OPENCOURSEWARE

BEKG 2452
NUMERICAL METHODS
Errors

Ser Lee Loh[a], Irma Wani binti Jamaludin[a]

Fakulti Kejuruteraan Elektrik

Universiti Teknikal Malaysia Melaka

Lesson Outcome

Upon completion of this lesson, the student should be able to:

1. Measure accuracy of approximations.
2. Compute the size of vectors and matrices.

1.1 Introduction to Numerical Methods

What Are Numerical Methods?

Methods of approximation

1.1 Introduction to Numerical Methods

Two main factors to be considered when applying numerical methods:

- **How** to obtain the approximations?
- **How accurate** are the approximations?

Error is the difference between the true value and the approximate value.

1.2 Floating Point Number System

- Fractional quantities are typically represented in computers using floating-point form.
- The decimal floating-point representation of a number x is given as

$$x = \pm m \times 10^{n}$$

where

m is the mantissa, $0.1 \le m < 1$

n is an integer

1.2 Floating Point Number System

A computer with a **four decimal digit floating point arithmetic** means that the number of **digits is limited to four after decimal point**. Hence, all the numbers are presented in the form

$$(\pm\, 0.\times\times\times\times)\, 10^{n}$$

Example:

1. $2 = (+0.2000)10^{1}$
2. $-0.004 = (-0.4000)10^{-2}$
3. $-\frac{2}{3} = (-0.6667)10^{0}$

1.3 Measuring Errors

Error = |true value – approximate value|

$$= |X_T - X_A|$$

Relative error = $\left|\frac{\text{error}}{\text{true value}}\right| = \left|\frac{X_T - X_A}{X_T}\right|$

True percent relative error = $\left|\frac{\text{true value} - \text{approximate value}}{\text{true value}}\right| \times 100\%$

Approximate percent relative error

= $\left|\frac{\text{present approximation} - \text{previous approximation}}{\text{present approximation}}\right| \times 100\%$

1.3 Measuring Errors

Example 1.1:

The number 3.14159 is approximated as 3.1416. Find the:

(a) Error

(b) Relative error

(c) True percent relative error

Solution:

(a) Error=$|X_T - X_A|$= 3.14159 – 3.14 = 0.00159

(b) Relative error $= \left|\frac{X_T - X_A}{X_T}\right| = \left|\frac{0.00159}{3.14159}\right|$=0.000506

(c) True percent relative error

$$=\left|\frac{X_T - X_A}{X_T}\right| \times 100 = \left|\frac{0.00159}{3.14159}\right| \text{x}100 = 0.0506\%$$

1.4 Norms

- A **norm** is a real-valued function that provides a strictly positive **size** or ‘**length**’ of vectors and matrices.

1.4.1 L_1 norm

Definition:

- The ***L_1* norm** on $\mathbb{R}^n$ is given by

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$

1.4.1 L_1 norm

Example 1.2:

Let $\mathbf{x} = (-1, 2, -3)^{\mathrm{T}}$. Find the L_1 norm of $\mathbf{x}$.

Solution:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{3} |x_i| = |-1| + |2| + |-3| = 6$$

1.4.1 L_1 norm

Example 1.3:

Let $\mathbf{x} = (-1, 2, -3, 4)^{\mathrm{T}}$. Find the L_1 norm of $\mathbf{x}$.

Solution:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{4} |x_i| = |-1| + |2| + |-3| + 4 = 10$$

1.4.2 L_2 norm (Euclidean norm)

Definition:

- The ***L_2* norm** on $\mathbb{R}^n$ is given by

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$

1.4.2 L_2 norm (Euclidean norm)

Example 1.4:

Let $\mathbf{x} = (-1, 2, -3)^{\mathrm{T}}$. Find the L_2 norm of $\mathbf{x}$.

Solution:

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{3} x_i^2\right)^{1/2} = (1 + 4 + 9)^{1/2} = \sqrt{14}$$

1.4.2 L_2 norm (Euclidean norm)

Example 1.5:

Let $\mathbf{x} = (-1, 2, -3, 4)^{\mathrm{T}}$. Find the L_2 norm of $\mathbf{x}$.

