

Capítulo 3

Sistemas de equações lineares

3.1 Breve referência histórica

No século I da era cristã, foi publicado na China um livro intitulado *Jiuzhang Suanshu* (Os Nove Capítulos da Arte Matemática). Pela influência que causou em toda a matemática oriental, essa obra é muitas vezes comparada aos *Elementos* de Euclides. Os 'Nove Capítulos' são constituídos por 246 problemas de aritmética e geometria mas a sua referência neste capítulo, dedicado à resolução de sistemas lineares, tem a ver com o facto de aí ter sido descrita uma forma sistemática de resolver sistemas lineares com coeficientes positivos. Como curiosidade refira-se que as operações eram efectuadas com o auxílio de pequenos paus dispostos numa folha de papel. Através de manipulações sobre esses paus a técnica proposta era em todo semelhante ao método da decomposição de Gauss (apresentado, somente, no século XIX!). As operações eram efectuadas sobre os coeficientes do sistema o que confere à técnica um estatuto de uma proto-álgebra linear. De notar que os chineses, já desde essa altura, usavam um sistema de numeração de posição, com recurso ao uso de um quadrado em branco para representar o zero.

No século XIII o matemático chinês Zhu Shijie publicou uma obra intitulada *Suanxue Quimeng* (Introdução à Ciência do Cálculo) onde aprefeioou o método de resolução de sistemas lineares proposto nos 'Nove Capítulos', não se libertando, contudo, do recurso aos pauzinhos para efectuar as contas. Este livro teve também o mérito de ter sido estudado, muito mais tarde, pelo matemático japonês Seki Takakazu (1642?-1708). Inspirado por esta obra, Takakazu generalizou a álgebra chinesa libertando-a do recurso aos paus. O cálculo proposto por este ilustre matemático (na obra *Kaikendai no Hô*) não restringe o número de incógnitas e estabelece regras gerais, em vez de resolver casos particulares. Outra contribuição importante de Takakazu foi a introdução da noção de determinante no seu livro *Kaifukudai no Hô*.

A noção de determinante, cuja teoria foi estudada de forma sistemática por Charles Jacobi (1804-1851), precedeu a de matriz que só foi considerada como um ente matemático por William Rowan Hamilton (1805-1856) no seu livro *Lectures on Quaternions*. No entanto, a resolução de sistemas de equações lineares já tinha sido considerada por vários autores no ocidente, sendo de destacar o contributo de Carl Friedrich Gauss (1777-1875). Na sua obra *Theoria Motus* (1809) Gauss apresentou uma técnica de resolução de sistemas lineares (surgidos no contexto de um problema de mínimos quadrados) que não é mais do que o método de eliminação que todos conhecemos e que hoje tem o seu nome. Mais tarde, o mesmo

Gauss apresentou um processo iterativo para resolver sistemas lineares de grande dimensão, antecipando o procedimento conhecido por método de Gauss-Seidel. Além dos contributos de Gauss, são de salientar os trabalhos de Joseph Louis Lagrange (1736-1813), Pierre Simon Laplace (1749-1827) e Charles Jacobi (1804-1851). Refira-se que Jacobi, baseado nos trabalhos de Gauss sobre mínimos quadrados, efectuou vários trabalhos sobre sistemas lineares tendo também influenciado um seu aluno alemão Ludwig Seidel. Estes dois matemáticos resolveram vários problemas usando métodos iterativos que foram baptizados com os seus nomes.

A teoria de matrizes foi posteriormente aperfeiçoada por Arthur Cayley (1821-1895) e James Joseph Sylvester (1814-1899) (a quem se deve a introdução do termo 'matriz'), tendo progredido de forma espectacular até ao início do século XX. Foi para a teoria de matrizes que o físico Heisenberg apelou, em 1925, quando fundou a mecânica dos quanta.

3.2 Introdução

O problema que pretendemos resolver neste capítulo consiste em determinar o valor de $x \in \mathbb{R}^n$ tal que $Ax = b$, sendo $b \in \mathbb{R}^n$ e $A \in \mathcal{M}_n(\mathbb{R})$, onde $\mathcal{M}_n(\mathbb{R})$ denota o conjunto das matrizes reais de ordem n . Para resolver este problema iremos supor que a matriz A é invertível ou, o que é equivalente, que o sistema é possível e determinado.

Há muitos sistemas físicos que podem ter como modelo sistemas de equações lineares. Suponhamos, por exemplo, um camião a atravessar uma ponte cuja estrutura é constituída por barras de ferro. O peso camião e da ponte são forças que são contrabalançadas pelas exercidas nas extremidades que seguram a ponte. Essas forças são propagadas ao longo de toda a estrutura e, em cada nodo (locais onde as barras de ferro seguram a estrutura) a resultante das forças deve ser nula. Se decomposermos as forças nas componentes horizontal (x) e vertical (y) temos, em cada nodo ($i = 1, 2, \dots$), as equações:

$$\begin{aligned} \text{soma } x \text{ das forças} &= 0, \text{ no nodo } i, \\ \text{soma } y \text{ das forças} &= 0, \text{ no nodo } i. \end{aligned}$$

As forças, em cada barra, podem assim ser determinadas. Como há forças conhecidas (peso do camião, peso das barras, etc), o sistema a resolver é não homogéneo.

