Finding the Smallest Absolute Eigenvalue of the Square Matrix

Hnin Oo Lwin*

Abstract

The aim of this paper is to find the smallest absolute eigenvalue of the square matrix. Firstly LU Factorization and Doolittle Method are presented. Then the smallest absolute eigenvalue of the square matrix is found by using the Inverse Power Method. The example is illustrated for square matrix of order 3.

Keywords: LU Factorization, Doolittle Method, Inverse Power Method.

Introduction

When the smallest absolute eigenvalue of the square matrix A is distinct, its value can be found using a variation of the power method called the **inverse power method**. This involves finding the largest absolute eigenvalue of the inverse matrix A,(A^{-1}), which is the smallest absolute eigenvalue of matrix A .

LU Factorization

The useful factorization for square matrices is $A = LU,$ where L and U are lower and upper triangular matrices, respectively. The LU factorization method is solving systems of the linear algebraic equations.

Matrices can be factored into the product of two other matrices in an infinite number of ways. Thus,

$$A = BC, \text{ where A, B and C are matrices.} \quad (1)$$

When B and C are lower triangular and upper triangular matrices, respectively,(1) becomes

$$A = LU. \quad (2)$$

Specifying the diagonal elements of either L or U makes the factoring unique. The procedure based on unity elements on the major diagonal of L is called the **Doolittle method**. The procedure based on unity elements on the major diagonal of U is called the **Crout method**.

We consider the linear system, $A\mathbf{x} = \mathbf{b}.$ Let A be factored into the product LU, as illustrated in (2). The linear system becomes

$$LU\mathbf{x} = \mathbf{b}. \quad (3)$$

Multiplying (3) by L^{-1},

$$L^{-1}LU\mathbf{x} = L^{-1}\mathbf{b}$$

$$IU\mathbf{x} = L^{-1}\mathbf{b}$$

$$U\mathbf{x} = L^{-1}\mathbf{b}. \quad (4)$$

* Dr, Professor, Department of Mathematics, Banmaw University

We define the vector $\mathbf{b}'$ as follows:

$$\mathbf{b}' = \mathrm{L}^{-1}\mathbf{b}. \tag{5}$$

Multiplying (5) by L,

$$\mathrm{L}\mathbf{b}' = \mathrm{L}\mathrm{L}^{-1}\mathbf{b} = \mathrm{I}\mathbf{b}$$

$$\mathrm{L}\mathbf{b}' = \mathbf{b}. \tag{6}$$

Substituting (5) into (4) yields

$$\mathrm{U}\mathbf{x} = \mathbf{b}'. \tag{7}$$

Equation (6) is used to transform the **b** vector into the $\mathbf{b}'$ vector, and (7) is used to determine the solution vector **x**. Since (6) is lower triangular, forward substitution is used to solve for $\mathbf{b}'$. Since (7) is upper triangular, back substitution is used to solve for **x**.

Matrix Inverse by the Doolittle Method

As an application of LU factorization, it can be used to evaluate the inverse of matrix A, that is, A^{-1}. The matrix inverse is calculated in a column by column manner using unit vectors for the right-hand-side vector **b.** Thus, if $\mathbf{b}_1^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$, $\mathbf{x}_1$ will be the first column of A^{-1}. The succeeding columns of A^{-1} are calculated by letting $\mathbf{b}_2^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}$, $\mathbf{b}_3^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \end{bmatrix}$, and so on, and $\mathbf{b}_{\mathrm{n}}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}$.

Inverse Power Method

We recall the linear eigenproblem:

$$\mathrm{A}\mathbf{x} = \lambda\mathbf{x}.$$

Multiplying $\mathrm{A}\mathbf{x} = \lambda\mathbf{x}$ by A^{-1} gives

$$\mathrm{A}^{-1}\mathrm{A}\mathbf{x} = \mathrm{I}\mathbf{x} = \mathbf{x} = \lambda\mathrm{A}^{-1}\mathbf{x}. \tag{8}$$

Rearranging (8) yields an eigenproblem for A^{-1}. Thus

$$\mathrm{A}^{-1}\mathbf{x} = \left(\frac{1}{\lambda}\right)\mathbf{x} = \lambda_{\text{inverse}}\,\mathbf{x}. \tag{9}$$

The eigenvalues of matrix A^{-1}, that is, λ_{inverse}, are the reciprocals of the eigenvalues of matrix A. The eigenvectors of matrix A^{-1} are the same as the eigenvectors of matrix A. The power method can be used to find the largest absolute eigenvalue of matrix A^{-1}, λ_{inverse}. The reciprocal of that eigenvalue is the smallest absolute eigenvalue of matrix A. The LU method is used to solve the inverse eigenproblem instead of calculating the inverse matrix A^{-1}. The power method applied to matrix A^{-1} is given by

$$\mathrm{A}^{-1}\mathbf{x}^{(\mathrm{k})} = \mathbf{y}^{(\mathrm{k}+1)}. \tag{10}$$

Multiplying (10) by A gives

$$\mathrm{A}\,\mathrm{A}^{-1}\mathbf{x}^{(\mathrm{k})} = \mathrm{I}\mathbf{x}^{(\mathrm{k})} = \mathbf{x}^{(\mathrm{k})} = \mathrm{A}\mathbf{y}^{(\mathrm{k}+1)}, \tag{11}$$

which can be written as

$$A\mathbf{y}^{(k+1)} = \mathbf{x}^{(k)}. \tag{12}$$

Equation (12) is in the standard form $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x} = \mathbf{y}^{(k+1)}$ and $\mathbf{b} = \mathbf{x}^{(k)}$. Thus, for a given $\mathbf{x}^{(k)}, \mathbf{y}^{(k+1)}$ can be found by the Doolittle LU method. The procedure is as follows:

(i) Solve for L and U such that $LU = A$ by the Doolittle LU method.

(ii) Assume $\mathbf{x}^{(0)}$. Designate a component of $\mathbf{x}$ to be unity.

(iii) Solve for $\mathbf{x}'$ by forward substitution using the equation $L\mathbf{x}' = \mathbf{x}^{(0)}$.

(iv) Solve for $\mathbf{y}^{(1)}$ by back substitution using the equation $U\mathbf{y}^{(1)} = \mathbf{x}'$.

(v) Scale $\mathbf{y}^{(1)}$ so that the unity component is unity. Thus, $\mathbf{y}^{(1)} = \lambda_{\text{inverse}}^{(1)}\mathbf{x}^{(1)}$.

(vi) Repeat step 3 to 5 with $\mathbf{x}^{(1)}$. Iterate to convergence. At convergence,

$\lambda = \dfrac{1}{\lambda_{\text{inverse}}}$, and $\mathbf{x}^{(k+1)}$ is the corresponding eigenvector. The inverse power method algorithm is as follows:

$$L\mathbf{x}' = \mathbf{x}^{(k)}$$

$$U\mathbf{y}^{(k+1)} = \mathbf{x}'$$

$$\mathbf{y}^{(k+1)} = \lambda_{\text{inverse}}^{(k+1)}\,\mathbf{x}^{(k+1)}.$$

Example for the square matrix of order 3

We consider the square matrix of order 3,

$$A = \begin{bmatrix} 8 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 13 \end{bmatrix}.$$

We assume that $\mathbf{x}^{(0)} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

We scale the first component of $\mathbf{x}$ to unity.

The first step is to solve for L and U by the Doolittle LU method.

Let $A = LU$. Then

$$\begin{bmatrix} 8 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 13 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}\begin{bmatrix} U_{11} & U_{12} & U_{13} \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix}$$

$$= \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ L_{21}U_{11} & L_{21}U_{12} + U_{22} & L_{21}U_{13} + U_{23} \\ L_{31}U_{11} & L_{31}U_{12} + L_{32}U_{22} & L_{31}U_{13} + L_{32}U_{23} + U_{33} \end{bmatrix}.$$

Comparing coefficients, we have

$$\begin{aligned} U_{11} &= 8, & U_{12} &= -2, & U_{13} &= -2, \\ L_{21}U_{11} &= -2, & L_{21}U_{12} + U_{22} &= 4, & L_{21}U_{13} + U_{23} &= -2, \\ L_{31}U_{11} &= -2, & L_{31}U_{12} + L_{32}U_{22} &= -2, & L_{31}U_{13} + L_{32}U_{23} + U_{33} &= 13. \end{aligned}$$

Solving these relations, we get

$$\begin{aligned} U_{11} &= 8, & U_{12} &= -2, & U_{13} &= -2, \\ L_{21} &= -\frac{1}{4}, & U_{22} &= \frac{7}{2}, & U_{23} &= -\frac{5}{2}, \\ L_{31} &= -\frac{1}{4}, & L_{32} &= -\frac{5}{7}, & U_{33} &= \frac{75}{7}. \end{aligned}$$

So,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \\ -\frac{1}{4} & -\frac{5}{7} & 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} 8 & -2 & -2 \\ 0 & \frac{7}{2} & -\frac{5}{2} \\ 0 & 0 & \frac{75}{7} \end{bmatrix}.$$

We have to solve for $\mathbf{x}'$ by forward substitution using $L\mathbf{x}' = \mathbf{x}^{(0)}$. Thus

$$\begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \\ -\frac{1}{4} & -\frac{5}{7} & 1 \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

From this we have three relations

$$\begin{aligned} x_1' &= 1 \\ -\frac{1}{4}x_1' + x_2' &= 1 \\ -\frac{1}{4}x_1' - \frac{5}{7}x_2' + x_3' &= 1. \end{aligned}$$

Solving these relations, we get

$$x_1' = 1, \quad x_2' = \frac{5}{4}, \quad x_3' = \frac{15}{7} \text{ , and}$$

$$\mathbf{x}' = \begin{bmatrix} 1 & \frac{5}{4} & \frac{15}{7} \end{bmatrix}^T.$$

We have to solve for $\mathbf{y}^{(1)}$ by backward substitution using $\mathrm{U}\mathbf{y}^{(1)} = \mathbf{x}'$. Thus

$$\begin{bmatrix} 8 & -2 & -2 \\ 0 & \frac{7}{2} & -\frac{5}{2} \\ 0 & 0 & \frac{75}{7} \end{bmatrix} \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ y_3^{(1)} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{5}{4} \\ \frac{15}{7} \end{bmatrix}.$$

From this we have three relations

$$\begin{aligned} 8y_1^{(1)} - 2y_2^{(1)} - 2y_3^{(1)} &= 1 \\ \frac{7}{2}y_2^{(1)} - \frac{5}{2}y_3^{(1)} &= \frac{5}{4} \\ \frac{75}{7}y_3^{(1)} &= \frac{15}{7}. \end{aligned}$$

Solving these relations, we get

$$y_3^{(1)} = 0.2, \quad y_2^{(1)} = 0.5, \quad y_1^{(1)} = 0.3 \text{ , and}$$

$$\mathbf{y}^{(1)} = \begin{bmatrix} 0.3 & 0.5 & 0.2 \end{bmatrix}^{\mathrm{T}}.$$

We scale $\mathbf{y}^{(1)}$ so that the unity component is unity. Thus

$$\mathbf{y}^{(1)} = \begin{bmatrix} 0.3 \\ 0.5 \\ 0.2 \end{bmatrix} = 0.3 \begin{bmatrix} 1 \\ 1.666667 \\ 0.666667 \end{bmatrix},$$

so $\lambda_{\text{inverse}}^{(1)} = 0.3$ and

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1 & 1.666667 & 0.666667 \end{bmatrix}^{\mathrm{T}}.$$

Next we have to solve for $\mathbf{x}'$ by forward substitution using $\mathrm{L}\mathbf{x}' = \mathbf{x}^{(1)}$. Thus

$$\begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \\ -\frac{1}{4} & -\frac{5}{7} & 1 \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} 1 \\ 1.666667 \\ 0.666667 \end{bmatrix}.$$

From this we have three relations

$$\begin{aligned} x_1' &= 1 \\ -\frac{1}{4}x_1' + x_2' &= 1.666667 \\ -\frac{1}{4}x_1' - \frac{5}{7}x_2' + x_3' &= 0.666667. \end{aligned}$$

Solving these relations, we get

$$x_1' = 1, \quad x_2' = 1.916667, \quad x_3' = 2.285715 \text{ , and}$$

$$\mathbf{x}' = \begin{bmatrix} 1 & 1.916667 & 2.285715 \end{bmatrix}^{\mathrm{T}}.$$

We have to solve for $\mathbf{y}^{(2)}$ by backward substitution using $\mathrm{U}\mathbf{y}^{(2)} = \mathbf{x}'$. Thus

$$\begin{bmatrix} 8 & -2 & -2 \\ 0 & \frac{7}{2} & -\frac{5}{2} \\ 0 & 0 & \frac{75}{7} \end{bmatrix} \begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \\ y_3^{(2)} \end{bmatrix} = \begin{bmatrix} 1 \\ 1.916667 \\ 2.285715 \end{bmatrix}.$$

From this we have three relations

$$8y_1^{(2)} - 2y_2^{(2)} - 2y_3^{(2)} = 1$$

$$\frac{7}{2} y_2^{(2)} - \frac{5}{2} y_3^{(2)} = 1.916667$$

$$\frac{75}{7} y_3^{(2)} = 2.285715.$$

Solving these relations, we get

$$y_3^{(2)} = 0.213333, \quad y_2^{(2)} = 0.7, \quad y_1^{(2)} = 0.353333 \text{ , and}$$

$$\mathbf{y}^{(2)} = \begin{bmatrix} 0.353333 & 0.7 & 0.213333 \end{bmatrix}^{\mathrm{T}}.$$

We scale $\mathbf{y}^{(2)}$ so that the unity component is unity. Thus

$$\mathbf{y}^{(2)} = \begin{bmatrix} 0.353333 \\ 0.7 \\ 0.213333 \end{bmatrix} = 0.353333 \begin{bmatrix} 1 \\ 1.981133 \\ 0.603773 \end{bmatrix},$$

so $\lambda_{\text{inverse}}^{(2)} = 0.353333$ and

$$\mathbf{x}^{(2)} = \begin{bmatrix} 1 & 1.981133 & 0.603773 \end{bmatrix}^{\mathrm{T}}.$$

The results of the first two iterations presented above and subsequent iterations (thirteen iterations) are presented in following table.

Table

k	$\lambda_{inverse}$	x_1	x_2	x_3
0		1	1	1
1	0.3	1	1.666667	0.666667
2	0.353333	1	1.981133	0.603773
3	0.382264	1	2.094439	0.597566
4	0.393346	1	2.130399	0.598460
5	0.396978	1	2.141252	0.599235
6	0.398094	1	2.144466	0.599554
7	0.398429	1	2.145406	0.599663
8	0.398526	1	2.145684	0.599692
9	0.398556	1	2.145764	0.599707
10	0.398565	1	2.145785	0.599711
11	0.398567	1	2.145792	0.599711
12	0.398568	1	2.145794	0.599712
13	0.398568	1	2.145794	0.599712

The final solution for the smallest absolute eigenvalue λ and the corresponding eigenvector $\mathbf{x}$ are

$$\lambda = \frac{1}{\lambda_{inverse}} = \frac{1}{0.398568} = 2.508982 \quad \text{and}$$

$$\mathbf{x} = \begin{bmatrix} 1 & 2.145794 & 0.599712 \end{bmatrix}^T .$$

Acknowledgements

I am grateful to **Dr Aung Kyaw Thin**, Rector and **Dr Aye Aye Han**, Pro-Rector, Banmaw University. I am very thankful to **Dr Kyaw Nyunt**, Professor (Retired), Department of Mathematics, Mandalay University, for his continuous encouragement on doing research.

References

Kreyszig, E., "**Advanced Engineering Mathematics**", John Wiley and Sons, Singapore, (1999).

Mathews, J.H., "**Mathematics, Science and Engineering**", Second Edition, California State University, United States of America, (1992).

Sastry, S.S., "**Introductory Methods of Numerical Analysis**", Third Edition, New Delhi, (1999).

Fundamental Theorems for Banach Spaces

M Roi Seng[1], Swe Mon Oo[2], Hnin Yee San[3]

Abstract

Banach space is very interesting area in functional analysis. So, in this paper, we discuss the important fundamental theorems for Banach space, namely, the open mapping theorem, the uniform boundedness principle and the closed graph theorem.

Introduction

Much of the theory of Banach space is base on these related results, called the open mapping theorem, uniform boundedness principle and the closed graph theorem. These are the cornerstone of the theory of Banach spaces. Firstly, we present the concept and some terminology of Banach space. And then we state and prove the important fundamental theorems for Banach Spaces.

Banach Spaces

In this section, we present some basic definitions and notions of Banach spaces. Now, we recall the concept of normed vector space which is the important tools of Banach space.

Definition

Let X be a vector space over the field K of real or complex numbers. A function $\|.\|: X \times X \to R$ is a **norm** if it is satisfied the following conditions:

(N_1) $\|x\| \geq 0, \|x\| = 0$ if and only if $x = 0, \forall x \in X,$

(N_2) $\|\alpha x\| = |\alpha| \, \|x\|, \forall \alpha \in K, \forall x \in X,$

(N_3) $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in X.$

A pair $(X, \| \ \|)$ is called a **normed vector space** (or) simply a **normed space**.

