Scientific Research

Applied Mathematics

Scientific Research

TABLE OF CONTENTS

Volume 1   Number 4   **October 2010**

**The figure on the front cover is from the article published in Applied Mathematics, 2010, Vol.1, No.4, pp.
301-306 by Alexander M. Formalskii.**

Applied Mathematics (AM)

Journal Information

SUBSCRIPTIONS

SERVICES

COPYRIGHT

PRODUCTION INFORMATION

Scientific
Research

Some Models of Reproducing Graphs:
II Age Capped Vertices

Richard Southwell, Chris Cannings

School of Mathematics and Statistics, *University of Sheffield*, *Sheffield*, *United Kingdom*
E-mail: *bugzsouthwell@yahoo.com*
Received June 26, 2010; *revised August* 10, 2010; *accepted August* 12, 2010

Abstract

In the prequel to this paper we introduced eight reproducing graph models. The simple idea behind these models is that graphs grow because the vertices within reproduce. In this paper we make our models more realistic by adding the idea that vertices have a finite life span. The resulting models capture aspects of systems like social networks and biological networks where reproducing entities die after some amount of time. In the 1940's Leslie introduced a population model where the reproduction and survival rates of individuals depends upon their ages. Our models may be viewed as extensions of Leslie's model-adding the idea of network joining the reproducing individuals. By exploiting connections with Leslie's model we are to describe how many aspects of graphs evolve under our systems. Many features such as degree distributions, number of edges and distance structure are described by the golden ratio or its higher order generalisations.

1. Introduction

Networks are everywhere, wherever a system can be thought of as a collection of discrete elements, linked up in some way, networks occur. With the acceleration of information technology more and more attention is being paid to the structure of these networks, and this has led to the proposal of many models [1-3].

In many situations networks grow-expanding in size as material is produced from the inside, not added from outside. To study network growth we introduced a class of pure reproduction models [4,5], where networks grow because the vertices within reproduce. These models can be applied to many situations where entities are introduced which derive their connections from pre existing elements. Most obviously they could be used to model social networks, collaboration networks, networks within growing organisms, the internet and protein-protein interaction networks. One of our systems (model 3) has also been introduced independently [6], proposed as a model for the growth of online social networks.

In our pure reproduction models networks grow endlessly in a deterministic fashion. This allows a rigorous analysis, but costs a degree of realism. Nature includes birth and death and entities may be destroyed for reasons of conflict, crowding or old age. In this paper we consider age; and extend our models by including vertex mortality.

2. The Models

In [5] we defined a set $\{F_m : m \in \{0,1,\cdots,7\}\}$ of eight different functions F_m which map graphs to graphs. $F_m(G)$ is the graph obtained by simultaneously giving each of G's vertices an offspring vertex and then adding edges according to some rule. The connections given to offspring depend upon the binary representation $\alpha\beta\gamma$ of m (*i.e.* $m = 4\alpha + 2\beta + \gamma$) as follows:

$\gamma = 1 \Leftrightarrow$ offspring are connected to their parent's neighbours,

$\beta = 1 \Leftrightarrow$ offspring are connected to their parents,

$\alpha = 1 \Leftrightarrow$ offspring are connected to their parent's neighbour's offspring.

In our age capped reproduction models we think of the vertices as having ages. Graphs grow under these models exactly as before, except that vertices grow and then die when their age exceeds some pre-specified integer Q. Our new update operator $T_{m,Q}$ is defined so that $T_{m,Q}(G)$ is the graph obtained by taking the graph G and performing the following process;

1) Increase the age of each vertex by one.

2) Give every vertex an age zero offspring, born with connectivity dependant upon m, as above (*i.e.* the new graph is $F_m(G)$).

3) Remove every vertex with age greater than the age cap Q.

We are interested in the sequence $\{G_t\}$ of graphs which evolve from an initial structure $G_0 = (V_0, E_0)$ in such a way that $G_{t+1} = T_{m,Q}(G_t)$, $\forall t \geq 0$. We always suppose that initial vertices have age $\leq Q$.

3. The Number of Vertices

The number of vertices $|G_t|$ in G_t is deeply connected with the golden ratio and its generalisations. The number n_i^t of age i vertices in G_t can be conveniently descri-bed in terms of Leslie matrices.

In Leslie's population model [7,8] individuals of age i have a survival rate s_i and fertility rate f_i. The expected number of individuals of a given age, at a given time, is kept track of via repeated multiplication of the state vector with the 'Leslie matrix'

$$\begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_Q \\ s_0 & 0 & 0 & \cdots & 0 \\ 0 & s_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & s_Q & 0 \end{pmatrix}.$$

In our case $n^{t+1} = L.n^t$ where $n^t = \left(n_0^t, n_1^t, \cdots, n_Q^t\right)^T$ and L is the Leslie matrix with $s_i = f_i = 1$, $\forall i$.

L is a primitive matrix with characteristic polyno-mial

$$\sum_{i=0}^{Q} x^i = x^{Q+1}$$

and principle eigenvalue λ_Q (also known as the $Q+1$ step Fibbonacci constant). The golden ratio is $\lambda_1 = \dfrac{1 + 5^{1/2}}{2}$, $\lambda_2 = 1.8393$, $\lambda_3 = 1.9276$, $\lambda_4 = 1.96559$ and $\lambda_Q \to 2$ as $Q \to \infty$. When t is large $n^{t+1} = \lambda_Q n^t$ where $n^t = c\left(1, \left(\lambda_Q\right)^{-1}, \left(\lambda_Q\right)^{-2}, \cdots, \left(\lambda_Q\right)^{-Q}\right)^T$ is the stable age distribution and c is a constant which depends upon the initial state of the system.

Let $d_i = (\delta_{i,0}, \delta_{i,1}, ..., \delta_{i,Q})^T$ where $\delta_{i,j}$ is the Krone-cker delta. The n step Fibonacci numbers $f_i^{[n]}$ are natural generalisations of the famous Fibonacci numbers [9] which can be generated by repeatedly multiplying d_0 by L. When G_0 is age zero (*i.e.* all its vertices have zero age) $n_i^t = |G_0|.\left(L^t d_0\right)_i$, where $\left(L^t.d_0\right)_i = f_{t+1-i}^{[Q+1]}$. In such a case G_t will have $|G_0|.f_{t+2}^{[Q+1]}$ vertices.

4. Binary Strings

As we update our graphs, their vertex sets will grow, and a good way to keep track of these vertex sets is to use binary strings. Suppose v is a vertex of G_0. When we update G we write $(v,0)$ and $(v,1)$ to denote v's offspring, and v itself (respectively), in the graph G_1. This means, for example, that $((v,0),0)$ is the grand child of $((v,1),1)$ in G_2. We use short hand by omit-ting the parenthesis, so for example we write $((v,0),0)$ as $v00$. An example of the evolution of model 2 is shown in **Figure 1**.

When our age cap $Q = \infty$ an initial graph $G_0 = (V_0, E_0)$ will evolve in exactly the same way as in pure repro-duction *i.e.* $G_t = F_m^t(G_0)$; this will have vertex set $V_0 \times \{0,1\}^t$ and edge set as specified in [5]. When Q is finite the situation is more complex, but our binary string notation allows us to keep track of the ages of vertices in a convenient way.

Let ab denote the concatenation of binary strings a and b and let a^t denote the string obtained by conca-tenating a with itself t times. Suppose $v(a01^n)$ is a vertex of G_t for some $a \in \{0,1\}^{t-1-n}$. Now $v(a0)$ is a new born offspring in G_{t-n} and every subsequent 1 in $v(a01^n)$'s name corresponds to an update within which
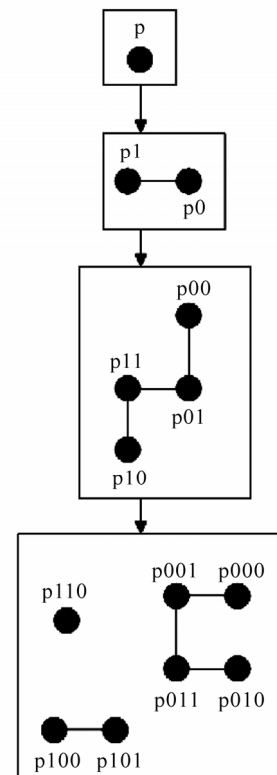

Figure 1. A depiction of the evolution of model 2 when $Q = 2$, starting with an isolated vertex named p.

this vertex survives as a parent and gets older by one. It follows that $v(a01^n)$ will be an age n vertex in G_t.

Theorem 1

If every vertex of the initial graph $G_0 = (V_0, E_0)$ is age zero then G_t will be the subgraph of $F_i^t(G_0)$ induced upon the vertex set $V_0 \times W_Q^t$, where $W_Q^t \subseteq \{0,1\}^t$ denotes the set of all t length binary strings which do not contain a run of $Q+1$ consecutive 1's.

Proof

Suppose G_t is the subgraph of $F_i^t(G_0)$ induced upon $V_0 \times W_Q^t$, as is clearly the case when $t=0$. An age n vertex va in G_t, with $v \in V_0$, will produce an offspring $v(a0)$ in G_{t+1}. va will also survive to become $v(a1)$ iff $n < Q$. Such an a must be of the form 01^n or 1^n. It follows that G_{t+1} will be the subgraph of $F_i^{t+1}(G)$ induced upon $V_0 \times X$ where X is the set of all ax with $a \in W_Q^t$ and $x \in \{0,1\}$ such that $a_{t-n+1} a_{t-n+2} \cdots a_t x \neq 1^{Q+1}$. Clearly X is W_Q^{t+1} so we can use induction with t to prove the result $\square$.

If the initial graph holds vertices of non-zero age; G_t can be obtained by taking the structure described in theorem 1 and removing every vertex of the form $v(1^n a)$, where n plus the age of v (in G_0) is greater than Q.

5. How Edges Connect Vertices of Different Ages

To keep track of the number of edges of G_t it helps to consider how vertices of different ages link to one another. Let S denote the Leslie matrix with all survival rates set at one and all fertility rates set at zero. Let F denote the Leslie matrix with all survival rates set at zero and all fertility rates set at one (note $L = S + F$). Let us define the age sampling vector $X = (X_0, X_1, \cdots, X_Q)^T$ of a vertex to be such that X_i is the number of neighbours it has of age i.

Applying the $T_{m,Q}$ update will cause an age $n \leq Q$ vertex to have an offspring with age sampling vector

$$o_{m,n}(X) = (\gamma S + \alpha F).X + \beta(1 - \delta_{n,Q}).d_{n+1} \quad (1)$$

and also, provided $n < Q$, this vertex will also survive the $T_{m,Q}$ update to become a parent with age sampling vector

$$p_m(X) = (S + \gamma F).X + \beta.d_0. \quad (2)$$

Equations (1) and (2) describe how the age sampling vector of a vertex determines the age sampling vector of itself and its offspring on the next time step. Repeatedly using these equations allows us to understand how the history of a vertex relates to its connectivity. The sequence of zeros and ones in a tell us the sequence of birth and survival stages which lead to the creation of a vertex

va in G_t. Given this information one can compute the age sampling vector of va by performing the corresponding sequence of $o_{m,n}$ and p_m operations-starting with the age sampling vector of the initial vertex, v, in G_0.

Let $e_{i,j}^t$ denote the number of edges of G_t that connect vertices of age i to vertices of age j. We consider how $e_{i,j}^t$ evolves in order to describe the growth rate of the number of edges in the different models. The vector $e_i^t = (e_{i,0}^t, e_{i,1}^t, \cdots, e_{i,Q}^t)$ is equal to the sum of the age sampling vectors of all of G_t's age i vertices and hence satisfies the equations

$$e_0^{t+1} = \sum_{j=0}^{Q} \left((\gamma S + \alpha F)e_j^t + \beta(1 - \delta_{j,Q})n_j^t.d_{j+1}\right), \quad (3)$$

$$i > 0 \Rightarrow e_i^{t+1} = (S + \gamma F).e_{i-1}^t + \beta.n_{i-1}^t.d_0. \quad (4)$$

Since the graphs we are concerned with are undirected we have $e_{i,j}^t = e_{j,i}^t, \ \forall i, j$.

The average asymptotic rates of increase of the minimal and maximal degrees for the different models are given in **Table 1**. We use the term *average* because, under some models, these extremal degrees increase at varying rates dependant upon the time modulo $Q+1$. These rates where found by determining which binary string describes a vertex with maximal (or minimal) degree and using Equations (1) and (2). For example, suppose the initial graph G_0 is age zero, and holds a vertex v with maximal degree, $deg(v)$, also suppose $t = n.(Q+1)+c$, for $c \leq Q$. When $m=1$ the vertex va, with $a = (1^Q 0)^n 1^c$, will have maximal degree in G_t. This vertex will have age sampling vector $L^c(S.L^Q)^n (deg(v).d_0)$. The degree of the vertex with this form will increase by $(\lambda_{Q-1})^Q$ every subsequent $Q+1$ time steps, and so it follows that the average asymptotic rates of increase of the maximal degree when $m=1$ is $(\lambda_{Q-1})^{Q/(Q+1)}$.

Table 1. A table showing the average asymptotic growth rates of the minimal and maximal degrees under the different models m. The notation LIN(x) indicates that the extremal degrees increase linearly (as opposed to exponentially) with time with gradient x.

m	growth rate of the minimal degree	growth rate of the maximal degree
0	0	0
1	0	$(\lambda_{Q-1})^{Q/(Q+1)}$
2	0	1
3	1	$(\lambda_{Q-1})^{Q/(Q+1)}$
4	1	1
5	λ_Q	λ_Q
6	LIN(Q)	LIN($Q+1$)
7	λ_Q	λ_Q

6. Connectivity, Degrees and Distances in Specific Models

In this section we will focus on reproduction mechanisms with $m \in \{1, 2, 3, 5, 6, 7\}$, one after another, and discuss the development of: connected components, number of edges, degree distributions, average path length and diameter. We do not discuss the dynamics when $m = 0$ or $m = 4$ because they are relatively uninteresting.

Before we discuss the specifics it is worth pointing out an effect that occurs under many models. We say that a graph is *age mixed* when each of it edges connect a pair of vertices with different ages.

If $\alpha = 0$ and $t > Q$ then G_t will be age mixed. The reason is that when $\alpha = 0$ offspring are not born connected to one another. So when $t > Q$ all of the initial vertices will be dead, and G_t will never again produce linked vertices with the same age.

Saying that G_t is age mixed has many implications, for example it means that G_t has chromatic number $\leq Q$ because its vertices may be coloured according to their ages.

6.1. Aspects of Model 1

Suppose $m = 1$ and we begin with a connected graph. G_t will typically consist of a growing connected component and lots of isolated vertices.

In the special case when $t > Q = 1$, updates do not cause the connected part of G_t's structure to change. The reason is that G_t is age mixed and every new born vertex either has a dead parent, which it replaces, or no surviving neighbours.

Suppose $G_0 = (V_0, E_0)$ is an age zero graph with $u \in V_0$. Any vertex ua of G_t will be isolated iff a holds a run of $Q + 1$ consecutive zeros.

To see this note that theorem 1, together with results from [5], imply that vb will be a neighbour of ua iff $\{u, v\} \in E_0$, $b \in W_Q^t$ and $a_i = 0 \Rightarrow b_i = 1$, $\forall i$. Now if a holds a run of $Q + 1$ consecutive zeros then this means b holds a run of $Q + 1$ consecutive ones, which means $b \notin W_Q^t$, so no neighbour vb can actually exist. On the other hand if a does not hold a run of $Q + 1$ consecutive zeros, and $\{u, v\} \in E_0$ then $\{ua, vb\} \in E_t$, where $b_i = 1 - a_i$, $\forall i$.

Let Y_i^t denote the set of t length binary strings of the form ax^i where a is any $t - i$ length string which does not hold a run of $Q + 1$ consecutive 0's or a run of $Q + 1$ consecutive 1's, and $x \in \{0, 1\}: x \neq a_{t-i}$. By our argument above the number of non-isolated vertices in G_t will be $|G_0| \sum_{i=1}^{Q} |Y_i^t|$. For example consider a string $100110 \in Y_1^6$ (when $Q = 2$), this string can be

thought of as being responsible for generating the strings $1001101 \in Y_1^7$ and $1001100 \in Y_2^7$ at the next time step. Following similar reasoning one can see that, for generic Q, we have the difference equation;

$$
\begin{pmatrix} |Y_1^{t+1}| \\ |Y_2^{t+1}| \\ |Y_3^{t+1}| \\ \vdots \\ |Y_Q^{t+1}| \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} |Y_1^t| \\ |Y_2^t| \\ |Y_3^t| \\ \vdots \\ |Y_Q^t| \end{pmatrix},
$$

which involves the $Q \times Q$ Leslie matrix. It follows that the number of non-isolated vertices in G_t increases asymptotically at a rate of λ_{Q-1} (whilst the total number of vertices grows at a rate of λ_Q), meaning eventually almost every vertex of G_t will be isolated.

Although vertices do get destroyed during the the $T_{1,Q}$ they are always replaced by offspring which link to their old neighbours. The only way that a component of G_0 could be disconnected under the update (in a non-trivial way) is if G_0 has a cutset of edges that connect pairs of age Q vertices. This can only happen during the first Q updates.

Regarding the edges, Equations (3) and (4) lose their dependance upon n_i^t and $t > Q \Rightarrow e_{i,i}^t = 0, \forall i$. Given these considerations we can reduce (3) and (4) to the following system of linear difference equations:

$\forall i, j \in \{1, 2, \cdots, Q\}$ such that $i < j$,

$$e_{0,i}^{t+1} = \sum_{n=0}^{i-2} e_{n,i-1}^t + \sum_{n=i}^{Q} e_{i-1,n}^t \tag{5}$$

$$e_{i,j}^{t+1} = e_{i-1,j-1}^t. \tag{6}$$

We can cast this system as a matrix difference equation which describes the evolution of

$$
\begin{pmatrix} e_{0,1}^t \\ e_{0,2}^t \\ \vdots \\ e_{0,Q}^t \\ e_{1,2}^t \\ e_{1,3}^t \\ \vdots \\ e_{1,Q}^t \\ e_{2,3}^t \\ \vdots \\ e_{2,Q}^t \\ \vdots \\ \vdots \\ e_{Q-1,Q}^t \end{pmatrix} \tag{7}
$$

The matrix which describes how (7) changes is primitive with principle eigenvalue μ_Q so the number of edges in G_t increases at a rate of μ_Q asymptotically. $\mu_1 = 1$, $\mu_2 = \lambda_2 = 1.8393$, $\mu_3 = 2.3336$ and $\mu_4 = 2.6077$. In general μ_Q increases with Q and $\mu_\infty = 3$.

Let $D_t(n)$ denote the number of vertices of degree n in G_t. Computer simulations suggest that when $1 \ll t < n$ we have $D_t\left(\lfloor \lambda_{Q-1}.n \rfloor\right) = \left(\lambda_{Q-1}\right)^{-1}.D_t(n)$ so it appears at the high end, as if the distribution obeys a geometric law. Whilst it seems there is some pattern in the degree distribution at the high degrees, the behaviour of the distribution of the lower degree vertices is more mysterious. For example it appears that when $Q > 1$ there will be less degree 1 vertices than degree 2 vertices when t is large.

Global notions of distance (such as diameter) do not really make sense when $m = 1$ because the structure is disconnected, with many isolated vertices.

6.2. Aspects of Model 2

Introducing an age cap into the $m = 2$ model leads to fascinating self replicative behaviour. Whatever graph we begin with we end up with a set of special tree graphs that grow up and break into more tree graphs. Let S_Q^t denote the graph obtained by starting with an age zero isolated vertex and evolving updating it with $T_{2,Q}$, t times. This graph will have vertex set W_Q^t and a pair of vertices a, b will be adjacent iff $\exists k \in \{1, 2, \cdots, t\}$ such that $\forall i \in \{1, 2, \cdots, t\}$ we have $i < k \Rightarrow a_i = b_i$, $i = k \Rightarrow a_i \neq b_i$ and $i > k \Rightarrow a_i = b_i = 1$.

To understand the self replicative behaviour in S_Q^t it helps to understand the self similarity of S_∞^t, the oldest and most central vertex of which is 1^t. Consider any neighbour $1^{t-n-1}01^n$ of 1^t. Since structures grow out of every vertex in the same way; the subgraph, X_n^t, induced upon the vertices $\{1^{t-n-1}0x : x \in \{0,1\}^n\}$ which grew out of $1^{t-n-1}01^n$, over n time steps, will be age-isomorphic to the graph S_∞^n which grew out of the initial vertex, over n time steps (by *age-isomorphic* we mean there is a one to one mapping, from one vertex set to the other, which preserves the adjacency, non-adjacency and ages of the vertices).

More generally $S_Q^t = S_\infty^t$ when $t \leq Q$, and S_Q^{Q+1} is the graph obtained by taking S_Q^{Q+1} and removing the oldest vertex, 1^{Q+1}. Since S_Q^{t+1} is a tree, the removal of 1^{Q+1} causes the graph to break into numerous components, namely $X_0^{Q+1}, X_1^{Q+1}, \cdots, X_Q^{Q+1}$. Since X_n^{Q+1} is age-isomorphic to S_∞^n, it follows that S_Q^{Q+1} consists of $Q+1$ different connected components, one age-isomorphic to S_∞^n for each $n \leq Q$. Each of these connected components will evolve in the same manner-growing until the age of its central vertex exceeds Q, at which point it will

fragment into yet more of these special trees.

Any initial graph will evolve to become a set of these trees after $Q+1$ time steps. The reason is that when $t = Q+2$ all of the initial vertices will have died. This means the oldest surviving ancestor of any vertex in G_{Q+2} will be a vertex which was born when $t > 0$. If a pair of vertices lie within the same connected component in G_{Q+2} then they will have the same oldest surviving ancestor, and it follows that every connected component of G_{Q+2} is a tree structure which grew out of a vertex that was not initially present, and is hence age isomorphic to S_∞^n for $n \leq Q$.

Let C_n^t denote the number of connected components of G_t which are age-isomorphic to S_∞^n for $n \leq Q$. The vector $C^t = (C_0^t, C_1^t, \cdots, C_Q^t)^T$ satisfies the matrix equation $C^{t+1} = \Gamma.C^t$, where

$$\Gamma = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \end{pmatrix}.$$

is equivalent to the transpose of the Leslie matrix L. It follows that when t is large C_n^t will increase at a rate of λ_Q and the probability that a random connected component is age-isomorphic to S_∞^i will be

$$p_i = \frac{\sum_{x=0}^{i} \lambda_Q^{-x}}{\sum_{y=0}^{Q} \sum_{x=0}^{y} \lambda_Q^{-x}} = \frac{1 - \lambda_Q^{-(i+1)}}{Q + 1 + \left(\frac{1 - \lambda_Q^{-(Q+1)}}{1 - \lambda_Q}\right)}. \tag{8}$$

The number of edges is described by the equations:

$$e_0^{t+1} = \sum_{j=0}^{Q-1} n_j^t . d_{j+1}, \tag{9}$$

$$i > 0 \Rightarrow e_i^{t+1} = S.e_{i-1}^t + n_{i-1}^t . d_0. \tag{10}$$

When $t > Q$ we will have $e_{i,i}^t = 0$, $\forall i$. This implies that $\forall i, j \in \{0, 1, \cdots, Q\} : i < j$ we have $e_{i,j}^t = n_{j-i-1}^{t-i-1}$ and so the number of edges in G_t increases at a rate of λ_Q asymptotically.

We can gain the asymptotic form of the degree distribution of G_t. First note that the graph S_∞^i has 2^i vertices. The number of degree k vertices in S_∞^i will be

$$\Delta_k^i = \left(\lfloor 2^{i-k} \rfloor + \delta_{k,i}\right)\left(1 - \delta_{k,0}\right) + \delta_{i,0}.\delta_{k,0}. \tag{11}$$

Now the probability $\dfrac{D_t(k)}{|G_t|}$ that a randomly selected vertex of G_t will be of degree k will be equal to [the probability that a randomly selected vertex of G_t be-

longs to a connected component isomorphic to S_∞^i] times [the probability that a randomly selected vertex of S_∞^i will have degree k], summed over all $i \in \{0, 1, \cdots, Q\}$.

For large t the probability that a randomly selected vertex of G_t belongs to a connected component isomorphic to S_∞^i is

$$\frac{p_i.2^i}{\sum\limits_{r=0}^{Q} p_r.2^r} = \frac{2^i\left(1 - \lambda_Q^{-(i+1)}\right)}{\Omega_Q} \qquad (12)$$

where

$$\Omega_Q = 2^{Q+1} - 1 + \lambda_Q^{-1} \left(\frac{1 - \left(\dfrac{2}{\lambda_Q}\right)^{Q+1}}{\left(\dfrac{2}{\lambda_Q}\right) - 1} \right). \qquad (13)$$

The probability that a randomly selected vertex of S_∞^i will be of degree k will be

$$\frac{\left(\lfloor 2^{i-k} \rfloor + \delta_{k,i}\right)\left(1 - \delta_{k,0}\right) + \delta_{i,0}.\delta_{k,0}}{2^i}. \qquad (14)$$

Hence as $t \to \infty$ we have $\dfrac{D_t(k)}{|G_t|}$ will be equal to

$$\sum\limits_{i=0}^{Q} \frac{2^i\left(1 - \lambda_Q^{-(i+1)}\right)\left(\lfloor 2^{i-k} \rfloor + \delta_{k,i}\right)\left(1 - \delta_{k,0}\right) + \delta_{i,0}.\delta_{k,0}}{\Omega_Q.2^i}. \quad (15)$$

Suppose t is large.

$$\frac{D_t(0)}{|G_t|} = \frac{1 - \lambda_Q^{-1}}{\Omega_Q}, \qquad (16)$$

$$\frac{D_t(Q)}{|G_t|} = \frac{2\left(1 - \lambda_Q^{-(Q+1)}\right)}{\Omega_Q}, \qquad (17)$$

If $1 \le k \le Q-1$ then $\dfrac{D_t(k)}{|G_t|}$ will be

$$\frac{2}{\Omega_Q} \left(2^{Q-k} - \lambda_Q^{-(k+1)} - \lambda_Q^k . \left(\frac{\left(\dfrac{2}{\lambda_Q}\right)^{Q-k} - 1}{\left(\dfrac{2}{\lambda_Q}\right) - 1} \right) \right), \qquad (18)$$

and if $k \notin \{0, 1, 2, \cdots, Q\}$ then $\dfrac{D_t(k)}{|G_t|} = 0$.

Once again we do not discuss distances because global notions of distance do not really make sense upon graphs which constantly disconnect.

6.3. Aspects of Model 3

Growth model 3 produces complicated structures; we

can say a little about their connectivity using reasoning like that used when $m = 1$. Since newborn vertices are never linked, $t > Q$ implies that G_t will not hold any linked vertices with the same age. If G_0 is connected then G_t will usually be connected. When G_0 has a cutset of edges connecting pairs of vertices with the same age then $T_{3,Q}(G_0)$ will be disconnected. This is the only way that structures can become disconnected, and it can only happen during the first Q updates.

With respect to edge numbers, there are many similarities in the way that G_t evolves when $m = 1$ and $m = 3$. The only difference is that when $m = 3$ offspring are connected to their parents, and this means that the equations which describe the evolution of $e_{i,j}^t$ gain a dependance upon the number of vertices. When $t > Q$ we will have $e_{i,i}^t = 0$, $\forall i$, and we will hence have that:

$\forall i, j \in \{1, 2, \cdots, Q\}$ such that $i < j$,

$$e_{0,i}^{t+1} = \sum\limits_{n=0}^{i-2} e_{n,i-1}^t + \sum\limits_{n=i}^{Q} e_{i-1,n}^t + n_{i-1}^t \qquad (19)$$

$$e_{i,j}^{t+1} = e_{i-1,j-1}^t \qquad (20)$$

In the $m = 1$ case the number of edges increase asymptotically at a rate of μ_Q. The $m = 3$ case is similar except that the number of edges is bolstered by the number of vertices n_i^t, which increases at a lesser rate of λ_Q. For large t the effect of these additional edges is hence negligible and the number of edges again increases at a rate of μ_Q.

Like the $m = 1$ case computer simulations again suggest that when $1 \ll t < n$ we have

$$D_t\left(\lfloor \lambda_{Q-1}.n \rfloor\right) = \left(\lambda_{Q-1}\right)^{-1}.D_t(n), \qquad (21)$$

so the distribution again obeys a geometric law at the high end.

When $Q = 1$ we can describe the evolution of the degree distribution exactly for any initial graph G_0 that is age mixed with no isolated vertices. Applying $T_{3,1}$ to G_0 is equivalent to changing the age of each vertex (from 0 to 1, and from 1 to 0) and then, for each age 1 vertex v, adding an age zero vertex that is only adjacent to v.

Let $N_x^t(d)$ denote the number of vertices of age x and degree d in G_t. $\forall t \ge 0$ we have

$$N_1^{t+1}(1) = 0, \qquad (22)$$

$$N_0^{t+1}(1) = N_1^t(1) + \sum\limits_{i=1}^{\infty} N_0^t(i) = N_1^t(1) + n_0^t, \qquad (23)$$

$$d > 1 \Rightarrow N_0^{t+1}(d) = N_1^t(d), \qquad (24)$$

$$N_1^{t+1}(d) = N_0^t(d-1). \qquad (25)$$

Solving these equations implies $\forall t, d > 0 \quad \forall x \in \{0, 1\}$

that when $d > \dfrac{t+x+1}{2}$ we have

$$N_x^t(d) = (1 - \delta_{1,x}\delta_{1,d})N_{t+x\bmod 2}^0\big(d - \lfloor (t+x)/2 \rfloor\big), \quad (26)$$

when $d = \dfrac{t+x+1}{2}$ we have

$$N_x^t(d) = n_0^0 + N_1^0(1), \quad (27)$$

and when $d < \dfrac{t+x+1}{2}$ we have

$$N_x^t(d) = n_0^{t+x+1-2d}. \quad (28)$$

When we introduce mortality our graphs seem to get longer. Diameter and average path length become greater. This is a result of the death of old vertices (which tend to be more central), this decreases the ease with which one can travel between the extremities.

Let L_t denote the sum of $d(u,v)$ for each ordered pair of vertices (u,v) in G_t. The average length l_t is equal to $L_t.|G_t|^{-2}$. Let $U_{i,j}^t$ denote the sum of the distances $d(u,v)$ between each ordered pair of vertices (u,v) from G_t such that either u is of age i, v is of age j or u is of age j, v is of age i. When $Q = 1$ the average length is given by

$$l_t = \frac{U_{0,0}^t + U_{0,1}^t + U_{1,1}^t}{\left(n_0^t + n_1^t\right)^2}. \quad (29)$$

When $m = 3$ it seems as if both the average length and diameter of G_t increase linearly with t whenever Q is finite. In the special case where $Q = 1$ we can gain an exact description.

Suppose that $G = (V, E)$ is a connected age mixed graph; if u and v are age zero vertices of G then, after applying the $T_{3,1}$ update, $d(u1,v1) = d(u,v)$, $d(u0,v1) = d(u,v)+1$ and (provided $u \neq v$), $d(u0,v0) = d(u,v)+2$. If u is age zero and v is age one in $T_{3,1}(G)$ then after the update we have $d(u1,v0) = d(u,v)$ and $d(u0,v0) = d(u,v)+1$. If u and v are both of age one then after the update $d(u0,v0) = d(u,v)$. The diameter of G_t will increase by two every two time steps and moreover the system obeys the equations

$$U_{0,0}^{t+1} = U_{0,0}^t + U_{0,1}^t + U_{1,1}^t + 2.n_0^t\left(n_0^t + n_1^t - 1\right), \quad (30)$$

$$U_{0,1}^{t+1} = U_{0,1}^t + 2\left(U_{0,0}^t + n_0^t.n_0^t\right) \quad (31)$$

$$U_{1,1}^{t+1} = U_{0,0}^t. \quad (32)$$

These equations imply that L_{t+1} is equal to

$$\begin{aligned} &2.L_t + 2.L_{t-1} - L_{t-2} + 4\left(n_0^t\right)^2 \\ &+ 2.n_1^t\left(3.n_0^t - 2\right) + 2.n_1^{t-1}\left(1 - n_0^{t-1}\right), \end{aligned} \quad (33)$$

which means that when t is large, the average length

increases linearly with

$$l_{t+1} = l_t + \frac{8 + 14.\lambda_1}{10 + 15.\lambda_1}. \quad (34)$$

For $Q > 1$ we have that $T_{3,1}^t(G)$ is a partial subgraph of $T_{3,Q}^t(G)$ which is a partial subgraph of $T_{3,\infty}^t(G)$. This implies that the curve which describes l_t, for generic Q is bounded below by a constant (because of the $Q = \infty$ case, see [5]) and bounded above by a straight line.

6.4. Aspects of Model 5

In this case, when G_0 is age zero, G_t may be obtained by replacing each vertex v of G_0 with a cluster C_v of $f_{t+2}^{[Q+1]}$ isolated vertices, and then connecting each vertex of C_u to each vertex of C_v whenever u and v where adjacent in G_0. It follows that

$$D_t(n) = f_{t+2}^{[Q+1]}.D_0\left(\frac{n}{f_{t+2}^{[Q+1]}}\right). \quad (35)$$

Equations (3) and (4) which describes the development of the edges may be cast as the matrix equation

$$\begin{pmatrix} e_0^{t+1} \\ e_1^{t+1} \\ e_2^{t+1} \\ \vdots \\ e_Q^{t+1} \end{pmatrix} = \begin{pmatrix} L & L & L & \cdots & L \\ L & 0 & 0 & \cdots & 0 \\ 0 & L & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & L & 0 \end{pmatrix} \begin{pmatrix} e_0^t \\ e_1^t \\ e_2^t \\ \vdots \\ e_Q^t \end{pmatrix}.$$

The matrix involved is clearly the Kronecker product of L with itself, it is hence primitive with principle eigenvalue λ_Q^2. It follows that the asymptotic growth rate of the number of edges will be λ_Q^2.

Suppose our initial graph is connected, non-trivial and age zero. G_t can be obtained by replacing each vertex by a cluster of $f_{t+2}^{[Q+1]}$ vertices. This means every ordered pair (u,v) such that $d(u,v) = k$, in the initial graph gives rise to $f_{t+2}^{[Q+1]}.f_{t+2}^{[Q+1]}$ ordered pairs, spaced by distance k, in G_t. In addition to this, every cluster adds $2.f_{t+2}^{[Q+1]}\left(f_{t+2}^{[Q+1]} - 1\right)$ to the total distance, by the fact that every pair of distinct vertices within a given cluster will be spaced by distance 2. It follows that the total distance of G_t will be

$$L_t = f_{t+2}^{[Q+1]}.f_{t+2}^{[Q+1]}L_0 + 2.f_{t+2}^{[Q+1]}\left(f_{t+2}^{[Q+1]} - 1\right)|G_0|. \quad (36)$$

This means (irrespective of Q) that when t is large the average length approaches the constant $l_t = l_0 + \dfrac{2}{|G_0|}$.

The diameter of G_t will be the maximum of the diameter

of G_0 and 2.

6.5. Aspects of Model 6

In this case G_t will be a connected graph that can be obtained by taking the t dimensional hypercube graph and removing some vertices. The next graph in a sequence can be obtained by fusing together previous structures. For example when $t > Q = 1$, G_{t+1} can be obtained by taking the disjoint union of G_t and G_{t-1}, choosing an isomorphism f from G_{t-1} to the subgraph of G_t induced upon its age zero vertices (such an isomorphism always exists) and adding an edge from each v vertex of G_{t-1} to $f(v)$. The age one vertices of G_{t+1} will be those which came from G_{t-1}, vertices which came from G_t will be age zero.

The number of edges $\| G_t \|$, in G_t, satisfies $e_{0,0}^{t+1} = \| G_t \|$, $e_{i,j}^{t+1} = e_{i-1,j-1}^t$ and $e_{0,i}^{t+1} = n_{i-1}^t$, $\forall i, j > 0$. This implies

$$\| G_{t+1} \| = \sum_{i=0}^{Q} \sum_{j=i}^{Q} e_{i,j}^{t+1}, \tag{37}$$

we can split the sums to get

$$\| G_{t+1} \| = \sum_{i=0}^{Q} e_{i,i}^{t+1} + \sum_{i=0}^{Q-1} \sum_{j=i+1}^{Q} e_{i,j}^{t+1} \tag{38}$$

making substitutions we find

$$\| G_{t+1} \| = \sum_{i=0}^{Q} \| G_{t-i} \| + \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-i-1} n_j^{t-i}. \tag{39}$$

When t is large the minimal degree of G_t becomes large-implying that the average degree also becomes large. This implies

$$\sum_{i=0}^{Q} \| G_{t-i} \| \sum_{i=0}^{Q-1} | G_{t-i} | > \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-i-1} n_j^{t-i} \tag{40}$$

and so the asymptotic growth rate of the number of edges will be λ_Q.

Determination of the degree distribution when $m = 6$ appears to be a difficult problem. Although some progress can be made when $Q = 1$ the resulting formulae are long and complicated.

With respect to distances it appears that the diameter and average length of G_t increase linearly when t is large. We can show explicitly that this is the case when $Q = 1$.

We say a graph is zero spanning if there is a shortest path between each pair of age zero vertices that only passes through age zero vertices. Updating any connected graph with $T_{6,1}$ will always yield a zero spanning graph. Supposing that G_t is a zero spanning graph, if u and v are age zero vertices of G_t then after updating with $T_{6,1}$ we will have $d(u1, v1) = d(u0, v0) = d(u, v)$ and

$d(u0, v1) = d(u, v) + 1$. If u is age zero and v is age one in G then after the update we will have $d(u1, v0) = d(u, v) + 1$ and $d(u0, v0) = d(u, v)$. If u and v are both age one vertices of G then after updating we will have $d(u0, v0) = d(u, v)$. This implies that the system obeys the equations:

$$U_{0,0}^{t+1} = U_{0,0}^t + U_{0,1}^t + U_{1,1}^t \tag{41}$$

$$U_{0,1}^{t+1} = U_{0,1}^t + 2\left(U_{0,0}^t + n_0^t \left(n_0^t + n_1^t \right) \right) \tag{42}$$

$$U_{1,1}^{t+1} = U_{0,0}^t \tag{43}$$

These equations imply that as $t \to \infty$ the average length increases linearly with

$$l_{t+1} = l_t + \frac{2}{5}$$

The reasoning behind this is very similar to that when $Q = 1$ and $m = 3$. The diameter of G_t will increase by one every time step once the graph becomes zero spanning.

6.6. Aspects of Model 7

When our initial graph G_0 is age zero G_t may be obtained by replacing each vertex v of G_0 with a complete graph K_v on $f_{t+2}^{[Q+1]}$ vertices, and then connecting each vertex of K_u to each vertex of K_v whenever u and v where adjacent in G_0. It follows that

$$D_t(n) = f_{t+2}^{[Q+1]} . D_0 \left(\frac{n + 1 - f_{t+2}^{[Q+1]}}{f_{t+2}^{[Q+1]}} \right). \tag{44}$$

With respect to the edges this case is similar to the $m = 5$ case, except that there is an extra dependence upon n_i^t caused by the presence of edges linking offspring to their parents.

$$\begin{pmatrix} e_0^{t+1} \\ e_1^{t+1} \\ e_2^{t+1} \\ \vdots \\ e_Q^{t+1} \\ n^{t+1} \end{pmatrix} = \begin{pmatrix} L & L & L & \cdots & L & S \\ L & 0 & 0 & \cdots & 0 & B^0 \\ 0 & L & 0 & \cdots & 0 & B^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & L & 0 & B^{Q-1} \\ 0 & 0 & \cdots & 0 & 0 & L \end{pmatrix} \begin{pmatrix} e_0^t \\ e_1^t \\ e_2^t \\ \vdots \\ e_Q^t \\ n^t \end{pmatrix}$$

where B^n is the $(Q+1) \times (Q+1)$ matrix such that $\forall i, j \in \{0, 1, \cdots, Q+1\}$ we have $B_{i,j}^n = 0$, except that $B_{0,n}^n = 1$.

In the $m = 5$ case the number of edges increase asymptotically at a rate of λ_Q^2. This $m = 7$ case is similar except that the number of edges is bolstered by the number of vertices n_i^t, which increases at a lesser rate of λ_Q. For large t the effect of these additional edges is hence negligible and the number of edges again increases at a rate of λ_Q^2.

Suppose our initial graph is connected, non-trivial and age zero. G_t can be obtained by replacing each vertex by a complete graph on $f_{t+2}^{[Q+1]}$ vertices. This means every ordered pair (u,v) such that $d(u,v)=k$, in the initial graph gives rise to $f_{t+2}^{[Q+1]} \cdot f_{t+2}^{[Q+1]}$ ordered pairs, spaced by distance k, in G_t. In addition to this, every cluster adds $f_{t+2}^{[Q+1]} \cdot \left(f_{t+2}^{[Q+1]} - 1 \right)$ to the total distance, by the fact that every pair of distinct vertices within a given cluster will be spaced by distance 1. It follows that the total distance of G_t will be

$$L_t = f_{t+2}^{[Q+1]} \cdot f_{t+2}^{[Q+1]} L_0 + f_{t+2}^{[Q+1]} \left(f_{t+2}^{[Q+1]} - 1 \right) | G_0 |. \quad (45)$$

Interestingly this means that when t is large l_t loses its dependance upon Q and approaches the constant $l_t = l_0 + \dfrac{1}{|G_0|}$. The diameter of G_t will be equal to the diameter of G_0.

7. Discussion

We have discussed many properties of age capped models, however many open problems remain. These include describing degree distribution when $m \in \{1,3,6\}$ and demonstrating the linearity of average length when $m \in \{3,6\}$ (for generic Q).

There are many directions in which our models may be expanded. As highlighted by theorem 1, our models may be regarded as an extension of pure reproduction models by adding restrictions upon the language of binary strings which the vertices can possess. Many other restrictions could be considered, e.g. forbidding the subword $1^{Q+1}0$ (which would correspond to saying vertices of age $>Q$ become infertile).

Our models can be viewed as an extension of Leslie's population model, introducing the idea of a network which connects the reproducing individuals. We will further develop this connection by considering the evolution of generic Leslie matrices (so that individuals of a given age can have differing numbers of offspring and chances of survival). Taking this approach and considering connectivity as stochastic (so that α, β and γ are probabilities, rather then binary integers) should yield models which directly simulate the development of animal social networks and other phenomena.

This paper demonstrates how our original reproducing graph models can be generalised in different directions whilst remaining analytically tractable. Perhaps the main reason these models are amenable to analysis is that the growth of one part of a graph is not influenced by the structure of another. This spatial independence allows one to understand the evolution of generic structures by studying the evolution of simple ones.

There are many extensions of these models that it would be interesting to consider. In the future papers we will discuss the fascinating dynamics which can ensue when game theory is incorporated into these models. In this case we lose the spatial independence and dynamics of immense complexity become possible. It is also possible to extend many of the results here to cases where individuals produce several offspring-connected up in different ways. This kind of generalisation allows one model how the social networks of specific types of organisms grow in a more direct way.

8. Acknowledgements

9. References

[1] P. Erdös and A. Rényi, "On Random Graphs. I," *Publicationes Mathematicae*, Vol. 6, 1959, pp. 290-297.

[2] G. U. Yule, "A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F. R. S.," *Philosophical Transactions of the Royal Society of London*, B, Vol. 213, 1925, pp. 21-87.

[3] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, Vol. 393, No. 6684, 1998, pp. 440-442.

[4] R. Southwell and C. Cannings, "Games on Graphs that Grow Deterministically," *Proceedings of International Conference on Game Theory for Networks GameNets* '09, Istanbul, Turkey, 2009, pp. 347-356.

[5] R. Southwell and C. Cannings, "Some Models of Reproducing Graphs: I Pure Reproduction," *Journal of Applied Mathematics*, Vol. 1, No. 3, 2010, pp. 137-145.

[6] A. Bonato, N. Hadi, P. Horn, P. Praalat and C. Wand, "Models of On-Line Social Networks," To appear in *Internet Mathematics*, 2010.

[7] P. H. Leslie, "The Use of Matrices in Certain Population Mathematics," *Biometrika*, Vol. 30, 1945, pp. 183-212.

[8] P. H. Leslie, "Some Further Notes on the Use of Matrices in Population Mathematics," *Biometrika*, Vol. 35, No. 3-4, 1948, pp. 213-245.

[9] T. D. Noe, "Primes in Fibonacci n-step and Lucas n-step Sequences," *Journal of Integer Sequences*, Vol. 8, 2005, pp. 1-12.

AM

Scientific
Research

Transformation of Nonlinear Surface Gravity Waves under Shallow-Water Conditions

Iftikhar B. Abbasov

Taganrog Technological Institute, Southern Federal University, Taganrog, Russia
E-mail: *iftikhar_abbasov@mail.ru*

Abstract

This article describes transformation of nonlinear surface gravity waves under shallow-water conditions with the aid of the suggested semigraphical method. There are given profiles of surface gravity waves on the crests steepening stages, their leading edges steepening. There are discussed the spectral component influence on the transformation of surface wave profile.

1. Introduction

Surface gravity waves under shallow-water conditions were of great interest for many researchers many years ago. In spite of their magnetic view, they are rather hard to describe. The research problems of surface waves in the near-shore region within shallow-water model are discussed in [1,2]. [3] is concerned with nonlinear waves in strongly dispersive media. Transformation of surface waves in the near-shore region is discussed in [4]. The nonlinear dynamics of surface gravity waves in long-wave approximation is described in [5,6]. [6] analysed the process of surface wave deformation under shallow-water conditions, estimated spectral amplitude of nonlinear wave with its steepness.

In works [7,8] give mathematic simulation of gravity waves in approximation of shallow water. [8] is concerned with numerical simulation of water waves within nonlinear-dispersive model of shallow water taking account of bed topography. Experimental data on influence of the dispersion effects on the propagation of nonlinear surface waves in inshore area are given in [9].

2. Statement of the Problem

In work [10] discussed the problem of nonlinear surface waves propagation under shallow-water conditions. Shallow-water equation taking account of quadratic nonlinearity was solved by the method of successive approximations.

Either the shallow-water equation or the method of successive approximations is the basis of nonlinear wave research. In the middle of the 19th century G. Airy described tide waves and G. Stokes described waves of finite amplitude with this equation and this method [11]. However, the generating process of higher harmonics, resulting in wave crest sharpening, did not always meet energy conservation law. When higher harmonics are growing, the decrease of principal wave energy is ignored. The exact solution of shallow-water equation without dispersion and damping appears to be Riemannian invariants, based on different propagation velocity of wave crests and troughs [12,13]. In addition, shallow-water equation ignores dispersion effect because of its weakening on shallow water.

Surface gravity waves on shallow water in approximation of small nonlinearity and small dispersion are described by well known Cortvega and De Vries equation [11,12]. In the equation small nonlinearity, resulting in wave steepening, and small dispersion, resulting in wave diffusion, compensate each other. Thereat, stationary nonlinear wave (named as a cnoidal wave) occurs, which propagates without form change at a constant velocity. However, form stability of cnoidal waves doesn't allow retracing the dynamics of nonlinear surface gravity waves with the propagation on the shallow water.

In the development of earlier studies this work describes the transformation process of surface wave profile with the propagation on the shallow water. There was an attempt to describe not only the initial wave crest sharp-

ening but further steepening of its leading edge. One uses rather understandable and vivid semigraphical method based on energy conservation law given in [10].

3. Research Method

Method of successive approximations is the straightest way of the solution of nonlinear Equations [14]. Here, the small parameter expansion $\varepsilon(\varepsilon \ll 1)$ is used, restricting to the first two terms under the restriction $u^{(n+1)} \ll u^{(n)}$.