A resolução de um problema envolvendo sistemas lineares pode dividir-se em três etapas:

1. formulação do modelo matemático (calcular a matriz A);
2. cálculo dos agentes exteriores (calcular o vector b);
3. resolução do sistema linear.

Os dois primeiros passos dependem, obviamente, do conhecimento do problema físico (tipo de material, leis físicas, etc); o terceiro passo pode ser equacionado e resolvido separadamente, usando um conveniente algoritmo matemático. Uma vez que este último passo aparece como pertencente a um algoritmo bastante mais vasto é essencial que seja calculado de forma eficiente.

Existem duas grandes classes de métodos para resolver sistemas de equações lineares: os métodos directos, que já foram estudados, em parte, na disciplina de Álgebra Linear e para os quais iremos fazer uma breve revisão; e os métodos iterativos que iremos estudar com mais pormenor neste capítulo. Antes porém, vamos apresentar algumas classes de matrizes que irão ser consideradas.

3.3 Classes de matrizes

Existem vários tipos de matrizes que poderão aparecer na prática e que iremos analisar apenas algumas delas.

- Matrizes densas e matrizes esparsas

Um matriz com muitos elementos nulos diz-se **esparsa**; caso contrário diz-se que a matriz é **densa**. Sistemas com matrizes esparsas modelam sobretudo problemas onde existem princípios de influência local. Note-se que, no caso da ponte, as equações em cada nodo apenas envolvem as barras que aí se encontram. O seu número é o mesmo quer a ponte tenha 50 metros e, digamos, 10 barras, ou 5 km e 1000 barras. Assim, para uma ponte grande, a maioria dos coeficientes da matriz são nulos.

Como caso particular das matrizes esparsas temos as matrizes **banda** e, dentro dessa classe, as chamadas matrizes **tridiagonais**. As matrizes banda são aquelas cujos elementos não nulos se concentram apenas num conjunto de diagonais paralelas à diagonal principal.

- Matrizes simétricas

As matrizes simétricas são aquelas que coincidem com a sua transposta.

- Matrizes estritamente diagonal dominantes (EDD)

Uma matriz $A = (a_{ij})_{i,j=1}^n$ diz-se **estritamente diagonal dominante por linhas** se

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|$$

e **estritamente diagonal dominante por colunas** se

$$|a_{jj}| > \sum_{i=1, i \neq j}^n |a_{ij}|.$$

A matriz diz-se **estritamente diagonal dominante** se for estritamente diagonal dominante por linhas ou estritamente diagonal dominante por colunas.

3.4 Métodos directos: revisão

Consideremos, de novo, o problema de determinar o vector $x \in \mathbb{R}^n$ tal que $Ax = b$, sendo $b \in \mathbb{R}^n$ e $A \in \mathcal{M}_n(\mathbb{R})$, onde $\mathcal{M}_n(\mathbb{R})$. O primeiro tipo de métodos que iremos considerar resolver este problema são os chamados **métodos directos**. Estes métodos são aqueles que, supondo não haver erros de arredondamento ou quaisquer outros, nos permitem obter a solução exacta do problema num número finito de operações aritméticas.

Os métodos directos baseiam-se no processo de eliminação de Gauss que consiste em transformar o sistema $Ax = b$ num sistema equivalente $Ux = c$, onde U é uma matriz triangular superior, através de **operações elementares** efectuadas na **matriz ampliada** (lembrar

estes conceitos dados na disciplina de Álgebra Linear). O sistema a resolver pode ser escrito na forma

$$\begin{cases} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1n}x_n = c_1 \\ \phantom{u_{11}x_1} + u_{22}x_2 + \cdots + u_{2n}x_n = c_2 \\ \phantom{u_{11}x_1} \phantom{+ u_{22}x_2} + \cdots + \phantom{u_{2n}x_n} = c_3 \\ \phantom{u_{11}x_1} \phantom{+ u_{22}x_2} + u_{nn}x_n = c_n \end{cases}$$

e a sua resolução, caso $u_{ii} \neq 0$, $i = 1, \dots, n$, é feita de acordo com o algoritmo seguinte.

Algoritmo 3.1 Resolução de um sistema triangular superior

Ler c_i , $i = 1, \dots, n$;

Ler u_{ij} , $i = 1, \dots, n$, $j = i, \dots, n$;

$x_n := c_n/u_{nn}$;

Para i de $n - 1$ até 1 fazer

$$x_i := (b_i - \sum_{j=i+1}^n u_{ij}x_j) / u_{ii};$$

Escrever x_i , $i = 1, \dots, n$;

O método da eliminação de Gauss tem como desvantagem a alteração do valor dos termos independentes ($c \neq b$). Para contornar esse problema, temos o chamado método da triangularização. Este método consiste em decompor a matriz A do sistema a resolver na forma

$$A = LU,$$

em que L é uma matriz triangular inferior e U uma matriz triangular superior, com $u_{ii} = 1$, $i = 1, \dots, n$. A forma de obter esta decomposição foi igualmente vista na disciplina de Álgebra Linear. Após obtida a decomposição, a resolução do sistema é feita em duas etapas:

1. resolver $Ly = b$;
2. resolver $Ux = y$.

Notemos que em cada etapa temos que resolver um sistema triangular.

Exercício 3.4.1 Supondo determinada a decomposição $A = LU$, obtenha o algoritmo que permita resolver o sistema $Ax = b$ pelo método da triangulação.

