a uma incipiente vantagem comparativa de uma delas, o que leva no

longo prazo ao domínio de todo o mercado pela mesma. Como será demonstrado a seguir, existe uma linha divisória nas regiões de vantagem comparativa, de um lado a área de domínio da firma 1 e do outro o domínio da firma 2. Exatamente nessa linha divisória não existe vantagem para qualquer das firmas. Assim, uma perturbação exatamente nesta linha implica em um retorno ao ponto de equilíbrio. Ou seja, na verdade, o equilíbrio é metaestável, e pode ser representado por um ponto de sela, assunto do próximo capítulo.

Passo 3 – Obtenção das isóclinas

No passo 1 obtemos os estados estacionários, onde $\frac{dx}{dt} = 0$ e $\frac{dy}{dt} = 0$. Aqui analisamos essas derivadas em separado. Queremos determinar os conjuntos de pontos onde $\frac{dx}{dt} = 0$, independente do valor da outra derivada e vice-versa. Note que, se $\frac{dx}{dt} = 0$ e $\frac{dy}{dt} \neq 0$, só o y varia e, assim, obtêm-se as regiões onde as isóclinas são verticais. Similarmente, se $\frac{dy}{dt} = 0$ e $\frac{dx}{dt} \neq 0$, só x varia e as isóclinas são horizontais.

Para isóclina vertical temos:

$$\frac{dx}{dt} = (4 - x - y)x = 0, \text{ o que implica em } x = 0 \text{ ou } y = 4 - x.$$

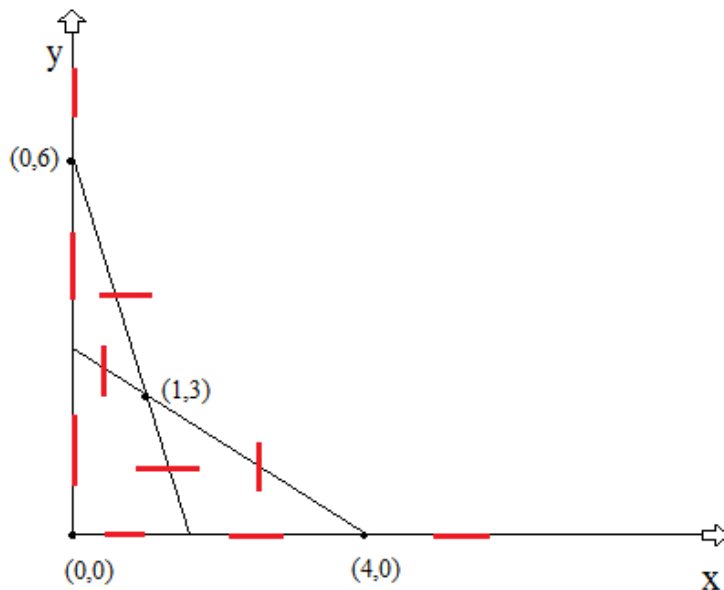
Para isóclina horizontal:

$$\frac{dy}{dt} = (6 - 3x - y)y = 0, \text{ o que implica em } y = 0 \text{ ou } y = 6 - 3x.$$

Passo 4 – Obtenção do diagrama preliminar e análise dos sinais das derivadas nas diferentes regiões.

Toda a informação obtida nos passos anteriores é transferida para um diagrama, como mostrado no diagrama 1.

DIAGRAMA 1
Primeiros passos de um diagrama de fase



As isóclinas separam quatro regiões distintas. Devem-se analisar os sinais de $\frac{dx}{dt}$ e $\frac{dy}{dt}$ para cada uma delas. Para tanto, escolhem-se pontos arbitrários em cada uma das quatro áreas. Por exemplo:

para a região A, ponto $(\frac{1}{2}, \frac{1}{2})$, $\frac{dx}{dt} > 0$ e $\frac{dy}{dt} > 0$;

para a região B, $(\frac{1}{10}, 5)$, $\frac{dx}{dt} < 0$ e $\frac{dy}{dt} > 0$;

para a região C, $(3, \frac{1}{10})$, $\frac{dx}{dt} > 0$ e $\frac{dy}{dt} < 0$; e

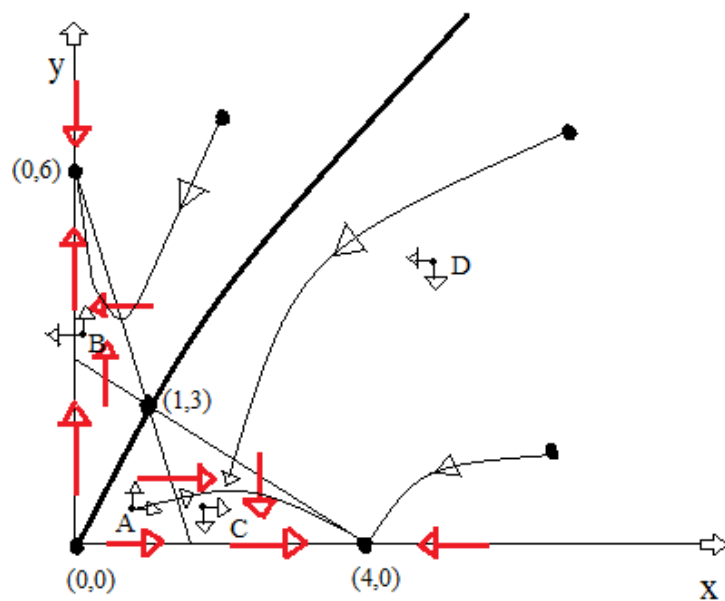
para a região D, $(10, 10)$, $\frac{dx}{dt} < 0$ e $\frac{dy}{dt} < 0$.

Essa informação também é transferida para o diagrama, como mostra o diagrama 2.

Passo 5 – Esquematização de curvas características

Como último passo, desenham-se curvas características no diagrama. Note que o diagrama é dividido em duas regiões. Uma área é de domínio da firma 1, onde essa firma tem vantagem comparativa sobre a outra. Na outra área, o domínio é da firma 2. A linha divisória entre essas áreas passa pelos equilíbrios instável $(0, 0)$ e metaestável $(1, 3)$.

DIAGRAMA 2
Diagrama de fase



CAPÍTULO 7 - ELEMENTOS DE ANÁLISE

Esse sétimo capítulo tem como objetivo apresentar brevemente alguns pontos introdutórios de Análise Matemática. Recomendam-se para discussões mais abrangentes as seguintes referências: Lima (1995 e 2004) e também Simon e Blume (1994). Dentre os vários conceitos desse tema foram selecionados alguns deles que são importantes para o entendimento do cálculo de várias variáveis no formato discutido nos capítulos subsequentes, ou que são diretamente relacionados à teoria microeconômica moderna. Mais especificamente, serão discutidos os seguintes conceitos: i) Sequências de números reais; ii) Sequências de Cauchy; iii) Conjuntos abertos; iv) Conjuntos fechados; v) Conjuntos compactos; vi) Funções contínuas e teorema de Weierstrass. Os conceitos serão discutidos basicamente para a reta, pois a discussão fica mais simples e direta. Em dimensões superiores, o raciocínio é, em geral, análogo. Somente para conjuntos abertos é que será apresentada uma breve discussão no \mathbb{R}^n , pois esse conceito é utilizado em definições do cálculo de várias variáveis. Assume-se que o leitor tem familiaridade com conceitos introdutórios de geometria analítica e do cálculo diferencial de uma e de várias variáveis.

1. Sequência de números reais

Seguem duas definições que serão utilizadas na discussão sobre sequências de números reais.

Supremo e ínfimo

Um conjunto $X \subset \mathbb{R}$ diz-se limitado superiormente quando para todo $x \in X$ existe $b' \in \mathbb{R}$ tal que $x \leq b'$. Diz-se que b' é cota superior de X . A menor das cotas superiores é chamada de **supremo de X** ou **supX**.

Seguindo um raciocínio análogo, um conjunto $X \subset \mathbb{R}$ é limitado inferiormente, se existe $a' \in \mathbb{R}$ tal que $a' \leq x$. Diz-se que a' é cota inferior de X . A maior dentre essas cotas inferiores é o **ínfimo de X** ou **infX**.

Além desses dois conceitos, outro que será utilizado na própria definição de sequência é o conceito de **números naturais**. Aqui utilizamos o conjunto dos números naturais sem o termo zero: $N^* = \{1, 2, 3, \dots\}$.

Sequências de números reais

Uma sequência é uma função $x: N^* \rightarrow \mathbb{R}$ que associa a cada número natural N^* , um número real x_n , chamado n-ésimo termo da sequência. Pode-se representar uma sequência como $(x_1, x_2, \dots, x_n, \dots)$, $(x_n)_{n \in N}$ ou, simplesmente, (x_n) . Seguem dois exemplos de sequência:

- a) $(1/n)_{n \in N} = (1, 1/2, 1/3, \dots)$
- b) $((-1)^n)_{n \in N} = (-1, 1, -1, \dots)$

Seja N' um subconjunto infinito de N^* . Por exemplo, os pares ou os ímpares. Dada uma sequência $x = (x_n)_{n \in N}$, uma subsequência de x é uma restrição da função x a um subconjunto infinito N' dos números naturais e escreve-se a subsequência como $(x_n)_{n \in N'}$.

Dadas as duas sequências anteriores, obtemos, por exemplo, as subsequências com $N' = \text{ímpares}$ e com $N' = \text{pares}$:

$$\begin{aligned} \text{a)} \quad & (1/n)_{n \in N'} = (1, 1/3, 1/5, \dots) \\ & (1/n)_{n \in N''} = (1/2, 1/4, 1/6, \dots) \\ \text{b)} \quad & ((-1)^n)_{n \in N'} = (-1, -1, \dots) \\ & ((-1)^n)_{n \in N''} = (1, 1, \dots) \end{aligned}$$

Note que na sequência descrita no exemplo a), os termos se aproximam de zero à medida que $n \in N^*$ aumenta. Isto é, $\lim x_n = 0$. Observe, ainda, que, para o exemplo b), os termos oscilam entre -1 e 1, ou seja, não tendem para um valor específico.

Diz-se que a primeira dessas sequências tem limite, enquanto que isso não ocorre para a segunda. Segue uma definição formal de limite de sequência e outras definições relacionadas.

Limite de sequência - Diz-se que o número real a é o limite da sequência (x_n) quando para todo número real $\varepsilon > 0$, dado arbitrariamente, pode-se obter $n_0 \in N$ tal que todos os termos x_n com índice $n > n_0$, cumprem a condição $|x_n - a| < \varepsilon$.

Diz-se ainda que uma sequência que tem limite é uma **sequência convergente** e converge para o valor do limite da sequência.

Sequência limitada - Uma sequência (x_n) diz-se limitada superiormente quando existe $b \in \mathbb{R}$ tal que $x_n \leq b$ para todo $n \in N$. Uma sequência (x_n) diz-se limitada inferiormente quando existe $a \in \mathbb{R}$ tal que $a \leq x_n$ para todo $n \in N$. Uma sequência é limitada se é limitada superiormente e inferiormente.

Note que as duas sequências descritas nos exemplos a) e b) são limitadas. Todos os termos da sequência $(1/n)_{n \in N}$ assumem valores entre 0 e 1. Para a sequência $((-1)^n)_{n \in N}$, os valores estão limitados entre -1 e 1.

Teorema 1 - Toda sequência convergente é limitada.

Demonstração. Dado que a sequência é convergente, ela converge para $a = \lim (x_n)$. Tomando $\varepsilon > 0$, vemos, por definição de limite de sequência, que existe $n_0 \in N$ tal que $n > n_0$ implica $x_n \in (a - \varepsilon, a + \varepsilon)$. Sejam α o menor e β o maior elemento do seguinte conjunto finito $\{x_1, x_2, \dots, x_{n_0}, a - \varepsilon, a + \varepsilon\}$. Como todos os termos de (x_n) estão contidos no intervalo $[\alpha, \beta]$, (x_n) é limitada.

Note que a recíproca não é verdadeira. Uma sequência limitada, como a do exemplo b) acima, não necessariamente é convergente.

Uma sequência (x_n) é uma **sequência monótona** quando, para todo $n \in N$, tem-se $x_n \leq x_{n+1}$, ou seja, a sequência é não decrescente, ou quando, também para todo $n \in N$, tem-se $x_n \geq x_{n+1}$, isto é, a sequência é não crescente.

Teorema 2 - Toda sequência monótona limitada é convergente.

Demonstração. Seja (x_n) monótona, digamos não decrescente, ou seja, $x_n \leq x_{n+1}$, e limitada. Escreve-se o conjunto $X = \{x_1, x_2, \dots, x_n, \dots\}$ e $a = \sup X$. Afirmamos que $a = \lim(x_n)$. Com efeito, dado $\varepsilon > 0$, o número $a - \varepsilon$ não é cota superior de X , pois a menor das cotas superiores é justamente $\sup X = a$. Logo, existe $n_0 \in \mathbb{N}$ tal que $a - \varepsilon < x_{n_0} < a$. Assim, como a sequência é monótona não decrescente, para $n > n_0$ temos $a - \varepsilon < x_{n_0} \leq x_n < a < a + \varepsilon$ e, assim, por definição de limite de sequência, a sequência é convergente, pois $\lim(x_n) = a$. Note que um raciocínio análogo pode ser feito para sequências não crescentes.

Teorema de Bolzano-Weierstrass – Toda sequência limitada de números reais possui uma subsequência convergente.

Demonstração. Baseado no que foi demonstrado no teorema 2, basta mostrar que toda sequência limitada (x_n) possui uma subsequência monótona, pois esta é convergente. A demonstração será feita em duas etapas, cada qual com um exemplo.

Digamos que um termo x_n da sequência é destacado quando $x_n \geq x_p$ para todo $p > n$. Seja $D \subset \mathbb{N}$ o conjunto dos índices n tais que x_n é um termo destacado. Se D for um conjunto infinito, $D = \{n_1 < n_2 < \dots < n_k < \dots\}$ então a subsequência $(x_n)_{n \in D}$ será monótona não crescente.

Segue um exemplo. Tome a sequência limitada $X = \{1, \frac{2}{1}, 1, \frac{3}{2}, 1, \frac{4}{3}, \dots\}$.

Para $n = 1$, $x_1 = 1$. Assim, para $p > 1$, temos para todo p , $x_1 \geq x_p$? Não, isso não ocorre, e o termo não é destacado.

Prosseguindo para $n = 2$, temos $x_2 = \frac{2}{1}$. Para $p > 2$, temos para todo p , $x_2 \geq x_p$? Isso ocorre e o termo é destacado.

Seguindo esse raciocínio, observamos que os termos ímpares não seriam destacados e os de número par seriam. Assim, seria obtida a subsequência $(x_n)_{n \in D}$ monótona não crescente, na verdade

decrescente, $Y = \{\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \dots\}$.

Entretanto, se D for finito, seja $n_1 \in \mathbb{N}$ maior do que todos os $n \in D$. Então x_{n_1} não é destacado. Logo, existe algum $n_{n_2} > n_{n_1}$ com $x_{n_2} > x_{n_1}$. Por sua vez, x_{n_2} também não é destacado, então existe algum $n_3 > n_2$ com $x_{n_2} < x_{n_3}$. Seguindo esse raciocínio infinitamente temos uma subsequência monótona crescente.

Segue um exemplo. Tome a sequência limitada $X = \{1, \frac{1}{2}, 1, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots\}$.

Para $n = 1$, $x_1 = 1$. Assim, para $p > 1$, temos para todo p , $x_1 \geq x_p$? Isso ocorre e o termo é destacado.

Prosseguindo para $n = 2$, temos $x_2 = \frac{1}{2}$. Para $p > 2$, temos para todo p , $x_2 \geq x_p$? Isso não ocorre e o termo não é destacado.

Seguindo esse mesmo raciocínio para os demais termos seria obtido um conjunto D finito com somente o primeiro e terceiro termos: $D = \{1, 1\}$.

Tomando o quarto termo da sequência $x_4 = \frac{2}{3}$, que não é destacado. Logo, existe algum $n > n_4$ com $x_n > x_4$, no caso $x_5 = \frac{3}{4}$. Seguindo esse raciocínio infinitamente temos uma subsequência monótona crescente: $Y = \{\frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots\}$.

2. Sequências de Cauchy

Até aqui sempre que discutimos a convergência de sequências utilizamos diretamente o conceito de limite da mesma. A convergência de sequências também pode ser relacionada com o conceito de sequência de Cauchy.

Sequência de Cauchy - Diz-se que (x_n) é uma sequência de Cauchy quando, para todo $\varepsilon > 0$ dado arbitrariamente, existe $n_0 \in \mathbb{N}$ tal que $m, n > n_0 \Rightarrow |x_m - x_n| < \varepsilon$.

Teorema 3 - Qualquer sequência convergente é de Cauchy

Demonstração. Se a sequência (x_n) é convergente com $\lim(x_n) = a$, então, por definição, para todo $\varepsilon > 0$ dado arbitrariamente, existe $n_0 \in \mathbb{N}$ tal que $m > n_0 \Rightarrow |x_m - a| < \varepsilon/2$, e também $n > n_0 \Rightarrow |x_n - a| < \varepsilon/2$. Somando esses dois termos, temos: $m, n > n_0 \Rightarrow |x_n - a| + |x_m - a| < \varepsilon$.

Usando a desigualdade triangular, $|A + B| \leq |A| + |B|$, temos que $\varepsilon > |x_n - a| + |x_m - a| = |x_n - a| + |a - x_m| \geq |(x_n - a) - (a - x_m)| = |x_n - x_m|$.

Assim, ficamos com: para todo $\varepsilon > 0$ dado arbitrariamente, existe $n_0 \in \mathbb{N}$ tal que $m, n > n_0 \Rightarrow |x_n - x_m| < \varepsilon$, que é a definição de sequência de Cauchy.