Example

The vector space R^n of all n-tuples $x = (x_1, \ldots, x_n)$ of real number is a real normed space with the normed defined by

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

1 Dr, Associate Professor, Department of Mathematics, Banmaw University
2 Assistant Lecturer, Department of Mathematics, Banmaw University
3 Tutor, Department of Mathematics, Banmaw University

Definition

A complete normed vector space X is called a **Banach space**.

Example

The vector space R^n and C^n of all n-tuples $x=(\xi_1,\ldots,\xi_n)$ of real and complex numbers. These are Banach spaces with norm defined by

$$\|x\|=\left(\sum_{i=1}^{n}|\xi_i|^2\right)^{\frac{1}{2}}=\sqrt{|\xi_1|^2+\ldots+|\xi_n|^2}.$$

Theorem

A subspace Y of a Banach space X is complete if and only if the set Y is closed in X.

Proof

Let X be a Banach space and $Y\subset X$.

Suppose that Y is a complete in X.

Let $x\in\overline{Y}$, then there is a sequence $(x_n)\in Y$ which converges to x.

Since every convergent sequence in a normed space is Cauchy, then (x_n) converges to x in Y.

Therefore $\overline{Y}\subset Y$. But $Y\subset\overline{Y}$, then $Y=\overline{Y}$.

Hence Y is closed.

Conversely, let Y be closed in X.

We take a Cauchy sequence (x_n) in Y and $x\in\overline{Y}$.

Then there exists a sequence (x_n) in Y such that (x_n) which is convergent to x.

Since Y is closed, then $Y=\overline{Y}$ and we have $x\in Y$.

Cauchy sequence (x_n) was arbitrary, then Y is complete.

Definition

A **linear operator** T is an operator such that the Domain D(T) of T is a vector space and the range R(T) lies in a vector space over the same field, for all $x,y\in D(T)$ and scalar α,

(i) $T(x+y)=Tx+Ty$,

(ii) $T(\alpha x)=\alpha Tx$.

Theorem

Let X and Y be vector spaces, both real and complex. Let $T:D(T)\to Y$ be a linear operator with domain $D(T)\subset X$ and the range $R(T)\subset Y$. Then

(i) the inverse $T^{-1}: R(T) \to D(T)$ exists if and only if $Tx = 0 \Rightarrow x = 0$.

(ii) If T^{-1} exists, it is a linear operator.

(iii) If $\dim D(T) = n < \infty$ and T^{-1} exists, then $\dim R(T) = \dim D(T)$.

Theorem

Let $T: D(T) \to Y$ be a linear operator, where $D(T) \subset Y$ and X, Y are normed spaces. Then

(i) T is continuous if and only if T is bounded.

(ii) If Y is continuous at a single point, it is continuous.

Definition

Let X and Y be normed spaces and $T: D(T) \to Y$ a linear operator, where $D(T) \subset X$. The operator T is said to be **bounded** if there is a real number c such that for all $x \in D(T)$, $\|Tx\| \le c\|x\|$.

Theorem

If a normed space X is finite dimension, then every linear operator on X is bounded.

The Fundamental Theorems for Banach Spaces

In this section, we discuss the fundamental theorems for Banach spaces. Firstly, we present the Baire's category theorem and drive from it the uniform boundedness theorem as well as the open mapping theorem.

Definitions

A subset M of a metric space X is said to be

(i) **rare** in X if its closure $\overline{M}$ has no interior points,

(ii) **meager** in X if M is the union of countably many sets each of which is rare in X,

(iii) **nonmeager** in X if M is not meager in X.

Theorem (Baire's Category Theorem)

If a metric space $X \neq \varnothing$ is complete, it is nonmeager in itself. Hence if $X \neq \varnothing$ is complete and $X = \bigcup_{k=1}^{\infty} A_k$, ($A_k$ closed) then at least one A_k contains a nonempty open subset.

Theorem (Uniform Boundedness Theorem)

Let (T_n) be a sequence of bounded linear operator $T_n : X \to Y$ from a Banach space X into a normed space Y such that $\left(\|T_n\, x\|\right)$ is bounded for every $x \in X$, say

$$\|T_n\, x\| \le c_x,\ n = 1, 2, ...,$$

where c_x is a real number. Then the sequence of the norms $\|T_n\|$ is bounded, that is, there is a c such that

$$\|T_n\, x\| \le c,\ n = 1, 2, \ldots .$$

Proof

For every $k \in N$, we let $A_k \subset X$ be the set of all $x \in X$ such that $\|T_n\, x\| \le k$ for all n.

For any $x \in \overline{A_k}$, there is a sequence (x_j) in A_k converging to x. This mean that for every fixed n we have $\|T_n\, x_j\| \le k$ and obtain $\|T_n\, x\| \le k$ because T_n is continuous and so is the norm.

Hence $x \in A_k$ and A_k is closed.

This show that each $x \in X$ belong to some A_k.

Then $X = \bigcup_{k=1}^{\infty} A_k$.

Since X is complete, Baire's theorem implies that some A_k contains an open ball, say,

$B_0 = B(x_0, r) \subset A_{k_0}$.

Let $x \in X$ be arbitrary, not zero.

We set $z = x_0 + \gamma x,\ \gamma = \dfrac{r}{2\|x\|}$.

Then $\|z - x_0\| < r$, so that $z \in B_0$.

Thus we have $\|T_n\, z\| \le k_0$ for all n.

Also $\|T_n\, x_0\| \le k_0$, since $x_0 \in B_0$.

We obtain $x = \dfrac{1}{r}(z - x_0)$.

This yield for all n,

$$\|T_n\, x\| = \frac{1}{r}\|T_n(z - x_0)\| \le \frac{1}{r}\left(\|T_n\, z\| + \|T_n\, x_0\|\right)$$

$$\le \frac{4}{r}\|x\| k_0 .$$

Hence for all n,

$$\|T_n\| = \sup\|T_n\, x\| \le \frac{4}{r} k_0 = c, \text{ where } c = \frac{4k_0}{r}.$$

We shall now approach the second big theorem in this section, the open mapping theorem. It will be also based on Baire's Category theorem.

Definition

Let X and Y be metric spaces. Then $T: D(T) \to Y$ with domain $D(T) \subset X$ is called an **open mapping** if for every open set in $D(T)$, the image is an open set in Y.

Lemma

A bounded linear operator T from a Banach space X into a Banach space Y has the property that the image $T(B_0)$ of the open unit ball $B_0 = B(0,1) \subset X$ contains an open ball about $0 \in Y$.

Theorem (Open Mapping Theorem (or) Bounded Inverse Theorem)

A bounded linear operator T from a Banach space X into a Banach space Y is an open mapping. Hence if T is bijective, T^{-1} is continuous and thus bounded.

Proof

Let A be an open set in X.

We prove that for every open set $A \subset X$, the image T(A) is open in Y.

Let $y \in Tx \in T(A)$.

Since A is open, it continuous an open ball with center x.

Hence $(A - x)$ contains an open ball with center 0.

Let the radius of the ball be r and set $k = \frac{1}{r}$, so that $r = \frac{1}{k}$.

Then $k(A - x)$ contains the open unit ball B(0,1).

By the above lemma, we have $T\big(k(A - x)\big) = k\big[T(A) - Tx\big]$ contains an open ball about 0, and so that $T(A) - Tx$.

Hence T(A) contains an open ball about $Tx = y$. Since $y \in T(A)$ was arbitrary, then T(A) is open.

We have if $T^{-1}: Y \to X$ exists, then it is continuous.

Now we applying the preceding theorems in the above section, we have T^{-1} is linear.

Therefore T^{-1} is bounded.

Finally, we state the important closed graph theorem which is a closed linear operator on a Banach space is bounded.

Definition

Let X and Y be normed spaces and $T: D(T) \to Y$ is a linear operator with domain $D(T) \subset X$. Then T is called a **closed linear operator** if its graph

$G(T) = \{(x, y) \mid x \in D(T),\ y = T_x\}$ is closed in the normed space $X \times Y$, where the two algebraic operations of a vector space in $X \times Y$ are defined as used, that is,

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$
$$\alpha(x, y) = (\alpha x, \alpha y),$$

where α is a scalar and the norm $X \times Y$ is defined by $\|(x, y)\| = \|x\| + \|y\|$.

Theorem (Closed Graph Theorem)

Let X and Y be Banach spaces and $T: D(T) \to Y$ a closed linear operator, where $D(T) \subset X$. Then if D(T) is closed in X, the operator T is bounded.

Proof

We first show that $X \times Y$ with norm defined by the definition is complete.

Let (z_n) be Cauchy in $X \times Y$, where $z_n = (x_n, y_n)$.

Then for every $\varepsilon > 0$, there is an n such that
$\|z_n - z_m\| = \|x_n - x_m\| + \|y_n - y_m\| < \varepsilon \quad (m, n > N)$.

Hence (x_n) and (y_n) are Cauchy in X and Y, respectively.

Since X and Y are complete, then $x_n \to x$ and $y_n \to y$.

This implies that $z_n \to z = (x, y)$ with $m \to \infty$, we have $\|z_n - z\| \le \varepsilon$ for $n > N$.

Since the Cauchy sequence (z_n) was arbitrary, $X \times Y$ is complete.

By assumption, G(T) is closed in $X \times Y$ and D(T) is closed in X.

Therefore G(T) and D(T) are complete.

We now consider the mapping $P: G(T) \to D(T)$ by $(x, Tx) \mapsto x$ is linear.

P is bounded because

$\|P(x, Tx)\| = \|x\| \le \|x\| + \|Tx\| = \|(x, Tx)\|$.

Since P is bijective, the inverse mapping is

$P^{-1}: D(T) \to G(T)$ by $x \mapsto (x, Tx)$.

Since G(T) and D(T) are complete and we can apply the bounded inverse theorem, then P^{-1} is bounded.

So, $\|(x, Tx)\| \le b\|x\|$ for some b and for all $x \in D(T)$.

Then $\|Tx\| \le \|Tx\| + \|x\|$

$= \|(x, Tx)\|$

$\le b\|x\|$, for all $x \in D(T)$.

Hence T is bounded.

Acknowledgement

First of all we would like to express our gratitude to Dr Aung Kyaw Thin, Rector, University of Banmaw, for his encouragement and permission to conduct this research paper.

We wish to express to Dr Aye Aye Han, Pro-Rector, University of Banmaw, whose permission to us to carry out our present research.

We also thank to Dr Khin San Aye, Professor and Head, Department of Mathematics, University of Banmaw, for her encouragement.

Finally, we would like particular to thank Dr Hnin Oo Lwin, professor, Department of mathematics, University of Banmaw, for her assistance and helpful suggestions.

References

[1] Robert E, Megginson, "An Introduction to Banach Space Theory". Springer Verlag, New York, (1998).

[2] KrwinKreyszig, "Introductory Functional Analysis with Applications". John Wiley & Sons.Inc. Canada, (1978).

[3] Avner Friedman, "Foundations of Morden Analysis". Dover Publications, Inc. New York, (1970).

Linear Flow of Heat in the Solid Bounded by Two Parallel Planes

Nge Nge Khaing*

Abstract

Fourier Series is applied to study the linear flow of the heat in solid bounded by two parallel planes by giving different initial temperature. And then it is also solved by giving surface temperature. Finally, the temperature distributions in the slab are illustrated.

Introduction

In this paper we shall examine various cases of linear flow of heat in a solid bounded by a pair of parallel planes, usually $x = 0$ and $x = \ell$. This region we shall usually refer to briefly as the "slab $0 < x < \ell$". The results apply also to a rod of length ℓ with the same end conditions with no loss of heat from its surface.

Steady temperature

In the case of steady flow in a slab of conductivity K and thickness ℓ whose surfaces are kept at temperatures u_1 and u_2, the differential equation becomes

$$\frac{d^2u}{dx^2} = 0.$$

Thus $$\frac{du}{dx} = \text{constant} = \frac{u_2 - u_1}{\ell}.$$

Also the flux at any point is

$$j = -K\frac{du}{dx} = -K\frac{(u_2 - u_1)}{\ell} = \frac{u_1 - u_2}{R}, \qquad (1)$$

where $$R = \frac{\ell}{K}. \qquad (2)$$

The relation (1) is precisely analogous to Ohm's law for the steady flow of electric current; the flux j corresponds to the electric current, and the fall in temperature $u_1 - u_2$ to the fall in potential. This R may be called the thermal resistance of the slab.

Next suppose we have a composite wall composed of n slabs of thickness $\ell_1, \ell_2 \ldots, \ell_n$ and conductivities $K_1, K_2 \ldots, K_n$. If the slabs are in perfect thermal contact at their surfaces of separation, the fall of temperature over the whole wall will be the sum of the falls over the component slabs, and, since the flux is the same at every point, this sum is

$$\frac{j\ell_1}{K_1} + \frac{j\ell_2}{K_2} + \cdots + \frac{j\ell_n}{K_n} = (R_1 + R_2 + \cdots + R_n)\,j. \qquad (3)$$

* Dr, Associate Professor, Department of Mathematics, Banmaw University

This is equivalent to the statement that the thermal resistance of a composite wall is the sum of the thermal resistance of the separate layers, assuming perfect thermal contact between them.

If the conductivity K is a function of temperature, the differential equation is

$$\frac{d}{dx}\left(K\frac{du}{dx}\right)=0.$$

Thus the relation $-K\frac{du}{dx}=j,$ constant,

still holds. Integrating between the surface temperatures u_1 and u_2 of a slab of thickness ℓ, we have

$$-\int_{u_1}^{u_2} K\,du = j\ell,$$

and thus $$j=\frac{(u_1-u_2)}{\ell}K_{av} \qquad (4)$$

where $$K_{av}=\frac{1}{u_2-u_1}\int_{u_1}^{u_2} K\,du. \qquad (5)$$

Surface conditions independent of the time

Suppose that we have to satisfy

$$\nabla^2 u-\frac{1}{k}\frac{\partial u}{\partial t}=A(x,y,z) \qquad (6)$$

throughout the solid, with $u=f(x,y,z)$ initially, and $u=\phi(x,y,z)$ at the surface.

Put $$u=v+w, \qquad (7)$$

where v is a function of x, y, z only which satisfies

$$\nabla^2 v=A(x,y,z) \qquad (8)$$

throughout the solid and

$$v=\phi(x,y,z) \qquad (9)$$

at its surface, while w is a function of x, y, z, t, which satisfies

$$\nabla^2 w-\frac{1}{k}\frac{\partial w}{\partial t}=0, \qquad (10)$$

throughout the solid,

$$w=f(x,y,z)-v, \text{ initially} \qquad (11)$$

and $$w=0, \text{ at the surface.} \qquad (12)$$

Duhamel's Theorem

If $u = F(x, y, z, \lambda, t)$ represents the temperature at (x, y, z) at the time t in a solid in which the initial temperature is zero, while its surface temperature is $\phi(x, y, z, \lambda)$, then the solution of the equation (6) in which the initial temperature is zero, and the surface temperature is $\phi(x, y, z, t)$ is given by

$$u = \int_0^t \frac{\partial}{\partial t} F(x, y, z, \lambda, t - \lambda)\, d\lambda.$$

Proof:

When the surface temperature is zero from $t = -\infty$ to $t = 0$, and $\phi(x, y, z, \lambda)$ from $t = 0$ to $t = t$, we may say that the initial temperature is zero and the surface temperature is $\phi(x, y, z, \lambda)$, so that the temperature at the time t is given by

$$u = F(x, y, z, \lambda, t), \text{ when } t > 0.$$

Therefore, when the surface temperature is zero from $t = -\infty$ to $t = \lambda$, and $\phi(x, y, z, \lambda)$ from $t = \lambda$ to $t = t$, we have

$$u = F(x, y, z, \lambda, t - \lambda), \text{ when } t > \lambda.$$

Also, when the surface temperature is zero from $t = -\infty$ to $t = \lambda + d\lambda$, and $\phi(x, y, z, \lambda)$ from $t = \lambda + d\lambda$ to $t = t$, we have

$$u = F(x, y, z, \lambda, t - \lambda - d\lambda), \text{ when } t > \lambda + d\lambda.$$

Hence, when the surface temperature is zero from $t = -\infty$ to $t = \lambda$, $\phi(x, y, z, \lambda)$ from $t = \lambda$ to $t = \lambda + d\lambda$, and zero from $t = \lambda + d\lambda$ to $t = t$, we have

$$du = F(x, y, z, \lambda, t - \lambda) - F(x, y, z, \lambda, t - \lambda - d\lambda)$$

or ultimately $u = \frac{\partial}{\partial t} F(x, y, z, \lambda, t - \lambda)\, d\lambda, \ (t > \lambda).$

In this way, by breaking up the interval $t = 0$ to $t = t$ into these small intervals, and then summing the results thus obtained, we find the solution of the problem for the surface temperature $\phi(x, y, z, t)$ in the form

$$u = \int_0^t \frac{\partial}{\partial t} F(x, y, z, t - \lambda)\, d\lambda.$$

The Region $0 < x < \ell$ Whose Ends are Kept at Zero Temperature and Whose Initial Temperature is f(x)

The following special cases of $u(x, t) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi x}{\ell} e^{-k\frac{n^2\pi^2 t}{\ell^2}}$ are of interest:

(i) Constant initial temperature $f(x) = U_0$.