Em muitas situações práticas, o sistema linear a resolver é **tridiagonal**, isto é, a matriz A é da forma (suprimindo os zeros)

$$A = \begin{bmatrix} \beta_1 & \gamma_1 & & & \\ \alpha_2 & \beta_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \alpha_{n-1} & \beta_{n-1} & \gamma_{n-1} \\ & & & \alpha_n & \beta_n \end{bmatrix}$$

ou, noutra notação

$$A = \text{Tridiag}(\alpha, \beta, \gamma),$$

com $\alpha = [\alpha_2, \dots, \alpha_n]^T$, $\beta = [\beta_1, \dots, \beta_n]^T$ e $\gamma = [\gamma_1, \dots, \gamma_{n-1}]^T$.

É fácil de demonstrar (ver Álgebra Linear) que a decomposição $A = LU$ é dada pelas matrizes

$$L = \begin{bmatrix} l_1 & & & & & \\ \alpha_2 & l_2 & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & \alpha_{n-1} & l_{n-1} \\ & & & & & \alpha_n & l_n \end{bmatrix} \quad \text{e} \quad U = \begin{bmatrix} 1 & u_1 & & & & \\ & 1 & u_2 & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & 1 & u_{n-1} \\ & & & & & & 1 \end{bmatrix},$$

com os elementos l_i , $i = 1, \dots, n$ e u_i , $i = 1, \dots, n - 1$, dados de acordo com o seguinte algoritmo.

Algoritmo 3.2 Decomposição LU para matrizes tridiagonais

Ler α_i , $i = 2, \dots, n$;

Ler β_i , $i = 1, \dots, n$;

Ler γ_i , $i = 1, \dots, n - 1$;

$l_1 := \beta_1$;

Se $l_1 \neq 0$ então $u_1 := \gamma_1/l_1$ caso contrário parar;

Para i de 2 até $n - 1$ fazer

$l_i := \beta_i - \alpha_i u_{i-1}$;

Se $l_i \neq 0$ então $u_i := \gamma_i/l_i$ caso contrário parar;

$l_n := \beta_n - \alpha_n u_{n-1}$;

Escrever l_i , $i = 1, \dots, n$;

Escrever u_i , $i = 1, \dots, n - 1$.

Como facilmente se pode demonstrar, a resolução do sistema linear $Ax = b$, com A uma matriz tridiagonal cuja decomposição $A = LU$ é dada pelo exercício anterior, pode ser efectuada de acordo com o seguinte algoritmo.

Algoritmo 3.3 Resolução de um sistema tridiagonal

Ler a matriz A e o vector b ;

Determinar as matrizes L e U ;

(* Resolver $Ly = b$ *)

$y_1 := b_1/l_1$;

Para i de 2 até n fazer

$y_i := (b_i - \alpha_i y_{i-1})/l_i$;

(* Resolver $Ux = y$ *)

$x_n := y_n$;

Para i de $n - 1$ até 1 fazer

$x_i := y_i - u_i x_{i+1}$;

Escrever x_i , $i = 1, \dots, n$.

3.5 Métodos iterativos

Consideremos, de novo, o problema de determinar o valor de $x \in \mathbb{R}^n$ tal que $Ax = b$, sendo $b \in \mathbb{R}^n$ e $A \in \mathcal{M}_n(\mathbb{R})$ uma matriz invertível. Um método iterativo para resolver o sistema consiste em, partindo de uma aproximação inicial $x^{(0)}$ para a solução do sistema (que iremos denotar por x^*), gerar uma sucessão de vectores $\{x^{(k)}\}$ convergente para x^* .

Os métodos que iremos considerar pertencem à classe dos métodos do ponto fixo e são obtidos transformando o problema $Ax = b$ num outro, equivalente, de forma

$$x = Mx + d,$$

para uma determinada matriz de iteração M e um determinado vector d .

Observação 3.1 Para determinar M e d podemos, por exemplo, decompor a matriz A na forma $A = N - P$, com N uma matriz invertível (mais simples que A) e considerar $M = N^{-1}P$ e $b = N^{-1}b$. De facto, $Ax = b \Leftrightarrow Nx = Px + b \Leftrightarrow x = N^{-1}Px + N^{-1}b$.

Com esta transformação podemos escrever o método iterativo na forma

$$\begin{aligned} x^{(0)} & \text{ dado} \\ x^{(k+1)} & = Mx^{(k)} + d, \quad k = 0, 1, \dots \end{aligned} \quad (3.1)$$

O processo iterativo termina quando se cumprirem os critérios de paragem estabelecidos. Os critérios mais comuns são:

1. Critério do erro absoluto: $\|x^{(k)} - x^{(k-1)}\| \leq \varepsilon$;
2. Critério do erro relativo: $\|x^{(k)} - x^{(k-1)}\| \leq \varepsilon \|x^{(k)}\|$;
3. Critério do número máximo de iterações: $k = kmax$.

Os métodos iterativos são sobretudo usados para matrizes esparsas de grandes dimensões, que surgem frequentemente em problemas de análise de circuitos ou cálculos de estruturas. Para esses problemas, os métodos iterativos são competitivos face aos métodos directos. Para matrizes densas ou de pequena dimensão os métodos directos são mais vantajosos.