Teorema 4 - Qualquer sequência de Cauchy é convergente

A recíproca da demonstração do teorema 3 será feita em três etapas via Lemas 1, 2 e 3.

Lema 1 - Toda sequência de Cauchy é limitada.

Demonstração. Seja (x_n) uma sequência de Cauchy, então $m, n \geq n_0 \Rightarrow |x_n - x_m| < \varepsilon$.

Especificamente para $m = n_0$ isto também é verdadeiro. Assim, temos: $n > n_0 \Rightarrow |x_n - x_{n_0}| < \varepsilon$. Tome

$\varepsilon = 1$, então, temos: $n > n_0 \Rightarrow x_n \in (x_{n_0} - 1, x_{n_0} + 1)$. Seja b o menor e c o maior dentre os elementos do conjunto finito $\{x_1, x_2, \dots, x_{n_0-1}, x_{n_0} - 1, x_{n_0} + 1\}$. Então $x_n \in [b, c]$ para todo $n \in \mathbb{N}$, logo (x_n) é limitada.

Lema 2 – Toda sequência de Cauchy tem uma subsequência convergente.

Demonstração. Por **Bolzano-Weierstrass**, toda sequência limitada possui uma subsequência convergente, logo, o mesmo é válido para uma sequência de Cauchy, que, pelo Lema 1, é limitada.

Lema 3 – Se a sequência de Cauchy tem uma subsequência convergente com limite a , toda a sequência também converge para esse limite.

Demonstração. Dado que a sequência é de Cauchy, $m, n \geq n_0 \Rightarrow |x_m - x_n| < \varepsilon/2$. Tome x_k da subsequência convergente com $k \geq n_0$ tal que $|x_k - a| < \varepsilon/2$. Assim, utilizando a desigualdade triangular, para todo $m \geq n_0$ temos: $|x_m - a| = |(x_m - x_k) + (x_k - a)| \leq |x_m - x_k| + |x_k - a| < \varepsilon$, que, uma vez que não foi feita qualquer restrição para m , é a definição de sequência convergente.

As próximas três seções apresentam algumas noções topológicas relacionadas com conjuntos abertos, conjuntos fechados e conjuntos compactos.

3. Conjuntos abertos

Para definirmos o que é um conjunto aberto, inicialmente necessitamos da definição do que é um ponto interior.

Diz-se que o ponto a é **ponto interior** ao conjunto $X \subset \mathbb{R}$ quando existe um número $\varepsilon > 0$ tal que o intervalo aberto $(a - \varepsilon, a + \varepsilon)$ está contido em X .

Por exemplo, tome o conjunto $X = [0, 1] \subset \mathbb{R}$. O ponto $x = 1/2$ é interior a X , pois existe um número $\varepsilon > 0$ tal que o intervalo aberto $(1/2 - \varepsilon, 1/2 + \varepsilon)$ está contido em X . Por outro lado, o ponto $x = 1$ não é interior a X , pois não existe um número $\varepsilon > 0$ tal que o intervalo aberto $(1 - \varepsilon, 1 + \varepsilon)$ esteja contido em X .

Cabe aqui discutir esse mesmo conceito de ponto interior também para o \mathbb{R}^n , pois esse será utilizado posteriormente nos próximos capítulos, sobretudo em definições do cálculo de várias variáveis.

Um ponto a é ponto interior ao conjunto $X \subset \mathbb{R}^n$ quando existe um número $\varepsilon > 0$ tal que a bola aberta em torno de a , denominada bola-Epston ou bola- ε , $B_\varepsilon(a) \equiv \{x \in \mathbb{R}^n : \|x - a\| < \varepsilon\}$, está contido em X .

Segue um exemplo no \mathbb{R}^2 . Tome o conjunto $A = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. O ponto $(x, y) = (1/2, 1/2)$ é interior a A , pois existe um número $\varepsilon > 0$ tal que a bola aberta em torno de $(1/2, 1/2)$, no caso um círculo com centro $(1/2, 1/2)$ e raio ε , $B_\varepsilon(a) \equiv \{x \in \mathbb{R}^2 : \|(x, y) - (1/2, 1/2)\| < \varepsilon\}$ está contido em

A. Em contrapartida, o ponto $(x, y) = (1, 1)$ não é interior a A , pois não existe um número $\varepsilon > 0$ tal que a bola aberta em torno de $(1, 1)$, $B_\varepsilon(a) \equiv \{x \in \mathfrak{R}^2 : \|(x, y) - (1, 1)\| < \varepsilon\}$ esteja contida em A .

Um conjunto formado unicamente por pontos interiores é um **conjunto aberto**.

4. Conjuntos fechados

A discussão prossegue na reta real. Partimos da definição de ponto aderente, que é essencial na definição de conjunto fechado.

Ponto aderente - Diz-se que um ponto \mathbf{a} é aderente ao conjunto $X \subset \mathfrak{R}$ quando \mathbf{a} é limite de alguma sequência de pontos $x_n \in X$.

Seguem três exemplos de ponto aderente para $X_1 = (0, 1)$ e também para $X_2 = [0, 1]$.

Tome a sequência $(x_n) = \left(\frac{1}{n+1}\right)$. Note que todos os termos da sequência pertencem aos conjuntos definidos acima: $x_n \in X_1$ e $x_n \in X_2$. Além disso, note que a sequência é convergente com $\lim(x_n) = \mathbf{a} = 0$. Assim, por definição, esse ponto é ponto aderente de X_1 e X_2 .

Por sua vez, tome a sequência convergente $(x_n) = \left(\frac{n}{n+1}\right)$, com $x_n \in X_1$ e $x_n \in X_2$, e $\lim(x_n) = \mathbf{a} = 1$. Esse também é ponto aderente de X_1 e X_2 .

Por fim, tome a sequência $(x_n) = \left(\frac{n}{2n+1}\right)$, que tem como limite $\mathbf{a} = \frac{1}{2}$, sendo esse ponto aderente de X_1 e X_2 .

Conjunto fechado - Um conjunto diz-se fechado quando TODO ponto aderente a X pertence a X .

Por exemplo, o conjunto $X_1 = (0, 1)$ não é fechado, pois não contém todos os pontos aderentes. $\mathbf{a} = 0$ e $\mathbf{a} = 1$ são aderentes a este conjunto, mas não pertencem a ele. Por sua vez, $X_2 = [0, 1]$ é fechado, pois contém todos os seus pontos aderentes.

Teorema 5 - Um ponto \mathbf{a} é aderente ao conjunto X se, e somente se, toda vizinhança de \mathbf{a} contém algum ponto de X .

Demonstração. Seja \mathbf{a} ponto aderente a X . Como \mathbf{a} é ponto aderente de X , esse ponto é limite de alguma sequência convergente, onde $x_n \in X$ para todo $n \in \mathbb{N}$, $\mathbf{a} = \lim(x_n)$. Dada uma vizinhança V de \mathbf{a} , $V = (a - \varepsilon, a + \varepsilon)$, temos pela própria definição de sequência convergente que para todo n suficientemente grande $x_n \in V$. Como $x_n \in X$ e também $x_n \in V$, $V \cap X \neq \emptyset$. Reciprocamente, se toda vizinhança de \mathbf{a} contém algum ponto de X , podemos escolher em cada intervalo $(a - 1/n, a + 1/n)$, $n \in$

N , um ponto $x_n \in X$. Então, tomando $1/n = \varepsilon$, temos $|x_n - a| < 1/n = \varepsilon$, logo a sequência (x_n) é convergente e $\lim(x_n) = a$ e, consequentemente, a é aderente a X .

Teorema 6 – Um conjunto $F \subset \mathfrak{R}$ é fechado se, e somente se, seu complemento $A = \mathfrak{R} - F$ é aberto.

Demonstração. Seja F fechado, sendo $a \in A$, isto é, $a \notin F$. Pelo teorema 5, como $a \notin F$, existe uma vizinhança V de a que não contém pontos de F , isto é, $a \in V \subset A$. Assim, todo ponto de A é ponto interior a A e A é um conjunto aberto. Reciprocamente, se o conjunto A é aberto e o ponto a é aderente a $F = \mathfrak{R} - A$, então toda vizinhança de a contém pontos de F , logo a não é interior a A . Sendo A aberto, temos $a \notin A$, ou seja, $a \in F$. Assim, todo ponto aderente a F pertence a F , e F é fechado.

5. Conjuntos compactos

Um conjunto $X \subset \mathfrak{R}$ chama-se compacto quando é limitado e fechado.

Note que esse tipo de conjunto é muito importante para problemas de otimização. Por exemplo, qual é o máximo da função $f(x) = x$ definida, respectivamente, nos intervalos:

- a) não limitado $(-\infty, \infty)$?
- b) aberto $(0, 1)$?
- c) compacto $[0, 1]$?

Note que o máximo da função nestes intervalos é bem definido somente para o terceiro dentre eles, que é justamente o conjunto compacto.

Teorema 7 – Um conjunto $X \subset \mathfrak{R}$ é compacto se, e somente se, toda sequência de pontos em X possui uma subsequência que converge para algum ponto de X .

Demonstração.

Se X é compacto, ou seja, limitado, logo, por Bolzano-Weierstrass, toda sequência cujos pontos pertençam a X será necessariamente limitada, portanto, possuirá uma subsequência convergente. Como X também é fechado, ele contém todos seus pontos aderentes, e o limite da subsequência será um ponto de X . Reciprocamente, assumamos que $X \subset \mathfrak{R}$ não é limitado. Neste caso, para cada $n \in \mathbb{N}$ poderíamos encontrar $x_n \in X$ com $|x_n| > n$. A sequência (x_n) assim obtida não possuiria qualquer subsequência limitada, logo não teria subsequência convergente. Assim, se toda sequência de pontos $x_n \in X$ possui uma subsequência convergente, o conjunto não pode ser ilimitado, ou seja, é limitado. Como toda subsequência convergente tem como limite um ponto em X , X contém todos os seus pontos aderentes e é fechado. Logo X é compacto.

Seja X um conjunto compacto não vazio, dado que o conjunto é limitado, existem $\sup X$ e $\inf X$. Dado que o conjunto é fechado, ele contém todos seus pontos aderentes. Tome uma sequência monótona crescente, o limite dessa sequência é $b = \sup X \in X$. Utilizando raciocínio análogo para uma sequência

monótona decrescente, o limite é $a = \inf X \in X$. Ou seja, um conjunto compacto contém seu supremo e seu ínfimo.

6. Funções contínuas e teorema de Weierstrass

Problemas de otimização são centrais em muitos dos modelos econômicos. Assim, um ponto fundamental da análise econômica é determinar quais são os pontos $x \in X$ nos quais certa função real $f: X \rightarrow \mathbb{R}$ assume seu valor máximo ou seu valor mínimo. Para tanto, como primeiro requisito, esses pontos devem existir. Mas quais são as condições que garantem a existência de um ponto ótimo? O teorema de Weierstrass trata exatamente desse assunto. Antes, porém, de discutirmos esse teorema, apresentamos a definição de função contínua.

Função contínua - Uma função $f: X \rightarrow \mathbb{R}$, definida no conjunto $X \subset \mathbb{R}$, diz-se contínua no ponto $a \in X$ se, sempre que a sequência (x_n) com $x_n \in X$ convergir para a , implicar que a sequência $(f(x_n))$ também convergirá para $f(a)$. Diz-se que $f: X \rightarrow \mathbb{R}$ é uma função contínua quando f é contínua em todos os pontos.

Como vimos, segundo o teorema de Bolzano-Weierstrass, qualquer sequência limitada, por exemplo, contida em um conjunto compacto, tem uma subsequência convergente.

Teorema 8 – A imagem $f(X)$ de um conjunto compacto $X \subset \mathbb{R}$ por uma função contínua $f: X \rightarrow \mathbb{R}$ é um conjunto compacto.

Demonstração. Dado que X é compacto, sabemos que uma sequência (x_n) qualquer tem uma subsequência $(x_n)_{n \in \mathbb{N}'}$ convergente com limite $a \in X$. Tomando a definição de função contínua, sendo f contínua no ponto a , a sequência $(y_n)_{n \in \mathbb{N}'} = (f(x_n))_{n \in \mathbb{N}'}$ converge para $f(a)$. Além disso, colocando $b = f(a)$, temos que $b \in f(X)$. Pelo teorema 7, sabemos que um conjunto Y é compacto se, e somente se, toda sequência tem uma subsequência que converge para algum ponto de Y . Assim, a imagem $f(X)$ tem essa característica e é um conjunto compacto.

Teorema de Weierstrass – Seja $f: X \rightarrow \mathbb{R}$ contínua no conjunto compacto $X \subset \mathbb{R}$. Existem $x_0, x_1 \in X$ tais que $f(x_0) \leq f(x) \leq f(x_1)$ para todos $x \in X$.

Demonstração. Dado que $f(X)$ é compacto, o conjunto possui seu ínfimo, ou seja, seu elemento de menor valor ou mínimo global, $f(x_0)$, e seu supremo, isto é, seu elemento de maior valor, $f(x_1)$, seu máximo global.

O teorema de Weierstrass mostra que para que uma função tenha necessariamente ponto ótimo, ela deve ser contínua e definida em um domínio compacto. Os próximos três capítulos discutem pontos associados com os problemas de otimização.

CAPÍTULO 8 - OTIMIZAÇÃO NÃO CONDICIONADA

Os problemas de otimização, tanto os não condicionados como os condicionados, são centrais em teoria econômica. São muitos os exemplos de aplicação da otimização em Economia, tais como maximização de utilidade, maximização de lucro, minimização de custos, etc. Esse capítulo discute alguns conceitos teóricos e aplicados da otimização não condicionada. Os pontos a serem abordados seguem a seguinte ordem: i) Otimização não condicionada no cálculo de uma variável; ii) Otimização não condicionada no cálculo de várias variáveis: condições de primeira ordem; iii) A matriz Hessiana e o estudo de concavidade de funções; iv) Definidade de matrizes; v) Exemplos de otimização não condicionada no \mathfrak{R}^n ; vi) Aplicação da otimização não condicionada no \mathfrak{R}^n ; e vii) Conceitos mais teóricos. Problemas de otimização condicionada são apresentados no próximo capítulo. Assume-se que o leitor tenha conhecimentos relativamente sólidos em álgebra matricial e em cálculo de várias variáveis, pontos já discutidos em capítulos anteriores.

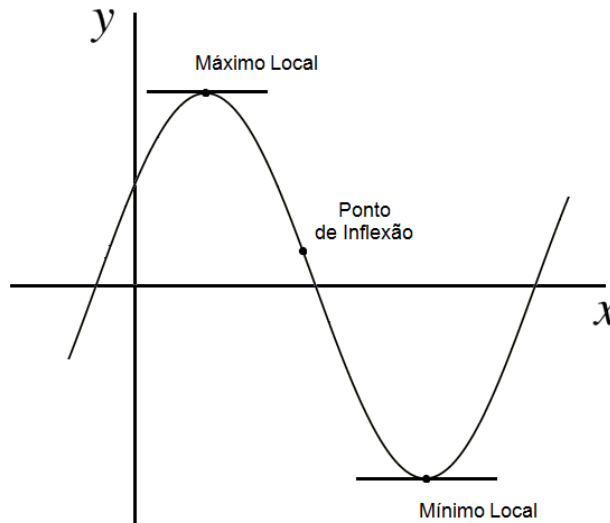
1. OTIMIZAÇÃO NÃO CONDICIONADA NO CÁLCULO DE UMA VARIÁVEL

Este capítulo tem como objetivo principal discutir a otimização não condicionada utilizando o cálculo de várias variáveis. Entretanto, como forma de iniciar essa discussão, esta seção apresenta conceitos associados a esse tópico do cálculo de uma variável, incluindo uma aplicação em Economia. O objetivo é retomar alguns conceitos discutidos em problemas de otimização não condicionada no cálculo de uma variável, que depois serão úteis nas apresentações posteriores.

Inicia-se a apresentação com as definições de máximo local, de mínimo local e de ponto de inflexão, temas normalmente abordados em cursos introdutórios de graduação. O diagrama 1 mostra exemplos desses três pontos em uma função de terceiro grau.

Um **ponto de máximo local** é aquele em que a função assume o maior valor dentre todos os pontos na vizinhança. O contrário ocorre com o **ponto de mínimo local**, em que a função assume o menor valor dentre todos os pontos na vizinhança. No **ponto de inflexão** ocorre uma troca de concavidade, em que esta passa de convexa para côncava ou vice-versa.

DIAGRAMA 1
Pontos de máximo, de mínimo e de inflexão no \mathbb{R}^2



Assumindo que as derivadas de 1ª e 2ª ordem existem, observa-se que a reta tangente, tanto no máximo como no mínimo, é horizontal, o que implica em $f'(x) = 0$, ou seja, os pontos são críticos. Para se determinar se um ponto crítico é de máximo ou de mínimo, faz-se uso da derivada segunda no estudo da concavidade da função. Note que no ponto de máximo local, a função é côncava, ou seja, $f''(x) \leq 0$. No ponto de mínimo local, a função é convexa, o que implica em $f''(x) \geq 0$. Além de máximos e mínimos, os pontos críticos podem ser pontos de inflexão. Entretanto, esses últimos não necessariamente são críticos, como mostra o exemplo do diagrama 1.

A discussão sobre máximos e mínimos locais é resumida a seguir.

Se $f'(x_0) = 0$, e:

- a) $f''(x_0) < 0$, o ponto x_0 é de máximo local
- b) $f''(x_0) > 0$, o ponto x_0 é de mínimo local
- c) $f''(x_0) = 0$, o ponto x_0 pode ser de máximo local, de mínimo local ou nenhum dos dois.

Nesse último caso, temos um ponto de inflexão, em que ocorre a troca de concavidade na função.

Como o objetivo aqui é discutir a otimização não condicionada no cálculo de várias variáveis, encerra-se aqui a apresentação de conceitos do cálculo de uma variável. A discussão tem continuidade com uma aplicação direta desses conceitos em um exemplo de maximização de lucro de uma firma.