The equation taking account of quadratic nonlinearity takes the form [12,13]:

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = -\varepsilon u\frac{\partial u}{\partial x} \qquad (1)$$

where $u = \vartheta_x$ —— the horizontal component of medium particle velocity, nonlinearity —— $\varepsilon = \dfrac{\vartheta_x}{c} = \dfrac{a}{H}$, the propagation velocity of gravity waves $c = \sqrt{gH}$, a —— the amplitude of the vertical shift of the free surface, H —— the fluid depth.

With the propagation of the initial harmonic wave on shallow water as a consequence of nonlinear effects, the second harmonic appears. With the growth of the second harmonic its interaction with first harmonic will get stronger. This interaction will lead to the exciting of the third, fourth, and so on, harmonics.

In the first approximation, solution of Equation (1) consists of primary waves and takes the form [10]:

$$\begin{aligned} u_{prim.}^{(1)}(x,t) &= \varepsilon u^{(1)}(x,t) + \varepsilon^2 u^{(2)}(x,t) = \\ &= \varepsilon U_0 \exp[ikx] - ikt\varepsilon^2 U_0^2 \exp[i2kx] + c.c. \end{aligned} \qquad (2)$$

In the second approximation the non-homogeneous linear equation is solved

$$\frac{\partial u_{sec.}^{(2)}}{\partial t} + c\frac{\partial u_{sec.}^{(2)}}{\partial x} = -\frac{\partial u_{prim.}^{(1)2}}{2\partial x}. \qquad (3)$$

After squaring and solution differentiation in the first approximation (2) we obtain the expression in the second approximation:

$$\begin{aligned} u_{sec.}^{(2)}(x,t) &= \Big[\big(t\varepsilon^2 U_0^2/2\big)\cdot ik_2 \exp[ik_2 x] + \\ &+ \big(-k^2 t^3 \varepsilon^4 U_0^4/2\big)\cdot ik_4 \exp[ik_4 x] + \\ &+ \big(-ikt^2\varepsilon^3 U_0^3\big)\cdot ik_3 \exp[ik_3 x] + \\ &+ \big(-ikt^2\varepsilon^3 U_0^3\big)\cdot ik_1 \exp[ik_1 x]\Big] + (c.c.) \end{aligned} \qquad (4)$$

where $k_n = nk$, $n = 1,2,3,4$ —— harmonic number.

As a consequence of interaction between principal wave k and second $2k$ secondary waves with double values $2k$, $4k$, and complex waves k and $3k$ will occur. With

time, higher harmonics in the spectrum lead to the distortion of wave profile.

The solution of shallow-water Equation (1) in two approximations for horizontal velocity of medium particles will take the form

$$\begin{aligned} u_{sec.}(x,t) &= \varepsilon u_{prim.}^{(1)}(x,t) + \varepsilon^2 u_{sec.}^{(2)}(x,t) = \\ &= \Big[\varepsilon^2 U_0 \exp[ikx] - ikt\varepsilon^3 U_0^2 \exp[2ikx]\Big] + \\ &+ \Big[kt^2\varepsilon^5 U_0^3 \cdot k_1 \exp[ik_1 x] + \big(t\varepsilon^4 U_0^2/2\big)\cdot ik_2 \exp[ik_2 x] + \\ &+ kt^2\varepsilon^5 U_0^3 \cdot k_3 \exp[ik_3 x] + \\ &+ \big(-k^2 t^3 \varepsilon^6 U_0^4/2\big)\cdot ik_4 \exp[ik_4 x] + \Big] + (c.c.) \end{aligned}$$

$$(5)$$

4. Simulation within the Gulf

To check the results obtained we will use hydrologic conditions of the Gulf of Taganrog of the Azov Sea. The mean depth of the gulf does not exceed 5 m, therefore, the shallow-water conditions will be satisfied with gravity waves with their lengths are not longer than 30 m, the gulf bed is flat, surface tension is absent, the effect of wind is not taken into account.

Before calculation one should mark some physical features of existent wave processes. In formula (5) all harmonic amplitudes are always growing because of secular terms. However, according to the formula, the first harmonic amplitude remains constant at that. Though, higher harmonics are fared energetically from the first harmonic.

It follows that the primary waves amplitude in formula (5) $u_{prim.}^{(1)}(x,t)$ should be getting decreased. That is why to research profile dispersions of a gravity wave it is necessary to consider these physical features of wave processes otherwise it will break the energy conservation law.

In the absence of dissipation, energies of primary and secondary waves should satisfy должны отвечать the energy conservation law

$$E_1^{(1)}(x) + E_2^{(1)}(x) = E_1^{(2)}(x) + E_2^{(2)}(x) + E_3^{(2)}(x) + E_4^{(2)}(x)$$

$$(6)$$

where $E_n^{(1)}(x)$, $E_n^{(2)}(x)$ —— energies of primary and secondary waves.

Let us follow variations the profile change of the gravity wave after coming into the gulf with the initial parameters: frequency $f = 0.09$ Hz; length $\lambda = 77.8$ m; initial steepness – $2a/\lambda = 0.014$; $a = 0.537$ m; $kH = 0.4$. According to **Figure 1(a)** for this wave instability appears at the distance: $x_{inst.}^{(2)} \approx 10$ km (i.e. in $t_{inst.}^{(2)} \approx 24$ min.). Though our model is correct only within $x \approx 6$ km, i.e. till, while the primary limitation $u^{(n+1)} \ll u^{(n)}$ is fulfilled,
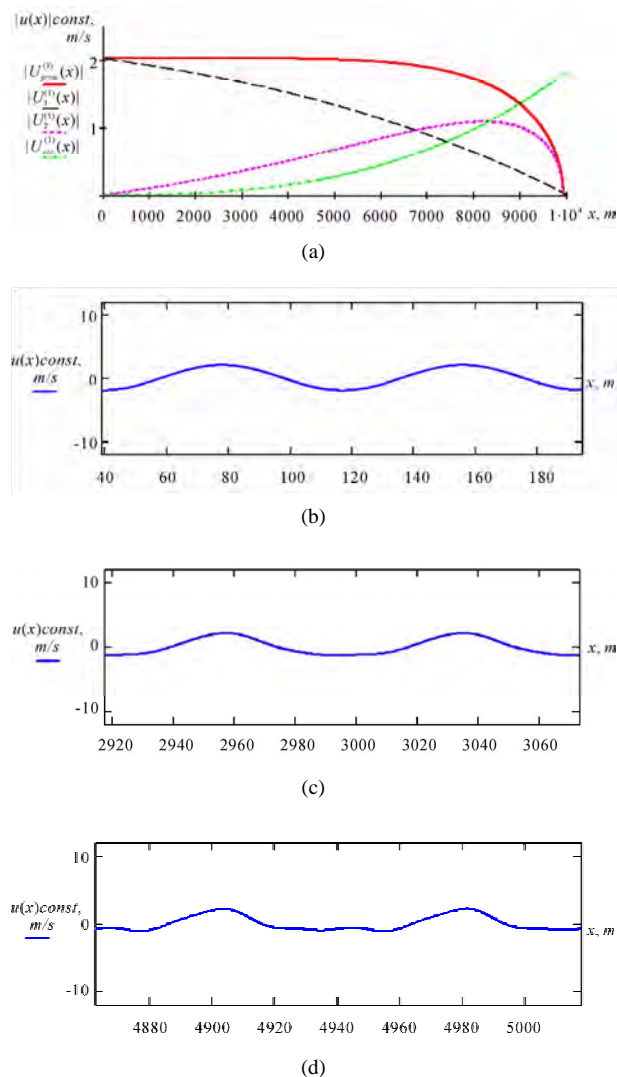
(a)

(b)

(c)

(d)

Figure 1. Primary and secondary wave velocities (a) and profiles (b, c, d) of the surface gravity wave: f **= 0.09 Hz;** λ **= 77.8 m; wave parameter –** kH **= 0.4; initial steepness –** $2a/\lambda$ **= 0.014; nonlinear parameter –** ε **= 0.107.**

when the value for the second approximation is by an order less than the value for the initial approximation.

Because of **Figure 1(a)** analysis, it is possible point out that the amplitude of primary waves and first harmonic falls with the growth of secondary waves summary amplitude. The amplitude of the second harmonic grows slowly, and to the moment of instability it falls down as there is a complete energy transfer from primary waves (principal wave and the second harmonic) to secondary waves.

With the growth of initial steepness the instability distance decreases, with the growth of the wave amplitude (with the constant steepness) this distance lengthens, as nonlinear parameter value decreases.

Let us follow the profile change of the initial gravity

wave within distance $x \approx$ 6 km. **Figures 1(b)**, **1(c)** and **1(d)** describe dependences of horizontal velocity change $u(x,t)$ on the distance run taking account of the formula (5) (expanded vertical scale). The wave with an initially cosinusoidal profile is gradually distorted in a run distance time, the crests steepen **Figure 1(c)**, and the wave troughs flatten. Thereafter there is a steepening of the wave leading edge **Figure 1(d)**. This is the result of the increasing influence of high frequency harmonics. The increasing of particle velocity on the crest leads to the further steepening of the wave leading edge. The breaking of such a wave in shallow water happens as a plunging or spilling breaker [15,16]. The breaking cause in our case is naturally liquid depth.

An additional point to emphasize is that the shallow-water condition could not be observed for the higher harmonics appear. It could lead to the propagation velocity dispersion that could cause braking of nonlinear process. But approaching the coast as the depth lessens, the shallow-water condition will be met better, therefore the wave breaks.

To analyse the secondary waves influence **Figure 2** gives graphs showing the growth of velocity amplitude of the secondary waves because of the distance. It clearly enough shows the relation between harmonics. Within the secondary waves, velocity amplitude of the fourth harmonic increases faster (because of the greatest degree of a secular term in the formula (5)), then there comes the third one, the first one and the second one. This conformity is observed for the gravity waves with other initial data as well.

5. Analysis and the Result Comparison

Analysing the constructed profiles of the surface gravity waves it is necessary to note their following features. From the field studies it is known that the profile of real surface waves is asymmetrical: the crests are steep and short, sharpened, and troughs are flattened and broad [11]. Undulations are asymmetrical as well: particles velocity below the crests is lower than below the troughs. Theses peculiarities are explained by Stokes theory and the ap-
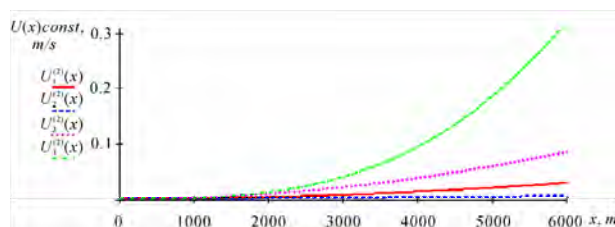

Figure 2. The increase of velocity amplitudes of secondary waves from the distance with: f **= 0.09 Hz;** λ **= 77.8 m;** kH **= 0.4;** $2a/\lambda$ **= 0.014;** ε **= 0.107.**

pearance of principal wave high frequency harmonics. For our study the surface waves obtain such profiles on the initial propagation phases, it is shown in **Figure 1(b)**. On this stage higher harmonics generation leads to the crests sharpening, at that the crests keep symmetrical relating to a vertical axis.

Beside the asymmetry type described above, there is another asymmetry type of the real surface wave profile. It is expressed in such a way that the advancing side of the transformed wave gets steeper than the wave tail [17-19]. This difference in steepness ratio increases with the approaching to the coast. This asymmetry type of the real surface wave profile is not described by Stokes theory as well as the major part of other theories. Though to describe the process of wave climb from the open sea to the coast G. Whitham offered the equation with an integral element based on Cortvega and De Vries equation. On the wave breaking stage the operation of an integral element becomes modest and the solutions of Whitham equation are analogue to the solutions of simplest equations with nonlinear nature (analogue to the Equation (1)), falling and breaking for a final time [12,20].

For our study the surface wave profiles in the third propagation area just as obtain the second asymmetry kind, **Figure 1(d)** At that, steepening becomes more and more intense with the growth of the wave length, *i.e.* with the increase of nonlinearity factor.

One of the explanations of steepening of crest advancing side is the following: higher harmonics gradually displace in phase relating to the principal wave [16,21,22]. When phase displacement approaches the value of π/2, the advancing side becomes almost vertical; as a result the wave could fall. In the obtained resultant expression for particle horizontal velocity (5) even harmonics amplitudes are imaginary. With the increase of secondary waves the imaginary is felt stronger than in the beginning as it leads to the phase displacement relative to the principal wave amplitude. It is necessary to note that there is no phase difference between the principal wave and higher harmonics in Stokes theory, *i.e.* it couldn't describe these processes.

To show the influence of even harmonics, especially the fourth one, the surface gravity wave profile, **Figure 3** gives graphs showing the horizontal velocity $u(x,t)$ of medium particles for the same propagation interval. The special influence of the fourth harmonic is connected with its rapid growth, it follows that the greatest energy-output ratio within secondary waves. **Figure 3(a)** gives the surface gravity wave profile, counted according to the formula (5), when the fourth harmonic in phase is behind the principal wave with π/2. **Figure 3(b)** gives the wave profile for the case, when the imaginary of the fourth harmonic amplitude is changed to be of real type, *i.e.* there is

(a)

(b)

(c)

(d)
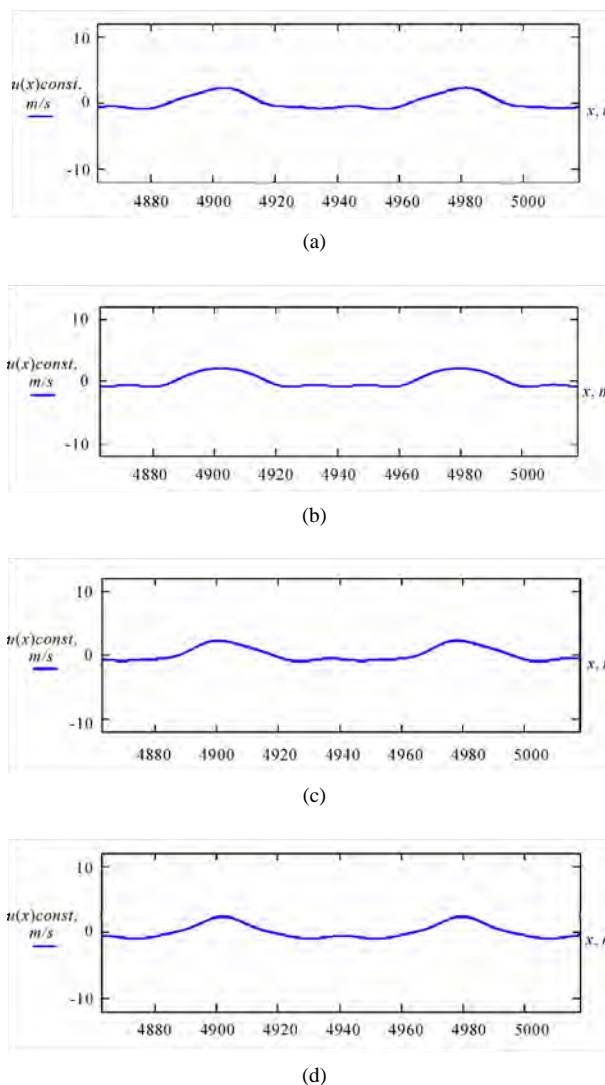
Figure 3. The fourth harmonic influence on the surface gravity wave profile: (a) is behind the principal wave with π/2; (b) subtracted in phase; (c) is ahead of the principal wave with π/2; (d) summarized in phase.

no the phase displacement between the harmonic and the principle wave. At that the fourth harmonic amplitude is negative, it follows that it is subtracted from the principle wave amplitude. It leads to the symmetric property of wave crest relating to vertical axis.

In formula (5) only the fourth harmonic amplitude is negative. The change of the negative sign to the positive one leads to the change of steepening of the wave crest front edge to the trailing edge. This case is given in **Figure 3(c)**, it reminds the profile of the reverse wave, the wave propagates backwards. In this case the fourth harmonic leads in phase the principle wave with π/2. If in formula (5) at the same time one changes the negative sign to the positive one, as well as the imaginary of the fourth harmonic changes to be of real type, the wave

profile takes the form as in **Figure 3(d)**.

In this case the fourth harmonic amplitude is summarized in phase with the principle wave amplitude, intensifying the wave crest. The given profile resembles Stokes waves of finite amplitude on the deep water (or trochoidal wave).

It is important to pay attention to the appearance of the local projection in the trough between the crests. This projection is also connected with the fourth harmonic, and it changes its location depending on the phase. Experimental observations with the trough between the principal wave crests in nearshore zone are given in [9,21,23].

6. Conclusions

As a consequence of the researches made, one could note that the suggested semigraphical method allows to describe not only the process of the initial sharpening of the surface waves during the propagation under shallow water conditions, and further steepening of their front edge. In our study we described the influence of the first four harmonics of the principle wave on profile transformation but the real wave spectrum is constant, and for more reasonable study it is necessary to record at least first eight harmonics. The record of higher harmonics will lead to more intense sharpening of wave crests and steepness growing of the wave leading edge.

7. References

[1]   D. H. Peregrine, "Long Waves on a Beach," *Journal of Fluid Mechanics*, Vol. 27, No. 0404, 1967, pp. 815-827.

[2]   N. E. Voltsinger, K. A. Klevanniy and E. H. Pelinovskiy, "Long-Wave Dynamics of Coastal Zone," *Gidrometeoizdat*, 1989, p. 272.

[3]   P. I. Naumkin and I. A. Shishmarev, "On the Existence and Breaking the Waves Described by the Whitham Equation," *Soviet Physics-Doklady*, Vol. 384, No. 31, 1986, pp. 90-95.

[4]   V. G. Galenin and V. V. Kuznetsov, "Simulation of Wave Transformation in Coastal Zone," *Water Resources*, Vol. 7, No. 1, 1980, pp. 156-165.

[5]   U. Kanoglu, "Nonlinear Evolution and Run Up-Run Down of Long Waves over a Sloping Beach," *Journal of Fluid Mechanics*, Vol. 513, 2004, pp. 363-372.

[6]   I. I. Didenkulova, N. Zaibo, A. A. Kurkin and E. N. Pelinovskiy, "Steepness and Spectrum of Nonlinear Deformed Wave under Water Conditions," *Izvestiya RAN Atmospheric and Oceanic Physics*, Vol. 42, No. 6, 2006, pp. 839-842.

[7]   N. A. Kudryashov, Yu. I. Sytsko and S. A. Chesnokov, "Mathematic Simulation of Gravity Waves in the Ocean in 'Shallow Water' Approximation," *Pisma v ZhETF*, Vol. 77, No. 10, 2003, p. 649.

[8]   A. A. Litvinenko and G. A. Khabakhpashev, "Computer Simulation of Nonlinear Considerably Long Two-dimensional Waves under Water Conditions in Basins with Sloping Floor," *Computer technologies*, Vol. 4, No. 3, 1999, pp. 95-105.

[9]   S. Yu. Kuznetsov and Ya. V. Saprykina, "Experimental Researches of Wave Group Evolution in the Coastal Sea Zone," *Oceanology*, Vol. 42, No. 3, 2002, p. 356-363.

[10] I. B. Abbasov, "Study and Simulation of Nonlinear Surface Waves under Shallow-Water Conditions," *Izvestiya RAN Atmospheric and Oceanic Physics*, Vol. 39, No. 4, 2003, pp. 506-511.

[11] H. Lamb, "Hydrodynamics," Dover, New York, 1930, p. 524.

[12] G. Whitham, "Linear and Nonlinear Waves," Wiley, New York, 1974, p. 622.

[13] L. M. Brekhovskikh and V. V. Goncharov, "Introduction to the Mechanics of Continuous Media," *Nauka*, 1982, p. 325.

[14] M. B. Vinogradova, O. V. Rudenko and A. P. Sukhorukov, "Wave Theory," *Nauka*, 1979, p. 383.

[15] A. K. Monin and V. P. Krasitskiy, "Phenomena at the Ocean Surface," *L. Gidrometeoizdat*, 1985, p. 375.

[16] I. O. Leontiev, "Coastal Dynamics: Waves, Streams, Burden Streams," *M.: GEOS*, 2001, p. 272.

[17] R. E. Flick, R. T. Guza and D. L. Inman, "Elevation and Velocity Measurements of Laboratory Shoaling Waves," *Journal of Geophysical Research*, Vol. 86, No. 5, 1981, pp. 4149-4160.

[18] M. S. Lonquet-Higgins, "Grest Instabilities of Gravity Waves, Part 1," *Journal of Fluid Mechanics*, Vol. 258, No. 1, 1994, pp. 115-129.

[19] V. Zaharov, "Weakly Non-Linear Waves on the Surface of an Ideal Finite Depth Fluid," *American Mathematical Society Transactions*, Series 2, Vol. 182, 1998, pp. 167-197.

[20] S. A. Gabov, "Introduction to the Theory on Nonlinear Waves," Moscow State Unversity, 1988, p.176.

[21] Y. Goda and K. Morinobu, "Breaking Wave Heights on Horizontal Bed Affected by Approach Slope," *Coastal Engineering Journal*, Vol. 40, No. 4, 1998, pp. 307-326.

[22] Ya. V. Saprykina, S. Yu. Kuznetsov, Zh. Tcherneva and N. Andreeva, "Space-Time Unsteadiness of the Amplitude-Phase Structure of Storm Waves in the Coastal Sea Zone," *Oceanology*, Vol. 49, No. 2, 2009, pp. 198-208.

[23] K. Kawasaki, "Numerical Simulation of Breaking and Post-Breaking Wave Deformation Process around a Submerged Breakwater," *Coastal Engineering Journal*, Vol. 41, No. 3-4, 1999, pp. 201-223.

A Characterization of Semilinear Surjective Operators and Applications to Control Problems*

Edgar Iturriaga, Hugo Leiva

Grupo de Matemática Aplicada, Departamento de Matemáticas, Universidad de Los Andes,
Mérida, Venezuela
E-mail: *iturri@ula.ve and hleiva@ula.ve*
Received April 28, 2010; revised July 30, 2010; accepted August 3, 2010

Abstract

In this paper we characterize a broad class of semilinear surjective operators $G_H : V \to Z$ given by the following formula $G_H w = Gw + H(w)$, $w \in V$, where V, Z are Hilbert spaces, $G \in L(V,Z)$ and $H : V \longrightarrow Z$ is a suitable nonlinear function. First, we give a necessary and sufficient condition for the linear operator G to be surjective. Second, we prove the following statement: If $Rang(G) = Z$ and H is a Lipschitz function with a Lipschitz constant h small enough, then $Rang(G_H) = Z$ and for all $z \in Z$ the equation $Gw + H(w) = z$ admits the following solution $w_z = G^*(GG^*)^{-1}(I + H \circ G^*(GG^*)^{-1})^{-1}z$. We use these results to prove the exact controllability of the following semilinear evolution equation $z' = Az + Bu(t) + F(t,z,u(t))$, $z \in Z, u \in U, t > 0$, where Z, U are Hilbert spaces, $A : D(A) \subset Z \longrightarrow Z$ is the infinitesimal generator of strongly continuous semigroup $\{T(t)\}_{t \geq 0}$ in Z, $B \in L(U,Z)$, the control function u belong to $L^2(0,\tau;U)$ and $F : [0,\tau] \times Z \times U \longrightarrow Z$ is a suitable function. As a particular case we consider the semilinear damped wave equation, the model of vibrating plate equation, the integrodifferential wave equation with Delay, etc.

Keywords: Semilinear Surjective Operators, Evolution Equations, Controllability, Damped Wave Equation

1. Introduction

In this paper we characterize a broad class of semilinear surjective operators

$G_H : V \longrightarrow Z$ given by the following formula

$$G_H w = Gw + H(w), w \in V, \qquad (1.1)$$

Where Z, V are Hilbert spaces, $G : V \longrightarrow Z$ is a bounded linear operator (continuous and linear) and $H : V \longrightarrow Z$ is a suitable non linear function in general nonlinear. First, we give a necessary and sufficient condition for the linear operator G to be surjective. Second, we prove the following statement: If $Rang(G) = Z$ and H is a Lipschitz function with a Lipschitz constant h small enough, then $Rang(G_H) = Z$ and for all $z \in Z$ the equation

$Gw + H(w) = z$, admits the following solution

$$w_z = G^*(GG^*)^{-1}(I + H \circ G^*(GG^*)^{-1})^{-1}z.$$

We apply our results to prove the exact controllability of the following semilinear evolution equation

$$z' = Az + Bu(t) + F(t,z,u(t)), z \in Z, u \in U, t > 0, \quad (1.2)$$

where Z and U are Hilbert spaces, $A : D(A) \subset Z \longrightarrow Z$ is the infinitesimal generator of strongly continuous semigroup $\{T(t)\}_{t \geq 0}$ in $Z, B \in L(U,Z)$, the control function u belong to $L^2(0,\tau;U)$ and $F : [0,\tau] \times Z \times U \longrightarrow Z$ is a suitable function. We give a necessary and sufficient condition for the exact controllability of the linear system

$$z' = Az + Bu(t), z \in Z, u \in U, t > 0. \qquad (1.3)$$

Under some conditions on F, we prove that the controllability of the linear system (1.3) is preserved by the semilinear system (1.2). In this case the control $u \in L^2(0,\tau;U)$ steering an initial state z_0 to a final state z_1 at time $\tau > 0$ (using the non linear system (1.2)) is given by the following formula:

$$u(t) = B^*T^*(\tau - t)W^{-1}(I + K)^{-1}(z_1 - T(\tau)z_0),$$

where $K : Z \to Z$ is non linear operator given by:

$$K\xi = \int_0^\tau T(\tau - s)F(s, z_\xi(s), (S\xi)(s))ds,$$

and $z_\xi(\cdot)$ is the solution of (1.2) corresponding to the

control u define by:

$$u(t) = (S\xi)(t) = B^*T^*(\tau - t)W^{-1}\xi, \quad t \in [0, \tau].$$

As an application we consider some control systems governed by partial differential equations, integrodifferential equations and difference equations that can be studied using these results. Particularly, we work in details the following controlled damped wave equation

$$\begin{cases} w_{tt} + cw_t - dw_{xx} = u(t,x) + f(t, u(t,x), w, w_t) \\ 0 < x < 1 \\ w(t,0) = w(t,1) = 0 \\ t \in IR \end{cases} \quad (1.4)$$

where $d > 0$, $c \geq 0$ the distributed control $u \in L^2(0, t_1; L^2(0,1))$ and the nonlinear term $f(t, w, v, u)$ is a function $f : [0, t_1] \times IR^3 \to IR$. A physical interpretation of the nonlinear term $f(t, u, w, w_t)$ could be as an eternal force like in the suspension bridge equation proposed by Lazer and McKenna (see [1]).

The novelty in this work lies in the following facts: First, the main results are obtained by standard and basic functional analysis such as Cauchy-Schwarz inequality, Hahn-Banach theorem, the open mapping theorem, etc. Second, the results are so general that can be apply to those control systems governed by evolutions equations like the one studied in [1-3] and [4]. Third, we find a formula for a control steering the system from the initial state z_0 to a final state z_1 on time $\tau > 0$, for both the linear and the nonlinear systems, which is very important from engineering point of view. Also, we present here a variational approach to find solutions of the semilinear equation $Gw + H(w) = z$ which is motivated by the one used to prove the interior controllability for some control system governed by PDE's, see [5]. Finally, these results can be used to motivate the study of semilinear range dense operator in order to characterize the approximate controllability of evolution equations.

2. Surjective Linear Operator

In this section we shall presents a characterization of surjective bounded linear operator. To this end, we denote by $L(V, Z)$ the space of linear and bounded operators mapping V to Z endow with the norm of the uniform convergence, and we will use the following lemma from [6] in Hilbert space:

Lemma 1. Let $G^* \in L(Z, V)$ be the adjoint operator of $G \in L(V, Z)$. Then the following statements holds:
1) $Rang(G) = Z \Longleftrightarrow \exists \gamma > 0$ such that

$$G^*z_V \geq \gamma z_Z, z \in Z.$$

2) $\overline{Rang(G)} = Z \Longleftrightarrow Ker(G^*) = \{0\}$.

In the same way as definition 4.1.3 from [7] we define the following concept:

Definition 1. The generalize controllability gramian of the operator G is define by:

$$W = GG^* : Z \to Z. \quad (2.1)$$

Theorem 1. An operator $G \in L(V, Z)$ is surjective if, and only if, the operator $W = GG^*$ is invertible. Under this condition, for all $z \in Z$ the equation

$$Gw = z, \quad (2.2)$$

admits the following solution

$$w_z = G^*(GG^*)^{-1}z = G^*W^{-1}z. \quad (2.3)$$

Moreover, this solution has minimum norm. *i.e.*,

$$w_z = \inf\{w : Gw = z, w \in V\}, \quad (2.4)$$

and $w_z = w$, $Gw = z \Longleftrightarrow w_z = w$.

Proof Suppose G is surjective. Then, from the foregoing Lemma there exists $\gamma > 0$ such that

$$G^*z_V^2 \geq \gamma^2 z_Z^2, z \in Z.$$

Therefore,

$$< GG^*z, z > = < Wz, z > \geq \gamma^2 z_Z^2 \quad z \in Z. \quad (2.5)$$

This implies that W is one to one. Now, we shall prove that W is surjective. That is to say

$$R(W) = \text{Range}(W) = Z.$$

For the purpose of contradiction, let us assume that $R(W)$ is strictly contained in Z. Using Cauchy Schwarz's inequality and (2.5) we get

$$Wz \geq \gamma^2 z_Z, \quad z \in Z$$

which implies that $R(W)$ is closed. Then, from Hahn Banach's Theorem there exists $z_0 \in Z$ with $z_0 \neq 0$ such that

$$\langle Wz, z_0 \rangle = 0, \ \forall z \in Z.$$

In particular, putting $z = z_0$ we get from (2.5) that

$$0 = \langle Wz_0, z_0 \rangle \geq \gamma^2 z_0{}_Z^2.$$

Then $z_0 = 0$, which is a contradiction. Hence, W is a bijection and from the open mapping Theorem W^{-1} is a bounded linear operator.

Now, suppose W is invertible. Then, given $z \in Z$ we shall prove the existence of $w \in V$ such that $Gw = z$. This w can be taking as follows

$$w_z = G^*W^{-1}z.$$

In fact,

$$Gw_z = GG^*W^{-1}z = GG^*(GG^*)^{-1}z = z.$$

Now, we shall see that the solution

$w_z = G^*(GG^*)^{-1}z = G^*W^{-1}z$ of the Equation (2.2) has minimum norm. In fact, let $w \in V$ such that $Gw = z$ and consider

$$w^2 = w_z + (w - w_z)^2 = w_z^2$$
$$+ 2\mathrm{Re}\langle w_z, w - w_z \rangle + w - w_z^2.$$

On the other hand,

$$\langle w_z, w - w_z \rangle = \langle G^*W^{-1}z, w - w_z \rangle =$$
$$= \langle W^{-1}z, Gw - Gw_z \rangle = \langle W^{-1}z, z - z \rangle = 0.$$

Hence, $w^2 - w_z^2 = w - w_z^2 \geq 0$.

Therefore, $w_z \leq w$, and $w_z = w$ if $w_z = w$.

Corollary 1. If an operator $G \in L(V, Z)$ is surjective, then the operator

$S : Z \longrightarrow V$ defined by:

$$S\xi = G^*W^{-1}\xi, \qquad (2.6)$$

is a right inverse of G. *i.e.*, $G \circ S = I$.

Definition 2. Under the condition of the above theorem the operator

$$S = G^*W^{-1} : Z \longrightarrow V,$$

is called the generalize steering operator.

Lemma 2. An operator $G \in L(V, Z)$ satisfies $\overline{Rang(G)} = Z$ if, and only if, $\overline{Rang(W)} = Z$.

Proof Suppose that $\overline{Rang(G)} = Z$. Then, from Lemma 1 part (2) we have that

$$\langle Wz, z \rangle > 0, \forall z \in Z, z \neq 0. \qquad (2.7)$$

For the purpose of contradiction, let us assume that

$$\overline{Rang(W)} \subset Z.$$

Then, from Hanh Banach's Theorem there exists $z_0 \neq 0$ such that

$$\langle Wz, z_0 \rangle = 0, \ \forall z \in Z.$$

In particular, if we put $z = z_0$, then $\langle Wz_0, z_0 \rangle = 0$, which contradicts (2.7).

Now, suppose that $\overline{Rang(W)} = Z$. Then, $Rang(GG^*) = Z$, and consequently $\overline{Rang(G)} = Z$.

2.1. Variational Method to Obtain Solutions

The Theorem 1 gave a formula for one solution of the system (2.2) which has minimum norma. But, it is not the only way allowing to build solutions of this equation. Next, we shall present a variational method to obtain solutions of (2.2) as a minimum of the quadratic functional $\jmath : Z \to IR$,

$$\jmath(\xi) = \frac{1}{2}G^*\xi^2 - \langle z, \xi \rangle, \forall \xi \in Z. \qquad (2.8)$$

Proposition 1. For a given $z \in Z$ the Equation (2.2) has a solution $w \in V$ if, and only if,

$$\langle w, G^*\xi \rangle - \langle z, \xi \rangle = 0, \quad \forall \xi \in Z. \qquad (2.9)$$

It is easy to see that (2.9) is in fact an optimality condition for the critical points of the quadratic functional $\jmath$ define above.

Lemma 3. Suppose the quadratic functional $\jmath$ has a minimizer $\xi_z \in Z$. Then,

$$w_z = G^*\xi_z, \qquad (2.10)$$

is a solution of (2.2).

Proof. First, observe that $\jmath$ has the following form

$$\jmath(\xi) = \frac{1}{2}\langle GG^*\xi, \xi \rangle - \langle z, \xi \rangle, \quad \forall \xi \in Z.$$

Then, if ξ_z is a point where $\jmath$ achieves its minimum value, we obtain that

$$\frac{d}{d\xi}\{\jmath\}(\xi_z) = GG^*\xi_z - z = 0.$$

So, $GG^*\xi_z = z$ and $w_z = G^*\xi_z$ is a solution of (2.2).

Remark 1. Under the condition of Theorem 1, the solution given by the formulas (2.10) and (2.3) coincide.

Theorem 2. The system (2.2) is solvable if, and only if, the quadratic functional $\jmath$ defined by (2.8) has a minimum for all $z \in Z$.

Proof Suppose (2.2) is solvable. Then, the operator G is surjective. Hence, from Lemma 1 there exists $\gamma > 0$ such that

$$G^*\xi^2 \geq \gamma^2\xi^2, \xi \in Z.$$

Then,

$$\jmath(\xi) \geq \frac{\gamma^2}{2}\xi^2 - z\xi, \xi \in Z.$$

Therefore,

$$\lim_{\xi \to \infty} \jmath(\xi) = \infty.$$

Consequently, $\jmath$ is coercive and the existence of a minimum is ensured. The other way of the proof follows as in proposition 1.

3. Surjective Semlinear Operators

In this section we shall look for conditions under which the semilinear operator

$G_H : V \longrightarrow Z$ given by:

$$G_H w = Gw + H(w), \quad w \in V, \qquad (3.1)$$

is surjective. To this end, we shall use the following theorem from non linear analysis.

Theorem 1. Let Z be a Banach space and $K : Z \to Z$ a Lipschitz function with a Lipschitz constant $k < 1$ and consider $\hat{G}(z) = z + Kz$. Then $\hat{G}$ is an homemorphisme whose inverse is a Lipschitz function with a Lipschitz constant $(1 - k)^{-1}$.

Theorem 2. If $Rang(G) = Z$ and H is a Lipschitz function with a Lipschitz constant h such that $hG^*(GG^*)^{-1} < 1$, then $Rang(G_H) = Z$ and for all $z \in Z$ the equation

$$G_H w = Gw + H(w) = z, \qquad (3.2)$$

admits the following solution

$$\begin{aligned} w_z &= G^*(GG^*)^{-1}(I + H \circ G^*(GG^*)^{-1})^{-1}z = \\ &= S(I + H \circ S)^{-1}z, \end{aligned} \qquad (3.3)$$

where $S = G^*(GG^*)^{-1}$.

Proof Suppose that $Rang(G) = Z$. Then, from Corollary 1 we know that the operator S define by (2.6) is a right inverse of G, so if we put $G_H = G_H \circ S$, we get the new operator

$$G_H \xi = G_H \circ S \xi = \xi + H(S\xi), \ \xi \in Z, \qquad (3.4)$$

where $S = G^*(GG^*)^{-1} = G^*W^{-1}$. Hence, if we define the operator $K : Z \longrightarrow Z$ by:

$$K\xi = H(S\xi) \qquad (3.5)$$

the operator G_H can be written as follows

$$G_H \xi = \xi + K\xi = (I + K)\xi, \ \xi \in Z. \qquad (3.6)$$

On the other hand, K is a Lipschitz function with a Lipschitz constant $\kappa \le hG^*(GG^*)^{-1} < 1$. Then, applying Theorem 1 we get the result.

Theorem 3. If $Rang(G) = Z$ and the operator K given by (3.5) is linear and $K \ge 0$, then $Rang(G_H) = Z$ and for all $z \in Z$ the equation

$$G_H w = Gw + H(w) = z, \qquad (3.7)$$

admits the following solution

$$w_z = G^*W^{-1}(I + K)^{-1}z.$$

Proof Since $G_H \circ S = I + K$, then

$$\langle (G_H \circ S)z, z \rangle > z^2, \ \forall z \in Z.$$

Then, in the same way as in the proof of Theorem 1 we get the result.

Corollary 1. Under the conditions of the above Theorems, the operator

$$\Gamma : Z \longrightarrow V \text{ define by:}$$

$$\begin{aligned} \Gamma z &= G^*(GG^*)^{-1}(I + H \circ G^*(GG^*)^{-1})^{-1}z = \\ &= S(I + H \circ S)^{-1}z, \end{aligned} \qquad (3.8)$$

is a right inverse of G_H. *i.e.*, $G_H \circ \Gamma = I$ d

Corollary 2. Under the conditions of the above

Theorems, the solution

$w_z = S(I + H \circ S)^{-1}z$ of the Equation (3.2) depends continuously on z. Moreover,

$$w_z - w_y{}_V \le \frac{S}{1 - hS}z - y_Z, \forall z, y \in Z.$$

4. Controllability of Semilinear Evolution Equations

In this section we shall characterize the exact controllability of the semilinear evolution equation

$$\begin{aligned} z' &= Az + Bu(t) + F(t, z, u(t)) \\ z &\in Z, u \in U, t > 0, \end{aligned} \qquad (4.1)$$

Where Z, U are Hilbert spaces, $A : D(A) \subset Z \longrightarrow Z$ is the infinitesimal generator of strongly continuous semigroup $\{T(t)\}_{t \ge 0}$ in Z, $B \in L(U, Z)$, the control function u belongs to $L^2(0, \tau; U)$ and $F : [0, \tau] \times Z \times U \longrightarrow Z$ is a suitable function.

4.1. Linear Systems

First, we shall study the controllability of the linear system (1.3), and to this end, for all $z_0 \in Z$ and $u \in L^2(0, \tau; U)$ the the initial value problem

$$\begin{cases} z' = Az(t) + Bu(t), \ t > 0 \\ z(0) = z_0, \end{cases} \qquad (4.2)$$

admits only one mild solution given by:

$$z(t) = T(t)z_0 + \int_0^t T(t - s)Bu(s)ds \ t \in [0, \tau]. \qquad (4.3)$$

Definition 1. (**Exact Controllability**) The system (1.3) is is said to be exactly controllable on $[0, \tau]$, if for all $z_0, z_1 \in Z$ there exists a control $u \in L^2(0, \tau; U)$ such that the solution $z(t)$ of (4.3) corresponding to u, verifies: $z(\tau) = z_1$.

Consider the following bounded linear operator:

$$G : L^2(0, \tau; U) \to Z, Gu = \int_0^\tau T(\tau - s)Bu(s)ds, \qquad (4.4)$$

whose adjoint operator $G^* : Z \longrightarrow L^2(0, \tau; U)$ is given by:

$$(G^*\xi)(s) = B^*T^*(\tau - s)\xi, \ \forall s \in [0, \tau], \ \forall \xi \in Z. \qquad (4.5)$$

Then, the gramian $W = GG^* : Z \to Z$ takes the following classical form

$$Wz = \int_0^\tau T(\tau - s)BB^*T^*(\tau - s)zds. \qquad (4.6)$$

Then, the following Theorem from [7](pg. 47, Theorem 4.17) is a characterization of the exact controllability of the linear system (1.3).

Theorem 1. For the system (1.3) we have the following condition for exact controllability.

System (1.3) is exactly controllable on $[0,\tau]$ if, and only if, any one of the following condition hold for some $\gamma > 0$ and all $z \in Z$:

1) $G(L^2(0,\tau;U)) = \text{Range}(G) = Z$,

2) $\langle Wz,z \rangle \geq \gamma \| z \|_Z^2$,

3) $\| G^*z \|_Z^2 := \int_0^\tau \| (Gz^*)(s) \|_U^2 ds \geq \gamma \| z \|_Z^2$

4) $\int_0^\tau \| B^*T^*(\tau-s)z \|_U^2 \geq \gamma \| z \|_Z^2$,

5) $Ker(G^*) = \{0\}$ and $Rang(G^*)$ is closed.

Remark 1. One can observe that the invertibility of the operator W is not proved in the foregoing theorem and, consequently, none formula for the control steering the system (4.2) from initial state z_0 to a final state z_1 on time $\tau > 0$ is given.

Now, we are ready to formulate and prove a new result on exact controllability of the linear system (1.3).

Theorem 2. The system (1.3) is exactly controllable on $[0,\tau]$ if, and only if, the operator W is invertible. Moreover, the control $u \in L^2(0,\tau;U)$ steering an initial state z_0 to a final state z_1 at time $\tau > 0$ is given by the following formula:

$$u(t) = B^*T^*(\tau-t)W^{-1}(z_1 - T(\tau)z_0). \quad (4.7)$$

Proof It follows directly from the above notation and applying Theorem 1.

Corollary 1. If the system (1.3) is exactly controllable, then the operator

$S : Z \longrightarrow L^2(0,\tau;U)$ define by:

$$S\xi = G^*W^{-1}\xi \text{ or } (S\xi)(s) = B^*T^*(\tau-s)W^{-1}\xi \quad (4.8)$$

is a right inverse of G. *i.e.*, $G \circ S = I$

In this case the Equation (2.2) takes the following form

$$\begin{aligned} Gu &= z \\ u &\in L^2(0,\tau;U), z \in Z \end{aligned} \quad (4.9)$$

and the quadratic functional $\jmath$ given by (2.8) can be written as follows

$$\Im(\xi) = \frac{1}{2}\int_0^\tau B^*T^*(\tau-s)\xi^2 ds - \langle z,\xi \rangle, \xi \in Z. (4.10)$$

The following results follow from Proposition 1, Lemma 3 and Theorem 1 respectively.

Proposition 1. For a given $z \in Z$ the Equation (4.9) has a solution $u \in L^2(0,\tau;U)$ if, and only if,

$$\int_0^\tau < u(t), B^*T^*(\tau-t)\xi > ds = <z,\xi>, \forall \xi \in Z. (4.11)$$

It is easy to see that (4.11) is in fact an optimality condition for the critical points of the quadratic functional $\Im$ define above.

Lemma 1. Suppose the quadratic functional $\Im$ has a minimizer $\xi_z \in Z$. Then,

$$u(t) = B^*T^*(\tau-t)\xi_z, t \in [0,\tau], \quad (4.12)$$

is a solution of (4.9).

Theorem 3. The system (1.3) is exactly controllable if, and only if, the quadratic functional $\Im$ define by (1) has a minimizer ξ_z for all $z \in Z$.

Moreover, under this condition we obtain that

$$u(t) = B^*T^*(\tau-t)\xi_z = B^*T^*(\tau-t)W^{-1}z, t \in [0,\tau], \quad (4.13)$$

and $\xi_z = W^{-1}z$.

4.2. Nonlinear System

We assume that F is good enough such that the Equation (4.1) with the initial condition $z(0) = z_0$ and a control $u \in L^2(0,\tau,U)$ admits only one mild solution given by:

$$\begin{aligned} z(t) &= T(t)z_0 + \int_0^t T(t-s)Bu(s)ds + \\ &\int_0^t T(t-s)F(s,z(s),u(s))ds, t \in [0,\tau] \end{aligned} \quad (4.14)$$

Definition 2. (**Exact Controllability**) The system (4.1) is said to be exactly controllable on $[0,\tau]$, if for all $z_0, z_1 \in Z$, there exists a control $u \in L^2(0,\tau;U)$ such that the corresponding solution z of (4.14) satisfies $z(\tau) = z_1$.

Define the following operator: $G_F : L^2(0,\tau;U) \longrightarrow Z$ by

$$\begin{aligned} G_F u &= \int_0^\tau T(\tau-s)Bu(s)ds + \int_0^\tau T(\tau-s)F(s,z(s),u(s))ds \\ &= Gu + \int_0^\tau T(\tau-s)F(s,z(s),u(s))ds \end{aligned} \quad (4.15)$$

where $z(t) = z(t,u)$ is the solution of (4.14) corresponding to the control u. Then, the following proposition is trivial and characterizes the exact controllability of (4.1).

Proposition 2. The system (4.1) is exactly controllable on $[0,\tau]$ if, and only if,

$$Rang(G_F) = Z.$$

So, in order to prove exact controllability of system (4.1) we have to verify the condition of the foregoing proposition. To this end, we need to assume that the linear system (1.3) is exactly controllable. In this case we know from corollary 1 that the steering operator S defined by (4.8) is a right inverse of G, so if we put $G_F = G_F \circ S$, we get the following representation:

$$G_F\xi = G_F \circ S\xi = \xi + \int_0^\tau T(\tau-s)F(s,z_\xi(s),(S\xi)(s))ds \quad (4.16)$$

where $z_\xi(\cdot)$ is the solution of (4.14) corresponding to

the control u define by:

$$u(t) = (S\xi)(t) = B^* T^* (\tau - t) W^{-1} \xi,\ t \in [0, \tau].$$

Hence, if we define the operator $K : Z \longrightarrow Z$ by:

$$K\xi = \int_0^\tau T(\tau - s) F(s, z_\xi(s), (S\xi)(s)) ds \qquad (4.17)$$

the operator G_F can be written as follows

$$G_F \xi = \xi + K\xi = (I + K)\xi \qquad (4.18)$$

Now, we shall prove some abstract results making assumptions on the operator K. After that, we will put conditions on the nonlinear term F that imply condition on K.

Theorem 4. If the linear system (1.3) is exactly controllable on $[0, \tau]$ and the operator K is globally Lipschitz with a Lipschitz constant $k < 1$, then the non linear system (4.1) is exactly controllable on $[0, \tau]$ and the control steering the initial state z_0 to the final state z_1 is given by:

$$u(t) = B^* T^* (\tau - t) W^{-1} (I + K)^{-1} (z_1 - T(\tau) z_0).$$

Proof It follows directly from Equation (3.6) and Theorem 1 or Theorem 2.

Theorem 5. If the system (1.3) is exactly controllable on $[0, \tau]$ and the operator K is linear with $K \geq 0$, then the system (4.1) is exactly controllable on $[0, \tau]$ and the control $u(t)$ steering the initial state z_0 to the final state z_1 is given by:

$$u(t) = B^* T^* (\tau - t) W^{-1} (I + K)^{-1} (z_0 - T(\tau) z_1)$$

Proof It follows directly from Equation (3).

The proof of the following lemma follows as in lemma 5.1 from [1].