Exemplo 3.2 Consideremos, por exemplo, o seguinte sistema

$$Ax = b \Leftrightarrow \begin{cases} 10x_1 - 2x_2 + 5x_3 = 13 \\ -x_1 + 3x_2 - x_3 = 1 \\ x_1 + 4x_2 + 2x_3 = 7 \end{cases}$$

que tem solução única $(x_1, x_2, x_3) = (1, 1, 1)$. Para converter o sistema na forma equivalente $x = Mx + d$ façamos

$$x = Mx + d \Leftrightarrow \begin{cases} x_1 = -\frac{1}{10}[-2x_2 + 5x_3 - 3] \\ x_2 = -\frac{1}{3}[-x_1 - x_3 - 1] \\ x_3 = -\frac{1}{2}[(x_1 + 4x_2 - 7)] \end{cases}.$$

Neste caso temos que a matriz M de iteração e o vector b são dados por

$$M = \begin{bmatrix} 0 & 2/10 & -5/10 \\ 1/3 & 0 & 1/3 \\ -1/2 & -4/2 & 0 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 13/10 \\ 1/3 \\ 7/2 \end{bmatrix}.$$

O método iterativo é dado na forma

$$x^{(k+1)} = Mx^{(k)} + d \Leftrightarrow \begin{cases} x_1^{(k+1)} &= -\frac{1}{10} [-2x_2^{(k)} + 5x_3^{(k)} - 3] \\ x_2^{(k+1)} &= -\frac{1}{3} [-x_1^{(k)} - x_3^{(k)} - 1] \\ x_3^{(k+1)} &= -\frac{1}{2} [x_1^{(k)} + 4x_2^{(k)} - 7] \end{cases}, \quad k = 0, 1, \dots,$$

sendo $x^{(0)} := [x_1^{(0)}, x_2^{(0)}, x_3^{(0)}]^T$ um valor dado.

O método descrito no exemplo anterior é conhecido por método de Jacobi e é um dos métodos que iremos estudar neste capítulo. Este método, quando aplicado ao sistema $Ax = b$, com $a_{ii} \neq 0$, $i = 1, \dots, n$, pode ser dado por

$$x_i^{(k+1)} = \frac{-1}{a_{ii}} \left[\sum_{j=1, j \neq i}^n a_{ij} x_j^{(k)} - b_i \right], \quad i = 1, \dots, n \quad (3.2)$$

Exercício 3.5.1 Escreva um algoritmo para determinar a solução aproximada de $Ax = b$ pelo método de Jacobi.

Um melhoramento ao método de Jacobi pode ser dado de acordo com o próximo exemplo.

Exemplo 3.3 Consideremos, de novo, o sistema linear dado no exemplo anterior e o processo iterativo:

$$\begin{cases} x_1^{(k+1)} &= -\frac{1}{10} [-2x_2^{(k)} + 5x_3^{(k)} - 3] \\ x_2^{(k+1)} &= -\frac{1}{3} [-x_1^{(k+1)} - x_3^{(k)} - 1] \\ x_3^{(k+1)} &= -\frac{1}{2} [x_1^{(k+1)} + 4x_2^{(k+1)} - 7] \end{cases}, \quad k = 0, 1, \dots,$$

sendo $x^{(0)} := [x_1^{(0)}, x_2^{(0)}, x_3^{(0)}]^T$ um valor dado. Como se pode ver, este método usa as componentes da nova aproximação logo após estas terem sido calculadas. Neste caso a matriz de iteração do método já não é tão simples de escrever.

O método descrito no exemplo anterior é chamado método de Gauss-Seidel. Quando aplicado à resolução numérica de $Ax = b$, com $a_{ii} \neq 0$, $i = 1, \dots, n$, o método é dado por

$$x_i^{(k+1)} = \frac{-1}{a_{ii}} \left[\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} + \sum_{j=i+1}^n a_{ij} x_j^{(k)} - b_i \right], \quad i = 1, \dots, n \quad (3.3)$$

A questão que se coloca é a de saber em que circunstâncias estes métodos são convergentes. Para estudar a convergência dos métodos iterativos para sistemas de equações lineares, há necessidade de introduzir o conceito de norma de matriz.

3.5.1 Normas de matrizes

Seja $\mathcal{M}_n(\mathbb{R})$ o conjunto das matrizes quadradas de ordem n com coeficientes reais. Como este conjunto é um espaço vectorial podemos nele definir uma norma.

Definição 3.4 *Seja $\|\cdot\|_v$ uma norma definida em \mathbb{R}^n . A aplicação $\|\cdot\| : \mathcal{M}_n(\mathbb{R}) \rightarrow \mathbb{R}_0^+$ tal que, para todo o $A \in \mathcal{M}_n(\mathbb{R})$,*

$$\|A\| = \sup_{\|x\| \neq 0} \frac{\|Ax\|_v}{\|x\|_v}$$

é designada norma matricial subordinada à norma vectorial $\|\cdot\|_v$.

Exercício 3.5.2 Mostre que a função estabelecida na definição anterior é uma norma, isto é, que verifica as seguintes propriedades:

1. $\forall A \in \mathcal{M}_n(\mathbb{R}), \quad \|A\| = 0 \Leftrightarrow A = 0$ (matriz nula),
2. $\forall A \in \mathcal{M}_n(\mathbb{R}), \forall \lambda \in \mathbb{R}, \quad \|\lambda A\| = |\lambda| \|A\|$,
3. $\forall A, B \in \mathcal{M}_n(\mathbb{R}), \quad \|A + B\| \leq \|A\| + \|B\|$,

Podemos definir as normas matriciais subordinadas às normas vectoriais dadas no capítulo anterior.