O diagrama 2 mostra as funções utilizadas nesse exemplo. $F(x)$, que representa o faturamento da firma, é definido como $F(x) = px$, em que p é o preço de venda e x é a quantidade vendida. Note que, como $p > 0$, a função é crescente. A função custo de produção, $C(x)$, é crescente e convexa, em

que x representa a quantidade produzida. O lucro é definido por: $\pi(x) = F(x) - C(x)$. O ponto de máximo é um ponto crítico dessa função: $\pi'(x) = F'(x) - C'(x) = 0$.

Como ilustração, assuma que $F(x) = x$ e que $C(x) = 0,2x^3$, como representado no diagrama. A função lucro é, portanto, $\pi(x) = x - 0,2x^3$.

Derivando e igualando a zero, temos:

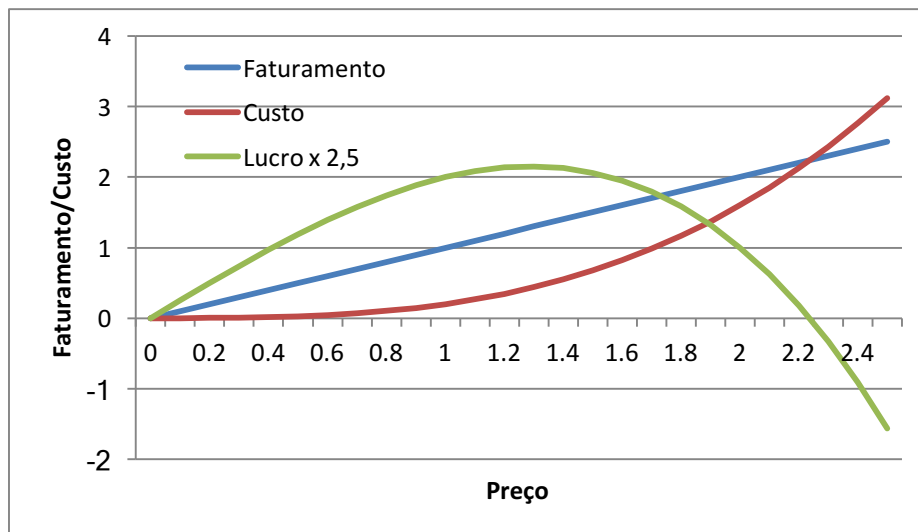
$$\pi'(x) = 1 - 0,6x^2 = 0$$

$$x = (5/3)^{1/2} \approx 1,29$$

Como derivada de segunda ordem, temos:

$$\pi''(x) = -1,2x < 0, \text{ pois } x > 0, \text{ o que implica que o ponto crítico é de máximo.}$$

DIAGRAMA 2
Maximização de lucro



2. OTIMIZAÇÃO NÃO CONDICIONADA NO CÁLCULO DE VÁRIAS VARIÁVEIS: CONDIÇÕES DE PRIMEIRA ORDEM

Como discutido anteriormente para o cálculo de uma variável, os pontos críticos podem ser pontos de máximo local, de mínimo local ou de inflexão. Para se determinar qual é o tipo de ponto crítico, vimos que devemos fazer o estudo da concavidade da função utilizando a derivada segunda.

Para a otimização não condicionada no cálculo de várias variáveis, seguimos um raciocínio análogo. Inicialmente, determinam-se os pontos críticos fazendo uso das derivadas parciais de 1ª ordem ou, em outros termos, das condições de primeira ordem. Esse é o ponto abordado nesta seção. Em seguida,

estuda-se a concavidade da função em cada um dos pontos críticos, ou seja, analisam-se as condições de segunda ordem, como será discutido em seções posteriores.

Em funções de uma variável, existem três possibilidades de pontos críticos, como mostrado no diagrama 1, e isso também ocorre em funções de várias variáveis. Os diagramas 3, 4 e 5 mostram essas possibilidades, respectivamente, um ponto de máximo, um ponto de mínimo e um ponto de sela. Em funções de uma variável, temos $f'(x) = 0$ nos pontos críticos. No \mathfrak{R}^n , assumindo que as derivadas parciais existam, elas são todas nulas nos pontos críticos.

DIAGRAMA 3
Ponto de máximo no \mathbb{R}^3

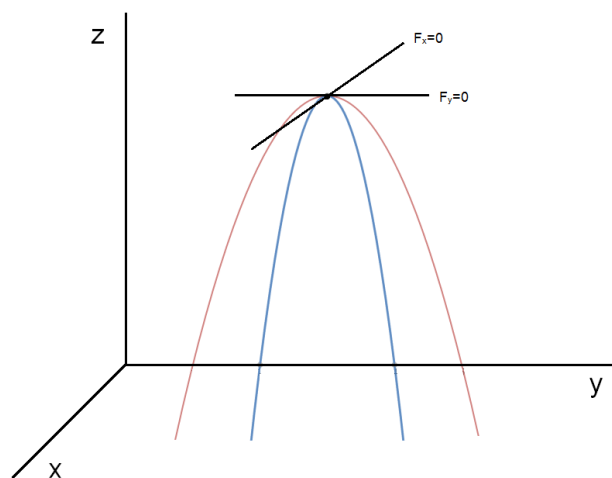


DIAGRAMA 4
Ponto de mínimo no \mathbb{R}^3

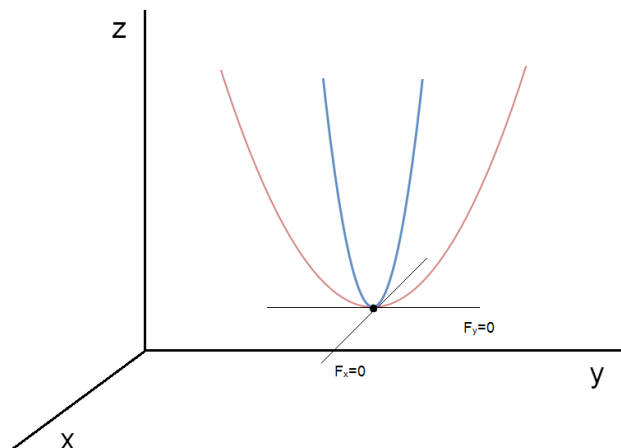
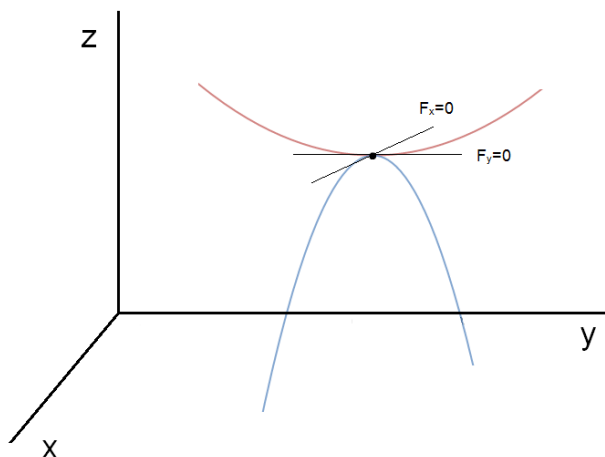


DIAGRAMA 5
Ponto de sela no \mathbb{R}^3



Segue uma definição um pouco mais formal sobre o ponto crítico em funções de várias variáveis.

Ponto crítico – Seja $F : U \rightarrow \mathbb{R}$ uma função definida no subconjunto U no \mathbb{R}^n , cuja derivada de 1ª ordem é contínua. Se x^* é máximo local ou mínimo local de F em U e se x^* é um ponto interior de U , então $\frac{\partial F}{\partial x_i}(x^*) = 0$ para $i = 1, \dots, n$

A seção 6 apresenta exemplos que ilustram esses conceitos e a seção 7 apresenta uma aplicação.

3. A MATRIZ HESSIANA E O ESTUDO DE CONCAVIDADE DE FUNÇÕES

Como vimos, em problemas de otimização com o cálculo de uma variável, normalmente determinamos se um ponto é de máximo ou de mínimo fazendo uso da derivada de segunda ordem. No \mathbb{R}^n , para diferenciar se um ponto crítico é de máximo, de mínimo ou ponto de sela, necessitamos das condições de 2ª ordem. A discussão dessas condições para o cálculo de várias variáveis passa pelo conceito de matriz Hessiana.

Uma matriz Hessiana de uma função é formada pelas derivadas parciais de segunda ordem. No caso de uma função de duas variáveis, $f(x, y)$, temos a seguinte matriz:

$$He = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}.$$

No caso de funções com mais variáveis, simplesmente acrescentamos linhas e colunas seguindo a mesma lógica de formação dessa matriz anterior.

Para $f(x, y, z)$ temos:

$$He = \begin{pmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{pmatrix}.$$

Para funções cujas derivadas de segunda ordem existem e são contínuas, as matrizes Hessianas são simétricas. Ou seja, nas matrizes acima $f_{xy} = f_{yx}$, $f_{xz} = f_{zx}$ e $f_{zy} = f_{yz}$.

A matriz Hessiana é muito utilizada no estudo sobre concavidade de funções de várias variáveis. Se essa matriz for:

- a) positivamente definida, a função é estritamente convexa;
- b) positivamente semidefinida, a função é convexa;
- c) negativamente definida, a função é estritamente côncava;
- d) negativamente semidefinida, a função é côncava;
- e) indefinida, a função não é côncava, nem convexa.

Portanto, na determinação se o ponto crítico é de máximo, de mínimo ou de sela, temos as condições de 2ª:

Condições de 2ª ordem – Seja $F: U \rightarrow \mathbb{R}$ uma função cujo domínio é o conjunto aberto U no \mathbb{R}^n e cuja derivada segunda é contínua. Assume-se que x^* é um ponto crítico, então, como condição suficiente, se a matriz Hessiana for:

- a) positivamente definida, x^* é ponto de mínimo estrito;

- b) negativamente definida, x^* é ponto de máximo estrito;
- c) indefinida, x^* é ponto de sela.

Resta saber como definir se uma matriz Hessiana é positivamente definida, positivamente semidefinida, negativamente definida, negativamente semidefinida ou indefinida, como será descrito na seção seguinte. Na sétima seção, mostra-se formalmente a relação entre definidade de matriz e determinação de ponto crítico.

4. DEFINIDADE DE MATRIZES

Esta seção discute o conceito de definidade de matrizes. Esse conceito é definido a partir das funções quadráticas. Segue uma definição sobre essas funções, e exemplos ilustrativos.

Funções quadráticas – Uma função quadrática no R^n é uma função de valor real da forma, $Q(x_1, x_2, \dots, x_n) = \sum_{i \leq j} a_{ij} x_i x_j$, que pode ser escrita na forma matricial como $X^T A X$, em que A é uma matriz simétrica e X é um vetor não-nulo.

Seguem dois exemplos para funções quadráticas de duas e três variáveis.

Exemplo 1 – Dada a função $Q(x, y) = a_{11}x^2 + a_{22}y^2$. Essa função pode ser representada por $X^T A X$, em que $X = \begin{pmatrix} x \\ y \end{pmatrix}$ e $A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}$.

Exemplo 2 – Dada a função $Q(x, y, z) = a_{11}x^2 + 2a_{12}xy + 2a_{13}xz + a_{22}y^2 + 2a_{23}yz + a_{33}z^2$. Essa função pode ser representada por $X^T A X$, com $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ e $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$.

Note que as matrizes A são simétricas, assim como as matrizes Hessianas. Utiliza-se o conceito de funções quadráticas para determinar a definidade da matriz.

Uma matriz simétrica $A_{n \times n}$ pode ser classificada da seguinte maneira:

- a) se $X^T A X > 0$ para todo $X \in R^n$, $X \neq 0$, a matriz é positivamente definida;
- b) se $X^T A X \geq 0$ para todo $X \in R^n$, $X \neq 0$, a matriz é positivamente semidefinida;
- c) se $X^T A X < 0$ para todo $X \in R^n$, $X \neq 0$, a matriz é negativamente definida;
- d) se $X^T A X \leq 0$ para todo $X \in R^n$, $X \neq 0$, a matriz é negativamente semidefinida;
- e) se $X^T A X > 0$ para alguns $X \in R^n$ e $X^T A X < 0$ para outros $X \in R^n$, a matriz é indefinida.

Os exemplos numéricos a seguir exemplificam cada uma dessas situações.

A função $Q(x, y) = x^2 + y^2$, com $Q(x, y) > 0$ para todo $X \neq 0$ no \Re^n . Em formato de matriz, temos a seguinte matriz: $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Ou seja, pela definição acima, $X^T A X > 0$, essa matriz é positivamente definida.

Para a função $Q(x, y) = (x + y)^2 = x^2 + 2xy + y^2$, temos $Q(x, y) \geq 0$ para todo $X \neq 0$ no \Re^n . Em formato de matriz, temos que $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $X^T A X \geq 0$, e a matriz A é positivamente semidefinida.

De forma similar, a função $Q(x, y) = -x^2 - y^2 < 0$ para todo $X \neq 0$ no \Re^n . Em formato de matriz, temos que $A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ é negativamente definida.

A função $Q(x, y) = -(x + y)^2 \leq 0$ para todo $X \neq 0$ no \Re^n . Assim, temos que a matriz $A = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$ é negativamente semidefinida.

A função $Q(x, y) = -x^2 + y^2$ é positiva para algum $X \neq 0$ no \Re^n e negativa para outros. Ou seja, $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ é indefinida.

Compare essas cinco matrizes com as respectivas matrizes Hessianas.

Para a função $Q(x, y) = x^2 + y^2$, temos: $He = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

A função $Q(x, y) = (x + y)^2 = x^2 + 2xy + y^2$ tem como Hessiana: $A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$.

A função $Q(x, y) = -x^2 - y^2$ tem como Hessiana a seguinte matriz: $He = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$.

Por fim, para a função $Q(x, y) = -(x^2 + y^2)$, temos: $He = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}$.

Para a função $Q(x, y) = -x^2 + y^2$, temos: $He = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$.

Todas as matrizes Hessianas são múltiplas das anteriormente mostradas. Ou seja, a matriz Hessiana apresenta a mesma definidade que a matriz da função quadrática.

Uma vez apresentadas as cinco definições e matrizes, resta saber como determinar de que tipo é uma matriz simétrica, em particular uma Hessiana. Apresentam-se dois métodos, sendo que o primeiro é baseado nos autovalores da matriz A . Seguem algumas relações entre autovalores e matrizes definidas:

- a) se todos os autovalores de uma matriz forem maiores que zero, a matriz é positivamente definida;
- b) se todos os autovalores forem maiores ou iguais a zero, a matriz é positivamente semidefinida;
- c) se todos os autovalores de uma matriz forem menores que zero, a matriz é negativamente definida;
- d) se todos os autovalores forem menores ou iguais a zero, a matriz é negativamente semidefinida;
- e) se algum dos autovalores for maior que zero e outro menor que zero, a matriz é indefinida.

Nos exemplos acima, partimos da definição de funções quadráticas para determinar a definidade da matriz. Entretanto, o mais usual é obter a matriz Hessiana de uma função qualquer e determinar a definidade desta a partir de um método.

Por exemplo, os autovalores da matriz Hessiana $He = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ são $\lambda_1 = 2 > 0$ e $\lambda_2 = 2 > 0$. Ambos maiores que zero, o que implica que a matriz é positivamente definida.

Além de autovalores, podemos utilizar o conceito de submatriz líder principal para determinar o tipo de matriz. Em alguns casos é mais simples utilizar um método e em outros a utilização do outro é mais fácil. A seção 7 mostra a associação entre esses métodos.

Seja A uma matriz $n \times n$. Uma submatriz líder principal de ordem k da matriz A é obtida deletando as últimas $n - k$ linhas e as últimas $n - k$ colunas da matriz A . Ou seja, para uma matriz $A_{n \times n}$ existe a submatriz líder principal de ordem $k = 0$, que é a própria matriz, a de ordem $k = 1$, até a de ordem $k = n - 1$.

O determinante de uma matriz líder principal é denominado menor líder principal.

Seja A uma matriz simétrica $n \times n$. Então,

- a) a matriz A é positivamente definida se, e somente se, todos os menores líderes principais forem estritamente positivos;
- b) a matriz A é negativamente definida se, e somente se, todos os menores líderes principais alternarem o sinal da seguinte forma: $|A_1| < 0$, $|A_2| > 0$, $|A_3| < 0$, etc., em que A_i é a submatriz líder principal de ordem i ;
- c) se menores líderes principais diferem de zero e não podem ser classificados nas duas letras anteriores, então A é indefinida.

A matriz A também pode ser definida como positivamente semidefinida e negativamente semidefinida utilizando conceitos similares.

Como exemplo de aplicação desse método, tome, por exemplo, a matriz Hessiana $He = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$.

Ela tem como submatriz líder principal de ordem $k = 0$, a própria matriz Hessiana: $A_2 = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$.

A submatriz líder principal de ordem $k = 1$ é $A_1 = (-2)$, em que a segunda coluna e segunda linha da matriz anterior foram eliminadas. Os determinantes das mesmas, isto é, os menores líderes principais são respectivamente $|A_2| = -4 < 0$ e $|A_1| = -2 < 0$.

Ambos são menores que zero, o que implica em uma matriz indefinida.

5. EXEMPLOS DE OTIMIZAÇÃO NÃO CONDICIONADA NO \mathbb{R}^n

Esta seção apresenta dois exemplos que abordam os conceitos discutidos nas seções anteriores. O primeiro deles com uma função de duas variáveis e o outro para uma função de três variáveis.

Exemplo 1 - $f(x, y) = x^2 + 2y^2$

A seção 2 apresentou as condições de 1ª ordem da otimização não condicionada no \mathbb{R}^n . No caso específico desse problema, temos duas derivadas parciais:

$$f_x = 2x$$

$$f_y = 4y$$

Igualando ambas a zero, obtemos um único ponto crítico:

$$2x = 0$$

$$4y = 0$$

O ponto crítico é $(0, 0)$.