Lemma 2. If F satisfies the Lipschitz condition

$$F(t, z_2, u_2) - F(t, z_1, u_1) \leq L\{z_2 - z_1 + u_2 - u_1\}$$

$$z_2, z_1 \in Z \quad u_2, u_1 \in U \quad t \in [0, \tau],$$

then

$$Kz_2 - Kz_1 \leq K(L) z_2 - z_1,\ z_2, z_1 \in Z$$

where $K(L) = L H_1 H_2$ and

$$H_1 = M[B + L] e^{ML\tau} \tau + 1, H_2 =$$
$$= M^2 B W^{-1} \tau, M = \sup_{0 \leq s \leq t \leq \tau} T(t - s).$$

Theorem 6. If F satisfies the foregoing Lipschitz condition, the linear system (1.3) is exactly controllable on $[0, \tau]$ and

$$L(M[B + L] e^{ML\tau} \tau + 1)(M^2 B W^{-1} \tau) < 1 \quad (4.19)$$

then the non linear system (4.1) is exactly controllable on $[0, \tau]$ and the control steering the initial state z_0 to the final state z_1 is given by:

$$u(t) = B^* T^* (\tau - t) W^{-1} (I + K)^{-1} (z_1 - T(\tau) z_0).$$

Proof From Lemma 2 we know that K is a Lipschitz function with a Lipschitz constant k given by:

$$k = L(M[B + L] e^{ML\tau} \tau + 1)(M^2 B W^{-1} \tau)$$

and from condition (4.19) we get that $k < 1$. Hence, applying Theorem 4 we complete the proof.

5. Applications and Further Research

In this section we consider some control systems governed by partial differential equations, integrodifferential equations and difference equations that can study using these results. Particularly, we work in details the controlled damped wave equation. Finally, we propose future investigations an open problem.

5.1. The Controlled Semilinear Damped Wave Equation

Consider the following control system governed by a $1D$ semilinear damped wave equation

$$\begin{cases} w_{tt} + c w_t - d w_{xx} = u(t, x) + f(t, u(t, x), w, w_t) \\ 0 < x < 1 \\ w(t, 0) = w(t, 1) = 0 \\ t \in IR \end{cases} \qquad (5.1)$$

where $d > 0$, $c \geq 0$, the distributed control $u \in L^2(0, t_1; L^2(0, 1))$ and the nonlinear term $f(t, w, v, u)$ is a function $f : [0, t_1] \times IR^3 \to IR$.

Abstract Formulation of the Problem.

Now we will choose the space in which problem (5.1) will be set as an abstract first order ordinary differential equation.

Let $X = L^2[0, 1]$ and consider the linear unbounded operator

$A : D(A) \subset X \to X$ defined by $A\varphi = -\varphi_{xx}$, where

$$D(A) = \{\varphi \in X : \varphi, \varphi_x, \text{are a.c}, \varphi_{xx} \in X; \varphi(0) = \varphi(1) = 0\}. \qquad (5.2)$$

The operator A has the following very well known properties:

1) The spectrum of A consists of only eigenvalues

$$0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n \to \infty,$$

each one with multiplicity one.

2) There exists a complete orthonormal set $\{\varphi_n\}$ of eigenvectors of A.

3) For all $x \in D(A)$ we have

$$Ax = \sum_{n=1}^\infty \lambda_n < x, \varphi_n > \varphi_n = \sum_{n=1}^\infty \lambda_n E_n x, \qquad (5.3)$$

where $< \cdot, \cdot >$ is the inner product in X and

$$E_n x = <x, \varphi_n> \varphi_n, \lambda_n = n^2 \pi^2$$
$$\text{and } \varphi_n(x) = \sqrt{2} \sin(n\pi x). \tag{5.4}$$

So, $\{E_n\}$ is a family of complete orthogonal projections in X and

$$x = \sum_{n=1}^{\infty} E_n x, x \in X.$$

4) $-A$ generates an analytic semigroup $\{e^{-At}\}$ given by:

$$e^{-At} x = \sum_{n=1}^{\infty} e^{-\lambda_n t} E_n x. \tag{5.5}$$

and

5) The fractional powered spaces X^r are given by:

$$X^r = D(A^r) = \left\{ x \in X : \sum_{n=1}^{\infty} (\lambda_n)^{2r} E_n x^2 < \infty \right\} \quad r \geq 0,$$

with the norm

$$x_r = A^r x = \left\{ \sum_{n=1}^{\infty} \lambda_n^{2r} E_n x^2 \right\}^{1/2}, x \in X^r$$

where

$$A^r x = \sum_{n=1}^{\infty} \lambda_n^r E_n x. \tag{5.6}$$

Also, for $r \geq 0$ we define $Z_r = X^r \times X$, which is a Hilbert Space with norm given by:

$$\left\| \begin{bmatrix} w \\ v \end{bmatrix} \right\|_{Z_r}^2 = w_r^2 + v^2.$$

Using the change of variables $w' = v$, the system (5.1) can be written as a first order systems of ordinary differential equations in the Hilbert space $Z_{1/2} = D(A^{1/2}) \times X = X^{1/2} \times X$ as:

$$z' = Az + Bu + F(t, z, u(t)), z \in Z_{1/2}, t \geq 0, \tag{5.7}$$

where

$$z = \begin{bmatrix} w \\ v \end{bmatrix}, B = \begin{bmatrix} 0 \\ I_X \end{bmatrix}, A = \begin{bmatrix} 0 & I_X \\ -dA & -cI_X \end{bmatrix}, \tag{5.8}$$

is an unbounded linear operator with domain $D(A) = D(A) \times X$ and

$$F(t, z, u) = \begin{bmatrix} 0 \\ f(t, u, w, v) \end{bmatrix} \tag{5.9}$$

and the function F: $[0, t_1] \times Z_{1/2} \times X \to Z_{1/2}$. Since $X^{1/2}$ is continuously included in X we obtain for all $z_1, z_2 \in Z_{1/2}$ and $u_1, u_2 \in X$ that

$$F(t, z_2, u_2) - F(t, z_1, u_1)_{Z_{1/2}}$$
$$\leq L\{z_2 - z_{1 \, 1/2} + u_2 - u_1\} \quad t \in [0, t_1]. \tag{5.10}$$

Throughout this section, without lose of generality, we will assume that

$$c^2 < 4d\lambda_1.$$

The following proposition follows from [8] and [1].

Proposition 1. The operator A given by (5.8), is the infinitesimal generator of strongly continuous group $\{T(t)\}_{t \in IR}$ in $Z_{1/2}$ given by:

$$T(t)z = \sum_{n=1}^{\infty} e^{A_n t} P_n z, z \in Z_{1/2}, \tag{5.11}$$

where $\{P_n\}_{n \geq 0}$ is a family of complete orthogonal projections on the Hilbert space $\mathbf{Z}_{\frac{1}{2}}$ given by:

$$P_n = diag(E_n, E_n), n \geq 1, \tag{5.12}$$

and

$$A_n = B_n P_n, B_n = \begin{pmatrix} 0 & 1 \\ -d\lambda_n & -c \end{pmatrix}, n \geq 1. \tag{5.13}$$

This group decays exponentially to zero. In fact, we have the following estimate

$$T(t) \leq M(c,d)e^{-\frac{c}{2}t}, t \geq 0, \tag{5.14}$$

where

$$\frac{M(c,d)}{2\sqrt{2}} = \sup_n \left\{ 2 \left| \frac{c \pm \sqrt{4d\lambda_n - c^2}}{\sqrt{c^2 - 4d\lambda_n}} \right|, (2+d) \left| \frac{\sqrt{\lambda_n}}{\sqrt{4d\lambda_n - c^2}} \right| \right\} \tag{5.15}$$

The proof of the following theorem follows in the same way as the one for Theorem 4.1 from [1].

Theorem 1. The system

$$\begin{cases} z' = Az + Buz \in Z_{1/2}, t > 0, \\ z(0) = z_0. \end{cases} \tag{5.16}$$

is exactly controllable on $[0, \tau]$.

Theorem 2. If the following estimate holds

$$L\left(M(c,d)[1+L]e^{M(c,d)L\tau}\tau + 1\right)\left(M(c,d)^2 W^{-1}\tau\right) < 1, \tag{5.17}$$

then the system (5.7) is exactly controllable on $[0, \tau]$.

Proof It follows from Theorem 6 one we observe that in this case $B \leq 1$.

5.2. Future Research

These results can be applied to the following class of second order diffusion system in Hilbert spaces

$$w'' + A_0 w = u(t) + f(t, w, u)$$
$$t > 0, \quad w \in W \quad u \in U$$

Where W, U are Hilbert spaces, the control $u \in L^2(0, \tau; U)$, $A_0 : D(A_0) \subset W \to W$ is an unbounded linear operator in W with the spectral decomposition:

$$A_0 w = \sum_{j=1}^{\infty} \lambda_j \sum_{k=1}^{\gamma_j} < \phi_{k,j}, w > \phi_{k,j} = \sum_{j=1}^{\infty} \lambda_j E_j w,$$

where $E_j w = \sum_{k=1}^{\gamma_j} < \phi_{k,j}, w > \phi_{k,j}$, $\{\varphi_{k,j}\}$ is a complete orthonormal set of eigenvectors of $-A_0$ correspondent to the eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_n \to \infty$ with multiplicity γ_n and $-A_0$ generates a strongly continuous semigroup $\{T(t)\}_{t \geq 0}$ given by:

$$T(t)w = \sum_{j=1}^{\infty} e^{-\lambda_j t} E_j w, \; w \in W, \; t \geq 0$$

and $f : [0, \tau] \times W \times U \longrightarrow W$ is a suitable function.

Examples of this class are the following well known systems of partial differential equations:

Example 1. The n D wave equation with Dirichlet boundary conditions

$$\begin{cases} \dfrac{\partial^2 w}{\partial^2 t} - \Delta w = u(t,x) + f(t, u(t,x), w) \\ t \geq 0, x \in \Omega \\[2mm] w(t,x) = 0 \\ t \geq 0, x \in \partial\Omega. \\[2mm] w(0,x) = \varphi_0(x) \\ \dfrac{\partial z}{\partial t}(0,x) = \psi_0(x), \\ x \in \Omega \end{cases} \quad (5.18)$$

where Ω is a sufficiently smooth bounded domain in IR^N, $u \in L^2\left([0,r]; L^2(\Omega)\right)$, $\varphi_0, \psi_0 \in L^2(\Omega)$ and f is a suitable function.

Example 2. The model of Vibrating Plate

$$\begin{cases} \dfrac{\partial^2 w}{\partial^2 t} + \Delta^2 w + = u(t,x) + f(t, u(t,x), w) \\ t \geq 0, x \in \Omega \\[2mm] w = \Delta w = 0 \\ t \geq 0, x \in \partial\Omega. \\[2mm] w(0,x) = \varphi_0(x) \\ \dfrac{\partial w}{\partial t}(0,x) = \psi_0(x) \\ x \in \Omega. \end{cases} \quad (5.19)$$

where Ω is a sufficiently smooth bounded domain in

IR^2, $u \in L^2\left([0,r]; L^2(\Omega)\right)$, $\varphi_0, \psi_0 \in L^2(\Omega)$ and f is a suitable function.

Others type of problems are the following control problems:

Example 3. Interior Controllability of the 1D Wave Equation

$$\begin{cases} y_{tt} - y_{xx} = 1_\omega u(t,x) + f(t, u(t,x), w, w_t) \\ \text{in } (0, \tau) \times (0,1) \\[2mm] y(t,0) = y(t,1) = 0, \quad t \in (0, \tau) \\ y(0,x) = y_0(x), \quad \text{in } [0,1] \end{cases} \quad (5.20)$$

where ω is an open nonempty subset of $[0,1]$, 1_ω denotes the characteristic function of the set ω, the distributed control $u \in L^2(0, \tau; L^2[0,1])$ and the nonlinear term $f(t, w, v, u)$ is a function $f : [0, t_1] \times IR^3 \to IR$. For the interior controllability of the linear wave equation one can see [5].

Example 4. Exact Controllability of Integrodifferential 1D Wave Equation with Delay.

$$\begin{cases} y_{tt} - y_{xx} = u(t,x) + \int_0^t p(s, y(s-r, x) ds \\ \text{in } (0, \tau) \times (0,1) \\[2mm] y(t,0) = y(t,1) = 0 \\ t \in (0, \tau) \\[2mm] y(s,x) = y_0(s,x) \\ \text{in } [-r, 0] \times [0,1] \\[2mm] y_t(s,x) = y_1(s,x) \\ \text{in } [-r, 0] \times [0,1] \end{cases} \quad (5.21)$$

where the distributed control $u \in L^2(0, \tau; L^2[0,1])$, $y_0, y_1 : [-r, 0] \times [0,1] \longrightarrow IR$ are continuous functions and the nonlinear term $p \in L^1([-r, \tau] \times [0,1])$.

Example 5. Exact controllability of Semilinear Difference Equations

$$\begin{aligned} z(n+1) &= A(n)z(n) + B(n)u(n) + \\ &\quad f(z(n), u(n)), n \in IN^*, z(0) = z_0 \end{aligned} \quad (5.22)$$

where Z, U are Hilbert spaces, $A \in l^\infty(IN, L(Z))$, $B \in l^\infty(IN, L(U, Z))$, $u \in l^2(IN, U)$, $L(U, Z)$ denotes the space of all bounded linear operators from U to Z and $L(Z, Z) = L(Z)$. The nonlinear term $f : Z \times U \longrightarrow Z$ is a continuous Lipschitzian function. That is to say: For all $z_2, z_1 \in Z$ and $u_1, u_2 \in U$ we have that

$$f(z_2, u_2) - f(z_1, u_1) \leq L\{z_2 - z_1 + u_2 - u_1\}. \quad (5.23)$$

AM

5.3. Open Problem

The solution of the following problem is very important, no only to study the approximate controllability of evolution Equation (1.2), it also can be used to solve others mathematical problems.

The problem can be formulate as follows: Let Z, W be Hilbert spaces, $G \in L(W,Z)$ and $H:W \longrightarrow Z$ is a suitable nonlinear function. When the following statement holds?

If $\overline{Rang(G)} = Z$ and H is a Lipschitz function with a Lipschitz constant h small enough, then $\overline{Rang(G+H)} = Z$ and for all $z \in Z$ there is a sequence $\{w_\alpha\}_{\alpha>0} \subset W$ such that equation

$$\lim_{\alpha \to 0^+} \left(Gw_\alpha + H(w_\alpha) \right) = z.$$

6. References

[1] H. Leiva, "Exact Controllability of the Suspension Bridge Model Proposed by Lazer and Mckenna," *Journal of Mathematical Analysis and Applications*, Vol. 309, No. 2, 2005, pp. 404-419.

[2] E. Iturriaga and H. Leiva, "A Necessary and Sufficient Condition for the Controllability of Linear Systems in Hilbert Spaces and Applications," *IMA Journal of Mathematics, Control and Information*, Vol. 25, No. 3, 2008, pp. 269-280.

[3] H. Leiva, "Exact Controllability of Semilinear Evolution Equation and Applications," *International Journal of Systems and communications*, Vol. 1, No. 1, 2008, pp. 1-12.

[4] H. Leiva and J. Uzcategui, "Exact Controllability of Semilinear Difference Equation and Application," *Journal of Difference Equations and Applications*, Vol. 14, No. 7, 2008, pp. 671-679.

[5] J. L. Lions, "Optimal Control of Systems Governed by Partial Differential Equations," Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen Band 170, 1971.

[6] R. F. Curtain and A. J. Pritchard, "Infinite Dimensional Linear Systems," Lecture Notes in Control and Information Sciences, Vol. 8, Springer Verlag, Berlin, 1978.

[7] R. F. Curtain and H. J. Zwart, "An Introduction to Infinite Dimensional Linear Systems Theory," Tex in Applied Mathematics, Vol. 21. Springer Verlag, New York, 1995.

[8] H. Leiva, "A Lemma on C_0-Semigroups and Applications PDEs Systems," *Quaestions Mathematicae*, Vol. 26, No. 3, 2003, pp. 247-265.

Scientific Research

On Robustness of a Sequential Test for Scale Parameter of Gamma and Exponential Distributions

Parameshwar V. Pandit[1], Nagaraj V. Gudaganavar[2]
[1]*Department of Statistics Bangalore University, Bangalore, India*
[2]*Department of Statistics Anjuman Arts, Science and Commerce, Dharwad, India*
E-mail: *panditpv*12@*gmail.com*

Abstract

The main aim of the present paper is to study the robustness of the developed sequential probability ratio test (SPRT) for testing the hypothesis about scale parameter of gamma distribution with known shape parameter and exponential distribution with location parameter. The robustness of the SPRT for scale parameter of gamma distribution is studied when the shape parameter has undergone a change. The similar study is conducted for the scale parameter of exponential distribution when the location parameter has undergone a change. The expressions for operating characteristic and average sample number functions are derived. It is found in both the cases that the SPRT is robust only when there is a slight variation in the shape and location parameter in the respective distributions.

1. Introduction

The robustness of sequential probability ratio test (SPRT) has been studied by several authors for various probability distributions when the distribution under consideration has undergone a change. Barlow and Proschan [1], Harter and Moore [2], Montagne and Singpurwalla [3], and others have studied this problem for various probability models.

In this paper, the problem of testing simple hypothesis against simple alternatives for scale parameter of the gamma distribution assuming shape parameter to be known is considered. The gamma distribution plays important role in many areas of the Statistics including areas of life testing and reliability. It is used to make realistic adjustment to exponential distribution in life-testing situations. The fact that a sum of independent exponentially distributed random variables has a gamma distribution, leads to the appearance of gamma distribution in the theory of random counters and other topics associated with precipitation processes.

In Section 2, we state the problem and develop SPRT for testing of hypothesis giving expressions for Operating Characteristic (OC) and Average Sample Number (ASN) functions. In Section 3, robustness of the devel-

oped SPRT with respect to OC functions when the distribution considered here has undergone a change in the shape parameter, has been studied. In Section 4 we have studied the robustness of SPRT for the scale parameter of exponential distribution with respect to OC function when the location parameter has undergone a change.

2. Materials and Methods

The set-up of the problem and SPRT
Let $X_1, X_2, \ldots.$ be a sequence of random variables from a gamma distribution with scale parameter $\theta \ (> 0)$ and shape parameter $\lambda \ (> 0)$, whose density function is given by

$$f(x:\theta,\lambda) = \frac{x^{\lambda-1}e^{-x/\theta}}{\theta^\lambda \Gamma \lambda}, x > 0 \qquad (1)$$

where it is assumed that λ is known. Suppose we want to test the null hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta = \theta_1 \ (> \theta_0)$. For this problem following SPRT is developed.

$$\text{Let } Z = \text{In} \frac{f(x;\theta_1)}{f(x;\theta_0)} = \lambda \text{In}\left(\frac{\theta_0}{\theta_1}\right) + \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)x \qquad (2)$$

Two numbers A and B $(0 < B < 1 < A)$ are chosen. At

nth stage accept H$_0$ if $\sum_{i=1}^{n} z_i \leq \ln B$, reject H$_0$ if

$\sum_{i=1}^{n} z_i \geq \ln A$, otherwise, continue sampling by taking (n + 1) th observation. Here Z_i is obtained by replacing X by X_i in (2). Let (α, β) be the desired strength of SPRT, then according Wald [4], A and B are approximately given by

$$A \approx \frac{1-\beta}{\alpha}, B \approx \frac{\beta}{1-\alpha}, \text{ where } \alpha \in (0,1) \text{ and } \beta \in (0,1).$$

The Operating Characteristic (OC) function $L(\theta)$ is given by

$$L(\theta) \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \quad (3)$$

Where $h(\theta)$ is the non-zero solution of

$$E\left(e^{h(\theta)z}\right) = 1 \quad (4)$$

From (1) and (2), since

$$E\left(e^{h(\theta)z}\right) = \left\{1 - h(\theta)\theta\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\right\}^{-\lambda} \left(\frac{\theta_0}{\theta_1}\right)^{h(\theta)\lambda}$$

we get from (4) that

$$\theta = \frac{1 - \left(\dfrac{\theta_0}{\theta_1}\right)^{h(\theta)}}{h(\theta)\left(\dfrac{1}{\theta_0} - \dfrac{1}{\theta_1}\right)} \quad (5)$$

The Average Sample Number (ASN) function is given by

$$E(N \mid \theta) = \frac{L(\theta)\ln B + [1 - L(\theta)]\ln A}{E(Z)} \quad (6)$$

Provided $E(Z) \neq 0$,
where

$$E(Z) = \lambda\left[\ln\left(\frac{\theta_0}{\theta_1}\right) + \theta\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\right] \quad (7)$$

The maximum value of ASN occurs in the neighbourhood of $\theta = \tilde{\theta}$, say where $\tilde{\theta}$ is the solution of $E_{\tilde{\theta}}(Z) = 0$ and this value is given by

$$E_{\tilde{\theta}}(N) = \frac{-\ln A . \ln B}{E_{\tilde{\theta}}(Z^2)} \quad (8)$$

It is easy to see that

$$\tilde{\theta} = \frac{\ln(\theta_1 / \theta_0)}{\left(\dfrac{1}{\theta_0} - \dfrac{1}{\theta_1}\right)} \quad (9)$$

and

$$E_{\tilde{\theta}}(Z^2) = \lambda\left\{\ln\left(\frac{\theta_0}{\theta_1}\right)\right\}^2 \quad (10)$$

Table 1 contains the values of $E_0(N)$, $E_1(N)$ and $E_{\tilde{\theta}}(N)$ for different values of α, β and ratios of and ratios of θ_0 and θ_1.

3. Results and Discussions

3.1. Robustness of the SPRT for Scale Parameter of Gamma Distribution

Let us suppose that there is misspecification for the shape parameter λ in the probability distribution. Then the pdf (1) becomes $f(x;\theta,\lambda*)$. To study the robustness of the SPRT developed in Section 2 with respect to OC function, consider $h(\theta)$ as the solution of the equation.

$$E_{\lambda*}\left(e^{h(\theta)z}\right) = 1$$

i.e. $\int_0^\infty \left[\frac{f(x;\theta_1,\lambda)}{f(x;\theta_0,\lambda)}\right]^{h(\theta)} f(x;\theta,\lambda*)dx = 1$

giving

$$\theta = \frac{1 - \left(\dfrac{\theta_0}{\theta_1}\right)^{\frac{\lambda}{\lambda*}h(\theta)}}{h(\theta)\left(\dfrac{1}{\theta_0} - \dfrac{1}{\theta_1}\right)} \quad (11)$$

For different values of θ, $h(\theta)$ is evaluated and the OC function is obtained. The robustness of SPRT with respect to ASN function can be studied by replacing the denominator of (7) by

$$E_{\lambda*}(Z) = \int_0^\infty zf(x;\theta,\lambda*)dx = \lambda \ln\left(\frac{\theta_0}{\theta_1}\right) + \left(\frac{1}{\theta_0} + \frac{1}{\theta_1}\right)\theta\lambda* \quad (12)$$

We consider the cases $\lambda > \lambda*$ and $\lambda < \lambda*$ to study the robustness of the SPRT. In **Table 2**, we present the ASN function for the cases $\lambda > \lambda*$ for the SPRT of testing the hypothesis H$_0$: $\theta = 45$ against H$_1$: $\theta = 50$. The values of OC function for the cases $\lambda > \lambda*$ and $\lambda < \lambda*$ respectively are plotted **Figure 1** and **Figure 2**.

3.2. Robustness of SPRT for Exponential Distribution

The random variable X is said to follow exponential distribution with location parameter θ and scale parameter λ

P. V. PANDIT *ET AL.*

Table 1. ASN values of SPRT for scale parameter of gamma distribution.

$\lambda = 5, \alpha = 0.01, \beta = 0.01$				$\lambda = 5, \alpha = 0.05, \beta = 0.05$			
θ_0/θ_1	$E_0(N)$	$E_1(N)$	$E_\theta(N)$	θ_0/θ_1	$E_0(N)$	$E_1(N)$	$E_\theta(N)$
0.1	0.64	0.13	0.80	0.1	0.38	0.08	0.33
0.2	1.11	0.38	1.63	0.2	0.65	0.22	0.67
0.3	1.79	0.80	2.91	0.3	1.05	0.47	1.20
0.4	2.85	1.54	5.03	0.4	1.68	0.91	2.07
0.5	4.66	2.94	8.79	0.5	2.74	1.73	3.61
0.6	8.13	5.78	16.18	0.6	4.78	3.40	16.64
0.7	15.89	12.53	33.20	0.7	19.35	17.37	13.63
0.8	38.92	33.54	84.81	0.8	22.90	19.73	34.82
0.9	168.02	156.62	380.42	0.9	198.87	192.16	156.20
$\lambda = 5, \alpha = 0.01, \beta = 0.01$				$\lambda = 5, \alpha = 0.05, \beta = 0.05$			
θ_0/θ_1	$E_0(N)$	$E_1(N)$	$E_\theta(N)$	θ_0/θ_1	$E_0(N)$	$E_1(N)$	$E_\theta(N)$
0.1	0.41	0.12	0.51	0.1	0.60	0.09	0.51
0.2	0.72	0.35	1.05	0.2	1.03	0.24	1.05
0.3	1.15	0.74	1.88	0.3	1.66	0.52	1.88
0.4	1.84	1.43	3.24	0.4	2.64	1.00	3.24
0.5	3.01	2.72	5.66	0.5	4.33	1.90	5.66
0.6	5.25	5.36	10.42	0.6	7.54	3.73	10.42
0.7	10.27	11.62	21.38	0.7	14.74	18.10	21.38
0.8	25.15	31.11	54.61	0.8	36.10	21.67	54.61
0.9	108.58	145.27	244.96	0.9	155.84	101.22	244.96

Table 2. ASN function for $\lambda < \lambda^*$ (H$_0$: $\theta = 45$, H$_1$: $\theta = 50$, $\alpha = \beta = 0.05$, $\lambda = 5$.

θ/λ^*	5.0	5.05	5.10	5.15	5.20	5.25
41.00	41.30	44.11	47.32	51.03	55.33	60.39
42.00	48.87	52.92	57.66	63.26	69.93	77.94
43.00	59.62	65.72	73.02	81.81	92.38	104.97
44.00	75.50	84.94	96.24	109.58	124.72	140.68
45.00	99.02	112.80	128.09	143.50	156.49	163.82
46.00	129.74	144.44	155.87	160.87	157.93	147.84
47.00	154.07	157.58	153.15	142.24	127.85	112.80
48.00	149.13	137.97	123.70	109.01	95.54	83.89
49.00	120.96	106.59	93.41	82.03	72.48	64.57
50.00	92.35	81.06	71.59	63.74	57.25	51.83
51.00	71.40	63.49	56.95	51.51	46.96	43.11
52.00	57.12	51.57	46.94	43.03	39.70	36.85
53.00	47.25	43.23	39.82	36.89	34.36	32.16
54.00	40.18	37.15	34.54	32.27	30.28	28.52
55.00	34.91	32.55	30.49	28.68	27.06	25.62

if its probability density function is given by

$$f(x; \theta, \lambda) = \lambda e^{-\lambda(x-\theta)}, \ x > 0, \ \lambda > 0 \qquad (13)$$

where it is assumed that θ is known.

Given a sequence of observations $X_1, X_2, \ldots$, from (13), we wish to test the null hypothesis H$_0$: $\lambda = \lambda_0$ against the alternative H$_1$: $\lambda = \lambda_1$ ($> \lambda_0$). We propose the following SPRT.

$$\text{Let } Z = \text{In}\left\{\frac{f(x; \lambda_1)}{f(x; \lambda_0)}\right\} = \text{In}\left(\frac{\lambda_1}{\lambda_0}\right) - (\lambda_1 - \lambda_0)(x - \theta) \quad (14)$$

We choose two numbers A and B such that $0 < B < 1 < A$. At n^{th} stage, accept H$_0$ if $\sum_{i=1}^{n} z_i \leq \text{In } B$, reject H$_0$ if $\sum_{i=1}^{n} Z_i \geq \ln A$, otherwise continue sampling by taking $(n + 1)^{\text{th}}$ observation. Here Z_i is obtained by replacing x by x_i in (3.2). If α and β are the probabilities of type I and type II errors respectively, then $A \cong \frac{1-\beta}{\alpha}, B \cong \frac{\beta}{1-\alpha}$ The OC function $L(\lambda)$ is given by
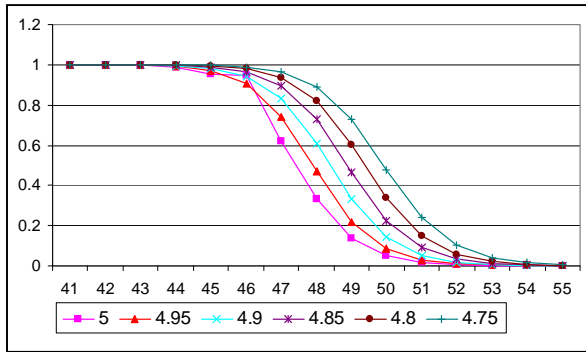
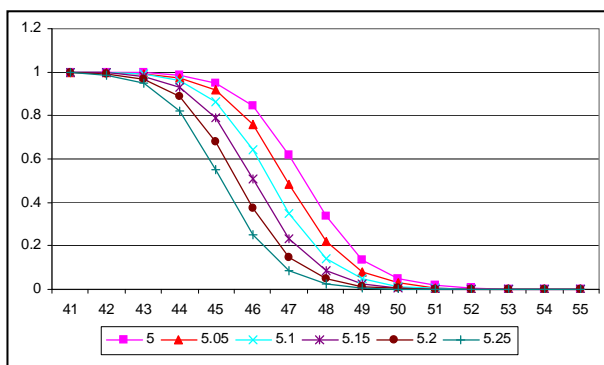**Figure 1. Graph of OC function for gamma distribution $\lambda < \lambda^*$.**

Figure 2. Graph of OC function for gamma distribution $\lambda > \lambda^*$.

Figure 3. Graph of OC function for exponential distribution $\theta < \theta^*$.
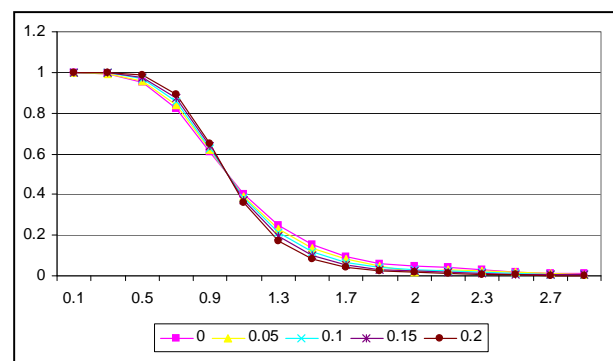
Figure 4. Graph of OC function for exponential distribution $\theta > \theta^*$.

$$L(\lambda) \approx \frac{A^{h(\lambda)} - 1}{A^{h(\lambda)} - B^{h(\lambda)}}, \qquad (15)$$

Where $h(\lambda)$ is the solution of $E\left[e^{h(\lambda)Z}\right] = 1$.

Here, $E\left[e^{h(\lambda)Z}\right] = \lambda\left(\dfrac{\lambda_1}{\lambda_0}\right)^{h(\lambda)}$ and hence we obtain

$$\lambda = \left(\frac{\lambda_0}{\lambda_1}\right)^{h(\lambda)}.$$

Let is suppose that the location parameter θ has undergone a change and the pdf (3.1) becomes $f(x; \theta^*; \lambda)$. To study the robustness of the SPRT developed here with respect to OC function, we consider $h(\lambda)$ as the solution of the equation

$$\int_0^\infty \left[\frac{f(x; \theta, \lambda_1)}{f(x; \theta, \lambda_0)}\right]^{h(\lambda)} f(x; \theta^*; \lambda)\,dx = 1$$

giving

$$h(\lambda) = \frac{-\ln \lambda}{(\lambda_1 - \lambda_0)(\theta^* - \theta) + \ln\left(\dfrac{\lambda_1}{\lambda_0}\right)} \qquad (16)$$

For different values of λ, $h(\lambda)$ is evaluated and OC function is obtained with the help of Equation (15).

The values of OC function for the cases $\lambda < \lambda^*$ and $\lambda > \lambda^*$ are plotted in **Figures 3** & **4** respectively.

4. Conclusions

Some Remarks

1) The values of OC function for the cases $\lambda > \lambda^*$ and $\lambda < \lambda^*$ respectively are plotted **Figure 1** and **Figure 2**. The figures indicate that for $\lambda > \lambda^*$ ($\lambda < \lambda^*$), the OC curve shifts to the right side (left side) of the curve for when $\lambda = \lambda^*$. From the figure it is clear that SPRT is robust for $\lambda^* = \lambda \pm 0.05$ as the deviation in OC function is insignificant. However, for other values of λ^* the test becomes highly non-robust.

2) We have considered the SPRT for testing the hypothesis H$_0$: $\lambda = 1$ versus H$_1$: $\lambda = 2$ for $\alpha = \beta = 0.05$. The values of OC function for the cases $\lambda < \lambda^*$ and $\lambda > \lambda^*$ are plotted in **Figures 3** & **4** respectively. It is exhibited from the table that the SPRT is non-robust even for $\lambda^* = \lambda \pm 0.05$. The ASN function for the case $\lambda < \lambda^*$ is tabulated in **Table 2**.

5. References

[1] R. E. Barlow and E. Proschan, "Exponential Life Test Procedures When the Distribution under Test has Monotone Failure Rate," *Journal of American Statistic Association*, Vol. 62, No. 318, 1967, pp. 548-560.

[2] L. Harter and A. H. Moore, "An Evaluation of Exponential and Weibull Test Plans," *IEEE Transactions on Reliability*, Vol. 25, No. 2, 1976, pp. 100-104.

[3] E. R. Montagne and N. D. Singpurwalla, "Robustness of Sequential Exponential Life-Testing Procedures," *Journal of American Statistic Association*, Vol. 80, No. 391, 1985, pp. 715-719.

[4] A. Wald, "Sequential Analysis," John Wiley, New York, 1947.

AM

Scientific
Research

Existence Solution for 5th Order Differential Equations under Some Conditions

Sayada Nabhan Odda
Department of Mathematics, Faculty of Computer Science,
Qassim University, Burieda, Saudi Arabia
E-mail: *nabhan*100@*yahoo.com*

Abstract

We study a nonlinear differential equations in the Banach space of real functions and continuous on a bounded and closed interval. With the help of a suitable theorems (fixed point) and some boundary conditions, the 5th order nonlinear differential equations has at least one positive solution.

Keywords: Fixed-Point Theorem, Banach Space, Nonlinear 5th Order Differential Equations

1. Introduction

In this paper, we are going to study the solvability nonlinear 5th order differential equations. We will look for solutions of that equation in the Banach space. The main tool used in our investigations to existence of positive solutions for the nonlinear 5th order boundary value problems. Let us mention that the theory of nonlinear differential equations has many useful applications in describing numerous events and problems of the real world. For example, nonlinear differential equations are often applicable in engineering, mathematical physics, economics and biology [1,2]. The nonlinear 5th order differential equations studied in this paper is an existence and nonexistence of positive solutions by using object of mathematical investigations [3-5]. The results presented in this paper seem to be new and original. They generalize several results obtained up to now in the study of nonlinear differential equations of several types.

2. Notation, Definition and Auxiliary Results

Theorem 2.1 [6,7]

Assume that U is a relatively open subset of convex set K in Banach space E. Let $N : \bar{U} \to K$ be a compact map with $o \in U$. Then either

1) N has a fixed point in $\bar{U}$; or
2) There $u \in U$ and $\lambda \in (0,1)$ such that $u = \lambda N u$.

Definition 2.1 An operator is called completely con-

tinuous if it is continuous and maps bounded sets into pre-compacts.

Definition 2.2 Let E be a real Banach space. A non-empty closed convex set $K \subset E$ is called cone of E if it satisfies the following conditions:

1) $x \in K$, $\sigma \geq o$ *implies* $\sigma x \in K$;
2) $x \in K$, $-x \in K$ *implies* $x = o$.

3. Main Result

In this section, we will study the existence and nonexistence of positive solutions for the nonlinear boundary value problem:

$$u^{(5)}(t) = f\left(t, u(t)\right), \ 0 < t < 1 , \qquad (3.1)$$

$$u(0) = u'(0) = u'''(0) = u^{(4)}(1) = 0, \qquad (3.2)$$

$$\alpha u'(1) + \beta u''(1) = 0 , \\ \text{where } \beta, \alpha \geq 0, \quad \alpha + \beta > 0 \qquad (3.3)$$

Theorem 3.1. Under conditions (3.2) and (3.3), Equation (3.1) has a unique solution.

Proof: Applying the Laplace transform to Equation (3.1) we get,

$$s^5 \bar{u}(s) - s^4 u(0) - s^3 u'(0) - s^2 u''(0) \\ - s u'''(0) - u^4(0) = \bar{y}(s) \qquad (3.4)$$

Where $\bar{u}(s)$ and $\bar{y}(s)$ is the Laplace transform of $u(t)$ and $y(t)$ respectively. The Laplace inversion of Equation (3.4) gives the final solution as:

$$u(t) = \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \int_0^1 \frac{t^2}{2!} f(s, u(s)) ds - \frac{\alpha}{[\alpha + \beta]3!} \int_0^1 \frac{t^2(1-s)^3}{2!} f(s, u(s)) ds$$

$$- \frac{\beta}{[\alpha + \beta]2!} \int_0^1 \frac{t^2(1-s)^2}{2!} f(s, u(s)) ds - \int_0^1 \frac{t^4}{4!} f(s, u(s)) ds + \int_0^t \frac{(t-s)^4}{4!} f(s, u(s)) ds \tag{3.5}$$

The proof is complete.

Defining $T : X \to X$ as:

$$Tu(t) = \frac{\alpha + (3)\beta}{[\alpha + \beta](3)!} \int_0^1 \frac{t^2}{2!} f(s, u(s)) ds - \frac{\alpha}{[\alpha + \beta](3)!} \int_0^1 \frac{t^2(1-s)^3}{2!} f(s, u(s)) ds$$

$$- \frac{\beta}{[\alpha + \beta](2)!} \int_0^1 \frac{t^2(1-s)^2}{2!} f(s, u(s)) ds - \int_0^1 \frac{t^4}{(4)!} f(s, u(s)) ds + \int_0^t \frac{(t-s)^4}{(4)!} f(s, u(s)) ds \tag{3.6}$$

Where X = C[0,1] is the Banach space endowed with the supper norm. We have the following result for operator T.

Lemma 3.1

Assume that $f : [0,1] \times R \to R$ is continuous function, then T is completely continuous operator.

Proof: It is easy to see that T is continuous. For $u \in M = \{u \in X; \|u\| \le l, \ l > 0\}$, we obtain,

$$Tu(t) = \left| \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \int_0^1 \frac{t^2}{2!} f(s, u(s)) ds - \frac{\alpha}{[\alpha + \beta]3!} \int_0^1 \frac{t^2(1-s)^3}{2!} f(s, u(s)) ds \right.$$

$$\left. - \frac{\beta}{[\alpha + \beta]2!} \int_0^1 \frac{t^2(1-s)^2}{2!} f(s, u(s)) ds - \int_0^1 \frac{t^4}{4!} f(s, u(s)) ds + \int_0^t \frac{(t-s)^4}{4!} f(s, u(s)) ds \right|$$

$$\le \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \int_0^1 \frac{t^2}{2!} |f(s, u(s))| ds + \frac{\alpha}{[\alpha + \beta]3!} \int_0^1 \frac{t^2(1-s)^3}{2!} |f(s, u(s))| ds$$

$$+ \frac{\beta}{[\alpha + \beta]2!} \int_0^1 \frac{t^2(1-s)^2}{2!} |f(s, u(s))| ds + \int_0^1 \frac{t^4}{4!} |f(s, u(s))| ds + \int_0^t \frac{(t-s)^4}{4!} |f(s, u(s))| ds$$

$$\le \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \frac{L}{2!} + \frac{\alpha}{[\alpha + \beta]4!} \frac{L}{2!} + \frac{\beta}{[\alpha + \beta]3!} \frac{L}{2!} + \frac{L}{4!} + \frac{L}{5!}$$

where $L = \max\limits_{0 \le t \le 1, \|u\| \le 1} |f(t, u(t))| + 1$,

so T(M) is bounded. Next we shall show the equicontinuity of $\overline{T(M)}$. $\forall u \in M, \forall \varepsilon > 0, t_1 < t_2 \in [0,1]$.

Let

$$\eta < Min \left\{ \begin{array}{l} 2! \dfrac{\varepsilon[\alpha + \beta]3!}{5L[\alpha + 3\beta]}, 2! \dfrac{\varepsilon[\alpha + \beta]4!}{5L\alpha}, \\ 2! \dfrac{\varepsilon[\alpha + \beta]3!}{5L\beta} \end{array} \right\}, \ \sigma < \left\{ \dfrac{\varepsilon(4)!}{5L} \right\} \ , \gamma < \left\{ \varepsilon \dfrac{5!}{5L} \right\}, \ t^2_2 - t^2_1 < \eta, \ t^4_2 - t^4_1 < \sigma, \ t^5_2 + t^5_1 < \gamma$$

$$|Tu(t_2) - Tu(t_1)| = \left| \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \int_0^1 \frac{t_2^2 - t_1^2}{2!} f(s, u(s)) ds - \frac{\alpha}{[\alpha + \beta]3!} \int_0^1 \frac{(t_2^2 - t_1^2)(1-s)^3}{2!} f(s, u(s)) ds \right.$$

$$\left. - \frac{\beta}{[\alpha + \beta]2!} \int_0^1 \frac{(t_2^2 - t_1^2)(1-s)^2}{2!} f(s, u(s)) ds - \int_0^1 \frac{t_2^4 - t_1^4}{4!} f(s, u(s)) ds \right.$$

$$\left. + \int_0^{t_2} \frac{(t_2 - s)^4}{4!} f(s, u(s)) ds - \int_0^{t_1} \frac{(t_1 - s)^4}{4!} f(s, u(s)) ds \right|$$

$$\le \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \frac{L(t_2^2 - t_1^2)}{2!} + \frac{\alpha}{[\alpha + \beta]4!} \frac{L(t_2^2 - t_1^2)}{2!} + \frac{\beta}{[\alpha + \beta](3)!} \frac{L(t_2^2 - t_1^2)}{2!} + \frac{L(t_2^4 - t_1^4)}{4!} + \frac{Lt_1^5}{5!} + \frac{Lt_1^5}{5!}$$

$$\le \frac{\alpha + 3\beta}{[\alpha + \beta]3!} \frac{L\eta}{2!} + \frac{\alpha}{[\alpha + \beta]4!} \frac{L\eta}{2!} + \frac{\beta}{[\alpha + \beta]3!} \frac{L\eta}{2!} + \frac{L\sigma}{4!} + \frac{L\gamma}{5!} \le \frac{\varepsilon}{5} + \frac{\varepsilon}{5} + \frac{\varepsilon}{5} + \frac{\varepsilon}{5} + \frac{\varepsilon}{5}.$$

Thus $\overline{T(M)}$ is equi-continuous. The Arzela-Ascoli theorem implies that the operator T is completely continuous.

Theorem 3.2

Assume that $f : [0,1] \times R \to R$ is continuous function, and there exist constants

$$0 < c_1 < \left(5\frac{2!(\alpha+\beta)4!}{n\alpha+(4)^2\beta}, \frac{5!}{6} \right), c_2 > 0, \text{ such that}$$

$|f(t,u(t))| \le c_1 |u| + c_2$ for all $t \in [0,1]$. Then the boundary value problem (3.1),(3.2)and(3.3) has a solution.

Proof: Following [8,9], we will apply the nonlinear alternative theorem to prove that T has one fixed point.

Let $\Omega = \{u \in X; \|u\| < R\}$, be open subset of *X*, where

$$R > \left\{ 4 \left\{ \begin{array}{c} \left[\frac{5\alpha+(4)^2\beta}{2!(4)!(\alpha+\beta)}c_1, \frac{6c_1}{5!}, \right] \\ \left[\frac{5\alpha+(4)^2\beta}{2!(4)!(\alpha+\beta)}c_2, \frac{6c_2}{5!} \right] \end{array} \right\} \right\}.$$

We suppose that there is a point $u \in \partial\Omega$ and $c_1 \in (0,1)$ such that $u = Tu$. So, for $u \in \partial\Omega$, we have:

$$Tu(t) = \left| \begin{array}{c} \frac{\alpha+3\beta}{[\alpha+\beta]3!}\int_0^1 \frac{t^2}{2!}f(s,u(s))ds - \frac{\alpha}{[\alpha+\beta]3!}\int_0^1 \frac{t^2(1-s)^3}{2!}f(s,u(s))ds \\ -\frac{\beta}{[\alpha+\beta]2!}\int_0^1 \frac{t^2(1-s)^2}{2!}f(s,u(s))ds - \int_0^1 \frac{t^4}{4!}f(s,u(s))ds + \int_0^t \frac{(t-s)^4}{4!}f(s,u(s))ds \end{array} \right|$$

$$\le \frac{\alpha+3\beta}{[\alpha+\beta]3!}\int_0^1 \frac{t^2}{2!}|f(s,u(s))|ds + \frac{\alpha}{[\alpha+\beta]3!}\int_0^1 \frac{t^2(1-s)^3}{2!}|f(s,u(s))|ds$$

$$+ \frac{\beta}{[\alpha+\beta]2!}\int_0^1 \frac{t^2(1-s)^2}{2!}|f(s,u(s))|ds + \int_0^1 \frac{t^4}{4!}|f(s,u(s))|ds + \int_0^t \frac{(t-s)^4}{4!}|f(s,u(s))|ds$$

So,

$$Tu(t) = \left| \begin{array}{c} \frac{\alpha+3\beta}{[\alpha+\beta]3!}\int_0^1 \frac{t^2}{2!}f(s,u(s))ds - \frac{\alpha}{[\alpha+\beta]3!}\int_0^1 \frac{t^2(1-s)^3}{2!}f(s,u(s))ds - \\ \frac{\beta}{[\alpha+\beta]2!}\int_0^1 \frac{t^2(1-s)^2}{2!}f(s,u(s))ds - \int_0^1 \frac{t^4}{4!}f(s,u(s))ds + \int_0^t \frac{(t-s)^4}{4!}f(s,u(s))ds \end{array} \right|$$

$$\le \frac{\alpha+3\beta}{[\alpha+\beta]3!}\int_0^1 \frac{t^2}{2!}(c_1|u(s)|+c_2)ds + \frac{\alpha}{[\alpha+\beta]3!}\int_0^1 \frac{t^2(1-s)^3}{2!}(c_1|u(s)|+c_2)ds$$

$$+ \frac{\beta}{[\alpha+\beta]2!}\int_0^1 \frac{t^2(1-s)^2}{2!}(c_1|u(s)|+c_2)ds + \int_0^1 \frac{t^4}{4!}(c_1|u(s)|+c_2)ds + \int_0^t \frac{(t-s)^4}{4!}(c_1|u(s)|+c_2)ds$$

$$\le \frac{\alpha+3\beta}{2![\alpha+\beta]3!}(c_1|u(s)|) + \frac{\alpha}{2![\alpha+\beta]4!}(c_1|u(s)|) + \frac{\beta}{2![\alpha+\beta]3!}(c_1|u(s)|) + \frac{1}{4!}(c_1|u(s)|) + \frac{1}{5!}(c_1|u(s)|)$$

$$\frac{\alpha+3\beta}{2![\alpha+\beta]3!}(c_2) + \frac{\alpha}{2![\alpha+\beta]4!}(c_2) + \frac{\beta}{2![\alpha+\beta]3!}(c_2) + \frac{1}{4!}(c_2) + \frac{1}{5!}(c_2) < \frac{R}{4} + \frac{R}{4} + \frac{R}{4} + \frac{R}{4} = R,$$

which implies that $\|T\| \ne R = \|u\|$, that is a contraction. Then the nonlinear alternative theorem implies that T has a fixed point $u \in \overline{\Omega}$, that is , problem(3.1),(3.2) and (3.3) has a solution $u \in \overline{\Omega}$.