Exercício 3.5.3 (Normas matriciais) Seja $A \in \mathcal{M}_n(\mathbb{R})$. Prove que as seguintes aplicações verificam as propriedades que caracterizam um norma:

- $\|A\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|$;
- $\|A\|_\infty = \max_{i=1, \dots, n} \sum_{j=1}^n |a_{ij}|$.

Exercício 3.5.4 Prove que as normas matriciais $\|\cdot\|_1$ e $\|\cdot\|_\infty$ são normas subordinadas às normas vectoriais $\|\cdot\|_1$ e $\|\cdot\|_\infty$, respectivamente.

Resolução: Consideremos a norma $\|\cdot\|_1$ para \mathbb{R}^n . Então, com $A \in \mathcal{M}_n(\mathbb{R})$, tem-se

$$\|A\|_1 = \sup_{\|x\| \neq 0} \frac{\|Ax\|_1}{\|x\|_1}$$

e

$$\|Ax\|_1 = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij} x_j \right| \leq \sum_{j=1}^n \sum_{i=1}^n |a_{ij}| |x_j| \leq \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}| \|x\|_1.$$

Concluimos, portanto, que

$$\|A\|_1 \leq \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|.$$

Temos então que (o máximo é atingido para um vector $x \in \mathbb{R}^n$ escolhido de forma conveniente)

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^n |a_{ij}|,$$

o que mostra que a norma anterior é subordinada à norma $\|\cdot\|_1$.

Consideremos agora em \mathbb{R}^n a norma $\|\cdot\|_\infty$ e seja $A \in \mathcal{M}_n(\mathbb{R})$. Então

$$\|A\|_\infty = \sup_{\|x\| \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}.$$

Atendendo a que

$$\|Ax\|_\infty = \max_{i=1,\dots,n} \left| \sum_{j=1}^n a_{ij}x_j \right| \leq \left(\max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}| \right) \max_{j=1,\dots,n} |x_j|$$

vem

$$\|A\|_\infty \leq \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|.$$

Atendendo a que o máximo é atingido para um vector $x \in \mathbb{R}^n$ escolhido de forma conveniente concluímos que a norma $\|A\|_\infty$ é subordinada à norma vectorial $\|\cdot\|_\infty$.

Provemos agora alguns resultados importantes referentes a normas matriciais.

Teorema 3.5 *Seja $\|\cdot\|$ uma norma em \mathbb{R}^n . Então, para $A \in \mathcal{M}_n(\mathbb{R})$, tem-se*

$$\|Ax\| \leq \|A\|\|x\|, \forall x \in \mathbb{R}^n,$$

em que $\|A\|$ é a norma de A subordinada à norma $\|\cdot\|$.

Demonstração: Notamos que se a norma $\|\cdot\|$ é subordinada a uma norma vectorial então, para $A \in \mathcal{M}_n(\mathbb{R})$, temos

$$\|A\| = \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} \geq \frac{\|Ax\|}{\|x\|}, \quad x \in \mathbb{R}^n, x \neq 0,$$

e portanto é válido o resultado. \square

Teorema 3.6 *Para $A, B \in \mathcal{M}_n(\mathbb{R})$, tem-se $\|AB\| \leq \|A\|\|B\|$.*

Demonstração: Atendendo à desigualdade demonstrada no teorema anterior, temos

$$\|AB\| = \sup_{\|x\| \neq 0} \frac{\|ABx\|}{\|x\|} \leq \|A\| \sup_{\|x\| \neq 0} \frac{\|Bx\|}{\|x\|} = \|A\|\|B\|,$$

o que prova o pretendido. \square

3.5.2 Convergência dos métodos iterativos

Vamos agora abordar a questão da convergência do método iterativo (3.1).

Definição 3.7 *Seja x^* a solução (única) de $Ax = b$ e $\{x^{(k)}\}$ uma sucessão de aproximações obtida pelo método (3.1). O método diz-se convergente se*

$$\lim_{k \rightarrow +\infty} e^{(k)} := \lim_{k \rightarrow +\infty} (x^* - x^{(k)}) = 0.$$

Caso contrário o método diz-se divergente. A $e^{(k)}$ chama-se erro (absoluto) da iteração k .

Notemos o seguinte resultado cuja demonstração é muito simples.

Teorema 3.8 *O método iterativo (3.1) converge, qualquer que seja a aproximação inicial $x^{(0)}$, se e só se*

$$\lim_{k \rightarrow +\infty} M^k = 0.$$

Demonstração: De facto

$$e^{(k)} = x^* - x^{(k)} = M e^{(k-1)}, \quad k = 1, 2, \dots$$

Assim

$$e^{(k)} = M^k e^{(0)}, \quad k = 1, 2, \dots, \quad (3.4)$$

e, como tal,

$$\lim_{k \rightarrow +\infty} e^{(k)} = 0 \Leftrightarrow \lim_{k \rightarrow +\infty} M^k = 0,$$

o que prova o pretendido. \square

Outro aspecto importante a determinar quando se lida com métodos iterativos tem a ver com a determinação de majorantes para o erro cometido. O teorema seguinte é, nesse sentido, muito importante. Além disso, estabelece uma condição suficiente de convergência mais útil que a referida no teorema anterior.