A seção 3 apresentou a matriz Hessiana da função. No caso desse exemplo, essa tem a seguinte forma:

$$He = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$$

Note que essa matriz é simétrica. Note também que a matriz Hessiana não muda para valores distintos de x e de y , ou seja, a matriz será definida de forma igual para todos os valores de x e y .

Como discutido na seção 4, deve-se definir qual é a definidade dessa matriz.

São pelo menos três os métodos para se discutir isso. Pela definição de função quadrática, podemos escrever uma função com $Z \in \mathbb{R}^2$, $Z = \begin{pmatrix} x \\ y \end{pmatrix}$, da seguinte forma:

$$Z^T M Z = (x \ y) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (x \ y) \begin{pmatrix} x \\ 2y \end{pmatrix} = 2x^2 + 4y^2$$

Note que $Z^T M Z > 0$ para todo $Z \in \mathbb{R}^n$, tal que $Z \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Ou seja, a matriz M é positivamente definida.

Pode-se também fazer uso dos autovalores. Como a matriz Hessiana já é diagonal, temos como autovalores $\lambda_1 = 2$ e $\lambda_2 = 4$, ambos maiores que zero, indicando que a matriz é positivamente definida.

Alternativamente, temos que a matriz líder principal de ordem $k = 0$, que é a própria matriz Hessiana, $A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, tem como determinante $|A_2| = 8 > 0$. Por sua vez, a matriz líder principal de ordem $k = 1$, $A_1 = (2)$, tem como determinante $|A_1| = 2 > 0$. Ou seja, ambos são positivos, indicando que a matriz é positivamente definida.

Concluindo: a função tem um único ponto crítico, a matriz Hessiana é a mesma para todos os valores de x e y , ela é positivamente definida, a função $f(x, y) = x^2 + 2y^2$ é estritamente convexa em todo o domínio, o ponto crítico é de mínimo local, na verdade, mínimo global, pois esse é o único ponto crítico em uma função estritamente convexa.

Exemplo 2 – $F(x, y, z) = -x^3 + 3xz + 2y - y^2 - 3z^2$.

Seguindo os mesmos passos, inicialmente, determinarmos os pontos críticos. Derivando a função parcialmente e igualando as derivadas a zero, temos:

$$F_x = -3x^2 + 3z = 0$$

$$F_y = 2 - 2y = 0$$

$$F_z = 3x - 6z = 0$$

Da segunda dessas equações, obtemos: $y = 1$.

Isolando z da primeira equação e substituindo na terceira, temos:

$$z = x^2$$

$$3x = 6z = 6x^2$$

$$x = 2x^2$$

$$x(1 - 2x) = 0$$

Daí decorre que:

$x = 0$, o que implica em $z = 0$; ou

$x = 1/2$, e $z = 1/4$.

Obtemos, portanto, dois pontos críticos: $(0, 1, 0)$ e $(1/2, 1, 1/4)$.

Uma vez obtidos esses pontos, verificamos se eles são pontos de máximo, de mínimo ou de sela com a utilização de matriz Hessiana:

$$He = \begin{pmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{pmatrix} = \begin{pmatrix} -6x & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{pmatrix}.$$

Note que, de forma distinta ao observado no exemplo anterior, a matriz Hessiana depende do valor de x . Portanto, devem-se substituir os pontos críticos na matriz para determinar a concavidade da função no ponto. Como forma de ilustração, para o primeiro dos pontos, faz-se uso dos autovalores e, para o segundo, utiliza-se o conceito de líder principal.

Para o primeiro dos pontos críticos $(0, 1, 0)$, a matriz Hessiana é a seguinte:

$$He = \begin{pmatrix} 0 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{pmatrix}.$$

Os autovalores dessa matriz são, respectivamente, $\lambda_1 = -2 < 0$, $\lambda_2 = -3 + 3\sqrt{2} > 0$ e $\lambda_3 = -3 - 3\sqrt{2} < 0$. Ou seja, temos autovalores negativos e um positivo. Portanto, a matriz é indefinida e o ponto crítico é de sela.

Para o segundo dos pontos críticos $(1/2, 1, 1/4)$, a matriz Hessiana toma o seguinte formato:

$$He = \begin{pmatrix} -3 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{pmatrix}.$$

Utilizando o conceito de matriz líder principal para determinar a definidade, temos:

$$|A_1| = |-3| = -3 < 0$$

$$|A_2| = \begin{vmatrix} -3 & 0 \\ 0 & -2 \end{vmatrix} = 6 > 0$$

$$|A_3| = \begin{vmatrix} -3 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -6 \end{vmatrix} = 18 > 0$$

Ou seja, a matriz é indefinida e o ponto crítico é também de sela.

6. APLICAÇÃO DA OTIMIZAÇÃO NÃO CONDICIONADA NO \mathfrak{R}^n

Esta seção apresenta um exemplo bastante simples de aplicação da otimização não condicionada no \mathfrak{R}^n em Economia. Assuma que uma firma produz dois produtos, cujas quantidades são definidas por x , e y , e cujos preços, p_x e p_y , são determinados exogenamente. O objetivo da firma é maximizar o lucro, assim como no exemplo mostrado na seção 1.

Como são dois os produtos, a receita é definida como $R(x, y) = p_x x + p_y y$. Para facilitar a ilustração, assumamos que $p_x = 10$ e $p_y = 18$. A função custo de produção é dada por $C(x, y) = 2x^2 + xy + 4y^2$. Portanto, como lucro, temos: $\pi(x, y) = 10x + 18y - (2x^2 + xy + 4y^2)$.

Primeiramente, determinamos os pontos críticos:

$$\pi_x = 10 - 4x - y = 0$$

$$\pi_y = 18 - x - 8y = 0$$

Reescrevendo e manipulando:

$$y = 10 - 4x$$

$$18 - x - 8(10 - 4x) = 0$$

$$31x = 62$$

$$x = 2$$

$$y = 2$$

Obtém-se, assim, um único ponto crítico (2, 2).

Utiliza-se a matriz Hessiana para verificar se ele é realmente ponto de máximo:

$$He = \begin{pmatrix} -4 & -1 \\ -1 & -8 \end{pmatrix}.$$

Daí, temos:

$$|A_2| = 33 > 0;$$

$$|A_1| = -4 < 0.$$

A matriz Hessiana é negativamente definida e o ponto crítico é de máximo.

7. CONCEITOS MAIS TEÓRICOS

Esta seção apresenta alguns conceitos mais teóricos referentes a pontos discutidos anteriormente. O objetivo é formalizar parte da apresentação das seções 3 e 4. A discussão desta seção se baseia em matrizes positivamente definidas. O raciocínio é análogo para matrizes negativamente definidas e indefinidas. Primeiro, discutem-se as condições de 2ª ordem, tema introduzido na seção 3 e, em seguida, relaciona-se autovalores e definidade de matrizes, tópico apresentado na seção 4.

7.1. Condições de 2ª ordem

Como vimos na discussão das condições de 2ª ordem, em uma função cuja derivada segunda é contínua e que tem x^* como ponto crítico, se a matriz Hessiana for positivamente definida em x^* , x^* é ponto de mínimo estrito. Segue uma demonstração de que o ponto crítico x^* , no qual a matriz Hessiana é positivamente definida, é realmente ponto de mínimo local. Uma demonstração mais formalmente rigorosa é descrita em Simon e Blume (1994).

Escreve-se a função F com uma aproximação de Taylor de 2ª ordem em torno de x^* :

$$F(x^* + h) = F(x^*) + DF(x^*)h + \frac{1}{2}h^T D^2F(x^*)h + R(h), \text{ com } R(h) \rightarrow 0 \text{ quando } h \rightarrow 0.$$

Dado que x^* é ponto crítico, então $DF(x^*) = 0$. Além disso, analisando a proximidade do ponto crítico, isto é, $h \rightarrow 0$, ficamos com:

$$F(x^* + h) - F(x^*) = \frac{1}{2}h^T D^2F(x^*)h.$$

Como F é uma função com derivada segunda contínua, $Q(h) = h^T D^2F(x^*)h$ é contínua.

Pela definição de matriz positivamente definida, $X^T D^2F(x^*)X > 0$ para todo $X \in \mathbb{R}^n$, $X \neq 0$, o que implica em $F(x^* + h) - F(x^*) > 0$, e x^* é mínimo estrito local.

7.2. Definidade de matrizes

Na seção 4, vimos que uma matriz, cujos autovalores são todos positivos, é positivamente definida. Nesta seção quer-se mostrar que $X^T AX > 0$ implica em $\lambda_i > 0$ para todo i , e vice-versa.

Como A é uma matriz simétrica, todos os autovalores são reais, e a matriz é diagonalizável. Ou seja, existe uma matriz P , para a qual P^{-1} existe, tal que $P^{-1}AP = D$. No caso particular da matriz simétrica, P é também uma matriz ortogonal, ou seja, $P^{-1} = P^T$.

Seja X um vetor não nulo e Y um vetor tal que $PY = X$. Daí, temos:

$$\begin{aligned} X^T AX &= (PY)^T A(PY) = Y^T P^T APY = Y^T (P^T AP)Y = Y^T (P^{-1}AP)Y = Y^T DY \\ &= Y^T \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} Y = \sum \lambda_i y_i^2 \end{aligned}$$

Note que, se todos os autovalores forem positivos, ou seja, $\lambda_i > 0$, necessariamente $X^T AX = \sum \lambda_i y_i^2 > 0$, e a matriz é positivamente definida.

Outra maneira de determinar a definidade de matrizes é através de submatrizes da matriz original. Segue uma demonstração para matrizes pequenas.

Assuma que a matriz seja simétrica e positivamente definida, ou seja, todos os autovalores são reais e positivos.

Para uma matriz 1 x 1: $A = (a)$, temos $|A_1| = a = \lambda_1$, como $\lambda_1 > 0$, $|A_1| > 0$.

Para uma matriz 2 x 2, $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$, temos:

$$|A_2| = ad - b^2 = \lambda_1 \lambda_2 > 0.$$

Isso implica que:

$$ad > 0 \text{ e, portanto, } a, d \neq 0.$$

Também implica em $tr(A_2) = a + d = \lambda_1 + \lambda_2 > 0$. Ou seja, pelo menos um dos elementos da diagonal principal tem que ser positivo. Entretanto, note que $ad > 0$, então ambos os elementos são positivos. Daí, também temos:

$$|A_1| = a > 0.$$

Ou seja, $|A_1| > 0$ e também $|A_2| > 0$.

CAPÍTULO 9 - OTIMIZAÇÃO CONDICIONADA

Os problemas de otimização não condicionados e condicionados são centrais em teoria econômica. A otimização não condicionada foi apresentada no capítulo anterior, que serve de base para o presente capítulo. Em problemas não condicionados, procuram-se determinar máximos e/ou mínimos de funções sem qualquer restrição quanto ao valor que as variáveis podem assumir. Ou seja, determinam-se quais são os pontos críticos da função tendo como possibilidade de escolha todo o domínio da função.

Entretanto, em muitos dos problemas de otimização em Economia, temos restrições, isto é, procuram-se determinar máximos e mínimos da função, sendo que os valores das variáveis estão sujeitos a um espaço restrito de todo o domínio da função. Ou seja, os valores das variáveis não podem assumir quaisquer valores no domínio e sim apenas aqueles que satisfazem a interseção das restrições. São muitos os exemplos de aplicações desse tipo de otimização em Economia, tais como maximização da utilidade sujeita a restrição orçamentária, etc. Esse nono capítulo discute justamente problemas de otimização condicionada.

Um problema típico de otimização condicionada tem o seguinte formato geral:

Maximize $f(x_1, \dots, x_n)$ sujeito a $h_i(x_1, \dots, x_n) = b_i$ e $g_j(x_1, \dots, x_n) \leq c_j$, onde f é a função objetiva que estamos otimizando, h_i são as restrições de igualdade e g_j , são as restrições de desigualdade.

A resolução de problemas de otimização apresenta diferentes metodologias dependendo do número de restrições e se essas são de igualdade ou de desigualdade. A discussão se inicia com a otimização com restrições de igualdade, primeiramente para funções de duas variáveis e em problemas de apenas uma restrição. Esses problemas são mais simples e servem de base para discussões posteriores. Em seguida, discutem-se os problemas com funções com mais variáveis e mais restrições de igualdade. Depois são discutidos os problemas que também incluem as restrições de desigualdade. Por fim, apresentam-se algumas discussões complementares.

Os pontos a serem abordados seguem a seguinte ordem: i) Introdução à otimização condicionada e a função de Lagrange; ii) Problemas de otimização condicionada com funções objetivas de duas variáveis e uma restrição de igualdade; iii) Otimização condicionada com mais de uma restrição de igualdade; iv) Maximização com restrições de desigualdade; v) Minimização com restrições de desigualdade; e vi) Tópicos adicionais. Assume-se que o leitor tem conhecimento prévio de álgebra linear, de cálculo de várias variáveis e de otimização não condicionada, pontos discutidos no primeiro, segundo e oitavo capítulos. Além disso, um conhecimento sobre pontos introdutórios de microeconomia também ajudam no entendimento deste capítulo.

1. INTRODUÇÃO À OTIMIZAÇÃO CONDICIONADA E A FUNÇÃO DE LAGRANGE

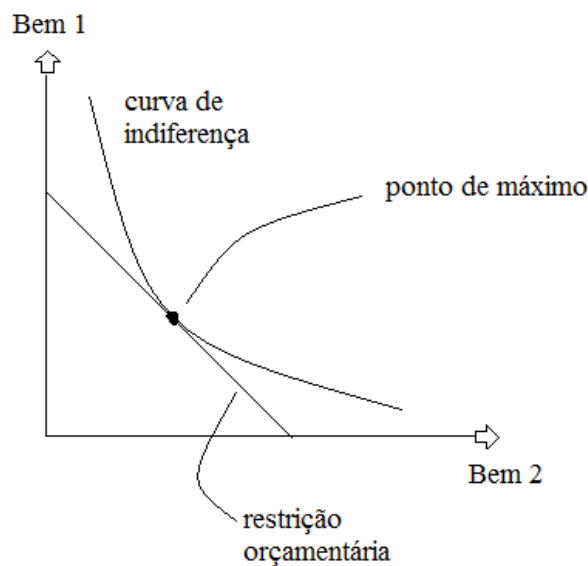
Essa seção apresenta conceitos introdutórios da otimização condicionada, em particular a função de Lagrange, que é uma poderosa ferramenta de resolução para problemas de maximização e minimização com restrições, tanto de igualdade como de desigualdade. Para tanto, inicialmente,

discutem-se os problemas de otimização mais simples, que são aqueles com funções objetivas de duas variáveis e uma restrição de igualdade. Esse tipo de problema tem o seguinte formato:

otimize $f(x, y)$ sujeito a $h(x, y) = b$.

Diversos problemas em microeconomia se utilizam desse arcabouço, como a maximização de utilidade com o consumo de dois bens, como representado no diagrama a seguir.

DIAGRAMA 1
Otimização com restrição de igualdade



Tendo como base esse problema e o diagrama 1, obtém-se a função de Lagrange, que servirá de base para a resolução dos problemas condicionados, que é a função de Lagrange ou Lagrangeano.

Note que no ponto de máximo, X^* , as inclinações da reta da restrição orçamentária, que é aqui genericamente escrita como $h(x, y) = b$, e da curva de indiferença, representada por $f(x, y) = C$, onde C é uma constante, são iguais. Ou seja, a curva tangencia a reta neste ponto.

O Lagrangeano será obtido a partir dessas inclinações. Inicialmente ambas são obtidas por derivação implícita.

Para a reta, temos:

$$\frac{dy}{dx} = - \frac{\frac{\partial h}{\partial x}(X^*)}{\frac{\partial h}{\partial y}(X^*)}.$$

Para a curva de indiferença, temos:

$$\frac{dy}{dx} = -\frac{\frac{\partial f}{\partial x}(X^*)}{\frac{\partial f}{\partial y}(X^*)}.$$

Igualando as duas inclinações, ficamos com:

$$-\frac{\frac{\partial f}{\partial x}(X^*)}{\frac{\partial f}{\partial y}(X^*)} = -\frac{\frac{\partial h}{\partial x}(X^*)}{\frac{\partial h}{\partial y}(X^*)}.$$

Essa relação é reescrita como:

$$\frac{\frac{\partial f}{\partial x}(X^*)}{\frac{\partial h}{\partial x}(X^*)} = \frac{\frac{\partial f}{\partial y}(X^*)}{\frac{\partial h}{\partial y}(X^*)} = \mu, \text{ onde } \mu \text{ é o multiplicador de Lagrange.}$$

Note que tanto as derivadas implícitas como essa última expressão só são válidas se os denominadores das razões forem diferentes de zero. Assim, para que toda a discussão seja válida, necessariamente:

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

Caso isso ocorra, podemos prosseguir na obtenção do Lagrangeano. O multiplicador de Lagrange tem significado econômico, porém esse ponto não é apresentado aqui. Para uma discussão ver Simon e Blume (1994).

Manipulando algebricamente as razões acima, obtemos:

$$\begin{aligned} \frac{\partial f}{\partial x}(X^*) - \mu \frac{\partial h}{\partial x}(X^*) &= 0 \\ \frac{\partial f}{\partial y}(X^*) - \mu \frac{\partial h}{\partial y}(X^*) &= 0 \end{aligned}$$

Note que essas equações podem ser reescritas fazendo uso do gradiente:

$$\begin{pmatrix} \frac{\partial f}{\partial x}(X^*) \\ \frac{\partial f}{\partial y}(X^*) \end{pmatrix} = \mu \begin{pmatrix} \frac{\partial h}{\partial x}(X^*) \\ \frac{\partial h}{\partial y}(X^*) \end{pmatrix} \Rightarrow \nabla f(X^*) = \mu \nabla h(X^*).$$

Note que essas equações garantem que as inclinações das funções $h(x, y) = b$ e $f(x, y) = C$ são iguais, como mostrado no diagrama. Entretanto, note que até aqui temos duas equações e três incógnitas, x , y e μ . Necessita-se de outra equação para que o sistema tenha solução única. Portanto, inclui-se uma terceira equação, que é a própria restrição, $h(x, y) - b = 0$, pois, assim, garante-se que essas inclinações são iguais e que o ponto X^* está contido na restrição.