Finally, we give an example to illustrate the results obtained in this paper.

Example: By using the Equation (3.1) with the conditions (3.2) and (3.3) to solve the boundary value problem

$$u^5(t) = \frac{u+1}{u^2+7} \tag{3.7}$$

and applying the theorem 3.2 with $\alpha = 1$ and $\beta = 1$,

we found $c_1 < \min\left(\frac{2!(\alpha+\beta)4!}{5\alpha+16\beta}, \frac{5!}{6} \right)$. So, we conclude that the problem (3.7) has a solution.

4. References

[1] R. P. Agarwal and D. Oregan, "Global Existence for Nonlinear Operator Inclusion," *Computers & Mathematics with Applications*, Vol. 38, No. 11-12, 1999, pp. 131 -139.

[2] R. P. Agarwal, D. Oregan and P. J. Y. Wong, "Positive

Solutions of Differential, Difference and Integral Equations," Kluwer Academic, Dordrecht, 1999.

[3] M. El-Shahed, "Positive Solutions for Nonlinear Singular Third Order Boundary Value Problem," *Communications in Nonlinear Science and Numerical Simulation*, Vol. 14, No. 2, 2009, pp. 424-429.

[4] D. Guo and V. Lakshmikantham, "Nonlinear Problems in Abstract Cones," Academic Press, San Diego, 1988.

[5] S. Li, "Positive Solutions of Nonlinear Singular Third-Order Two-Point Boundary Value Problem," *Journal of Mathematical Analysis and Applications*, Vol. 323, No. 1, 2006, pp. 413-425.

[6] H. Sun and W. Wen, "On the Number of Positive Solutions for a Nonlinear Third Order Boundary Value Prob-

lem," *International Journal of Difference Equations*, Vol. 1, No. 1, 2006, pp. 165-176.

[7] R. P. Agarwal, M. Meehan and D. Oregan, "Fixed Point and Applications," Cambridge University Press, Cambridge, 2001.

[8] B. Yang, "Positive Solutions for a Fourth Order Boundary Value Problem," *Electronic Journal of Qualitative Theory of differential Equations*, Vol. 3, No. 1, 2005, pp. 1-17.

[9] S. Q. Zhang, "Existence of Solution for a Boundary Value Problem of Fractional Order," *Acta Mathematica Scientica*, Vol. 26, No. 2, 2006, pp. 220-228.

Scientific
Research

Analyticity of Semigroups Generated by Singular Differential Matrix Operators

Ould Ahmed Mahmoud Sid Ahmed[1], Adel Saddi[2]

[1]*Department of Mathematics, College of Science, Aljouf University, Aljouf, Saudi Arabia*
[2]*Department of Mathematics, College of Education for Girls in Sarat Ebeidah, King Khalid University, Abha, Saudi Arabia*
E-mail: *sidahmed@ju.edu.sa, adel.saddi@fsg.rnu.tn*

Abstract

In this paper we prove the analyticity of the semigroups generated by some singular differential matrix operators of the form $\mathcal{A} = \dfrac{d^2}{dx^2} + \dfrac{B(x)}{x}\dfrac{d}{dx},$ in the Banach space $C([0, +\infty], M_d(C)),$ with suitable boundary conditions. To illustrate the work an example is discussed.

Keywords: Dissipative Operators, Positive Operators, Spectrum, Analytic Semigroups, Evolution Equations

1. Introduction

As we will see in the sequel the problem of characterizing operator matrices generating strongly continuous semigroups or analytic semigroups is quite difficult. The main problem consists in finding appropriate assumptions on the matrix entries allowing general results but still including the concrete examples we have in mind.

The evolution of a physical system in time is usually described in a Banach space by an initial value problem of the form

$$\begin{cases} \dfrac{dU(t)}{dt} - \mathcal{A}U(t) = 0, \quad t \geq 0 \\ U(0) = U_0. \end{cases} \quad (1.1)$$

Problem of type (1.1) is well posed in a Banach space X if and only if the operator $(\mathcal{A}, D(\mathcal{A}))$ generates a C_0-semigroup $(T_t)_{t>0}$ on X. Here the solution $U(t)$ is given by $U(t) = T_t U_0$ for the initial data $U_0 \in D(\mathcal{A})$. For operator semigroups we refer to [1-5] and to [6] for the theory of operator matrices. The harmonic analysis for a class of differential operators with matrix coefficients was treated in [7,8]. In this work we are interested in a generalization of the analyticity and the positivity of the semigroup generated by a matrix singular differential operator $(\mathcal{A}, D(\mathcal{A}))$. A similar study was realized in [9] for a class of differential operators with matrix coefficients and interface. For the scalar case we refer to [10].

This paper is organized as follows. In the second section we introduce some notations and give preliminaries results. In the third section we investigate some properties of the operator $(\mathcal{A}, D(\mathcal{A}))$ in particular we prove that it is closed, densely defined, dissipative and satisfies the positive minimum principle. Some spectral properties of the operator $(\mathcal{A}, D(\mathcal{A}))$ are obtained in this section. In section fourth we established the analyticity of the semigroup generated by an operator in the form $(\mathcal{A}, D(\mathcal{A}))$. In this case the problem (1.1) has a unique solution for all $U_0 \in X$. In section five we present a concrete example of application of the results obtained.

2. Notations and Preliminary Results

Let $M_d(C)$ be the space of $d \times d$ matrices with complex coefficients and I_d the identity matrix. The norm of a matrix $A = \left(a_{ij}\right)_{1 \leq i, j \leq d} \in M_d(C)$ will be defined by

$$\| A \| = \max_{1 \leq i, j \leq d} \left(| a_{ij} |\right) \quad (2.1)$$

The reason of this special choice will be justified in Lemma 2.1 and Lemma 2.2.

A matrix $A = \left(a_{ij}\right)_{1 \leq i, j \leq d} \in M_d(C)$ is called nonnegative (resp positive) if all entries a_{ij} are real numbers and nonnegative (resp positive). In this case we write $A \geq 0$, $(resp. A > 0)$ and we call $| A | = \left(| a_{ij} |\right)_{1 \leq i, j \leq d}$

the absolute value of A.

For $A = \left(a_{ij}\right)_{1\leq i, j\leq d} \in M_d(C)$ and $A = \left(a_{ij}\right)_{1\leq i, j\leq d} \in M_d(C)$, we will write

$$A \geq B \quad if \quad A - B \geq 0.$$

and

$$A > B \quad if \quad A - B > 0.$$

A matrix-valued functions is said to be continuous, differentiable or integrable if all its elements are continuous, differentiable or integrable functions. If the matrix $A(x)$ is integrable over $[a, b]$, then

$$\left\| \int_a^b A(x)dx \right\| \leq \int_a^b \|A(x)\| dx.$$

For two matrix functions $A(x) = \left(a_{ij}(x)\right)_{1\leq i, j\leq d}$ and $B(x) = \left(b_{ij}(x)\right)_{1\leq i, j\leq d}$, we shall write the negligibility $A(x) = o(B(x))$, $x \to x_0$ if, $a_{ij}(x) = o\left(b_{ij}(x)\right)$ $x \to x_0$ for all $1 \leq i, j \leq d$. Similarly we define the notions of domination $A(x) = O(B(x))$ and the equivalence $A(x) \sim B(x)$ (see for more details [11]).

Let X be the space $C([0, +\infty], M_d(C))$ of continuous, matrix-valued functions on $[0, +\infty]$. On the space X we define the norm $\|.\|_X$, by $\| f \|_X = \sup_{x\geq 0} \| f(x) \|$, for all $f \in X$.

Note that the normed space $(X, \|.\|_X)$ is a Banach space.

We denote by $C_\infty^k(]0, +\infty], M_d(C)), k = 1, 2, \cdots$ the space of all $k -$ times continuously differentiable matrix-valued functions U defined on $]0, +\infty[$ such that

$\lim_{x\to+\infty} U^{(p)}(x)$ exists and finite for all $0 \leq p \leq k$.

Consider a singular second order differential operator $(\mathcal{A}, D(\mathcal{A}))$ with matrix coefficients defined on X by

$$\mathcal{A}U = \frac{d^2}{dx^2}U + \frac{B(x)}{x}\frac{d}{dx}U,$$

where U and B are matrix-valued functions, with the domain

$$D(\mathcal{A}) = \{U \in C([0, +\infty], M_d(C)) \cap C_\infty^2(]0, +\infty],$$
$$M_d(C)), \lim_{x\to 0}\mathcal{A}U(x) = 0\}.$$

We assume that B is a real valued continuous and bounded matrix-valued function on $[0, +\infty[$ and if $B(0) = -I_d$ we add the assumption

$$R x^D \geq | B(x) - B(0) | \qquad (2.2)$$

in a neighborhood of 0, for nonnegative constants diagonal matrices R and D. Here x^D is the diagonal matrix with diagonal entries $\alpha_i = x^{d_i}$, $i = 1, 2\cdots, d$, where $d_i, i = 1, 2, \cdots, d$ are the diagonal entries of D.

We will now recall some results needed in the sequel.

Lemma 2.1 *Let* $A \in M_d(C)$. *The following properties hold*

1) $A \geq 0$ if and only if $AB \geq 0$ for all $B \geq 0$.

2) $| AB | \leq | A \| B |$, hence $| AB | \leq | A \| B |$ if $A \geq 0$.

Lemma 2.2 *Let* $A, B \in M_d(C)$. *The following properties hold*

1) $| A | \leq | B |$ implies $\|A\| \leq \|B\|$.

2) $\| A \| = \|A\|$.

3) $\| AE \| \leq d \| A \|$, where

$$E = \left(e_{ij}\right)_{1\leq i, j\leq d}, e_{ij} = 1, 1 \leq i, j \leq d.$$

Proposition 2.1 *Let* A *be a nonnegative matrix with spectral radius* $r(A)$.

1) The resolvent $R(\mu, A)$ is positive whenever $\mu > r(A)$.

2) If $| \mu | > r(A)$, then $| R(\mu, A) | \leq R(| \mu |, A)$.

Proof.

1) We use the Neumann series representation for the resolvent

$$R(\mu, A) = (\mu - A)^{-1} = \mu^{-1}(I_d - \mu^{-1}A)^{-1}$$
$$= \sum_{n\geq 0} \frac{A^n}{\mu^{n+1}} \quad for \quad | \mu | > r(A).$$

If $A \geq 0$ then $A^n \geq 0$, for all n, hence for $\mu > r(A)$, we have

$$R(\mu, A) = \lim_{n\to+\infty}\sum_{k=0}^n \frac{A^k}{\mu^{k+1}} \geq 0$$

since the finite sums are positive and convergence holds in every entry.

2) For $| \mu | > r(A)$, we have

$$| R(\mu, A) | = \left| \lim_{n\to+\infty}\sum_{k=0}^n \frac{A^k}{\mu^{k+1}} \right|$$
$$\leq \lim_{n\to+\infty}\sum_{k=0}^n \frac{A^k}{| \mu |^{k+1}} = R(| \mu |, A).$$

Theorem 2.1 ([12], *Perrons Theorem*)

If A is a nonnegative matrix, then $r(A)$ is an eigenvalue of A with positive eigenvector.

Definition 2.1

1) An operator $(S, D(S))$ on a Banach Lattice $(X, \geq)$ is called positive if

$$Su \geq 0 \quad for \ all \ u \in D(S)_+ = \{v \in D(S), v \geq 0\}.$$

2) A semigroup $(T_t)_{t>0}$ on X is positive if and only if T_t is positive for all $t > 0$.

Remark 2.1 *An operator* $(S, D(S))$ *defined on a* $C(K) -$ *space* (K *compact*) *satisfies the positive minimum principle if for every* $u \in D(S), u$ *positive and* $x \in K, u(x) = 0$, *then* $(Su)(x) \geq 0$.

Theorem 2.2 ([1])

Let $(S, D(S))$ be the generator of a semigroup $(T_t)_{t>0}$ on $C(K)$, then the semigroup is positive if and only if $(S, D(S))$ satisfies the positive minimum principle.

3. The Diagonal Case

In this section we assume that the matrix function B is diagonal.

3.1. Characterization of the Operator $(\mathcal{A}, D(\mathcal{A}))$

The proofs of Proposition 3.1 and Proposition 3.2 can be deduced from the scalar case given in Proposition 2.2 and Proposition 2.3 from [10].

Proposition 3.1 *Let* $-I_d > B(0)$ *and* $U \in X \cap C^2_\infty(]0,+\infty], M_d(C))$. *Then* $U \in D(\mathcal{A})$ *if and only if* $U \in \mathcal{C}^2([0,\tau], M_d(C))$ *for some* $\tau > 0$ *and* $U'(0) = U''(0) = 0$.

Proposition 3.2 *Let* $B(0) = -I_d$ *and* $U \in X \cap C^2_\infty(]0,+\infty], M_d(C))$. *Then* $U \in D(\mathcal{A})$ *if and only if* $U \in C^1([0,\tau]), M_d(C))$ *for some* $\tau > 0$, $U'(0) = 0$ *and*

$$\lim_{x \to 0} \frac{1}{x} \int_0^x (U''(x) - U''(t)) dt = 0.$$

Moreover, if $U \in D(\mathcal{A})$ then $U''(x) = o(\log xE)$ and $\frac{1}{x} U'(x) = o(logxE)$ as $x \to 0$, where E is the constant matrix introduced in Lemma 2.2.

Proposition 3.3 *The operator* $(\mathcal{A}, D(\mathcal{A}))$ *is a densely defined, closed, dissipative and satisfies the positive minimum principle.*

Proof. Put $B(x) = \begin{pmatrix} b_1(x) & 0 & \cdots & 0 \\ 0 & b_2(x) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_d(x) \end{pmatrix}$

and for $i = 1, 2 \cdots, d$, let $\mathcal{A}_i u = u'' + \dfrac{b_i(x)}{x} u'$

with

$D(\mathcal{A}_i) =$
$\{u \in C([0,+\infty], C) \cap \mathcal{C}^2_\infty(]0,+\infty], C), \lim_{x\to 0} \mathcal{A}_i u(x) = 0\}.$

Then $U = \left(u_{ij}\right)_{1 \le i, j \le d} \in D(\mathcal{A})$ if and only if

$u_{ij} \in D(\mathcal{A}_i)$ for all $i, j = 1, 2 \cdots, d$. Hence from ([10]. Lemma 2.4), the operator $(\mathcal{A}, D(\mathcal{A}))$ is a densely defined and closed.

Let us show that $(\mathcal{A}, D(\mathcal{A}))$ is dissipative:

$\lambda \| U \|_X \le \| \lambda U - \mathcal{A}U \|_X$, for $\lambda > 0$ and $U \in D(\mathcal{A})$

Let $U = \left(u_{ij}\right)_{1 \le i, j \le d} \in D(\mathcal{A})$ and $\lambda > 0$. According to [10], for all $1 \le i, j \le d$ we have

$$\lambda \| u_{ij} \|_\infty \le \| \left(\lambda u_{ij} - \mathcal{A}_i u_{ij}\right) \|_\infty$$

then

$$\lambda \| U \|_X \le \| \left(\lambda U - \mathcal{A}U\right) \|_X,$$

and hence the dissipativity holds.

In order to prove that $(\mathcal{A}, D(\mathcal{A}))$ satisfies the positive minimum principle, assume that $U = \left(u_{ij}\right)_{1 \le i, j \le d} \in D(\mathcal{A})$ and positive such that $U(x_0) = 0$ for some $x_0 \in [0,\infty[$. If $x_0 > 0$ then $\left(u_{ij}\right) \ge 0$ and $u_{ij}(x_0) = 0$, for all $i, j = 1, 2, \cdots, d$. Hence $u_{ij}'(x_0) = 0$ and $u_{ij}''(x_0) \ge 0$ for all $i, j = 1, 2, \cdots, d$. That means $\mathcal{A}U(x_0) \ge 0$. If $x_0 = 0$ then $\mathcal{A}U(x_0) = 0$.

3.2. Spectral Analysis of the Operator $(\mathcal{A}, D(\mathcal{A}))$

The purpose of next theorem is to deduce under reasonable hypothesis on the coefficients of B a precise description of the spectrum of the operator $(\mathcal{A}, D(\mathcal{A}))$.

Theorem 3.1 *If* $-I_d \ge B(0)$, *then the spectrum of* $(\mathcal{A}, D(\mathcal{A}))$ *is contained in* $]-\infty, 0]$.

Proof. Let $\lambda \in C, U = \left(u_{ij}\right)_{1 \le i, j \le d} \in D(\mathcal{A})$

and $V = (v_{ij})_{1 \le i, j \le d} \in X$.

We have $(\lambda - \mathcal{A})U = V \Leftrightarrow (\lambda - \mathcal{A}_i)u_{ij} = v_{ij}$

for all $i, j = 1, 2, \cdots, d$

$\Leftrightarrow u_{ij} = (\lambda - \mathcal{A}_i)^{-1} v_{ij}$ *for all* $i, j = 1, 2, \cdots, d$

and $\lambda \in \bigcap_{i=1}^{i=d} \rho(\mathcal{A}_i)$.

This is sufficient to note that the spectrum of $(\mathcal{A}, D(\mathcal{A}))$ verifies $\sigma(\mathcal{A}) \subset \bigcup_{i=1}^{i=d} \sigma(\mathcal{A}_i)$ and then the result hold by ([10], Lemma 2.6).

Theorem 3.2 *The operator* $(\mathcal{A}, D(\mathcal{A}))$ *with* $-I_d \ge B(0)$, *generates an analytic semigroup of angle* $\dfrac{\pi}{2}$. *Moreover, the semigroup is positive and contractive.*

Proof. For $0 < \theta < \pi$, put $\sum_\theta = \{z \in C^* ; | arg(z) |\le \pi - \theta\}$. It is clear from theorem 3.1 that $\sum_\theta \subset \rho(\mathcal{A})$ and then for $\lambda \in \sum_\theta$ and $i = 1, 2, \cdots, d$, the operator $(\lambda - \mathcal{A}_i)$ is invertible and verifies $\| (\lambda - \mathcal{A}_i)^{-1} \| \le \dfrac{C_i}{| \lambda |_i}$ for some positive constant C_i (see [10]). So $(\lambda - \mathcal{A})$ is invertible and veri-

fies $\| (\lambda - \mathcal{A})^{-1} \|_X \le \dfrac{\max\limits_{1 \le i \le d} C_i}{|\lambda|}$. Hence the operator

$(\mathcal{A}, D(\mathcal{A}))$ with $-I_d \ge B(0)$, generates an analytic

semigroup of angle $\dfrac{\pi}{2}$. For the positivity using the rela-

tion between the operators $\mathcal{A}_i, i = 1, 2, \cdots, d$ and $\mathcal{A}$
and from the fact that each operator $(\mathcal{A}_i, D(\mathcal{A}_i))$,
$i = 1, 2, \cdots, d$, generates an analytic positive and contrac-
tive semigroup, we deduce the result.

4. The Non Diagonal Case

In this section we consider the operator

$$\mathcal{A}U = U'' + \frac{B(x)}{x} U'$$

Assume that $B(x) = P(x)D(x)P^{-1}(x)$, where D is
diagonal matrix function and P is nonsingular matrix
function. If P is constant, put $V = P^{-1}U$, then

$$U = PV, \ U' = PV' \ and \ U'' = PV''$$

so we obtain

$$P^{-1}\mathcal{A}PV = \left(V'' + \frac{D(x)}{x} V' \right)$$

Hence, from Theorem 3.2 we can easily verify the
following

Proposition 4.1 *The operator* $(\mathcal{A}, D(\mathcal{A}))$ *with*
$-I_d \ge D(0)$, *generates an analytic semigroup of angle*
$\dfrac{\pi}{2}$.

Remark that the condition $-I_d \ge D(0)$ is equivalent
to the fact that the spectrum $\sigma(B(0))$ of the constant
matrix $B(0)$ verifies $\sigma(B(0)) \subseteq]-\infty, -1]$.

We turn now to the general case in which we proceed
with a perturbation argument. For this we recall the
following definition.

Definition 4.1 ([2].*Definition* 2.1).

Let $A : D(A) \subset X \to X$ be a linear operator on the
Banach space X . Any operator $B : D(B) \subset X \to X$ is
called A -bounded if $DA) \subseteq D(B)$ and if there exist
constants a, b in $\mathbb{R}_+$ such that

$$\| BU \| \le a \| AU \| + b \| U \| \ for \ all \ U \in D(A). \quad (5.1)$$

The A -bound of B is

$a_0 = \inf \{ a \ge 0 : there \ exists \ b \ge 0 \ such \ that \ (5.1) \ holds \}$

Proposition 4.2 ([2].*Theorem* 2.10).

Let the operator $(A, D(A))$ generates an analytic
semigroup on a Banach space X . Then there exists a
constant $\alpha > 0$ such that $(A + B, D(A))$ generates an

analytic semigroup for every A -bounded operator B
having A -bound $a_0 < \alpha$.

Introduce now the operators $(\mathcal{A}_0, D(\mathcal{A}_0))$ and
$(\mathcal{R}, D(\mathcal{R}))$ given by

$$\mathcal{A}_0 = \frac{d^2}{dx^2} + \frac{B(0)}{x} \frac{d}{dx} \ and \ \mathcal{R} = \frac{B(x) - B(0)}{x} \frac{d}{dx}$$

with,

$$D(\mathcal{R}) = D(\mathcal{A}_0) = \{ U \in C([0, +\infty],$$
$$M_d(C)) \cap C_\infty^2(]0, +\infty], M_d(C)), \lim_{x \to 0} \mathcal{A}_0 U(x) = 0 \}.$$

Then we have $\mathcal{A} = \mathcal{A}_0 + \mathcal{R}$ and if we choose
$D(\mathcal{A}) = D(\mathcal{A}_0)$, we obtain the principal result of the
paper.

Theorem 4.1 *Assume that $B(0)$ is diagonalizable
and $\sigma(B(0)) \subseteq]-\infty, -1]$ or $B(0) = -I_d$ and (2.2)
holds. Then the operator $(\mathcal{A}, D(\mathcal{A}))$, generates an analy-
tic semigroup of angle $\dfrac{\pi}{2}$.*

Proof. Let $U \in D(\mathcal{A})$ and observe that

$$\mathcal{A}U = \frac{d^2}{dx^2} U + \frac{B(x)}{x} \frac{d}{dx} U$$
$$= \frac{d^2}{dx^2} U + \frac{B(0)}{x} \frac{d}{dx} U + \frac{B(x) - B(0)}{x} \frac{d}{dx} U$$
$$= \mathcal{A}_0 U + \mathcal{A}U,$$

Let $\varepsilon > 0$, there exists $\delta > 0$ such that for all
$0 < x < \delta$ we have $\| B(x) - B(0) \| < \varepsilon$. The formula
$\dfrac{U'(x)}{x} = \int_0^1 U''(xt)dt$ implies that

$$\| RU(x) \| \le \varepsilon \| U'' \|_X, \ 0 < x < \delta.$$

On the other hand from the Taylor expansion to order
2 at $x > \delta$ and for all $U \in D(\mathcal{A})$ there exists a
constant $C_\varepsilon > 0$ such that

$$\| RU(x) \| \le \varepsilon \| U'' \|_X + C_\varepsilon \| U \|_X$$
$$\le \varepsilon \| \mathcal{A}_0 U \|_X + \varepsilon \| \frac{B(0)}{x} U' \|_X + C_\varepsilon \| U \|_X .$$

Since $B(0)$ is diagonalizable and $\sigma(B(0)) \subset]-\infty, -1[$
or B satisfies the condition (2.2) for $B(0) = -I_d$, the
map

$$U \mapsto \frac{B(0)}{x} \frac{d}{dx} U,$$

from $D(\mathcal{A})$ into $C([0, +\infty], M_d(C))$ is continuous
(see [10]. Remark 2.5), so we deduce that the operator
$\mathcal{R}$ is $\mathcal{A}$ -bounded with $\mathcal{A}$ -bound is equal to zero.
Hence, the desired result follows by applying Theorem
3.2 and Proposition 4.2.

5. Application and Example

Assume now that the operator $(\mathcal{A}, D(\mathcal{A}))$ satisfies the assumptions of Theorem 4.1, it generates so an analytic semigroup, and consider the evolution equation problem

$$\begin{cases} \dfrac{dU}{dt}(t) - \mathcal{A}U(t) = f(t), & t \geq 0 \\ U(0) = U_0 \end{cases} \tag{5.2}$$

Corollary 5.1 *If* $f = 0$ *then the problem* (5.2) *has a unique solution for all* $U_0 \in X$. *This solution is of infinitely continuously differentiable on* $]0, +\infty[$.

For general case we have by Pazy [3] and R.autry [13].

Corollary 5.2 *If* $f = (f_{ij})_{1 \leq i, j \leq d}$, *and for all* $i, j, f_{ij} \in L^1(]0, T[, C)$ *and for every* $t \in]0, T[$ *there is a* $\delta_t^{i,j}$ *and a continuous function* $\varphi_t^{i,j} : [0, +\infty[\rightarrow [0, +\infty[$ *such that*

$$\left| f_{ij}(t) - f_{ij}(s) \right| \leq \phi_t^{i,j}(|t - s|) \ and \ \int_0^{\delta_t^{i,j}} \frac{\phi_t^{ij}(\tau) d\tau}{\tau} < \infty.$$

Then the problem (5.2) has a classical solution.

EXAMPLE

Let the Banach space $X = C([0, +\infty], M_d(C))$ and $\alpha \geq 0$. Put $\phi(x) = x^\alpha$ and define the linear transformation P_ϕ on X in itself by $P_\phi U = U o \phi$. It is clear that P_ϕ is invertible and $(P_\phi)^{-1} = P_{\phi^{-1}}$.

Consider now the operator $(\mathcal{L}, D(\mathcal{L}))$ defined on X by

$$\mathcal{L}U = x^{2-2\alpha} U'' + x^{1-2\alpha} B(x) U'$$

with

$$D(\mathcal{L}) = \{ U \in X \cap C^2(]0, +\infty], M_d(C)),$$
$$\mathcal{L}U \in X, \lim_{x \to 0} (\mathcal{L}oP_\phi)U(x) = 0 \},$$

and $B \in X$ is a diagonal matrix real valued function satisfying $B(0) < (1 - 2\alpha)I_d$. A simple calculus gives

$$\mathcal{L} = \alpha^2 P_\phi \mathcal{K} (P_\phi)^{-1} \tag{5.3}$$

where

$$\mathcal{K}U = U'' + \frac{1}{\alpha} \left(\frac{\alpha - 1 + B(x^{\frac{1}{\alpha}})}{x} \right) U'.$$

Put

$$D(\mathcal{K}) = \{ U \in C([0, +\infty], M_d(C)) \cap C_\infty^2(]0, +\infty],$$
$$M_d(C)), \lim_{x \to 0} \mathcal{K}U(x) = 0 \}.$$

From Theorem 3.2 the operator $(\mathcal{K}, D(\mathcal{K}))$ generates an analytic semigroup of angle $\dfrac{\pi}{2}$ moreover, the semi-

group is positive and contractive. Hence the relation (5.3) implies that the operator $(\mathcal{L}, D(\mathcal{L}))$ generates an analytic semigroup of angle $\dfrac{\pi}{2}$ and the semigroup is contractive if $\alpha \geq 1$.

6. Acknowledgements

7. References

[1] R. Nagel, "One-Parameter Semigroups of Positive Operators," Lecture Notes in Math, Springer-Verlag, 1986.

[2] K. J. Engel and R. Nagel, "One-Parameter Semigroups for Linear Evolution Equations," Springer-Verlag, 2000.

[3] A. Pazy, "Semigroups of Linear Operators and Applications to Partial Differential Equations," Applied Math Sciences 44, Springer, 1983.

[4] K. Ito and F. Kappel, "Evolution Equations and Approximations," *Series on Advances in Mathematics for Applied Sciences*, Vol. 61, 2002.

[5] E. M. Ouhabaz, "Analysis of Heat Equations on Domains," Princeton University Press, New Jersey, 2005.

[6] K. J. Engel, "Operator Matrices and Systems of Evolution Equations." (Preprint).

[7] N. H. Mahmoud, "Partial Differential Equations with Matricial Coefficients and Generalised Translation Operators," *Transactions of the Americain Mathematical Society*, Vol. 352, No. 8, 2000, pp. 3687-3706.

[8] N. H. Mahmoud, "Heat Equations Associated with Matrix Singular Differential Operators and Spectral Theory," *Integral Transforms and Special Functions*, Vol. 15, No. 3, 2004, pp. 251-266.

[9] A. Saddi and O. A. M. S. Ahmed, "Analyticity of Semigroups Generated by a Class of Differential Operators with Matrix Coefficients and Interface," *Semigroup Forum*, Vol. 71, No. 1, 2005, pp. 1-17.

[10] G. Metafune, "Analyticity for Some Degenerate One-Dimentional Evolution Equations," *Studia Mathematica*, Vol. 127, No. 3, 1998, pp. 251-276.

[11] Z. S. Agranovich and V. A. Marchenko, "The Inverse Problem of Scattering Theory," Kharkov State University, Gordon and Breach, NewYork and London, 1963.

[12] R. B. Bapat and T. E. S. Raghavan, "Nonegative Matrices and Applications," Cambridge University Press, Cambridge, 1997.

[13] R. Dautray and J.-L. Lions, "Analyse Mathématique et Calcul Numérique pour les Sciences et les Techniques," Tome 3, Série Scientifique, Masson, 1985.

Scientific Research

Artificial Neural Networks Approach for Solving Stokes Problem

Modjtaba Baymani, Asghar Kerayechian, Sohrab Effati
Department of Mathematics, Ferdowsi University of Mashhad, Mashhad, Iran
E-mail: *mhbaymani@yahoo.com, krachian@gmail.com, s-effati@um.ac.ir*

Abstract

In this paper a new method based on neural network has been developed for obtaining the solution of the Stokes problem. We transform the mixed Stokes problem into three independent Poisson problems which by solving them the solution of the Stokes problem is obtained. The results obtained by this method, has been compared with the existing numerical method and with the exact solution of the problem. It can be observed that the current new approximation has higher accuracy. The number of model parameters required is less than conventional methods. The proposed new method is illustrated by an example.

1. Introduction

CFD stands for Computational Fluid Dynamics, a subgenre of fluid mechanics that uses computers (numerical methods and algorithms) to represent, or model, problems that engage fluid flows. CFD software is usually used to solve equations in a discretized way. The domain is transferred into a grid or mesh – a regular/irregular and 2D/3D surface of cells. After discretization, an equation solver runs to solve the equations of fluid motion (Euler equations, Navier-Stokes equations, etc.). Algorithms from numerical linear algebra, like: Gauss-Seidel, successive over relaxation, Krylov subspace method or algorithms from Multigrid family are typically used. These methods involve millions of calculations, so, as it can be easily observed, computing is time consuming. Also in many problems, even with parallel programming and supercomputers, only approximate solutions can be reached.

There are various optimization methods of computer science which can be used for CFD. Simulated Annealing, Genetic Algorithms, Evolution Strategy, Feed-Forward Neural Networks are popular and we reimplemented in many projects.

In [1] a framework is created for evolutionary optimization which is then tested on aerodynamic design example. The framework was based on covariance matrix adaptation, with the feed-forward neural network as an approximate fitness function. An aerodynamic design procedure which combines neural networks with polynomial fits is introduced in [2] and [3] discussed an arti-

ficial neural network which is an approximate model that is used for optimization of the blade geometry by simulated annealing method. Parallel stochastic search algorithm is introduced in [4] and tested on defining a shape of two airfoils. In this work, a performance neural network for solving Stokes equations is presented.

Lagaris, *et al.* [5] used artificial neural networks (ANN) for solving ordinary differential equations and partial differential equations for both boundary value and initial value problems. Canh and Cong [6] presented a new technique for numerical calculation of viscoelastic flow based on the combination of neural networks and Brownian dynamics simulation or stochastic simulation technique (SST). Hayati and Karami [7] used a modified neural network to solve the Berger's equation in one-dimensional quasilinear partial differential equation.

The Stokes equations describe the motion of a fluid in $R^n \ (n = 2 \ or \ 3)$. These equations are to be solved for an unknown velocity vector $u(x, y) = (u_i(x, y))_{1 \le i \le n} \in R^n$ and pressure $p(x, y) \in R$.

We restrict our attention here to incompressible fluids filling all of $R^n \ (n = 2)$ as follow:

$$
\begin{cases}
-\Delta u_1 + \dfrac{\partial p}{\partial x} = f_1 & in \ \Omega \subset R^2 \\[2mm]
-\Delta u_2 + \dfrac{\partial p}{\partial y} = f_2 & in \ \Omega \\[2mm]
\dfrac{\partial u_1}{\partial x} + \dfrac{\partial u_2}{\partial y} = 0 & in \ \Omega
\end{cases}
\tag{1}
$$

with boundary conditions:

$$u = (u_1, u_2) = (u^0{}_1, u^0{}_2) = u^0, \text{ on } \partial\Omega.$$

Here, u^0 is given, C^∞ divergence-free vector field on Ω, f_1, f_2 are the components of a given, externally applied force (e.g. gravity). The first and second equations of (1) are just Newton's law $f = ma$ for fluid element subject to the external force $f = (f_1, f_2)$ and to the forces arising from pressure and friction. The third equation of (1) says that the fluid is incompressible. For physically reasonable solutions, we want to make sure $u = (u_1, u_2)$ does not grow as $|(x, y)| \to \infty$. Hence we will restrict our attention to the force f and initial condition u^0 that satisfy

$$\left|\partial_x^\alpha u^0(x)\right| \le C_{\alpha K}\left(1 + |x|\right)^{-K}, \text{ on } R^n, \text{ for any } \alpha \text{ and some } K,$$

$$\left|\partial_x^\alpha f(x)\right| \le C_{\alpha K}\left(1 + |x|\right)^{-K}, \text{ on } R^n, \text{ for any } \alpha \text{ and some } K.$$

We accept a solution of (1) as physically reasonable only if it satisfies $p, u \in C^\infty(R^n)$ and

$$\int_\Omega |u(x)| dx < C \text{ (bounded energy)}.$$

2. Description of the Method

The usual proposed approach for problem (1) will be illustrated in terms of the following general partial differential equation:

$$G_1\left(x, \psi_1(x), \nabla^2 \psi_1(x), \frac{\partial\phi}{\partial x}(x)\right) = 0, \ x \in D$$

$$G_2\left(x, \psi_2(x), \nabla^2 \psi_2(x), \frac{\partial\phi}{\partial x}(x)\right) = 0, \ x \in D \qquad (2)$$

$$G_3\left(x, \frac{\partial\psi_1}{\partial x}(x), \frac{\partial\psi_2}{\partial y}(x)\right) = 0, \ x \in D$$

subject to certain boundary conditions (BC's) (for example Dirichlet and/or Neumann), where $x = (x_1, ..., x_n) \in R^n$, $D \subset R^n$ denotes the definition domain and $\psi_1(x), \psi_2(x), \phi(x)$ are the solutions to be computed.

If $\psi_{1t}(x, P_1)$, $\psi_{2t}(x, P_2)$, $\phi_t(x, P_3)$ denote trial solutions with adjustable parameters P_1, P_2, P_3, the problem (2) is transformed to

$$\min_{P_1, P_2, P_3} \sum_{x_i \in D} \left[\left(G_1(x_i)\right)^2 + \left(G_2(x_i)\right)^2 + \left(G_3(x_i)\right)^2\right] \qquad (3)$$

subject to the constraints imposed by the BC's.

In the proposed approach, the trial solutions $\psi_{1t}(x, P_1)$, $\psi_{2t}(x, P_2)$, $\phi_t(x, P_3)$ employ a feed forward neural network and parameters P_1, P_2, P_3 correspond to the weights and biases of the neural architecture. We choose trial functions $\psi_{1t}(x, P_1)$, $\psi_{2t}(x, P_2)$, $\phi_t(x, P_3)$ such that by

construction satisfy the BC's. This achieves by writing it as a sum of two terms

$$\psi_{1t}(x) = A_1(x) + F_1\left(x, N_1(x, P_1)\right)$$

$$\psi_{2t}(x) = A_2(x) + F_2\left(x, N_2(x, P_2)\right) \qquad (4)$$

$$\phi_t(x) = A_3(x) + F_3\left(x, N_3(x, P_3)\right)$$

where $N_k(x, P) = \sum_{i=1}^{H} v_i \sigma(z_i)$ and $z_i = \sum_{j=1}^{n} w_{ij} x_j + u_i$

$(k = 1, 2, 3)$ are single-outputs feed forward neural network with parameters P_1, P_2, P_3 and n input units fed with the input vector x. The terms $A_i(x)$ $(i = 1, 2, 3)$ contain no adjustable parameters and satisfy the boundary conditions.

The second terms $F_i(x, N_i(x, P_i))(i = 1, 2, 3)$ is constructed so as not to contribute to the BC's, since $\psi_{1t}(x, P_1)$, $\psi_{2t}(x, P_2)$, $\phi_t(x, P_3)$ must also satisfy the BC's. These terms employ a neural network whose weights and biases are to be adjusted in order to deal with the minimization problem. Note at this point that the problem has been reduced from the original constrained optimization problem to an unconstrained one (which is much easier to handle) due to the choice of the form of the trial solution that satisfies by construction the BC's. In the next section we present a systematic way to construct the trial solution, *i.e.*, the functional forms of both A_i and F_i. We treat several common cases that one frequently encounters in various scientific fields. As indicated by our experiments, the approach based on the above formulation is very effective and provides in reasonable computing time accurate solutions with impressive generalization (interpolation) properties.

3. Neural Network for Solving Stokes Equations

To solve problem (1) with $\Omega = [0,1] \times [0,1]$, we apply the operators $\dfrac{\partial}{\partial x}$ and $\dfrac{\partial}{\partial y}$ on the first and second equations respectively. Then we obtain:

$$-\Delta\left(\frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y}\right) + \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} = (f_1)_x + (f_2)_y \qquad (5)$$

Using the third equation in (1), the Equation (5) may be written as:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} = (f_1)_x + (f_2)_y \qquad (6)$$

this is the Poisson equation, and has infinitely many solutions. By imposing some boundary conditions, we are going to obtain an appropriate solution for Equation (6)

by the neural network.

The trial solution is written as

$$p_t(x, y) = A(x, y) + x(1 - x)y(1 - y)N(x, y, P) \quad (7)$$

where $A(x, y)$ is chosen so as to satisfy the BC, namely

$$\begin{aligned}A(x, y) &= (1 - x)h_0(y) + xh_1(y) \\ &+ (1 - y)\{g_0(x) - [(1 - x)g_0(0) + xg_0(1)]\} \\ &+ y\{g_1(x) - [(1 - x)g_1(0) + xg_1(1)]\}.\end{aligned} \quad (8)$$

where

$h_0(y) = p(0, y)$, $h_1(y) = p(1, y)$, $g_0(x) = p(x, 0)$ and $g_1(x) = p(x, 1)$.

Note that the second term of the trial solution does not affect the boundary conditions since it vanishes at the part of the boundary where Dirichlet BC's are imposed. In the above PDE problems the error to be minimized is given by

$$E(p) = \sum_i \left\{ \frac{\partial^2 p_t(x_i, y_i)}{\partial x^2} + \frac{\partial^2 p_t(x_i, y_i)}{\partial y^2} - F(x_i, y_i) \right\}^2 \quad (9)$$

where $F = (f_1)_x + (f_2)_y$ and (x_i, y_i) is a point in Ω.

By solving the optimization problem (9), the weights v_i, w_{ij}, u_i are obtained and then a trial solution for p_t is obtained. By substituting the trial solution p_t in the first equation of (1) we obtain:

$$\Delta u_1 = \frac{\partial(p_t)}{\partial x} - f_1 \quad (10)$$

which is a Poisson equation for u_1, by using (9) and by substituting u_{1t} and $\frac{\partial(p_t)}{\partial x} - f_1$ for or p and F, respectively, we can obtain a trial solution for u_{1t}. To obtain u_2, we can substitute the trial solution p_t in the second equation in (1) to obtain:

$$\Delta u_2 = \frac{\partial(p_t)}{\partial y} - f_2. \quad (11)$$

In a similar manner we can obtain u_{2t}.

4. Numerical Examples

In this section we present one example to illustrate the method. We used a multilayer perceptron having one hidden layer with five hidden units and one linear output unit. For a given input vector $x = (x_1, x_2)$ the output of the network is $N(x, P) = \sum_{i=1}^{5} v_i \sigma(z_i)$ where $z_i = \sum_{j=1}^{2} w_{ij}x_j + u_i$ and $\sigma(x) = \frac{1}{1 + e^{-x}}$. The exact analytic solution is known in advance. Therefore we test the

accuracy of the obtained solution.

Example. Consider the problem (1) with $\Omega = [0,1] \times [0,1]$. We choose f_1 and f_2 such that the exact solution for u_1, u_2 and p be as follows:

$$u_1 = 10x^2 y(1 - x)^2(1 - 3y + 2y^2)$$
$$u_2 = -10y^2 x(1 - y)^2(1 - 3x + 2x^2)$$
$$p = 5(x^2 - y^2).$$

The domain Ω is first discretized by uniform mesh of size $h = 1/3$ (4 points). This initial mesh is successively refined to produce meshes with sizes $h = 1/4$ and $h = 1/5$ (respectively 9 and 16 points). **Table 1** reports the maximum error at nodal points (Maximum error) at the training set points and the distances $\|p - p_t\|_{L_2}$ and $\|p - p_t\|_{H^{(1)}}$ between the exact solution and the training solution.

In **Figure 1** the error function $p - p_t$ for N = 25 is depicted which shows the solution is very accurate.

We used the Equation (10) and obtained the solution u_{1t}. **Table 2** reports the maximum error at the training set points and the distances $\|u_1 - u_{1t}\|_2$ and $\|u_1 - u_{1t}\|_{H^{(1)}}$ between the exact solution and the training solution.

In a similar manner we obtained u_{2t}. **Table 3** reports the maximum error at the training set points and the dis-

Table 1. Maximum error at training set points and the distances $\|p - p_t\|_2$ and $\|p - p_t\|_{H^{(1)}}$.

	N = 4	N = 9	N = 16
Maximum Error	0.0058	5.1007e-4	1.8433e-7
$\|p - p_t\|_2$	1.0025e-5	7.5171e-8	4.1647e-15
$\|p - p_t\|_{H^{(1)}}$	4.1569e-4	7.5120e-6	4.9101e-13

Table 2. Maximum error at training set points and the distances $\|u_1 - u_{1t}\|_2$ and $\|u_1 - u_{1t}\|_{H^{(1)}}$.

	N = 9	N = 16	N = 25
Maximum error	0.0372	2.0723e-8	1.733e-8
$\|u_1 - u_{1t}\|_2$	4.091e-4	4.487e-17	2.576e-17
$\|u_1 - u_{1t}\|_{H^{(1)}}$	0.040	6.4928e-15	1.632e-15

Table 3. Maximum error at training set points and the distances $\|u_2 - u_{2t}\|_2$ and $\|u_2 - u_{2t}\|_{H^{(1)}}$.

	N = 9	N = 16	N = 25
Maximum error	0.0379	2.6141e-05	2.6507e-07
$\|u_2 - u_{2t}\|_2$	4.0684e-04	1.2722e-010	1.3238e-14
$\|u_2 - u_{2t}\|_{H^{(1)}}$	0.0407	1.8825e-008	2.7010e-12

tances $\left\|u_2 - u_{2t}\right\|_2$ and $\left\|u_2 - u_{2t}\right\|_{H^{(1)}}$ between the exact solution and the training solution.

In **Figures 2** and **3** the error functions $u_1 - u_{1t}$ and $u_2 - u_{2t}$ for N = 25 is depicted which shows the solutions are very accurate (even between training points).

In **Figures 4-7** differential of the error functions $u_1 - u_{1t}$ and $u_2 - u_{2t}$ respect to x and y, for N = 25 are depicted, respectively, which show the solutions are differentiable.

Aman and Kerayechian [8] used linear programming-methods to solve the above problem. They converted the Stokes problem to a minimization problem, then by discretizing the minimization problem. They obtained a linear programming problem and solved it. **Table 4** present the comparison between the proposed method and with Aman – Kerayechian method for 25 points.
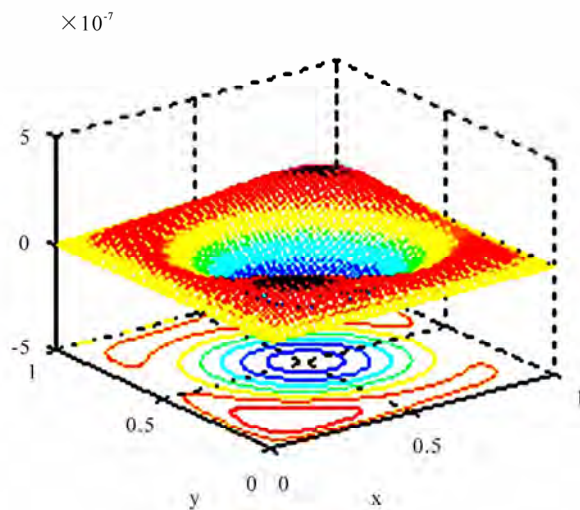
$\times 10^{-6}$

Figure 3. The error function $u_2 - u_{2t}$.

$\times 10^{-7}$

Figure 1. Error function $p - p_t$.

$\times 10^{-7}$
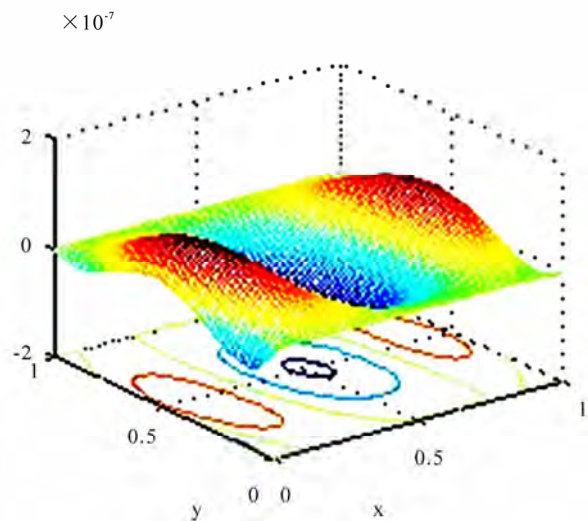
Figure 4. The differential of error function $u_1 - u_{1t}$ **respect to** x.

Table 4. The comparison of our proposed method with Aman – Kerayechian method.

Errors	The proposed method	Aman-Kerayechian method
Maximum error $u_1 - u_{1t}$	1.7335e-008	0.007885
$\left\|u_1 - u_{1t}\right\|_{H^{(1)}}$	1.6328e-015	0.082281
Maximum error $u_2 - u_{2t}$	2.6507e-007	0.007885
$\left\|u_2 - u_{2t}\right\|_{H^{(1)}}$	2.7010e-012	0.082281
Maximum error $p - p_t$	1.6392e-005	0.035104
$\left\|p - p_t\right\|_2$	2.2448e-011	0.027446

$\times 10^{-8}$

Figure 2. The error function $u_1 - u_{1t}$.

Figure 5. The differential of error function $u_1 - u_{1t}$ respect to y.