Teorema 3.9 *Se $\|M\| < 1$ então o método iterativo (3.1) converge, qualquer que seja a aproximação inicial escolhida, e tem-se que*

1. $\|e^{(k)}\| \leq \|M\|^k \|e^{(0)}\|;$
2. $\|e^{(k)}\| \leq \frac{\|M\|^k}{1 - \|M\|} \|x^{(1)} - x^{(0)}\|.$

Demonstração: Vamos demonstrar apenas o primeiro ponto. A demonstração da segunda parte segue um procedimento análogo ao efectuado na demonstração do Corolário 2.27 e fica ao cuidado dos alunos.

Demonstremos assim a primeira parte (a mais simples). Considerando normas em (3.4) obtemos

$$\|e^{(k)}\| \leq \|M\|^k \|e^{(0)}\|, \quad k = 1, 2, \dots$$

Tomando limites e atendendo a que $\|M\| < 1$ concluímos que o método é convergente. \square

O resultado anterior dá-nos apenas uma condição suficiente de convergência. Para estabelecer uma condição necessária e suficiente de convergência temos que usar a noção de raio espectral de uma matriz. Recordemos que o raio espectral de uma matriz $A \in \mathcal{M}_n(\mathbb{R})$, que denotaremos por $\rho(A)$, é definido por

$$\rho(A) = \max_{i=1, \dots, n} \{|\lambda_i| : \lambda_i \text{ é valor próprio de } A\}.$$

Diz-se que $\lambda \in \mathbb{C}$ é um valor próprio de A se existe um $z \in \mathbb{R}$, com $z \neq 0$, tal que $Az = \lambda z$. A z chamamos o vector próprio associado ao valor próprio λ .

Exercício 3.5.5 Seja A uma matriz real, não singular e de ordem n . Prove que se λ é um valor próprio de A então

$$\frac{1}{\|A^{-1}\|} \leq |\lambda| \leq \|A\|.$$

Estamos agora em condições de estabelecer o seguinte teorema.

Teorema 3.10 *O método iterativo (3.1) converge, qualquer que seja a aproximação inicial $x^{(0)}$ escolhida, se e só se $\rho(M) < 1$.*

Demonstração: Se $\rho(M) < 1$ pode demonstrar-se (ver Kress (1998)) que existe uma norma matricial tal que $\|M\| < 1$. Logo, pelo teorema anterior, o método converge.

Falta provar que se o método convergir então o raio espectral da matriz de iteração é menor que um. Vamos demonstrar este facto provando que se o raio espectral de M for maior ou igual a um podemos definir um processo iterativo de forma (3.1) divergente. De facto, se $\rho(M) \geq 1$ existe um valor próprio λ de M tal que $|\lambda| \geq 1$. Seja z o vector próprio associado a esse valor próprio. Considerando, em (3.1), a aproximação inicial $x^{(0)} = z$ e $d = z$ temos

$$x^{(1)} = Mz + z = \lambda z + z = (1 + \lambda)z.$$

Repetindo o processo temos, sucessivamente,

$$x^{(2)} = Mx^{(1)} + z = (1 + \lambda + \lambda^2)z, \quad \dots, \quad x^{(k)} = Mx^{(k-1)} + z = \left(\sum_{j=0}^k \lambda^j \right) z.$$

Como $|\lambda| \geq 1$ concluímos que o método iterativo assim definido é divergente. \square

3.5.3 Métodos de Jacobi e Gauss-Seidel

Vamos agora estudar a convergência dos métodos de Jacobi e Gauss-Seidel. Seja $Ax = b$, com $A \in \mathcal{M}_n(\mathbb{R})$ uma matriz invertível. Considerando

$$A = N - P, \quad \text{com } N \text{ uma matriz invertível,} \quad (3.5)$$

podemos definir (como vimos) o método iterativo

$$\begin{aligned} x^{(0)} & \text{ dado} \\ x^{(k+1)} & = Mx^{(k)} + d, \quad k = 0, 1, \dots, \end{aligned} \quad (3.6)$$

com $M = N^{-1}P$ e $d = N^{-1}b$.

A escolha dos diferentes métodos iterativos depende da forma como se define a partição (3.5). Para definir os métodos de Jacobi e Gauss-Seidel, considera-se

$$A = D - L - U,$$

em que

$$D = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix}$$

e

$$L = \begin{bmatrix} 0 & & & & \\ -a_{21} & 0 & & & \\ \vdots & \ddots & \ddots & & \\ -a_{n1} & \cdots & -a_{n,n-1} & 0 & \end{bmatrix}, \quad U = \begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ & \ddots & \ddots & \vdots \\ & & 0 & -a_{n-1,n} \\ & & & 0 \end{bmatrix}.$$

Método de Jacobi Para definir o método de Jacobi, temos sucessivamente

$$Ax = b \Leftrightarrow (D - L - U)x = b \Leftrightarrow Dx = (L + U)x + d.$$

Caso D seja invertível, temos que

$$x = D^{-1}(L + U)x + D^{-1}b.$$

O método de Jacobi é assim dado por (3.6), com $M = D^{-1}(L + U)$ e $d = D^{-1}b$.