Since

$$u(x,t) = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{\ell} e^{-k\frac{n^2\pi^2 t}{\ell^2}}$$

where

$$\begin{aligned} b_n &= \frac{2}{\ell}\int_0^{\ell} U_0 \sin\frac{n\pi x}{\ell}\,dx \\ &= \frac{2U_0}{n\pi}\left[1-\cos n\pi\right] \\ &= \begin{cases} 0 \text{ , if n is even} \\ \dfrac{4U_0}{n\pi}\text{, if n is odd.} \end{cases} \end{aligned}$$

The most general solution is

$$u(x,t) = \frac{4U_0}{\pi}\sum_{n=0}^{\infty}\frac{1}{(2n+1)} e^{\frac{k(2n+1)^2\pi^2 t}{\ell^2}} \sin\frac{(2n+1)\pi x}{\ell}. \qquad (13)$$

(ii) A linear initial distribution $f(x) = kx$.

$$u(x,t) = \frac{2\ell k}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n} e^{-\frac{kn^2\pi^2 t}{\ell^2}} \sin\frac{n\pi x}{\ell}. \qquad (14)$$

In general, it is a little more satisfactory to set out results for the symmetrical case of the slab $-\ell < x < \ell$ so that direct comparison with similar results for the sphere and cylinder is possible.

(iii) The slab $-\ell < x < \ell$ with constant initial temperature U_0. Changing the origin in (13) to the mid-point of the slab and replacing $\frac{1}{2}\ell$ by ℓ gives

$$\begin{aligned} a_n &= \frac{2}{\ell}\int_0^{t} U_0 \cos\frac{(2n+1)\pi x}{2\ell}\,dx \\ &= \frac{2}{\ell}\left[U_0 \sin\frac{(2n+1)\pi x}{2\ell}\frac{2\ell}{(2n+1)\pi}\right]_0^{\ell} \\ &= \frac{4U_0}{(2n+1)\pi}\sin\frac{(2n+1)\pi}{2} \\ &= 4U_0\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)}. \end{aligned}$$

Thus

$$u(x,t)=\frac{4U_0}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)}e^{\frac{-k(2n+1)^2\pi^2t}{4\ell^2}}\cos\frac{(2n+1)\pi x}{2\ell}. \tag{15}$$

The flux of heat j at the surface is

$$j=-K\left[\frac{\partial u}{\partial x}\right]_{x=\ell}=\frac{2KU_0}{\ell}\sum_{n=0}^{\infty}e^{-\frac{k(2n+1)^2\pi^2t}{4\ell^2}}. \tag{16}$$

(iv) For the region $-\ell<x<\ell$ with initial temperature $\frac{U_0(\ell-|x|)}{\ell}$ and zero surface temperature,

$$\begin{aligned}a_n&=\frac{1}{\ell}\int_{-\ell}^{\ell}\frac{U_0\ell-U_0|x|}{\ell}\cos\frac{(2n+1)\pi x}{2\ell}dx\\&=\frac{1}{\ell}\left[U_0\frac{2\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi x}{2\ell}\Bigg|_{-\ell}^{\ell}-\left\{\frac{U_0(-x)}{\ell}\frac{2\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi x}{2\ell}\Bigg|_{-\ell}^{0}\right.\right.\\&\quad-\int_{-\ell}^{0}-\frac{U_0}{\ell}\frac{2\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi x}{2\ell}dx+\frac{U_0x}{\ell}\frac{2\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi x}{2\ell}\Bigg|_{0}^{\ell}\\&\quad\left.\left.-\int_{0}^{\ell}\frac{U_0}{\ell}\frac{2\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi x}{2\ell}dx\right\}\right]\end{aligned}$$

$$\begin{aligned}a_n&=\frac{1}{\ell}\left[\frac{4U_0\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi}{2}-\frac{2U_0\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi}{2}+\frac{4U_0\ell}{(2n+1)^2\pi^2}\right.\\&\quad\left.-\frac{2U_0\ell}{(2n+1)\pi}\sin\frac{(2n+1)\pi}{2}+\frac{4U_0\ell}{(2n+1)^2\pi^2}\right]\\&=\frac{1}{\ell}\left[\frac{8U_0\ell}{(2n+1)^2\pi^2}\right]\\&=\frac{8U_0}{(2n+1)^2\pi^2}.\end{aligned}$$

Therefore

$$u(x,t)=\frac{8U_0}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}e^{\frac{-k(2n+1)^2\pi^2t}{4\ell^2}}\cos\frac{(2n+1)\pi x}{2\ell}. \tag{17}$$

(v) For the region $-\ell<x<\ell$ with initial temperature $\frac{U_0(\ell^2-x^2)}{\ell^2}$ and zero surface temperature,

$$u(x,t)=\frac{32U_0}{\pi^3}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^3}e^{\frac{-k(2n+1)^2\pi^2t}{4\ell^2}}\cos\frac{(2n+1)\pi x}{2\ell}. \qquad (18)$$

(vi) For the region $-\ell < x < \ell$ with initial temperature $U_0\cos(\frac{\pi x}{2\ell})$ and zero surface temperature,

$$u(x,t)=U_0\cos(\frac{\pi x}{2\ell})e^{-\frac{k\pi^2t}{4\ell^2}}. \qquad (19)$$

These results are interesting since they give a qualitative idea of the way in which heat is extracted from a slab with a given initial distribution of temperature.

Fig.1. Temperatures in the slab $0 < x < \ell$ with no flow at $x = 0$, zero temperature at $x = \ell$, and various initial distributions of temperature. The numbers on the curves are the values of kt/ℓ^2.

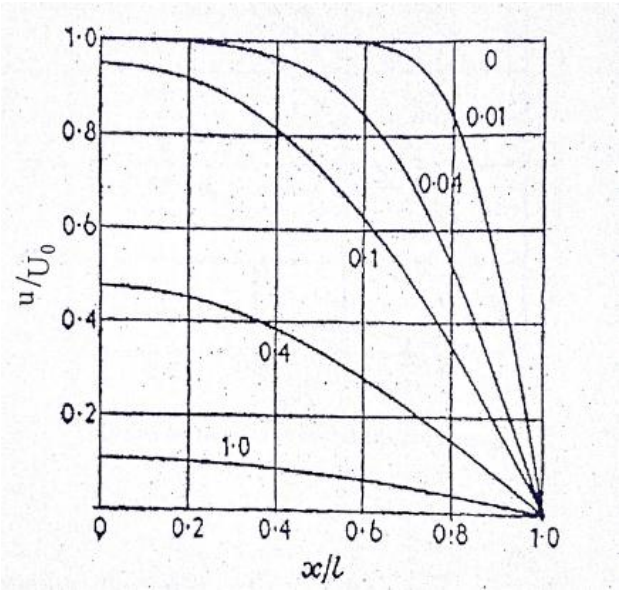

Figure:1(a) Constant initial temperature;

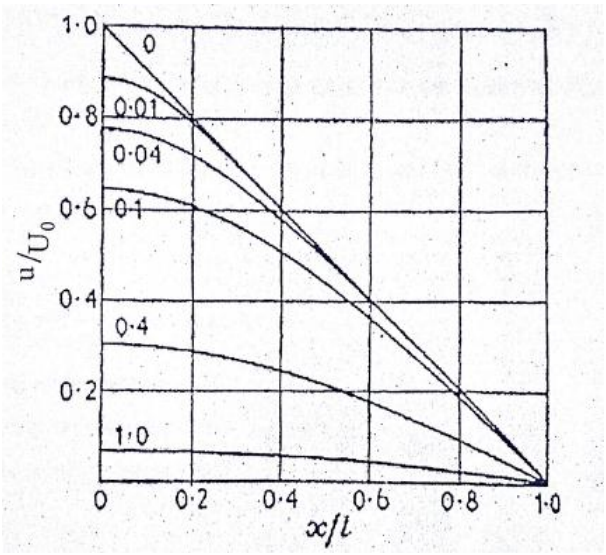

Figure:1(b) Linear initial temperature, (iv)

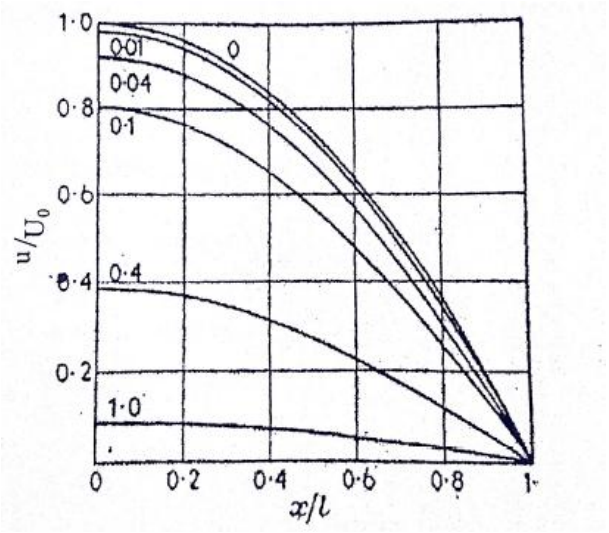

Figure:1(c) Parabolic initial temperature, (v)

The above figures illustrate that initial temperature on the surface begins to decrease to zero prior to the initial temperature inside the slab.

The Region $0 < x < \ell$ with the Initial Temperature f(x) and Whose Ends are Kept at Constant Temperature or Insulated

In the case in which the ends are kept at constant temperatures u_1 and u_2 we have the equations

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2} \quad (0 < x < \ell),$$

$$u = u_1, \text{ when } x = 0$$

$$u = u_2, \text{ when } x = \ell$$

$$\text{and } u = f(x), \text{ when } t = 0.$$

We reduce this to a case of steady temperature, and a case where the ends are kept at the zero temperature.

Put $u = v + w,$

where v and w satisfy the following equations:

$$\frac{d^2 v}{dx^2} = 0, \quad (0 < x < \ell),$$

$$v = u_1, \text{ when } x = 0,$$

$$v = u_2, \text{ when } x = \ell,$$

$$\text{and } \frac{\partial w}{\partial t} = k\frac{\partial^2 w}{\partial x^2}, \quad (0 < x < \ell),$$

$$w = 0 \text{ when } x = 0 \text{ and } x = \ell,$$

$$w = f(x) - v, \text{ when } t = 0.$$

We find at once that

$$v = u_1 + \frac{(u_2 - u_1)x}{\ell},$$

and it follows from equation $u(x,t) = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{\ell} e^{-k\frac{n^2\pi^2 t}{\ell^2}}$ that

$$w = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{\ell} e^{-\frac{kn^2\pi^2 t}{\ell^2}},$$

where $b_n = \frac{2}{\ell}\int_0^{\ell}\left[f(x') - \left\{u_1 + (u_2 - u_1)\frac{x'}{\ell}\right\}\right]\sin\frac{n\pi x'}{\ell}dx'$

$$= \frac{2}{\ell}\int_0^{\ell} f(x')\sin\frac{n\pi x'}{\ell}dx' - \frac{2}{\ell}\left[\left\{u_1 + (u_2 - u_1)\frac{x}{\ell}\right\}\left[(-\cos\frac{n\pi x'}{\ell})\frac{\ell}{n\pi}\right]_0^{\ell}\right.$$

$$\left. -\int_0^{\ell}(-\cos\frac{n\pi x'}{\ell})\frac{\ell}{n\pi}\left\{\frac{(u_2 - u_1)}{\ell}\right\}dx'\right]$$

$$= \frac{2}{\ell}\int_0^{\ell} f(x')\sin\frac{n\pi x'}{\ell}dx + \frac{2}{n\pi}(u_2\cos n\pi - u_1) + \frac{1}{n\pi}\frac{2}{\ell}(\sin\frac{n\pi x'}{\ell}\frac{\ell}{n\pi})\ell_0$$

$$= \frac{2}{n\pi}(u_2\cos n\pi - u_1) + \frac{2}{\ell}\int_0^{\ell} f(x')\sin\frac{n\pi x'}{\ell}dx$$

$$w = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{u_2\cos n\pi - u_1}{n}\sin\frac{n\pi x}{\ell}e^{-k\frac{n^2\pi^2 t}{\ell^2}}$$

$$+\frac{2}{\ell}\sum_{n=1}^{\infty}\sin\frac{n\pi x}{\ell}e^{-k\frac{n^2\pi^2 t}{\ell}}\int_0^{\ell} f(x')\sin\frac{n\pi x'}{\ell}dx.$$

Thus

$$u = u_1 + (u_2 - u_1)\frac{x}{\ell} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{u_2\cos n\pi - u_1}{n}\sin\frac{n\pi x}{\ell}e^{-\frac{kn^2\pi^2 t}{\ell^2}}$$

$$+\frac{2}{\ell}\sum_{n=1}^{\infty}\sin\frac{n\pi x}{\ell}e^{-\frac{kn^2\pi^2 t}{\ell^2}}\int_0^{\ell} f(x')\sin\frac{n\pi x'}{\ell}dx'. \qquad (20)$$

The simplest and most important case is that of region $-\ell < x < \ell$ with zero initial temperature and with the surfaces $x = \pm\ell$ kept at constant temperature U for $t > 0$. The solution, which follows immediately from (20) is

$$u = U - \frac{4U}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)}e^{-\frac{k(2n+1)^2\pi^2 t}{4\ell^2}}\cos\frac{(2n+1)\pi x}{2\ell}. \qquad (21)$$

Introducing the dimensionless parameters

$$T = \frac{kt}{\ell^2},\ \xi = \frac{x}{\ell}, \qquad (22)$$

(21) may be written in the form

$$\frac{u}{U}=1-\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)}e^{-\frac{(2n+1)^2\pi^2T}{4}}\cos\frac{(2n+1)\pi\xi}{2}, \qquad (23)$$

and the solution for all values of k, ℓ, t and x may be obtained from a family of curves in two dimensions. In figure 2, $\frac{u}{U}$ is plotted against ξ for various values of T.

For the case in which the end $x=0$ is insulated and the end $x=\ell$ is kept at U, the initial temperature being f(x), the solution is

$$u=U+\frac{2}{\ell}\sum_{n=0}^{\infty}e^{-\frac{k(2n+1)^2\pi^2t}{4\ell^2}}\cos\frac{(2n+1)\pi x}{2\ell}\left\{\frac{2\ell(-1)^{n+1}U}{(2n+1)\pi}+\int_0^{\ell}f(x')\cos\frac{(2n+1)\pi x'}{2\ell}dx'\right\}. \qquad (24)$$

If the initial temperature is f(x) and both ends $x=0$ and $x=\ell$ are thermally insulated the solution is

$$u=\frac{1}{\ell}\int_0^{\ell}f(x')\,dx'+\frac{2}{\ell}\sum_{n=1}^{\infty}e^{-\frac{kn^2\pi^2t}{\ell^2}}\cos\frac{n\pi x}{\ell}\int_0^{\ell}f(x')\cos\frac{n\pi x'}{\ell}dx'. \qquad (25)$$

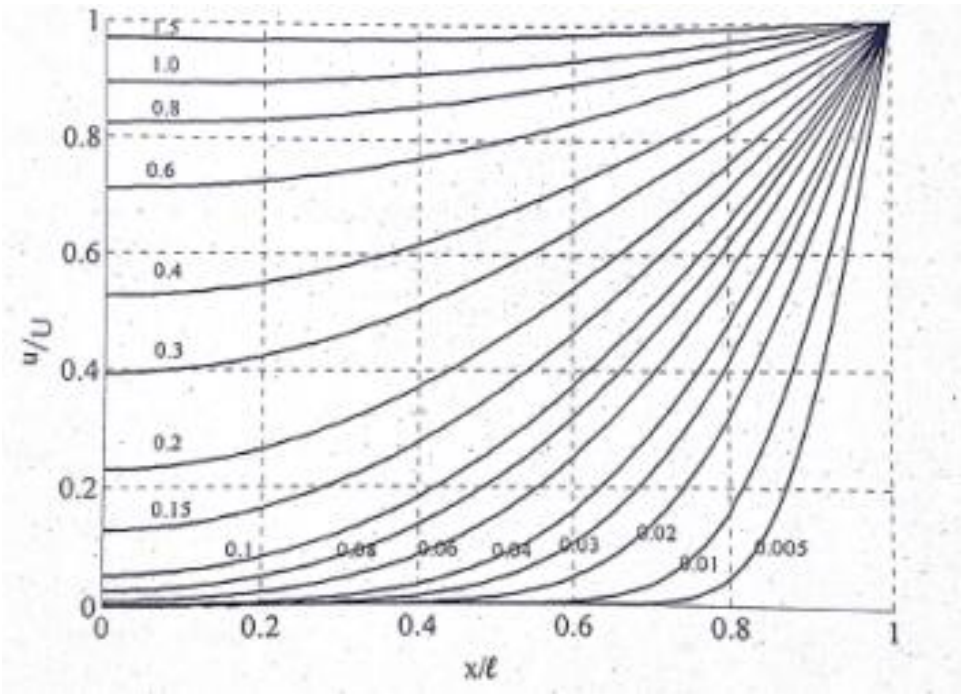

Fig.2. Temperature distribution at various times in the slab $-\ell<x<\ell$ with zero initial temperature and surface temperature U. The numbers on the curves are the values of kt/ℓ^2.

The figure illustrates that the temperature on the surface is gradually conducted to the centre of the slab and in the end the whole of the slab attains the temperature on the surface of the slab constantly.

Conclusion

We have seen that the temperature distribution in the rod whose surface temperatures and initial temperature are known can be obtained by using Fourier Series. Linear flow of heat can be calculated by giving different surface temperatures and different initial temperatures. Moreover, when the initial temperature is given to the system, the temperatures of the surfaces of the system decreases to zero first while the interior of the system gradually decreases and reaches zero lastly. On the other hand, if a slab of surface temperature zero is given a constant temperature to its surfaces then the temperature on the surfaces is gradually conducted to the interior and eventually the whole system attains the same temperature. These results may be applied in the engineering practice.