De posse dessas três equações, define-se a função Lagrangeana:

$$L(x, y, \mu) \equiv f(x, y) - \mu(h(x, y) - b).$$

Note que as condições de primeira ordem dessa função são justamente as três equações do problema de otimização condicionada:

$$L_x = f_x(X^*) - \mu h_x(X^*) = 0$$

$$L_y = f_y(X^*) - \mu h_y(X^*) = 0$$

$$L_\mu = -(h(X^*) - b) = 0$$

Assim, obtém-se um sistema de 3 equações e três incógnitas e calculam-se os pontos críticos da função Lagrangeana. Uma vez obtidos esses pontos, deve-se determinar se eles são de máximo, de mínimo ou nenhuma dessas alternativas. Em geral, em problemas de maximização, verificamos em qual dos pontos críticos a função assume o maior valor. A próxima seção apresenta dois exemplos, sendo o primeiro genérico e o segundo aplicado. Um procedimento alternativo mais formal, porém mais trabalhoso, para determinar se um ponto crítico é de máximo, de mínimo ou nenhum dos dois é a utilização das condições de segunda ordem com o uso da matriz hessiana orlada, como discutido na sexta seção desse capítulo.

2. PROBLEMAS DE OTIMIZAÇÃO CONDICIONADA COM FUNÇÕES OBJETIVAS DE DUAS VARIÁVEIS E UMA RESTRIÇÃO DE IGUALDADE

Essa seção apresenta um exemplo genérico e outro aplicado de otimização condicionada com funções objetivas de duas variáveis e uma restrição de igualdade. As seções 3, 4 e 5 estendem essa discussão para problemas mais sofisticados.

Exemplo genérico – Maximize $f(x, y) = x + y$, sujeito a restrição $x^2 + y^2 = 1$

Note que o objetivo desse problema é obter o máximo da função $f(x, y)$, sendo que os valores das variáveis x e y são restritos à circunferência de raio 1, como determinado na restrição. Para resolver esse problema, inicialmente, escreve-se a restrição no formato $h(x, y) - b = 0$:

$$x^2 + y^2 - 1 = 0$$

Depois, monta-se o Lagrangeano do problema:

$$L(x, y, \lambda) = f(x, y) - \lambda(h(x, y) - b) = x + y - \lambda(x^2 + y^2 - 1).$$

Em seguida, obtêm-se as condições de 1ª ordem:

$$L_x = 1 - \lambda(2x) = 0$$

$$L_y = 1 - \lambda(2y) = 0$$

$$L_\lambda = -(x^2 + y^2 - 1) = 0$$

Reescrevem-se essas equações da seguinte forma:

$$1 = 2x\lambda$$

$$1 = 2y\lambda$$

$$x^2 + y^2 = 1$$

Dado que $\lambda \neq 0$, como mostram as duas primeiras equações, isola-se $\frac{1}{\lambda}$ de ambas e iguala-as:

$$\frac{1}{\lambda} = 2x = 2y \Rightarrow x = y$$

Substituindo essa relação na terceira das equações, temos:

$$2x^2 = 1$$

$$x = \pm \frac{1}{2^{1/2}} = y$$

Assim, obtemos dois pontos candidatos a máximo: $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ e $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$.

Substituindo na função objetiva, ficamos com:

$$f\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) = \sqrt{2}$$

$$f\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) = -\sqrt{2}$$

O primeiro ponto é o de máximo, pois a função assume um maior valor neste.

Exemplo aplicado – Um indivíduo quer maximizar a utilidade de suas férias, cuja função utilidade é dada por $U(x, y) = x^2 y$, onde x são dias na praia e y são dias na montanha. Este tem R\$ 1000 reais para gastar nas férias, cada dia na praia sai por 50 reais e cada dia na montanha custa 100.

Desta forma, a restrição orçamentária é escrita e simplificada como:

$$50x + 100y = 1000$$

$$x + 2y = 20$$

$$x + 2y - 20 = 0$$

Depois, monta-se a função de Lagrange e obtêm-se as condições de primeira ordem:

$$L(x, y, \mu) = x^2 y - \mu(x + 2y - 20)$$

$$L_x = 2xy - \mu = 0$$

$$L_y = x^2 - 2\mu = 0$$

$$L_\mu = -(x + 2y - 20) = 0$$

Se qualquer das variáveis for nula, a utilidade é zero. Portanto, partindo da observação que se ambas forem positivas, a utilidade também será, assume-se que no ponto de máximo x e y são positivos. Daí decorre das duas primeiras equações que:

$$\mu = 2xy = \frac{x^2}{2}$$

$$y = \frac{x}{4}$$

Substituindo na terceira das equações, temos:

$$x + 2y = 20$$

$$x + \frac{x}{2} = 20$$

$$x = 40/3 \Rightarrow y = 10/3$$

O indivíduo irá passar 40/3 dias na praia e 10/3 dias na montanha.

3. OTIMIZAÇÃO CONDICIONADA COM MAIS DE UMA RESTRIÇÃO DE IGUALDADE

A seção 1 apresentou o Lagrangeano a partir de um problema com duas variáveis e uma restrição de igualdade. Como vimos, essa função pode ser utilizada quando: $\left(\frac{\partial h}{\partial x}(X^*), \frac{\partial h}{\partial y}(X^*) \right) \neq (0,0)$.

Para o caso com n variáveis e m restrições de igualdade, com $m < n$, o procedimento é análogo. Monta-se o Lagrangeano:

$$L(x_1, \dots, x_n, \mu) \equiv f(x_1, \dots, x_n) - \mu_1(h_1(x_1, \dots, x_n) - b_1) - \dots - \mu_m(h_m(x_1, \dots, x_n) - b_m).$$

Essa função pode ser utilizada se as condições de qualificação de restrição não degenerada são satisfeitas. Como essa verificação é mais sofisticada do ponto de vista teórico, a discussão é realizada na seção 6.

Segue um exemplo, onde se assume que essa função pode ser utilizada.

Exemplo – Determine os pontos de máximo e de mínimo da função $f(x, y, z) = xyz$ sujeita às restrições $h_1 \equiv x^2 + y^2 = 1$ e $h_2 \equiv x + z = 1$.

Seguindo os mesmos passos já descritos, monta-se o Lagrangeano do problema:

$$L(x, y, z, \mu_1, \mu_2) = xyz - \mu_1(x^2 + y^2 - 1) - \mu_2(x + z - 1).$$

De posse deste, obtêm-se cinco equações com as condições de 1ª ordem que serão utilizadas para definir os valores das cinco incógnitas:

$$L_x = yz - 2x\mu_1 - \mu_2 = 0$$

$$L_y = xz - 2y\mu_1 = 0$$

$$L_z = xy - \mu_2 = 0$$

$$L_{\mu_1} = -(x^2 + y^2 - 1) = 0$$

$$L_{\mu_2} = -(x + z - 1) = 0$$

Depois de obtido esse sistema com cinco equações e incógnitas, basta resolvê-lo. Como os procedimentos são simples, porém tediosos, eles ficam a cargo do leitor.

4. MAXIMIZAÇÃO COM RESTRIÇÕES DE DESIGUALDADE

As seções anteriores apresentaram problemas de otimização com restrições de igualdade. Entretanto, em muitos dos problemas de otimização em Economia as restrições de desigualdade aparecem naturalmente. Por exemplo, em problemas de maximização de utilidade, o consumo de qualquer bem é necessariamente maior ou igual a zero. No exemplo descrito na seção 2, dado o formato da função utilidade, os consumos eram todos maiores que zero e a inclusão das restrições de desigualdade não foi necessária. Entretanto, como veremos em exemplos, dependendo do formato da função utilidade, os pontos de máximo podem ser soluções de canto com valores nulos para alguma das variáveis.

Para resolver problemas com restrição de desigualdade usamos uma técnica um pouco mais sofisticada do que a utilizada para problemas apenas com restrições de igualdade. Inicialmente, discutimos o método de resolução para um problema de maximização de uma função de duas variáveis sujeita a uma única restrição de desigualdade. A resolução de problemas mais sofisticados é feita a partir da extensão direta dos conceitos discutidos aqui. Portanto, partimos do seguinte problema:

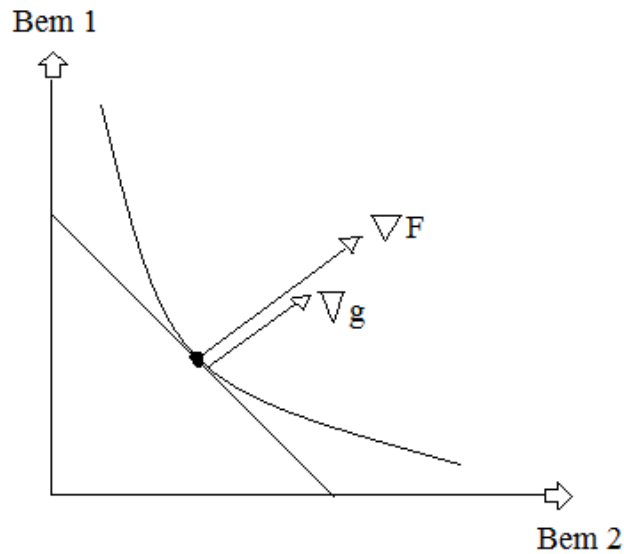
maximize $f(x, y)$ sujeito a $g(x, y) \leq c$.

Para problemas desse tipo temos duas possibilidades. Na primeira delas, o máximo é obtido na fronteira da restrição, $g(x, y) = c$. Diz-se que a restrição é ativa e o problema é exatamente o mesmo que o discutido nas seções anteriores. Alternativamente, o máximo pode ocorrer em um ponto interior do espaço da restrição, $g(x, y) < c$. Diz-se que a restrição é não ativa.

O diagrama abaixo exemplifica o problema com restrição ativa. Note que nesse caso o problema é exatamente igual ao observado para uma restrição de igualdade, ou seja, as inclinações da curva e da reta são iguais no ponto de máximo. Portanto, as equações de 1ª ordem, $L_x = L_y = 0$, continuam válidas. Essas equações foram também escritas como $\nabla f(X^*) = \mu \nabla g(X^*)$ na seção 1. Como essa relação com os gradientes foi obtida com uma restrição de igualdade, utilizou-se a letra μ . No caso das restrições de desigualdade, faz-se uma pequena mudança de nomenclatura para diferenciar os problemas com restrição de igualdade daqueles com restrição de desigualdade. Utiliza-se a letra λ para o multiplicador de Lagrange em vez de μ e reescreve-se a relação como $\nabla f(X^*) = \lambda \nabla g(X^*)$.

Como discutido no segundo capítulo, o gradiente de uma função indica a direção na qual a inclinação da função no ponto é a máxima possível. Pelo diagrama, observa-se que os gradientes da função objetiva e da restrição são vetores paralelos no ponto de máximo. Ou seja, ambas as funções crescem de forma mais rápida na mesma direção. Assim, pode-se escrever um gradiente como função do outro, desde que um deles seja multiplicado por um parâmetro positivo. Como conclusão, temos que para uma restrição ativa, $\nabla f(X^*) = \lambda \nabla g(X^*)$ e $\lambda > 0$.

DIAGRAMA 2
Otimização com restrição de desigualdade ativa

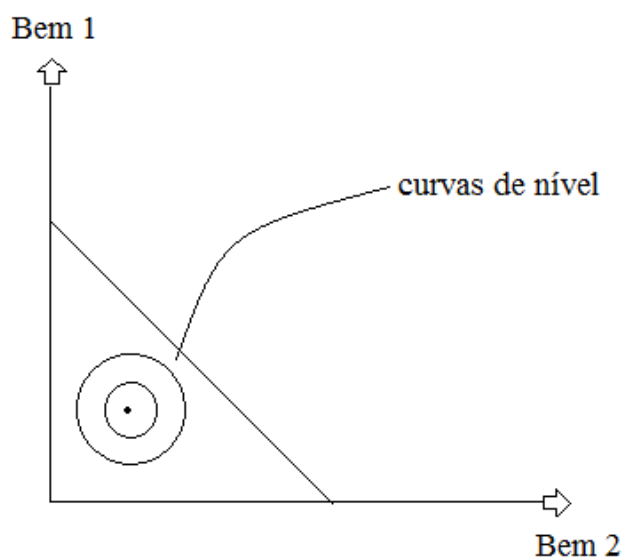


Quando temos uma restrição não ativa, o máximo ocorre no interior do espaço da restrição, como mostra o diagrama 3. Ou seja, o problema é similar a maximização não condicionada. Isto é, as derivadas parciais da função objetiva no máximo são nulas, $f_x(X^*) = f_y(X^*) = 0$, como discutido no capítulo anterior. Assim, o gradiente da função objetiva toma a seguinte forma no máximo:

$\nabla f(X^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Dada a relação $\nabla f = \lambda \nabla g$ apresentada acima, como ∇g não é necessariamente nulo, devemos ter $\lambda = 0$ para que a relação seja satisfeita.

Isso indica que se a restrição for não ativa, como $\lambda = 0$, pode-se excluir a restrição do Lagrangeano. O problema se torna não condicionado com relação à restrição de desigualdade. Note que caso seja esse o caso, as derivadas parciais do Lagrangeano $L_x = L_y = f_x = f_y = 0$, continuam nulas.

DIAGRAMA 3
Otimização com restrição de desigualdade inativa



A tabela a seguir resume toda a discussão com relação às possibilidades de restrição ativa e não ativa:

TABELA 1
Comparação entre restrições ativas e não ativas

Restrição	λ	$g(x, y) - c$	$L_x = L_y$
Ativa	> 0	$= 0$	0
Não ativa	$= 0$	< 0	0

Concluindo a discussão metodológica a respeito da maximização condicionada com restrição de desigualdade, inicialmente devem-se obter os pontos que satisfazem a relação $L_x = L_y = 0$, pois essas condições satisfazem as restrições ativas e não ativas. Em seguida, como não sabemos de antemão se a restrição é ativa ou não, baseados na tabela acima, temos três equações referentes à restrição de desigualdade que devem ser satisfeitas:

$$\lambda \geq 0$$

$$g(x, y) - c \leq 0$$

$$\lambda(g(x, y) - c) = 0$$

Utilizando essas equações, garante-se que ambas as possibilidades estão sendo analisadas.

Segue um exemplo aplicado de todo esse procedimento, muito similar ao exemplo analisado na seção 2.

Um indivíduo quer maximizar a utilidade de suas férias, cuja função utilidade é dada por $U(x, y) = x^2 + y$, onde x são dias na praia e y são dias na montanha. Este tem R\$ 1000 reais para gastar nas férias, cada dia na praia sai por 50 reais e cada dia na montanha custa 100.

Note que a restrição orçamentária não mudou: $50x + 100y = 1000$, ou de forma simplificada, $x + 2y = 20$. A função utilidade foi ligeiramente mudada. No exemplo anterior a função utilidade era um produto, onde no ponto de máximo tínhamos $x > 0$ e $y > 0$. Como a função utilidade é neste novo exemplo representada por uma soma, tanto x como y podem assumir o valor zero em um eventual máximo, pois essa função pode assumir valores positivos, mesmo que uma das variáveis seja zero. Assim, incluem-se duas restrições de desigualdade no problema, $x \geq 0$ e $y \geq 0$, pois o número de dias passados em cada localidade é maior ou igual a zero.

O primeiro passo na obtenção do Lagrange é escrever as restrições de desigualdade como:

$$-x \leq 0 \text{ e } -y \leq 0.$$

Note que esse é o formato que a restrição de desigualdade aparece em toda discussão metodológica de maximização com esse tipo de restrição.

Em seguida, monta-se o Lagrangeano:

$$L(x, y, \mu, \lambda_1, \lambda_2) = x^2 + y - \mu(x + 2y - 20) - \lambda_1(-x) - \lambda_2(-y).$$

Esse problema tem restrições de igualdade e de desigualdade, portanto, como forma de diferenciação entre elas no Lagrangeano, para as primeiras é incluído o multiplicador de Lagrange representado por μ e para as segundas, λ .

Inicialmente o problema se resolve como os descritos nas seções anteriores com os problemas com restrição de igualdade:

$$L_x = 2x - \mu + \lambda_1 = 0$$

$$L_y = 1 - 2\mu + \lambda_2 = 0$$

$$L_\mu = -(x + 2y - 20) = 0$$

Note que para a restrição de igualdade, obtém-se a derivada parcial com relação a μ . Entretanto, para as restrições de desigualdade, o processo é totalmente diferente. Para cada uma das restrições, escrevem-se as três equações descritas anteriormente, que foram baseadas na tabela 1.

Assim, ficamos com seis equações, três para cada uma das restrições:

$$\lambda_1 \geq 0 \quad x \geq 0 \quad \lambda_1 x = 0$$

$$\lambda_2 \geq 0 \quad y \geq 0 \quad \lambda_2 y = 0$$

A resolução do problema passa, portanto, por essas nove equações. Reescrevem-se as três primeiras de forma que não tenham sinais negativos, pois isso facilita a interpretação das mesmas:

$$2x + \lambda_1 = \mu$$

$$1 + \lambda_2 = 2\mu$$

$$x + 2y = 20$$

Depois se verifica se existe um número isolado nas duas primeiras, pois isso também facilita a obtenção do ponto de máximo condicionado. Note que na segunda equação aparece o número 1 de forma isolada. Parte-se da constatação óbvia que $1 > 0$.