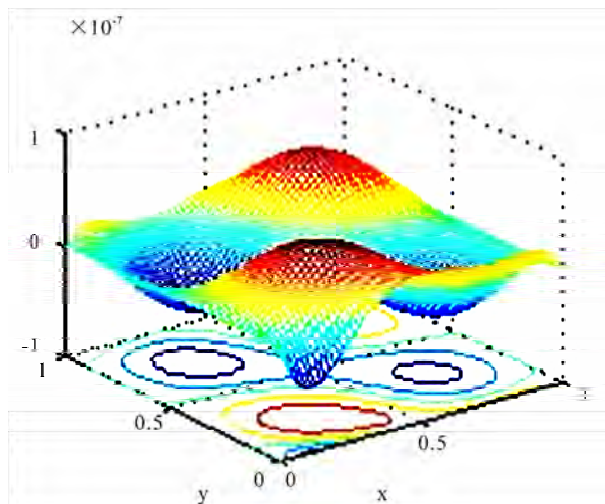

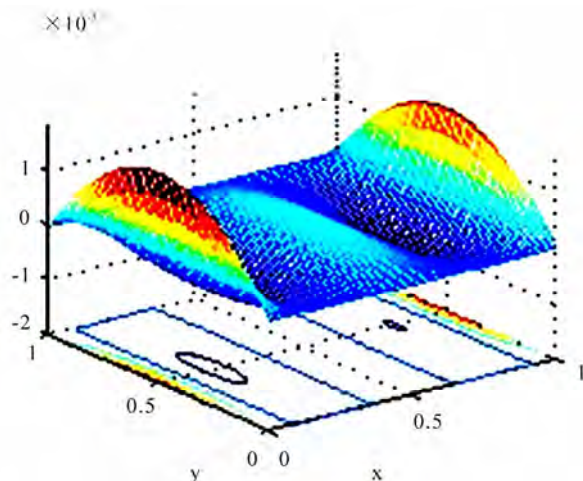**Figure 7. The differential of error function $u_2 - u_{2t}$ respect to y.**

Figure 6. The differential of error function $u_2 - u_{2t}$ respect to x.

5. Conclusions

In this paper a new method based on ANN has been applied to find solution for Stokes equation. The solution via ANN method is a differentiable, closed analytic form easily used in any subsequent calculation. The neural network here allows us to obtain fast solution of Stokes equation starting from randomly sampled data sets and refined it without wasting memory space and therefore reducing the complexity of the problem.

If we compare the results of the numerical methods (see [8]) with our method, we see that our method has some small error. Other advantage of this method, the solution of the Stokes problem is available for each arbitrary point in training interval (even between training points). Indeed, 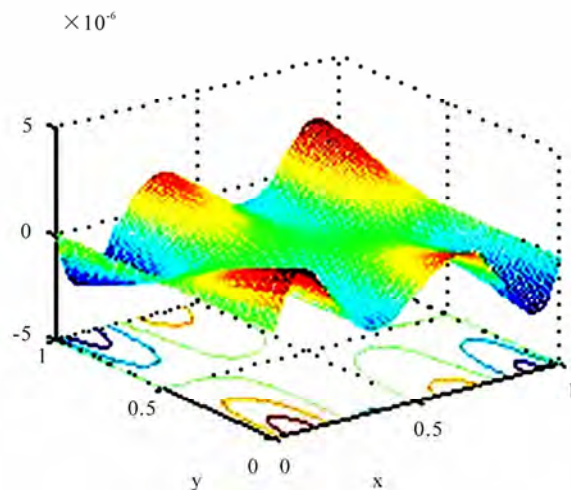after solving the Stokes problem, we obtain an approximated function for the solution and so we can calculate the answer at every point immediately.

6. References

[1] Y. C. Jin, M. Olhofer and B. Sendhoff, "A Framework for Evolutionary Optimization with Approximate Fitness Functions, Evolutionary Computation," *IEEE Transactions*, Vol. 6, No. 5, 2002, pp. 481-494.

[2] M. H. Man, R. Nateri and K. Madavan, "Aerodynamic Design Using Neural Networks," *American Institute of Aeronautics and Astronautics Journal*, Vol. 38, No. 1, 2000, pp. 173-182.

[3] S. Pierret, "Turbo Machinery Blade Design Using a Navier-Stokes Solver and Artificial Neural Network," *ASME Journal of Turbomach*, Vol. 121, No. 3, 1999, pp. 326-332.

[4] T. Ray, H. Tsai and C. Tan, "Effects of Solver Fidelity on a Parallel Search Algorithm's Performance for Airfoil Shape Optimization Problems," 9*th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, Georgia, 2002, pp. 1816-1826.

[5] I. E. Largris and A. Likas, "Artificial Neural Networks for Solving Ordinary and Partial Differential Equations," *IEEE transaction on neural networks*, Vol. 9, No. 5, 1998, pp. 987-1000.

[6] D. Tran-Canh and T. Tran-Cong, "Computation of Viscoelastic Flow Using Neural Networks and Stochastic Simulation," *Korea-Australia Rheology Journal*, Vol. 14, No. 4, 2002, pp. 161-174.

[7] M. Hayati and B. Karami, "Feedforward Neural Network for Solving Partial Differential Equations," *Journal of Applied Sciences*, Vol. 7, No. 19, 2007, pp 2812-2817.

[8] M. Aman and A. Kerayechian, "Solving the Stokes Problem by Linear Programming Methods," *International Mathematical Journal*, Vol. 5, No. 1, 2004, pp. 9-22.

Scientific
Research

Entire Large Solutions of Quasilinear Elliptic Equations of Mixed Type*

Hongxia Qin[1], Zuodong Yang[1,2]

[1]*Institute of Mathematics, School of Mathematical Sciences, Nanjing Normal University,*
Nanjing, China
[2]*College of Zhongbei, Nanjing Normal University, Nanjing, China*
E-mail: *zdyang_jin@263.net*

Abstract

In this paper, the existence and nonexistence of nonnegative entire large solutions for the quasilinear elliptic equation $div\left(|\nabla u|^{m-2}\nabla u\right) = p(x)f(u) + q(x)g(u)$ are established, where $m \geq 2$, f and g are nondecreasing and vanish at the origin. The locally Hölder continuous functions p and q are nonnegative. We extend results previously obtained for special cases of f and g.

Keywords: Entire Solutions, Large Solutions, Quasilinear Elliptic Equations

1. Introduction

In this paper, we consider the problem

$$\begin{cases} div\left(|\nabla u|^{m-2}\nabla u\right) = p(x)f(u) + q(x)g(u), \text{ in } R^N (N \geq 3) \\ u(x) \to \infty, \quad as \quad |x| \to \infty \end{cases}$$

(1)

where $m \geq 2$, $f, g \in C([0,\infty),[0,\infty)) \bigcap C^1((0,\infty),[0,\infty))$, the locally Hölder continuous functions p and q are nonnegative on $\mathbf{R}^N$. In addition, we assume that

$$f(0) = g(0) = 0; f'(t) \geq 0, g'(t) \geq 0,$$
$$f(t)g(t) > 0, \text{ for } \forall t > 0$$

(2)

We call nonnegative solutions of (1) entire large solutions. In fact, this problem appears in the study of non-Newtonian fluids [1,2] and non-Newtonian filtration [3,4], such problems also arise in the study of the sub-sonic motion of a gas [5], the electric potential in some bodies [6], and Riemannian geometry [7].

Large solutions of the problem

$$\begin{cases} \Delta u(x) = f(u(x)), \quad x \in \Omega, \\ u|_{\partial\Omega} = \infty, \end{cases}$$

(3)

where Ω is a bounded domain in $R^N (N \geq 1)$ have been extensively studied, see [8-20]. A problem with $f(u) = e^u$ and $N = 2$ was first considered by Bieber-

bach [13] in 1916. Bieberbach showed that if Ω is a bounded domain in R^2 such that $\partial\Omega$ is a C^2 submanifold of R^2, then there exisSts a unique $u \in C^2(\Omega)$ such that $\Delta u = e^u$ in Ω and $|u(x) - \ln(d(x))^{-2}|$ is bounded on Ω. Here $d(x)$ denotes the distance from a point x to $\partial\Omega$. Rademacher [17], using the idea of Bieberbach, extended the above result to a smooth bounded domain in R^3. In this case the problem plays an important role, when $N = 2$, in the theory of Riemann surfaces of constant negative curvature and in the theory of automorphic functions, and when $N = 3$, according to [17], in the study of the electric potential in a glowing hollow metal body. Lazer and McKenna [6] extended the results for a bounded domain Ω in $R^N (N \geq 1)$ satisfying a uniform external sphere condition and the non-linearity $f = f(x,u) = p(x)e^u$, where $p(x)$ is continuous and strictly positive on $\overline{\Omega}$. Lazer and McKenna [12] obtained similar results when Δ is replaced by the Monge-Ampere operator and Ω is a smooth, strictly convex, bounded domain. Similar results were also obtained for $f = p(x)u^a$ with $a > 1$. Posteraro [16], for $f(u) = -e^u$ and $N \geq 2$, proved the estimates for the solution $u(x)$ of the problem (1,2) and for the measure of Ω comparing with a problem of the same type defined in a ball. In particular, when $N = 2$, Posteraro [16] obtained an explicit estimate of the minimum of $u(x)$ in terms of the measure of Ω:

$$\min_{\Omega} u(x) \geq \ln(8\pi / |\Omega|).$$

The existence, but not uniqueness, of solutions of the problem (3) with f monotone was studied by Keller [18]. For $f(u) = -u^a$ with $a > 1$, the problem (3) is of interest in the study of the sub-sonic motion of a gas when $a = 2$ (see [15]) and is related to a problem involving super-diffusion, particularly for $1 < a \le 2$ (see [21,22]). Pohozaev [15] proved the existence, but not uniqueness, for the problem (1.2) when $f(u) = -u^2$. For the case where $f(u) = -u^{(N+2)/(N-2)}(N > 2)$, Loewener and Nirenberg [20] proved that if $\partial\Omega$ consists of a disjoint union of finitely compact C^∞ manifolds, each having co-dimension less than $N/2 + 1$, then there exists a unique solution of the problem (3). The uniqueness was established for $f(u) = -u^a$ with $a > 3$, when $\partial\Omega$ is a C^2-submanifold and Δ is replaced by a more general second-order elliptic operator, by Kondrat'ev and Nikishkin [19]. Marcus and Veron [14] proved the uniqueness for $f(u) = -u^a$ with $a > 1$, when $\partial\Omega$ is compact and is locally the graph of a continuous function defined on an $(N-1)$-dimensional space.

In [23], the authors considered the problem of existence and nonexistence of positive entire large solutions of the semilinear elliptic equation

$$\Delta u = p(x)u^\alpha + q(x)u^\beta, \quad 0 < \alpha \le \beta.$$

Recently [24], which is to extend some of these results to a more general the problem

$$\begin{cases} \Delta u = p(x)f(u) + q(x)g(u) \ in \ R^N, N \ge 3, \\ u(x) \to \infty \quad as \ |x| \to \infty. \end{cases}$$

Quasilinear elliptic problems with boundary blowup

$$\begin{cases} div\left(|\nabla u|^{m-2}\nabla u\right) = f(u(x)), \ x \in \Omega, \\ u\,|_{\partial\Omega} = \infty, \end{cases} \tag{4}$$

have been studied, see [9,25,26] and the references therein. Diaz and Letelier [10] proved the existence and uniqueness of large solutions to the problem (4) both for $f(u) = u^\gamma, \gamma > m-1$(super-linear case) and $\partial\Omega$ being of the class C^2. Lu, Yang and E.H.Twizell [25] proved the existence of Large solutions to the problem (4) both for $f(u) = u^\gamma, \gamma > m-1, \Omega = R^N$ or Ω being a bounded domain (super-linear case) and $\gamma \le m-1, \Omega = R^N$ (sub-linear case) respectively.

Recently [27], which is to extend some results of [28] to the following quasilinear elliptic problem

$$\begin{cases} div\left(|\nabla u|^{m-2}\nabla u\right) = p(x)f(u), \ in \ \Omega \\ u(x) \to \infty, \quad on \ \partial\Omega \end{cases} \tag{5}$$

where $\Omega \subseteq \mathbf{R}^N$, the non-negative function $p(x) \in C(\overline{\Omega})$, and the continuous function f satisfies (2) and the Keller-Osserman condition

$$\int_1^\infty [F(s)]^{-1/m} ds = \infty, \ F(s) = \int_0^s f(t)dt \tag{6}$$

then the author also consider the nonexistence for the non-negative non-trivial entire bounded radial solution on $\mathbf{R}^N$ of (5) when p satisfies

$$\int_0^\infty (tp_*(t))^{1/(m-1)} dt = \infty, \ p_*(t) = \min_{|x|=t} p(x). \tag{7}$$

On the other hand, if f does not satisfy (6), that is $\int_1^\infty [F(s)]^{-1/m} ds < \infty$, we can obtain from Lemma 2.4 in [29] that

$$\int_1^\infty \frac{1}{f^{1/(m-1)}(s)} ds < \infty \tag{8}$$

which is also shown in [30]. In this paper, we will require the above integral to be infinite, that is

$$\int_1^\infty \frac{1}{f^{1/(m-1)}(s)} ds = \infty \tag{9}$$

which is a very important condition in our main results. Furthermore, motivated by the results of [24], we also admit the following condition which is opposite to (7), that is

$$\int_0^\infty (tp^*(t))^{1/(m-1)} dt < \infty, \ p^*(t) = \max_{|x|=t} p(x). \tag{10}$$

As far as the authors know, however, there are no results which contain the existence criteria of positive solutions to the problem (1). In this paper, we prove the existence of the positive large solutions for the problem (1). When $p = 2$, the related results have been obtained by A.Lair and A.Mohammed [24]. The main results of the present paper contain extension of the results in [24] and complement of the results in [10,25,26].

The plan of the paper is as follows. In Section 2, for the convenience of the reader we give some basic lemmas that will be used in proving our results. In Section 3 we state and prove the main results. Section 4 contains some consequences of the main theorems, and other results. In Section 5 we present an Appendix where we state and prove three lemmas needed for proofs in previous sections.

2. Preliminary

In this section, we give some results that we shall use in the rest of the paper.

Lemma 2.1.(Weak comparison principle)(see [25]) Let Ω be a bounded domain in $\mathbf{R}^N$ ($N \ge 2$) with smooth boundary $\partial\Omega$ and $\theta : (0,\infty) \to (0,\infty)$ is continuous and non-decreasing. Let $u_1, u_2 \in W^{1,m}(\Omega)$ satisfy

$$\int_\Omega |\nabla u_1|^{m-2} \nabla u_1 \nabla \psi dx + \int_\Omega \theta(u_1)\psi dx$$

$$\le \int_\Omega |\nabla u_2|^{m-2} \nabla u_2 \nabla \psi dx + \int_\Omega \theta(u_2)\psi dx$$

AM

for all non-negative $\psi \in W_0^{1,m}(\Omega)$. Then the inequality

$$u_1 \le u_2 \quad on \quad \partial\Omega$$

implies that

$$u_1 \le u_2 \quad in \quad \Omega.$$

Lemma 2.2. Let f verify (9), and $\rho : [0, \infty) \to [0, \infty)$ be continuous. Then

$$\begin{cases} (r^{N-1}\Phi_m(v'(r)))' = r^{N-1}\rho(r)f(v), & r > 0 \\ v(0) = \alpha, \quad v'(0) = 0 \end{cases} \quad (11)$$

admits a non-negative solution v defined on $[0, \infty)$, where $\Phi_m(s) = |s|^{m-2} s$. If in addition f is nondecreasing and ρ satisfies (7), then $v(r) \to \infty$ as $r \to \infty$.

Proof. First we note that (11) has a solution $v \in C^1(0, R)$ for a maximal $0 < R \le \infty$. As a consequence of (7) we claim that $R = \infty$. By way of contradiction, let us suppose that $0 < R < \infty$ instead. Then we must have $v(r) \to \infty$ as $r \to R^-$. Let

$$\phi(t) := \sup\{\rho(s) : 0 < s \le t\}, \quad t > 0$$

Then ϕ is nondecreasing, and clearly $\rho(t) \le \phi(t)$ for $t > 0$. Integrating Equation (11) from 0 to r yields

$$\Phi_m(v'(r)) = r^{1-N} \int_0^r s^{N-1}\rho(s)f(v(s))ds \quad (12)$$

From (12) we see that $v'(r) \ge 0$, therefore, v is a non-decreasing function and we can obtain from (12) that $\Phi_m(v'(r)) \le \dfrac{r}{N}\varphi(r)f(v(r))$. Then we can obtain

$$\int_0^r \frac{v'(t)}{f^{1/(m-1)}(v(t))}dt \le \int_0^r \left(\frac{t}{N}\right)^{1/(m-1)}\varphi^{1/(m-1)}(t)dt$$

That is

$$\int_\alpha^{v(r)} \frac{1}{f^{1/(m-1)}(s)}ds \le \int_0^r \left(\frac{t}{N}\right)^{1/(m-1)}\varphi^{1/(m-1)}(s)ds$$

Letting $r \to R$, and recalling that $v(r) \to \infty$, we conclude that

$$\int_\alpha^\infty \frac{1}{f^{1/(m-1)}(s)}ds \le \left(\frac{R}{N}\right)^{1/(m-1)}\int_0^R \varphi^{1/(m-1)}(s)ds$$

which is an obvious contradiction. Thus, indeed v is defined on $(0, \infty)$.

We now show that $v(r) \to \infty$ as $r \to \infty$. For this we will use (7) on ρ. Integrating the equation in (11) we find

$$v(r) = \alpha + \int_0^r \left(t^{1-N}\int_0^t s^{N-1}\rho(s)f(v(s))ds\right)^{1/(m-1)}dt$$

$$\ge \left(\frac{f(\alpha)}{N}\right)^{1/(m-1)}\int_0^r (t\rho_*(t))^{1/(m-1)}dt$$

That is

$$v(r) \ge C(m, \alpha, N)\int_0^r (t\rho_*(t))^{1/(m-1)}dt, \quad r > 0$$

and as a consequence of (7) we conclude that $v(r) \to \infty$ as $r \to \infty$.

3. Main Theorems

In this section, we will state the first of our main results.

Theorem 3.1. Under the following hypotheses

$$(H1)\psi(t) = \int_1^t \frac{1}{f(s)}ds, \quad t > 0;$$

$$(H2)\int_0^\infty \left(\left(p^*\right)^{1/(m-1)}(t) - \left(p_*\right)^{1/(m-1)}(t)\right)$$

$$\left(tf\left(\phi(\overline{p}(t))\right)\right)^{1/(m-1)}dt < \infty, \quad \overline{p}(t) = \int_0^t sp_*(s)ds,$$

where ϕ is the inverse of ψ;

$$(H3)\int_0^\infty \left(tq^*(t)g\left(\phi(\overline{p}(t))\right)\right)^{1/(m-1)}dt < \infty$$

Let f and g satisfy (2). Furthermore, assume that (9) and (10) hold. If p satisfies (7), then (1) admits a solution.

Proof. Let v be an entire radial large solution of $div(|\nabla v|^{m-2}\nabla v) = p_*(|x|)f(v)$ such that $v(0) = \alpha$ for some $0 < \alpha < 1$. This is possible by Lemma 2.2, since f satisfies (9) and p_* satisfies (7). Thus v is a super-solution of (1). We proceed to construct a sub-solution u of (1) such that $u \le v$ on $\mathbf{R}^N$. Then by the standard regularity argument for elliptic problems, it is a straight forward argument to prove that (1) would have a solution w such that $u \le w \le v$ on $\mathbf{R}^N$. For each positive integer n, let u_n be a solution of

$$\begin{cases} div(|\nabla u|^{m-2}\nabla u) \\ = p^*(|x|)f(u) + q^*(|x|)g(u), 0.4cm \ in \ B_n, \\ u(x) = v, 0.4cm \ on \quad \partial B_n, \end{cases} \quad (13)$$

where $B_n = B(0, n)$ is the ball of radius n centered at the origin. That such a solution exists is shown in Lemma 5.2 of Appendix. Then we note that each u_n is a radial solution and that

$$0 < u_{n+1} \le u_n \le v, \quad on \quad B_n.$$

Let

$$u(x) := \lim_{n \to \infty} u_n(x), x \in R^N.$$

Since each u_n is radial, it follows that u is radial as well. By a standard argument we can show that u is a solution of the differential equation in (1). Clearly $u \le v$ on $\mathbf{R}^N$. So We only prove that u is nontrivial and that $u(x) \to \infty$ as $|x| \to \infty$.

Recalling that u_n and v are radial and that $u_n(n) = v(n)$ we see that

$$u_n(0) + \int_0^n (t^{1-N} \int_0^t s^{N-1}(p^*(s)f(u_n) + q^*(s)g(u_n))ds)^{1/(m-1)} dt$$

$$= v(0) + \int_0^n \left(t^{1-N} \int_0^t s^{N-1} p_*(s)f(v)ds\right)^{1/(m-1)} dt \geq v(0) + \int_0^n \left(t^{1-N} \int_0^t s^{N-1} p_*(s)f(u_n)ds\right)^{1/(m-1)} dt$$

for $\forall x \geq 0, m \geq 2$, we can use the inequality $(1+x)^{1/(m-1)} \leq 1 + x^{1/(m-1)}$, then we obtain

$$u_n(0) + \int_0^n \left(t^{1-N} \int_0^t s^{N-1} p^*(s)f(u_n)ds\right)^{1/(m-1)} dt + \int_0^n \left(t^{1-N} \int_0^t s^{N-1} q^*(s)g(u_n)ds\right)^{1/(m-1)} dt$$

$$-\int_0^n \left(t^{1-N} \int_0^t s^{N-1} p_*(s)f(u_n)ds\right)^{1/(m-1)} dt \geq v(0) = \alpha$$

that is

$$u_n(0) + \int_0^n (t^{(1-N)/(m-1)} \left((\int_0^t s^{N-1} p^*(s)f(u_n)ds)^{1/(m-1)} - (\int_0^t s^{N-1} p_*(s)f(u_n)ds)^{1/(m-1)} \right) dt$$

$$+ \int_0^n \left(t^{1-N} \int_0^t s^{N-1} q^*(s)g(u_n)ds\right)^{1/(m-1)} dt \geq v(0) = \alpha$$

Since $p^*(s)$ is increasing and $p_*(s)$ is decreasing, so

$$\int_0^n t^{(1-N)/(m-1)} \left(\left(\int_0^t s^{N-1} p^*(s)f(u_n)ds\right)^{1/(m-1)} - \left(\int_0^t s^{N-1} p_*(s)f(u_n)ds\right)^{1/(m-1)} \right) dt$$

$$\leq \int_0^n t^{(1-N)/(m-1)}((p^*)^{1/(m-1)}(t)\left(\int_0^t s^{N-1} f(u_n(s))ds\right)^{1/(m-1)} - (p_*)^{1/(m-1)}(t)\left(\int_0^t s^{N-1} f(u_n(s))ds\right)^{1/(m-1)}) dt$$

$$\leq \int_0^n \left((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t)\right) t^{(1-N)/(m-1)} \left(\int_0^t s^{N-1} f(u_n(s))ds\right)^{1/(m-1)} dt$$

$$\leq \int_0^n \left((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t)\right)\left(\int_0^t f(u_n(s))ds\right)^{1/(m-1)} dt$$

$$\leq C_1(m) \int_0^n \left((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t)\right)(tf(u_n(t)))^{1/(m-1)} dt$$

and

$$\int_0^n \left(t^{1-N} \int_0^t s^{N-1} q^*(s)g(u_n)ds\right)^{1/(m-1)} dt \leq C_2(m) \int_0^n \left(tq^*(t)g(u_n(t))\right)^{1/(m-1)} dt$$

Therefore we get

$$u_n(0) + C_1(m,N)\int_0^n \left((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t)\right)(tf(u_n))^{1/(m-1)} dt + C_2(m,N)\int_0^n \left(tq^*(t)g(u_n)\right)^{1/(m-1)} dt \geq \alpha \quad (14)$$

Now, let $\phi(t)$ be the inverse of the increasing function defined in (9). We note that $\phi(t) \geq 1$ for all $t \geq 0$. Furthermore, we have

$$\phi'(t) = f(\phi(t)), \phi''(t) = f'(\phi(t))f(\phi(t)), t > 0.$$

Let w be an entire large solution of $\Delta w = p_*(|x|)$ such that $w(0) = 0$. Set $a(x) := \phi(w(x))$. Then $\Delta a \geq p_*(|x|)f(a)$. Since

$a(0) = \phi(w(0)) \geq 1 > \alpha = v(0) > 0$, we invoke Lemma 2.1 in [24] to conclude that $v(x) \leq a(x)$ for all $x \in \mathbf{R}^N$.

Moreover, $w(r) \leq \overline{p}(r) := \int_0^r t p_*(t)dt$, we have

$v(x) \leq \phi(\overline{p}(|x|))$.

Now, recalling that $u_n \leq v$ for all $n \in \mathbf{N}$ we see that

$$tq^*(t)g(u_n(t)) \leq tq^*(t)g(v(t)) \leq tq^*(t)g(\phi(\overline{p}(t))),$$

$$((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t))(tf(u_n(t)))^{1/(m-1)}$$

$$\leq ((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t))(tf(v(t)))^{1/(m-1)}$$

$$\leq ((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t))(tf(\phi(\overline{p}(t))))^{1/(m-1)}$$

Take note of (9) and (10), we invoke the Lebesgue dominated convergence theorem to infer from (14) that

$$u(0)$$

$$+C_1(m,N)\int_0^n \left((p^*)^{1/(m-1)}(t) - (p_*)^{1/(m-1)}(t)\right)(tf(u))^{1/(m-1)} dt$$

$$+C_2(m,N)\int_0^n \left(tq^*(t)g(u)\right)^{1/(m-1)} dt \geq \alpha > 0.$$

This shows that u is nontrivial. Now we note that

$$u_n(r) = u_n(0)$$

$$+\int_0^r \left(t^{1-N} \int_0^t s^{N-1}\left(p^*(s)f(u_n) + q^*(s)g(u_n)\right)ds\right)^{1/(m-1)} dt$$

$$\geq \int_0^r \left(t^{1-N} \int_0^t s^{N-1} p^*(s)f(u_n)ds\right)^{1/(m-1)} dt.$$

Recalling that $u_n \leq v$ for all n, we invoke the Lebesgue dominated convergence theorem again, on letting $n \to \infty$

$$u(r) \geq \int_0^r \left(t^{1-N} \int_0^t s^{N-1} p^*(s)f(u)ds\right)^{1/(m-1)} dt.$$

Since u is nontrivial we see that $u(\frac{r_0}{2}) > 0$ for some $r_0 > 0$. Thus for $r > r_0$, we have

$$u(r) \geq f^{1/(m-1)}\left(u(\frac{r_0}{2})\right)\int_{r_0}^{r}\left(t^{1-N}\int_{0}^{t}s^{N-1}p^*(s)ds\right)^{1/(m-1)}dt$$

then

$$u(r) \geq f^{1/(m-1)}\left(u(\frac{r_0}{2})\right)\left(\frac{1}{N}\right)^{1/(m-1)}\int_{r_0}^{r}(tp_*(t))^{1/(m-1)}dt$$

Therefore, as a consequence of (7) we see that $u(x) \to \infty$ as $|x| \to \infty$.

To show our next main result, now we set p is c-positive on Ω (*i.e.*, for any $x_0 \in \Omega$ satisfying $p(x_0) = 0$, there exists a domain Ω_0 such that $x_0 \in \Omega_0, \overline{\Omega}_0 \subset \Omega$, and $p(x) > 0$ for all $x \in \partial\Omega_0$.) we know that p is c-positive on $\mathbf{R}^N$ if and only if there is a sequence Ω_n of smooth bounded domains with $\Omega_n \subseteq \Omega_{n+1}$ for each n such that $\bigcup_{n=1}^{\infty}\Omega_n = \mathbf{R}^N$ and p is c-positive on each Ω. It is easy to see that if γ is a non-negative and locally Hölder continuous function in $\mathbf{R}^N$ that satisfies (10), then the following problem admits a positive solution.

$$\begin{cases} -div(|\nabla w|^{m-2}\nabla w) = \gamma(x), & x \in \mathbf{R}^N \\ w(x) \to 0, & |x| \to \infty \end{cases} \quad (15)$$

In fact

$$v(x) = \int_{|x|}^{\infty}\left(t^{1-N}\int_{0}^{t}s^{N-1}\gamma^*(s)ds\right)^{1/(m-1)}dt.$$

is a super-solution of (15) such that $v(x) \to 0$ as $|x| \to \infty$. On the other hand, 0 is a sub-solution of (15), (See [31], Lemma 3) the assertion follows.

Theorem 3.2. Suppose f and g satisfy (2). If (1) has a solution, f satisfies (8) and p is c-positive in $\mathbf{R}^N$, (3) admits a solution. Conversely, if $f + g$ satisfies (8), (15) admits a non-negative solution with $\gamma(x) = p(x) + q(x)$ and $\xi(x) := \min\{p(x), q(x)\}$ is c-positive, then (1) has a solution.

Proof. Let $\{\Omega_n\}$ be a sequence of bounded smooth domains in $\mathbf{R}^N$ as provided in the definition of the c-positivity of p.

Suppose (1) has a solution, say v is a solution. For each n, the problem

$$\begin{cases} div(|\nabla u|^{m-2}\nabla u) = p(x)f(u), & x \in \Omega_n \\ u(x) = \infty, & x \in \partial\Omega_n \end{cases} \quad (16)$$

has a solution(see [29]). For each positive integer n, let u_n be a solution of (16). Then by Lemma 2.1 it follows that

$$v(x) \leq u_{n+1}(x) \leq u_n(x), \quad x \in \Omega_n.$$

A standard procedure (for example, see [30]) can be used to show that

$$u(x) := \lim_{n \to \infty}u_n(x), \quad x \in \mathbf{R}^N,$$

is the desired solution of (3). For the converse, we let u_n be a solution of the problem

$$\begin{cases} div(|\nabla u|^{m-2}\nabla u) = p(x)f(u) + q(x)g(u), & x \in \Omega_n \\ u(x) = \infty, & x \in \partial\Omega_n \end{cases} \quad (17)$$

The existence of such a solution is demonstrated in Lemma 5.3 of Appendix. It easily follows that the sequence $\{u_n\}$ is a non-increasing sequence. Let

$$u(x) = \lim_{n \to \infty}u_n(x), \quad x \in \mathbf{R}^N.$$

A standard argument shows that u is a solution of the quasilinear equation in (17). Thus we need only show that u is nontrivial and that $u(x) \to \infty$ as $|x| \to \infty$. For this we consider the following function

$$\psi(t) = \int_{t}^{\infty}\frac{1}{h^{1/(m-1)}(s)}ds, \quad t > 0, \quad (18)$$

where $h(t) := f(t) + g(t)$. Obviously, (18) is finite for all $t > 0$. We also notice that

$$\psi'(t) = -\frac{1}{h^{1/(m-1)}(t)} < 0, \quad \psi''(t) = \frac{h'(t)}{(m-1)(h^{1/(m-1)}(t))^m} > 0$$

Now fix $\varepsilon > 0$, and let $v_n(x) = \psi(u_n(x) + \varepsilon)$, $x \in \Omega_n$. Note the sequence v_n is nondecreasing. Moreover, a simple computation shows that

$$-div(|\nabla v_n|^{m-2}\nabla v_n) = |\psi'(u_n + \varepsilon)|^{m-1}div(|\nabla u_n|^{m-2}\nabla u_n)$$

$$-(m-1)|\psi'(u_n + \varepsilon)|^{m-2}\psi''(u_n + \varepsilon)|\nabla u_n|^m$$

$$\leq |\psi'(u_n + \varepsilon)|^{m-1}(p(x)f(u_n) + q(x)g(u_n))$$

$$= \frac{p(x)f(u_n) + q(x)g(u_n)}{h(u_n + \varepsilon)} \leq p(x) + q(x)$$

We can also note that $v_n = 0$ on $\partial\Omega_n$. Let w be a solution of (15). Thus by Lemma 2.1 we see that $v_n \leq w$ on Ω_n for all n, letting $n \to \infty$, and then $\varepsilon \to 0$ we see that $\psi(u) \leq w$ on $\mathbf{R}^N$. Thus $\psi(u(x)) \to 0$ as $|x| \to \infty$, that is $u(x) \to \infty$ as $|x| \to \infty$.

4. Consequences and Related Results

We can obtain some consequences of the main theorems, and other results that are of independent interest.

Theorem 4.1. Let f and g be continuous, nondecreasing functions such that $f + g$ satisfies (9), and Suppose $p + q$ is nontrivial. If there is a solution to

$$\begin{cases} div(|\nabla u|^{m-2}\nabla u) \leq p(x)f(u) + q(x)g(u), & x \in R^N, N \geq 3 \\ u(x) \to \infty, & as \ |x| \to \infty \end{cases}$$

$$(19)$$

then $p + q$ satisfies (7).

Proof. Let u be a solution of (19). Let v be a solution of the initial value (11) with $\rho = (p+q)^*$, f replaced by $f+g$ and $\alpha = \alpha_0$ where α_0 is chosen such that $\alpha_0 > u(0)$. Since $f+g$ satisfies (9), we recall from Lemma 2.2 that v is defined on $[0,\infty)$. Then $w(x) = v(|x|)$ is a solution of

$$div(|\nabla w|^{m-2} \nabla w) = (p+q)^*(|x|)(f(w(x))+g(w(x))),$$

and hence $div(|\nabla w|^{m-2} \nabla w) \geq pf(w)+qg(w)$ on $\mathbf{R}^N$. Since $v'(r) > 0$ we see that $v(r) \to A$ as $r \to \infty$, for some $0 < A \leq \infty$. Assume that $A < \infty$ so that $w(x) \leq A$ for all $x \in \mathbf{R}^N$. Since $u(x) \to \infty$ as $|x| \to \infty$, we see that for some R, we have $w(x) \leq A \leq u(x), |x| \geq R$. Thus $w(x) \leq u(x)$ on $|x| = R$ and therefore by Lemma 2.1 we find that $w(x) \leq u(x)$ on $B(0,R)$. But this contradicts the choice that $w(0) > u(0)$. So we have $A = \infty$, then $w(x) \to \infty, as |x| \to \infty$. From Equation (11) we find

$$\begin{aligned} v'(r) &= \left(r^{1-N} \int_0^r t^{N-1}(p+q)^*(t)(f+g)(v(t))dt \right)^{1/(m-1)} \\ &\leq \left(r^{1-N}(f+g)(v(r))\int_0^r t^{N-1}(p+q)^*(t)dt \right)^{1/(m-1)} \end{aligned} \quad (20)$$

Dividing (20) through by $(f+g)^{1/(m-1)}(v(r))$ and integrating the resulting inequality on $(0,r)$ we have

$$\int_0^r \frac{v'(t)}{(f+g)^{1/(m-1)}(v(t))} dt$$
$$\leq \int_0^r \left(t^{1-N} \int_0^t s^{N-1}(p+q)^*(s)ds \right)^{1/(m-1)} dt$$

That is

$$\int_{\alpha_0}^{v(r)} \frac{1}{(f+g)^{1/(m-1)}(t)} dt$$
$$\leq \left(\frac{1}{N} \right)^{1/(m-1)} \int_0^r (t(p+q)^*(t))^{1/(m-1)} dt$$

Letting $r \to \infty$ and recalling that $f+g$ satisfies (9), the claim is proved.

As a consequence of Theorem 3.1 and Theorem 4.1 we also obtain the following corollaries.

Corollary 1. Suppose (2) and (9) hold for f. Further, let p satisfy (10). (3) admits a solution if and only if p satisfies (7).

Proof. If p satisfies (7) then Theorem 3.1, with $q(x) = 0$ shows that (3) has a solution. The converse follows from Theorem 4.1 on taking $q(x) \equiv 0$ again.

The next corollary provides sufficient conditions for the existence and nonexistence of solutions to (1) when both p and q satisfy (7).

Corollary 2. Suppose f and g satisfy (2) and p and q satisfy (7). If $f+g$ satisfies (9), then (1) has no solution. On the other hand, (1) admits a solution if $f+g$ satisfies (8) and the function $\xi(x) \equiv \min\{p(x), q(x)\}$ is c-positive on $\mathbf{R}^N$.

Proof. By way of contradiction, we can obtain the first statement from Theorem 4.1. Since $p+q$ satisfies (10) and the remark noted just before Theorem 3.2 shows that (15) admits a solution with $b = p+q$. Thus the second part of the corollary is an immediate consequence of Theorem 3.2.

5. Appendix

In this appendix we state and prove results that have been used in the proofs of the main results of the paper.

We start by proving the existence of a solution to the following Dirichlet problem on a bounded smooth domain Ω in $\mathbf{R}^N$.

$$\begin{cases} div(|\nabla u|^{m-2} \nabla u) = p(x)f(u)+q(x)g(u), & in\ \Omega, \\ u(x) = \varphi(x), & on\ \partial\Omega. \end{cases} \quad (21)$$

Lemma 5.1. Let $\Omega \subseteq \mathbf{R}^N$ be a smooth bounded domain and let f and g satisfy (2). Let $\phi \in C^2(\partial\Omega)$ be positive. If v is a positive super-solution of (21), then the problem (21) has a solution u such that $0 < u \leq v$ on Ω.

Proof. Let $\beta := \min_{x \in \partial\Omega} \phi(x)$. Obviously, $\beta > 0$. Now we set

$$\varphi(t) = \int_0^t (h(s))^{1/(m-1)} ds,$$

where $h(s) = f(s)+g(s)$ for all $s \geq 0$. An application of L'Hôpital's Rule shows that $\varphi(t) \leq t$ for all $0 < t < \varepsilon$ and some $\varepsilon > 0$. Without of generality we can suppose that $0 < \varphi(\varepsilon) \leq \beta$. Finally, let z be a solution of the Dirichlet problem

$$\begin{cases} div(|\nabla z|^{m-2} \nabla z) = p(x)+q(x), & in\ \Omega, \\ z(x) = \varepsilon, & on\ \partial\Omega. \end{cases}$$

Then the maximum principle shows that $0 < z(x) \leq \varepsilon$ on Ω, we define $w(x) := \varphi(z(x))$ for all $x \in \overline{\Omega}$, we note that $w(x) \leq z(x)$ for all $x \in \overline{\Omega}$. A simple computation shows that

$$\varphi'(t) = h^{1/(m-1)}(t) > 0, \quad \varphi''(t) = \frac{h'(t)}{(m-1)(h^{1/(m-1)}(t))^{m-2}} > 0$$

and

$$\begin{aligned} &div(|\nabla w|^{m-2} \nabla w) \\ &= |\varphi'|^{m-1} div(|\nabla z|^{m-2} \nabla z) + (m-1)|\varphi'|^{m-2} \varphi'' |\nabla z|^m \\ &\geq |\varphi'|^{m-1}(p(x)+q(x)) = (f(z)+g(z))(p(x)+q(x)) \\ &\geq (f(w)+g(w))(p(x)+q(x)) \geq p(x)f(w)+q(x)g(w) \end{aligned}$$

and $w(x) \leq \varphi(\varepsilon) \leq \beta \leq \phi(x)$ for $x \in \partial\Omega$. Thus w is a sub-solution of (21) and v is a super-solution of (21) such that $w \leq \varphi \leq v$ on $\partial\Omega$. By the maximum principle we note that $w \leq v$ on Ω. Thus by lemma 1 in [31] we

conclude that (21) has a solution u such that $w \le u \le v$ which is what we want to show.

The following lemma was used in the proof of Theorem 3.1.

Lemma 5.2. Let $a,b \in C([0,\infty),[0,\infty))$ and B be a ball in $\mathbf{R}^N$ centered at the origin. If f and g are nondecreasing on $[0,\infty)$, then given a positive constant δ, there exists a radial solution to the problem

$$\begin{cases} div(|\nabla u|^{m-2}\nabla u) = a(|x|)f(u)+b(|x|)g(u), & in \ B, \\ u(x) = \delta, & on \ \partial B, \end{cases} \tag{22}$$

Proof. Let $\{a_k\}$ and $\{b_k\}$ be decreasing sequences of Hölder continuous functions which converge uniformly on B to a and b respectively(See [32]). Then by Lemma 5.1, for each k there exists a nonnegative solution u_k of

$$\begin{cases} div(|\nabla u_k|^{m-2}\nabla u_k) \\ = a_k(|x|)f(u_k)+b_k(|x|)g(u_k), & in \ B, \\ u_k(x) = \delta, & on \ \partial B. \end{cases}$$

Since the sequence $\{a_k\}$ and $\{b_k\}$ are decreasing,it is easy to show that $\{u_k\}$ is increasing. Of course,it is also bounded above by δ. Thus it converges, and assume $u_k \to u$. Since u_k satisfies the integral equation

$$u_k(r) = u_k(0) +$$

$$\int_0^r \left(t^{1-N} \int_0^t s^{N-1} \left(a_k(s)f(u_k(s))+b_k(s)g(u_k(s)) \right)ds \right)^{1/(m-1)} dt$$

the function u will satisfy the integral equation

$$u(r) = u(0) +$$

$$\int_0^r \left(t^{1-N} \int_0^t s^{N-1} \left(a(s)f(u(s))+b(s)g(u(s)) \right)ds \right)^{1/(m-1)} dt.$$

Since $u_k(R) = \delta$ for each k, where R is the radius of the ball B, it is clear that $u(R) = \delta$. Thus u is a nonnegative solution of problem (22) on B as claimed.

Finally we state and prove a lemma that was used in the proof of Theorem 3.2.

Lemma 5.3. Let $\Omega \subseteq \mathbf{R}^N$ be smooth. Suppose f and g satisfy (2). If f satisfies (6) and p is c-positive on Ω, then the problem

$$\begin{cases} div(|\nabla u|^{m-2}\nabla u) = p(x)f(u)+q(x)g(u), & x \in \Omega, \\ u(x) = \infty, & x \in \partial\Omega, \end{cases} \tag{23}$$

has a solution. Similarly, if instead of requiring f to satisfy (6), we require only $f+g$ to satisfy (6), and require $\xi := \min\{p,q\}$ to to be c-positive on Ω, then (23) has a solution.

Proof. Since p is c-positive and f satisfies (6), let v be a large solution of $div(|\nabla v|^{m-2}\nabla v) = p(x)f(v)$ on Ω (see [29]). Now for each positive integer k, let w_k be a solution (See Lemma 5.1) of

$$\begin{cases} div(|\nabla w|^{m-2}\nabla w) \\ = (p(x)+q(x))(f(w)+g(w)), & x \in \Omega, \\ w(x) = k, & x \in \partial\Omega, \end{cases}$$

By Lemma 2.1 we see that

$$w_k(x) \le w_{k+1}(x) \le v(x), \ x \in \Omega$$

If $w(x) = \lim_{k\to\infty} w_k(x)$, then by a standard procedure we conclude that w is a solution of $div(|\nabla w|^{m-2}\nabla w) = (p(x)+q(x))(f(w)+g(w))$ on Ω such that $w \le v$. Since w is a sub-solution, and v is a super-solution of the differential equation in (23), we conclude that (23) has a solution u with $w \le u \le v$ (See [31]).

Similarly.We can obtain the second part by defining v in this case as a large solution of $div(|\nabla v|^{m-2}\nabla v) = \xi(x)(f(v)+g(v))$ on Ω and the argument is as the same as the previous process.

6. References

[1] G. Astrita and G. Marruci, "Principles of Non-Newtonian Fluid Mechanics," McGraw-Hill, 1974.

[2] L. k. Martinson and K. B. Pavlov, "Unsteady Shear Flows of a Conducting Fluid with a Rheological Power Law," *Magnitnaya Gidrodinamika*, Vol. 7, No. 2, 1971, pp. 50-58.

[3] J. R. Esteban and J. L. Vazquez, "On the Equation of Turbulent Filteration in One-Dimensional Porous Media," *Non-Linear Analysis archive*, Vol. 10, No. 3, 1982, pp. 1303-1325.

[4] A. S. Kalashnikov, "On a Nonlinear Equation Appearing in the Theory of Nonstationary Filtration," Trudy Seminara I.G. Petrovski, Russian, 1978.

[5] S. L. Phhozaev, "The Dirichlet Problem for the Equation $\Delta u = u^2$," *Soviet mathematics-Doklady*, Vol. 1, No. 2, 1960, pp. 1143-1146.

[6] A. C. Lazer and P. J. Mckenna, "On a Problem of Bieberbach and Rademacher," *Non-Linear Analysis archive*, Vol. 21, No. 5, 1993, pp. 327-335.

[7] K.-S. Cheng and W.-M. Ni, "On the Structure of the Conformal Scalar Curvature Equation on $\mathbf{R}^N$," *Indiana University Mathematic Journal*, Vol. 41, No. 1, 1992, pp. 261-278.

[8] V. Anuradha, C. Brown and R. Shivaji, "Explosive Nonnegative Solutions to Two Point Boundary Value Problems," *Non-Linear Analysis archive*, Vol. 26, No. 3, 1996, pp. 613-630.

[9] S.-H. Wang, "Existence and Multiplicity of Boundary Blow-Up Nonnegative Solutions to Two-Point Boundary Value Problems," *Non-Linear Analysis archive*, Vol. 42, No. 1, 2000, pp. 139-162.

[10] G. Diaz and R. Letelier, "Explosive Solutions of Qua-

silinear Elliptic Equations: Existence and Uniqueness," *Non-Linear Analysis archive*, Vol. 20, No. 1, 1993, pp. 97-125.

[11] A. C. Lazer and P. J. McKenna, "On a Singular Nonlinear Elliptic Boundary-Value Problem," *Proceedings of American Mathematic Society*, Vol. 111, No. 3, 1991, pp. 721-730.

[12] A. C. Lazer and P. J. McKenna, "On Singular Boundary Value Problems for the Monge-Ampere Operator," *Journal of Mathematical Analysis Applications*, Vol. 197, No. 2, 1996, pp. 341-362.

[13] L. Bieberbach, "$\Delta u = e^u$ und die automorphen Funktionen," *Mathematische Annalen*, Vol. 77, No. 1, 1916, pp. 173-212.

[14] M. Marcus and L. Veron, "Uniqueness of Solutions with Blow-Up at the Boundary for a Class of Nonlinear Elliptic Equation," *Comptes rendus de l'Académie des sciences*, Vol. 317, No. 2, 1993, pp. 559-563.

[15] S. L. Pohozaev, "The Dirichlet Problem for the Equation $\Delta u = u^2$," *Soviet mathematics-Doklady*, Vol. 1, No. 2, 1960, pp. 1143-1146.

[16] M. R. Posteraro, "On the Solutions of the Equation $\Delta u = e^u$ Blowing up on the Boundary," *Comptes rendus de l'Académie des sciences*, Vol. 322, No. 2, 1996, pp. 445-450.

[17] H. Rademacher, "Einige Besondere Problem Partieller Differentialgleichungen," In: *Die Differential-und Integralgleichungen, der Mechanik und Physikl*, Rosenberg, New York, 1943, pp. 838-845.

[18] J. B. Keller, "On Solutions of $\Delta u = f(u)$," *Communications on Pure and Applied Mathematics*, Vol. 10, No. 4, 1957, pp. 503-510.

[19] V. A. Kondrat'ev and V. A. Nikishken, "Asymptotics, near the Boundary, of a Singular Boundary-Value Problem for a Semilinear Elliptic Equation," *Differential Equations*, Vol. 26, No. 1, 1990, pp. 345-348.

[20] C. Loewner and L. Nirenberg, "Partial Differential Equations Invariant under Conformal or Projective Transformations," In: *Contributions to Analysis* (*A Collection of Paper Dedicated to Lipman Bers*), Academic Press, New York, 1974, pp. 245-272.

[21] E. B. Dynkin, "Superprocesses and Partial Differential Equations," *Annals of Probability*, Vol. 21, No. 3, 1993, pp. 1185-1262.