Acknowledgement

I would like to express my thanks to Dr Aung Kyaw Thin, Rector, Banmaw University, for his permission and encouragement during the course of the research work. I am also grateful to Dr Aye Aye Han, Pro-rector, Banmaw University, for her encouragement to submit this paper. I also thank to Professor Dr Khin San Aye, Head of the Department of Mathematics, Professor Dr Hnin Oo Lwin and my seniors, Department of Mathematics, Banmaw University who give me their invaluable advices.

References

Carslaw, H. S. & Jaeger, J. C., "*Conduction of Heat in Solids*", Clarendon Press, (1959).

Churchill, R. V., "*Complex Variable and Applications*", McGraw-Hill, (1960).

Jaeger, J. C., "*An Introduction to Applied Mathematics*", Oxford University Press, (1951).

Tyn Myint-U, "*Partial Differential Equations of Mathematical Physics*", Elsevier Science Publishing Company, Ltd., (1980).

Properties of the Pascoletti-Serafini Scalarization

Nu Wai Lwin Tun*

Abstract

In this paper, Pascoletti-Serafini scalarization of the mutiobjective optimization problems(MOP) is discussed. Then, the properties of the Pascoletti-Serafini scalarization are investigated.

Key words: mutiobjective optimization, scalarization, K-minimality.

Introduction

For determining solutions of the multiobjective optimization problem (MOP)

$$\min f(x) \qquad (\text{MOP})$$

subject to the constraints

$$g(x) \in C,$$

$$h(x) = 0_q,$$

$$x \in S$$

with the constraint set $\Omega = \{x \in S \mid g(x) \in C, h(x) = 0_q\}$ a wide-spread approach is the transformation of this problem to a scalar-valued parameter dependent optimization problem. This is done for instance in the weighted sum method. There the scalar problems

$$\min_{x \in \Omega} \sum_{i=1}^{m} \omega_i f_i(x)$$

with weights $\omega \in K^* \setminus \{0_m\}$ and K^* the dual cone to the cone K,

that is $K^* = \{y^* \in R^m \mid (y^*)^T y \geq 0 \text{ for all } y \in K\},$ are solved.

Another scalarization especially for calculating EP-minimal points is based on the minimization of only one of the m objectives while all the other objectives are transformed into constraints by introducing upper bounds. This scalarization is called **ε**-constraint method and is given by

$$\min f_K(x)$$

subject to the constraints

$$f_i(x) \leq \varepsilon_i, i = \{1, ..., m\} \setminus \{k\},$$

$$x \in \Omega.$$

Here the parameters are the upper bounds $\varepsilon_i, i = \{1, ..., m\} \setminus \{k\}$ for a $k \in \{1, ..., m\}$.

The ε-constraint method as given in this problem is only suited for the calculation of EP-minimal points.

* Dr, Lecturer, Department of Mathematics, Banmaw University

Yet problems arising in applications are often non-convex. Further it is also of interest to consider more general partial orderings than the natural ordering.

Thus I concentrate on a scalarization by Pascoletti and Serafini. An advantage of this scalarization is that many other scalarization approaches as the weighted sum method or the ε-constraint method are included in this more general formulation.

Pascoletti-Serafini Scalarization

Pascoletti and Serafini propose the following scalar optimization problem with parameters $a \in R^m$ and $r \in R^m$ for determining minimal solutions of (MOP) with respect to the cone K:

$$(SP(a,r)) \qquad \min t$$

subject to the constraints

$$a + tr - f(x) \in K,$$

$$g(x) \in C,$$

$$h(x) = 0_q,$$

$$t \in R, x \in S.$$

This problem has the parameter dependent constraint set

$$\sum(a,r) = \{(t,x) \in R^{n+1} \mid a + tr - f(x) \in K, x \in \Omega\}.$$

It is assumed that the cone K is a nonempty closed pointed convex cone. The formulation of this scalar optimization problem corresponds to the definition of K-minimality. A point $\bar{x} \in \Omega$ with $\bar{y} = f(\bar{x})$ is **K-minimal** if

$$(\bar{y} - K) \cap f(\Omega) = \{\bar{y}\}, \text{ (see Figure 1 for } m = 2 \text{ and } K = R_+^2).$$

For $K = R_+^m$ the K-minimal points are also called **Edgeworth-Pareto-minimal (EP-minimal) points** according to Edgeworth and Pareto.

Let K be a pointed ordering cone with $\text{int}(K) \neq \emptyset$. A point $\bar{x} \in \Omega$ is a **weakly minimal solution** of (MOP) with respect to K if

$$(\bar{y} - \text{int}(K)) \cap f(\Omega) = \emptyset.$$

If the problem (SP(a, r)) as follows

$$\min t$$

subject to the constraints

$$f(x) \in a + tr - K,$$

$$x \in \Omega,$$

$$t \in R,$$

for solving this problem, the ordering cone $-K$ is moved in direction $-r$ on the line $a + tr$ starting in the point a till the set $(a + tr - K) \cap f(\Omega)$ is reduced to the empty set. The smallest

value $\bar{t}$ for which $(a+tr-K)\bigcap f(\Omega)\neq\varnothing$ is the minimal value of (SP (a, r)) (see Figure 2 with $m=2$ and $K=R_+^2$).

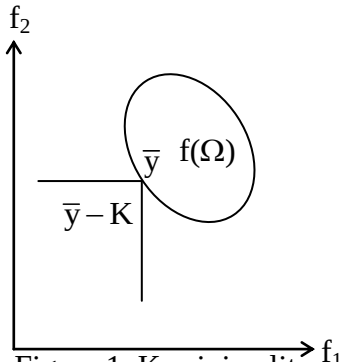

Figure 1. K-minimality.

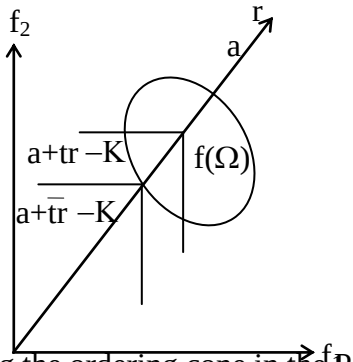

Figure 2. Moving the ordering cone in the Pascoletti-Serafini problem.

The scalar problem (SP(a,r)) features all important properties a scalarization approach for determining minimal solutions of (MOP) should have. If $(\bar{t},\bar{x})$ is a minimal solution of (SP(a,r)) then the point $\bar{x}$ is an at least weakly K-minimal solution of the multiobjective optimization problem (MOP) and by a variation of the parameters $(a,r)\in R^m\times R^m$ all K-minimal points of (MOP) can be found as solutions of (SP(a,r)). I will investigate these important properties.

Properties of the Pascoletti-Serafini Scalarization

It is examined that the following properties of this Pascoletti-Serafini Scalarization scalarization. It is assumed that K is nonempty closed pointed ordering cone in R^m.

Consider that the scalar optimization problem (SP(a,r)) to the multiobjective optimization problem (MOP). Let $int(K)\neq\varnothing$.

(i) Let $\bar{x}$ be a weakly K-minimal solution of the multiobjective optimization problem (MOP), then $(0,\bar{x})$ is a minimal solution of (SP(a,r)) for the parameter $a=f(\bar{x})$ and for arbitrary $r\in int(K)$.

(ii) Let $\bar{x}$ be a K-minimal solution of the multiobjective optimization problem (MOP), then $(0,\bar{x})$ is a minimal solution of (SP(a,r)) for the parameter $a=f(\bar{x})$ and for arbitrary $r\in K\setminus\{0m\}$.

(iii) Let $(\bar{t}, \bar{x})$ be a minimal solution of the scalar problem (SP(a,r)), then $\bar{x}$ is a weakly K-minimal solution of the multiobjective optimization problem (MOP) and $a + \bar{t}r - f(\bar{x}) \in \partial K$ with ∂K the boundary of the cone K.

(iv) Let $\bar{x}$ be a locally weakly K-minimal solution of the multiobjective optimization problem (MOP), then $(0, \bar{x})$ is a local minimal solution of (SP(a,r)) for the parameter $a = f(\bar{x})$ and for arbitrary $r \in \text{int}(K)$.

(v) Let $\bar{x}$ be a locally K-minimal solution of the multiobjective optimization problem (MOP), then $(0, \bar{x})$ is a local minimal solution of (SP(a,r)) for the parameter $a = f(\bar{x})$ and for arbitrary $r \in K \setminus \{0m\}$.

(vi) Let $(\bar{t}, \bar{x})$ be a local minimal solution of the scalar problem (SP(a,r)), then $\bar{x}$ is a locally weakly K-minimal solution of the multiobjective optimization problem (MOP) and $a + \bar{t}r - f(\bar{x}) \in \partial K$.

Proof:

(i) Set $a = f(\bar{x})$ and choose $r \in \text{int}(K)$ arbitrarily. Then the point $(0, \bar{x})$ is feasible for (SP(a, r)). It is also minimal, because otherwise there exists a feasible point (t', x') with $t' < 0$ and a $k' \in K$ with

$$a + t'r - f(x') = k'.$$

Hence I have $f(\bar{x}) = f(x') + k' - t'r$. It is $k' - t'r \in \text{int}(K)$ and thus it follows $f(\bar{x}) = f(x') + \text{int}(K)$, in contradiction to $\bar{x}$ weakly K-minimal.

(ii) Set $a = f(\bar{x})$ and choose $r \in K \setminus \{0m\}$ arbitrarily. Then the point $(0, \bar{x})$ is feasible for (SP(a, r)).

It is also a minimal solution because otherwise there exists a scalar $t' < 0$ and a point $x' \in \Omega$ with (t', x') feasible point for (SP(a, r)), and a $k' \in K$ with $a + t'r - f(x') = k'$. This leads to

$$f(\bar{x}) = f(x') + k' - t'r \in f(x') + K.$$

Because of the K-minimality of $\bar{x}$ I conclude $f(\bar{x}) = f(x')$ and thus $k' = t'r$. Due to the pointedness of the ordering cone K, $k' \in K$ and $t'r \in -K$ it follows $k' = t'r = 0_m$, in contradiction to $t' < 0$ and $r \neq 0_m$.

(iii) Assume $\bar{x}$ is not weakly K-minimal. Then there is a point $x' \in \Omega$ and a $k' \in \text{int}(K)$ with $f(\bar{x}) = f(x') + k'$. As $(\bar{t}, \bar{x})$ is a minimal solution and hence feasible for (SP (a, r)) there is a $\bar{k} \in K$ with $a + \bar{t}r - f(\bar{x}) = \bar{k}$.

Because of $\bar{k} + k' \in \text{int}(K)$ there is an $\varepsilon > 0$ with $\bar{k} + k' - \varepsilon r \in \text{int}(K)$. Then I conclude from $a + \bar{t}r - f(x') = \bar{k} + k'$,

$a + (\bar{t} - \varepsilon)r - f(x') \in \text{int}(K)$.

Then the point $(\bar{t} - \varepsilon, x')$ is feasible for (SP (a, r)), too, with $(\bar{t} - \varepsilon) < \bar{t}$, in contradiction to $(\bar{t}, \bar{x})$ minimal.

As $(\bar{t}, \bar{x})$ is a minimal solution and hence feasible for (SP(a,r)) there is a $\bar{k} \in K$ with $a + \bar{t}r - f(\bar{x}) = \bar{k}$.

Since $\bar{k} \in K, \bar{k} \notin (R^m \setminus K)$. Then $\bar{k} \notin \text{int}(R^m \setminus K)$ clearly.

Suppose that $\bar{k} \in \text{int}(K)$. Then there is an $\varepsilon > 0$ with $\bar{k} - \varepsilon r \in \text{int}(K)$.

Hence $a + (\bar{t} - \varepsilon)r - f(\bar{x}) \in \text{int}(K)$.

Then the point $(\bar{t}-\varepsilon,\bar{x})$ is feasible for (SP(a,r)), too, with $(\bar{t}-\varepsilon)<\bar{t}$, in contradiction to $(\bar{t},\bar{x})$ minimal.

Thus $\bar{k}\notin \text{int}(K)$. Hence $\bar{k}\in \partial K$.

Therefore $a+\bar{t}r-f(\bar{x})\in \partial K$.

(iv) It is assumed that $(0,\bar{x})$ is not a local minimal solution of (SP(a, r)). Then in any neighborhood $U=U_t\times U_x\subset R^{n+1}$ of $(0,\bar{x})$ there exists a feasible point (t',x') with $t'<0$ and a $k'\in K$ with $a+t'r-f(x')=k'$.

With $a=f(\bar{x})$ I get $f(\bar{x})=f(x')+k'-t'r$.

Since $k'-t'r\in \text{int}(K)$, we have $f(\bar{x})\in f(x')+\text{int}(K)$ and because the neighborhood U_x is arbitrarily chosen, cannot be locally weakly K-minimal.

(v) With the same arguments as in the preceding proof we conclude again that if there exists a feasible point (t',x') with $t'<0$ and x' in a neighborhood of $\bar{x}$, this leads to $f(\bar{x})=f(x')+k'-t'r$ with $r\in K\setminus\{0m\}$. Hence I have $f(\bar{x})=f(x')+K\setminus\{0_m\}$, in contradiction to $\bar{x}$ locally K-minimal.

(vi) Let $U=U_t\times U_x\subset R^{n+1}$ be a neighborhood such that $(\bar{t},\bar{x})$ is a local minimal solution of (SP (a, r)). Then there exists a $\bar{k}\in K$ with

$$a+\bar{t}r-f(\bar{x})=\bar{k}.$$

It is assumed that $\bar{x}$ is not a locally weakly K-minimal point of the multiobjective optimization problem (MOP). Then there exists no neighborhood $\bar{U}_x$ of $\bar{x}$ such that

$$f(\Omega\cap\bar{U}_x)\cap(f(\bar{x})-\text{int}(K))=\varnothing.$$

Hence for U_x there exists a point $x'\in\Omega\cap U_x$ with $f(x')\in f(\bar{x})-\text{int}(K)$ and thus there is a $k'\in\text{int}(K)$ with $f(x')=f(\bar{x})-k'$.

Together with $a+\bar{t}r-f(\bar{x})=\bar{k}$ I get $f(x')=a+\bar{t}r-\bar{k}-k'$. Because of $\bar{k}+k'\in\text{int}(K)$ there exists an $\varepsilon>0$ with $\bar{t}-\varepsilon\in U_t$ and $\bar{k}+k'-\varepsilon r\in\text{int}(K)$.

I conclude $f(x')=a+(\bar{t}-\varepsilon)r-(\bar{k}+k'-\varepsilon r)$

and thus $(\bar{t}-\varepsilon,x')\in U$ is feasible for (SP(a,r)) with $\bar{t}-\varepsilon<\bar{t}$ in contradiction to $(\bar{t},\bar{x})$ a local minimal solution.

Let $U=U_t\times U_x\subset R^{n+1}$ be a neighborhood such that $(\bar{t},\bar{x})$ is a local minimal solution of (SP(a,r)). Then there exists a $\bar{k}\in K$ with

$$a+\bar{t}r-f(\bar{x})=\bar{k}.$$

Since $\bar{k}\in K,\bar{k}\notin(R^m\setminus K)$. Then $\bar{k}\notin\text{int}(R^m\setminus K)$ clearly.

Suppose that $\bar{k}\in\text{int}(K)$. Then there is an $\varepsilon>0$ with $\bar{k}-\varepsilon r\in\text{int}(K)$.

Hence $a+(\bar{t}-\varepsilon)r-f(\bar{x})\in\text{int}(K)$.

Thus $(\bar{t}-\varepsilon,\bar{x})\in U$ is feasible for (SP(a,r)), with $(\bar{t}-\varepsilon)<\bar{t}$, in contradiction to $(\bar{t},\bar{x})$ a local minimal solution.

Thus $\bar{k}\notin\text{int}(K)$. Hence $\bar{k}\in\partial K$.

Therefore, $a+\bar{t}r-f(\bar{x})\in\partial K$.

Conclusion

It is noted that for the statement of the properties (ii) I need the pointedness of the ordering cone K. This is for instance not the case for the properties (iii) and also that it is not a consequence of the properties (iii) that I get always a weakly K-minimal point $\overline{x}$ by solving (SP(a, r)) for arbitrary parameters. It is possible that the problem (SP(a, r)) has no minimal solution at all as in the example shown in Figure 3 for the case $m = 2$ and $K = R_+^2$. There the minimal value of (SP(a,r)) is not bounded from below.

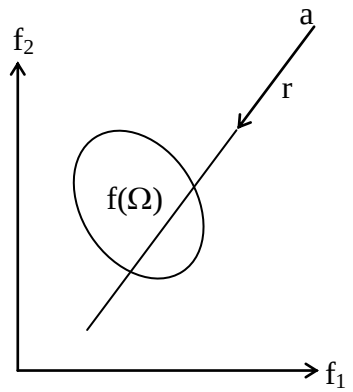

Figure 3. For $K = R_+^2$ there exists no minimal solution of problem (SP(a,r)).