Como $\lambda_2 \geq 0$, temos:

$$1 + \lambda_2 > 0.$$

Daí resulta que: $2\mu > 0$, $\mu > 0$, $2x + \lambda_1 > 0$.

Como $\lambda_1 x = 0$, temos:

$$x > 0 \text{ ou } \lambda_1 > 0.$$

1) Se $\lambda_1 > 0$:

como $\lambda_1 x = 0$, $x = 0$.

Daí, temos:

$$2y = 20$$

$$y = 10$$

Substituindo o ponto encontrado na função utilidade, temos:

$$U(0,10) = 10$$

Esse é o primeiro candidato a máximo. Em seguida, verificamos os outros candidatos para definirmos qual é o ponto de máximo condicionado.

2) Se $x > 0$:

como $\lambda_1 x = 0$, $\lambda_1 = 0$.

$$2x = \mu$$

$$x = \mu/2$$

Neste ponto do problema temos mais de um caminho a seguir. Deve-se, portanto, fazer hipóteses. Por exemplo, como $y \geq 0$, temos: $y > 0$ ou $y = 0$. Note que outras hipóteses podem ser feitas e o resultado final de maximização é o mesmo.

2a) Se $y = 0$:

$$x = 20$$

Substituindo o ponto na função utilidade:

$$U(20,0) = 400$$

Esse é o segundo candidato a máximo. Como o valor da utilidade deste é maior do que no candidato anterior, ficamos com esse momentaneamente e descartamos o primeiro.

2b) Se $y > 0$:

como $\lambda_2 y = 0$, $\lambda_2 = 0$.

Então:

$$1 = 2\mu$$

$$\mu = \frac{1}{2}$$

$$x = \frac{\mu}{2} = \frac{1}{4}.$$

$$\frac{1}{4} + 2y = 20$$

$$y = \frac{80-1}{8} = \frac{79}{8}.$$

$$U\left(\frac{1}{4}, \frac{79}{8}\right) = \frac{1}{16} + \frac{79}{8} = \frac{159}{16}$$

Esse é o terceiro candidato a máximo e como ele é inferior ao anterior, ele é descartado. O ponto de máximo é, portanto, $(20,0)$. O indivíduo irá passar todo o tempo na praia, 20 dias no total.

5. MINIMIZAÇÃO COM RESTRIÇÕES DE DESIGUALDADE

O processo utilizado para a minimização com restrições de desigualdade é similar ao descrito para maximização, com apenas uma pequena modificação. Diferentemente dos problemas de maximização, onde as restrições de desigualdade são escritas no formato $g(x, y) - c \leq 0$ quando inseridas no Lagrangeano, nos problemas de minimização essas são escritas no formato $g(x, y) - c \geq 0$. Segue uma ilustração sobre as diferenças entre os problemas de maximização e de minimização.

Por exemplo, em um problema de maximização da utilidade, $U(x, y)$, assuma para efeitos ilustrativos que a restrição orçamentária é de desigualdade, $g(x, y) - c \leq 0$. Em um problema similar, pode-se minimizar os gastos orçamentários, $g(x, y)$, sujeitos a uma utilidade mínima $U(x, y) - c \geq 0$. O importante aqui é notar que no problema de maximização a restrição é escrita no formato $g(x, y) - c \leq 0$, e no problema de minimização essa é escrita como $U(x, y) - c \geq 0$.

Segue o mesmo exemplo aplicado apresentado na seção anterior, mas agora em um formato de minimização dos gastos orçamentários.

Um indivíduo quer minimizar seus gastos sabendo que cada dia na praia sai por 50 reais e cada dia na montanha custa 100. A utilidade de suas férias é dada por $U(x, y) = x^2 + y$, sendo que seu nível mínimo de utilidade demandada é 100.

Ou seja, essa restrição de utilidade mínima é escrita como $x^2 + y - 100 \geq 0$. Novamente, incluem-se as restrições de desigualdade no problema, $x \geq 0$ e $y \geq 0$, pois o indivíduo pode minimizar os custos ficando apenas na praia ou apenas na montanha. Note que todas as restrições já estão no formato para serem incorporadas ao Lagrangeano.

Em seguida, monta-se essa função:

$$L(x, y, \mu, \lambda_1, \lambda_2) = 50x + 100y - \lambda_1(x^2 + y - 100) - \lambda_2(x) - \lambda_3(y).$$

Note que são três restrições de desigualdade. Obtêm-se as equações para a resolução do problema.

$$L_x = 50 - 2x\lambda_1 - \lambda_2 = 0$$

$$L_y = 100 - \lambda_1 - \lambda_3 = 0$$

$$\lambda_1 \geq 0 \quad x^2 + y - 100 \geq 0 \quad \lambda_1(x^2 + y - 100) = 0$$

$$\lambda_2 \geq 0 \quad x \geq 0 \quad \lambda_2 x = 0$$

$$\lambda_3 \geq 0 \quad y \geq 0 \quad \lambda_3 y = 0$$

Novamente, reescrevem-se as duas primeiras equações eliminando os sinais negativos:

$$50 = 2x\lambda_1 + \lambda_2$$

$$100 = \lambda_1 + \lambda_3$$

Esse problema deve ser resolvido a partir dessas últimas 11 equações. A questão é, por onde começar? Devem-se fazer hipóteses e existem inúmeras possibilidades que chegarão ao mesmo resultado final. Uma boa opção é partir de λ_1 , pois esse termo aparece nas duas últimas equações.

1) Se $\lambda_1 = 0$:

$$\lambda_2 = 50$$

$$\lambda_3 = 100$$

Como $\lambda_2 x = 0$ e $\lambda_3 y = 0$, tem-se que: $x = 0$ e $y = 0$.

Porém, caso isso ocorra, a equação $x^2 + y - 100 \geq 0$ não é satisfeita. Assim, descarta-se essa possibilidade, pois todas as 11 equações acima devem ser satisfeitas no ponto de mínimo.

Então $\lambda_1 > 0$:

Como $\lambda_1(x^2 + y - 100) = 0$, tem-se que:

$$x^2 + y = 100.$$

Até aqui temos um caminho único. Por onde continuar? De novo, temos diferentes caminhos e continuamos com hipóteses para $\lambda_2 \geq 0$:

A) Se $\lambda_2 > 0$:

Como $\lambda_2 x = 0$, tem-se que $x = 0$, o que resulta em $y = 100$.

Daí, temos que o custo das férias,

$C(x, y) = 50x + 100y$, nesse ponto é:

$$C(0, 100) = 10000.$$

Esse é o primeiro candidato a mínimo. Vejamos os demais.

B) Se $\lambda_2 = 0$:

$$50 = 2x\lambda_1$$

Daí, temos que $x > 0$, e nada mais sabemos. Vamos então ter que fazer outra hipótese, lembrando que já estamos em uma hipótese inicial.

B1) Se $\lambda_3 > 0$:

tem-se que $y = 0$, o que resulta em $x = 10$.

Obtém-se um segundo ponto $(10, 0)$, cujo custo é:

$$C(10,0) = 500.$$

Esse é o segundo candidato a mínimo e como ele é muito mais barato que o anterior, ficamos momentaneamente com ele e descartamos o primeiro.

B2) Se $\lambda_3 = 0$:

$$50 = 2x\lambda_1$$

$$100 = \lambda_1$$

Daí, obtemos o valor de $x = \frac{1}{4}$.

Sabendo que $x^2 + y = 100$, temos:

$$y = 100 - \frac{1}{16} = \frac{1599}{16}$$

$$C\left(\frac{1}{4}, \frac{1599}{16}\right) = \frac{50}{4} + \frac{159900}{16}.$$

Esse é o terceiro candidato a mínimo e como ele é muito mais caro que o anterior, ficamos definitivamente com o ponto de mínimo nos custos das férias $(10,0)$. Ou seja, 10 dias na praia.

6. TÓPICOS ADICIONAIS

Essa última seção do capítulo apresenta dois tópicos adicionais que enriquecem a discussão anterior, mas que não são imprescindíveis na maioria dos problemas aplicados. O primeiro deles é referente às condições de segunda ordem para problemas condicionados. Como vimos anteriormente, muitos dos problemas com restrições são resolvidos sem que seja necessária a utilização dessas condições. O segundo tópico abordado é uma discussão um pouco mais teórica sobre a qualificação da restrição não degenerada, indicando as condições necessárias para o uso do Lagrangeano.

6.1. Otimização condicionada: condições de segunda ordem

Como discutido no capítulo anterior, quando discutimos as condições de 2ª ordem para problemas não condicionados, utilizamos a matriz hessiana. Para problemas condicionados, de forma relativamente similar, utiliza-se a matriz hessiana orlada.

A apresentação dessa matriz será feita a partir do exemplo genérico apresentado na seção dois:

maximize $f(x, y) = x + y$, sujeito a restrição $x^2 + y^2 = 1$.

O Lagrangeano é $L(x, y, \lambda) = x + y - \lambda(x^2 + y^2 - 1)$, e foram obtidos dois pontos a partir das condições de primeira ordem: $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ e $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$.

Esses pontos podem ser analisados a partir das condições de segunda ordem com a matriz hessiana orlada. Por brevidade, analisa-se apenas o primeiro desses pontos, que, como já sabemos, é ponto de máximo. A matriz hessiana orlada tem o seguinte formato:

$$He_{orlada} = \begin{pmatrix} 0 & \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial y} \\ \frac{\partial h}{\partial y} & \frac{\partial^2 L}{\partial y \partial x} & \frac{\partial^2 L}{\partial y^2} \end{pmatrix}.$$

No caso desse exemplo, ficamos com:

$$He_{orlada} = \begin{pmatrix} 0 & 2x & 2y \\ 2x & -2\lambda & 0 \\ 2y & 0 & -2\lambda \end{pmatrix}$$

Utilizando a relação descrita na seção 2, obtém-se o valor de λ para o ponto específico.

$$\frac{1}{\lambda} = 2x$$

$$\lambda = \frac{1}{2x} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

Daí, substituindo na matriz, temos:

$$He_{orlada} = \begin{pmatrix} 0 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \\ \sqrt{2} & 0 & -\sqrt{2} \end{pmatrix}$$

Como analisar essa matriz?

Seja a matriz hessiana orlada com n variáveis e k restrições. Cheque os sinais dos determinantes das últimas $n - k$ submatrizes, ou seja, cheque os sinais dos $n - k$ menores líderes principais:

- i) Se o determinante da matriz orlada tem o mesmo sinal que $(-1)^n$ e os sinais dos determinantes das submatrizes se alternam, então, o ponto é de máximo.
- ii) Se o determinante da matriz orlada tem o mesmo sinal que $(-1)^k$ e os sinais dos determinantes das submatrizes são iguais a este primeiro, então, o ponto é de mínimo.
- iii) Se essas relações são violadas por determinantes não nulos, o ponto não é de máximo nem de mínimo.

No caso desse exemplo temos $n = 2$ variáveis e $k = 1$ restrições e deve-se verificar $n - k = 2 - 1 = 1$, um único determinante, que é o de toda matriz hessiana orlada:

$$\begin{vmatrix} 0 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \\ \sqrt{2} & 0 & -\sqrt{2} \end{vmatrix} = 4\sqrt{2} > 0$$

Verifica-se que a primeira hipótese é a verdadeira e o ponto é de máximo. Seguindo esse mesmo procedimento pode-se verificar o outro ponto crítico.

A generalização dessa matriz para problemas com mais variáveis e restrições é direta. Segue o exemplo apresentado na seção 3, que tem o seguinte Lagrangeano:

$$L(x, y, z, \mu_1, \mu_2) = xyz - \mu_1(x^2 + y^2 - 1) - \mu_2(x + z - 1).$$

O primeiro passo é obter a matriz hessiana orlada para o problema com três variáveis e duas restrições:

$$He_{orlada} = \begin{pmatrix} 0 & 0 & \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} & \frac{\partial h_1}{\partial z} \\ 0 & 0 & \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} & \frac{\partial h_2}{\partial z} \\ \frac{\partial h_1}{\partial x} & \frac{\partial h_2}{\partial x} & \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial y} & \frac{\partial^2 L}{\partial x \partial z} \\ \frac{\partial h_1}{\partial y} & \frac{\partial h_2}{\partial y} & \frac{\partial^2 L}{\partial x \partial y} & \frac{\partial^2 L}{\partial y^2} & \frac{\partial^2 L}{\partial y \partial z} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_2}{\partial z} & \frac{\partial^2 L}{\partial x \partial z} & \frac{\partial^2 L}{\partial y \partial z} & \frac{\partial^2 L}{\partial z^2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 2x & 2y & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 2x & 1 & -2\mu_1 & z & y \\ 2y & 0 & z & -2\mu_1 & x \\ 0 & 1 & y & x & 0 \end{pmatrix}.$$

De posse dos pontos obtidos com as condições de 1ª ordem pelo sistema 5 x 5 apresentado na seção 3, obtém-se uma matriz 5 x 5 para cada deles. Em seguida, calcula-se o determinante de uma matriz, pois

são três variáveis, $n = 3$, e duas restrições, $k = 2$, $n - k = 3 - 2 = 1$, ou seja, toda a matriz 5×5 , e segue-se a mesma regra já enunciada.

6.2. Qualificação da restrição não degenerada

Quando discutimos a obtenção da função de Lagrange para o problema com duas variáveis e uma restrição de igualdade, vimos que para que essa função exista, necessariamente:

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

No caso de mais restrições de igualdade, a obtenção da qualificação da restrição não degenerada (NDCQ na sigla em inglês), indicando que o Lagrangeano pode ser usado, é obtida de forma análoga, utilizando a matriz Jacobiana:

$$Dh(X^*) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(X^*) & \dots & \frac{\partial h_1}{\partial x_n}(X^*) \\ \dots & \dots & \dots \\ \frac{\partial h_m}{\partial x_1}(X^*) & \dots & \frac{\partial h_m}{\partial x_n}(X^*) \end{pmatrix}.$$

Primeiro, deve-se ter $m < n$, ou seja, o número de restrições de igualdade deve ser menor que o número de variáveis. Além disso, o posto da matriz deve ser máximo, isto é, igual a m , e diz-se que a matriz satisfaz NDCQ.

Se essa matriz tem posto máximo, as linhas são independentes e não podem ser escritas como função das demais. Ou seja, a condição NDCQ implica que:

$$(1) \quad \alpha_1 \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(X^*) \\ \dots \\ \frac{\partial h_1}{\partial x_n}(X^*) \end{pmatrix} + \dots + \alpha_m \begin{pmatrix} \frac{\partial h_m}{\partial x_1}(X^*) \\ \dots \\ \frac{\partial h_m}{\partial x_n}(X^*) \end{pmatrix} \neq \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \text{ com } \alpha_1, \dots, \alpha_m \text{ não todos nulos.}$$

Caso isso seja verdadeiro, quer-se mostrar que o Lagrangeano pode ser utilizado. Ou seja, que é válida a expressão:

$$(2) \quad \begin{pmatrix} \frac{\partial f}{\partial x_1}(X^*) \\ \dots \\ \frac{\partial f}{\partial x_n}(X^*) \end{pmatrix} - \mu_1 \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(X^*) \\ \dots \\ \frac{\partial h_1}{\partial x_n}(X^*) \end{pmatrix} - \dots - \mu_m \begin{pmatrix} \frac{\partial h_m}{\partial x_1}(X^*) \\ \dots \\ \frac{\partial h_m}{\partial x_n}(X^*) \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}$$

Vejamos como relacionar essas duas expressões.

Faça $f(X^*) = C_0$ e considere o sistema:

$$f(X^*) = C_0$$

$$h_1(X^*) = C_1$$

...

$$h_m(X^*) = C_m$$

Sabe-se que o ponto de máximo, X^* , é solução do sistema formado pelas m equações das restrições, pois esse ponto é obtido pela interseção dos mesmos.

Pense que em um sistema que também incluía a primeira dessas equações e também assuma que C_0, \dots, C_m são variáveis exógenas. Esse novo sistema tem a seguinte matriz Jacobiana $(m + 1) \times n$:

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(X^*) & \dots & \frac{\partial f}{\partial x_n}(X^*) \\ \frac{\partial h_1}{\partial x_1}(X^*) & \dots & \frac{\partial h_1}{\partial x_n}(X^*) \\ \dots & \dots & \dots \\ \frac{\partial h_m}{\partial x_1}(X^*) & \dots & \frac{\partial h_m}{\partial x_n}(X^*) \end{pmatrix}$$

Assuma que essa última matriz tem posto máximo igual a $m + 1$. Se ela tem posto máximo, as linhas são independentes e temos uma nova solução obtida a partir da perturbação de C_0 para $C_0 + \varepsilon$ com $\varepsilon > 0$. Assim, as últimas m equações do sistema estão satisfeitas, bem como a primeira. Entretanto, vemos que a nova solução com o sistema perturbado X^{**} é tal que $f(X^{**}) > f(X^*)$, uma vez que $C_0 + \varepsilon > C_0$. Isso contradiz o fato que X^* é ponto de máximo no conjunto das restrições.

Assim, conclui-se que a matriz acima não tem posto máximo e as linhas não são independentes. Portanto, pode-se escrevê-las como:

$$(3) \quad \alpha_0 \begin{pmatrix} \frac{\partial f}{\partial x_1}(x^*) \\ \dots \\ \frac{\partial f}{\partial x_n}(x^*) \end{pmatrix} + \alpha_1 \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(x^*) \\ \dots \\ \frac{\partial h_1}{\partial x_n}(x^*) \end{pmatrix} + \dots + \alpha_m \begin{pmatrix} \frac{\partial h_m}{\partial x_1}(x^*) \\ \dots \\ \frac{\partial h_m}{\partial x_n}(x^*) \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \text{ com } \alpha_0, \alpha_1, \dots, \alpha_m \text{ não todos}$$

nulos.