Solution:

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{4} x_i^2\right)^{1/2} = (1 + 4 + 9 + 16)^{1/2}$$

$$= \sqrt{30}$$

1.4.3 L_p norm (Holder L_p norm or p-norm)

Definition:

- The ***L_p* norm** on $\mathbb{R}^n$ is given by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$

1.4.3 L_p norm (Holder L_p norm or p-norm)

Example 1.6:

Let $\mathbf{x} = (-1, 2, -3)^{\mathrm{T}}$ and $p = 3$. Find the L_p norm of $\mathbf{x}$.

Solution:

$$\|\mathbf{x}\|_3 = \left(\sum_{i=1}^{3} |x_i|^3\right)^{1/3} = (1 + 8 + 27)^{1/3} = \sqrt[3]{36}$$

1.4.3 L_p norm (Holder L_p norm or p-norm)

Example 1.7:

Let $\mathbf{x} = (-1, 2, -3, 4)^{\mathrm{T}}$ and $p = 4$. Find the L_p norm of $\mathbf{x}$.

Solution:

$$\|\mathbf{x}\|_4 = \left(\sum_{i=1}^{4} |x_i|^4\right)^{1/4} = (1 + 16 + 81 + 256)^{1/4}$$

$$= \sqrt[4]{354}$$

1.4.4 1-norm matrix

Definition:

Let $A \in \mathbb{R}^{n\times m}$ be an $n\times m$ matrix and let i, j entry of a matrix $A \in \mathbb{R}^{n\times m}$ be denoted by a_{ij}. Then the **1-norm** of a matrix is the maximum of the column sums of the absolute values of the entries of the matrix.

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

1.4.4 1-norm matrix

Example 1.8:

Find the 1-norm of A given:

$$A = \begin{bmatrix} 1 & -2 & 3 \\ -6 & 5 & 4 \\ 7 & -8 & -9 \end{bmatrix}$$

Solution:

$$\|A\|_1 = \max\{1+6+7, 2+5+8, 3+4+9\}$$
$$= \max\{14, 15, 16\}$$
$$= 16$$

1.4.4 1-norm matrix

Example 1.9:

Find the 1-norm of $A = \begin{bmatrix} 2 & -1 & 1 \\ -5 & 4 & 1 \end{bmatrix}$.

Solution:

$$\|A\|_1 = \max\{\mathbf{2} + \mathbf{5}, \mathbf{1} + \mathbf{4}, \mathbf{1} + \mathbf{1}\}$$

$$= \max\{7,5,2\}$$

$$= 7$$

1.4.5 ∞-norm matrix

Definition:

Let $A \in \mathbb{R}^{n\times m}$ be a $n\times m$ matrix and let the i,j entry of a matrix $A \in \mathbb{R}^{n\times m}$ be denoted a_{ij}. Then the **∞-norm** of a matrix is the maximum of the row sums of the absolute values of the entries of the matrix.

$$||A||_\infty = \max_i \sum_{j=1}^{m} |a_{ij}|$$

1.4.5 ∞-norm matrix

Example 1.10:

Find the ∞-norm of A given

$$A = \begin{bmatrix} 1 & -2 & 3 \\ -6 & 5 & 4 \\ 7 & -8 & -9 \end{bmatrix}$$

Solution:

$$\|A\|_\infty = \max\{\mathbf{1} + \mathbf{2} + \mathbf{3}, \mathbf{6} + \mathbf{5} + \mathbf{4}, \mathbf{7} + \mathbf{8} + \mathbf{9}\}$$
$$= \max\{6, 15, 24\} = 24$$

1.4.5 ∞-norm matrix

Example 1.11:

Find the ∞-norm of $A = \begin{bmatrix} 2 & -1 & 1 \\ -5 & 4 & 1 \end{bmatrix}$.

Solution:

$$\begin{aligned} \|A\|_\infty &= \max\{\mathbf{2+1+1}, \mathbf{5+4+1}\} \\ &= \max\{4, 10\} \\ &= 10 \end{aligned}$$