Acknowledgement

I would like to express my thanks to Dr Aung Kyaw Thin, Rector, Banmaw University, for his permission and encouragement during the course of the research work. I am also grateful to Dr Aye Aye Han, Pro-rector, Banmaw University, for her encouragement to submit this paper. I also thank to Professor Dr Khin San Aye, Head of the Department of Mathematics, Professor Dr Hnin Oo Lwin and my seniors, Department of Mathematics, Banmaw University who give me their invaluable advices.

References

Eichfelder, G., "Adaptive Scalarization Methods in Multiobjective Optimization", Springer Verlag, Berlin, 2008.

Jahn, J., "Mathematical Vector Optimization in Partially Ordered Linear Spaces", Lang, Frankfurt, 1986.

Pascoletti, A. and Serafini, P., "*Scalarizing Vector Optimization Problems*", J.

Optimization Theory Appl., 42(4):499-524, 1984.

Sawaragi, Y., Nakayama, H., and Tanino, T., "Theory of Multiobjective Optimization", number 176 in mathematics in science and engineering, Academic Press, London, 1985.

Numerical Approximation for Definite Integral

Ohnmar[1], Khin Aye Win[2]

Abstract

The aim of this paper is the process of finding or evaluating the values of the definite integral from a set of numerical values of the integrand. Four kinds of approximate quadrature formulae are studied. The Trapezoidal rule, Simpson's One-third rule, Simpson's Three-eighth rule and Weddle's rule are expressed. Finally the approximate values of some problems are obtained by using these formulae.

Keywards: Forward differences, general quadrature formula, Simpson's rule.

Forward Differences

Let $y = f(x)$ be a function. Let $y = y_0, y_1, ..., y_n$ denote the set of values taken by the function. Then $y_1 - y_0, y_2 - y_1, ..., y_n - y_{n-1}$ are called the first differences of the function y. If we denote these differences by $\Delta y_0, \Delta y_1, ..., \Delta y_{n-1}$, we have

$$\Delta y_0 = y_1 - y_0$$

$$\Delta y_1 = y_2 - y_1$$

$$\vdots$$

$$\Delta y_{n-1} = y_n - y_{n-1}.$$

If we derive differences of these first differences, we obtain second differences of the function y. Now, if we denote these differences by $\Delta^2 y_0, \Delta^2 y_1, ..., \Delta^2 y_{n-2}$,

we have

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = y_2 - 2y_1 + y_0$$

$$\Delta^2 y_1 = \Delta y_2 - \Delta y_1 = y_3 - 2y_2 + y_1$$

$$\vdots$$

$$\Delta^2 y_{n-2} = \Delta y_{n-1} - \Delta y_{n-2} = y_n - 2y_{n-1} + y_{n-2}.$$

Similarly, the third differences are

$$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0 = y_3 - 3y_2 + 3y_1 - y_0$$

$$\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1 = y_4 - 3y_3 + 3y_2 - y_1$$

and so on.

1Lecturer, Department of Mathematics, Banmaw University
2Lecturer, Department of Mathematics, Banmaw University

In general, the k^{th} forward difference of y at x_j is given by

$$\Delta^k y_j = \Delta^{k-1} y_{j+1} - \Delta^{k-1} y_j .$$

The following table shows how the forward differences are formed.

j	Argument x	Entry y	First Differences Δy	Second Differences $\Delta^2 y$	Third Differences $\Delta^3 y$	Fourth Differences $\Delta^4 y$
0	x_0	y_0				
			Δy_0			
1	x_1	y_1		$\Delta^2 y_0$		
			Δy_1		$\Delta^3 y_0$	
2	x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$
			Δy_2		$\Delta^3 y_1$	
3	x_3	y_3		$\Delta^2 y_2$		
			Δy_3			
4	x_4	y_4				

General Quadrature Formula

Let $I = \int_a^b y\,dx$, where $y = f(x)$.

We suppose that f(x) is given for certain equidistant values of x:

$$x_0,\ x_0 + h,\ x_0 + 2h,\ \ldots .$$

Let the range (a, b) be divided into n equal parts, each of width h so that $b - a = nh$.

Let $x_0 = a,\ x_1 = x_0 + h = a + h,\ x_2 = a + 2h,\ \ldots,\ x_n = a + nh = b$.

We have assumed that the $n+1$ ordinates $y_0,\ y_1,\ \ldots,\ y_n$ are equidistant.

Therefore

$$I = \int_a^b y\,dx = \int_{x_0}^{x_0+nh} y_x\,dx = \int_0^n y_{x_0+hu}\,h\,du, \quad \text{where } u = \frac{x - x_0}{h},\ dx = h\,du .$$

Then

$$I = h\int_0^n \left[y_0 + u\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 y_0 + ... \right] du$$

$$= h\left[uy_0 + \frac{u^2}{2}\Delta y_0 + \left(\frac{u^3}{3} - \frac{u^2}{2}\right)\frac{\Delta^2 y_0}{2!} + \left(\frac{u^4}{4} - \frac{3u^3}{3} + \frac{2u^2}{2}\right)\frac{\Delta^3 y_0}{3!} + ... \right]_0^n$$

$$= h\left[ny_0 + \frac{n^2}{2}\Delta y_0 + \left(\frac{n^3}{3} - \frac{n^2}{2}\right)\frac{\Delta^2 y_0}{2!} + \left(\frac{n^4}{4} - n^3 + n^2\right)\frac{\Delta^3 y_0}{3!} \right.$$

$$\left. + ... \text{ up to } (n+1) \text{ terms} \right].$$

This is the **general quadrature formula**. A number of formulae can be deduced from this by putting $n = 1, 2, 3, ...$

(i)Trapezoidal Rule

We put $n = 1$ in the general quadrature formula and neglect second and higher differences. We get

$$\int_{x_0}^{x_0+h} y\,dx = h\left[y_0 + \frac{1}{2}\Delta y_0 \right] = h\left[y_0 + \frac{y_1 - y_0}{2} \right] = h\left[\frac{y_0 + y_1}{2} \right].$$

Similarly,

$$\int_{x_0+h}^{x_0+2h} y\,dx = h\left[\frac{y_1 + y_2}{2} \right],$$

$$\vdots,$$

$$\int_{x_0+(n-1)h}^{x_0+nh} y\,dx = h\left[\frac{y_{n-1} + y_n}{2} \right].$$

Adding these n integrals, we obtain

$$\int_{x_0}^{x_0+nh} y\,dx = h\left[\frac{y_0 + y_1}{2} \right] + h\left[\frac{y_1 + y_2}{2} \right] + ... + h\left[\frac{y_{n-1} + y_n}{2} \right]$$

$$= h\left[\frac{1}{2}(y_0 + y_n) + (y_1 + y_2 + ... + y_{n-1}) \right].$$

This is equal to "distance between two consecutive ordinates × [means of the first and the last ordinate + sum of all the intermediate ordinates]".

This rule is called as the **Trapezoidal rule**.

(ii) Simpson's One-third Rule

We put $n = 2$ in the general quadrature formula and neglect third and higher differences, we get

$$\int_{x_0}^{x_0+2h} y\,dx = h\left[2y_0 + 2\Delta y_0 + \left(\frac{8}{3} - 2\right)\frac{\Delta^2 y_0}{2!}\right]$$

$$= h\left[2y_0 + 2(y_1 - y_0) + \frac{3}{2}\left(\frac{y_2 - 2y_1 + y_0}{2}\right)\right]$$

$$= h\left[2y_1 + \frac{1}{3}(y_2 - 2y_1 + y_0)\right]$$

$$\int_{x_0}^{x_0+2h} y\,dx = \frac{h}{3}(y_0 + 4y_1 + y_2).$$

Similarly,

$$\int_{x_0+2h}^{x_0+4h} y\,dx = \frac{h}{3}(y_2 + 4y_3 + y_4),$$

$$\vdots,$$

$$\int_{x_0+(n-2)h}^{x_0+nh} y\,dx = \frac{h}{3}(y_{n-2} + 4y_{n-1} + y_n).$$

Adding all these integrals, we have when n is

even $\int_{x_0}^{x_0+nh} y\,dx = \frac{h}{3}(y_0 + y_n) + 4(y_1 + y_3 + ... + y_{n-1}) + 2(y_2 + y_4 + ... + y_{n-2})$.

This formula is known as **Simpson's one-third rule**.

(iii) Simpson's Three-eighth Rule

We put $n = 3$ in the general quadrature formula and neglect fourth and higher differences, we get

$$\int_{x_0}^{x_0+3h} y\,dx = h\left[3y_0 + \frac{3^2}{2}\Delta y_0 + \left(\frac{3^3}{3} - \frac{3^2}{2}\right)\frac{\Delta^2 y_0}{2!} + \left(\frac{3^4}{4} - 3^3 + 3^2\right)\frac{\Delta^3 y_0}{3!}\right]$$

$$= h\left[3y_0 + \frac{9}{2}(y_1 - y_0) + \frac{9}{4}(y_2 - 2y_1 + y_0) + \frac{3}{8}(y_3 - 3y_2 + 3y_1 - y_0)\right]$$

$$\int_{x_0}^{x_0+3h} y\,dx = h\left[\frac{3}{8}y_0 + \frac{9}{8}y_1 + \frac{9}{8}y_2 + \frac{3}{8}y_3\right] = \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + y_3)\ .$$

Similarly,

$$\int_{x_0+3h}^{x_0+6h} y\,dx = \frac{3h}{8}[y_3 + 3y_4 + 3y_5 + y_6].$$

Adding such expressions as these form x_0 to x_n where n is a multiple of 3, we have

$$\int_{x_0}^{x_0+nh} y\,dx = \frac{3h}{8}[(y_0+y_n)+3(y_1+y_2+y_4+y_5+...+y_{n-1}) +2(y_3+y_6+...+y_{n-3})].$$

This is known as **Simpson's three-eight rule**.

(iv) Weddle's Rule

Put $n=6$ in the general quadrature formula and neglect all differences above the sixth, we obtain

$$\int_{x_0}^{x_0+6h} ydx = h\left[6y_0+18\Delta y_0+27\Delta^2 y_0+24\Delta^3 y_0+\frac{123}{10}\Delta^4 y_0 +\frac{33}{10}\Delta^5 y_0+\frac{41}{140}\Delta^6 y_0\right].$$

If we replace the last term by $\frac{3}{10}\Delta^6 y_0$, the error mode is $\frac{h}{140}\Delta^6 y_0$ which will be negligible if the value of h is such that the sixth differences are small. With this assumptions, we have on replacing the last term by $\frac{3}{10}\Delta^6 y_0$,

$$\begin{aligned}\int_{x_0}^{x_0+6h} ydx &= h[6y_0+18(y_1-y_0)+27(y_2-2y_1+y_0)+24(y_3-3y_2+3y_1-y_0)\\ &\quad+\frac{123}{10}(y_4-4y_3+6y_2-4y_1+y_0)+\frac{33}{10}(y_5-5y_4+10y_3-10y_2+5y_1-y_0)\\ &\quad+\frac{3}{10}(y_6-6y_5+15y_4-20y_3+15y_2-6y_1+y_0)]\\ &= h\left[\frac{3}{10}y_0+\frac{15}{10}y_1+\frac{3}{10}y_2+\frac{18}{10}y_3+\frac{3}{10}y_4+\frac{15}{10}y_5+\frac{3}{10}y_6\right]\\ &= \frac{3h}{10}[y_0+5y_1+y_2+6y_3+y_4+5y_5+y_6].\end{aligned}$$

Similarly,

$$\int_{x_0+6h}^{x_0+12h} y\,dx = \frac{3h}{10}[y_6+5y_7+y_8+6y_9+y_{10}+5y_{11}+y_{12}],$$

$$\vdots,$$

$$\int_{x_0+(n-6)h}^{x_0+nh} y\,dx = \frac{3h}{10}[y_{n-6}+5y_{n-5}+y_{n-4}+6y_{n-3}+y_{n-2}+5y_{n-1}+y_n].$$

Adding all these integrals, we have if n is a multiple of 6,

$$\int_{x_0}^{x_0+nh} y\,dx = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + 2y_6 + 5y_7 + y_8 + ...].$$

This formula is known as **Weddle's rule**.

Some Examples for Approximate Quadrature

(A) Example

We will evaluate $\int_4^{5.2} \log x\ dx$.

Let $I = \int_4^{5.2} \log x\ dx$.

Then $f(x) = \log x$.

Since $h = 0.2$, we have the following table for the values of x and y.

Table 2

i	0	1	2	3	4	5	6
x_i	4.0	4.2	4.4	4.6	4.8	5.0	5.2
y_i	1.3863	1.4351	1.4816	1.5261	1.5686	1.6094	1.6487

(i) By Trapezoidal rule, we have

$$I = \frac{h}{2}(y_0 + y_6) + 2(y_1 + y_2 + y_3 + y_4 + y_5)$$

$$= \frac{0.2}{2}[(1.3863 + 1.6487) + 2(1.4351 + 1.4816 + 1.5261 + 1.5686 + 1.6094)]$$

$$= 1.82766.$$

(ii) By Simpson's one-third rule, we have

$$I = \frac{h}{3}(y_0 + y_6) + 4(y_1 + y_3 + y_5) + 2(y_2 + y_4)$$

$$= \frac{0.2}{3}[(1.3863 + 1.6487) + 4(1.4351 + 1.5261 + 1.6094) + 2(1.4816 + 1.5686)]$$

$$= 1.82785.$$

(iii) By Simpson's three-eight rule, we have

$$I = \frac{3h}{8}(y_0 + y_6) + 3(y_1 + y_2 + y_4 + y_5) + 2y_3$$

$$= \frac{3(0.2)}{8}[(1.3863 + 1.6487) + 3(1.4351 + 1.4816 + 1.5686 + 1.6094) + 2(1.5261)]$$

$$= 1.82785.$$

(iv) By Weddle's rule, we have

$$I = \frac{3h}{10}\ y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6)$$
$$= \frac{3(0.2)}{10}[1.3863 + 5(1.4351) + 1.4816 + 6(1.5261) + 1.5686 + 5(1.6094) + 1.6487]$$
$$= 1.827858.$$

(B) Example

We will find the value of $\int_0^6 f(x)\,dx$, using four kinds of approximate quadrature formulae from the following data.

Table 3

x_i	0	1	2	3	4	5	6
$y_i = f(x_i)$	6.9897	7.4036	7.7815	8.1291	8.4510	8.7506	9.0309

(i) By using Trapezoidal rule,

$$\int_0^6 f(x)\,dx = h\left[\frac{1}{2}(y_0+y_6)+(y_1+y_2+y_3+y_4+y_5)\right]$$
$$= 1\left[\frac{1}{2}(6.9897+9.0309)+(7.4036+7.7815+8.1291+8.4510+8.7506)\right]$$
$$= 48.5261.$$

(ii) By using Simpson's one-third rule,

$$\int_0^6 f(x)\,dx = \frac{h}{3}\left[(y_0+y_6)+4(y_1+y_3+y_5)+2(y_2+y_4)\right]$$
$$= \frac{1}{3}\left[(6.9897+9.0309)+4(7.4036+8.1291+8.7506)+2(7.7815+8.4510)\right]$$
$$= 48.5396.$$

(iii) By using Simpson's three-eighth rule,

$$\int_0^6 f(x)\,dx = \frac{3h}{8}\left[(y_0+y_6)+3(y_1+y_2+y_4+y_5)+2y_3\right]$$
$$= \frac{3(1)}{8}\left[(6.9897+9.0309)+3(7.4036+7.7815+8.4510+8.7506)+2(8.1291)\right]$$
$$= 48.5395875.$$

(iv) By using Weddle's rule,

$$\int_0^6 f(x)\,dx = \frac{3h}{10}\left[y_0+5y_1+y_2+6y_3+y_4+5y_5+y_6\right]$$

$$= \frac{3(1)}{10}\left[6.9897+5(7.4036)+7.7815+6(8.1291)+8.4510+5(8.7506)+9.0309\right]$$

$$= 48.53961.$$

Acknowledgements

We are greatly thankful to Dr Aung Kyaw Thin, Rector and Dr Aye Aye Han, Pro-Rector of Banmaw University for their permission to write in this journal. We would like to express our thanks to Dr Khin San Aye, Professor and Head, Department of Mathematics, Banmaw University for her kind encouragement to submit this paper. We would like to thank Professor Dr Hnin Oo Lwin, Department of Mathematics, Banmaw University for her helpful advice to do this paper.

References

Sastry, S.S., "**Introductory Method of Numerical Analysis"**, Third Edition, New Delhi, (1999).

Saxena, H.C., "**Finite Differences and Numerical Analysis"**, Chand& Company Ltd., New Delhi, Ind, (2003)

Numerical Techniques for Unconstrained Optimization

Su May Win*

Abstract

In this paper, the optimization problem in standard form is introduced. The real problem is transformed to the mathematical model. Then, the constraint problem and unconstraint problem are illustrated. Especially, the unconstraint problems are solved.

Key Words: Kuhn-Tucker conditions, unconstraint problem, matlab.

Introduction

Optimization is an essential part of design activity in all major disciplines. These disciplines are not restricted to engineering. In product development, competition demands producing economically relevant products with embedded quality. Today, globalization demands that additional dimensions such as location, language and expertise must also merit consideration as new constraints in the development process. Improved production and design tools coupled with inexpensive computational resources have made optimization an important part of the process. Even in the absence of a tangible product, optimization ideas provide the ability to define and explore problems.