Comparando essa expressão com (1), verifica-se que necessariamente $\alpha_0 \neq 0$. Define-se $\mu_i \equiv -\alpha_i / \alpha_0$ e reescreve-se a relação (3) exatamente como (2). Assim, temos uma demonstração que podemos usar o Lagrangeano em problemas de otimização com múltiplas restrições de igualdade se a matriz Jacobiana com as restrições tiver posto máximo.

Como exemplo de aplicação dessa discussão, utiliza-se o exemplo mostrado na seção três, com o problema de maximização de $f(x, y, z) = xyz$ sujeito às restrições $h_1 \equiv x^2 + y^2 = 1$ e $h_2 \equiv x + z = 1$.

O primeiro passo é determinar se as restrições são qualificadas. Obtemos, assim, a matriz Jacobiana das mesmas e verificamos se a matriz tem posto máximo.

$$Dh(x^*) = \begin{pmatrix} \frac{\partial h_1}{\partial x}(X^*) & \frac{\partial h_1}{\partial y}(X^*) & \frac{\partial h_1}{\partial z}(X^*) \\ \frac{\partial h_2}{\partial x}(X^*) & \frac{\partial h_2}{\partial y}(X^*) & \frac{\partial h_2}{\partial z}(X^*) \end{pmatrix} = \begin{pmatrix} 2x & 2y & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Essa matriz não terá posto máximo somente se $x = 0$ e $y = 0$. Entretanto, esse ponto não pertence a h_1 e está fora da interseção das restrições. Ou seja, no espaço de restrições do problema de otimização dado pela interseção de h_1 e h_2 , no ponto de máximo X^* , x e y não podem ser zero simultaneamente. Portanto, as restrições são qualificadas e o Lagrangeano pode ser utilizado.

Note que para restrições de desigualdade a condição NDCQ é a mesma que a observada acima para as restrições de igualdade, porém se leva em conta apenas as restrições ativas, uma vez que as demais são não ativas e não interferem no problema de otimização.

Por exemplo, no problema já descrito na seção 4, onde um indivíduo quer maximizar $U(x, y) = x^2 + y$ sujeito às restrições $x + 2y = 20$, $x \geq 0$ e $y \geq 0$. O ponto de máximo foi obtido em $(20, 0)$. Portanto, entre essas restrições e usando o ponto de máximo, depreende-se que a primeira é de igualdade, a segunda é não ativa e a terceira é ativa. Daí, temos como matriz Jacobiana, onde a segunda restrição não é incluída:

$$Dh(x^*) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \text{ matriz de posto máximo.}$$

CAPÍTULO 10 - FUNÇÕES CÔNCAVAS E FUNÇÕES QUASE CÔNCAVAS

O capítulo discute as funções côncavas e convexas, e também as funções quase côncavas e quase convexas, sendo que todas elas aparecem regularmente em problemas econômicos, como na otimização não condicionada, na análise de riscos e na teoria microeconômica. Entretanto, antes de discutirmos os conceitos referentes a essas funções, devem-se introduzir alguns conceitos iniciais referentes às funções homogêneas, às transformações monotônicas, às características ordinais e cardinais e os conjuntos convexos, que servirão de base para a discussão posterior. Em seguida, discutem-se as funções côncavas, quase côncavas e pseudocôncavas. Para tanto, o capítulo foi dividido em seis seções: i) Funções homogêneas, transformações monotônicas e propriedades ordinais e cardinais; ii) Conjuntos convexos; iii) Funções côncavas e funções convexas; iv) Propriedades de funções côncavas; e v) Funções quase côncavas. Assume-se que o leitor tenha conhecimentos sólidos em álgebra linear, em cálculo de várias variáveis, e em otimização não condicionada e condicionada, temas discutidos em capítulos anteriores. Além disso, o leitor deve ter conhecimentos básicos de microeconomia.

1. Funções homogêneas, transformações monotônicas e propriedades ordinais e cardinais.

Esta seção apresenta os conceitos de funções homogêneas, de transformações monotônicas e de propriedades ordinais e cardinais, conceitos estes utilizados nas discussões posteriores sobre funções côncavas.

As funções homogêneas aparecem naturalmente em muitos problemas econômicos, como em funções de utilidade e funções de produção. Por exemplo, nas funções de produção $Q(L, K) = L^{1/3} K^{2/3}$, $Q(L, K) = LK$ e $Q(L, K) = LK^2$, que são respectivamente funções homogêneas de grau 1, 2 e 3.

Uma **função homogênea** de grau k é aquela que satisfaz a seguinte relação:

$$F(tx_1, \dots, tx_n) = t^k F(x_1, \dots, x_n).$$

Note que para a função $Q(L, K) = L^{1/3} K^{2/3}$, temos:
 $Q(tL, tK) = (tL)^{1/3} (tK)^{2/3} = t(L^{1/3} K^{2/3}) = tQ(L, K)$. Ou seja, como já enunciado, essa função é homogênea de grau 1.

Um polinômio será uma função homogênea se todos seus monômios forem de mesmo grau. Por exemplo, a função $f(x, y) = xy + x^2 + x^3 y^{-1}$ é homogênea de grau 2, pois todos os monômios são desse grau, enquanto $f(x, y) = xy + x$ não é homogênea, porque existem monômios de diferentes graus.

Outro conceito que será utilizado na discussão posterior sobre funções côncavas e quase côncavas é o de transformação monotônica. Tome uma função estritamente crescente $g: I \rightarrow \mathfrak{R}$, em que $I \subset \mathfrak{R}$. Por exemplo, as funções $g_1(x) = x + 1$, $g_2(x) = e^x$, $g_3(x) = \ln x$ e $g_4(x) = x^2$, sendo que para as duas últimas $I = \mathfrak{R}^+$.

Diz-se que $g: I \rightarrow \mathfrak{R}$ é uma transformação monotônica de I . Além disso, pode-se também aplicar uma transformação monotônica em funções. Seja $U = U(x_1, \dots, x_n)$ uma função real de n variáveis, então $g \circ U(x_1, \dots, x_n)$ é uma transformação monotônica de U .

Como exemplo, tome uma função de utilidade homogênea de grau 2, $U(x, y) = xy$, e as transformações monotônicas descritas acima. Quando essas transformações são aplicadas a essa função obtemos respectivamente as seguintes funções: $V_1(x, y) = xy + 1$, $V_2(x, y) = e^{xy}$, $V_3(x, y) = \ln(xy)$ e $V_4(x, y) = x^2 y^2$. Note que as três primeiras são não homogêneas e a última é homogênea de grau 4. Ou seja, uma transformação monotônica de uma função homogênea não necessariamente gera uma função homogênea, isto é, a propriedade de homogeneidade não é mantida por qualquer transformação monotônica.

De posse desses conceitos, define-se o que é uma característica ordinal e o que é uma característica cardinal. Uma característica é ordinal se toda transformação monotônica preserva essa característica. Caso contrário, a característica é cardinal.

Por exemplo, a homogeneidade de funções não é ordinal, pois não necessariamente é uma propriedade mantida por uma transformação monotônica. Portanto, a homogeneidade é uma propriedade cardinal. Segue um exemplo de propriedade ordinal. Dadas duas curvas de indiferença de $U = U(x, y)$, tais que $U(x_1, y_1) > U(x_2, y_2)$. Note que, quando se faz qualquer transformação monotônica nessa função de utilidade, isso implica em $V(x_1, y_1) > V(x_2, y_2)$. Isto é, a ordem de preferência definida pela função de utilidade é preservada por uma transformação monotônica, ou seja, essa é uma característica ordinal.

2. Conjuntos convexos

O conceito de conjunto convexo é utilizado em muitas definições que são diretamente aplicadas a modelos econômicos, como os conceitos de função côncava e de função convexa, apresentados posteriormente. Note que não se deve confundir uma função convexa com um conjunto convexo, eles são conceitos totalmente distintos. Segue uma definição desse último.

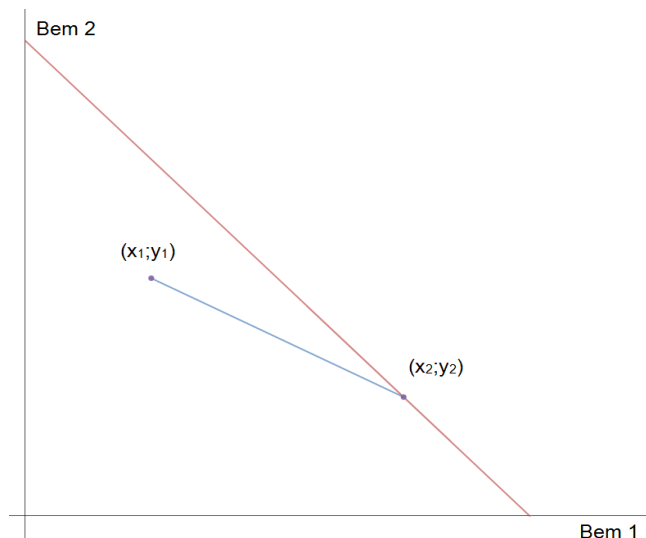
Um conjunto $X \subset \mathfrak{R}^n$ é convexo se $\alpha x + (1 - \alpha)x' \in X$ sempre que $x, x' \in X$ e $\alpha \in [0, 1]$.

Como exemplo seguem duas aplicações da teoria microeconômica. Por simplicidade de visualização, se expressa o conjunto X em diagramas no \mathfrak{R}^2 .

O diagrama 1 representa uma restrição orçamentária e duas cestas de bens quaisquer $(x_1, y_1) \in X$ e $(x_2, y_2) \in X$, que podem ser adquiridas sujeitas a essa restrição. Note que uma média ponderada das cestas, $\alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2) \in X$, também pode ser adquirida sujeita à mesma restrição orçamentária.

DIAGRAMA 1

Restrição orçamentária

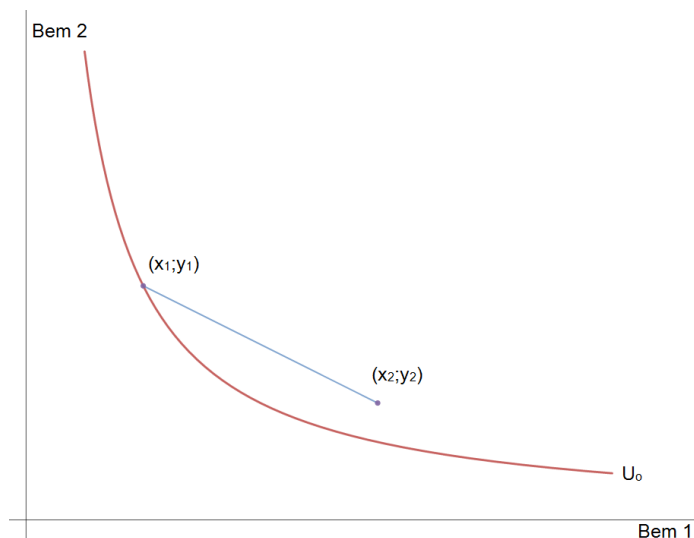


O diagrama 2 representa uma curva de indiferença e determina um conjunto tal que a utilidade é igual ou superior ao valor da utilidade determinada pela curva de indiferença: $U(x, y) \geq U_0$. Definem-se duas cestas de bens pertencentes a esse conjunto: $U(x_1, y_1) \geq U_0$ e $U(x_2, y_2) \geq U_0$. Note que uma média ponderada das duas cestas também pertence ao conjunto: $U[\alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2)] \geq U_0$.

Note por esses diagramas que a definição de conjuntos convexos pode ser descrita como: sempre que dois pontos pertencerem a um conjunto convexo, a reta que liga os mesmos também pertencerá.

DIAGRAMA 2

Conjunto definido acima de uma curva de indiferença de uma função côncava



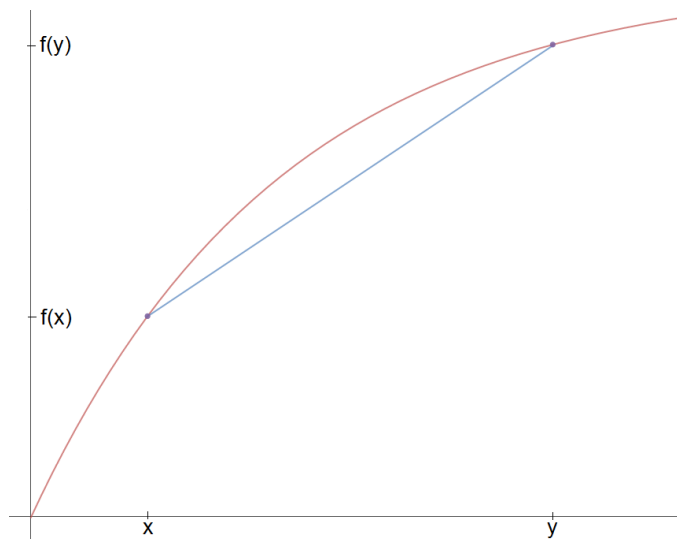
3. Funções côncavas e funções convexas

Aqui serão apresentados alguns conceitos sobre funções côncavas e funções convexas. Em geral, discutem-se pontos referentes somente às primeiras, uma vez que o raciocínio é análogo para as segundas. Segue, portanto, a definição de função côncava. Para convexo, basta trocar o sinal de \geq pelo de \leq .

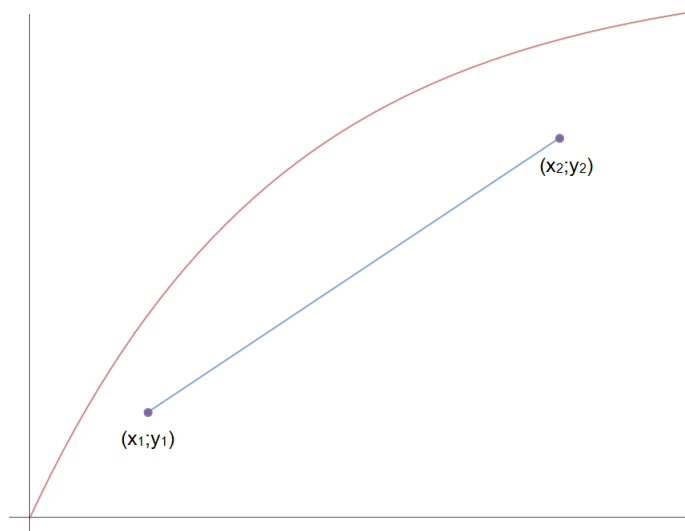
Uma função de valor real definida em um conjunto convexo $X \subset \mathbb{R}^n$ é côncava se, para todo $x, y \in X$ e para todo $\alpha \in [0, 1]$, temos: $f[\alpha y + (1 - \alpha)x] \geq \alpha f(y) + (1 - \alpha)f(x)$.

Uma interpretação geométrica dessa definição de função côncava é bastante útil. Segue um diagrama no \mathbb{R}^2 . Os valores da função côncava em qualquer ponto entre x e y são iguais ou superiores à reta com os pontos da média ponderada dos dois pontos da função $f(x)$ e $f(y)$.

DIAGRAMA 3
Função côncava



Segue outra interpretação geométrica que utiliza os conceitos de função côncava e de conjunto convexo. Por simplicidade, novamente se utiliza um diagrama no \mathbb{R}^2 . Uma função é côncava se, e somente se, o conjunto definido abaixo do gráfico for convexo.

DIAGRAMA 4**Função côncava e conjunto convexo**

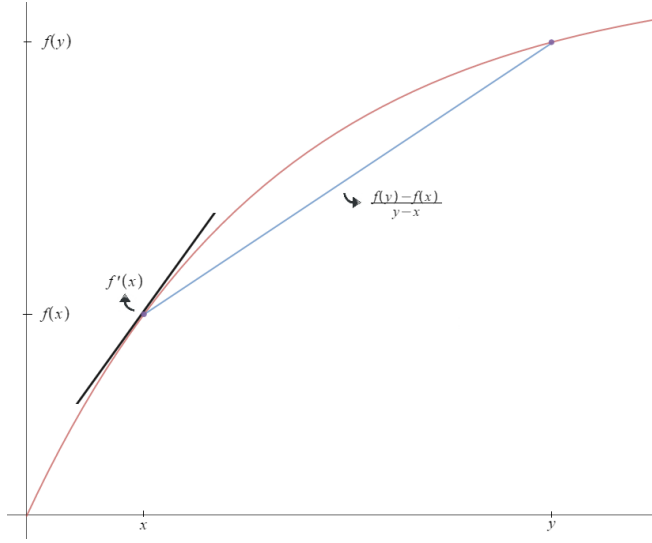
Acima foi definida a função côncava e foram descritas interpretações geométricas para essa definição. Podem-se utilizar também critérios do cálculo para definir se uma função é côncava ou não. Vejamos inicialmente no \mathbb{R}^2 e, em seguida, o raciocínio será generalizado para o \mathbb{R}^n .

Teorema 1 – Seja f uma função definida no intervalo I cuja derivada primeira é contínua. Então f é côncava se, e somente se, $f(y) - f(x) \leq f'(x)(y - x)$ para todo $x, y \in I$.

Note que a relação apresentada no diagrama pode ser escrita para $(y - x) > 0$, como

$$\frac{f(y) - f(x)}{y - x} \leq f'(x).$$

Antes da prova do teorema é mostrado um diagrama com essa relação.