Método de Gauss-Seidel Para definir o método de Gauss-Seidel, consideramos

$$Ax = b \Leftrightarrow (D - L - U)x = b \Leftrightarrow (D - L)x = Ux + d.$$

Caso $D - L$ seja invertível ou, o que é equivalente, caso D seja invertível, temos que

$$x = (D - L)^{-1}Ux + (D - L)^{-1}b.$$

O método de Gauss-Seidel é assim dado por (3.6), com $M = (D - L)^{-1}U$ e $d = (D - L)^{-1}b$.

O estudo da convergência destes métodos iterativos pode ser efectuado de acordo com os resultados estabelecidos na secção anterior. Assim, se D for invertível, temos que, qualquer que seja a aproximação inicial escolhida:

- Método de Jacobi
 1. $\rho(D^{-1}(L + U)) < 1 \Leftrightarrow$ o método converge;
 2. $\|D^{-1}(L + U)\| < 1 \Rightarrow$ o método converge;

- Método de Gauss-Seidel

1. $\rho((D - L)^{-1}U) < 1 \Leftrightarrow$ o método converge;
2. $\|(D - L)^{-1}U\| < 1 \Rightarrow$ o método converge.

Para o caso particular dos métodos de Jacobi e Gauss-Seidel existe um resultado de convergência específico que pode ser útil na prática. Esse resultado é dado no próximo teorema, cuja demonstração irá ser feita apenas para o caso do método de Jacobi.

Teorema 3.11 *Se A é uma matriz estritamente diagonal dominante por linhas (ou por colunas), então os métodos de Jacobi e Gauss-Seidel convergem para a única solução do sistema $Ax = b$, qualquer que seja a aproximação inicial escolhida.*

Demonstração: Vamos efectuar a demonstração apenas para o caso do método de Jacobi. A matriz M de iteração do método é dada por

$$M = D^{-1}(L + U) = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \dots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & 0 \end{bmatrix}.$$

Como A é estritamente diagonal dominante por linhas temos que

$$\|M\|_{\infty} = \max_{i=1,\dots,n} \left\{ \frac{1}{|a_{ii}|} \sum_{j=1, j \neq i}^n |a_{ij}| \right\} < 1.$$

Assim, pelo Teorema 3.9 temos que o método de Jacobi converge para a solução de $Ax = b$, qualquer que seja a aproximação inicial escolhida. \square

Observação 3.12 *Note-se que o teorema anterior nos dá apenas uma condição suficiente de convergência. Assim, se a matriz do sistema a resolver for estritamente diagonal os métodos de Jacobi e Gauss-Seidel irão gerar sucessões de aproximações convergentes para a sua solução; caso contrário, nada poderemos afirmar quanto à convergência dessas sucessões.*

3.6 Condicionamento de matrizes

Consideremos um sistema possível e determinado $Ax = b$ e seja \bar{b} o vector obtido a partir de b considerando perturbações numéricas nas suas componentes. Esta situação é frequente quando o vector dos termos independentes representa medições. Sejam $e(\bar{b})$ e $r(\bar{b})$ respectivamente os erros absoluto e relativo de \bar{b} . Vejamos de que modo estes erros influenciam os erros absoluto e relativo de \bar{x} , sendo \bar{x} a solução do sistema $A\bar{x} = \bar{b}$.

Se o sistema $Ax = b$ é possível e determinado, então A é invertível e portanto $x = A^{-1}b$. Consideremos agora o sistema em que o vector dos termos independentes tem as componentes afectadas de erro, i.e, $A\bar{x} = \bar{b}$. Temos

$$r(x) = \frac{\|e(x)\|}{\|x\|} = \frac{\|A^{-1}(b - \bar{b})\|}{\|x\|} = \frac{\|A^{-1}e(b)\|}{\|x\|}.$$

Como

$$\|b\| = \|Ax\| \leq \|A\|\|x\|$$

concluimos que

$$\|x\| = \|A\|^{-1}\|b\|.$$

Utilizando esta desigualdade em $r(x)$, deduzimos

$$\|r(x)\| \leq \frac{\|A^{-1}\|\|e(b)\|}{\|A\|^{-1}\|b\|} = \|A^{-1}\|\|A\|\|r(b)\|.$$

A $\|A^{-1}\|\|A\|$ chamamos número de condição da matriz A e é denotado por $\text{cond}(A)$.

Do exposto podemos afirmar que se o número de condição de A for pequeno (de preferência menor ou igual a um) pequenas perturbações no vector dos termos independentes conduzem a pequenas perturbações no vector solução. Neste caso dizemos que o sistema $Ax = b$ é estável e que a matriz do A é bem condicionada. Se o número de condição for muito grande o sistema $Ax = b$ diz-se instável e a matriz do A mal condicionada.

Quando se pretende resolver sistemas lineares instáveis usando métodos directos, a solução obtida vem, frequentemente, afectada de erro. Nessa medida, é usual considerar métodos mistos, isto é, métodos iterativos que consideram como aproximação inicial a solução obtida pelo método directo. O método iterativo funciona assim como um corrector do resultado obtido pelo método directo.