The standard form

These definitions allow us to assemble the general mathematical model for optimization through the various design functions:

$$\begin{aligned}
&\text{Minimize} && f(x_1, x_2, \ldots, x_n)\\
&\text{Subject to:} && h_1(x_1, x_2, \ldots, x_n) = 0\\
& && h_2(x_1, x_2, \ldots, x_n) = 0\\
& && \vdots\\
& && h_l(x_1, x_2, \ldots, x_n) = 0\\
& && g_1(x_1, x_2, \ldots, x_n) \le 0\\
& && g_2(x_1, x_2, \ldots, x_n) \le 0\\
& && \vdots\\
& && g_m(x_1, x_2, \ldots, x_n) \le 0\\
& && x_i^l \le x_i \le x_i^u \;;\; i = 1, 2, \ldots, n.
\end{aligned}$$

The same problem can be expressed concisely using this notation:

$$\text{Minimize} \quad f(x_1, x_2, \ldots, x_n)$$

* Lecturer, Department of Mathematics, Banmaw University

Subject to : $h_k(x_1, x_2, \ldots, x_n) = 0, \quad k = 1, 2, \ldots, l$

$g_j(x_1, x_2, \ldots, x_n) \leq 0, \quad j = 1, 2, \ldots, m$

$x_i^l \leq x_i \leq x_i^u \; ; \; i = 1, 2, \ldots, n.$

The general model in vector notation (length of vector is shown in the definition):

Minimize $f(X);[X]_n$

Subject to : $[h(X)]_l = 0$

$[g(X)]_m \leq 0$

$X^{\text{low}} \leq X \leq X^{\text{up}}$.

The previous mathematical model expresses the standard optimization problem in natural language as:

Minimize the objective function *f*, subject to *l* equality constraints, m inequality constraints, with the *n* design *x* variables lying between prescribed lower and upper limits.

The techniques will apply to the problem described in the standard form which refers to a single objective function. To solve any specific design problem, it is required to reformulate the problem in this manner so that the methods can be applied directly. The techniques are developed progressively, starting from the standard model without constraints. For example, the unconstrained problem will be explored first. The equality constraints are considered next followed by the inequality constraints and finally, the complete model. This represents a natural progression, as prior information is used to develop additional conditions that need to be satisfied by the solution in these instances.

Mathematical modeling

In this section, the two design problems will be translated to the standard form after first identifying the mathematical model. The second problem requires information from a course in mechanics and materials. This should be within the scope of most engineering students.

Example

New consumer research with deference to the obesity problem among the general population, suggests that people should drink no more than about 0.25 liter of soda pop at a time. The fabrication cost of the redesigned soda can is proportional to the surface area and can be estimated at $ 1.00 per square centimeter of the material used. A circular cross-section is the most plausible, given current tooling available for manufacture. For aesthetic reason, the height must be at least twice the diameter. Studies indicate that holding comfort requires a diameter between 5 and 8 cm. Create a design that will cost the least.

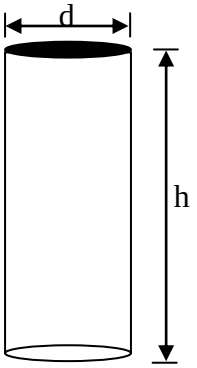

Figure 1. Design of a new beverage can.

The diameter *d* and the height *h* in the cylindrical are sufficient to describe the soda can. What about the thickness *t* of the material of the can? What are the assumptions for the design problem? Is *t* small enough to be ignored in the calculation of the volume of soda in the can? Another important assumption could be that the can will be made using a given stock roll. Another one is that the material required for the can will include only the cylindrical surface area and the area of the bottom. The top will be fitted with an end cap that will provide the mechanism by which the soda can be poured. The top is not part of this design problem. In the first attempt at developing the mathematical, we could start out by considering the quantities identified, including the thickness, as design variables:

Design variables: *d, h, t.*

Reviewing the statement of the design problem, one of the parameters is the cost of material per unit area that is given as $ 1.00 per square meter. Let us identify the cost of material per unit area as constant *C*. During the search for the optimal solution this quantity will be held at the given value. That this value changes then our cost of the can will correspondingly change. This is what we mean by a design parameter. Typically, change in parameters will require a new solution to the optimization problem:

Design parameter : *C.*

The design functions will include the computation of the volume enclosed by the can and the surface area of the material used. The volume in the can is $\pi d^2 h/4$. The surface area is $\pi d\, h + \pi d^2/4$. The aesthetic constraint requires that $h \geq 2d$. We can formally set up the optimization problem:

Minimize $f(d, h, t): C(\pi d\, h + \pi d^2/4)$ (1)

Subject to : $h_1(d, h, t): \pi d^2 h/4 - 250 = 0$ (2)

$g_1(d, h, t): 2d - h \leq 0$ (3)

$5 \leq \mathrm{d} \leq 8;\ \ 4 \leq \mathrm{h} \leq 20;\ \ \ 0.001 \leq \mathrm{t} \leq 0.01.$

In the mathematical model of the optimization problem for the first example the values of the design variables are expected to be expressed in consistent units-centimeters. It is the responsibility of the designer to ensure correct and accurate problem formulation including dimensions and units. The cost *C* was originally expressed as meter squared. Hence a scaling factor must be used in (1). We will call this C_1.

Intuitively, there is some concern with the problem as expressed by the equations even though the description is valid. How can the value of the design variable *t* be established? The variation in *t* does not affect the design. Changing the value of *t* does not change f, h_1 or g_1. Hence it cannot be a design variable. If this were a serious design example then the cans have to be designed for impact, stacking strength, and stresses occurring during transportation and handling. In that case *t* will probably be a critical design variable. This will require several additional structural constraints in the problem. Moreover, it is likely that development of these functions will not be a simple exercise. This could serve as an interesting extension to this problem for homework or project. The new mathematical model for the optimization problem after dropping *t*, and expressing [*d, h*] as [x_1 , x_2] becomes:

Minimize $f(x_1, x_2): C_1(\pi x_1 x_2 + \pi x_1^2/4)$

Subject to: $h_1(x_1, x_2): \pi x_1^2 x_2/4 - 250 = 0$

$g_1(x_1, x_2): 2x_1 - x_2 \le 0$

$5 \le x_1 \le 8; \quad 4 \le x_2 \le 20.$

The problem represented by Equations is the mathematical model for the design problem expressed in the standard form. For this problem, simple geometrical relations were sufficient to set up the optimization problem.

Example

My PC Company has decided to invest $ 12 million in acquiring several new component placement machines to manufacture different kinds of motherboards for a new generation of personal computers. Three models of these machines are under consideration. Total number of operators available is 100 because of the local labor market. A floor-space constraint needs to be satisfied because of the different dimensions of these machines. Additional information relating to each of the machines is given in Table 1. The company wishes to determine how many of each kind is appropriate to maximize the number of boards manufactured per day.

Table 1. Component Placement Machines

Machine Model	Board Types	Boards/ Hour	Operators/ Shift	Operable Hours/Day	Cost/ Machine
A	10	55	1	18	400,000
B	20	50	2	18	600,000
C	18	50	2	21	700,000

The number of machines of each model needs to be determined. Let x_1 represent the number of component placement machines of model A. Similarly, x_2 will be associated with model B, and x_3 with model *C*.

Design variables: x_1, x_2, x_3.

The values in the tables can be regarded as parameters. For this problem, it is more useful to work with the values directly and hence identifying them a parameters does not serve any useful purpose. The information in Table 1 is used to set up the design functions in terms of the design variables directly. An assumption is made that all machines are run for three

shifts. The utilization of each machine is the number of boards per hour times the number of hours the machine operates per day. The optimization problem can be assembled in the following form:

Maximize $f(X): 990\,x_1 + 900\,x_2 + 1050\,x_3$

or

Minimize $-f(X): -990\,x_1 - 900\,x_2 - 1050\,x_3$

Subject to: $g_1(X): 400{,}000x_1 + 600{,}000x_2 + 700{,}000x_3 \le 12{,}000{,}000$

$g_2(X): 3x_1 - x_2 + x_3 \le 30$

$g_3(X): 3x_1 + 6x_2 + 6x_3 \le 100$

$x_1 \ge 0; \quad x_2 \ge 0; \quad x_3 \ge 0.$

Equations express the mathematical model of the problem. In this problem there is no product being designed. Here, a strategy for placing order for the number of machines is being determined.

Unconstrained optimization

This section illustrates many numerical techniques for multivariable unconstrained optimization. Although unconstrained optimization is not a common occurrence in engineering design, nevertheless the numerical techniques included here demonstrate interesting ideas and also capture some of the early intensive work in the area of design optimization.

Problem definition

The unconstrained optimization problem requires only the objective function. This paper has chosen to emphasize an accompanying set of side constraints to restrict the solution to an acceptable design space/region. It is the responsibility of the designer to include the side constraints as part of the exploration of the optimum. The problem can be defined as follows:

Minimize $f(X); [X]_n$

Subject to : $x_i^{\ell} \le x_i \le x_i^{u}$; $i = 1, 2, \ldots, n.$

Example

We will follow a nontrivial problem with both minimum and maximum present within the design space, typical of real problems, so that we can get feedback if there is any problem with the technique. Real problems are not usually well behaved and will require intervention, even with a good code.

Minimize $f(X) = f(x_1, x_2) = 3 \sin(0.5 + 0.25\, x_1 x_2) \cdot \cos(x_1)$

Subject to : $0 \le x_1 \le 5;\ 0 \le x_2 \le 8$;

Necessary and sufficient conditions

The necessary and sufficient conditions for unconstrained problem also called **the Kuhn-Tucker conditions** in the above example require the gradients must vanish at the solution :

$$\frac{\partial f}{\partial x_1} = 0.75 \cos (0.5 + 0.25\, x_1x_2).\, x_2 \cos x_1 - 3 \sin(0.5 + 0.25\, x_1x_2) \sin x_1 = 0 \qquad (4)$$

$$\frac{\partial f}{\partial x_2} = 0.75 \cos (0.5 + 0.25\, x_1x_2) .\, x_1 \cos x_1 = 0. \qquad (5)$$

Equations look difficult to solve. A little persistence can be rewarding. Three possible solutions to the problem can be deduced from (5)

(i) $x_1 = 0$

(ii) $\cos x_1 = 0$

(iii) $\cos (0.5 + 0.25\, x_1x_2) = 0.$

We will consider

$0.5 + 0.25\, x_1x_2 = \pm$ n $\frac{\pi}{2}$, where n is integers.

Substituting in (4) will require

$x_1 = \pm\, m\pi,$

where m is integers. For a solution in the positive quadrant;

$$x_1 = \pi = 3.1416\ ;\ x_2 = \frac{0.5\pi - 0.5}{0.25\pi} = 1.3633.$$

Example

We consider that use of Kuhn-Tucker conditions

Minimize : $f(X) = f(x_1, x_2) = 3 + (x_1 - 1.5\, x_2)^2 + (x_2 - 2)^2$

Subject to : $0 \le x_1 \le 5,\ \ 0 \le x_2 \le 5;$

We calculate

$$\frac{\partial f}{\partial x_1} = 2x_1 - 3x_2 = 0$$

$$\frac{\partial f}{\partial x_2} = -3x_1 + 6.5x_2 - 4 = 0.$$

We get $x_1 = 3, x_2 = 2$

$$f(x_1, x_2) = 3 + (x_1 - 1.5\, x_2)^2 + (x_2 - 2)^2$$

$$f(3, 2) = 3.$$

This solution will be verified through MATLAB using symbolic and numerical calculations. The nearly exact solution is red spots.