DIAGRAMA 5**Critério do cálculo para a função côncava**

Prova do teorema 1– Parte-se da definição de função côncava:

$$f[\alpha y + (1 - \alpha)x] \geq \alpha f(y) + (1 - \alpha)f(x), \text{ onde } \alpha \in [0, 1].$$

Rescreve-se o termo à esquerda como $\alpha(f(y) - f(x)) + f(x)$ e o da direita como $f[x + \alpha(y - x)]$. Assim, essa relação se torna:

$$\alpha(f(y) - f(x)) + f(x) \leq f(x + \alpha(y - x)).$$

Reescrevendo essa última relação, temos:

$$f(y) - f(x) \leq \frac{f(x + \alpha(y - x)) - f(x)}{\alpha} = \frac{f(x + \alpha(y - x)) - f(x)}{\alpha(y - x)}(y - x).$$

Tomando o limite $\alpha \rightarrow 0$, temos:

$$f(y) - f(x) \leq \lim_{\alpha \rightarrow 0} \frac{f(x + \alpha(y - x)) - f(x)}{\alpha(y - x)}(y - x) = f'(x)(y - x).$$

Ou seja, se a definição de função côncava for satisfeita, a relação do cálculo proposta também será. Reciprocamente, assumindo que a relação proposta no teorema é satisfeita, $f(y) - f(x) \leq f'(x)(y - x)$, fazemos as seguintes mudanças de variável, $y = x$ e $x = (1 - \alpha)x + \alpha y$.

Assim, obtemos:

$$f(x) - f((1 - \alpha)x + \alpha y) \leq f'((1 - \alpha)x + \alpha y)(x - ((1 - \alpha)x + \alpha y)) = -\alpha f'((1 - \alpha)x + \alpha y)(y - x).$$

De forma análoga, mas somente com a mudança $x = (1 - \alpha)x + \alpha y$, obtemos:

$$f(y) - f((1 - \alpha)x + \alpha y) \leq f'((1 - \alpha)x + \alpha y)(y - ((1 - \alpha)x + \alpha y)) = (1 - \alpha)f'((1 - \alpha)x + \alpha y)(y - x).$$

Multiplicando a primeira das expressões por $(1 - \alpha)$ e a segunda por α e somando as duas, obtemos:

$$f[\alpha y + (1 - \alpha)x] \geq \alpha f'(y) + (1 - \alpha)f'(x).$$

A generalização do teorema acima para funções de várias variáveis é direta:

Teorema 2 – Seja F uma função definida no conjunto convexo $U \subset \mathfrak{R}^n$, cujas derivadas parciais de primeira ordem são contínuas. Então F é côncava se, e somente se, $F(y) - F(x) \leq DF(x)(y - x)$ para todo $x, y \in U$.

Esse teorema permite escrever o seguinte corolário:

Corolário 1 - Dada uma função convexa, cujas derivadas parciais de primeira ordem são contínuas, definida no conjunto convexo $U \subset \mathfrak{R}^n$. Seja $x_0 \in U$, então $DF(x_0)(y - x_0) \leq 0$ implica em $F(y) \leq F(x_0)$. Se isso ocorre para todo $y \in U$, x_0 é ponto de máximo global.

Esse corolário mostra a relação direta entre funções côncavas e a otimização não condicionada discutida no oitavo capítulo. Seja f côncava e definida em um subconjunto $X \subset \mathfrak{R}^n$, aberto e convexo. Se x_0 é ponto crítico, ele é ponto de máximo, pois se $DF(x_0) = 0$, $F(y) \leq F(x_0)$ para todo $y \in U$.

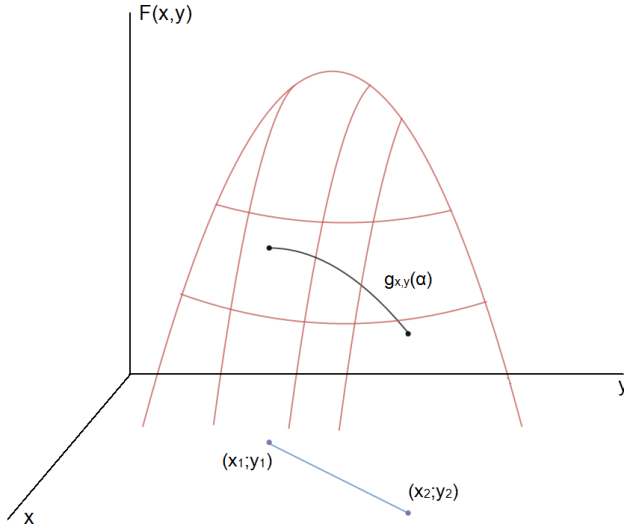
Também como vimos nesse capítulo, pode-se estudar a concavidade de funções a partir da matriz hessiana.

Teorema 3 – Seja f uma função definida em um subconjunto convexo de $X \subset \mathfrak{R}^n$. Assumindo que as derivadas de segunda ordem são contínuas, f é côncava se, e somente se, a matriz hessiana for negativamente semidefinida para todo $x \in U$.

Prova – Dados x e y , pontos arbitrário de U . Defina $g_{x,y}(\alpha) \equiv f(\alpha y + (1 - \alpha)x)$, como mostra o diagrama 6.

DIAGRAMA 6

Segmento de uma função côncava



Sabe-se que uma função é côncava se, e somente se, toda a restrição da mesma em um segmento de linha for uma função côncava de uma variável. Então f será côncava se, e somente se, toda $g_{x,y}(\alpha)$ for côncava. Isso equivale a dizer que $\frac{d^2}{d\alpha^2} g_{x,y}(\alpha) \leq 0$, em todo segmento de linha.

Reescreve-se $g_{x,y}(\alpha)$ e deriva-se:

$$g_{x,y}(\alpha) \equiv f(\alpha(y-x) + x)$$

$$\frac{d}{d\alpha} g_{x,y}(\alpha) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(z)(y_i - x_i), \text{ em que } z \text{ é um ponto arbitrário em } U.$$

Derivando novamente:

$$\begin{aligned} \frac{d^2}{d\alpha^2} g_{x,y}(\alpha) &= \sum_{j=1}^n \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(z)(y_i - x_i)(y_j - x_j) \\ &= \sum_{j=1}^n \sum_{i=1}^n (y_j - x_j) \frac{\partial^2 f}{\partial x_i \partial x_j}(z)(y_i - x_i) = (y-x)^T D^2 f(z)(y-x) = A^T D^2 f(z) A. \end{aligned}$$

Como vimos no oitavo capítulo, se $D^2 f$ é negativamente semidefinida então, por definição,

$$A^T D^2 f(z) A \leq 0 \text{ o que implica } \frac{d^2}{d\alpha^2} g_{x,y}(\alpha) \leq 0.$$

Assim, $g_{x,y}(\alpha)$ é côncava para todo $x, y \in U$ e, conseqüentemente, f também é côncava.

Reciprocamente, assuma que f seja côncava. Como f é côncava, temos $\frac{d^2}{d\alpha^2} g_{x,y}(\alpha) \leq 0$.

Dado que U é aberto, pois as derivadas de segunda ordem existem para todo $x \in U$, existe $t_0 > 0$ tal que $y = z + t_0 v$, onde z é um ponto arbitrário em U e v é uma direção arbitrária no \mathbb{R}^n .

$$\text{Então: } \frac{d^2}{d\alpha^2} g_{z,y}(0) = (y - z)^T D^2 f(z)(y - z) = (t_0 v)^T D^2 f(z)(t_0 v) = (t_0)^2 [v^T D^2 f(z)v] \leq 0.$$

Ou seja, $v^T D^2 f(z)v \leq 0$ e $D^2 f(z)$ é negativamente semidefinida para todo z em U .

4. Propriedades de funções côncavas

As funções côncavas (e convexas) têm várias propriedades que as tornam muito úteis em Economia. Citam-se três delas.

1. Uma função côncava multiplicada por uma constante positiva continua sendo côncava. Se a multiplicação for feita por uma constante negativa, ela passa a ser convexa.

Demonstração – Partindo da definição de função côncava, $f[\alpha y + (1 - \alpha)x] \geq \alpha f(y) + (1 - \alpha)f(x)$, multiplica-se a função por uma constante positiva, b^2 , temos:

$$b^2 f[\alpha y + (1 - \alpha)x] \geq \alpha b^2 f(y) + (1 - \alpha)b^2 f(x).$$

Trocando $b^2 f(\cdot)$ por $g(\cdot)$, obtém-se novamente a definição de função côncava, mas para $g(\cdot)$

2. Médias ponderadas de funções côncavas são côncavas. Ou seja, se $a, b > 0$ e $f(x)$ e $g(x)$ forem côncavas, então $af(x) + bg(x)$ também é côncava.

Esse fato tem aplicação direta em Economia. Dada uma função de bem estar social $W(x_1, x_2, \dots, x_n) = \sum a_i u_i(x_1, x_2, \dots, x_n)$, em que x_i são os bens consumidos, u_i são as utilidades individuais e a_i são constantes não negativas. Se as funções de utilidades individuais forem côncavas, isso implica em uma função de bem estar social também côncava.

3. Funções côncavas têm curvas de nível que limitam por baixo conjuntos convexas.

Sejam x e y pontos pertencentes ao subconjunto delimitado por baixo pela curva de nível de uma função côncava, $f(x) \geq f(x_0)$ e $f(y) \geq f(x_0)$, como já exemplificado pelo diagrama 2. Então, por definição de função côncava, temos:

$$f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y) \geq \alpha f(x_0) + (1 - \alpha)f(x_0) = f(x_0).$$

Ou seja, dados dois pontos quaisquer \mathbf{x} e \mathbf{y} pertencentes ao subconjunto delimitado por baixo por uma curva de nível de uma função côncava, o ponto $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$ também pertence ao conjunto e o conjunto é convexo.

5. Funções quase côncavas

As funções quase côncavas são obtidas a partir de uma transformação monotônica de funções côncavas. Segue um exemplo. A função de produção com elasticidade de substituição constante (CES), $Q(x, y) = (a_1x^r + a_2y^r)^{1/r}$, com $r \in (0,1)$ e $a_1, a_2 > 0$, é quase côncava. Vejamos como que essa função é obtida a partir de uma transformação monotônica de uma função côncava.

Note que x^r e y^r são côncavas, pois $x, y > 0$ e $r \in (0,1)$. Como as constantes são positivas, $a_1, a_2 > 0$, como vimos na seção anterior, a_1x^r e a_2y^r também são côncavas. Além disso, a soma de funções côncavas é côncava. Assim, conclui-se que $f(x, y) = a_1x^r + a_2y^r$ é côncava. Dada uma transformação monotônica $z \mapsto z^{1/r}$, obtemos $Q(x, y)$, que é quase côncava.

Como vimos na primeira seção deste capítulo, existem propriedades ordinais e cardinais. Como apresentado, a ordem de preferência entre curvas de indiferença é uma propriedade ordinal e é preservada em transformações monotônicas.

De maneira distinta, a propriedade de concavidade é uma propriedade cardinal, ou seja, pode não ser preservada por uma transformação monotônica. Por exemplo, uma reta $f(x) = x$ é uma função côncava (e também convexa). Faça uma transformação monotônica da reta com a função $g(x) = e^x$ e, assim, obtém-se $g(f(x)) = e^x$, que é uma função convexa. Concluindo, uma função côncava ao sofrer uma transformação monotônica pode perder a propriedade de concavidade.

Diferentemente da propriedade de concavidade que é cardinal, a quase concavidade é uma propriedade ordinal, ou seja, uma função quase côncava que sofre uma transformação monotônica continua sendo uma função quase côncava. Por exemplo, parte-se de uma função côncava, $f(x, y)$, que sofre uma transformação monotônica $z \mapsto g(z)$ e transforma-se em uma quase côncava.

Se essa nova função sofrer uma segunda transformação monotônica $z \mapsto h(z)$, pode-se representar todo o processo como uma única transformação monotônica, incluindo as duas funções anteriores de forma composta $z \mapsto h \circ g(z)$. Essa última é aplicada a função côncava original, $f(x, y)$, e tem como resultado uma função quase côncava. Ou seja, uma função quase côncava que sofre uma transformação monotônica continua sendo quase côncava, e a propriedade de quase concavidade é ordinal.

Já vimos como obter uma função quase côncava a partir de uma côncava. Segue uma definição de funções quase côncavas.

Uma função definida em um subconjunto convexo $U \subset \Re^n$ é quase côncava se para todo número real a e $x, y \in U$, quando $f(x) \geq a$ e $f(y) \geq a$ implica em $f(\alpha x + (1-\alpha)y) \geq a$ para todo $\alpha \in [0,1]$. Ou seja, o conjunto $C_a^+ \equiv \{x \in U; f(x) \geq a\}$ é um conjunto convexo.

Dessa definição resulta que, para todo $\alpha \in [0,1]$ e para qualquer $x, y \in U$, $f(x) \geq f(y)$ implica $f(\alpha x + (1-\alpha)y) \geq f(y)$.

Note que a aplicação desse conceito na teoria microeconômica é direta, como será descrito em maiores detalhes no fim dessa seção.

De posse das definições acima, se afirma que toda função côncava é quase côncava, mas a recíproca não é verdadeira. Ou seja, as funções côncavas são um subconjunto das funções quase côncavas.

Prova – Partimos da definição de função côncava, $f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y)$.

Seja $f(x) \geq f(y)$, e assim temos:

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) \geq \alpha f(y) + (1-\alpha)f(y) = f(y).$$

Essa é uma das definições de função quase côncava. Assim, demonstrou-se que toda função côncava é também quase côncava.

Contrariamente, partimos de uma das definições de quase côncava:

$$f(x) \geq f(y) \text{ implica } f(\alpha x + (1-\alpha)y) \geq f(y).$$

Assim, se $f(x) \geq f(y)$ temos:

$$f(\alpha x + (1-\alpha)y) \geq f(y) = \alpha f(y) + (1-\alpha)f(y) \leq \alpha f(x) + (1-\alpha)f(y).$$

Note que dados os sinais das desigualdades, o argumento é inconcluso. Ou seja, não necessariamente uma função quase côncava será uma função côncava.

Seguem dois exemplos de funções quase côncavas que não são côncavas. Basta partir de uma função côncava e realizar uma transformação monotônica em que a propriedade de concavidade não seja preservada.

Exemplo 1 – Note que uma reta $f(x) = x$ é uma função côncava e, portanto, como toda função côncava, ela é também quase côncava. Seguindo o exemplo já descrito, dada uma transformação monotônica $z \mapsto e^z$, obtém-se a função $h(x) = e^x$. Essa função é uma função quase côncava, mas não é côncava, na verdade é convexa.

Exemplo 2 – Seja $f(x) = \ln x$ para $x > 0$, que é uma função côncava e, portanto, quase côncava. Dada uma transformação linear $z \mapsto e^{z^2}$, obtém-se a função $h(x) = x^2$, que é quase côncava, porém, não é côncava.

Assim como foi feito para funções côncavas, existem critérios do cálculo para definir se uma função é quase côncava.

Assuma que F seja uma função definida em um conjunto aberto convexo $U \subset \mathfrak{R}^n$, cujas derivadas primeiras são contínuas. Se F for quase côncava, então $F(y) \geq F(x)$ implica em $DF(x)(y - x) \geq 0$.

Se a função for quase convexa, então $F(y) \leq F(x)$ implica em $DF(x)(y - x) \leq 0$.

Prova para funções quase côncavas – Assuma que F seja quase côncava e que $F(y) \geq F(x)$. Então, por definição de quase côncava, $F(\alpha x + (1 - \alpha)y) \geq F(x)$.

Assim, temos que para $\alpha \in (0, 1]$:

$$\frac{F(\alpha x + (1 - \alpha)y) - F(x)}{\alpha} \geq 0$$

$$\frac{F(\alpha x + (1 - \alpha)y) - F(x)}{\alpha(y - x)}(y - x) \geq 0.$$

Tomando o limite $\alpha \rightarrow 0$, ficamos com: $DF(x)(y - x) \geq 0$.

Note que a recíproca também é verdadeira. Entretanto, em vez de uma prova formal, seguem dois exemplos.

Exemplo 1 – A função $f(x) = e^x$ tem derivada positiva, $f'(x) = e^x > 0$. Note que se $f(y) \geq f(x)$ implica em $y \geq x$. Ou seja, $f(y) \geq f(x)$ implica em $f'(x)(y - x) \geq 0$.

Exemplo 2 – A função $f(x) = x^3$ também tem derivada positiva, $f'(x) = 3x^2 \geq 0$. Note que se $f(y) \geq f(x)$, também temos $y - x \geq 0$, o que implica em $f'(x)(y - x) \geq 0$.

Vimos na primeira seção do capítulo que a ordem de preferência entre duas cestas de bens é preservada por uma transformação monotônica. Vimos na seção anterior que curvas de nível de funções côncavas limitam por baixo conjuntos convexos. Nesta seção, vimos que uma função côncava que sobre uma transformação monotônica se transforma em uma função quase côncava. Esses três pontos em conjunto indicam que curvas de nível de funções quase côncavas limitam por baixo conjuntos convexos, como descrito na primeira forma de definição de funções quase côncavas. Esse ponto é particularmente relevante em muitos problemas de microeconomia, em que podemos ter, por exemplo, funções de utilidade quase côncavas cujas curvas de indiferença limitam inferiormente conjuntos convexos.

BIBLIOGRAFIA

Banks, R. Growth and diffusion phenomena. Berlin: Springer-Verlag, 1994.

Chiang, A e Wainwright, K. Matemática para economistas, 5ª reimpressão, Rio de Janeiro: Elsevier, 2006.

Gandolfo, G. *Economic Dynamics*, Berlin: Springer-Verlag, 1997.

Golgher, A. e Martins, R. *Matemática: questões da ANPEC resolvidas 1993-2007*, Belo Horizonte: Editora da UFMG, 2008.

Kreps, D.A *Course in Microeconomic Theory*, Princeton: Princeton University press, 1990.

Lima, E. *Análise Real*, vol. 1, Rio de Janeiro: IMPA, 2004.

_____. *Curso de Análise*, vol. 1, Rio de Janeiro: IMPA, 1995.

Mas-Colell, A., Dennis, M. e Whinston, J. *Microeconomic Theory*, Oxford: Oxford University Press, 1995

Simon, C. e Blume, L. *Mathematics for Economists*. New York: NORTON, 1994.

Stewart, J. *Cálculo*, São Paulo: Thomson, 5ª edição, vol 1 e 2, 2006.