[22] E. B. Dynkin and S. E. Kuznetsov, "Superdiffusions and Removable Singularities for Quasilinear Partial Differential Equations," *Communications on Pure and Applied Mathematics*, Vol. 49, No. 2, 1996, pp. 125-176.

[23] A. V. Lair, "Large Solutions of Mixed Sublinear/Superlinear Elliptic Equations," *Journal of Mathematical Analysis Applications*, Vol. 346, No. 1, 2008, pp. 99-106.

[24] A. V. Lair and A. Mohammed, "Entire Large Solutions of Semilinear Elliptic Equations of Mixed Type," *Communications on Pure and Applied Analysis*, Vol. 8, No. 5, 2009, pp. 1607-1618.

[25] Q. S. Lu, Z. D. Yang and E. H. Twizell, "Existence of Entire Explosive Positive Solutions of Quasi-linear Elliptic Equations," *Applied Mathematics and Computation*, Vol. 148, No. 2, 2004, pp. 359-372.

[26] Z. D. Yang, "Existence of Explosive Positive Solutions of Quasilinear Elliptic Equations," *Applied Mathematics and Computation*, Vol. 177, No. 2, 2006, pp. 581-588.

[27] J. L. Yuan and Z. D. Yang, "Existence of Large Solutions for a Class of Quasilinear Elliptic Equations," *Applied Mathematics and Computation*, Vol. 201, No. 2, 2008, pp. 852-858.

[28] A. V. Lair, "Large Solutions of Semilinear Elliptic Equations under the Keller-Osserman Condition," *Journal of Mathematical Analysis Applications*, Vol. 328, No. 2, 2007, pp. 1247-1254.

[29] Z. D. Yang, B. Xu and M. Z. Wu, "Existence of Positive Boundary Blow-up Solutions for Quasilinear Elliptic Equations via Sub and Supersolutions," *Applied Mathematics and Computation*, Vol. 188, No. 1, 2007, pp. 492-498.

[30] Z. D. Yang, "Existence of Entire Explosive Positive Radial Solutions for a Class of Quasilinear Elliptic Systems," *Journal of Mathematical Analysis Applications*, Vol. 288, No. 2, 2003, pp. 768-783.

[31] H. H. Yin and Z. D. Yang, "New Results on the Existence of Bounded Positive Entire Solutions for Quasilinear Elliptic Systems," *Applied Mathematics and Computation*, Vol. 190, No. 1, 2007, pp. 441-448.

[32] A. V. Lair and A. W. Shaker, "Classical and Weak Solutions of a Singular Semilinear Elliptic Problem," *Journal of Mathematical Analysis Applications*, Vol. 211, No. 2, 1997, pp. 371-385.

Scientific
Research

On the Design of Optimal Feedback Control for Systems of Second Order

Alexander M. Formalskii

Institute of Mechanics of Moscow Lomonosov State University, Michurinskii prospect, Moscow, Russia
E-mail: formal@imec.msu.ru

Abstract

A difficult but important problem in optimal control theory is the design of an optimal feedback control, *i.e.*, the design of an optimal control as function of the phase (state) coordinates [1,2]. This problem can be solved not often. We study here the autonomous nonlinear system of second order in general form. The constraints imposed on the control input can depend on the phase (state) coordinates of the system. The goal of the control is to maximize or minimize one phase coordinate of the considered system while other takes a prescribed in advance value. In the literature, optimal control problems for the systems of second order are most frequently associated with driving both phase coordinates to a prescribed in advance state. In this statement of the problem, the optimal control feedback can be designed only for special kind of systems. In our statement of the problem, an optimal control can be designed as function of the state coordinates for more general kind of the systems. The problem of maximization or minimization of the swing amplitude is considered explicitly as an example. Simulation results are presented.

1. Mathematical Model of the Considered System

Let the motion of the studied object under control be governed by a system of two nonlinear autonomous differential equations of the form

$$\dot{x} = f_1(x, y, u), \quad \dot{y} = f_2(x, y, u), \qquad (1)$$

the dot denotes, as usual, derivative with respect to time.

For example, a controllable mechanical system with one degree of freedom is described by similar differential equations. In this case, x is positional coordinate and

$$f_1(x, y, u) = y \qquad (2)$$

is the velocity of the object (linear or angular), function $f_2(x, y, u)$ is the generalized force divided by the object's mass or moment of inertia.

Let for each piecewise continuous vector function $u(t)$, system (1) with initial conditions from some region of the phase plane (x, y) has a unique solution $x(t)$, $y(t)$. We assume that the control parameter u belongs to a given set $U(x, y)$ depending on the state coordinates x and y. In other words, a vector piecewise con-

tinuous function $u(t)$ is assumed to be an admissible control, if

$$u(t) \in U(x(t), y(t)). \qquad (3)$$

Here $x(t)$, $y(t)$ is the solution of Equations (1) with $u = u(t)$. If set $U(x, y)$ depends on state coordinates x, y, then condition (3) can be checked for a given piecewise continuous control function $u(t)$, in general, only by finding the solution of the system (1) with this control.

Assume in what follows that function $f_1(x, y, u)$ does not vanish. To be definite, let

$$f_1(x, y, u) > 0. \qquad (4)$$

Under condition (4), the coordinate x can only increase with time. If (1) is a mathematical model of a mechanical system with one degree of freedom, then equality (2) takes place and inequality (4) holds in upper half of the phase plane (x, y).

We rewrite system (1) in the form of a first-order equation

$$\frac{dy}{dx} = \frac{f_2(x, y, u)}{f_1(x, y, u)} = f(x, y, u). \qquad (5)$$

Let

$$x(0) = x_0, \ y(0) = y_0 \tag{6}$$

be initial conditions for system (1) or Equation (5). To be definite, assume that $y_0 > 0$.

We do not formulate here in the first Section all conditions on the system (1), on the set $U(x, y)$ (see relation (3)). It is difficult to specify in advance all these conditions. We formulate new assumptions during the problem consideration as need arises.

2. Sets of Reachability

Assume that, in the phase plane (x, y), every trajectory $y(x)$ that starts from point (6) and corresponds to an admissible control function $u(t)$, intersects the axis $Y = 0$ at some finite time t and for some finite coordinate x. Note that time t and coordinate x have their own values for each admissible control function $u(t)$. Consider the set of all possible admissible control functions $u(t)$ and the set of corresponding trajectories $y(x)$, obtained under these controls. More precisely, consider only the portions of these trajectories that start from point (6) and terminate on the axis of abscissas $Y = 0$. The collection of these curves covers a set of points that form a reachable set [3] or so-called integral funnel [4,5]. This set of reachability D is schematically shown in **Figure 1**.

3. Boundaries of Reachable Set

Let us consider control that maximizes derivative $\dfrac{dy}{dx}$ over variable u at the point (x, y). This control maximizes the function $f(x, y, u)$ over argument u on the right-hand side of Equation (5) and it has the form

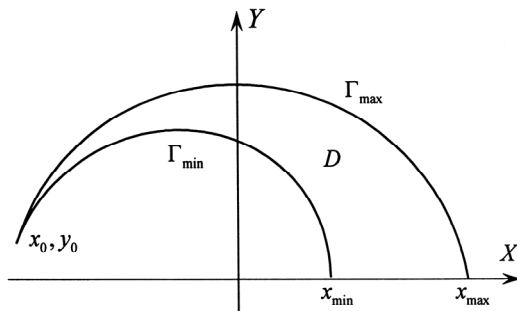$$u = u_{\max}(x, y) = \arg\left[\max_{u \in U(x, y)} f(x, y, u) \right]. \tag{7}$$

Figure 1. Reachabe set D.

We assume here function $f(x, y, u)$ and set $U(x, y)$ such that the maximum in (7) exists and is unique in each point of the phase plane in some domain including the reachable set D. We assume also that the solution to system (1) with initial conditions (6) under control (7) yields a piecewise continuous function $u(t)$, *i.e.*, an admissible control function. Let $y = y_{\max}(x)$ be the solution to the equation

$$\frac{dy}{dx} = f\left[x, y, u_{\max}(x, y) \right] \tag{8}$$

with initial conditions (6). Denote by $\Gamma_{\max}$ the part of the trajectory $y = y_{\max}(x)$ for $x_0 \le x \le x_{\max}$, where $x_{\max}$ is the first value of argument x, at which function $y = y_{\max}(x)$ vanishes $(y_{\max}(x) = 0)$. Now we will show that the curve $\Gamma_{\max}$ is the upper boundary of reachable set D (see **Figure 1**).

Given any control function $u^*(x, y) \ne u_{\max}(x, y)$, assume that the trajectory of Equation (5) starting from some point $(x, y) \in \Gamma_{\max}$, lies above by curve $\Gamma_{\max}$. Then, at this point, we have the inequality

$$f\left[x, y, u^*(x, y) \right] \ge f\left[x, y, u_{\max}(x, y) \right], \tag{9}$$

or the solution $y = y_{\max}(x)$ of Equation (8) is not unique. However, inequality (9) contradicts condition (7), while the solution $y = y_{\max}(x)$ of Equation (8) starting at point (6) is unique by assumption.

Now consider control function that minimizes derivative $\dfrac{dy}{dx}$ over parameter u at the point (x, y), *i.e.*, minimizes the function $f(x, y, u)$ over argument u on the right-hand side of Equation (5):

$$u = u_{\min}(x, y) = \arg\left[\min_{u \in U(x, y)} f(x, y, u) \right]. \tag{10}$$

We assume here function $f(x, y, u)$ and set $U(x, y)$ such that the minimum in (10) exists and is unique in the phase plane in some domain including the reachable set D. Let the solution to system (1) with initial conditions (6) under control (10) yields a piecewise continuous function $u(t)$, *i.e.*, an admissible control function.

Let $y = y_{\min}(x)$ be the solution to the equation

$$\frac{dy}{dx} = f\left[x, y, u_{\min}(x, y) \right] \tag{11}$$

with initial conditions (6). Denote by $\Gamma_{\min}$ the part of the trajectory $y = y_{\min}(x)$ for $x_0 \le x \le x_{\min}$, where $x_{\min}$ is the first value of argument x, at which function $y = y_{\min}(x)$ vanishes $(y_{\min}(x) = 0)$.

Applying the considerations similar to that used for the curve $\Gamma_{\max}$, we can prove that curve $\Gamma_{\min}$ is the lower boundary of reachable set D.

4. Statement of the Problem and its Solution

The problem is to find an admissible control, under which the coordinate x reaches its maximum when the coordinate y vanishes at first time (from the beginning of the motion). This maximization problem can be symbolically written as

$$\max_{u \in U(x,y)} [x] \text{ at } y = 0. \tag{12}$$

It should be accented that coordinate x has to be maximized not at a prescribed time but at the time when coordinate y vanishes. When Equations (1) describe the motion of a mechanical system with one degree of freedom and equality (2) holds, the condition $y = 0$ means that the velocity of motion vanishes. In this case, the goal is to maximize the deviation of the x-coordinate from the initial position by the time when the velocity of motion vanishes.

Along with the formulated above problem, we also consider problem to find an admissible control, under which the coordinate x reaches its minimum when the coordinate y vanishes at first time (from the beginning of the motion):

$$\min_{u \in U(x,y)} [x] \text{ at } y = 0. \tag{13}$$

A maximization problem of type (12) was considered, for example, in paper [6] and in other works of the same authors. In these works, the conditions for the absolute stability of bilinear systems were found, by constructing a control that maximally "swings" the system.

The analysis performed in Section 3 implies that in the reachable set D, there is no a trajectory with so large x-coordinate as $x_{\max}$ (see **Figure 1**). Consequently the maximum value of x-coordinate is equal to $x_{\max}$ and the control $u = u_{\max}(x,y)$ (see formula (7)) solves problem (12).

The control $u = u_{\min}(x,y)$ (see formula (10)) solves problem (13), and the minimal value of x-coordinate is equal to $x_{\min}$ (see **Figure 1**).

5. More General Case

More general problems than those discussed above are the maximization or minimization of the coordinate x at the time when coordinate y first takes a prescribed in advance value $\overline{y}$, which may be nonzero. If $\overline{y} < y_0$, then, as before, the control $u = u_{\max}(x,y)$ (see formula (7)) is optimal for the maximization problem, while the control $u = u_{\min}(x,y)$ (see formula (10)) is optimal for the minimization problem. If $\overline{y} > y_0$ (see **Figure 2**), then the maximum of the coordinate x is reached under
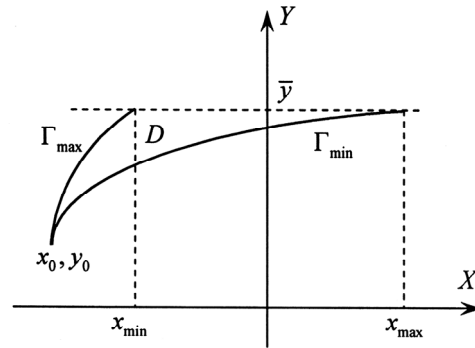

Figure 2. Reachabe set D.

the control $u = u_{\min}(x,y)$, while the minimum of the coordinate x is reached under the control $u = u_{\max}(x,y)$. This is explained by the fact that, for $\overline{y} > y_0$, the upper boundary $\Gamma_{\max}$ of the reachable set D intersects the line $y = \overline{y}$ at a smaller value of coordinate x, than the lower boundary $\Gamma_{\min}$.

Formulas (7) or (10) can be considered for the formulated above problems as *a local maximum or minimum principle*.

In the next section, the obtained above results are illustrated by constructing an optimal control of the motion of the swing.

6. Maximization and Minimization of the Swing Amplitude

As a model of a swing with a human on it, we consider a physical pendulum of mass μ with a point particle of mass M moving along it (see **Figure 3**).

In **Figure 3**, x is the deflection angle of the pendulum from the vertical line (we consider that $-\pi < x < \pi$), u denote the distance OM between points M and O ($u = OM$). Assume that distance u is a control parameter. Unlike the general case, here this control parameter u is a scalar. The distance u can vary within bounded limits:

$$u_0 \le u \le u_1 \ (u_0, u_1 = const, \ u_0 < u_1). \tag{14}$$

Let J denote the moment of inertia of the pendulum (without point participle M) relative to the point of suspension O, ρ denote the distance from the suspension joint O to the pendulum's center of mass C ($\rho = OC$), g is the gravity acceleration.

According to the principle of moment of angular momentum relative to joint O, the nonlinear equation of motion of the swing has the form [7]:

$$\frac{d}{dt}\left[\left(J + Mu^2\right)\frac{dx}{dt}\right] = -\left(Mu + \mu\rho\right)g\sin x - c\frac{dx}{dt}. \tag{15}$$

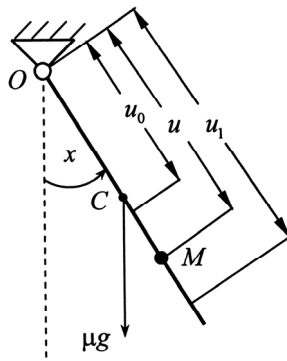Here $\left(J + Mu^2\right)\dot{x}$ is the angular momentum of the

Figure 3. The scheme of the swing.

system relative to suspension point O, c is the coefficient of viscous resistance (for example, viscous friction in the suspension joint O).

Let y denote the angular momentum $\left(J + Mu^2\right)\dot{x}$. Then the second-order Equation (15) can be rewritten as a system of two first-order equations of form (1):

$$\dot{x} = \frac{y}{J + Mu^2},$$
$$\dot{y} = -\left(Mu + \mu\rho\right)g\sin x - \frac{cy}{J + Mu^2}. \tag{16}$$

Let the initial state of system (16) be specified as

$$x(0) < 0, \; y(0) = 0. \tag{17}$$

The problem is to find a law of variation of the distance u subject to inequalities (14) at which the deflection of the angle x is maximal at the instant θ when angular momentum y (and, hence, the velocity $\dot{x}$) vanishes $\left(y(\theta) = 0, \; \dot{x}(\theta) = 0\right)$ first from the start of the motion. In other words, the goal is to maximize the deviation of the swing from the vertical line (amplitude of the swing) at the end of the first half-period of its oscillations. We consider initial value $x(0)$ of angle x sufficiently close to zero and assume that for this angle $x(0)$ and each admissible control $u(t)$ the corresponding instant θ exists when $y(\theta) = 0$.

It follows from second Equation (16) that $y > 0$ (*i.e.*, $\dot{x} > 0$), if $0 < t < \theta$. Then on the time interval $0 < t < \theta$, system (16) can be rewritten in a form similar to (5), namely

$$\frac{dy}{dx} = -c - \frac{\left(Mu + \mu\rho\right)\left(J + Mu^2\right)g\sin x}{y}. \tag{18}$$

According to the above results, maximizing the right-hand side of Equation (18) over argument u on interval (14) yields an optimal control law for the swing on a half-period of oscillations for which $y > 0$

$$u = \begin{cases} u_1, & \text{if } x < 0 \\ u_0, & \text{if } x > 0. \end{cases}$$

This control law means that distance $u = OM$ is maximal as possible when $x < 0$ and minimal as possible when $x > 0$.

The next (second) half-period of oscillations, for which $y < 0$ (*i.e.*, $\dot{x} < 0$), can be considered by analogy with the first half-period. As a result, we conclude that the optimal control of the swing on each half-period is described by the relations

$$u = \begin{cases} u_1, & \text{if } x\dot{x} < 0 \\ u_0, & \text{if } x\dot{x} > 0. \end{cases} \tag{19}$$

Figure 4 shows in the phase plane $(x, \dot{x})$ the synthesis picture of optimal rocking control $u(x, \dot{x})$ (19). Under control (19), the point particle M instantaneously moves up to the stop when the swing goes through the lower position and moves down to the stop when the swing maximally deviates from the vertical, *i.e.*, when its angular velocity $\dot{x}$ vanishes. Thus, the problem of optimal rocking of the swing is solved.

The control given by formula (19) was considered in the book [8] but without any discussion of its optimality.

Now let us consider the problem of optimal damping of the swing at the end of each half-period of its oscillations. To design the optimal damping control we have to solve the problem (13) for the right-hand side of Equation (18). After solving this problem we conclude that the deviation of the swing from the vertical at the end of each oscillation half-period is minimal under the control:

$$u = \begin{cases} u_1, & \text{if } x\dot{x} > 0 \\ u_0, & \text{if } x\dot{x} < 0. \end{cases} \tag{20}$$

In **Figure 5**, the synthesis picture of optimal damping
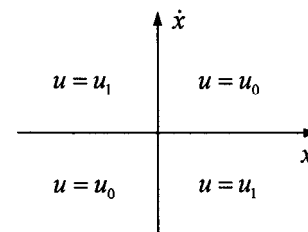

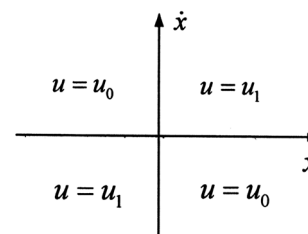**Figure 4. Design of the optimal rocking feedback control.**

Figure 5. Design of the optimal damping feedback control.

control $u(x, \dot{x})$ (20) is shown in the phase plane $(x, \dot{x})$. Under control (20), the point particle M instantaneously moves down to the stop when the swing goes through the lower position and moves up to the stop when the swing maximally deviates from the vertical, *i.e.*, when its angular velocity $\dot{x}$ vanishes.

So, the feedback control law (20) for the optimal damping of the swing is contrary to the feedback control law (19) for the optimal rocking of the swing.

7. Simulation of the Optimal Swing Motion

Consider Equations (16) with the following numerical parameters:

$$\mu = 5 \, kg, \ J = 26.67 \, kg \cdot m^2, \ M = 70 \, kg, \ u_0 = 3 \, m, \quad (21)$$
$$u_1 = 3.75 \, m, \ \rho = 2 \, m, \ c = 2 \, N \cdot m \cdot s.$$

Here we assume that pendulum is a homogeneous beam of the length $2\rho = 4 \, m$, and consequently

$$J = \frac{1}{3}\mu(2\rho)^2 = \frac{4}{3}\mu\rho^2.$$

In **Figure 6**, the graphs of angle x, of angular velocity $\dot{x}$, of control u as functions of time are shown. These functions are obtained solving equations of motion (16) under optimal *rocking* feedback control (19) with parameters (21) and initial conditions $x(0) = -0.1$, $y(0) = 0$.

Figure 6 shows that under control $u(x, \dot{x})$ (19) the amplitude of the swing increases. The *relay* control function $u(t)$ instantaneously switches from value u_1 to value u_0 when *angle x becomes zero* and switches from value u_0 back to value u_1 when the *velocity $\dot{x}$ becomes zero*. Between the shift points, control parameter $u = const$.

Figure 7 shows the phase portrait of the rocking motion in the plane $(x, \dot{x})$.

In **Figures 6** and **7**, we observe the jumps of angular velocity $\dot{x}$ at the instants when control $u(t)$ switches from value u_1 to value u_0. At these instants, the moment of inertia $J + Mu^2$ of the swing together with the point particle M (with a human) *decreases* and the angular velocity $\dot{x}$ *increases* due to the conservation of the angular momentum $y = (J + Mu^2)\dot{x}$, which is not zero at these times.

In **Figure 8**, the graphs of angle x, of angular velocity $\dot{x}$, of control u as functions of time are shown. These functions are obtained solving equations of motion (16) under optimal *damping* feedback control (20) with parameters (21) and initial conditions $x(0) = -\pi/2 \approx -1.57$, $y(0) = 0$.

Figure 8 shows that under control $u(x, \dot{x})$ (19) the amplitude of the swing decreases. The *relay* control function $u(t)$ instantaneously switches from value u_1 to

Figure 6. Graphs of angle $x(t)$, angular velocity $\dot{x}(t)$ and function $u(t)$ for rocking feedback control.

Figure 7. Phase portrait in the plane $(x, \dot{x})$ for rocking feedback control.
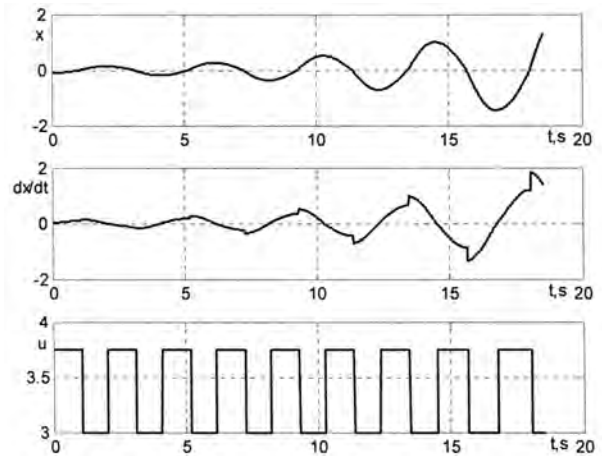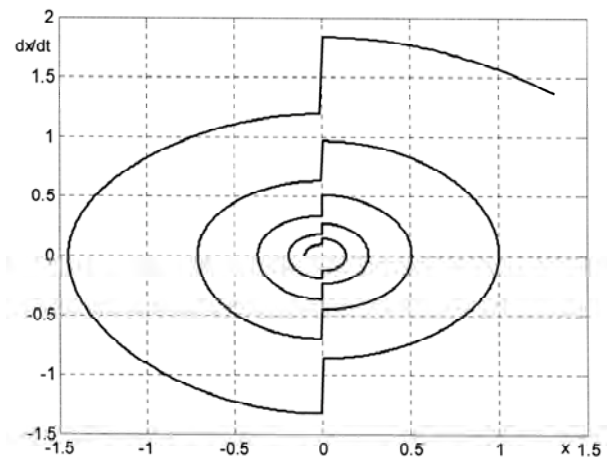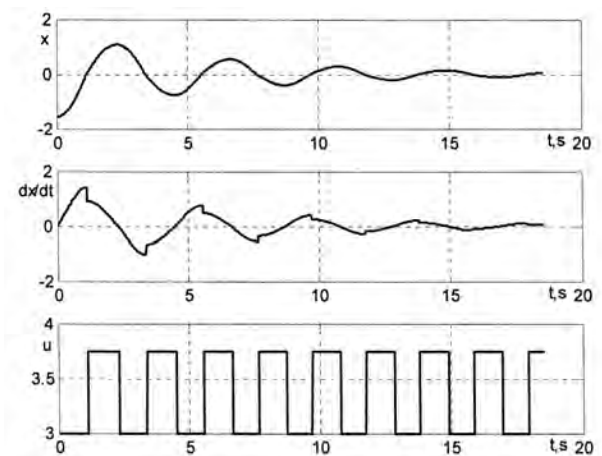

Figure 8. Graphs of angle $x(t)$, angular velocity $\dot{x}(t)$ and function $u(t)$ for damping feedback control.
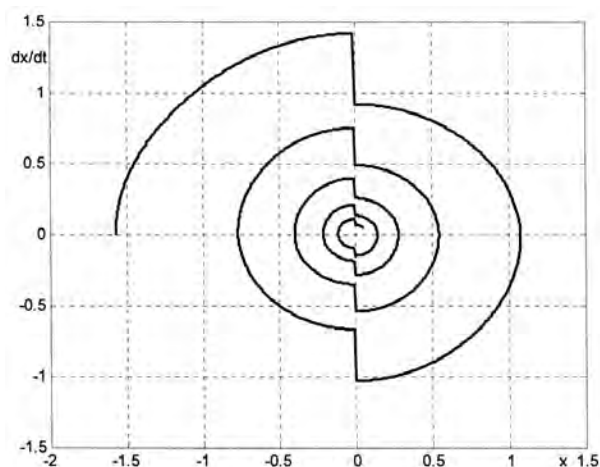
Figure 9. Phase portrait in the plane $(x, \dot{x})$ **for damping feedback control.**

value u_0 when the *velocity $\dot{x}$ becomes zero*, and switches in the opposite direction when *angle x becomes zero*. Between the shift points, the distance u remains constant.

Figure 9 shows the phase portrait of the damping motion in the plane $(x, \dot{x})$.

In **Figures 8** and **9**, we observe the jumps of angular velocity $\dot{x}$ at the times when control $u(t)$ switches from value u_0 to value u_1. At these times, the moment of inertia $J + Mu^2$ of the swing together with the point particle M (with a human) *increases* instantaneously and the angular velocity $\dot{x}$ *decreases* also instantaneously due to the conservation of the angular momentum $y = \left(J + Mu^2\right)\dot{x}$, which is not zero.

8. Acknowledgements

9. References

[1]   L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko, "The Mathematical Theory of Optimal Processes," Wiley Interscience, New York, 1962.

[2]   V. G. Boltyanskii, "Mathematical Methods of Optimal Control," Holt, Rinehart & Winston, 1971.

[3]   A. M. Formalskii, "Controllability and Stability of Systems with Restricted Control Resources," Nauka, Moscow, 1974 (in Russian).

[4]   I. A. Sultanov, "Studying the Control Processes Obeying Equations with Underdefinite Parameters," *Automation and Remote Control*, No. 10, October 1980, pp. 30-41.

[5]   A. G. Butkovskiy, "Phase Portrait of Control Dynamical Systems," Kluwer, 1991.

[6]   V. V. Alexandrov and V. N. Jermolenko, "On the Absolute Stability of Second-Order Systems," *Bulletin of Moscow University*, *Series* 1, *Mathematics and Mechanics*, No. 5, October 1972, pp. 102-108.

[7]   E. K. Lavrovskii and A. M. Formalskii, "Optimal Control of the Pumping and Damping of a Swing," *Journal of Applied Mathematics and Mechanics*, Vol. 57, No. 2, April 1993, pp. 311-320.

[8]   K. Magnus, "Schwingungen Eine einfurung in die theoretische behandlung von schwingungsproblemen," D. G. Teubner Stuttgart, 1976.

AM

Scientific
Research

New Periodic Solitary Wave Solutions for a Variable-Coefficient Gardner Equation from Fluid Dynamics and Plasma Physics

Mohamed Aly Abdou

*Theoretical Research Group, Physics Department, Faculty of Science, Mansoura University,
Mansoura, Egypt*
E-mail: *m_abdou_eg@yahoo.com*
Received May 6, 2010; *revised August* 17, 2010; *accepted August* 20, 2010

Abstract

The Gardner equation with a variable-coefficient from fluid dynamics and plasma physics is investigated. Different kinds of solutions including breather-type soliton and two soliton solutions are obtained using bilinear method and extended homoclinic test approach. The proposed method can also be applied to solve other types of higher dimensional integrable and non-integrable systems.

1. Introduction

In nonlinear science, many important phenomena in various fields can be describe by the nonlinear evolution equations. Seeking exact solutions of nonlinear partial differential equations is of great significance as it appears that these (NLPDEs) are mathematical models of complex physics phenomena arising in physics, mechanics, biology, chemistry and engineers. In order to help engineers and physicists to better understand the mechanism that governs these physical models or to better provide knowledge to the physical problem and possible applications, a vast variety of the powerful and direct methods have been derived. Various powerful methods for obtaining explicit travelling solitary wave solutions to nonlinear equations have proposed such as [1-8].

One of the most exciting advances of nonlinear science and theoretical physics has been a development of methods to look for exact solutions for nonlinear partial differential equations. A search of directly seeking for exactly solutions of nonlinear equations has been more interest in recent years because of the availability of symbloic computation Mathematica or Maple. These computer systems allow us to perform some complicated and tedious algebraic and differential calculations on a computer.

Much attention has been paid to the variable coeffi-

cients nonlinear equation which can describe many nonlinear phenomena more realistically than their constant coefficient ones [9]. The Gardner equation, or extended KdV equation can describe various interesting physics phenomens, such as the internal waves in a stratified Ocean [10], the long wave propagation in an inhomogeneous two-layer shallow liquid [11] and ion acoustic waves in plasma with negative ion [12], we consider a generalized variable-coefficient Gardner equation [13]

$$\phi_t + f(t)\phi_{xxx} + g(t)\phi\phi_x + h(t)\phi^2\phi_x + r(t)\phi_x + \tau(t)\phi = 0, \quad (1)$$

where $\phi(x,t)$ is a function of x and t. The coefficients $f(t)$, $g(t)$, $h(t)$, $r(t)$ and $\tau(t)$ are differential functions of t. Equaion (1) is not completely integrable in the sense of the inverse scattering scheme it contains some important special cases:

In case of $h(t)=0$, $r(t)=0$ and $\tau(t)=0$, Equation (1) reduces to

$$\phi_t + f(t)\phi_{xxx} + g(t)\phi\phi_x = 0, \quad (2)$$

and

$$f(t) = g(t)\left[a + b\int f(t)dt \right]. \quad (3)$$

Equation (2) possesses the Painleve property [14,15]. The Bäcklund transformation, Lax pair, similarity reduction and special analytic solution of Equation (2) have been obtained [16-20].

For $g(t) = -6a(t)$, $h(t) = -6r$ and $r(t) = \tau(t) = 0$,

Equation (1) reduces to

$$\phi_t - 6a(t)\phi\phi_x - 6r\phi^2\phi_x + f(t)\phi_{xxx} = 0, \quad (4)$$

which describe strong and weak interactions of different mode internal solitary waves, etc. When $f(t) = r$, $g(t) = 6\alpha$, $h(t) = 6\beta$, $r(t) = \tau(t) = 0$, Equation (1) becomes the constant-coefficient Gardner equation

$$\phi_t + 6\alpha\phi\phi_x + 6\beta\phi^2\phi_x + r\phi_{xxx} = 0, \quad (5)$$

where r, β and α are constants. It is widely applied to physics and quantum fields, such as solid state physics, plasma physics, fluid dynamics and quantum field theory.

When $g(t) = 6$, $f(t) = 1$, $h(t) = r(t) = 0$, Equation (1) reduces to constant coefficient KdV equation

$$\phi_t + 6\phi\phi_x + \phi_{xxx} + \tau(t)\phi = 0, \quad (6)$$

which possesses the Painleve property. If $\tau(t) = 0$ or $\tau(t) = \dfrac{1}{2(t - t_0)}$, it corresponds to the well known cylindrical KdV equation.

The structure of this paper is organized as follows; In Section 2, with symbolic computation, the bilinear form of Equation (1) are obtained. In order to illustrate the proposed method, we consider for a variable-coefficient Gardner equation from fluid dynamics and plasma physics and new periodic wave solutions are obtained which included periodic two solitary solution, doubly periodic solitary solution. Finally, conclusion and discussion are given in Section 3.

2. Bilinear Form of the Gardner Equation with Variable Coefficients

Making use the dependent variable transformation as

$$\phi(x,t) = k(t)\frac{\partial}{\partial x}w(x,t), \quad (7)$$

into Equation (1) and integrating once with respect to x, admits to [13]

$$k'(t)w + k(t)w_t + f(t)k(t)w_{xxx} + \frac{1}{2}g(t)k^2(t)w_x^2 + \frac{1}{3}h(t)k^3(t)w_x^3 + r(t)k(t)w_x + \tau(t)k(t)w = 0 \quad , \quad (8)$$

with the integration constant to zero. Then introducing the transformation

$$w(x,t) = \arctan\left[\frac{v(x,t)}{u(x,t)}\right], \quad (9)$$

where $u(x,t)$ and $v(x,t)$ are differential functions of x and t into (8) yields [13]

$$[k'(t) + \tau(t)k(t)]\arctan\left[\frac{v(x,t)}{u(x,t)}\right] + k(t)\frac{D_t v.u}{v^2 + u^2}$$

$$+ f(t)k(t)\left[\frac{D_x^3 v.u}{v^2 + u^2} - 3\frac{D_x v.u[D_x^2(u.u + v.v)]}{(v^2 + u^2)^2}\right.$$

$$\left. - 8(\frac{D_x v.u}{v^2 + u^2})^3\right] + \frac{1}{2}g(t)k^2(t)\left(\frac{D_x v.u}{v^2 + u^2}\right)^2 \quad (10)$$

$$+ \frac{1}{3}h(t)k^3(t)\left(\frac{D_x v.u}{v^2 + u^2}\right)^3 + r(t)k(t)\left(\frac{D_x v.u}{v^2 + u^2}\right) = 0,$$

where the prime denotes the derivative with respect to t, and D_t, D_x, D_x^2 and D_x^3 are the bilinear derivative operators [7] defined as

$$D_x^m D_t^n f(x,t).g(x,t)$$
$$= \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^m\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^n[f(x,t)g(x',t')]_{x'=x,t'=t}. \quad (11)$$

Decopling Equation (10), we obtain [13]

$$k'(t) + \tau(t)k(t) = 0, \quad (12)$$

$$-8f(t)k(t) + \frac{1}{3}h(t)k^3(t) = 0, \quad (13)$$

$$\left[D_t + f(t)D_x^3 + r(t)D_x\right]v.u = 0, \quad (14)$$

$$3f(t)D_x^2(u.u + v.v] = \frac{1}{2}g(t)k(t)D_x v.u, \quad (15)$$

Via Equations (12) and (13), we have the following relations

$$k(t) = c_0 e^{-\int\tau(t)dt}, f(t) = \frac{1}{24}h(t)k^2(t) \quad (16)$$

where c_0 is a nonzero arbitrary constant. That is to say, through the dependent variable transformation

$$\phi(x,t) = c_0 e^{-\int\tau(t)dt}\left[\arctan\frac{v(x,t)}{u(x,t)}\right]_x. \quad (17)$$

Equation (1) is transformed into its bilinear form, *i.e.*, Equations (14) and (15) under constraint (16). To solve the reduced Equations (14) and (15) using the extended homoclinic test function [21-29], we suppose a solution of Equations (14) and (15) as follows

$$v(x,t) = e^{a_1 x + b_1 t} + p_1\cos[a_2 x + b_2 t] + q_1 e^{-a_1 x - b_1 t}, \quad (18)$$

and

$$u(x,t) = e^{a_1 x + b_1 t} + p_2\cos[a_2 x + b_2 t] + q_2 e^{-a_1 x - b_1 t}, \quad (19)$$

where p_i, q_i, a_i, b_i $(i = 1,2)$ are parameters to be determined later.

Substituting Equations (18) and (19) into Equations (14) and (15), and equating all coefficients of $[e^{j(a_1 x + b_1 t)}$

$(j = -1, 0, 1)$, $\cos(a_2 x + b_2 t)$, $\sin(a_2 x + b_2 t)$] to zero, we get the set of algebraic equation for p_i, q_i, a_i, b_i $(i = 1, 2)$. Solve the set of algebraic equations with the aid of Maple, we have many solutions, in which the following solutions are

Case (1):

$$b_2 = 0, \ q_2 = 0, \ q_1 = 0, \ b_1 = -f(t)a_1^3 - r(t)a_1,$$

$$r(t) = r(t), \ p_2 = p_2, \ p_1 = p_1, \ a_2 = 0, \ a_1 = a_1, \quad (20)$$

$$g(t) = \frac{12 f(t)a_1(p_2 + p_1)}{k(t)(p_1 - p_2)}$$

Case (2):

$$b_2 = b_2, \ q_2 = q_2, \ q_1 = q_1, \ b_1 = -4f(t)a_1^3 - r(t)a_1,$$

$$r(t) = r(t), \ p_2 = 0, \ p_1 = 0, \ a_2 = a_2, \ a_1 = a_1, \quad (21)$$

$$g(t) = \frac{4 f(t)a_1(q_2 + q_1)}{k(t)(q_1 - q_2)}$$

Case (3):

$$p_2 = 0, \ p_1 = 0, \ q_2 = q_2, \ q_1 = q_1,$$

$$b_2 = b_2, \ b_1 = b_1 a_1 = a_1, \ a_2 = 0, \quad (22)$$

$$r(t) = -\frac{4f(t)a_1^3 + b_1}{a_1}, \ g(t) = -\frac{24 f(t)a_1(q_2 + q_1)}{k(t)(q_1 - q_2)}$$

Case (4):

$$g(t) = 0, \ p_1 = p_1, \ b_2 = b_2, \ a_1 = a_1, \ a_2 = a_2, \ q_2 = \frac{a_2^2 p_1^2}{4a_1^2},$$

$$q_1 = \frac{a_2^2 p_1^2}{4a_1^2}, \ b_1 = \frac{a_1(b_2 + 2f(t)a_2^3 + 2f(t)a_2 a_1^2)}{a_2},$$

$$r(t) = -\frac{b_2 - f(t)a_2^3 + 3f(t)a_1^2 a_2}{a_2}, \ p_2 = -p_1$$

$$\quad (23)$$

Case (5):

$$b_2 = a_2(4f(t)a_2^2 - r(t)), \ a_2 = a_2, \ p_1 = p_1, \ p_2 = p_2,$$

$$r(t) = r(t), \ b_1 = ia_2(4f(t)a_2^2 - r(t)), \ a_1 = ia_2,$$

$$g(t) = \frac{24if(t)a_2(p_2 + p_1)}{k(t)(p_1 - p_2)},$$

$$q_1 = \frac{1}{8}p_1^2 + \frac{1}{4}p_1 p_2 - \frac{1}{8}p_2^2, \ q_2 = -\frac{1}{8}p_1^2 + \frac{1}{4}p_1 p_2 + \frac{1}{8}p_2^2$$

$$\quad (24)$$

Using Equation (20), Equations (18) and (19) can be written as

$$v(x,t) = e^{a_1 x + [-f(t)a_1^3 - r(t)a_1]t} + p_1, \quad (25)$$

and

$$u(x,t) = e^{a_1 x + [-f(t)a_1^3 - r(t)a_1]t} + p_2. \quad (26)$$

Inserting Equations (25) and (26) into Equation (17), admits to the new solitary wave solution of Equation (1) as

$$\phi(x,t) = c_0 e^{-\int \tau(t)dt} \left[\arctan \frac{v(x,t)}{u(x,t)} \right]_x \quad (27)$$

With the aid of Equation (21), Equations (18) and (19) yields

$$v(x,t) = e^{a_1 x + [-4f(t)a_1^3 - r(t)a_1]t} + q_1 e^{-a_1 x - [-4f(t)a_1^3 - r(t)a_1]t}, \quad (28)$$

and

$$u(x,t) = e^{a_1 x + [-4f(t)a_1^3 - r(t)a_1]t} + q_2 e^{-a_1 x - [-4f(t)a_1^3 - r(t)a_1]t}, \quad (29)$$

Knowing Equations (28) and (29) with Equation (17), we have the solitary wave solution of Equation (1) as

$$\phi(x,t) = c_0 e^{-\int \tau(t)dt} \left[\arctan \frac{v(x,t)}{u(x,t)} \right]_x \quad (30)$$

In view of case (3), Equations (18) and (19) reads

$$v(x,t) = e^{a_1 x + b_1 t} + + q_1 e^{-a_1 x - b_1 t}, \quad (31)$$

and

$$u(x,t) = e^{a_1 x + b_1 t} + q_2 e^{-a_1 x - b_1 t}. \quad (32)$$

Inserting Equations (31) and (32) into Equation (17), admits to the new solitary wave solution of Equation (1) as

$$\phi(x,t) = c_0 e^{-\int \tau(t)dt} \left[\arctan \frac{v(x,t)}{u(x,t)} \right]_x \quad (33)$$

Via Equation (23) with Equations (18) and (19), we have

$$v(x,t) = e^{a_1 x + \left[\frac{a_1(b_2 + 2f(t)a_2^3 + 2f(t)a_2 a_1^2)}{a_2} \right]t} + \left[\frac{a_2^2 p_1^2}{4a_1^2} \right] \cos[a_2 x + b_2 t]$$

$$+ \left[\frac{a_2^2 p_1^2}{4a_1^2} \right] e^{-a_1 x - \left[\frac{a_1(b_2 + 2f(t)a_2^3 + 2f(t)a_2 a_1^2)}{a_2} \right]t},$$

$$\quad (34)$$

and

$$u(x,t) = e^{a_1 x + \left[\frac{a_1(b_2 + 2f(t)a_2^3 + 2f(t)a_2 a_1^2)}{a_2} \right]t} + \left[\frac{a_2^2 p_1^2}{4a_1^2} \right] \cos[a_2 x + b_2 t]$$

$$+ \left[\frac{a_2^2 p_1^2}{4a_1^2} \right] e^{-a_1 x - \left[\frac{a_1(b_2 + 2f(t)a_2^3 + 2f(t)a_2 a_1^2)}{a_2} \right]t},$$

$$\quad (35)$$

Using Equations (34) and (35), admits to the new solitary wave solution of Equation (1) as

$$\phi(x,t) = c_0 e^{-\int \tau(t)dt} \left[\arctan \frac{v(x,t)}{u(x,t)} \right]_x \qquad (36)$$

According to case (5), Equations (18) and (19) becomes

$$v(x,t) = e^{ia_2 x + \left[ia_2(4f(t)a_2^2 - r(t)) \right]t} + p_1 \cos[a_2 x$$
$$+ [a_2(4f(t)a_2^2 - r(t))]t] + q_1 e^{-ia_2 x - \left[ia_2(4f(t)a_2^2 - r(t)) \right]t}, \qquad (37)$$

and

$$u(x,t) = e^{ia_2 x + \left[ia_2(4f(t)a_2^2 - r(t)) \right]t} + p_2 \cos[a_2 x$$
$$+ [a_2(4f(t)a_2^2 - r(t))]t] + q_2 e^{-ia_2 x - \left[ia_2(4f(t)a_2^2 - r(t)) \right]t}. \qquad (38)$$

By means of Equations (36) and (37) with Equation (17) we have a new solitary wave solutions as

$$\phi(x,t) = c_0 e^{-\int \tau(t)dt} \left[\arctan \frac{v(x,t)}{u(x,t)} \right]_x \qquad (39)$$

3. Conclusions

In this paper, with the aid of two methods, namely, bilinear form and the extended homoclinic test approach, we obtain breather-type soliton and two soliton solutions for the Gardner equation with a variable-coefficient from fluid dynamics and plasma physics. The results reported here show that the extended homoclinic test approach is very effective in finding exact solitary wave solutions for nonlinear evolution equations with variable coefficients.

Finally, it is worthwhile to mention that, the proposed method is reliable and effective can also be applied to solve other types of higher dimensional integrable and non-integrable systems.

4. Acknowledgements

5. References

[1] M. Ablowitz and P. A. Clarkson, "Soliton, Nonlinear Evolution Equations and Inverse Scattering," Cambridge University Press, New York, 1991.

[2] S. A. El-Wakil and M. A. Abdou, "New Applications of Adomian Decomposition Method," *Chaos, Solitons and Fractals*, Vol. 33, No. 2, 2007, pp. 513-522.

[3] S. A. El-Wakil and M. A. Abdou, "New Exact Travelling Wave Solutions of Two Nonlinear Physical Models," *Nonlinear Analysis*, Vol. 68, No. 2, 2008, pp. 235-245.

[4] J.-H. He and M. A. Abdou, "New Periodic Solutions for Nonlinear Evolution Equations Using Exp Function Method," *Chaos, Solitons and Fractals*, Vol. 34, No. 5, 2007, pp. 1421-1429.

[5] M. A. Abdou and S. Zhang, "New Periodic Wave Solutions via Extended Mapping Method," *Communication in Nonlinear Science and Numerical Simulation*, Vol. 14, No. 1, 2009, pp. 2-11.

[6] M. A. Abdou, "On the Variational Iteration Method," *Physics Letters A*, Vol. 366, No. 1-2, 2007, pp. 61-68.

[7] R. Hirota, "Direct Method in Soliton Theory," In: R. K. Bullough and P. J. Caudrey, Ed., *Solitons*, Springer, Berlin, 1980, pp. 1-196.

[8] M. A. Abdou, "Generalized Solitary and Periodic Solutions for Nonlinear Partial Differential Equations by the Exp-Function Method," *Journal of Nonlinear Dynamics*, Vol. 52, No. 1-2, 2008, pp. 1-9.

[9] G. M. Wei, Y. T. Gao and X. G. Xu, "Painlevé Analysis and Transformations for a Generalized Two-Dimensional Variable-Coefficient Burgers Model from Fluid Mechanics, Acoustics and Cosmic-Ray Astrophysics," N*uovo Cimento B*, Vol. 121, No. 4, 2006, pp. 327-342.

[10] P. E. P. Holloway, E. Pelinovsky, T. Talipova and B. Barnes, "A Nonlinear Model of Internal Tide Transformation on the Australian North West Shelf," *Journal of Physical Oceanography*, Vol. 27, No. 6, 1997, pp. 871-896.

[11] J. A. Gear and R. Grimshaw, "A Second-Order Theory for Solitary Waves in Shallow Fluids," *Physics of Fluid*, Vol. 26, No. 14, 1983, 16 pages.

[12] S. Watanabe, "Ion Acoustic Soliton in Plasma with Negative Ion," *Journal of Physical Society of Japan*, Vol. 53, 1984, pp. 950-956.

[13] X.-G. Xu, X. Meng, Y. Gao and X. Wen, "Analytic N-Solitary-Wave Solution of a Variable-Coefficient Gardner Equation from Fluid Dynamics and Plasma Physics," *Applied Mathematics and Computation*, Vol. 210, No. 2, 2009, pp. 313-320.

[14] N. Joshi, "Painlevé Property of General Variable- Coefficient Versions of the Korteweg-De Vries and Non-Linear Schrödinger Equations," *Physics Letter A*, Vol. 125, No. 9, 1987, p. 456.

[15] R. Grimshaw, *Proceedings of the Royal Society of London*, Series A, 1979, p. 359.

[16] Z. X. Chen, B. Y. Guo and L. W. Xiang, "Complete Integrability and Analytic Solutions of a Kdv-Type Equation," *Journal of Mathematical Physics*, Vol. 31, 1990, p. 2851.

[17] W. P. Hong and Y. D. Jung, "Auto-Bäcklund Transformation and Analytic Solutions for General Variable-Coefficient Kdv Equation," *Physics Letters A*, Vol. 257, No. 3-4, 1999, pp. 149-157.