MATLAB Code

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Example 2.1.3
% Numerical Techniques for Unconstrained Optimization
% Optimzation with MATLAB, Section 2.1
% symbolic calculation
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
syms xx1 xx2 ff gradx1 gradx2 hess
ff = 3 + (xx1 - 1.5*xx2)*(xx1 -1.5*xx2) + (xx2 -2 )*(xx2 - 2);
% gradients
gradx1 = diff(ff,xx1);
gradx2 = diff(ff,xx2);
% solution to FOC
optimum = solve(gradx1, gradx2);
fprintf('FOC --- \n')
fprintf('x1* = '),disp(double(optimum.xx1))
fprintf('x2* = '),disp(double(optimum.xx2))
% applying second order condition
fprintf('\n\nSOC -----')
hess=[diff(gradx1,xx1)  diff(gradx1,xx2); diff(gradx2,xx1) diff(gradx2,xx2)];
fprintf('\nHessian \n '),disp(hess)
% calculate eigen value
evalue = eig(hess);
fprintf('\neigen value \n'),disp(evalue)
% graphical optimization
x1=0:0.1:5;
x2 = 0:0.1:5;
[X1 X2] = meshgrid(x1,x2);
f = 3 +(X1 - 1.5*X2).*(X1 - 1.5*X2) + (X2- 2).^2;
c1 = contour(x1,x2,f, ...
[3.1 3.25 3.5 4 6 10 15 20 25],'k');
clabel(c1);
grid
xlabel('x_1')
ylabel('x_2')
title('Example 2.1.3')
```

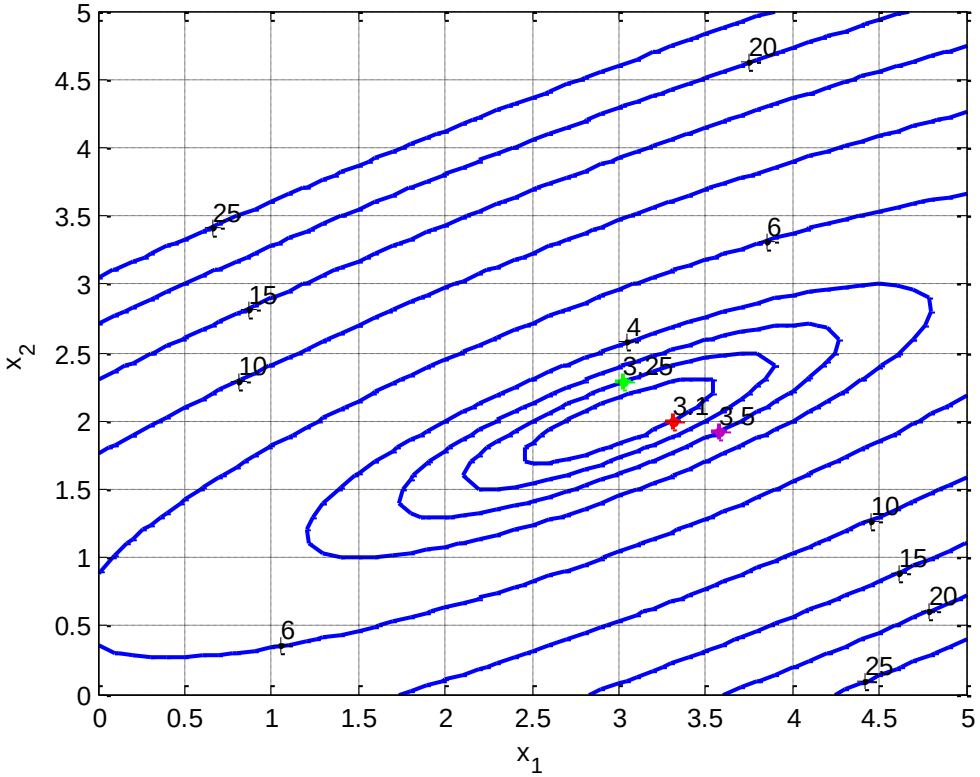

Figure 2. Graphical Solution.

Conclusion

We can construct the real problem in mathematical model and solve the problem that we construct. We get the graphical solution of the problem by using MATLAB. It is noted that the numerical solution and graphical solution are the same.

Acknowledgement

I would like to express my thanks to Dr Aung Kyaw Thin, Rector, Banmaw University, for his permission and encouragement during the course of the research work. I am also grateful to Dr Aye Aye Han, Pro-rector, Banmaw University, for her encouragement to submit this paper. I also thank to Professor Dr Khin San Aye, Head of the Department of Mathematics, Professor Dr Hnin Oo Lwin and my seniors, Department of Mathematics, Banmaw University who give me their invaluable advices.

References

Bronson, R., *Operations Research*, SOS, McGraw-Hill, 1983

Venkataraman P, *Applied Optimization with Matlab Programming*, Second Edition John Wiley & Sons, Inc., 2009

Application of Euler Cycle

Swe Mon Oo[1], Hnin Yee San[2]

Abstract

In this paper,some definitions of graph theory are presented. And then definitions, examples and theorems related to Euler cycle are discussed. Finally, postman problem for weighted graph is solved by using the algorithm.

Basic definitions and concepts of graph theory

A **graph** $G=(V,E)$ consists of two sets: a finite nonempty set V of elements called **vertices** and a finite set E of elements called **edges**. The vertex set of G is denoted by **V(G)**, while the edge set of G is denoted by **E(G)**. The edge $e=(u,v)$ or uv is said to **join** the vertices u and v. If uv is an edge of a graph G, then u and v are **adjacent** vertices, while u and e are **incident** as are v and e. The distinct edges associated with the same pair of vertices are called **parallel edges**. The cardinality of the vertex set of a graph G is called the **order** of G, which is denoted by $\nu(G)$ and the cardinality of its edges set is **size** of G, which is denoted by $\varepsilon(G)$. The **degree** of a vertex v in a graph G is the number of edges of G incident with v, which is denoted by d(v). The graph $G=(V,E)$ is a **subgraph** of the graph $G'=(V',E')$ if and only if $V\subseteq V'$ and $E\subseteq E'$. Let $G=(V,E)$ be a graph. With edge e of G let there be associated a real number, called its **weight**. Then G together with weights on its edges is called a **weighted graph**. Let u and v be vertices of a graph G. A **u-v walk** W of G is a finite, alternating sequence $W: u=u_0,e_1,u_1,e_2,\ldots,u_{k-1},e_k,u_k=v$ of vertices and edges, beginning with vertex u and ending with vertex v, such that $e_1=u_{i-1}u_i$ for $i=1,2,\ldots,k$. The number k (the number of occurrences of edges) is called the **length** of W. A **trivial walk** contains no edges. A u-v walk is **closed** or **open** depending on whether $u=v$ or $u\neq v$. **A u-v trail** is a u-v walk in which no edge is repeated, while **u-v path** is a u-v walk in which no vertex is repeated. A path with n vertices is denoted by P_n. A closed trail is a **cycle (circuit)** if all its vertices except the end vertices are distinct. A cycle with n vertices is denoted by C_n. A graph G is **connected** if there exists u-v path in G for every two vertices u and v in G. A graph that is not connected is **disconnected**. For a connected graph G, **the distance d(u, v)** between two vertices u and v as the minimum of the lengths of the u-v path of G.

A simple path in a graph G is called **Euler path** if it is traverse every edge of a graph exactly once. **An Euler cycle** (or circuit) is a cycle that traverses every edge of a graph exactly once. An **Euler trail** in a graph G is a closed trail containing all the edges of G. A graph which contains either Euler path or Euler circuit is called **Eulerian** graph.

1Assistant Lecturer, Department of Mathematics, Banmaw University

2Tutor, Department of Mathematics, Banmaw University

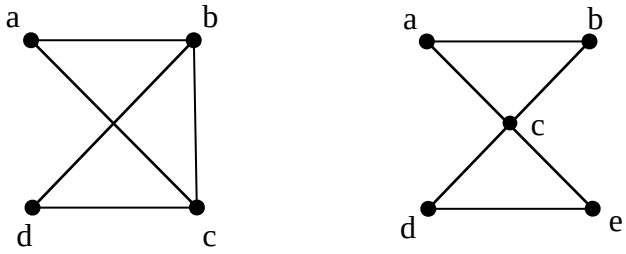

Figure 1

In Figure 1, the left graph has an Euler path b, a, c, d, b, c and the right graph has an Euler cycle a, c, d, e, c, b, a.

Properties of Euler Cycle

Theorem

A non-empty connected graph is Eulerian if and only if it has no vertices of odd degree.

Proof

Let G be Eulerian, and C be an Euler cycle of G with origin (and terminus) u. Each time a vertex v occurs as an internal vertex of C, two of the edges incident with v are accounted for. Since an Euler cycle contains every edge of G, d(v) is even for all $v \neq u$. Similarly, since C starts and ends at u, d(u) is also even. Thus G has each vertex is of even degree.

Conversely, suppose that G is a non-Eulerian connected graph with at least one edge and no vertices of odd degree. Choose such a graph G with as few edges as possible. Since each vertex of G has degree at least two, G contains a closed trail. Let C be a closed trail of maximum possible length in G. By assumption, C is not an Euler cycle of G and so $G - E(C)$ has some component G' with $\varepsilon(G') > 0$. Since C is itself Eulerian, it has no vertices of odd degree, thus the connected graph G' also has no vertices of odd degree. Since $\varepsilon(G') < \varepsilon(G)$, it follows from the choice of G that G' has an Euler cycle C'. Now, because G is connected, there is a vertex v in $V(C) \cap V(C')$, and we may assume, without loss of generality, that v is the origin and terminus of both C and C'. But the CC' is closed trail of G with $\varepsilon(CC') > \varepsilon(C)$, contradicting the choice of C.

Theorem

A connected graph has an Euler trail if and only if it has exactly two vertices of odd degree.

Proof

If G has an Euler trail, then as in the proof of above theorem, each vertex other than the origin and terminus of trail has even degree. Conversely, suppose that G is a nontrivial connected graph with at most two vertices of odd degree. If G has no such vertices then, by above theorem, G has a closed Euler path. Otherwise, G has exactly two vertices u and v, of odd degree. In this case, let $G + e$ denote the graph obtained from G by the addition of a new edge e joining u and v. Clearly, each vertex of $G + e$ has an Euler path $v_0 e_1 v_1 \ldots e_{\varepsilon+1} v_{\varepsilon+1}$, where $e_1 = e$. The trail $v_1 e_2 v_2 \ldots e_{\varepsilon+1} v_{\varepsilon+1}$ is an Euler trail of G.

Constructive algorithm

Constructive algorithm used to the prove Euler's theorem and to find an Euler cycle or path in an Eulerian graph.A graph with two vertices of odd degree. The graph with its edges labeled according to their order of appearance in the path found. Steps that kept in mind while traversing Euler graph are first to choose any vertex u of G. Start traversing through edges that no visited yet until a cycle is formed. Record the cycle and remove the edges it consists of. If there is an unvisited edge's then start the first step until whole graph is traversed.

Two cycles emerged by traversing one of them and insert other when common vertex found. The result is a new cycle. For an Eulerian graph that must contain two vertices with odd degree. Otherwise no Euler path can be found. Start from a vertex of odd degree and thus ensure that every vertex has an even degree.

Example: Illustrations of constructive algorithm to find Euler cycle, consider the graph.

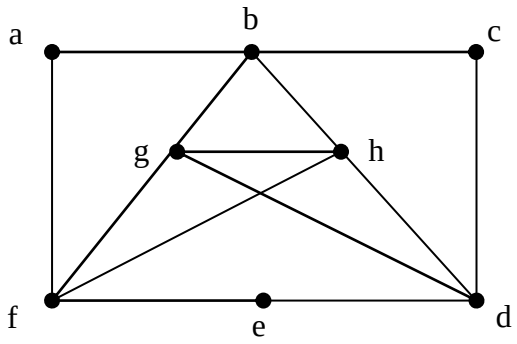

Figure 2

Step 1: Check to start that the graph is connected and all vertices are of even degree, findcycle.

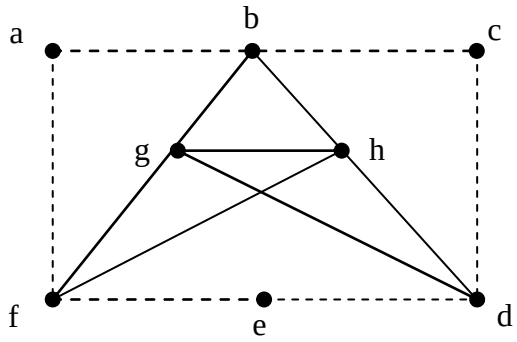

Figure 3

In Figure 3, cycle of length 6 such as f, a, b, c, d, e, f.

Step 2: Remove the edges involved in this cycle from the graph.

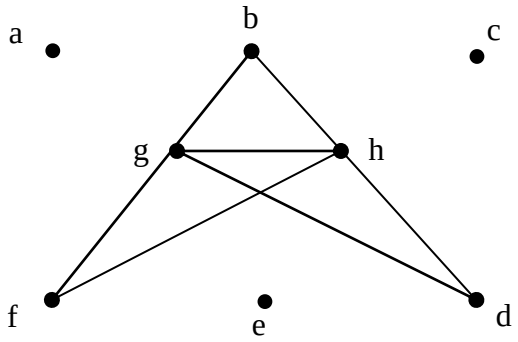

Figure 4

In the 'smaller' graph that remains, the vertices must still all be of even degree. Cycle of length 7 such as f, g, b, h, d, g, h, f.

Step 3: Merge the cycles from step 2 into the cycle in step 1 at appropriate points.
Cycle of length 13 such as f, a, b, c, d, e, f, g, b, h, d, g, h, f.

Example: An Euler trail exists for the graph in Figure 5.

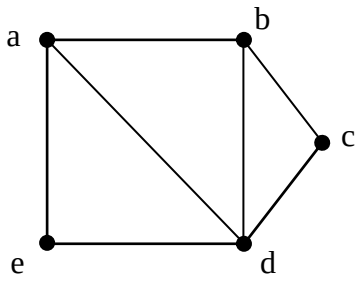

Figure 5

Euler trail is a, e, d, c, b, d, a, b.

An algorithm for constructing an Euler path is the following:

1. If the graph is connected and all its vertices are of even degree then construct an Euler cycle (necessarily it's an Euler trail). Otherwise, if the graph is connected and has exactly two vertices are of odd degree, identify those vertices as the initial vertices and end vertices of the Euler trail to be constructed and remove the edges along a trail joining them. Find an Euler cycle in what remains.
2. If the cycle obtained is written using its initial vertices, then the edges of the trail can simply be added to the end of the cycle to give an Euler trail for the original graph.

Example: An Euler cycle for the graph same as in Figure 5 can be found by Constructive algorithm.

Noting that $d(a) = 3$ and $d(b) = 3$ identifies these as the start and end of the trail. The edge(a,b) (which acts a trail from a to b) is thus removed from the graph to start leaving the 'new' graph shown below:

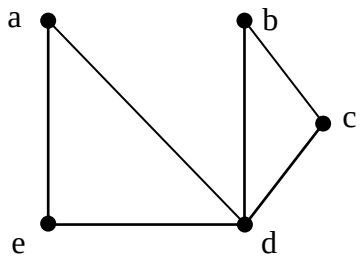

Figure 6

Cycle of length 6 such as a, e, d, c, b, d, a. Attaching the edge (a, b) to the end of this cycle gives as an Euler path for the original graph: a, e, d, c, b, d, a, b.

The Postman problem for weighted graph

A part of his duties, a postman starts from his office, visits every street at least once, delivers the mail and comes back to the office. Suggest a route of minimum distance. Algorithm was designed to give the shortest route that was required for any network or road map, including those that had more than two vertices with odd numbers of edges emerging from them. If a road map or network had more than two nodes that were of odd order, the odd order nodes had to be paired together, and the shortest distance found between the pairs. If the network was traversed, these edges would have to be repeated to give the shortest distance to travel all edges and return to the starting point.

The algorithm is as follows:

Step 1: Find all nodes with odd numbers of edges emerging from them.

Step 2: These nodes should be listed in all the possible pairings.

Step 3: Find the shortest distance between each vertex of the pairs, then add up the total for each set of pairs.

Step 4: Choose the set of pairs with the minimum total distance. This added on to the sumof all the edges to give the shortest distance to traverse the network and return to the original point of entry.

Example: A postman has to go around the following route starting and finishing at A, his postal depot. He has to go along each road, shown as edges, once and only once in graph.

Let $A =$ Postal Depot,

$B =$ Thiri Market,

$C =$ CB Bank,

$D =$ Star Hotel,

$E =$ Win Electronic Centre,

$F =$ Fisheries Shop.

Figure 8 is a weighted graph. Where weight defined between two vertices.

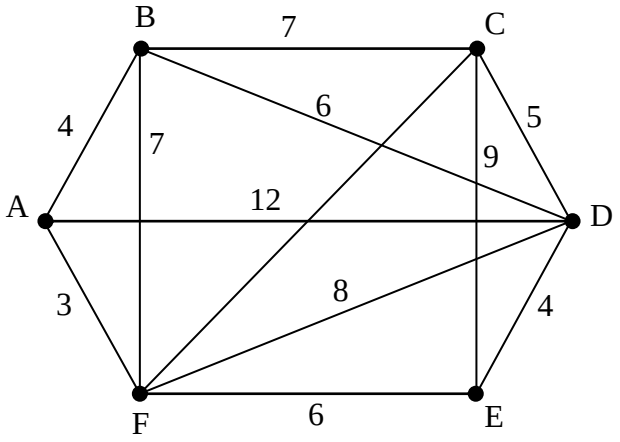

Figure 7

Following the steps of the route inspection algorithm:

Step 1:

Odd Vertices	Degree
A	3
D	5
E	3
F	5

Step 2 & 3: Possible pairings of odd vertices AD and EF shortest route AD and EFdistance $12+6=18$.

Possible pairings of odd vertices AE and DF shortest route AFE and DF distance $3+6+8=17$.

Possible pairings of odd vertices AF and DE shortest route AF and DE distance $3+4=7$.

Step 4: From the above calculations it is clear that to create an Eulerian graph that is the shortest possible edges AF and DE must be repeated.

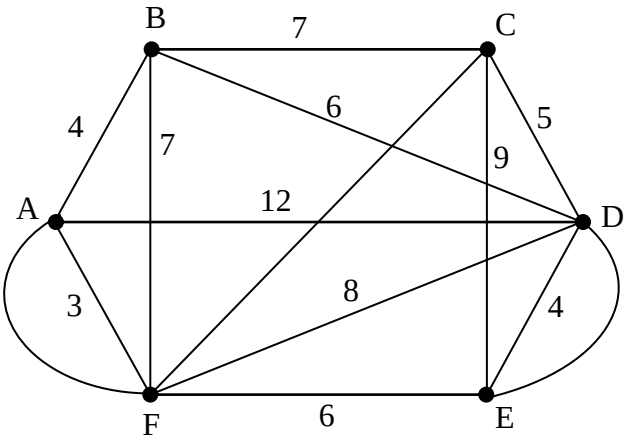

Figure 8

A possible Euler cycle that starting and finishing at A:

A, F, E ,D, C, B, A, D, E, C, F, D, B, F, A.

The total length of this route is

$$3+6+4+5+7+4+12+7+9+6+8+10+\text{Repeated edges}(4+3)=78.$$

Acknowledgements

We thank to Dr Aung Kyaw Thin, Rector and Dr Aye Aye Han, Pro-Rector of Banmaw University for the permission to the journal. We would like to express our thanks to Dr Khin San Aye, Professor and Head, Department of Mathematics, Banmaw University for her kind encouragement to submit this paper. We would like to thank Professor Dr Hnin Oo Lwin, Department of Mathematics, Banmaw University for her helpful advice to do this paper.