3.7 Exercícios de aplicação à engenharia

Exercício 3.7.1 Considere um barra horizontal fixa numa extremidade e livre no restante do seu comprimento. Um modelo discreto de forças na barra conduz ao sistema de equações lineares $Ax = b$, onde A é a matriz quadrada de ordem n , com estrutura de banda, dada por

$$\begin{bmatrix} 9 & -4 & 1 & 0 & \cdots & \cdots & 0 \\ -4 & 6 & -4 & 1 & \ddots & & \vdots \\ 1 & -4 & 6 & -4 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 & -4 & 5 & -2 \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}.$$

O vector b é dado pela carga que é imposta à barra (incluindo o seu próprio peso), e o vector x representa a deformação da barra que queremos determinar. Consideremos a barra sujeita a uma carga uniforme dada por $b_i = 1$, para $i = 1, \dots, n$. Considerando $n = 100$, resolva o sistema usando métodos directos e métodos iterativos, comparando a sua eficácia.

Exercício 3.7.2 Num artigo intitulado 'Population Waves', Journal of the Burma Research Society 31 (1941), 1-18, DS527.B85387, H. Bernadelli (o seu autor) apresenta um estudo sobre um hipotético tipo de escaravelho que tem um tempo de vida natural de 3 anos. A fêmea desta espécie possui uma taxa de sobrevivência de $\frac{1}{2}$ no primeiro ano de vida, uma taxa de $\frac{1}{3}$ do segundo para o terceiro ano, e dá à luz uma média de seis novas fêmeas antes de morrer no final do terceiro

ano. Para mostrar a contribuição de cada fêmea, num sentido probabilístico, para a população total de fêmeas da espécie poderemos usar uma matriz. Denotemos por a_{ij} , elemento da matriz $A = (a_{ij})_{i,j=1}^3$, a contribuição que uma determinada fêmea de escarvelho de idade j dá para população de fêmeas do ano seguinte de idade i . Consideremos essa matriz dada por

$$A = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

1. Qual a contribuição que uma fêmea dá para a população do ano n (inteiro positivo) após o seu nascimento?
2. Use as conclusões obtidas no ponto anterior para determinar a contribuição de uma determinada fêmea num ano futuro, supondo uma população inicial de 6000 fêmeas de escarvelho em cada um das três faixas etária.
3. Construa a matriz A^{-1} e diga qual o seu significado no que diz respeito à população desta espécie.

Exercício 3.7.3 Suponhamos que num sistema biológicos existem n espécies de animais e m tipos de alimento. Nesse sistema, seja x_j o número de animais da espécie j , para $j = 1, \dots, n$; b_i representa a quantidade diária de comida disponível do tipo i , $i = 1, \dots, m$; e a_{ij} a quantidade de alimento i consumido (em média) por um animal do tipo j . O sistema linear

$$Ax = b \Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

representa a situação de equilíbrio, isto é, a situação onde existe uma quantidade diária de comida disponível exactamente igual ao consumo médio diário de cada espécie.

1. Considerando

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

$x = [1000, 500, 350, 400]^T$, e $b = [3500, 2700, 900]^T$, haverá comida suficiente para satisfazer o consumo médio diário das espécies de animais?

2. Qual o número máximo de animais de cada espécie que pode ser individualmente acrescentada ao sistema de modo a que as reservas de comida ainda sejam suficientes para o consumo?
3. Se a espécie 1 se extinguisse, qual o acréscimo individual das restantes espécies suportado pelo sistema?
4. Se a espécie 2 se extinguisse, qual o acréscimo individual das restantes espécies suportado pelo sistema?

Exercício 3.7.4 O estudo das cadeias alimentares é um tópico importante na determinação da propagação e acumulação de poluentes ambientais na matéria viva. Suponhamos que a cadeia alimentar possui três elos. O primeiro elo é constituído pelos vegetais que podem ser dos tipos v_1, v_2, \dots, v_n ; o segundo elo é constituído pelos herbívoros h_1, h_2, \dots, h_m , que se alimentam à custa dos vegetais; o terceiro elo é constituído pelos carnívoros c_1, c_2, \dots, c_k , que dependem inteiramente dos herbívoros para a sua alimentação. As coordenadas a_{ij} da matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

representam o número de total de plantas do tipo v_i comidas pelos herbívoros da espécie h_j , enquanto o elemento b_{ij} da matriz

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{bmatrix}$$

descreve o número de herbívoros da espécie h_i que foram devorados pelos carnívoros do tipo c_j .

1. Mostre que o número de plantas do tipo v_i que eventualmente vão terminar no corpo do carnívoro c_j é dado pelo elemento que figura na linha i e na coluna j da matriz AB .
2. Qual o significado físico associado às matrizes A^{-1} , B^{-1} , e $(AB)^{-1} = B^{-1}A^{-1}$?

3.8 Referências bibliográficas

- K.E. Atkinson (1989), *An Introduction to Numerical Analysis*, 2th ed., John Wiley, New York.
- R.L. Burden e J.D. Faires (1988), *Numerical Analysis*, 4th ed., PWS-Kent, Boston.
- S.D. Conte e C. de Boor (1980), *Elementary Numerical Analysis*, 3th ed., McGraw-Hill, New York.
- I. Grattan-Guinness (1997), *Mathematical Sciences*, Fontana Press, London.
- J.A. Ferreira e M.F. Patrício (1999), *Análise Numérica, Textos de Apoio*, DMUC, Coimbra.
- R. Kress (1998), *Numerical Analysis*, Springer-Verlag, New York.
- M. Rosa (1992), *Tópicos de Análise Numérica, Textos de Apoio*, DMUC, Coimbra.
- J. Stoer e R. Bulirsch (1980), *Introduction to Numerical Analysis*, Springer-Verlag, Berlin.