[18] E. J. Fan, "Auto-Bäcklund Transformation and Similarity Reductions for General Variable Coefficient Kdv Equations," *Physics Letters A*, Vol. 294, No. 1, 2002, pp. 26-30.

[19] G. Xu, X. H. Meng, Y. T. Gao and X. Wen, "Analytic N-Solitary Wave Solution for a Variable-Coefficient Gardner Equation from Fluid Dynamics and Plasma Physics," *Applied Mathematics and Computation*, Vol. 210, No. 2, 2009, pp. 313-320.

[20] Z. D. Dai, Z. J. Liu and D. L. Li, "Exact Periodic Solitary-Wave Solutions for the Kdv Equation," *Chinese Physics Letters*, Vol. 25, No. 5, 2008, pp. 1531-1533.

[21] Z. D. Dai, M. R. Jiang, Q. Y. Dai and S. L. Li, "Homoclinic Bifurcation for the Boussinesq Equation with Even Constraints," *Chinese Physics Letters*, Vol. 23, No. 5, 2006, pp. 1065-1067.

[22] Z. D. Dai, J. Liu and D. L. Li, "Applications of HTA and EHTA to the YTSF Equation," *Applied Mathematics and Computation*, Vol. 207, No. 2, 2009, pp. 360-364.

[23] Z. D. Dai, J. Liu, X. P. Zeng and Z. J. Liu, "Periodic Kink-Wave and Kinky Periodic-Wave Solutions for the Jimbo-Miwa Equation," *Physics Letters A*, Vol. 372, No. 38, 2008, pp. 5984-5986.

[24] Z. Dai, L. Song, H. Fu and X. Zeng, "Exact Three Wave Solutions for the KP Equation," *Applied Mathematics and Computation*, Vol. 216, No. 5, 2010, pp. 1599-1604.

[25] C. Wang, Z. D. Dai and L. Liang, "Exact Three Wave Solution for Higher Dimensional Kdv Equation," *Applied Mathematics and Computation*, Vol. 216, No. 2, 2010, pp. 501-505.

[26] W. Chuan, Z. D. Dai, M. Gui and L. S. Qing, "New Exact Periodic Solitary Wave Solutions for New (2 + 1)-Dimensional Kdv Equation," *Communications in Theoretical Physics*, Vol. 52, No. 2, 2009, pp. 862-864.

[27] D. L. Li and J. X. Zhao, "New Exact Solutions to the (2 + 1)-Diomensional Ito Equation, Extended Homoclinic Test Technique," *Applied Mathematics and Computation*, Vol. 215, No. 5, 2009, pp. 1968-1974.

[28] Z.-D. Dai, C.-J. wang, S.-Q. Lin, D. L. Li and G. Mu, "The Three Wave Method for Nonlinear Evolution Equations," *Nonlinear Science Letter A*, Vol. 1, No. 5, 2010, pp. 77-82.

Scientific Research

Some Notes on the Distribution of Mersenne Primes

Sibao Zhang[1], Xiaocheng Ma[1], Lihang Zhou[2]

[1]*Department of Mathematics, Kashgar Teachers College, Kashgar, China*
[2]*Department of Computer Science, Guangdong Technical College of Water Resources and Electric Engineering, Guangzhou, China*
E-mail: sibao98@sina.com

Abstract

Mersenne primes are a special kind of primes, which are always an important content in number theory. The study of Mersenne primes becomes one of hot topics of the nowadays science. It has not settled that whether there exist infinite Mersenne primes. And several of conjectures on the distribution of it provided by scholars. Starting from the Mersenne primes known about, in this paper we study the distribution of Mersenne primes and argued against some suppositions by data analyzing.

Keywords: Mersenne Primes, Distribution, Zhou Conjecture, Number Theory

1. Introduction

In 300 B.C , Euclid, famed ancient Greek mathematician proved that there are infinitude prime numbers by contradiction, and raised that a small amount of prime numbers could be expressed in the formulation of $2^p - 1$ (*p* is a prime number). After that, many famous mathematicians including mathematical masters like Fermat, Descartes, Leibniz, Goldbach, Euler, Gauss all researched into the prime numbers of this kinds of special formulation. But M. Mersenne, the French mathematician in 17th century and founder of Institute of France, was the first one to research into the prime number of $2^p - 1$ formulation deeply and systematically. These kinds of formulation like $2^p - 1$ is named as "Mersenne number" and expressed as M_p to commemorate him. If a Mersenne number is a prime number, then is named as "Mersenne prime". The Mersenne prime seems like very simple, but the calculation of it is very complex. As the increase of index *p*, the calculation increases more complex. It not only need advanced theory and practiced skills, but arduous calculation is needed to validate whether a Mersenne number is prime or not. People have only discovered 47 Mersenne primes in the past 2300 years [1].

It should be noted that we have only settled the sequence of first 39 primes, while the last 8 primes left. That is, there is no other prime number *p* which makes $2^p - 1$ to be a prime number in the range of $2 \le p \le 13\ 466\ 917$. And we still cant assure that whether exists other prime number *p* which makes $2^p - 1$ to be prime

number in the range of $20\ 996\ 011 \le p \le 43\ 112\ 609$. Though we have not discovered other prime *p* which makes $2^p - 1$ to be a prime number by checked the range at least once, but twice could confirm the seating arrangement [2].

It is not known whether the set of Mersenne primes is infinite. But researching on the attribution of Mersenne Primes is very important for the seeking of new Mersenne primes and exploring whether exist infinitude Mersenne primes. From the known Mersenne primes, the distribution of these special kinds of prime number is either sparse or dense in the positive integer, so very anomalous. In the long-term exploring, mathematicians have advanced some kind of suppositions. For example, mathematicians like Shanks from England, Bertrand from France, Ramanujan from India, Gilles from America and Brillhart from Germany have all supposed the distribution of Mersenne primes. The common point of their supposition is that they all presented as asymptotic expression.

In 1992, Haizhong Zhou [3], famed Chinese mathematician and linguist advanced the precise formulation of the Mersenne prime distribution: If $2^{2^n} < p < 2^{2^{n+1}}$ (*n* = 0, 1, 2, 3,…), then the amount of Mersenne primes is $2^{n+1} - 1$. The accurate and beautiful expression is made by him. Zhou conjecture has not been proved or disproved, and becomes a well-known mathematical problem [4].

Some researchers have raised the suppositions for the distribution of Mersenne primes. In this article we would

argue against the deduction and raise different views by data analyzing for several suppositions.

2. Questions and View Points

2.1. Questions

In 1995, based on Zhou conjecture, Suwen Chen tried to make a further discussion on the distribution of Mersenne primes [5]. In the paper he defined that: Sequence **R** consists of primes p which makes $2^p - 1$ to be prime and numbers like 2^{2^k} ($k = 0, 1, 2, \ldots$) in ascending, and the item No. n marked **R**(n); **P** is the sequence consists of primes p which makes $2^p - 1$ primes in ascending, while the item No. n marked **P**(n); **Q** is the sequence consists of numbers like $Q(n) = 2^{2^n}$ ($n = 0, 1, 2, 3, \ldots$). By analyzing the first 35 items settled by **R** (n), conjectures proposed like this:

CONJECTURE 1

1) $R(2^{n+1}) = Q(n) = 2^{2^n}$;

2) $n/2 - 1 < \log_2 R(n) < n/2 + 1$;

3) $\lim_{n \to \infty} \dfrac{\log_2 R(n)}{n} = \dfrac{1}{2}$.

Then, raise doubts on item 2). we shall list the related data as following for convenience.

Attention to this data:

WHEN $n = 20$ and **R**(n) = 2 203, GET $\log_2 R(n)$ = 11.10525, **n/2** = 10.00000, **n/2** + 1 = 11.00000;

WHEN $n = 37$ and **R**(n) = 756 839, GET $\log_2 R(n)$ = 19.52963, **n/2** = 18.50000, **n/2** + 1 = 19.50000;

WHEN $n = 41$, **R**(n) = 2 976 221, GET $\log_2 R(n)$ = 21.50505, **n/2** = 20.50000, **n/2** + 1 = 21.50000;

WHEN $n = 43$, **R**(n) = 6 972 593, GET $\log_2 R(n)$ = 22.73326, **n/2** = 21.50000, **n/2** + 1 = 22.50000;

WHEN $n = 44$, **R**(n) = 13 466 917, GET $\log_2 R(n)$ = 23.68292, **n/2** = 22.00000, **n/2** + 1 = 23.00000

The location of Mersenne primes ignored for they have not been settled when $20\,996\,011 \leq p \leq 43\,112\,609$. The data listed above satisfied the formula $\log_2 R(n) > n/2 + 1$, and item 2) in the conjecture follows the formula $\log_2 R(n) < n/2 + 1$ which contradicts the result $\log_2 R(n) > n/2 + 1$, so it concludes that the formula $\log_2 R(n) < n/2 + 1$ in conjecture $n/2 - 1 < \log_2 R(n) < n/2 + 1$ is wrong.

In Chen's paper, the author presupposed that $2^{\frac{n+3}{2}} < P(n) < 2^{\frac{n+7}{2}}$, while $P(31) \leq P(n) \leq P(58)$ according to conjecture $n/2 - 1 < \log_2 R(n) < n/2 + 1$. According to Chen's deduction, a great difference occurred on some numbers. It will be subscribed by the location-settled Mersenne primes as following.

WHEN **R**(32) = 756839, $2^{19.5} < 741456$, GET 756

839 < 741 456, impossible;

WHEN **R**(36) = 2976221, $2^{21.5} < 2965821$, GET 2 976221 < 2 965 821, impossible;

WHEN **R**(38) = 6 972 593, $2^{22.5} < 5931642$, GET 6 972 593 < 5 931 642, impossible;

2.2. Different Views

Scholars proposed many conjectures on how to confirm the quantity of Mersenne primes in certain range. In 1980, Lenstra and Pomerance [6] independently presupposed the quantity of Mersenne primes while less than x, which would be $(e^\gamma / \log 2) \log \log x$, as $\gamma = 0.5772$ is the Euler's constant. Based on that, Wagstaff presupposed a conjecture in 1983, as following:

CONJECTURE 2 [7]

1) IF the quantity of Mersenne primes which less than x is $\pi_M(x)$, Then

$$\pi_M(x) \approx \frac{e^\gamma}{\ln 2} \log \log x = (2.5695\cdots) \ln \ln x,$$

As γ is the Euler's constant;

2) the expect value of Mersenne primes M_q is about $e^\gamma = 1.7806\cdots$, while $x < q < 2x$;

3) the probability of M_q is a prime number is about

$$\frac{e^\gamma}{\ln 2} \cdot \frac{\ln aq}{\ln 2} = (2.5695\cdots)\frac{\ln aq}{q},$$

as $a = \begin{cases} 2 & q \equiv 3 \pmod 4 \\ 6 & q \equiv 1 \pmod 4 \end{cases}$.

This conjecture explains the probability of M_q is a Mersenne prime in the precondition of q, and also pointed the quantity of Mersenne primes in certain range. It has been confirmed that there are 39 Mersenne primes when $p \leq 13466917$. However, the number

$$\pi_M(x) \approx \frac{e^\gamma}{\ln 2} \log \log 13466917$$
$$= (2.5695\cdots) \ln \ln 13466917 \approx 7.190360$$

quite differs from the actual situation which could be thought that the conjecture 1) is not advisable. Based on conjecture 3), take $q = 521$ and $q = 257$ in Mersenne conjecture as an example, for 521*257 1(mod4), while $a = 6$, then we get

$$(2.5695\cdots)\frac{\ln aq}{q} = (2.5695\cdots)\frac{\ln(6 \cdot 521)}{521} \approx 0.039691$$

$$(2.5695\cdots)\frac{\ln aq}{q} = (2.5695\cdots)\frac{\ln(6 \cdot 257)}{257} \approx 0.073397$$

It's well known that M_{521} is a Mersenne prime, but M_{257}. It could meet a big mistake to some extent if we

AM

Table 1. Relationship between *R*(*n*) and *n* [5].

n	*R*(*n*)	*P*, *Q*	log₂*R*(*n*)	*n*/2	*n*	*R*(*n*)	*P*, *Q*	log₂*R*(*n*)	*n*/2
1	2	*P*(1)	1.00000	0.50000	27	11 213	*P*(23)	13.45288	13.50000
2	2	*Q*(0)	1.00000	1.00000	28	19 937	*P*(24)	14.28316	14.00000
3	3	*P*(2)	1.58496	1.50000	29	21 701	*P*(25)	14.40547	14.50000
4	4	*Q*(1)	2.00000	2.00000	30	23 209	*P*(26)	14.50240	15.00000
5	5	*P*(3)	2.32193	2.50000	31	44 497	*P*(27)	15.44142	15.50000
6	7	*P*(4)	2.80735	3.00000	32	65 536	*Q*(4)	16.00000	16.00000
7	13	*P*(5)	3.70044	3.50000	33	86 243	*P*(28)	16.39612	16.50000
8	16	*Q*(2)	4.00000	4.00000	34	110 503	*P*(29)	16.75373	17.00000
9	17	*P*(6)	4.08746	4.50000	35	132 049	*P*(30)	17.01071	17.50000
10	19	*P*(7)	4.24793	5.00000	36	216 091	*P*(31)	17.72128	18.00000
11	31	*P*(8)	4.95420	5.50000	37	756 839	*P*(32)	19.52963	18.50000
12	61	*P*(9)	5.93074	6.00000	38	859 433	*P*(33)	19.71303	19.00000
13	89	*P*(10)	6.47573	6.50000	39	1 257 787	*P*(34)	20.26246	19.50000
14	107	*P*(11)	6.74147	7.00000	40	1 398 269	*P*(35)	20.41521	20.00000
15	127	*P*(12)	6.89968	7.50000	41	2 976 221	*P*(36)	21.50505	20.50000
16	256	*Q*(3)	8.00000	8.00000	42	3 021 377	*P*(37)	21.52677	21.00000
17	521	*P*(13)	9.02514	8.50000	43	6 972 593	*P*(38)	22.73326	21.50000
18	607	*P*(14)	9.24555	9.00000	44	13 466 917	*P*(39)	23.68292	22.00000
19	1 279	*P*(15)	10.32080	9.50000	45	20 996 011	*P*(?)	24.32361	22.50000
20	2 203	*P*(16)	11.10525	10.00000	46	24 036 583	*P*(?)	24.51873	23.00000
21	2 281	*P*(17)	11.15545	10.50000	47	25 964 951	*P*(?)	24.63006	23.50000
22	3 217	*P*(18)	11.65150	11.00000	48	30 402 457	*P*(?)	24.85768	24.00000
23	4 253	*P*(19)	12.05427	11.50000	49	32 582 657	*P*(?)	24.95760	24.50000
24	4 423	*P*(20)	12.11081	12.00000	50	37 156 667	*P*(?)	25.14712	25.00000
25	9 689	*P*(21)	13.24213	12.50000	50	42 643 801	*P*(?)	25.34583	25.50000
26	9 941	*P*(22)	13.27918	13.00000	52	43 112 609	*P*(?)	25.36161	26.00000

Note: The symbol ? means the location of those Mersenne primes haven't been settled.

judge from higher probability.

3. Conclusions

Starting from analyzing of the known Mersenne primes, different ideas proposed about the distributional conjectures of Mersenne primes, which would be beneficial to the studies on the quantity of Mersenne primes and the distribution of its index prime *p*.

4. References

[1] Q. Chen and P. Zhang, "A Math Treasure: Mersenne Primes," *Encyclopedic Knowledge*, Vol. 32, No. 5, 2009, p. 22.

[2] K. H. Rosen, "Elementary Number Theory and it's Ap-

plications," Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Addision-Wesley, 2005.

[3] H. Z. Zhou, "The Distribution of Mersenne Primes," *Acta Scientiarum Naturalium Universitatis Sunyatseni*, Vol. 31, No. 4, 1992, pp. 121-122.

[4] J. Z. Zhang, "Zhou Conjecture Reveals the Beauty of Mathematics," In: J. Z. Zhang, *et al.*, Ed., 100 *Cases of Scientific and Technological Achievements in Three Decades*: 1978-2008, Children's Press, Wuhan: Hubei, 2008, pp. 8-9.

[5] S. W. Chen, "Conjecture on Distribution of Mersenne Primes," *Huanghuai Journal*, Vol. 11, No. 4, 1995, pp. 44-46.

[6] C. Pomerance, "Recent Developments in Primality Testing," *Mathematical Intelligencer*, Vol. 3, No. 3, 1980, pp. 97-105.

[7] S. Wagstaff, "Divisors of Mersenne Numbers," *Mathematics and Computation*, Vol. 40, No. 38, 1983, pp. 385-397.

◆◆ Scientific
◆◆ Research

Using Differential Evolution Method to Solve Crew Rostering Problem

Budi Santosa, Andiek Sunarto, Arief Rahman
Indusrial Engineering, Institut Teknologi Sepuluh Nopember,Kampus ITS Sukolilo Surabaya, Indonesia
E-mail: *budi_s@ie.its.ac.id*

Abstract

Airline crew rostering is the assignment problem of crew members to planned rotations/pairings for certain month. Airline companies have the monthly task of constructing personalized monthly schedules (roster) for crew members. This problem became more complex and difficult while the aspirations/criterias to assess the quality of roster grew and the constraints increased excessively. This paper proposed the differential evolution (DE) method to solve the airline rostering problem. Different from the common DE, this paper presented random swap as mutation operator. The DE algorithm is proven to be able to find the near optimal solution accurately for the optimization problem. Through numerical experiments with some real datasets, DE showed more competitive results than two other methods, column generation and MOSI (the one used by the Airline). DE produced good results for small and medium datasets, but it still showed reasonable results for large dataset. For large crew rostering problem, we proposed decomposition procedure to solve it in more efficient manner using DE.

Keywords: Differential Evolution, Crew Scheduling, Pairing, Rostering

1. Introduction

Development of crews rostering plan which be able to produce the high utility of crews become the priority in human resources department in airline industry. It is estimated that the use of optimization software for airline could save more than US $20 million per year [1]. Saving 1% in crew utilization can save cost largely. Though airline crews scheduling became attention in many operation research literature such as [1-7] but airline crews scheduling remains to become the main attention for many researchers due to its level of complexity and difficulty to solve. Therefore, methods and approaches which are used to solve it are continuously developed to get better result both in optimality side and speed of computational time. Generally, solving airline crew scheduling is done by decomposition approach [8-10]), it devides problem into crew pairing and crew rostering. Crew pairing is done to get initial feasible solution, that is sequence of flight which begin and end at the same home base. Crew rostering assigns pairings which were arranged for the certain month to set of crews based on individual calender.

Decomposition approach is very effective to solve the

difficult and complex problem but this method loss the global treatment since crew pairing and crew rostering done separately. Some other researchers developed the integrated approach to overcome obstacle, such as Souai and Thegem [5], where crew pairing and rostering were done simultantly to get a better optimality level.

Many optimization methods have been developed to solve crew scheduling to increase roster quality and to improve computational time such as simulated annealing [11], genetic algorithm [5], tree search algorithm [12], hybrid genetic algorithm [13] and GASA hybrid algorithm [14].

This research focused on developing differential evolution algorithm applied on intelligent airline crew rostering system. This paper is organized as follows. The second section reviews differential evolution (DE). Section 3 describes the problem statement. In Section 4, we describe our methodology. Section 5 explains the experimental setting and the results. Section 6 concludes the results.

2. Differential Evolution

Differential evolution is an evolutionary population-based

algorithm proposed by Storn and Price [15,16]. Since its initiation in 1995, DE has shown its performance as a very effective global optimizer. DE originated with Genetic Annealing (GA) Algorithm. Since GA was very slow and effective control parameters were hard to determine, the modification of the GA algorithm were made. DE uses a floating-point instead of bit-string encoding and arithmatic operations instead of logical ones. DE differs significantly from the evolutionary algorithms in the sense that distance and direction information from the current population is used to guide the search process. DE uses the differences between two randomly chosen vectors (individuals) as the base to form a third vector (individual), referred to as the target vector. Trial solutions are generated by adding weighted difference vectors to the target vector. This process is referred to as the mutation operator where the target vector is mutated. The next step is recombination or crossover which is applied to produce an offspring. This new individual is accepted only if it improves on the fitness of the parent individual. The basic DE algorithm is described in more detail below [16].

2.1. Initialization

In this step, a set of initial solutions are genereted randomly. A random number generator assigns each variable of each vector value from within specified range, lower bound, $\mathbf{b}_L$, and upper bound, $\mathbf{b}_U$. For example the initial value ($g = 0$) of the jth variable of ith vector is:

$$\mathbf{x}_{j,i,0} = \text{rand}_j(0,1).\left(\mathbf{b}_{j,U} - \mathbf{b}_{j,L}\right) + \mathbf{b}_{j,L} \qquad (1)$$

where, $\text{rand}_j(0,1)$, returns a uniformly distributed random number from within the range [0,1].

2.2. Mutation

DE mutates and recombines the population to produce a population of N-trial vectors. In particular, differential mutation adds a scaled, randomly sampled, vector difrence of third vector. To combine three different ranmly chosen vectors to create the mutant vector, $\mathbf{v}_{i,g}$, the following equation is used:

$$\mathbf{v}_{i,g} = \mathbf{x}_{r0,g} + \mathbf{F}.\left(\mathbf{x}_{r1,g} - \mathbf{x}_{r2,g}\right) \qquad (2)$$

where the scale factor, $\mathbf{F} \in (0, 1+)$ is a positive real number that control the rate at which the population evolve. The vector index, r_0, r_1, and r_2, can be chosen randomly and meet $r_0 \neq r_1 \neq r_2 \neq i$.

2.3. Crossover

DE employs the uniformly crossover. Crossover builds

trial vectors out of parameter value that have been copied from two different vectors. In particular, DE crosses each vector with a mutant vector to creat $\boldsymbol{u}_{i,g}$.

$$\boldsymbol{u}_{i,g} = \boldsymbol{u}_{j,i,g} = \begin{cases} v_{j,i,g}, if\,(rand_j\,(0,1) \leq Cr, or\, j = j\max \\ x_{j,i,g}, otherwise \end{cases}$$

$$(3)$$

The crosover probability, $Cr \in [0,1]$, is user-defined value that controls the fraction of parameter value that are copied from the mutant.

2.4. Selection

If the trial vector, $\boldsymbol{u}_{i,g}$, has an equal or lower objective function value (better fitness value) than that of its target vector, $\boldsymbol{x}_{i,g}$, it replaces the target vector in the next generation; otherwise, the target retains its place in the population at least one more generation.

$$\boldsymbol{x}_{i,g+1} = \begin{cases} \boldsymbol{u}_{i,g}, if\ f(u_{i,g}) \leq f(x_{i,g}) \\ \boldsymbol{x}_{i,g}, otherwise \end{cases} \qquad (4)$$

Once the new population created, the process of mutation, recombination, and selection are repeated until the optimum solution is achieved or prespesified termination criterion is satisfied, e.g., the number of generations reaches preset maximum, g_{max}.

DE has been applied in many field successfully. In 1995, DE has been used by Ken to solve 5-dimension Chebyshev model. By the time, Ken modified *genetic annealing* algorithm with *differential mutation operator*. Different from *genetic annealing*, DE has not found some difficulty to find the coefficient even 33-dimension Chebyshev.

Tasgetiren [15] used a discrete differential evolution (DDE) algorithm to solve the single machine total earliss and tardiness penalties with a common due date. A new binary swap mutation operator called *Bswap* is prented. In addition, the DDE algorithm is hybridized with a local search algorithm to further improve the performance of the DDE algorithm. The performance of the proposed DDE algorithm is tested on 280 benchmark instances ranging from 10 to 1000 jobs from the OR Library. The computational experiments showed that the proposed DDE algorithm has generated better results than those in the literature in terms of both solution quality and computational time.

A genetic differential evolution (GDE) was derived from the differential evolution (DE) and incorporated with the genetic reproduction mechanisms, namely crossover and mutation used to solve traveling salesman problems (TSP). The Greedy Subtour Crossover (GSX) was employed to generate an offspring to denote the difference of the parents. A modified ordered crossover (MOX)

was employed to perform mutation to generate trial vector with a user defined parameter, the parameter were used to control the rates of the target vector components and the mutated vector components in the trial vector. Moreover, a 2-opt local search was implemented to enhance local search performance. GDE was implemented to the well-known TSP with 52, 100 and 200 cities with variable parameters. Based on analysis and discussion on the results, typical values of the parameters were given, with which GDE provided effective and robust performance [18].

Omran and Salman [19] improved Differential evolution by combining with chaotic search, opposition-based learning, and quantum mechanics, called CODEQ, to solve constrained optimization problems. The performance of the proposed approach when applied to five constrained benchmark problems is investigated and compared with other approaches proposed in the literature. The experiments conducted show that CODEQ provides excellent results with the added advantage of no parameter tuning.

3. Problem Statement

There are two main processes in the airline crew planning. These two process are pairing and rostering. Pairing, is the step where the flying activities are created. The flight timetable is used as input to form sequences of flights, known as pairings. The timetable horizon usually covers a period of 4-6 weeks. The main objective of this process is to utilize the minimum number of crew to cover the complete timetable. Rostering is the step where the created pairings are assigned to actual crew (pilot and stewardess), with regard to the qualifications and previously assigned activities, referred to as pre-assignments, of the crew. The objective is to find feasible assignments that minimize costs and match the notion of quality of life for the crew imposed by the airline [9]. In this context, problem of *airline crew scheduling* generally varies among different countries. Especially in rostering

problem since each country may impose different set of regulations, rules and policies. The speed of roster construction is a critical matter in *airline crew scheduleing*. In [9], it is stated that for monthly planning, solution should be obtained within 15-20 minutes. While, for shorter horizon planning, such as daily planning, where there are some changes to roster, solutions should be obtained within 1-5 minutes. Therefore, solving optimization problem of airline crew scheduling in time manner become very important.

In this paper two main problems are addressed. First, how to mathematically formulate the problem of airline crew scheduling. This formulation includes the construction of objective function and spesific set of constraints which influence the roster quality. We modified the model which is developed by Lučić and Teodorović [11] based on the real condition of rostering in the airline under study. We modified the single crew and single aircraft scheduling model becomes multiple crews and multiple aircrafts model. We also added open time criteria to the objective function. Second, how to solve the model using *differential evolution* approach with considering the simplicity of the model, quality of the resulting *roster*, and computational time.

3.1. Index

k is index for kind of crews $k = 1, \cdots, K$. For example, k = 1 is for Pilot F-100; k = 2 is for Pilot CN-235; and k = 8 is for stewardess and etc.

i is index for numbers of crew members $(1, \cdots, m_k)$. For example $m_5 = 17$ is the number of crew members for Boeing 737-200, and $m_8 = 55$ is the number of stewardess of Boeing 737-200.

j is index rotation/pairing which assigned to crew members $(1, \cdots, n_k)$. For example $n_5 = 82$, is the number of pairings for Being 737-200.

l is number of days in a month $(1, \cdots, 31)$.

3.2. Parameters

d_{jk} is length of rotation-j which assigned to crew k. (in hours).

$$p_{ilk} = \begin{cases} 1, \text{if member } i \text{ from crew } k \text{ can be assigned to day } l \\ 0, \text{ otherwise} \end{cases}$$

$$q_{jlk} = \begin{cases} 1, \text{if } rotation / pairing \ j \text{ assigned to crew } k \text{ start to day } l \\ 0, \text{ otherwise} \end{cases}$$

$d_{max,k}$ is maximum flight times crew k for one month;
v_{jk} is numbers of take-off rotation j assigned to crew k;
$v_{max,k}$ is maximum take-off in one month;
$D_{min,jk}$ is minimum of numbers of crews k needed to

complete rotation j;
t_{jk} is numbers of duty period needed by crew k to complete rotation j;
$t_{max,k}$ is maximum of flying day before free day.

$$\tilde{n}_{rsk} = \begin{cases} 1, \text{if rotation } r \text{ overlap with rotation } s \text{ when assigned to crew } k \\ 0, \text{ otherwise} \end{cases}$$

3.3. Objectives Function

The objective function of this airline crew rostering is minimizing three terms of criterias.

Cost of roster

Cost of roster paid by the airline company to crew is variable cost. By assumption that salary per hour is same to all crew, cost of roster can be represented by actual flying hours.

$$\min \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} d_{jk} x_{ijk} \quad k = 1, \cdots, K \tag{5}$$

Deviation of flying days between crew members

Let $\bar{t}_k$ be the average flying days per month for crew member k, then

$$\bar{t}_k = \frac{\sum_{i=1}^{m_k} \sum_{j=1}^{n_k} t_{jk} x_{ijk}}{m_k} \quad k = 1, \cdots, K \tag{6}$$

The deviation of total flying days per month can be formulated as

$$\min \sum_{i=1}^{m_k} \left| \sum_{j=1}^{n_k} t_{jk} x_{ijk} - \bar{t}_k \right|^p \quad k = 1, \cdots, K \tag{7}$$

where p is positive integer. In this paper we use $p = 1$.

Open time

Open time is days when a crew member does not have flying duty. If there are 31 days in a scheduling month, then *open time* for crew member k can be formulated as

$$\min \sum_{i=1}^{m_k} \left(31 - \sum_{j=1}^{n_k} t_{jk} x_{ijk} \right) \quad k = 1, \cdots, K \tag{8}$$

3.4. Constraints

There are some constraints which must be satisfied when constructing a roster. The following are the constraints used:

Flight time constraint

Maximum flying hours for pilot and co-pilot is 110 hours per month, and for cabin crew is 120 hours per month. So $d_{max,k} = 110$ for $k = 1, \cdots, 7$ and $d_{max,k} = 120$ for $k = 8$.

$$\sum_{j=1}^{n_k} d_{jk} x_{ijk} \leq d_{\max,k} \quad i = 1, ..., m_k, \ k = 1, \cdots, K \tag{9}$$

Duty period constraint

Maximum duty period allowed to crew member k is 21 days.

$$\sum_{j=1}^{n_k} t_{jk} x_{ijk} \leq 21 \quad i = 1, \cdots, m_k, \ k = 1, \cdots, K \tag{10}$$

Numbers of *take-off*

Numbers of maximum *take off* allowed to pilot is 90,

then $v_{max,k} = 90$. But, cabin crew have no take off constraint.

$$\sum_{j=1}^{n_k} v_{jk} x_{ijk} \leq v_{\max,k} \quad i = 1, \cdots, m_k, \ k = 1, \cdots, K \tag{11}$$

Numbers of crew reqirement

Every rotation needs minimum numbers of crew.

$$\sum_{i=1}^{m_k} x_{ijk} \leq D_{\min, jk} \quad j = 1, \cdots, n_k, \ k = 1, \cdots, K \tag{12}$$

Free day constraint

Every crew member must be given free days after 7 flying days.

$$\sum_{j=1}^{n_k} t_{jk} x_{ijk} \sum_{l=p}^{p+7} q_{jkl} \leq 7 \quad i = 1, \cdots, m_k, \ p = 1, \cdots, 23, \ k = 1, \cdots, 8 \tag{13}$$

Rotation without free day

When crew members complete this rotation not allowed to have free day.

$$\sum_{j=1}^{n_k} x_{ijk} = \sum_{j=1}^{n_k} x_{ijk} \sum_{l=1}^{31} q_{jkl} \prod_{s=l}^{l+t_{jk}-1} p_{isk} \quad i = 1, \cdots, m_k, \ k = 1, \cdots, K \tag{14}$$

No overlap constraint

Two rotations in series may not be overlap each other. It means that precedence rotation must be finished when following rotation will start.

$$x_{ijk} \sum_{s=1}^{n_k} \rho_{jsk} x_{isk} (s - j) = 0$$
$$i = 1, \cdots, m_k, \ j = 1, \cdots, n_k, \ k = 1, \cdots, K \tag{15}$$

Airline rostering problem is optimization problem with many constraints. It falls to constrained optimization problem. To use DE to solve this original problem we have to transform the problem into an unconstrained optimization problem. We use external penalty function method [20,21] to do this transformation. Basically, this method incurs big penalty while the solution violated any constraints. The resulting constrained optimization problem is as follows

$$\min \beta_1 \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} d_{jk} x_{ijk} + \beta_2 \sum_{i=1}^{m_k} \left| \sum_{j=1}^{n_k} t_{jk} x_{ijk} - \bar{t}_k \right|^p$$
$$+ \beta_3 \sum_{i=1}^{m_k} \left(31 - \sum_{j=1}^{n_k} t_{jk} x_{ijk} \right) + \left(r_1 C_1 + r_2 C_2 + r_3 C_3 + r_4 C_4 \right) \tag{16}$$
$$+ \left(r_5 C_5 + r_6 C_6 + r_7 C_7 + r_8 C_8 \right)$$

where:

$$C_1 = \sum_{i=1}^{m_k} \left(\max \left(0, \sum_{j=1}^{n_k} d_{jk} x_{ijk} - d_{\max,k} \right) \right)^2 \quad k = 1, \cdots, K$$

$$C_2 = \sum_{i=1}^{m_k} \left(\max\left(0, \sum_{j=1}^{n_k} v_{jk} x_{ijk} - v_{\max,k} \right) \right)^2 \quad k = 1, \cdots, K$$

$$C_3 = \sum_{j=1}^{n_k} \left(\max\left(0, \sum_{i=1}^{m_k} x_{ijk} - D_{\min,jk} \right) \right)^2 \quad k = 1, \cdots, K$$

$$C_4 = \sum_{i=1}^{m_k} \left(\max\left(0, \sum_{j=1}^{n_k} t_{jk} x_{ijk} \sum_{l=p}^{p+7} q_{jkl} - 7 \right) \right)^2 \quad k = 1, \cdots, K$$

$$C_5 = \sum_{i=1}^{m_k} \left(\max\left(0, \sum_{j=1}^{n_k} x_{ijk} - \sum_{j=1}^{n_k} x_{ijk} \sum_{l=1}^{31} q_{jkl} \prod_{s=l}^{l+t_{jk}-1} p_{isk} \right) \right)^2$$
$$k = 1, \cdots, K$$

$$C_6 = \sum_{i=1}^{m_k} \left(-\min\left(0, \sum_{j=1}^{n_k} x_{ijk} - \sum_{j=1}^{n_k} x_{ijk} \sum_{l=1}^{31} q_{jkl} \prod_{s=l}^{l+t_{jk}-1} p_{isk} \right) \right)^2$$
$$k = 1, \cdots, K$$

$$C_7 = \sum_{i=1}^{m_k} \left(\max\left(0, x_{ijk} \sum_{s=1}^{n_k} \rho_{jsk} x_{isk} (s-j) \right) \right)^2 \quad k = 1, \cdots, K$$

$$C_8 = \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} \left(-\min\left(0, x_{ijk} \sum_{s=1}^{n_k} \rho_{jsk} x_{isk} (s-j) \right) \right)^2 \quad k = 1, \cdots, K$$

where β_1, β_2, and β_3 are weight coeficients of the objective function, $r_1,...,r_8$ are penalties which are given since model violate any constraints where $r_1,...,r_8 \to \infty$ will assure that algorithm will satisfy constraint early before considering objective function. Equation (16) becomes the fitness function of DE method. If we assume cost of roster more important than cost of deviation of flying day and more important than open time, then β_1, β_2, and β_3 are selected carefully where $\beta_1 \gg \beta_2 \gg \beta_3$ and assure followed inequality is true :

$$\beta_1 \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} d_{jk} x_{ijk} \gg \beta_2 \sum_{i=1}^{m_k} \left| \sum_{j=1}^{n_k} t_{jk} x_{ijk} - \bar{t}_k \right|^p$$
$$\gg \beta_3 \sum_{i=1}^{m_k} \left(31 - \sum_{j=1}^{n_k} t_{jk} x_{ijk} \right) \quad (17)$$

Term (17) will assure hirarchical ordering of solving iteratively by DE.

4. Solving the Model Using DE

In this paper we applied this model on a real case taken from MNA (Merpati Nusantara Airlines), Ltd., a private airline company based in Indonesia [22,23]. To solve the above problem we pursue the following steps of DE.

4.1. Initialization

Initial solution x_{ijk} is defined by generating n_pop binary

random numbers (0 or 1) of dimension $n_k \times m_k$. Let $X_{np,0}$ be the initial solution population np, then

$$X_{np,0} = rand_{np}[0;1] \quad np = 1..., n\ pop \quad (18)$$

Rand$_{np}$[0;1] is binary random 0 or 1 of population np.

4.2. Mutation

Different from those which usually used as a mutation operator in DE as indicated in Equation (2), this paper introduce a random swap as a mutation operator. Let r_0 be random number between 0 and 1 with $n_k \times m_k$ dimension for each population np, and $v_{np,r0,g}$ be element of solution V at column r_0 at generation g. If $W_{np,g-1}$ is the best population for generation g-1 and $w_{np,r0,g-1}$ is the element at column r_0 of $W_{np,g-1}$, then mutant of generation g is defined as

$$v_{np,r0,g} = \begin{cases} (W_{np,r0,g-1} + 1) \bmod(2), \text{if } r_0 < c_m \\ W_{np,r0,g-1}, \text{otherwise} \end{cases} \quad (19)$$

where c_m is mutation probability ($0 < c_m < 1$) which represents mutation power imposed to the best population of previous generation. Term $W_{np,r0,g-1} \, mod(2)$ means what value left after dividing $W_{np,r0,g-1}$ by 2. The selection of c_m must be done carefully because too small c_m can cause old solution is difficult to exit from local optimum. While, too big c_m causes noise solution such that fast convergence toward global optima can not be achieved. Selecting c_m accurately becomes the successful key of implementing this algorithm.

As an illustration how this mutation operator works, notice the following example. Suppose we have 3 pilots and 4 pairings, use $c_m = 0.2$.

Suppose we have W_0 (current solution):

$$W_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Entry $(i,j) = 1$ indicates a pilot i be placed in pairing j, entry$(i,j) = 0$, otherwise. After generating randomly, suppose we got:

$$r_0 = \begin{bmatrix} 0,10 & 0,40 & 0,08 & 0,70 \\ 0,30 & 0,20 & 0,90 & 0,15 \\ 0,85 & 0,05 & 0,25 & 0,55 \end{bmatrix}$$

To have a new solution, apply Equation (19). By comparing r_0 and c_m for each corresponding entry in W_0 and r_0, we have matrix W'_0, where each entry (typed in bold) should be changed since $r_0 < c_m$.

$$W'_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

The changes should be made as follows

Cell(1,1) $v = (w_0 + 1) \bmod(2) = (1 + 1) \bmod(2) = 2 \bmod(2) = \mathbf{0}$

Cell(1,3) $v = (w_0 + 1) \bmod(2) = (0 + 1) \bmod(2) = 1 \bmod(2) = \mathbf{1}$

Cell(2,4) $v = (w_0 + 1) \bmod(2) = (1 + 1) \bmod(2) = 2 \bmod(2) = \mathbf{0}$

Cell(3,2) $v = (w_0 + 1) \bmod(2) = (0 + 1) \bmod(2) = 1 \bmod(2) = \mathbf{1}$

The other entries of W_0 remains the same as $r_o \geq c_m$. Therefore, we obtain the new solution

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

We follow this process each time the algorithm is on the mutation step.

4.3. Crossover

Crossover changes over parent solution $X_{np,g}$ by mutant solution $V_{np,g}$ to construct a new solution $U_{np,g}$. Crossover is done by defining threshold probability $(0 < c_r < 1)$ for mutant to change current solution. Then we generate *n_pop* random numbers $(0,1)$. If the random number $< c_r$, then the mutant replaces the current solution and otherwise.

$$U_{np,g} = \begin{cases} V_{np,g}, & if\ rand_{np}(0,1) < c_r \\ X_{np,g}, & otherwise \end{cases} \quad (20)$$

4.4. Selection

This process is done by comparing the fitnesses of the parent solution and the new solution which is produced from crossover process. The new population will replace the old population only if the new population has better fitness than that of the old population. The fitness function refers to Equation (16). Then, the solution of the next generation $X_{np,g+1}$ can be obtained from this formula:

$$X_{np,g+1} = \begin{cases} U_{np,g}, & if\ f(U_{np,g}) < f(X_{np,g}) \\ X_{np,g}, & otherwise \end{cases} \quad (21)$$

where $f(U_{np,g})$ and $f(X_{np,g})$ are the fitness of $U_{np,g}$ and $X_{np,g}$ respectively.

The constructed mathematical model consists of some aspirations/criterias as the objective function and some constraints. The objective function includes minimum flying times, deviation of flying days, and open time.

Some constraints which are considered when constructing a roster include overlap, crew requirement of pairing, free days before seven days of flying days, maximum flying times, and maximum numbers of take off.

5. Experiments and Analysis

We used Matlab to implement DE algorithm. The experiments were done using the datasets shown in **Table 1**. The datasets consist of pairings, numbers of crews, type of fleed and rules. Datasets are divided into two categories: small and large datasets. Small dataset consists of assignment of F-100, CN-235, DHC-6, and Cassa 212 pilot. While, large dataset consists of assignment of Boeing 737-200 pilot and stewardess.

The experiments aim to assign crew members in which pairings. The results of the experiments on small datasets were compared with column generation [22,23] and MOSI (method used by the company). For large datasets we included exact decomposition as a comparative method [23].

Probability of mutation (c_m) and probability of crosser (c_r) are the key succes factors for DE implementation. Therefore, these two parameters should be determined carefully. The precise parameter values will lead to the global optimal solution. In this paper, these parameters settings were done through trial process. The best c_m, is determined by using one dataset (Cassa 212) with *npop* = 50, *gmax* = 50, $c_r = 0.7$ and the experiments were done with 5 replications. **Table 2** shows the objective function values for different c_m. The best value is typed in bold. We found that $c_m = 0.05$ is the best one.

Using $c_m = 0.05$, the same procedure was done to find the best c_r. The results are shown in **Table 3**.

Next, using combination of c_m and c_r we tried to determine the best values of these two parameters. Through experiments, we obtained the best combination between c_m and c_r are 0.1 and 0.5 as shown in bold in **Table 4**.

Criteria in the objective function are weighted based on their significances. In this paper we put roster cost minimization as the most significance followed by deviation of minimum flying days and open time with weights of 10^4, 10^2 and 1 respectively. Other parameter needs to set up is pinalty for the constraints. Penalties for overlap and rotation without free day, number of pilot requirements and day off constraints are 10^{15}, 10^{13} and 10^{11} respectively based on their signifances. While, the penalties for flying day, flying times, and numbers of take off are set to 10^6 as these constraints are assumed to be the same important.

Experimental results for small datasets are indicated in **Table 5**. For the flying time criterion, DE spent more

Table 1. Characteristic of the datasets.

No	Name of aircraft	Type of crews	Numbers of crews	Numbers of pairings
1	F-100	Pilot	5	4
2	CN-235	Pilot	4	8
3	DHC-6	Pilot	3	10
4	Cassa-212	Pilot	6	13
5	Boeing 737-200	Pilot	17	82
6	Boeing 737-200	stewardess	55	114

Table 2. Average objective function values on different c_m.

No	c_m	Average objective function value ($\times 10^{13}$)
1	0.000	9,980
2	0.025	41.00
3	**0.050**	**3.00**
4	0.075	22.60
5	0.100	7.60
6	0.125	48.20
7	0.150	122.20
8	0.175	140.80
9	0.200	163.40

Table 3. Average objective function values on different c_r.

No	c_r	Average objective function value ($\times 10^{13}$)
1	0.300	22.00
2	0.350	43.20
3	**0.400**	**21.80**
4	0.450	80.60
5	0.500	42.60
6	0.550	41.00
7	0.600	22.60
8	0.650	40.60
9	0.700	23.40

time than two other methods. The reason is because DE did not violate pilot requirement constraint. This is different with column generation and MOSI which violated pilot requirement constraint. It shows that DE, column generation, and MOSI produced the same performance to meet the roster constraints. Except for Cassa

212, DE can meet all of the constraints. While, column generation and MOSI violate pilot requirements constraints. Column generation and MOSI just assign 1 out of 2 required pilots to pairing 8981. From roster quality side, DE is only inferior for flying times criterion for Cassa 212 aircraft. The other methods show the same results for other assignments.

Table 6 shows the total deviation of average flying days for three methods. We see that DE is superior for all assignments except for Cassa 212. This proves that DE produced good roster quality. From **Table 6**, we see that DE produced better deviation of flying days than other methods for three assignments and MOSI is superior for Cassa 212.

Table 4. Average objective function values on different combination of c_m and c_r.

No	c_m	c_r	Average objective function value ($\times 10^{13}$)
1	0.05	0.400	21.80
2	0.05	0.5	42.60
3	0.05	0.6	22.60
4	0.05	0.7	3.00
5	0.1	0.4	22.20
6	**0.1**	**0.5**	**2.40**
7	0.1	0.6	23.20
8	0.1	0.7	7.60

Tabel 5. Comparison of flying time.

Name of Aircraft	Fying time		
	DE	Column Generation	MOSI
F-100	66.0	66.0	66.0
CN-235	92.0	92.0	92.0
DHC-6	156.0	156.0	156.0
Cassa 212	251.0	**242.0**	**242.0**

Tabel 6. Comparison of deviation of flying days.

Name of Aircraft	Fying time		
	DE	Column Generation	MOSI
F-100	**7.2**	14.4	14.4
CN-235	**2.0**	12.0	4.0
DHC-6	**8.0**	10.7	11.3
Cassa 212	11.3	21.0	**6.0**

       AM

For the open time criteria, as shown in **Table 7**, DE produced superior results compared to the other methods for Cassa 212 and the same results for other assignments.

Dataset of pilots and stewardess assignment on Boeing 737-200 aircraft has a larger size. This dataset consists of 17 pilots with 82 pairings and 55 stewardess with 114 pairings. It required high number of iterations and computational time to obtain near optimal solution if random initial solutions were used. We proposed a sequence of partial optimization and total optimization techniques. In partial optimization, the dataset is splitted into several smaller sets and the corresponding optimization problems were solved by DE method. At the end of this stage, all of solutions were combined all together to form initial solution for the total optimization problem. Setting the initial solution through this process decreased the computational time significantly.

For the pilot assignment of Boeing 737-200, number of pairings is much higher than the available pilots. Therefore the resulting flight schedules violate several constraints. The results for this assignment are shown in **Table 8**. The table presents the flying days and the actual difference produced by each pilot. In this case, we compared 6 methods namely differential evolution (DE), column generation, MOSI, exact decomposition with 21 flying days (DEC21), exact decomposition with 20 flying days (DEC20) and exact decomposition with 19 flying days (DEC19). The results are presented in **Table 8**.

Table 7. Comparison of open time.

Name of Aircraft	Open time		
	DE	Column Generation*	MOSI*
F-100	137.0	137.0	137.0
CN-235	88.0	88.0	88.0
DHC-6	50.0	50.0	50.0
Cassa 212	**119.0**	123.0	123.0

Table 8. Comparison flying days and differences.