References

Bondy, J.A, and Murty, U.S.R., "Graph Theory with Applications", Macmillan Press Ltd, London, 1976.

Thulasiraman, K. and Swamy, M.N.S., "Graph: Theory and Alogrithms", John Wiley and Sons, Inc., New York, 1992.

Kumar. A., "A Study on Euler Graph and It's Applications", International Journal of Mathematics Trends and Technology (IJMTT)-Volume 43 Number 1-March 2017.

Applications of Vertex Coloring

Hnin Yee San[1] , Swe Mon Oo[2]

Abstract

In this paper, some definitions and concepts of graph theory are presented. And then definitions, examples and theorems concerned with vertex coloring are stated. Finally, some applications of vertex coloring are discussed with illustrative examples.

Some definitions and concepts of graph theory

A **graph** $G=(V,E)$ consists of a finite nonempty set V, called the set of **vertices** and a set E of unordered pairs of distinct vertices, called the set of **edges**. Two vertices of a graph are **adjacent** if they are joined by an edge. The **degree** of vertex v in a graph denoted by d(v) is the number of edges incident to v. We denote $\delta(G)$ and $\Delta(G)$ by the **minimum and maximum degree** of vertices of G respectively. The edge $e=(u,v)$ or uv is said to **join** the vertices u and v. If e has just one endpoint, e is called a **loop**. If e_1 and e_2 are two different edges that have the same endpoints, then we call e_1 and e_2 are **parallel edges**. A graph is **simple** if it has no loops and parallel edges.

The graph $G'=(V',E')$ is a **subgraph** of the graph $G=(V,E)$ if $V'\subseteq V$ and $E'\subseteq E$. In this case we write $G'\subseteq G$. $G'\subseteq G$, but $G'\neq G$ we write $G'\subset G$ and called G' is a **proper subgraph** of G. The removal of a vertex v_i from a graph G results in that subgraph of G consisting of all vertices of G except v_i and all edges not incident with v_i. This subgraph is denoted by $G-v_i$. Thus $G-v_i$ is the **maximal subgraph** of G not containing v_i.

A **walk** in G is a finite non-null sequence $w=v_0e_1v_1e_2v_2\ldots e_kv_k$, whose terms are alternately vertices and edges, such that, for $1\le i\le k$, the ends of e_i are v_{i-1} and v_i. We say that w is a walk from v_0 to v_k, or a v_0v_k-walk. The vertices v_0 and v_k are called the **origin** and **terminus** of w respectively, and $v_1,v_2,\ldots,v_{k-1}$ are its **internal vertices**. The integer k, the number of edges in it is the **length** of w. A walk is called a **path** if there are no vertex repetitions. A v_0v_k-walk is said to be **closed** if $v_0=v_k$. A closed walk is said to be a **cycle (circuit)** if $k\ge 3$ and $v_0e_1v_1e_2v_2\ldots v_{k-1}$ is a path. A cycle of length k is called a **k-cycle**; a k-cycle is odd or even according as k is odd or even.

Two vertices u and v of G are said to be **connected** if there is a uv-path in G. If there is a uv-path in G for any distinct vertices u and v of G, then G is said to be **connected**, otherwise G is **disconnected**. If vertices u and v are connected in G, the distance between u and v in G, denoted by $d_G(u,v)$, is the length of a shortest uv-path in G. A **bipartite graph** is one whose vertex set can be partitioned into two subsets X and Y, so that each edge has one end in X and one end in Y; such a partition (X, Y) is called a **bipartition** of the graph.

1Tutor, Department of Mathematics, Banmaw University

2Assistant Lecturer, Department of Mathematics, Banmaw University

Vertex Colorings

Definitions

A **vertex coloring** of a graph is an assignment $f : V \to C$ from its vertex set to a codomain set C whose elements are called colors.

For any positive integer k, a **vertex k-coloring** is a vertex-coloring that uses exactly k different colors.

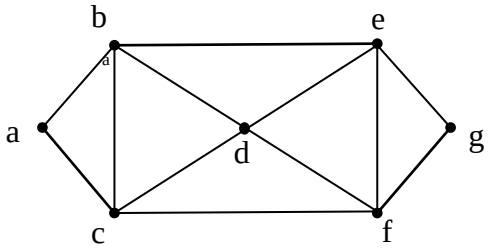

Figure 1 The simple graph G

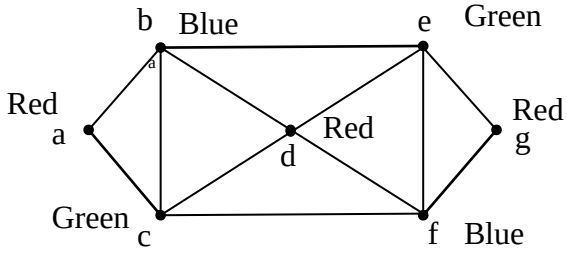

Figure 2 The simple graph G

Definitions

A **proper vertex coloring** of a graph is a vertex-coloring such that the endpoints of each edge are assigned two different colors. A graph is said to be **vertex k-colorable** if it has proper vertex k-coloring.

Example

A proper vertex 4-coloring of a graph G is shown in Figure 3.

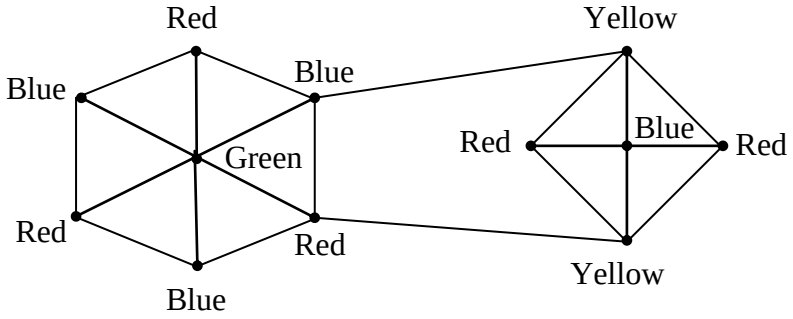

Figure 3 A proper vertex 4-coloring of a graph G

Definition

The **vertex chromatic number** of G, denoted by $\chi(G)$, is the minimum number of different colors required for a proper vertex-coloring of G.

Theorem

A graph is bipartite if and only if its chromatic number is at most 2.

Proof

If the graph G is bipartite, with parts V_1 and V_2, then a coloring of G can be obtained by assigning red to each vertex in V_1 and blue to each vertex in V_2. Since there is no edges from any vertex in V_1 to any other vertex in V_1, and no edge from any vertex in V_2 to any other vertex in V_2, no two adjacent vertices can receive the same color. Therefore, the chromatic number of bipartite graph is at most 2.

Conversely, if we have a 2-coloring of a graph G, let V_1 be the set of vertices that receive color 1 and let V_2 be the set of vertices that receive color 2. Since adjacent vertices must receive different colors, there is no edge between any two vertices in the same part. Thus, the graph is bipartite.

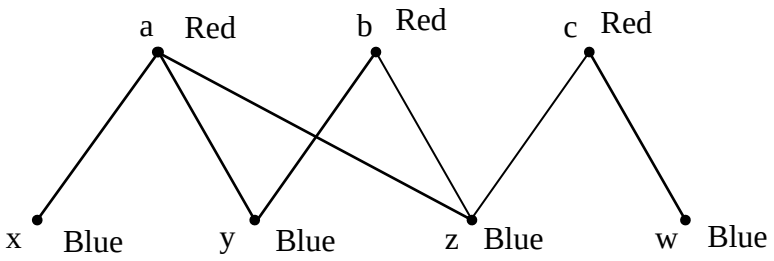

Figure 4 A coloring of bipartite graph

Definition

The **complete graph** on n vertices, for $n \geq 1$, which we denote K_n, is a graph with n vertices and an edge joining every pair of distinct vertices.

Example

We can find the chromatic number of K_n. A coloring of K_n can be constructed using n colors by assigning a different color to each vertex. No two vertices can be assigned the same color, since every two vertices of this graph are adjacent. Hence, the chromatic number of $K_n = n$.

Example

A coloring of K_5 using five colors is shown in Figure 5.

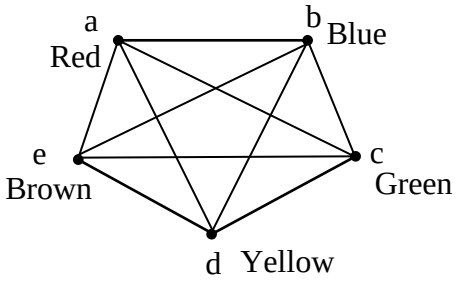

Figure 5 A coloring of K_5

Theorem

If K_n is a subgraph of G, then $\chi(G) \geq n$.

Proof

In any coloring, the vertices of the K_n contained in G must receive different colors, since they are all adjacent to each other. Thus at least n colors are required.

Definition

The **complete bipartite graph** $K_{m,n}$, where m and n are positive integers, is the graph whose vertex set is the union $V = V_1 \cup V_2$ of disjoint sets of cardinalities m and n, respectively, and whose edge set is $\{uv | u \in V_1 \text{ and } v \in V_2\}$.

Example

We can find the chromatic number of the complete bipartite graph $K_{m,n}$, where m and n are positive integers.

The number of colors needed may seem to depend on m and n. However, only two colors are needed. Color the set of m vertices with one color and the set of n vertices with a second color. Since edges connect only a vertex from the set of m vertices and a vertex from the set of n vertices, no two adjacent vertices have the same color.

Example

A coloring of $K_{3,4}$ with two colors is displayed in Figure 6.

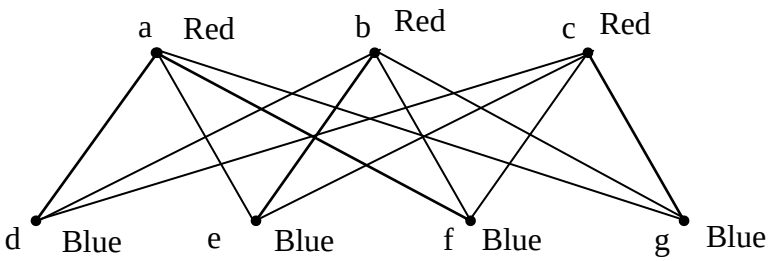

Figure 6 A coloring of $K_{3,4}$

Some applications of Vertex Colorings

Example

Suppose that Mg Aung, Mg Ba, Mg Hla, Mg Kyaw, Mg Lin and Mg Tin are planning a ball. Mg Aung, Mg Chit and Mg Hla constitute the publicity committee. Mg Chit, Mg Hla and Mg Lin are the refreshment committee. Mg Lin and Mg Aung make up the facilities committee. Mg Ba, Mg Chit, Mg Hla and Mg Kyaw form the decorations committee. Mg Kyaw, Mg Aung, and Mg Tin are the music committee. Mg Kyaw, Mg Lin, and Mg Chit form the clean up committee. We can find the number of meeting times, in order for each committee to meet once.

We will construct a graph model with one vertex for each committee. Each vertex is labeled with the first letter of the name of the committee that is represents. Two vertices are adjacent whenever the committees have at least one member in common. For example, P is

adjacent to R, since Mg Chit is on both the publicity committee and the refreshment committee.

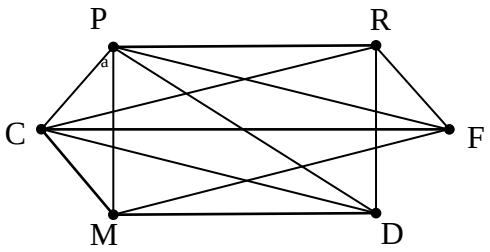

Figure 7 Graph model for the ball committee problem

To solve this problem, we have to find the chromatic number of G. Each of the four vertices P, D, M and C is adjacent to each of the other. Thus in any coloring of G these four vertices must receive different colors say 1, 2, 3 and 4, respectively, as shown in Figure 8. Now R is adjacent to all these except M, so it cannot receive the colors 1, 2, or 4. Similarly vertex F cannot receive colors 1, 3, or 4, nor can it receive the color that R receive. But we could color R with color 3 and F with color 2.

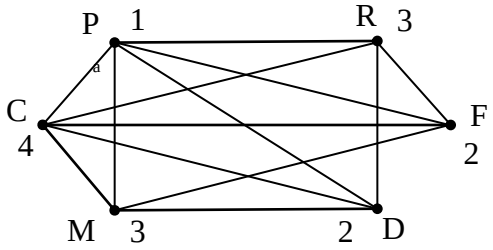

Figure 8 Using a coloring to ball committee problem

This complete a 4-coloring of G. Since the chromatic number of this graph is 4, four meeting times are required. The publicity committee meets in the first time slot, the facilities committee and the decorations committee meet in the second time slot, the refreshment and music committee meet in the third time slot, and the clean up committee meets last.

Example

We can find a schedule of the final exams for Math 5201, Math 5202, Math 5203, Math 5204, Math 5205, Math 5206, Math 5207, and Math 5208, using the fewest number of different time slots, if there are no student taking both Math 5201 and Math 5208, both Math 5202 and Math 5208, both Math 5209 and Math 5205, both Math 5209 and Math 5206, both Math 5201 and Math 5202, both Math 5201 and Math 5203, and Math 5203 and Math 5209, but there are students in every other combination of courses. This scheduling problem can be solved using a graph model, with vertices representing courses and with an edge between two vertices if there is a common student in the courses they represent. Each time slot for a final exam is represented by a different color. A scheduling of the exams corresponds to a coloring of the associated graph.

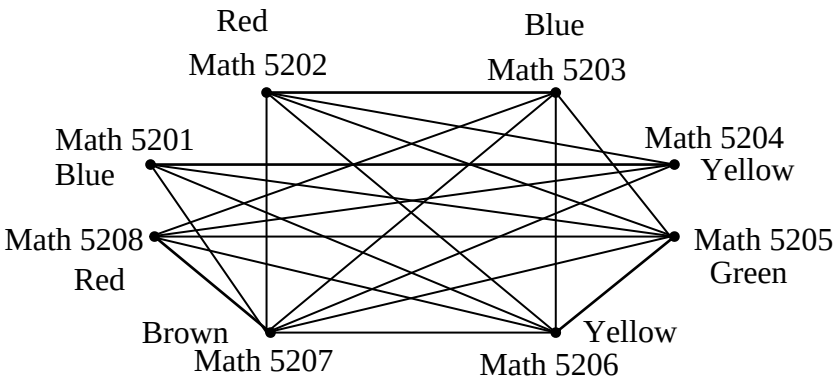

Figure 10 Using a coloring to scheduling of final exams

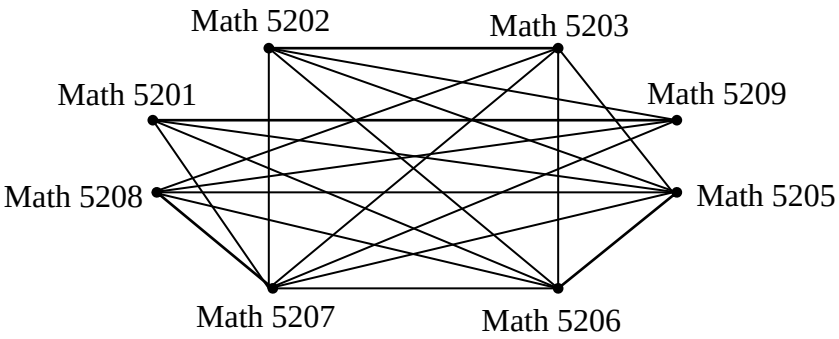

Figure 9 The graph representing the scheduling of final exams

Since $\{\text{Math 5202, Math 5203, Math 5205, Math 5206, Math 5207}\}$ forms a complete subgraph of G, $\chi(G) \geq 5$. Thus five time slots are needed. A coloring of the graph using five colors and the associated schedule are shown in Figure 10.

Time Period	Courses
I	Math 5202, Math 5208
II	Math 5201, Math 5203
III	Math 5204, Math 5206
IV	Math 5205
V	Math 5207

Example

Television channels are assigned to broadcasting stations by a governmental agency. Obviously, two stations in geographic proximity must get different channels, to avoid reception interference. Suppose that the rule has been adopted that stations within 140 miles of each other (as the crow files) must have different channels. The grid in Figure 11 shows the locations of 15 hypothetical stations. Each square is 50 miles on one side. We can find channels, they be assigned to comply with rule.

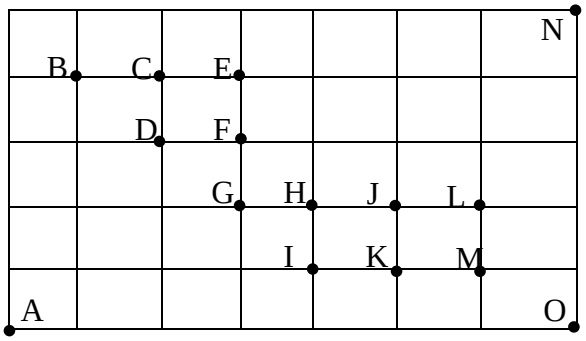

Figure 11 Locations of television stations A to O

We will construct a graph model. The vertices of the graph are the stations; two vertices are adjacent in the graph if the stations are within 140 miles of each other. Using the Pythagorean theorem to compute distances, we obtain the graph G shown in Figure 12. For example $d(HM) = \sqrt{50^2 + 100^2} \approx 112 \leq 140$, so HM is an edge of our graph, but $d(GL) = 150 > 140$, so no edge joins vertex G and vertex L.

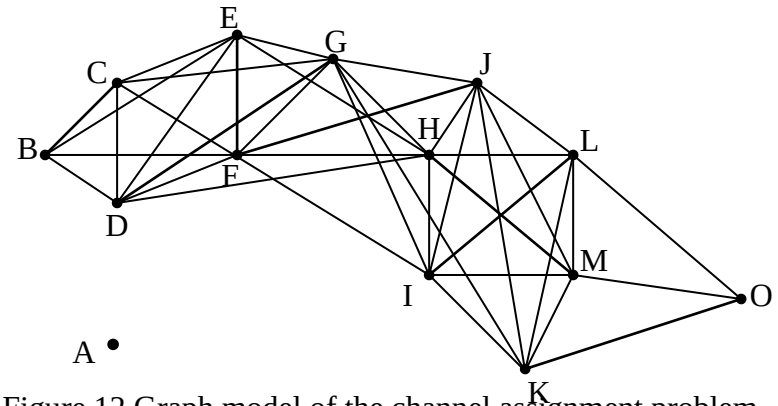

Figure 12 Graph model of the channel assignment problem

Now an assignment of channels to stations is precisely a coloring of G. The shannels are the colors. Thus to solve this problem, we have to find a coloring of G that uses as few colors as possible.

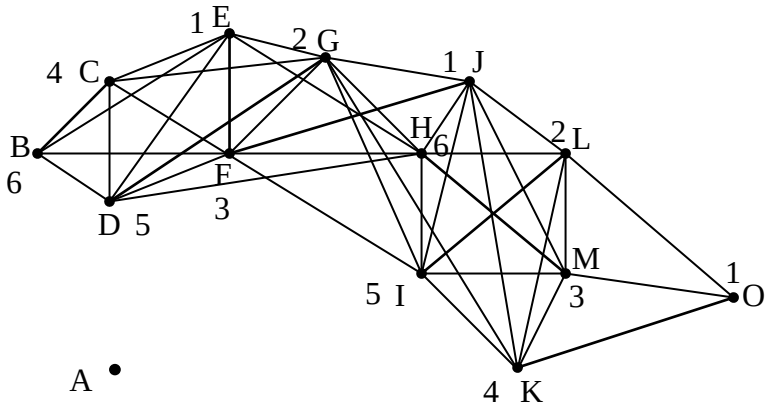

Figure 13 Using a coloring to channel assignment problem

Since $\{H, I, J, K, L, M\}$ forms a complete subgraph of G, $\chi(G) \geq 6$. Thus six channels are both necessary and sufficient, and any 6-coloring of G determines an allowable assignment of channels to stations.

Conclusion

We can know that some real life problems such as minimum meeting number problem, timetable scheduling problem and channel assignment problem can be solved by using properties of vertex coloring.

Acknowledgements

We thank to Dr Aung Kyaw Thin, Rector and Dr Aye Aye Han, Pro-Rector of Banmaw University for the permission to the journal. We would like to express our thanks to Dr Khin San Aye, Professor and Head, Department of Mathematics, Banmaw University for her kind encouragement to submit this paper. We would like to thank Professor Dr Hnin Oo Lwin, Department of Mathematics, Banmaw University for her helpful advice to do this paper.

References

Bondy, J.A, and Murty, U.S.R., "Graph Theory with Applications", Macmillan Press Ltd, London, 1976.

Parthasarathy, K. R., "Basic Graph Theory", McGraw-Hill, New Delhi, 1994.

Verstraete, J., "Map Coloring Theorems", UCSD La Jolla CA Jacques@ ucsd. Edu.