Pilot	Differential Evolution		Column Generation*		MOSI*		Method DEC21**		Method Dec20		Method Dec19	
	flying day	diff.	flying day	diff.	flying day	diff.	flying day	diff.	Fying days	diff	Fying daya	diff
1	21	0	20	0	21	0	20	0	21	0	22	1
2	20	0	18	0	21	0	22	**1**	23	**2**	22	**1**
3	20	0	20	0	23	2	26	**5**	23	**2**	24	**3**
4	21	0	20	0	22	**1**	24	**3**	23	**2**	23	**2**
5	21	0	17	0	18	0	17	0	18	0	17	0
6	21	0	20	0	22	**1**	24	**3**	23	**2**	23	**2**
7	22	**1**	21	0	16	0	18	0	17	0	20	0
8	21	0	26	**5**	21	0	22	**1**	24	**3**	22	**1**
9	20	0	16	0	17	0	20	0	18	0	19	0
10	18	0	18	0	16	0	13	0	16	0	14	0
11	22	**1**	23	**2**	22	**1**	19	0	19	0	19	0
12	21	0	21	0	21	0	22	**1**	22	**1**	21	**0**
13	20	0	24	**3**	24	**3**	22	**1**	21	**0**	20	**0**
14	21	0	23	**2**	24	**3**	20	0	20	0	22	1
15	20	0	20	0	19	0	21	0	21	0	21	0
16	20	0	22	**1**	21	0	21	0	21	0	22	1
17	22	**1**	22	**1**	23	2	18	0	18	0	17	0
Total	351	3	351	14	351	13	349	15	348	12	348	12
Avg Std.D	20.65 0.99		20.65 2.57		20.65 2.57		20.53 3.04		20.47 2.43		20.47 2.40	
Number of pilot violate flying. days	3		6		7		7		8		8	

We see from the table that all of the methods violated flying days constraint. DE assigned smoother flying days than other methods based on the standard deviation values. DE reached the smallest standard deviation. We also recorded that DE produced the smallest number of pilots violating flying days constraint, that is 3 pilots. While, column generation, MOSI, DEC21, DEC20 and DEC19 respectively produced 6, 7, 7, 8 and 8 pilots violating the constraint.

Different from those of pilot assignments, in stewardess assignment of Boeing 737-200, we compared DE, column generation and MOSI. **Table 9** shows that DE only violated crew requirement constraint while column generation and MOSI violated both crew requirement and flying days constraints. We can see in **Table 9** that column generation and MOSI assigned 6 and 7 stewardesses exceeding the maximum flying days, 21 days.

While, in the pilot assignment of Boeing 737-200, as shown in **Table 10**, DE is superior for minimum deviation of flying days.

In the assignment of Boeing 737-200 stewardess, DE is superior by its minimum deviation of flying days compared to other methods. But, DE is inferior for flying time and open time criteria.

6. Conclusions

We have investigated the use of DE in solving airline

Tabel 9. Violation of flying days constraint for assignment of Boeing 737-200 stewardess.

Constraints	Method		
	DE	Column Gen.	MOSI
Flying days	-	√	√
Flying times	-	-	-
Take off	-	-	-
Overlap	-	-	-
Free day	-	-	-
Requirement of pilots	√	√	√

Tabel 10. Roster quality of pilot assignment of Boeing 737-200.

Criteria	Methods			
	DE	Column Gen.*	MOSI*	DEC21**
Flying times	1,114.0	1,114.0	1,114.0	1,114.0
Dev. of fly. days	**13.1**	33.6	34.5	38.5
Open time	176.0	176.0	176.0	178.0

Tabel 11. Roster quality of stewardess assignment of Boeing 737-200.

Criteria	Methods		
	DE	Column Generation	MOSI*
Flying times	3,488.0	**3,285.0**	**3,285.0**
Dev. of fly. days	**92.0**	110.0	106.2
Open time	693.0	**658.0**	**658.0**

crew scheduling problem. Generally, the rostering problem in MNA has characteristic indifferent from other airline companies in terms of roster constraint and roster quality. Selecting mutation and crossover probability accurately became the successful key to implement DE. The best mutation and crossover probability are 0.1 and 0.5 respectively. Different from using DE in general, in this paper we introduced random swap as mutation operator. For small datasets, DE was proven more competitive then column generation and MOSI, based on constraints fulfillment and roster quality. In the assignment of Boeing 737 pilot, DE produced smoother flying days and the least pilots which violated flying days constraint compared to other methods. For the stewardess assignment of Boeing 737-200, DE violated the least constraints compared to column generation and MOSI. DE produced superior deviation of flying days criterion but it is still inferiror for iwo other criteria.

7. References

[1] R. Anbil, J. J. Forrest and W. R. Pulleyblanck, "Column Generation and the Airline Crew Pairing Problem," Documentation Mathematica Extra, Journal der Deutschen Mathematiker Vereinigung Volume ICM, III, 1998, pp. 677-686.

[2] C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh and P. H. Vance, "Branch and Price: Column Generation for Solving Huge Integer Programs," *Operation Research*, Vol. 46, No. 3, 1998, pp. 316-329.

[3] I. Gershkoff, G. W. Graves, R. D. Mc Bridge, D. Anderson and D. Mahidhara, "Flight Crew Scheduling," *Management Science*, Vol. 39, No. 6, 1993, pp. 736-745.

[4] K. L. Hoffman and M. Padberg, "Solving Airline Crew Scheduling Problems by Branch and Cut," *Management Science*, Vol. 39, No. 6, 1993, pp. 657-682.

[5] N. Souai, and J. Thegem, "Genetic Algorithm Based Approach for the Integrated Airline Crew-Pairing and Rostering Problem," *European Journal of Operational Research*, Vol. 199, No. 3, 2009, pp. 674-683.

[6] N. Kohl, "Application of OR and CP Techniques in a Real World Crew Scheduling System," *In Proceedings of CP-AI-OR*'00: 2*nd International Workshop on Integration of AI and OR Techniques in Constraint Program-*

ming for Combinatorial Optimization Problems, Paderborn, Germany, 8-10 March 2000, pp. 105-108.

[7] N. Kohl and S. E. Karisch, "Integrating Operations Research and Constraint Programming Techniques in Crew Scheduling," *In Proceedings of the* 40*th Annual AGIFORS Symposium*, Istanbul, Turkey, 20-25 August 2000.

[8] S. C. K. Chu, "Generating, scheduling, and Rostering of Shift Crew-Duties: Applications at the Hongkong International Airport," *European Journal of Operational Research*, Vol. 177, No. 3, 2007, pp. 1764-1778.

[9] C. P. Medard and N. Sawhney, "Airline Crew Scheduling from Planning to Operation," *European Journal of Operational Research*, Vol. 183, No. 3, 2005, pp. 1013-1027.

[10] P. H. Vance, C. Barnhart, E. L. Johnson and G. L. Nemhauser, "Airline Crew Scheduling: A New Formulation and Decomposition Algorithm," *Operation research*, Vol. 45. No. 2, 1997, pp. 188-200.

[11] P. Lučić, and D. Teodorović, "Simulated Annealing for the Multi-Objective Aircrew Rostering Problem," *Transportation Research Part A*, Vol. 33, No. 1, 1999, pp. 19-45.

[12] D. Levine, "Application of a Hybrid Genetic Algorithm to Airline Crew Scheduling," *Computer Operations Research*, Vol. 23, No. 6, 1996, pp. 547-558.

[13] J. E. Beasly and B. Cao, "A Tree Search Algorithm for the Crew Scheduling Problem," *European Journal of Operational Research*, Vol. 94, No. 3, 1996, pp. 517-526.

[14] Z. Yinghui, R. Yunbao and Z. Mingtian, "GASA Hybrid Algorithm Applied in Airline Crew Rostering System," *Tsinghua Science and Thechnology*, Vol. 12, No. S1, 2007, pp. 225-259.

[15] R. Storn and K. Price, "Differential Evolution: A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces," Technical Report TR-95-012, International Computer Science Institute, 1995.

[16] V. P. Kennet, M. S. Rainer and A. L. Jouni, "Differential Evolution: A Practical Approach to Global Optimization," Springer-Verlag, Berlin Heidelberg, 2005.

[17] M. F. Tasgetiren, Q. K. Pan, Y. C. Liang and P. N. Suganthan, "A Discrete Differential Evolution Algorithm for the Total Earliness and Tardiness Penalties with a Common Due Date on Single Machine," *Proceedings of the* 2007 *IEEE Symposium on Computational Intelligence in Scheduling*, 2007, pp. 271-278.

[18] L. Jian, P. Chen and Z. M. Liu, "Solving Traveling Salesman Problems by Genetic Differential Evolution with Local Search," Workshop on Power Electronics and Intelligent Transportation System, 2008

[19] M. G. H. Omran and A. Salman, "Constrained Optimization Using CODEQ," *Chaos, Solitons and Fractals*, Vol. 42, No. 2, 2009, pp. 662-668.

[20] R. L. Fox, "Optimization Methods for Engineering Design," Addison-Wesley, Reading, Massachusetts, 1971.

[21] J. H. Cassis and L. A. Schmit, "On Implementation of the Extended Interior Penalty Function," *International Journal for Numerical Methods in Engineering*, Vol. 10, No. 1, 1976, pp. 3-23.

[22] Z. Labiba, "Aplikasi metode column generation dalam penyelesaian penugasan kru maskapai penerbangan: studi kasus di PT. Merpati Nusantara Airlines," Tesis magister teknik, Jurusan Teknik Industri ITS, Surabaya, (in Bahasa Indonesia), 2006.

[23] A. Rusdiansyah, Y. D. Mirenani and Z. Labiba, "Pemodelan dan penyelesaian permasalahan penjadwalan pilot dengan metode eksak dekomposisi," *Jurnal Teknik Industri*, Vol. 9, No. 2, Desember 2007, pp. 112-124.

Scientific
Research

Effect of Non-Homogeneity on Thermally Induced Vibration of Orthotropic Visco-Elastic Rectangular Plate of Linearly Varying Thickness

Arun Kumar Gupta, Pooja Singhal
Department of Mathematics, M.S. College, Saharanpur, Uttar Pradesh, India
E-mail: gupta_arunnitin@yahoo.co.in, poojaacad@yahoo.in
Received August 3, 2010; revised August 30, 2010; accepted September 3, 2010

Abstract

The analysis presented here is to study the effect of non-homogeneity on thermally induced vibration of orthotropic visco-elastic rectangular plate of linearly varying thickness. Thermal vibrational behavior of non-homogeneous rectangular plates of variable thickness having clamped boundary conditions on all the four edges is studied. For non-homogeneity of the plate material, density is assumed to vary linearly in one direction. Using the method of separation of variables, the governing differential equation is solved. An approximate but quite convenient frequency equation is derived by using Rayleigh-Ritz technique with a two-term deflection function. Time period and deflection at different points for the first two modes of vibration are calculated for various values of temperature gradients, non-homogeneity constant, taper constant and aspect ratio. Comparison studies have been carried out with non-homogeneous visco-elastic rectangular plate to establish the accuracy and versatility.

Keywords: Non-Homogeneous, Orthotropic, Visco-Elastic, Variable Thickness, Rectangular Plate, Vibration, Thermal Gradient

1. Introduction

Thermal effect on vibration of non-homogenous viscoelastic plates are of great interest in the field of engineering such as for better designing of gas turbines, jet engine, space craft and nuclear power projects, where metals and their alloys exhibits visco-elastic behavior. Therefore, for these reason such structures are exposed to high intensity heat fluxes and thus material properties undergo significant changes, in particular the thermal effect on the modules of elasticity of the material can not be taken as negligible.

Plates of variable thickness have been extensively used in Civil, Electronic, Mechanical, Aerospace and Marine Engineering applications. The practical importance of such plates has made vibration analysis essential especially since the vibratory response needs to be accurately determined in design process in order to avoid resonance excited by internal or external forces.

Visco-elastic, as its name implies, is a generalization of elasticity and viscosity. The ideal linear elastic element is the spring. When a tensile force is applied to it,

the increase in the distance between its two ends is proportional to the force. The ideal linear viscous element is the dashpot.

The plate type structural components in aircraft and rockets have to operate under elevated temperatures that cause non-homogeneity in the plate material *i.e.* elastic constants of the materials becomes functions of space variables. In an up-to-date survey of literature, authors have come across various models to account for non-homogeneity of plate materials proposed by researchers dealing with vibration but none of them consider non-homogeneity with thermal effect on orthotropic visco-elastic plates.

Free vibration of visco-elastic orthotropic rectangular plates was discussed by Sobotka [1].Gupta and Khanna [2] discussed vibration of viscoelastic rectangular plate with linearly thickness variations in both directions. Leissa's monograph [3] contains an excellent discussion of the subject of vibrating plates with elastic edge support. Several authors [4,5] have studied the thermal effect on vibration of homogeneous plates of variable thickness but no one considered thermal effect on vibra-

tion on non-homogeneous rectangular plates of varying thickness. Tomar and Gupta [6-8] solved the vibration problem of orthotropic rectangular plate of varying thickness subjected to a thermal gradient. Gupta, Lal and Sharma [9] discussed the vibration of non-homogeneous circular plate of nonlinear thickness variation by a quadrature method. Gupta, Johri and Vats [10] solved the problem of thermal effect on vibration of non-homogeneous orthotropic rectangular plate having bi-directional parabolically varying thickness. Gupta, Kumar and Gupta [11] studied the vibration of visco-elastic orthotropic parallelogram plate with a linear variation of thickness. Recently, Gupta and Kumar [12] solved the vibration problem of non-homogeneous visco-elastic rectangular plate of linearly varying thickness subjected to linearly thermal effect. Free vibration of a clamped visco-elastic rectangular plate having bi-direction exponentially varying thickness were studied by Gupta, Khanna and Gupta [13] .Gupta, Aggarwal, Gupta, Kumar and Sharma [14] discussed the non-homogeneity on free vibration of orthotropic visco-elastic rectangular plate of parabolic varying thickness. Subsequent review article of Bhasker and Kaushik [15] are best source for problems involving rectangular plates fall into three distinct categories: 1) plates with all edges simply supported; 2) plates with a pair of opposite edges simply supported; 3) plates which do not fall into any of the above categories.

Rectangular plates have wide applications in civil structures, electrical engineering, marine industry and mechanical engineering. The dynamic characteristics of rectangular plates are important to engineering designs. An analysis is presented in this paper is to study the effect of non-homogeneity on thermally induced vibration of orthotropic visco-elastic rectangular plate of linearly varying thickness. It is clamped supported on all the four edges. The assumption of small deflection and linear orthotropic visco-elastic properties are made. It is further assumed that the visco-elastic properties of the plates are of the Kelvin type. For this the material constants of alloy 'Duralium' is used for the calculation of numerical values. Time period and deflection for the first two modes of vibration are calculated for the various values of thermal gradients, non-homogeneity constant, aspect ratio and taper constant.

2. Analysis

The equation of motion of a visco-elastic orthotropic rectangular plate of variable thickness may be written in the form, as by Sobotka [1]

$$\frac{\partial^2 M_x}{\partial x^2} + 2\frac{\partial^2 M_{xy}}{\partial x \partial y} + \frac{\partial^2 M_y}{\partial y^2} = \rho h \frac{\partial^2 w}{\partial t^2} \quad (1)$$

Here M_x, M_y and M_{xy} are moments per unit length of plate, ρ is mass per unit volume, h is thickness of plate and w is displacement at time t.

The expression for M_x, M_y, M_{xy} are given by

$$M_x = -\tilde{D}\left[D_x \frac{\partial^2 w}{\partial x^2} + D_1' \frac{\partial^2 w}{\partial y^2}\right]$$

$$M_y = -\tilde{D}\left[D_1' \frac{\partial^2 w}{\partial x^2} + D_y \frac{\partial^2 w}{\partial y^2}\right] \quad (2)$$

$$M_{xy} = -2\tilde{D}D_{xy}\frac{\partial^2 w}{\partial x \partial y}$$

where

$$D_x = \frac{E_x h^3}{12(1 - v_x v_y)}$$

is called the flexural rigidity of the plate in *x*-direction,

$$D_y = \frac{E_y h^3}{12(1 - v_x v_y)}$$

is called the flexural rigidity of the plate in *y*-direction,

$$D_{xy} = \frac{G_{xy} h^3}{12}$$

is called the torsion rigidity, and

$$D_1' = v_x D_y \ (= v_y D_x)$$

Here $\tilde{D}$ is Rheological operator and $E_x \& E_y$ are the modules of elasticity in *x*- and *y*-direction respectively, v_x and v_y are the Poisson ratios & G_{xy} is the shear modulus.

Taking deflection w as a product of two functions as:

$$w = w(x, y, t) = W(x, y)T(t) \quad (3)$$

where *W*(*x*, *y*) is the function of coordinates in *x*, *y* and *T*(*t*) is a time function.

Using Equation (3) in Equations (1) & (2) and then equating both sides of equation comes to a constant, say p^2, one gets two separate differential equations as follows:

$$\left[D_x \frac{\partial^4 W}{\partial x^4} + D_y \frac{\partial^4 W}{\partial y^4} + 2H \frac{\partial^4 W}{\partial x^2 \partial y^2} + 2\frac{\partial H}{\partial x}\frac{\partial^3 W}{\partial x \partial y^2}\right.$$

$$+ 2\frac{\partial H}{\partial y}\frac{\partial^3 W}{\partial x^2 \partial y} + 2\frac{\partial D_x}{\partial x}\frac{\partial^3 W}{\partial x^3} + 2\frac{\partial D_y}{\partial y}\frac{\partial^3 W}{\partial y^3}$$

$$+ \frac{\partial^2 D_x}{\partial x^2}\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 D_y}{\partial y^2}\frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 D_1'}{\partial x^2}\frac{\partial^2 W}{\partial y^2} \quad (4)$$

$$\left. + \frac{\partial^2 D_1'}{\partial y^2}\frac{\partial^2 W}{\partial x^2} + 4\frac{\partial^2 D_{xy}}{\partial x \partial y}\frac{\partial^2 W}{\partial x \partial y}\right] - \rho h p^2 W = 0$$

and

$$\ddot{T} + p^2 \tilde{D} T = 0 \qquad (5)$$

where

$$H = D_1^{'} + 2D_{xy}$$

Equation (4) is a differential equation of motion for orthotropic rectangular plate of variable thickness and (5) is a differential equation of time functions of free vibration of viscoelastic rectangular orthotropic plate.

The temperature dependence of the modulus of elasticity for orthotropic materials is given by

$$E_x = E_1 \left(1 - \gamma\tau\right)$$
$$E_y = E_2 \left(1 - \gamma\tau\right) \qquad (6)$$
$$G_{xy} = G_0 \left(1 - \gamma\tau\right)$$

and temperature distribution along the length *i.e.* in the x-direction,

$$\tau = \tau_o \left(1 - x / a\right) \qquad (7)$$

where τ denotes the temperature excess above the reference temperature at any point at distance x/a and τ_0 denotes the temperature excess above reference temperature at the end, *i.e.* for $x = 0$. Here E_1 and E_2 are values of the Young's moduli respectively along the x and y axis at the reference temperature *i.e.* at $\tau = 0$

The modulus variation (6) in view of expressions (7) becomes

$$E_x\left(x\right) = E_1 \left[1 - \alpha\left(1 - x/a\right)\right]$$
$$E_y\left(x\right) = E_2 \left[1 - \alpha\left(1 - x/a\right)\right] \qquad (8)$$
$$G_{xy}\left(x\right) = G_0 \left[1 - \alpha\left(1 - x/a\right)\right]$$

where $\alpha = \gamma\, \tau_o\, (0 \leq \alpha < 1)$, known as thermal gradient.

The expression for the strain energy V and Kinetic energy P in the plate are [3]

$$\frac{1}{2}\int_0^a \int_0^b \left[D_x (W_{,xx})^2 + D_y (W_{,yy})^2 + 2D_1 W_{,xx} W_{,yy} \right. \\ \left. + 4D_{xy}(W_{,xy})^2 \right] dxdy \qquad (9)$$

$$P = \frac{1}{2} p^2 \int_0^a \int_0^b \rho h W^2 dx dy \qquad (10)$$

The thickness and density varies linearly in the x-direction only, so let us assume

$$h = h_o \left(1 + \beta\, x/a\right) \qquad (11)$$

and $\rho = \rho_o \left(1 + \alpha_1 x/a\right) \qquad (12)$

where β is the taper constant and α_1 is non-homogeneity constant.

3. Solution and Frequency Equation

To find the solution, we use Rayleigh-Ritz technique. In this method, one requires maximum strain energy be equal to the maximum Kinetic energy. So, it is necessary for the problem under consideration that

$$\delta\left(V - P\right) = 0 \qquad (13)$$

for arbitrary variations of W satisfying relevant geometrical boundary conditions are

$$W = W_{,x} = 0 \text{ at } x = 0, a \\ W = W_{,y} = 0 \text{ at } y = 0, b \qquad (14)$$

and the corresponding two term deflection function is taken as [6]

$$W = \left[\left(x/a\right)\left(y/b\right)\left(1 - x/a\right)\left(1 - y/b\right)\right]^2 \\ \left[A_1 + A_2\left(x/a\right)\left(y/b\right)\left(1 - x/a\right)\left(1 - y/b\right)\right] \qquad (15)$$

The non-dimensional variables are

$$X = x/a, \ Y = y/b, \ \overline{W} = W/a, \ \overline{h} = h/a, \\ \overline{\rho} = \rho/a \qquad (16)$$

$$E_1^* = E_1 / \left(1 - v_x v_y\right), \ E_2^* = E_2 / \left(1 - v_x v_y\right) \\ E^* = v_x E_2^* = v_y E_1^* \qquad (17)$$

By using Equations (8), (11) and (12) in (9) and (10), one gets

$$P = \frac{1}{2}\rho_o p^2 \overline{h_o} a^5 \int_0^1 \int_0^{b/a} [(1 + \alpha_1 X)(1 + \beta X)\overline{W}^2]dXdY) \qquad (18)$$

and

$$V = R\int_0^1 \int_0^{b/a} [\{1 - \alpha(1 - X)\}(1 + \beta X)^3 \\ \{(\overline{W}_{,XX})^2 + (E_2^* / E_1^*)(\overline{W}_{,YY})^2 + \\ (2E^* / E_1^*)\overline{W}_{,XX}\overline{W}_{,YY} + \\ (4G_o / E_1^*)(\overline{W}_{,XY})^2\}]dXdY \qquad (19)$$

where $R = \frac{1}{2}(E_1^* \overline{h_o}^3 / 12)a \qquad (20)$

On substituting the values of P and V from Equations (18) and (19) in Equation (13), we get

$$\left(V_1 - \lambda^2 p^2 P_1\right) = 0 \qquad (21)$$

$$V = R\int_0^1 \int_0^{b/a} [\{1 - \alpha(1 - X)\}(1 + \beta X)^3 \\ \{(\overline{W}_{,XX})^2 + (E_2^* / E_1^*)(\overline{W}_{,YY})^2 + \\ (2E^* / E_1^*)\overline{W}_{,XX}\overline{W}_{,YY} + \\ (4G_o / E_1^*)(\overline{W}_{,XY})^2\}]dXdY \qquad (22)$$

$$P_1 = \int_0^1 \int_0^{b/a} \left[\{(1 + \alpha_1 X)(1 + \beta X)\overline{W}^2 \} \right] dXdY \qquad (23)$$

where $\quad \lambda^2 = \dfrac{12a^4 \rho_o}{E_1^* \overline{h_o}^2} \qquad (24)$

Equation (21) involves the unknown A_1 and A_2 arising due to the substitution of $W(x,y)$ from Equation (15). These two constants are to be determined from Equation (21) as follows:

$$\frac{\partial}{\partial A_n}(V_1 - \lambda^2 p^2 P_1) = 0 \qquad (25)$$

where $n = 1,2$

On simplifying (25) we get

$$b_{n1}A_1 + b_{n2}A_2 = 0 \qquad (26)$$

where $n = 1, 2$, b_{n1}, b_{n2} involves parametric constants and the frequency parameter p. For a non-trivial solution, the determinant of the coefficient of Equation (26) must be zero. So, we get the frequency equation as

$$\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = 0 \qquad (27)$$

On solving Equation (27) one gets a quadratic equation in p^2, which gives two values of p^2. On substituting the value of $A_1 = 1$, by choice, in Equation (15) one get $A_2 = -b_{11}/b_{12}$ and hence W becomes:

$$W = \left[XY \frac{a}{b}(1 - X)\left(1 - Y\frac{a}{b}\right) \right]^2$$
$$\left[1 + \left(-\frac{b_{11}}{b_{12}}\right) XY\left(\frac{a}{b}\right)(1 - X)\left(1 - Y\frac{a}{b}\right) \right] \qquad (28)$$

4. Time Functions of Vibration of Viscoelastic Plates

The expression for Time function of free vibrations of visco-elastic plates of variable thickness can be derived from Equation (5) that depends upon visco-elastic operator $\tilde{D}$ and which for Kelvin's Model can be taken as:

$$\tilde{D} \equiv \left\{ 1 + \left(\frac{\eta}{G}\right)\left(\frac{d}{dt}\right) \right\} \qquad (29)$$

where η is visco-elastic constant and G is shear modulus. Taking temperature dependence of viscoelastic constant η and shear modulus G is the same form as that of Young's moduli, we have

$$G(\tau) = G_0(1 - \gamma_1\tau),$$
$$\eta(\tau) = \eta_0(1 - \gamma_2\tau) \qquad (30)$$

where G_0 is shear modulus and η_0 is visco-elastic constant at some reference temperature *i.e.* at $\tau = 0$, γ_1 and γ_2 are slope variation of τ with G and η respectively. Substituting the value of τ from Equation (7) and using Equation (16) in Equation (30), one gets:

$$G = G_0[1 - \alpha_5(1 - X)] \qquad (31)$$

where $\alpha_5 = \gamma_1\tau_0$, $0 \le \alpha_5 < 1$
and $\eta = \eta_0[1 - \alpha_4(1 - X)]$
where $\alpha_4 = \gamma_2\tau_0$, $0 \le \alpha_4 < 1$

Here α_4 and α_5 are thermal constants.

Substituting Equations (29) and (31) in Equation (5), one gets:

$$\ddot{T} + p^2 k \dot{T} + p^2 T = 0 \qquad (32)$$

Where $\quad k = \dfrac{\eta}{G} = \dfrac{\eta_0[1 - \alpha_4(1 - X)]}{G_0[1 - \alpha_5(1 - X)]} \qquad (33)$

Equation (32) is a second order differential equation in time function T. The solution of which comes out to be

$$T(t) = e^{a_1 t}[C_1 \cos b_1 t + C_2 \sin b_1 t] \qquad (34)$$

where $a_1 = -p^2 k/2$,

$$b_1 = p\sqrt{1 - \left(\frac{pk}{2}\right)^2} \qquad (35)$$

where C_1 and C_2 are constants of integration, which can be determined easily from initial conditions of the plate.

Let us take initial conditions as

$$T = 1 \text{ and } \dot{T} = 0 \text{ at } t = 0 \qquad (36)$$

Using Equation (36) in Equation (34), we have $C_1 = 1$ and

$$C_2 = p^2 k/2b_1 = -a_1/b_1 \qquad (37)$$

Using Equation (37) in Equation (34), one has

$$T(t) = e^{a_1 t}[\cos b_1 t + (-a_1/b_1)\sin b_1 t] \qquad (38)$$

Thus deflection w may be expressed by using Equation (38) and Equation (28) in Equation (3)

$$w = \left[XY \frac{a}{b}(1 - X)\left(1 - Y\frac{a}{b}\right) \right]^2$$
$$\left[1 + \left(-\frac{b_{11}}{b_{12}}\right) XY\left(\frac{a}{b}\right)(1 - X)\left(1 - Y\frac{a}{b}\right) \right] \qquad (39)$$
$$\times e^{a_1 t}[\cos b_1 t + (-a_1/b_1)\sin b_1 t]$$

Time period of the vibration of the plate is given by

$$K = 2\pi/P \qquad (40)$$

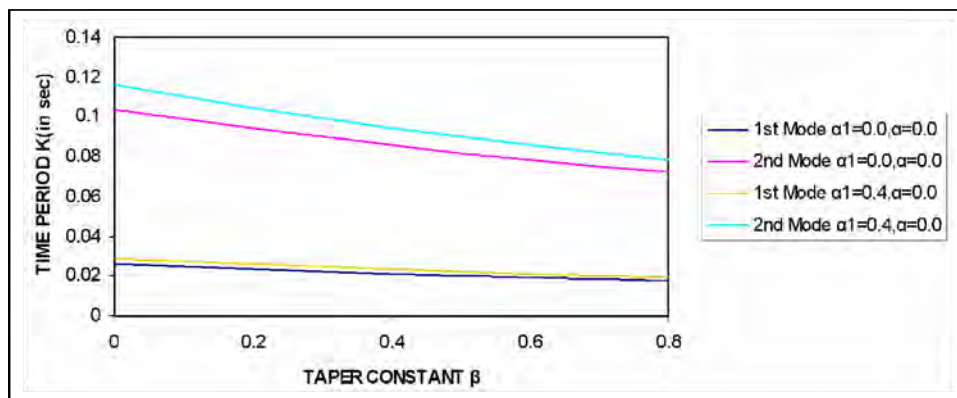where p is the frequency given by Equation (27).

Figure 1. Variation of time period with taper constant of visco-elastic non homogeneous rectangular plate of linearly varying thickness.
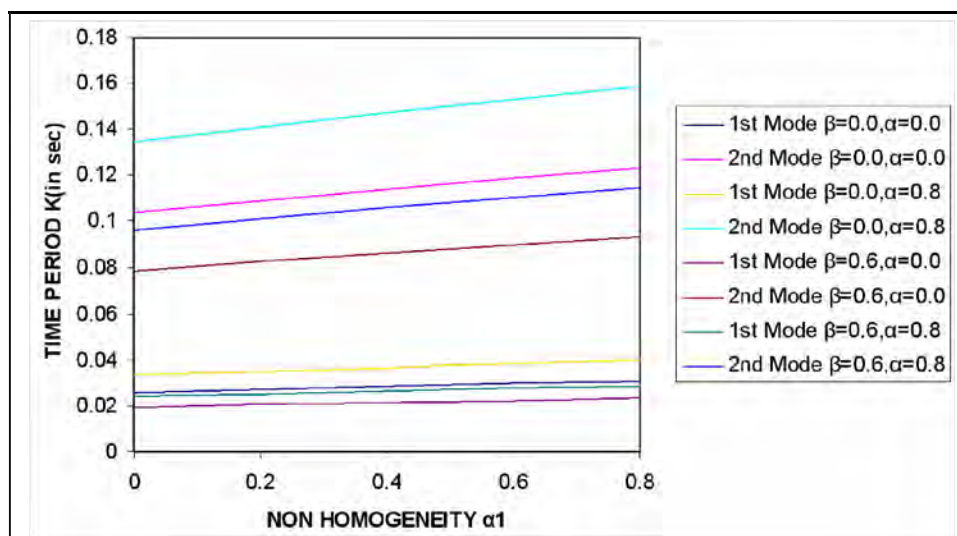

Figure 2. Variation of time period with non homogeneity constant of visco-elastic non homogeneous rectangular plate of linearly varying thickness.
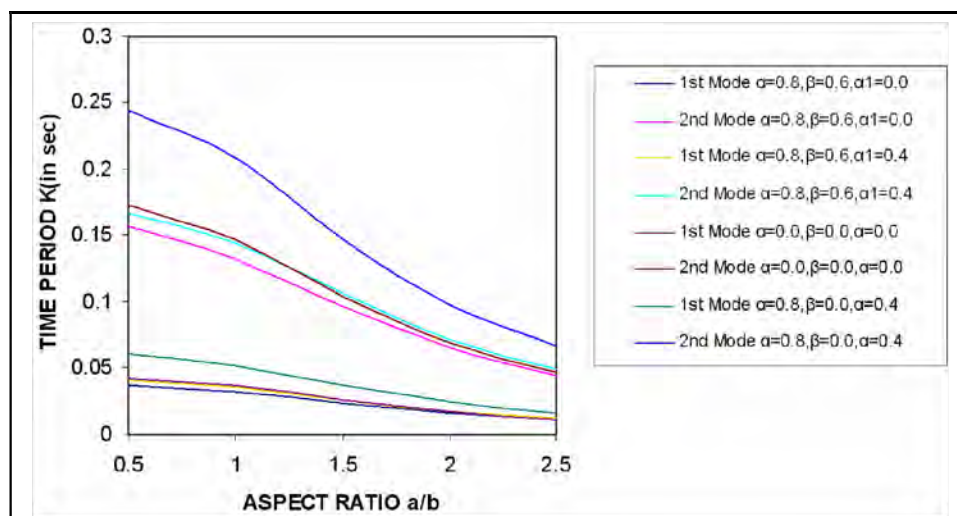

Figure 3. Variation of time period with aspect ratio of visco-elastic non homogeneous rectangular plate of linearly varying thickness.
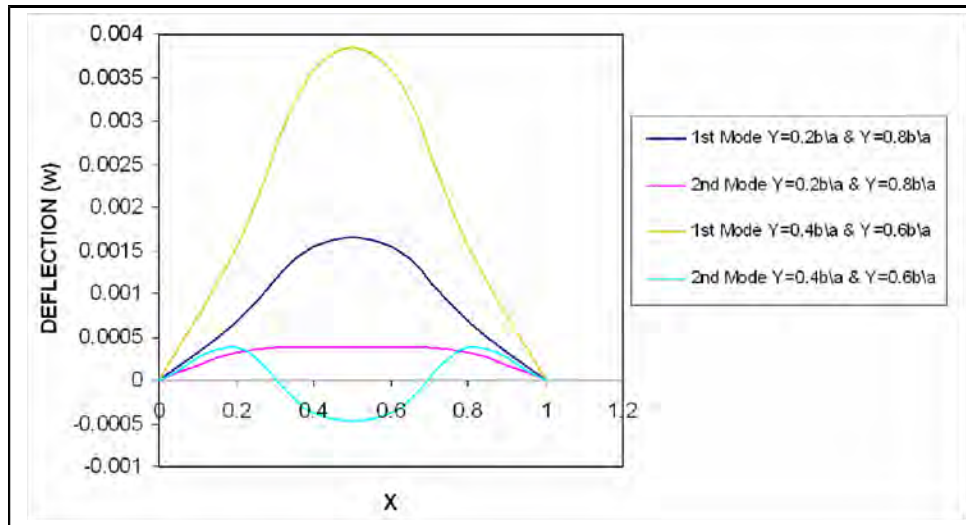
AM

Figure 4. Transverse deflection w Vs X of visco-elastic non homogeneous rectangular plate of linearly varying thickness at initial time 0.K having constants combination as $\alpha = 0.0$, $\beta = 0.0$, $\alpha_1 = 0.0$, $\alpha_4 = 0.3$, $\alpha_5 = 0.2$.
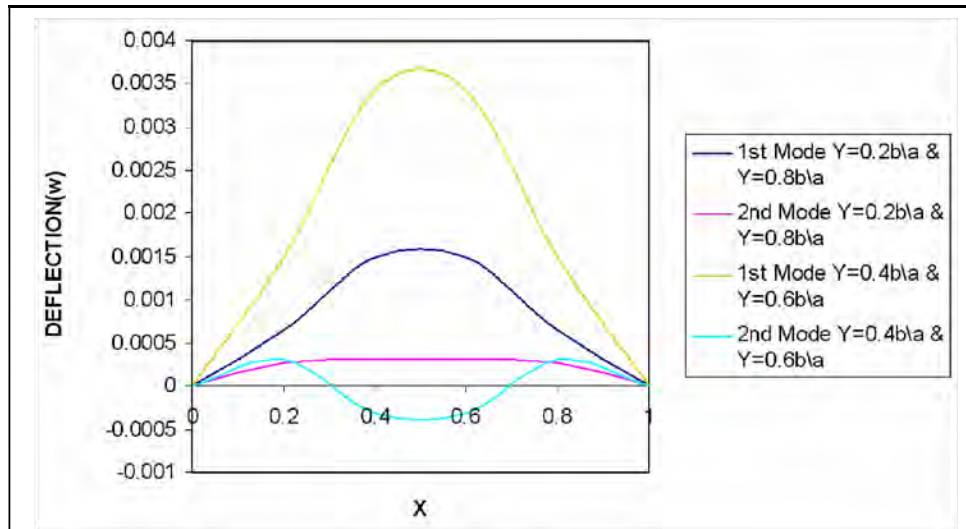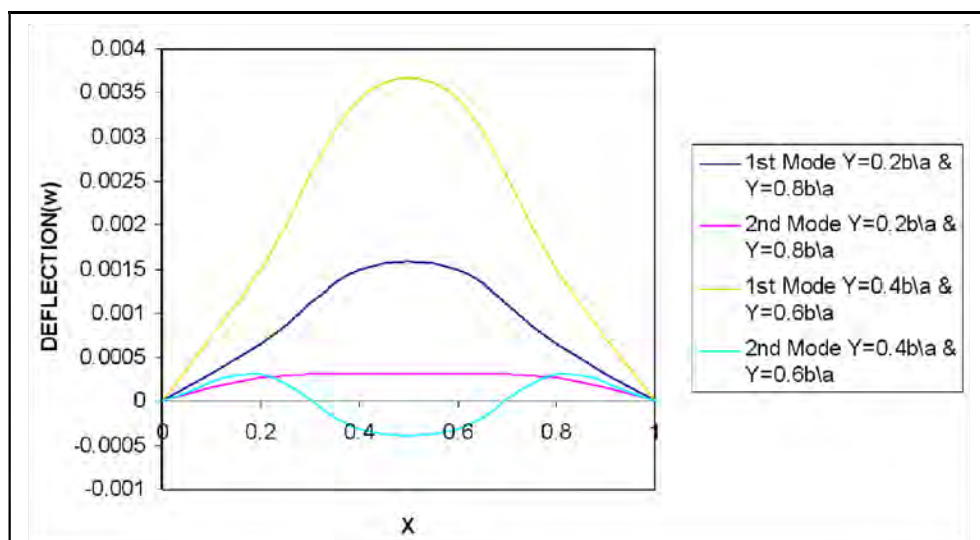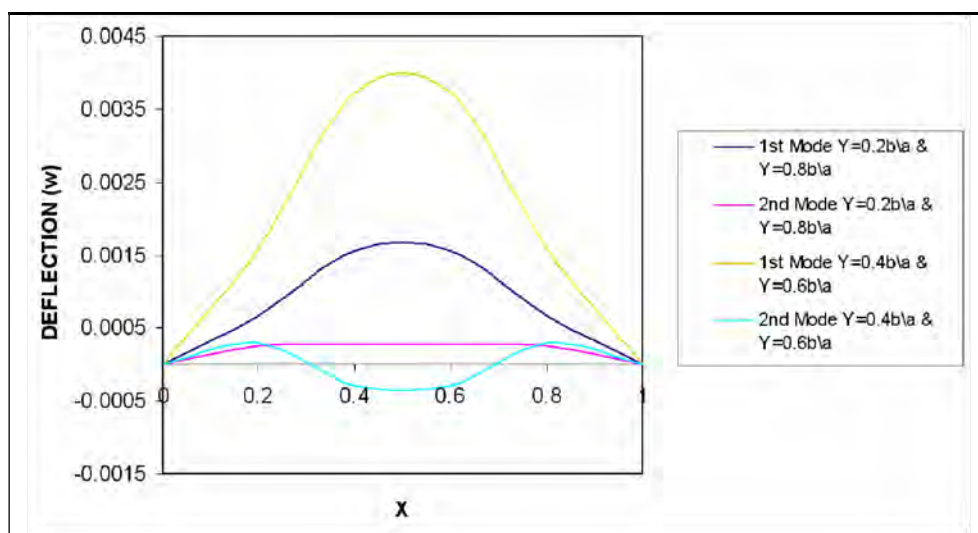

Figure 5. Transverse deflection w Vs X of visco-elastic non homogeneous rectangular plate of linearly varying thickness at initial time 0.K having constants combination as $\alpha = 0.0$, $\beta = 0.6$, $\alpha_1 = 0.0$, $\alpha_4 = 0.3$, $\alpha_5 = 0.2$.

5. Results and Discussions

The orthotropic material parameters have been taken as [3]

$$E_2^* / E_1^* = 0.32$$

$$E^* / E_1^* = 0.04$$

$$G_o / E_1^* = 0.09$$

$$\eta_o / G_o = 0.000069$$

$h = 0.01$ (plate thickness)

$\rho_o = 3 \times 10^5$ (mass density per unit volume of the plate material)

for calculating the values of this period K and deflect-tion w for a orthotropic visco-elastic rectangular plate for different values of taper constant β , thermal constant (α, α_4, α_5), non homogeneity constant α_1 and aspect ratio a/b at different points for first two modes of vibrations.

Figure 1 shows the result of time period K for different values of taper constant β and fixed thermal constant $\alpha = 0$ and aspect ratio a/b = 1.5 for two values of non – homogeneity constant α_1 are 0.0 and 0.4 for first two modes of vibration. It can be seen that time period (K) decreases when taper constant (β) increases for first two modes of vibration.

Figure 2 shows the result of time period K for first two modes of vibration for different values of non-ho-

AM

mogeneity constant α_1 and fixed aspect ratio a/b = 1.5 and four combinations of taper constant β and thermal constant α are

$$\beta = 0.0 , \alpha = 0.0$$
$$\beta = 0.0 , \alpha = 0.8$$
$$\beta = 0.6 , \alpha = 0.0$$
$$\beta = 0.6 , \alpha = 0.8$$

It can be seen that time period K increases when non-homogeneity constant increases for first two modes of vibration.

Figure 3 shows the result of time period K for different aspect ratio and four combinations of thermal con-

stant α, taper constant β and non-homogeneity constant α_1 *i.e.*

$$\alpha = 0.8 , \beta = 0.6, \alpha_1 = 0.0$$
$$\alpha = 0.8 , \beta = 0.6, \alpha_1 = 0.4$$
$$\alpha = 0.0 , \beta = 0.0, \alpha_1 = 0.0$$
$$\alpha = 0.8 , \beta = 0.0, \alpha_1 = 0.4$$

It can be seen that time period K decreases when aspect ratio increases for first two modes of vibration.

Figures 4-**7** show the result of deflection for first two modes of vibration for different X, Y and fixed aspect ratio a/b = 1.5 for initial time 0.K and 5.K for the following combination of thermal constants (α, α_4, α_5),

Figure 6. Transverse deflection w Vs X of visco-elastic non homogeneous rectangular plate of linearly varying thickness at time 5.K having constants combination as α = 0.8, β = 0.0, α_1 = 0.4, α_4 = 0.3, α_5 = 0.2.

Figure 7. Transverse deflection w Vs X of visco-elastic non homogeneous rectangular plate of linearly varying thickness at time 5.K having constants combination asα = 0.8, β = 0.6, α_1 = 0.0, α_4 = 0.3, α_5 = 0.2.

taper constant β and non-homogeneity constant α_1.

Results are compared with isotropic plate [12] and found to be in very close agreement.

6. References

[1] Z. Sobotka, "Free Vibration of Visco-Elastic Ortho-tropic Rectangular Plates," *Acta Technica*, Vol. 23, No. 6, 1978, pp. 678-705.

[2] A. K. Gupta and A. Khanna, "Vibration of Viscoelastic Rectangular Plate with Linearly Thickness Variations in Both Directions," *Journal of Sound and Vibration*, Vol. 301, No. 3-5, 2007, pp. 450-457.

[3] A. W. Leissa, "Vibration of Plate," NASA SP-60, 1969.

[4] S. R. Li and Y. H. Zhou, "Shooting Method for Non Linear Vibration and Thermal Buckling of Heated Orthotropic Circular Plates," *Journal of Sound and Vibration*, Vol. 248, No. 2, 2001, pp. 379-386.

[5] D. V. Bambill, C. A. Rossil, P. A. A. Laura and R. E. Rossi, "Transverse Vibrations of an Orthotropic Rectangular Plate of Linearly Varying Thickness and with a Free Edge," *Journal of Sound and Vibration*, Vol. 235, No. 3, 2000, pp. 530-538.

[6] J. S. Tomar and A. K. Gupta, "Effect of Thermal Gradient on Frequencies of Orthotropic Rectangular Plate Whose Thickness Varies in Two Directions," *Journal of Sound and Vibration*, Vol. 98, No. 2, 1985, pp. 257-262.

[7] J. S. Tomar and A. K. Gupta, "Thermal Effect on Frequencies of an Orthotropic Rectangular Plate of Linearly Varying Thickness," *Journal of Sound and Vibration*, Vol. 90, No. 3, 1983, pp. 325-331.

[8] J. S. Tomar and A. K. Gupta, "Effect of Exponential Temperature Variation on Frequencies of an Orthotropic Rectangular Plate of Exponentially Varying Thickness," *Proceeding of the Workshop on Computer Application in Continum Mechanics*, Roorkee, 11-13 March 1986, pp. 183-188.

[9] U. S. Gupta, R. Lal and S. Sharma, "Vibration Analysis of Non-Homogenous Circular Plate of Nonlinear Thickness Variation by Differential Quadrature Method," *Journal of Sound and Vibration*, Vol. 298, No. 4-5, 2006, pp. 892-906.

[10] A. K. Gupta, T. Johri and R. P. Vats, "Thermal Effect on Vibration of Non-Homogeneous Orthotropic Rectangular Plate Having Bi-Directional Parabolically Varying Thickness," *Proceeding of International Conference in World Congress on Engineering and Computer Science* 2007 (*WCECS* 2007), San Francisco, 24-26 October 2007, pp. 784-787.

[11] A. K. Gupta, A. Kumar and D. V. Gupta, "Vibration of Visco-Elastic Orthotropic Parallelogram Plate with Linearly Thickness Variation," *Proceeding of International Conference in World Congress on Engineering and Computer Science* 2007 (*WCECS* 2007), San Francisco, 24-26 October 2007, pp. 800-803.

[12] A. K. Gupta and L. Kumar, "Thermal Effect on Vibration of Non-Homogenous Visco-Elastic Rectangular Plate of Linear Varying Thickness," *Meccanica*, Vol. 43, No. 1, 2008, pp. 47-54.

[13] A. K. Gupta, A. Khanna and D. V. Gupta, "Free Vibration of Clamped Visco-Elastic Rectangular Plate Having Bi-Directional Exponentially Thickness Variations," *Journal of Theotrocial and Applied Mechanics*, Vol. 47, No. 2, 2009, pp. 457-471.

[14] A. K. Gupta, N. Aggarwal, D. V. Gupta, S. Kumar and P. Sharma, "Study of Non-Homogeneity on Free Vibration of Orthotropic Visco-Elastic Rectangular Plate of Parabolic Varying Thickness," *Advanced Studies of Theory Physics*, Vol. 4, No. 10, 2010, pp. 467-486.

[15] K. Bhasker and B. Kaushik, "Simple and Exact Series Solutions for Flexure of Orthotropic Rectangular Plates with any Combination of Clamped and Simply Supported Edges," *Composite Structure*, Vol. 63, No. 1, 2004, pp. 63-81.
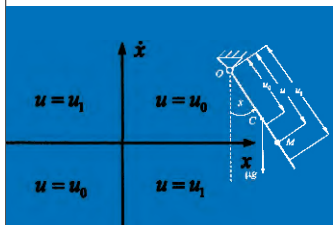
Call for Papers

Applied Mathematics (AM)

Applied Mathematics (*AM*) is an international journal dedicated to the latest advancement of applied mathematics. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas.

Subject Coverage

This journal invites original research and review papers that address the following issues. Topics of interest include, but are not limited to:

- Approximation Theory
- Chaos Theory
- Combinatorics
- Complexity Theory
- Computability Theory
- Control Theory
- Cryptography
- Discrete Geometry
- Dynamical Systems
- Financial Mathematics
- Game Theory
- Graph Theory
- Information Theory
- Mathematical Biology
- Mathematical Chemistry
- Mathematical Economics
- Mathematical Physics
- Mathematical Psychology
- Mathematical Sociology
- Numerical Analysis
- Operations Research
- Optimization
- Probability Distribution
- Probability Theory
- Statistics
- Stochastic Processes
- Theoretical Computer Science

We are also interested in short papers (letters) that clearly address a specific problem, and short survey or position papers that sketch the results or problems on a specific topic. Authors of selected short papers would be invited to write a regular paper on the same topic for future issues of *Applied Mathematics*.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

http://www.scirp.org/journal/am   E-mail: am@scirp.org

TABLE OF CONTENTS

Volume 1   Number 4   October 2010